



Universiteit
Leiden
The Netherlands

A Bayesian (meta-)regression model for treatment effects on the risk difference scale

Thomassen, D.; Steyerberg, E.; Cessie, S. le

Citation

Thomassen, D., Steyerberg, E., & Cessie, S. le. (2023). A Bayesian (meta-)regression model for treatment effects on the risk difference scale. *Statistics In Medicine*, 42(11), 1741-1759. doi:10.1002/sim.9697

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](#)

Downloaded from: <https://hdl.handle.net/1887/3720819>

Note: To cite this publication please use the final published version (if applicable).

A Bayesian (meta-)regression model for treatment effects on the risk difference scale

Doranne Thomassen¹  | Ewout Steyerberg¹  | Saskia le Cessie^{1,2} 

¹Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

²Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

Correspondence

Doranne Thomassen, Postzone S-05-S,
P.O. box 9600, 2300 RC Leiden.
Email: d.thomassen@lumc.nl

In clinical settings, the absolute risk reduction due to treatment that can be expected in a particular patient is of key interest. However, logistic regression, the default regression model for trials with a binary outcome, produces estimates of the effect of treatment measured as a difference in log odds. We explored options to estimate treatment effects directly as a difference in risk, specifically in the network meta-analysis setting. We propose a novel Bayesian (meta-)regression model for binary outcomes on the additive risk scale. The model allows treatment effects, covariate effects, interactions and variance parameters to be estimated directly on the linear scale of clinical interest. We compared effect estimates from this model to (1) a previously proposed additive risk model by Warn, Thompson and Spiegelhalter (“WTS model”) and (2) backtransforming the predictions from a logistic model to the natural scale after regression. The models were compared in a network meta-analysis of 20 hepatitis C trials, as well as in the analysis of simulated single trial settings. The resulting estimates diverged, in particular for small sample sizes or true risks close to 0% or 100%. Researchers should be aware that modelling untransformed risk can yield very different results from default logistic models. The treatment effect in participants with such extreme predicted risks weighed more heavily on the overall treatment effect estimate from our proposed model compared to the WTS model. In our network meta-analysis, this sensitivity of our proposed model was needed to detect all information in the data.

KEYWORDS

absolute risk modelling, Bayesian regression, network meta-analysis

1 | INTRODUCTION

Faced with a newly diagnosed patient, clinicians must consider which treatment will provide the largest risk reduction in their individual patient. To answer this question requires statistical methods that quantify treatment effects on a clinically relevant scale. Heterogeneity in treatment effects (HTE) is a scale-dependent concept. For binary outcomes, treatment effects may be quantified on several scales, including the odds ratio (OR), the absolute risk difference (RD) and the relative risk (RR) scales. If a treatment effect is assumed constant across patients on one scale, the effect generally varies on the other scales. The choice of effect measure impacts the relation of absolute treatment benefit to baseline risk. This

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Statistics in Medicine* published by John Wiley & Sons Ltd.

relation is often monotonic, but not necessarily so: in the case of effect modification, an increased effect on one scale may correspond to a decreased effect on another.¹

There are diverging views on the most appropriate scale of treatment effects in (meta-)analysis of binary outcomes. Arguments include mathematical properties, the presence of between-trial heterogeneity in treatment effects, clinical relevance, and ease of interpretation. There is typically less heterogeneity on ratio scales,²⁻⁴ making effects on these scales more generalisable and, in the case of meta-analysis, more consistent. Some argue that an effect measure should be chosen based purely on the best fit to the data, or least between-trial heterogeneity in meta-analysis.⁵ Others suggest that the choice of effect measure should be based on both the data and clinical considerations regarding the relation of treatment benefit to baseline risk.^{2,6} Recently, guidelines for modelling between-patient HTE in RCT analyses were published in the form of the PATH (Predictive Approaches to Treatment effect Heterogeneity) statement.^{7,8} The PATH authors recommend measuring benefits on the “natural” RD scale to guide clinical decision making. This is in line with the PRISMA-statement for reporting meta-analyses,⁹ which proposes the absolute risk scale as the most easily interpretable.

In this paper, we focus on (meta-)regression methods to obtain treatment effects on the absolute risk scale: risk differences. One approach is to use logistic regression and transform treatment effects in the form of log odds ratios to risk differences if an estimate of baseline risk is available. This was recently done in a Bayesian network meta-analysis (NMA),¹⁰ where both personalised risks and treatment effects were incorporated following the PATH risk modelling approach. The log odds scale has attractive mathematical properties, though it is not always suitable for situations with predicted risks very close to 0 or 1. This occurs, for example, in trial data where almost all patients experience the outcome in both treatment groups.

Recently, we performed a network meta-analysis where 100% success rates were an important feature of the data. In a logistic NMA model, the use of weakly informative priors was not sufficient to obtain convergence for all model parameters. In addition, all model parameters were defined on the logit scale, while the clinicians were only interested in the natural scale. This meant, for instance, that the between-study variance was not on the scale of interest to the clinicians. The same applied to any treatment-by-covariate interactions, which may influence the end results in different ways due to the noncollapsibility of the log odds. Since the original clinical focus of the NMA was on “personalised” treatment effects on the natural scale, we were motivated to explore options to model these directly.

This led to the development of an alternative model for benefit estimation directly on the natural risk scale. Any model that regresses on the natural risk scale must somehow keep the predicted risk within $[0, 1]$. Standard methods for the maximum likelihood (ML) estimation of binomial generalised linear (mixed) models (GL(M)M) with identity link have no such safety net and can fail to attain convergence. The Bayesian MCMC framework offers more flexibility to define bounds on the linear predictor at each iteration. Warn, Thompson and Spiegelhalter (WTS) bound the risk parameter domain using $\min()$ and $\max()$ functions in the meta-analysis of aggregated binary outcomes.¹¹ Their method has been applied in a few meta-analyses¹²⁻¹⁴ and is referenced in meta-analysis guidelines.¹⁵ It has been developed further for the analysis of cluster-randomised trials.¹⁶ However, the model runs into difficulty with empirical success percentages close to 0 or 100%. As a remedy, WTS propose a small continuity correction on the aggregated data. It is not immediately obvious how such a continuity correction may be applied to individual patient data (IPD) in HTE modelling.

We aimed to develop an improved absolute risk model that is suitable for IPD MA and NMA, and that can deal with extreme risk estimates. We hereto propose an alternative Bayesian (meta-)regression model for binary outcomes on the natural risk scale. The model has a true identity link function and it deals well with empirical success in (almost) all patients of a trial arm. Our proposed model, as well as the logistic and WTS model are defined formally in Section 2. We compare the models' resulting estimates in an IPD NMA of the example dataset that motivated this research (Section 3). To further investigate the differences between the models, we examine simulated scenarios (Section 4). We then discuss our results and give recommendations in (Section 5).

2 | MODELS TO OBTAIN ABSOLUTE RISK DIFFERENCES

2.1 | Data setting

We will define all models very generally, so that they can be applied to a network meta-analysis. Simpler forms can be applied in direct pairwise meta-analysis and to the analysis of single trials. Here, assume we have data from n_s studies, with a total of n_p participants. In total, n_t treatments are compared, though not all n_s studies necessarily have arms for all n_t treatments. For each study $j \in \{1, \dots, n_s\}$, $r_j \in \{1, \dots, n_t\}$ is defined as the study-specific reference treatment. For

each participant, we assume to have observed a binary outcome and possibly covariate values of interest. With y_{ijk} , we denote the outcome for participant $i \in \{1, \dots, n_p\}$, who participated in study $j \in \{1, \dots, n_s\}$ and was given treatment $k \in \{1, \dots, n_t\}$.

In our notation above and model definitions below, we assume that individual participant data (IPD) are available. The models are easily adjusted to aggregate data (AD) by replacing the individual binary outcome y_{ijk} with y_{jk} , the aggregate binomial outcome in a trial arm that received treatment k and was part of study j .

2.2 | NMA using generalised linear models

In all models below, y_{ijk} is assumed to be randomly sampled from a binomial distribution with size 1 and success probability p_{ijk} : $y_{ijk} \sim \text{Binomial}(1, p_{ijk})$, and $p_{ijk} = g^{-1}(\alpha_j + \delta_{j,r_j,k})$, where $\delta_{j,r_j,k} = 0$ for $k = r_j$. Here, g is the link function, which is the logit or the identity in this paper.

The regression equation for y_{ijk} contains study-specific regression parameters:

- α_j : the intercept of study j , representing baseline risk in models with identity link or baseline log odds in logistic models for study j under the reference treatment;
- $\delta_{j,r_j,k}$: the study-specific treatment effect of treatment k compared to the reference treatment r_j .

In the case of meta-analysis of multiple trials, study-specific treatment effects may be combined into overall effects in two ways. They can be assumed identical across studies and equal to the overall effect ($\delta_{j,r_j,k} = d_{r_j,k}$), which is often referred to as a common effects (CE) meta-analysis model. Alternatively, study-specific effects may be assumed to be an exchangeable sample from a common distribution, usually Gaussian. In this case, the effects are independent conditional on their mean and between-study variance: $\delta_{j,r_j,k} \sim \mathcal{N}(d_{r_j,k}, \tau^2)$. In frequentist terms, this type of model is usually called a random effects (RE) model. Both the overall mean and between-study variance are defined on the scale of the link function. When the number of studies per comparison is small, as in the NMA of our example, it is convenient to assume the between-study variance τ^2 to be constant across treatment comparisons. When there are studies with more than two arms, the correlation between study arms can be accounted for by specifying a multivariate normal distribution with non-zero covariances.¹⁵ The dependence structure of the study-specific intercepts α_j can be chosen freely. We chose to model the α_j as marginally independent across studies, so that the meta-analysis was fully stratified by study.

In the NMA, indirect comparisons of interest can be specified through consistency assumptions, again on the scale defined by the link function: $d_{r_j,k} = d_{1k} - d_{1r_j}$, $d_{11} = 0$. In what follows, we denote the scale of the overall treatment effect (risk difference or difference in log odds) with a superscript (RD or log, respectively).

2.3 | Obtaining risk difference estimates using logistic regression

Our objective is to estimate the risk difference that treatment k yields compared to the overall reference 1, denoted with RD_{1k} . The logistic model yields log odds ratios as treatment effects. Their transformation to the absolute scale requires ℓ_0 , the log odds (or logit risk) of the outcome for a patient who receives the reference treatment:

$$\text{RD}_{1k} = \frac{\exp[\ell_0 + d_{1k}^{\text{log}}]}{\exp[\ell_0 + d_{1k}^{\text{log}}] + 1} - \frac{\exp[\ell_0]}{\exp[\ell_0] + 1}. \quad (1)$$

The reference log odds ℓ_0 may be supplied to the model as a point estimate or in the form of a distribution. They may be estimated from internal data, for instance using the study-specific baseline log odds α_j . Alternatively, an estimate could be obtained from external data, such as observational studies. In our analysis of the Hepatitis C data, we used a point estimate based on all patients in the dataset that received the reference treatment. As an example, suppose we are analysing a single trial with two treatment arms. For each patient i , we record an outcome $y_i \sim \text{Binomial}(1, p_i)$ and only one covariate $x_i \in \{0, 1\}$ that indicates the treatment assigned to patient i . An translation of (1) into JAGS model code is given below for this situation. Here, we have chosen a weakly informative $\mathcal{N}(0, 5^2)$ prior on the

treatment effect on the log odds scale, and an uninformative $\mathcal{N}(0, 100^2)$ prior on the log odds of the outcome in the control group.

```

model {
  for(i in 1:length(y)) {
    y[i] ~ dbinom(p[i], 1)
    logit(p[i]) <- mu + logoddsdelta*x[i]
  }

  logit(pnull) <- mu
  logit(ptot) <- mu + logoddsdelta
  deltaRD <- ptot - pnull

  logoddsdelta ~ dnorm(0, 0.04) #variance=1/precision=1/0.04=25
  mu ~ dnorm(0, 0.0001)
}

```

2.4 | Modelling untransformed risk directly

In binomial models with an identity link, RD_{1k} is equal to the treatment effect d_{1k}^{RD} , so no additional transformations are necessary. An important difficulty in the estimation of the additive model, though, is that the probabilities p_{ijk} have unit interval support. The regression coefficients, by contrast, are not necessarily restricted to produce linear predictors within $[0, 1]$. This is not a problem with the logistic model, as the support of log odds comprises the entire real line. The Bayesian framework with MCMC estimation provides the flexibility to define additional constraints in the regression equations that keep the linear predictor in $[0, 1]$.

We will discuss two Bayesian implementations of a binomial regression model with identity link. Both implementations essentially aim to fit the same theoretical additive risk model. When writing out the likelihood of a binomial GLM with identity link, an indicator function $\mathbb{1}_{\{0 \leq \text{linear predictor} \leq 1\}}$ would make the bounds of the model parameters explicit in the expression. As such, the support for each model parameter depends on the data as well as on the other parameters. To fit the model, this indicator function needs to be translated into code. As the parameter bounds may vary for each observed patient, it is not generally possible to translate them into priors. Warn, Thompson and Spiegelhalter (WTS)^{11,15} translate the parameter bounds into code using min and max functions (Section 2.4.1). We found that placing a logit function on both sides of the regression equation (Section 2.4.2) also translates the parameter bounds, with slightly different properties than the WTS approach.

2.4.1 | The Warn, Thompson, and Spiegelhalter (WTS) model

One option is the model published by Warn, Thompson, and Spiegelhalter,^{11,15} which we adapted slightly to IPD context:

$$\begin{aligned}
 p_{ijk} &= \alpha_j^{RD} + \min \left[\max \left(\delta_{j,r,k}^{RD}, -\alpha_j^{RD} \right), \left(1 - \alpha_j^{RD} \right) \right], \\
 \delta_{j,r,k} &= 0 \quad \text{for } k = r_j.
 \end{aligned}
 \tag{2}$$

Equation (2) is akin to equation (4) in the WTS paper.¹¹ However, in equation (5) in WTS the analysis takes place on the treatment group level, and exactly two treatment arms are allowed per trial. Equation (2) above is a slightly generalised version with three index levels: i for individual patients, j for trials and k for treatments in the (N)MA. These indices allow for the application of the model to other settings than a two-treatment meta-analysis, such as NMA, IPD MA, and trial analysis with individual covariates. If a more elaborate linear predictor is of interest, its intercept may replace α_j^{RD} in (2), and $\delta_{j,r,k}^{RD}$ may be replaced by the rest of the linear predictor.

In Equation (2), the parameters α_j^{RD} represent the “baseline” probability of the outcome in the reference arm of each trial j . This probability is by definition in the unit interval, hence priors with unit interval support are appropriate for α_j^{RD} .

When such priors are used, Equation (2) reduces to $p_{ijk} = \alpha_j^{\text{RD}}$ for patients in the reference arm of trial j as follows:

$$\begin{aligned} p_{ijk} &= \alpha_j^{\text{RD}} + \min \left[\max \left(\delta_{j,rjk}^{\text{RD}}, -\alpha_j^{\text{RD}} \right), \left(1 - \alpha_j^{\text{RD}} \right) \right] \\ &= \alpha_j^{\text{RD}} + \min \left[\max \left(0, -\alpha_j^{\text{RD}} \right), \left(1 - \alpha_j^{\text{RD}} \right) \right] && \text{(since } \delta_{j,rjk}^{\text{RD}} = 0 \text{ for } k = r_j) \\ &= \alpha_j^{\text{RD}} + \min \left[0, \left(1 - \alpha_j^{\text{RD}} \right) \right] && \text{(since } \alpha_j^{\text{RD}} \geq 0 \text{ by its prior)} \\ &= \alpha_j^{\text{RD}} + 0 && \text{(since } \alpha_j^{\text{RD}} \leq 1 \text{ by its prior).} \\ &= \alpha_j^{\text{RD}} \end{aligned}$$

The parameters $\delta_{j,rjk}^{\text{RD}}$ in equation (2) represent the difference in risk of the outcome under treatment k versus the reference in trial j . As the total risk for each patient is in $[0, 1]$, the risk difference is bounded between $-\alpha_j^{\text{RD}}$ and $1 - \alpha_j^{\text{RD}}$. Generally, in an additive risk model, the bounds for each parameter depend on the data as well as the other parameters. WTS translate this constraint to BUGS code using min and max functions: when $\delta_{j,rjk}^{\text{RD}}$ is such that the sum $-\alpha_j^{\text{RD}} + \delta_{j,rjk}^{\text{RD}}$ would drop below 0 or exceed 1, the second term of the sum is set to $-\alpha_j^{\text{RD}}$ or $1 - \alpha_j^{\text{RD}}$, respectively, so that the sum equals 0 or 1, respectively.

As such, the WTS model does not constrain the linear predictor $\alpha_j^{\text{RD}} + \delta_{j,rjk}^{\text{RD}}$ per se; instead, it constrains p_{ijk} . The min and max functions set p_{ijk} to 0 or 1 when the linear predictor reaches below 0 or above 1. Hence, (2) allows treatment effect estimates that yield a linear predictor outside $[0, 1]$, though the predicted risks remain within $[0, 1]$. Neither p_{ijk} nor the linear predictor is transformed to a nonlinear scale, so treatment effects from this model have an interpretation as risk differences. However, because of the min and max functions, p_{ijk} and the linear predictor are not identical.

We illustrate (2) on the example trial of Section 2.3, where the data consists of a binary outcome y_i and a binary treatment indicator x_i for each patient i . JAGS model code for this simple case is provided below. As previously, we defined a weakly informative prior on the (risk difference) treatment effect and a uniform prior on the probability of the outcome in the control group.

```
model {
  for(i in 1:length(y)) {
    y[i] ~ dbinom(p[i], 1)
    p[i] <- alpha + min(max(lp[i], (0-alpha)), (1-alpha))
    lp[i] <- delta*x[i] #linear predictor without intercept
  }
  delta ~ dnorm(0, 1) T(-1, 1)
  alpha ~ dunif(0, 1)
}
```

When the WTS model is applied in a meta-analysis, the dependence structure of the within-study parameters within and between the studies may be defined in several ways, as explained in Section 2.2. In the examples in this paper, we assume α_j^{RD} to be independent across studies, whereas $\delta_{j,rjk}^{\text{RD}}$ are assumed exchangeable with constant between-study variance τ^2 .

2.4.2 | A proposed alternative

As an alternative to the WTS model, we also assume a linear risk model $p_{ijk} = \alpha_j^{\text{RD}} + \delta_{j,rjk}^{\text{RD}}$, where we propose using a logit transformation on both sides of the regression equation:

$$\log \left(\frac{p_{ijk}}{1 - p_{ijk}} \right) = \log \left(\frac{\alpha_j^{\text{RD}} + \delta_{j,rjk}^{\text{RD}}}{1 - [\alpha_j^{\text{RD}} + \delta_{j,rjk}^{\text{RD}}]} \right),$$

$$\delta_{j,rjk} = 0 \quad \text{for } k = r_j. \quad (3)$$

The logit on the right-hand side of the equation constrains the linear predictor $\alpha_j^{\text{RD}} + \delta_{j,r,k}^{\text{RD}} \mathbb{1}_{\{k \neq r_j\}}$ within $[0, 1]$. Theoretically, this likelihood is only defined for parameter combinations that result in linear predictor values in $[0, 1]$, as the logit of a number below 0 or above 1 is not on the real line. Therefore, no posterior mass is placed on parameter combinations resulting in probabilities outside $[0, 1]$. In our implementation, this resulted in such parameter combinations not being sampled by the Gibbs samplers JAGS¹⁷ and Stan,¹⁸ as implemented in R¹⁹ version 4.2.2. A transformation is also applied on the left-hand side of the equation. As the logit function is one-to-one, the transformation on both sides ensures an identity relation between p_{ijk} and the linear predictor.

We revisit the example of a two-arm trial with binary outcome y_i and a binary treatment indicator x_i . Approach (3) can be translated into JAGS code as follows, using the same priors as in the previous section.

```

model {
  for(i in 1:length(y)) {
    y[i] ~ dbinom(p[i], 1)
    logit(p[i]) <- log((alpha+lp[i]) / (1-(alpha+lp[i])))
    lp[i] <- delta*x[i] #linear predictor without intercept
  }
  delta ~ dnorm(0, 1) T(-1, 1)
  alpha ~ dunif(0, 1)
}

```

In the examples of (network) meta-analysis that follow, we assume α_j^{RD} to be independent and $\delta_{j,r,k}^{\text{RD}}$ to be exchangeable across studies with constant between-study variance τ^2 . Other assumptions are of course possible (Section 2.2).

2.5 | Priors and estimation

In the Bayesian context, the model is completed with prior distributions on the study-specific intercepts α_j , the overall effects $d_{1.}$, and the between-study SD τ . For the intercepts and overall effects, (N)MA guidelines have recommended using vague priors such as the $\mathcal{N}(0, 100^2)$ -distribution on log odds ratios.^{10,15} A downside of vague default priors is that they can introduce an unrealistic amount of variance. For instance, a $\mathcal{N}(0, 100^2)$ prior encodes the assumption that the treatment effect is almost as likely to lie between 20 and 30 as it is to lie between 0 and 10. Yet treatment effects on the log odds scale exceeding $[-5, 5]$ are quite extreme, corresponding to odds ratios exceeding $[0.007, 148]$. Moreover, risk difference treatment effects mathematically cannot exceed $[-1, 1]$. Prior distributions with smaller variances encode this information. In addition, informative priors can help regularise the model. We therefore opt for normal priors with mean 0 for all models, and SD 5 for log odds and SD 1 for risk differences. These priors are very weakly informative, but do not add an unrealistic amount of variance.

For each model, the posterior distribution of the model parameters can be estimated using MCMC, though there are some considerations to take into account. Firstly, the WTS model (2) and our proposed model (3) are nonstandard and require a large degree of flexibility in the model specification options of the MCMC method. In addition, some MCMC methods may be more appropriate to the shape of the parameter support and the correlation structure than others. In the additive models, the support of one parameter depends linearly on the data and on the other parameters, resulting in sharp edges of the likelihood. This can lead to problems with MCMC algorithms that use the shape of the posterior distribution to their advantage, such as the No U-Turn Sampler (NUTS) implemented in Stan.^{18,20} We have been able to estimate several versions (single trial, meta-analysis, network meta-analysis) of the model in the Gibbs sampler JAGS, and some in Stan as well. Although we have not done a formal comparison, we noticed that the Metropolis-within-Gibbs sampler in JAGS¹⁷ achieved convergence in all simulated scenarios of Section 4, while Stan did not. This may have to do with the “sharp edges” in the likelihood support of the additive risk models.

The convergence rate and autocorrelation of the chain will depend on the specific model and data structure at hand. Therefore, the parameters of the MCMC procedure (eg, the number of iterations, the number of burnin samples and the thinning rate) will need to be finetuned in each implementation to obtain decent effective sample sizes and to ensure convergence for the MCMC samples.

2.6 | Other applications

As special cases of NMA, the models described above can be applied to the analysis of a single trial as well as MAs with direct comparisons. Continuous and categorical covariates could be added, as we will show in some of our simulations in Section 4. Interaction terms could also be added in single trial analyses and (N)MA, though in (N)MA their interpretation warrants caution.²¹

3 | NMA OF HEPATITIS C DATA

Our alternative model was motivated by in an individual participant data (IPD) network meta-analysis (NMA) of the largest connected component of the TherapySelector (TS) database. The database contains data from 20 studies, comprising 5,842 hepatitis C (HCV) patients and 15 therapy regimens (Table A1). The primary outcome was Sustained Virologic Response (SVR, binary) and four categorical covariates were available for all patients. In the past few years, antiviral agents have entered the market that are almost 100% successful in repressing the hepatitis C virus. As a result, the HCV data contains several trial arms with SVR proportions close to or equal to 1. This aspect of the data played an important role in our modelling process.

The aim of the analysis was to estimate treatment effects with respect to overall reference treatment 1 (peginterferon- α plus ribavirin for 48 weeks), in the form of absolute risk differences, preferably conditional on patient covariates. To this end, we considered the three network meta-analysis approaches outlined above.

3.1 | Descriptives

The connectivity within the network was generally low (Figure 1). The therapies 8, 12, 14, and 15 were given to relatively small numbers of patients, potentially leading to larger uncertainty about these therapies in our analysis. There were several trial arms where (almost) all patients attained SVR during the study (Table A2). All analyses were complete case, intention-to-treat. The study samples were heterogeneous with respect to HCV genotype and proportion of patients with cirrhosis (Supplement, Table S1).

3.2 | Network meta-analysis

We applied the three models described in Section 2 in an NMA of the TS data. As in the single trial example described in Sections 2.3, 2.4.1 and 2.4.2, we used uninformative priors on the baseline log odds or risk, while weakly informative priors were defined for the treatment effects. All models were estimated using JAGS accessed from R version 4.0.2¹⁹ through R2jags.²² After some finetuning of the MCMC parameters, we ran two parallel chains of 50 000 iterations each, with a burn-in period of 5000 samples. A thinning factor of 20 was used, resulting in a total of 4500 saved iterations for each model. For both additive risk models, we found no signs of lack of MCMC convergence in the Gelman-Rubin statistics (all ≤ 1.01) nor in the traceplots. The Markov Chain resulting from our proposed model did suffer from more autocorrelation than the WTS model as reflected in lower effective sample sizes (minimum ESS 210 vs 1000). For the logistic model, three risk difference estimates had Gelman-Rubin statistics above 1.01 (RD4, RD8, RD13), raising concerns about their convergence, despite the weakly informative priors. All effective sample sizes were below 500, with the minimum ESS at 72 (RD5), suggesting high autocorrelation within the chains.

The posterior densities of the estimated risk differences for each treatment compared to the reference were summarised by their median and 95% central credible interval (Figure 2). Firstly, the credible intervals are generally wide, indicating large posterior uncertainty about the relative effectiveness of the treatments. Secondly, the estimated risk differences differ vastly depending on the model that was used. The between-model differences are most pronounced between the logistic model on the one hand and the two additive risk models on the other. Still, the two additive risk models disagree for almost half of the treatments.

The between-study SDs (τ) were estimated at 0.19 [0.01-0.82] (logistic model, τ on log(OR) scale), 0.03 [0.00-0.09] (model by Warn et al., τ on RD scale) and 0.05 [0.01-0.10] (our proposed model, τ on RD scale), respectively. This heterogeneity may contribute to the posterior uncertainty.

Therapy network for NMA

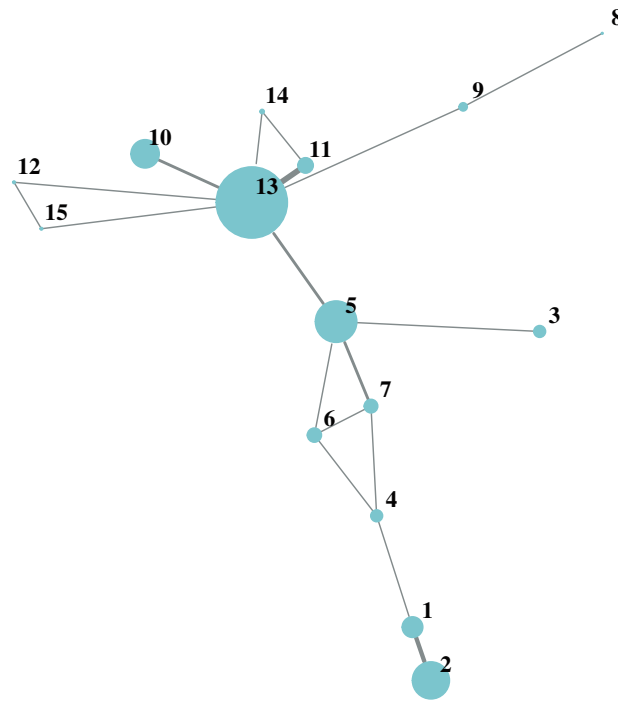


FIGURE 1 Graphical representation of the subnetwork of therapies that we analysed. The nodes in this graph each represent a unique therapy regimen. The relative size of the nodes indicates the total number of patients in the database who received this therapy regimen. The presence of an edge between two nodes indicates that these nodes have been compared directly in at least one study. The relative thickness of the edges indicates the number of studies comparing their respective endpoints.

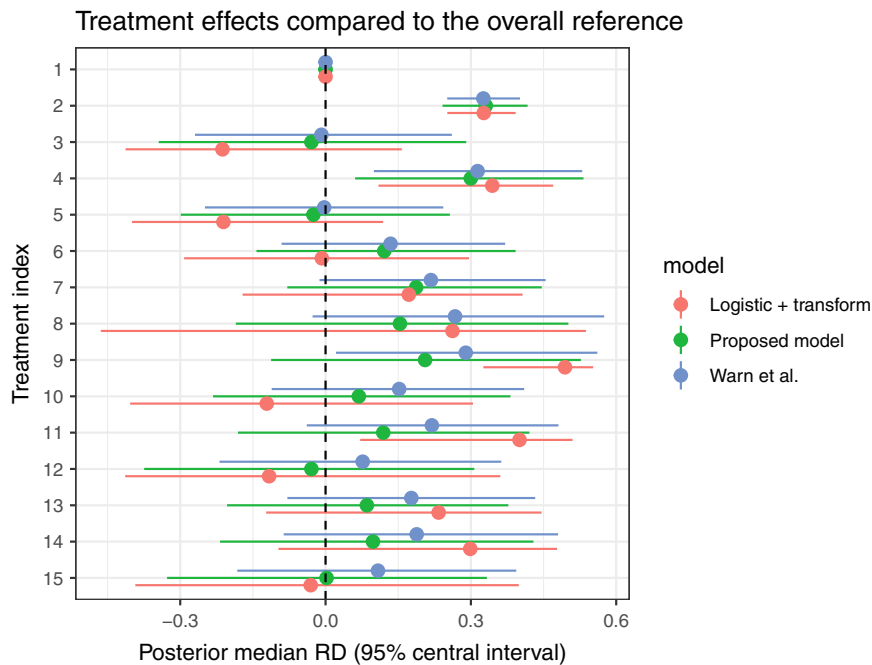


FIGURE 2 Graphical summaries of the posterior densities of the risk differences, after fitting the three models described in Section 2. The horizontal lines in the plots span a central 95% posterior interval and the dots represent posterior medians. As the outcome (sustained virologic response) is beneficial, higher RD values mean higher treatment effectiveness.

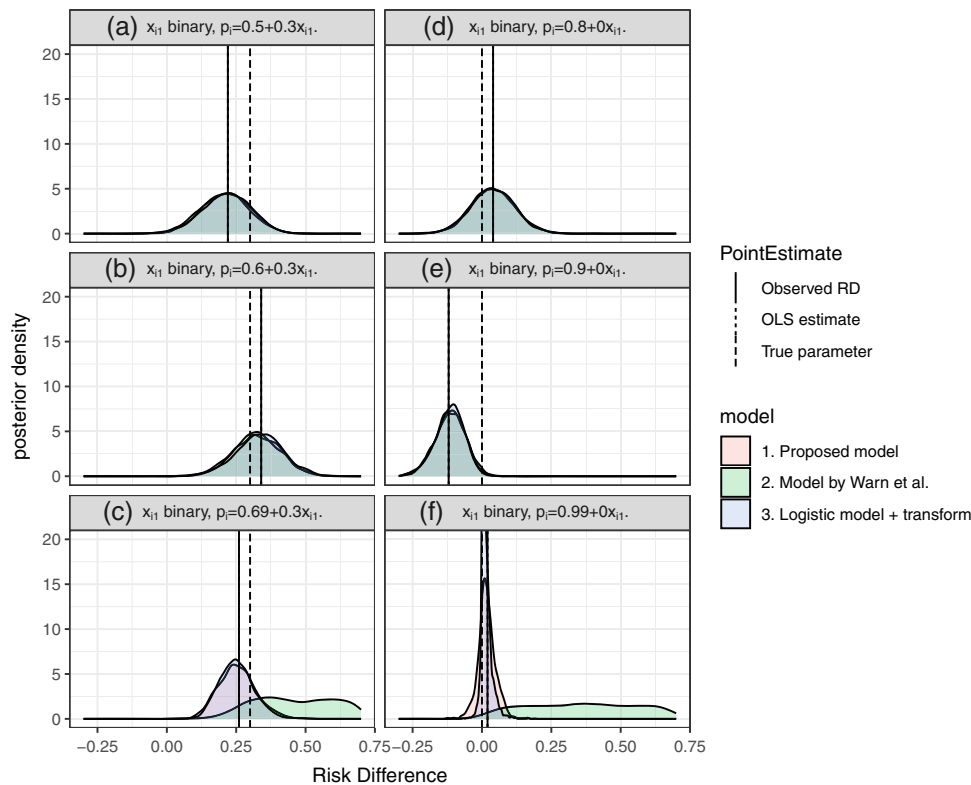


FIGURE 3 Estimated posterior density of the risk difference (RD) in simulated trials, produced by our proposed model, the logistic model and the WTS model. The sample size in each trial was 50 individuals per treatment arm, 100 in total. The lines of the OLS estimate and the observed risk difference overlap completely. Observed RD is defined as the observed success proportion in treatment group minus the observed success proportion in the control group. OLS, ordinary least squares.

The differences in between-study variance on the logit and natural scales of analysis, as well as the convergence problems of the logistic model may partly explain the variation in results across the models. To gain more insight into why and when the models obtain differing results, we simulated several scenarios, zooming in on less complicated settings in single trials and pairwise MA.

4 | SIMULATIONS

4.1 | Simulated scenarios

To understand the differences in behaviour of the three models on the TS data, we studied their results in simpler simulated settings. As our simulations were motivated by a network meta-analysis of trials, our simulated scenarios were inspired by the setup of a single trial, with x_{i1} denoting the binary treatment indicator. An important characteristic of the Hepatitis C data is the perfect success rate in some trial arms. We therefore chose our true parameters such that the true treatment effect remained mostly constant on the scale of the link function, and increased the intercept, moving the predicted risks p_i closer to 1 with every increase. Each simulation was performed with $n = 1000$, to limit the effect of sampling variability, as well as $n = 100$, to obtain an impression of the models' behavior in smaller samples.

We simulated data from the following GLM, assuming $i \in \{1, 2, \dots, n\}$:

$$y_i \sim \text{Binomial}(1, p_i), \quad (4)$$

$$p_i = g^{-1}(\alpha + \delta x_{i1} + \beta x_{i2}). \quad (5)$$

The link function g was defined as either the logit or the identity. In addition, we varied the number of independent variables, the true parameters α, δ, β , the total sample size n , and the distribution of x .

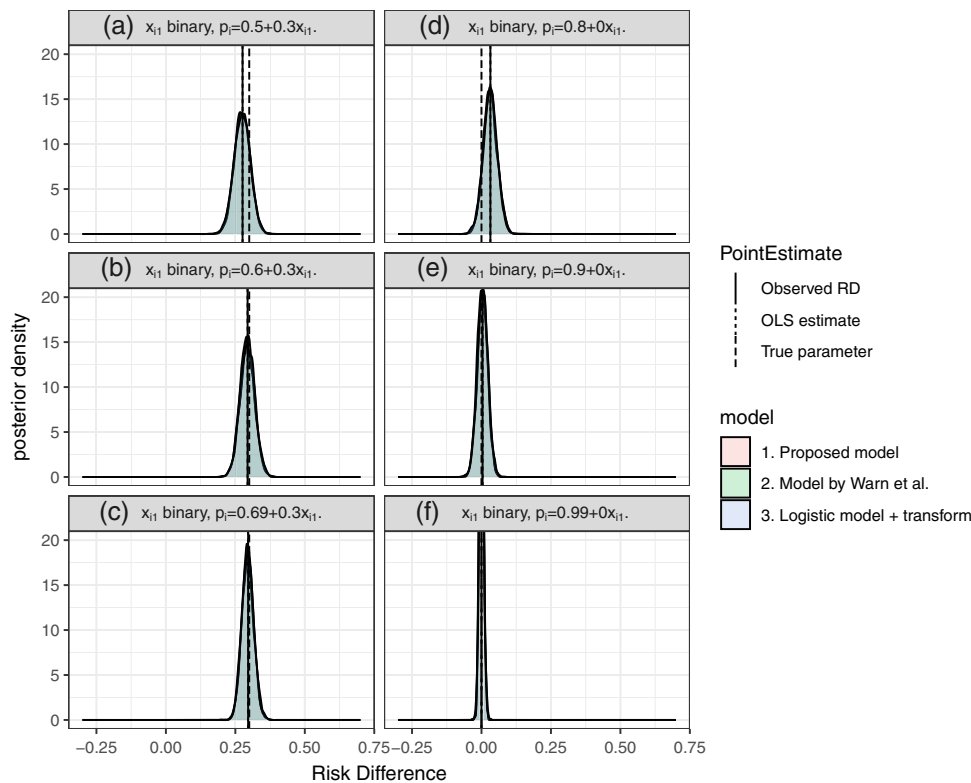


FIGURE 4 Estimated posterior density of the risk difference (RD) in simulated trials, produced by our proposed model, the logistic model and the WTS model. The sample size in each trial was 500 individuals per treatment arm, 1000 in total. The lines of the OLS estimate and the observed risk difference overlap completely. Observed RD is defined as the observed success proportion in treatment group minus the observed success proportion in the control group. OLS, ordinary least squares.

4.1.1 | One independent variable

We first examined scenarios with a single balanced treatment variable x_{i1} . In this case, the structure of the data is such that it can be fully captured in a crosstable, irrespective of the true link function. We therefore did not vary the true link and used the identity for ease of interpretation. The true risk difference between the treatment groups (δ) was first held fixed at 0.3, while varying the intercept $\alpha \in \{0.5, 0.6, 0.69\}$. In a second round of simulations, δ was held fixed at 0 while $\alpha \in \{0.8, 0.9, 0.99\}$ varied.

4.1.2 | Two independent variables

In all our simulations with two independent variables, we imitated the two-arm trial case with one balanced binary treatment variable and one covariate. The covariate x_{i2} was defined as $x_{i2} \sim \text{Bin}(n, 0.5)$ or $x_{i2} \sim \mathcal{U}[-1, 1]$, in both cases drawn independently of x_{i1} , mimicking treatment randomisation. For the binary and continuous x_{i2} , respectively, we simulated data where g was the identity and data where g was the logit. The following parameter combinations were specified:

- $x_{i2} \sim \text{Bin}(n, 0.5)$, g identity: $\delta = \beta = 0.3$, $\alpha \in \{0.2, 0.3, 0.39\}$;
- $x_{i2} \sim \text{Bin}(n, 0.5)$, g logit: $\delta = \beta = 1$, $\alpha \in \{-1, 0, 1\}$;
- $x_{i2} \sim \mathcal{U}[-1, 1]$, g identity: $\delta = 0.3$, $\beta = 0.2$, $\alpha \in \{0.3, 0.4, 0.49\}$;
- $x_{i2} \sim \mathcal{U}[-1, 1]$, g logit: $\delta = \beta = 1$, $\alpha \in \{-1, 0, 1\}$.

We chose the regression parameters to create a comparable range (or comparable proximity to the edges of $[0,1]$) of true risks between the two link functions. We keep the effect of the covariates (β , δ) constant and vary the intercepts to

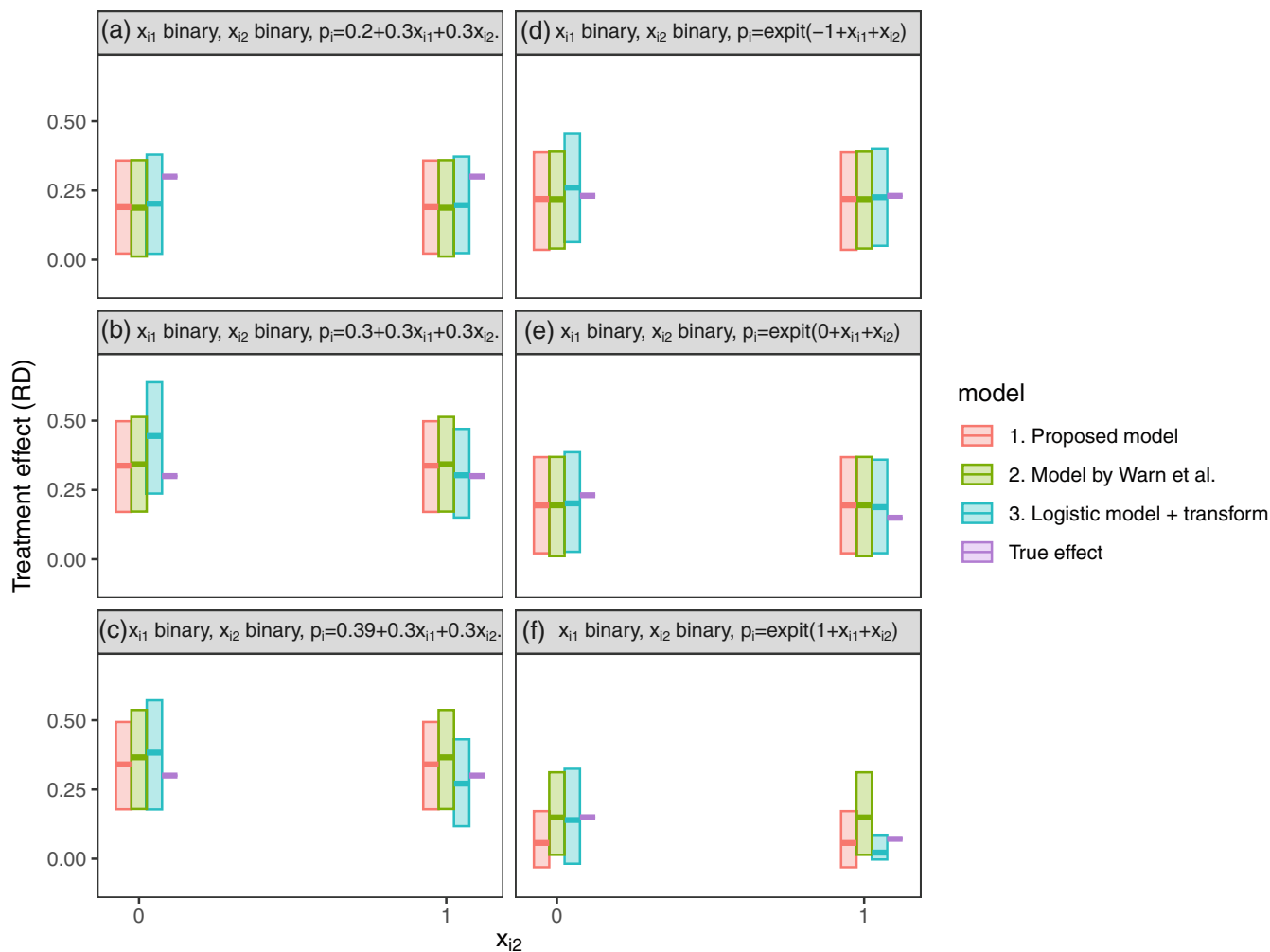


FIGURE 5 Estimated posterior mean and 95% central CrI of the risk difference (RD) in simulated trials across levels of a binary covariate x_{i2} . Results are shown for our proposed model, the logistic model and the WTS model. The sample size in each trial was 50 individuals per treatment arm, 100 in total.

move risks over the $[0,1]$ interval, from more in the middle towards the edges. Since the parameters are on different scales (logit vs natural) for the logistic and identity models, respectively, the parameter values that create this (comparable) range of true risks, differ, while the obtained risks are similar.

4.2 | Simulation results

In each simulated scenario, we fitted the three models described in section 2 using JAGS from R (4.2.2). For each model fit, we ran two Markov chains of 10 000 iterations each, with a burn-in period of 2000 iterations and a thinning factor of 8. Compared to the real data NMA, the simulation scenarios were relatively simple in terms of model complexity and data structure. Therefore, we started with a (comparatively) shorter burnin period, fewer iterations and the default thinning factor of R2jags. This yielded satisfactory MCMC diagnostics, so we did not change this. For the resulting Markov chains, we assessed convergence based on the Gelman-Rubin statistic. We considered a chain converged when the statistic was equal to or below 1.01. In addition, we registered the time that it took for the models to fit. For full RMarkdown reports of the simulations, including further model diagnostics, we refer the reader to our GitHub repository <https://github.com/DThomassen/Bayesian-RD-Regression>.

Our primary interest was in the estimated posterior distribution of the risk difference (RD) due to treatment, compared to the true risk difference. A “true” constant risk difference is only available in the scenarios where the true link function

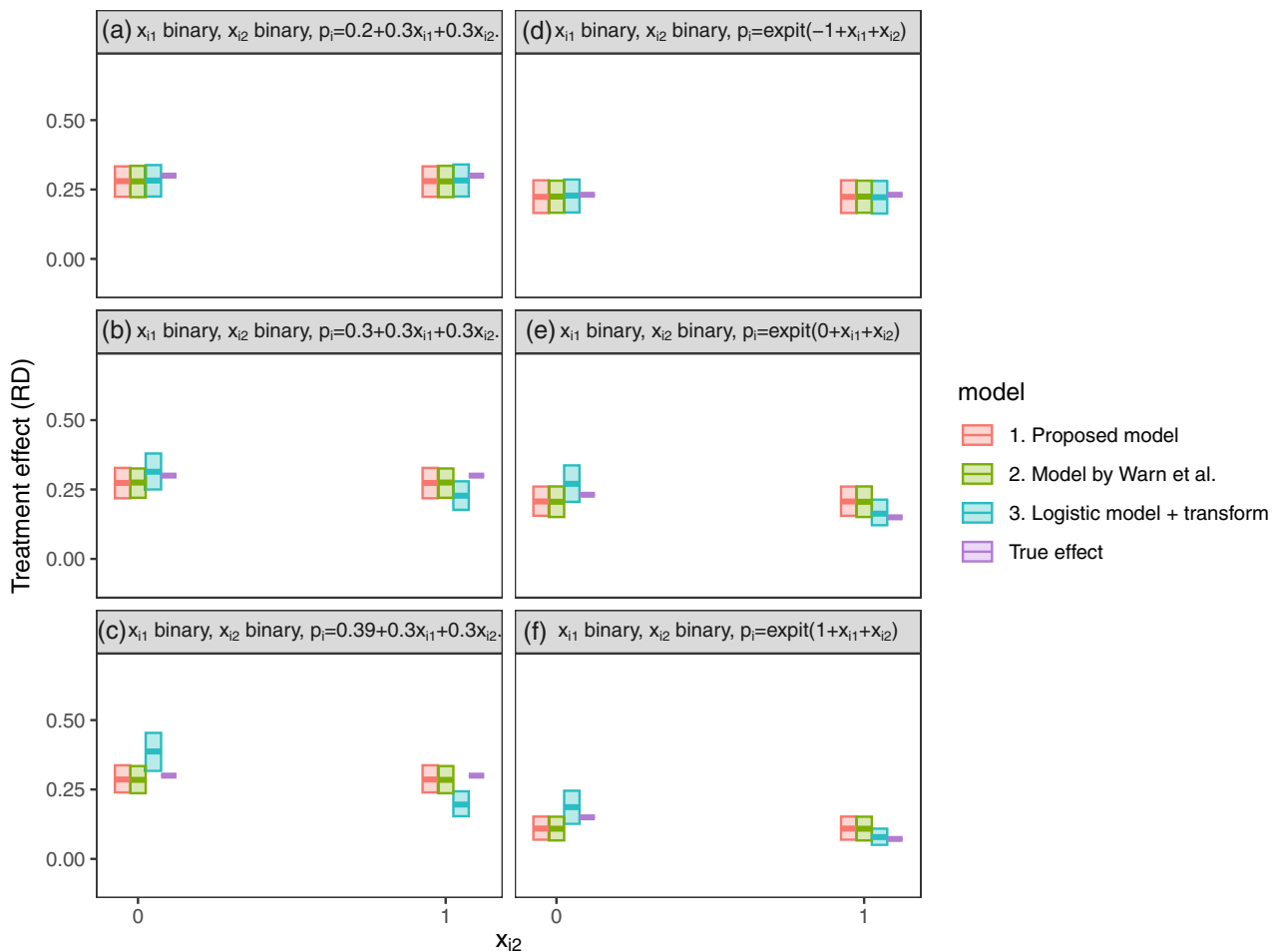


FIGURE 6 Estimated posterior mean and 95% central CrI of the risk difference (RD) in simulated trials across levels of a binary covariate x_{i2} . Results are shown for our proposed model, the logistic model and the WTS model. The sample size in each trial was 500 individuals per treatment arm, 1000 in total.

is the identity. Because we sampled each scenario only once, the data is subject to sampling variability with respect to the true risk difference. As additional comparators, we obtained the observed RD (ie, observed success proportion in treatment group minus observed success proportion in the control group) and the ordinary least squares (OLS) estimate of the risk difference.

4.2.1 | One binary independent variable

In the single binary treatment variable scenarios, all three models provided very similar results in most cases (Figures 3 and 4). All Gelman-Rubin statistics were well below 1.01. In the smaller sample scenarios (Figure 3), the sampling variability was larger, resulting in higher posterior variance and larger deviations of the observed RD from the true RD. In all binary treatment scenarios, the fitting time of the logistic models was roughly 150% of the fitting time of the additive risk models. The latter were very similar in fitting time.

A difference between the additive risk model results arose in cases where $n = 100$ and the true risk for one of the treatment groups was very close to one (Figure 3C,F). Here, the posterior RD resulting from the WTS model had large variance and most of its mass was above the observed risk difference. This results from the use of the `min()` and `max()` functions in the model. When the success rate in one treatment arm is 100%, any value of δ that exceeds the observed risk difference will result in a linear predictor larger than one. The `min()` function turns these linear predictors into predicted risks of 100% for that treatment arm, which is perfectly compatible with the data. For this reason, the Deviance Information Criterion (DIC) of the WTS model was similar to or even lower than the DIC of our proposed model in these scenarios. As a consequence, the WTS model can only learn from the data that the risk difference is likely equal to or

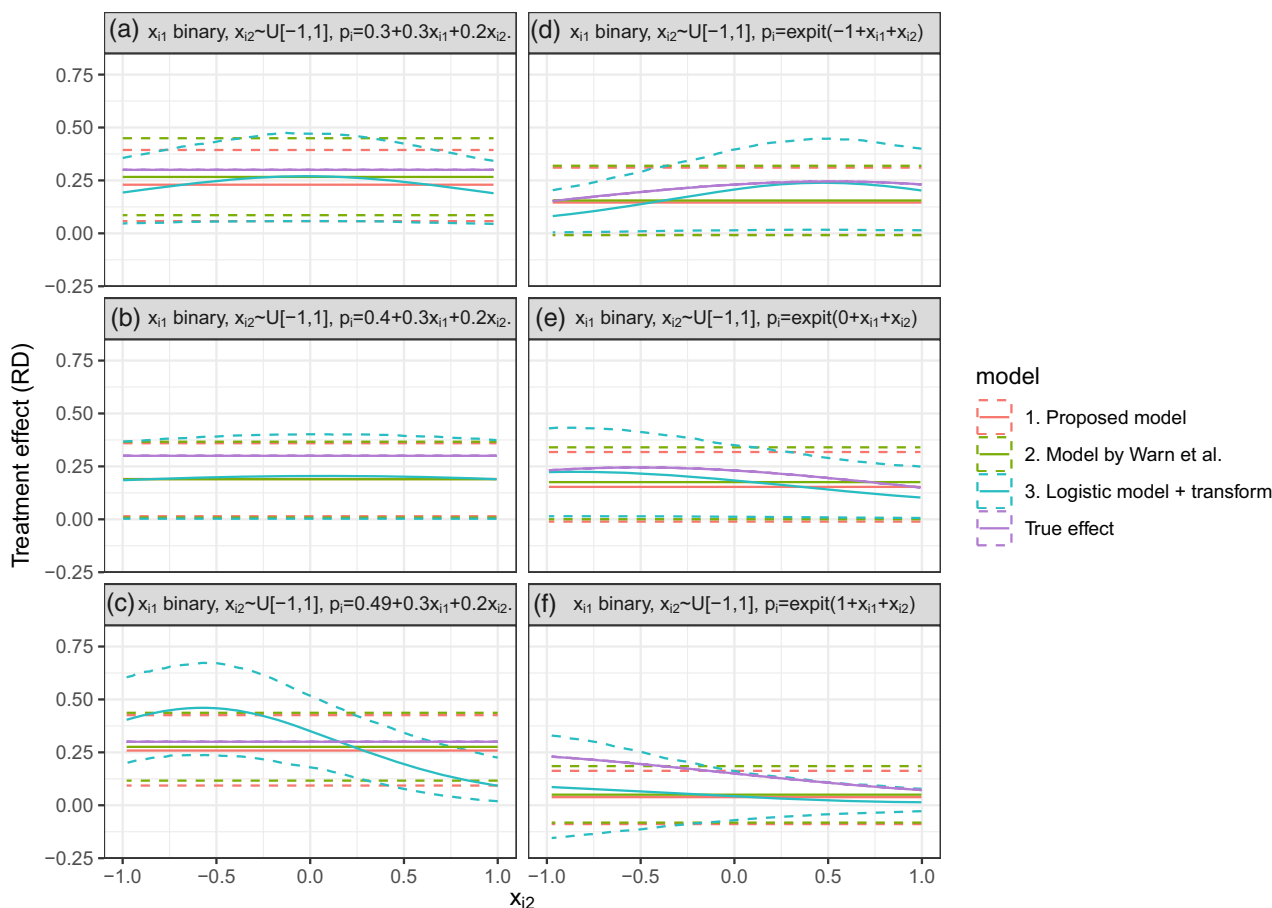


FIGURE 7 Estimated posterior mean and 95% central CrI of the risk difference (RD) in simulated trials across levels of a continuous covariate x_{i2} . Results are shown for our proposed model, the logistic model and the WTS model. The sample size in each trial was 50 individuals per treatment arm, 100 in total.

larger than the observed risk difference. Posterior mass is assigned to all risk difference values that exceed the observed risk difference, limited by the prior. The analogue holds for predicted risks of 0%.

In the scenario where $n = 100$ and the true risk was 99% in both trial arms (Figure 3F), the logistic model differed slightly from the additive risk models. The logistic model showed extreme effects on the log odds scale and a sharp peak in density around the observed risk difference after transformation to the RD scale. For the single trial RD point estimate, this is not necessarily a problem. But such extreme effects on the log odds scale can influence the overall results of MA or NMA, as the effects are often combined (weighed, averaged, added and subtracted) on the log odds scale. This may have played a role in the Hepatitis C NMA.

4.2.2 | Two independent variables

When we look at the results from the scenarios with one binary treatment variable and one additional independent variable (Figures 5–8), the main differences appeared between the logistic model on one side and the two additive models on the other. There was a home advantage: where the true link was the logit (Figure panels D–F), the logistic models were better able to capture the treatment-covariate-outcome relation and where the true link was the identity (Figure panels A–C), the natural risk models performed better.

This home advantage increased when variation of α moved the true risks to the more nonlinear part of the logistic function. For the scenarios with $x_{i2} \sim \mathcal{U}[-1, 1]$, this was less clear, because when α was lower, the minimum true risks got close to zero, whereas when α was higher, the maximum true risks approached one (Figures 7 and 8). So in most scenarios, the “home advantage” of the models was visible on either the lower or higher end of x_{i2} . When the true link is the logit, the linear effect of treatment (ie, the first derivative of g^{-1} (linear predictor) with respect to the treatment variable) becomes

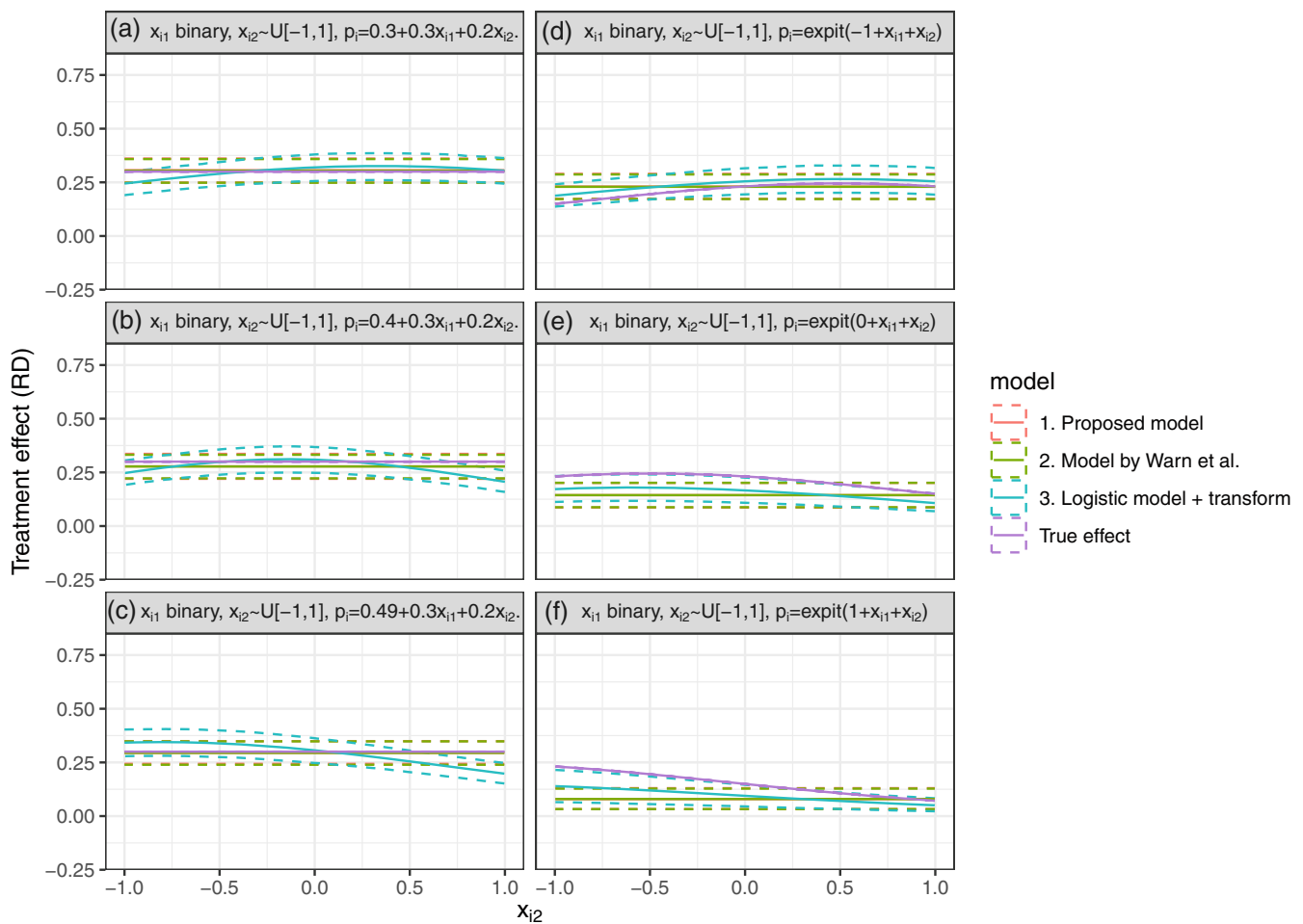


FIGURE 8 Estimated posterior mean and 95% central CrI of the risk difference (RD) in simulated trials across levels of a continuous covariate x_{i2} . Results are shown for our proposed model, the logistic model and the WTS model. The sample size in each trial was 500 individuals per treatment arm, 1000 in total.

smaller as the predicted risks approach 0 or 1 from 0.5. This home advantage was also visible in the DICs of the models, with differences in DIC ranging from 1 to 19 in favor of the true model as the predicted risks got closer to 0 or 1.

In the smaller samples ($n = 100$) contrasts also emerged between the additive risk models (Figures 5 and 7). Our proposed model generally resulted in slightly more conservative treatment effect estimates that were closer to the observed risk difference. Where the maximum true risk approached one, the WTS model yielded larger posterior variance and assigned posterior mass to RD values corresponding to linear predictors exceeding one (Figure 5C,F).

For binary x_{i2} , the two additive models were comparable in terms of fitting time, while the logistic model was roughly 15% slower. For uniform x_{i2} , the logistic model was the fastest, with the WTS model being about 30-60% slower, and our proposed model taking more than twice as long to fit.

4.3 | Link with the data example

In scenarios with 100% success in one trial arm, we noted that the WTS model put posterior mass on all beta values above the observed RD, within the constraints of the prior. Our model assigned mass somewhat symmetrically around the observed RD. If such a trial were part of a meta-analysis, the differences in posterior variance could impact the weight that is given to the trial. In the Hepatitis C data, there is an example of such a meta-analysis with one study that has a 100% successful arm. To compare treatment 13 to 5, two studies are available: ASTRAL-2 and ASTRAL-3. The observed risk differences in these studies are 0.06 and 0.17, respectively. ASTRAL-2 has one trial arm where 133/133 patients reach SVR.

	Logistic model	Additive risk model	
Assumptions about the data	<p>The relationship between the expected outcome and the treatment or other variables is additive on the logit scale</p> <p>Treatment-by-covariate interactions take place on the logit scale.</p> <p>In meta-analysis: the log odds ratio as a treatment effect is exchangeable between studies and between-study heterogeneity is modelled on the log odds ratio scale.</p>	<p>The relationship between the expected outcome and the treatment or other variables is additive on the natural scale</p> <p>Treatment-by-covariate interactions take place on the natural scale.</p> <p>In meta-analysis: the risk difference as a treatment effect is exchangeable between studies and between-study heterogeneity is modelled on the risk difference scale.</p>	
Research interest	<p>One is interested in/wants to detect treatment-by-covariate interactions on the logit scale</p> <p>In meta-analysis: one is interested in between-study heterogeneity on the logit scale</p>	<p>One is interested in/wants to detect treatment-by-covariate interactions on the natural scale</p> <p>In meta-analysis: one is interested in between-study heterogeneity on the natural scale</p>	
		<p>Our proposed approach</p> <p>The estimated treatment effect should be sensitive to patients with event probability close to 0 or 1 since they carry important information about the effect.</p>	<p>WTS approach - IPD</p> <p>The estimated treatment effect should not be restricted by 'extreme' patients with event probability close to 0 or 1.</p>
Technical considerations	<p>When predicted risks are close to 0 or 1, small effects in absolute risk may correspond to very large effects with small variances on the logit scale. In a meta-analysis, such trials may weigh heavily on the overall effect. Therefore, we do not recommend the use of the logistic model in meta-analyses where some trials have observed event rates of (almost) 0 or 100%.</p>	<p>When a 0% or 100% risk is observed in the data, the estimated posterior of the treatment effect will be skewed with a very large tail, leading to unstable point estimates. In this case, the use of the WTS approach is not advised.</p>	

FIGURE 9 Considerations in choosing between the three approaches discussed in this article.

When we analysed these trials individually, our proposed model as well as the logistic model resulted in a posterior mean close to 0.06 for ASTRAL-2, whereas the WTS model resulted in a flat posterior and a posterior mean around 0.4. The large posterior variance will lower the weight that is given to this trial in a meta-analysis. For ASTRAL-3, all three models give the same result.

When we look at the pairwise comparison within the network meta-analysis, we see that the three models yielded diverging summary estimates. Our model estimated the overall posterior RD around 0.12, so in between the estimates from the individual trials. The summary estimate from the WTS model was roughly 0.17, close to the estimate of the ASTRAL-3 trial. Apparently the weight of ASTRAL-2 in the MA was low. The summary estimate from the logistic model was around 0.22, exceeding the estimates from both individual trials, which is surprising. Factors such as heterogeneity and violations of the consistency assumption may be at play here. As the treatments 5 and 13 are very central in the network (Figure 1), these results may have seriously affected the other treatment comparisons in the NMA of Figure 2.

5 | DISCUSSION

We compared three methods to obtain estimates of risk differences in the (meta-)analysis of trials: a logistic model with transformation, a natural risk model by WTS and our proposed natural risk model. In the NMA of the Hepatitis C data, the models' results were quite different. In most simulated single trial scenarios, however, all four models provided similar results. Still, differences emerged when the true risks approached 0 or 100%. Here, the sigmoidal shape of the logistic function is most nonlinear and the 'home advantage' of the logistic versus the additive risk models is most pronounced. In some such cases, the logistic model became numerically unstable. This also occurred in our analysis of the Hepatitis C data.

In addition, our proposed model was more sensitive to what happened at the boundaries of the risk parameter support than the WTS model. This sensitivity can be both an advantage and a disadvantage. In the Hepatitis C data, there were several trials where much information about the treatment effect lay at these boundaries. The WTS model was not able to pick up this signal, whereas our proposed model was. In the case of a continuous variable with a few extreme values, the risk difference estimate for our model would be limited by these outliers: the linear predictor for the extreme data points should still lead to a linear predictor in the unit interval. The WTS model would be less affected by such outliers.

A limitation of our simulated scenarios is that they are by no means exhaustive. We primarily aimed to understand the diverging model results on the Hepatitis C data. Furthermore, we only explored scenarios where at least one of the models had the 'true' link function, while in reality, the structure of a dataset will almost never be exactly logistic or exactly linear. More systematic comparisons of these models are required to gain insight into their behaviour in a wider range of scenarios. To apply our proposed model in a different setting, we recommend examining simulated scenarios that are relevant to this setting.

We defined a first version of our proposed model in a trial and error process while working on a Bayesian NMA of the Hepatitis C data using JAGS. When developing it further, we decided to remain within this Bayesian MCMC framework, as it provided the flexibility to define our logit-logit model. We also tried fitting the model using Stan, but ran into problems with convergence. The angular shape of the likelihood likely plays a role here, as well as the specification of the priors. As the support of the likelihood for any additive risk model is data-dependent, it is not evident how to define a joint prior whose support perfectly fits the likelihood's support. We defined priors such that their support at least fully contains the likelihood support.

An often-cited downside of the Bayesian approach is the need for the specification of prior distributions, which introduces a degree of subjectivity. It is also possible to use the flexible MCMC framework to obtain maximum likelihood estimates. To this end, one can fit a "Bayesian" model with the desired likelihood function and a (possibly improper) uniform prior on the parameters. The posterior distribution is then proportional to the likelihood function and the posterior mode will be equal to the maximum likelihood estimate. In contrast to JAGS, Stan allows improper priors and may therefore be preferred for MCMC ML estimation.

On the other hand, the priors in a Bayesian model present a transparent way to restrict parameters and regularise the model. This is especially useful in the case of additive risk models, which can be computationally tricky. In our experience, regularising priors were necessary for convergence of the MCMC procedure in more complex settings like our network meta-analysis.

In our example NMA, we specified independent priors on the study-specific intercepts α_j , among α_j and δ_j , and among the overall treatment effects d . Conditionally on the data, these parameters are not independent. The assumption of marginal independence is commonly made, for example in Reference 15, but some have argued that it may not always be realistic.²³ To salvage this, a multivariate prior can be employed, or the model can be reparametrised such that the assumption of marginal independence becomes more likely to hold.^{24,25} The parametrisation of our proposed model could be optimised further, though this may come at the cost of the interpretability of the parameters.

The inclusion of covariates and interactions in the NMA model is necessary to obtain treatment effects conditional on patient covariates.¹⁰ Attention is needed on separating within-trial from between-trial interactions.²¹ Study-specific centering may be a pragmatic solution, which however complicates the interpretation of the summary treatment effect. Prediction of a treatment effect for a new patient that was not part of any study is an important direction for further investigation.

The compared models have their own properties and strengths. Which model is most suitable, depends on the specific modelling scenario. Considerations in choosing between the three approaches discussed above are summarised in Figure 9. Distinct criteria such as interpretability and best fit need not exclude each other. It is possible to model treatment effects on one scale for best fit and transform the results to another scale for interpretation. Mainly, a logistic model

and an additive risk model imply different assumptions about the data. These concern, for instance, the scale on which the treatment-outcome relation is assumed to be linear and the scale on which interactions are presumed to take place. In network meta-analysis, they also concern the scale on which consistency between the treatment effects across trials is assumed.

When one is interested in interactions or heterogeneity on the natural risk scale, an additive risk approach can be applied. As we have discussed, there are at least two viable approaches: the WTS approach and our proposed approach. The choice between the two additive risk models depends on how sensitive the model should be to what happens at the extremes of the probability parameter support. We suggest comparing both approaches in a sensitivity analysis. Our proposed model is preferable when datapoints with predicted risk in the outer regions of the 0%-100% range carry important information about the parameters to be estimated.

ACKNOWLEDGEMENTS

We thank TherapySelector for making the Hepatitis C data available in support of this study. In addition, we wish to thank prof. dr. Solko W. Schalm for introducing and collaborating on the network meta-analysis that motivated our research.

CONFLICT OF INTEREST STATEMENT

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

The simulation and analysis code that supports our results is openly available in figshare at <https://doi.org/10.6084/m9.figshare.22210885.v1>. The TherapySelector data that we analysed in our real data example is available in Table A2 of this paper.

ORCID

Doranne Thomassen  <https://orcid.org/0000-0002-9450-6015>

Ewout Steyerberg  <https://orcid.org/0000-0002-7787-0122>

Saskia le Cessie  <https://orcid.org/0000-0003-2154-4923>

REFERENCES

- Brumback B, Berg A. On effect-measure modification: relationships among changes in the relative risk, odds ratio, and risk difference. *Stat Med.* 2008;27(18):3453-3465. doi:10.1002/sim.3246
- Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med.* 2002;21(11):1575-1600. doi:10.1002/sim.1188
- Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Stat Med.* 2000;19(13):1707-1728. doi:10.1002/1097-0258(20000715)19:13<1707::AID-SIM491>3.0.CO;2-P
- Rhodes KM, Turner RM, Higgins JPT. Empirical evidence about inconsistency among studies in a pair-wise meta-analysis. *Res Synth Methods.* 2016;7(4):346-370. doi:10.1002/jrsm.1193
- Walter S. Choice of effect measure for epidemiological data. *J Clin Epidemiol.* 2000;53(9):931-939. doi:10.1016/S0895-4356(00)00210-9
- Caldwell DM, Welton NJ, Dias S, Ades A. Selecting the best scale for measuring treatment effect in a network meta-analysis: a case study in childhood nocturnal enuresis. *Res Synth Methods.* 2012;3(2):126-141. doi:10.1002/jrsm.1040
- Kent DM, Paulus JK, Van Klaveren D, et al. The predictive approaches to treatment effect heterogeneity (PATH) statement. *Ann Intern Med.* 2020;172(1):35-45.
- Kent DM, Van Klaveren D, Paulus JK, et al. The predictive approaches to treatment effect heterogeneity (PATH) statement: explanation and elaboration. *Ann Intern Med.* 2020;172(1):W1-W25.
- The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med.* 2009;151(4):W-65-W-94. doi:10.7326/0003-4819-151-4-200908180-00136
- Chalkou K, Steyerberg E, Egger M, Manca A, Pellegrini F, Salanti G. A two-stage prediction model for heterogeneous effects of treatments. *Stat Med.* 2021;40(20):4362-4375.
- Warn DE, Thompson SG, Spiegelhalter DJ. Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. *Stat Med.* 2002;21(11):1601-1623.
- Thompson S, Turner R, Warn D. Multilevel models for meta-analysis, and their application to absolute risk differences. *Stat Methods Med Res.* 2001;10(6):375-392.
- Sutton AJ, Abrams KR, Jones DR. Generalized synthesis of evidence and the threat of dissemination bias: the example of electronic fetal heart rate monitoring (EFM). *J Clin Epidemiol.* 2002;55(10):1013-1024. doi:10.1016/S0895-4356(02)00460-2
- Moran JL, Graham PL. Risk related therapy in meta-analyses of critical care interventions: Bayesian meta-regression analysis. *J Crit Care.* 2019;53:114-119. doi:10.1016/j.jcrc.2019.06.003

15. Dias S, Welton NJ, Sutton AJ, Ades A. *NICE DSU Technical Support Document 2: a Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials*. Sheffield, United Kingdom: Decision Support Unit, University of Sheffield; 2011.
16. Thompson SG, Warn DE, Turner RM. Bayesian methods for analysis of binary outcome data in cluster randomized trials on the absolute risk scale. *Stat Med*. 2004;23(3):389-410.
17. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*; 2003:1-10.
18. Stan Development Team. RStan: the R interface to Stan. 2020; R package version 2.21.2.
19. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2020.
20. Carpenter B, Gelman A, Hoffman MD, et al. Stan: a probabilistic programming language. *J Stat Softw*. 2017;76(1):1-32.
21. Riley RD, Debray TP, Fisher D, et al. Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: statistical recommendations for conduct and planning. *Stat Med*. 2020;39(15):2115-2137.
22. Su YS, Yajima M. *R2jags: using R to run 'JAGS'* 2015; R package version 0.5-7.
23. Parameterization GA, Modeling B, Gelman A. Parameterization and bayesian modeling. *J Am Stat Assoc*. 2004;99(466):537-545.
24. Gelman A, Bois F, Jiang J. Physiological pharmacokinetic analysis using population Modeling and informative prior distributions. *J Am Stat Assoc*. 1996;91(436):1400-1412.
25. Gelman A, Jakulin A, Pittau MG, Su YS. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat*. 2008;2(4):1360-1383. doi:10.1214/08-AOAS191

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Thomassen D, Steyerberg E, le Cessie S. A Bayesian (meta-)regression model for treatment effects on the risk difference scale. *Statistics in Medicine*. 2023;42(11):1741-1759. doi: 10.1002/sim.9697

APPENDIX . TABLES

TABLE A1 Patient numbers receiving each of the therapy regimens, split by cirrhosis status (Child score), HCV genotype and comorbidities, respectively.

#	Therapy regimen	No. of patients
1	PegIFN-Riba 48wk	422
2	PegIFN-Riba-SMPV 24wk	758
3	PegIFN-Riba 24wk	243
4	SOF-Riba-PegIFN 12wk	244
5	SOF-Riba 12wk	847
6	SOF-Riba 16wk	291
7	SOF-Riba 24wk	279
8	SOF-VELP-VOX-Riba 12wk	37
9	SOF-VELP-VOX 12wk	170
10	SOF-VELP-VOX 8wk	579
11	SOF-VELP-Riba 12wk	317
12	SOF-VELP-Riba 8wk	56
13	SOF-VELP 12wk	1454
14	SOF-VELP 24wk	90
15	SOF-VELP 8wk	55

Abbreviations: PegIFN, Peginterferon- α ; Riba, Ribavirin; SOF, Sofosbuvir; VELD, Velpatasvir; VOX, Voxilaprevir.

TABLE A2 Pairwise analysis results and study sample characteristics for all direct comparisons in the NMA network.

Study	n_1^a	SVR ₁ ^a	n_2^a	SVR ₂ ^a	RD	95%CI		log(OR)	95%CI		
1 vs 2											
Forns Gastroenterology 2014 Promise	132	53	250	198	-0.39	-0.49	-0.29	-1.74	-2.20	-1.27	
Jacobson Lancet 2014 QUEST-1	130	65	260	206	-0.29	-0.39	-0.19	-1.34	-1.79	-0.88	
Manns Lancet 2014 QUEST-2	134	67	248	200	-0.31	-0.40	-0.21	-1.43	-1.89	-0.96	
1 vs 4											
Lawitz Lancet Inf Dis 2013 PROTON	26	15	47	42	-0.32	-0.53	-0.11	-1.82	-3.03	-0.61	
3 vs 5											
Lawitz N Eng J Med 2013 FISSION	243	162	253	170	-0.01	-0.09	0.08	-0.02	-0.40	0.35	
4 vs 6											
Foster Gastroenterology 2015 BOSON	197	183	196	141	0.21	0.14	0.28	1.63	1.00	2.26	
4 vs 7											
Foster Gastroenterology 2015 BOSON	197	183	199	170	0.07	0.01	0.14	0.80	0.13	1.47	
5 vs 6											
Jacobson N Eng J Med 2013 FUSION	100	50	95	69	-0.23	-0.36	-0.09	-0.98	-1.57	-0.38	
5 vs 7											
Doss J Hepatol 2015	52	40	51	46	-0.13	-0.27	0.01	-1.02	-2.14	0.11	
Ruane J Hepatol 2015	31	21	29	27	-0.25	-0.44	-0.06	-1.86	-3.48	-0.24	
5 vs 13											
Foster N Eng J Med 2015 ASTRAL-2	131	123	133	133	-0.06	-0.10	-0.02	-2.91	-5.77	-0.05	
Foster N Eng J Med 2015 ASTRAL-3	280	218	277	264	-0.17	-0.23	-0.12	-1.75	-2.38	-1.13	
6 vs 7											
Foster Gastroenterology 2015 BOSON	196	141	199	170	-0.13	-0.21	-0.06	-0.83	-1.33	-0.32	
8 vs 9											
Lawitz Hepatology 2017 June	37	36	33	33	-0.03	-0.10	0.05	-1.01	-4.25	2.22	
9 vs 13											
Bourliere N Eng J Med 2017 POLARIS-4	137	136	125	110	0.11	0.05	0.17	2.92	0.88	4.96	
10 vs 13											
Jacobson Gastroenterology 2017 POLARIS-2	473	451	418	415	-0.04	-0.06	-0.02	-1.91	-3.12	-0.70	
Jacobson Gastroenterology 2017 POLARIS-3	106	104	106	104	0.00	-0.04	0.04	0.00	-1.98	1.98	
11 vs 13											
Curry N Eng J Med 2015 ASTRAL-4	87	82	90	75	0.11	0.02	0.20	1.19	0.13	2.25	
Esteban Gastroenterology 2018	101	99	98	92	0.04	-0.01	0.10	1.17	-0.45	2.80	
Pianko Ann Int Med 2015	160	78	80	77	0.01	-0.04	0.07	0.42	-1.40	2.24	
Takehara J Gastroenterol 2018	49	47	51	47	0.04	-0.05	0.13	0.69	-1.05	2.44	
11 vs 14											
Curry N Eng J Med 2015 ASTRAL-4	87	82	90	79	0.06	-0.02	0.15	0.83	-0.28	1.93	
12 vs 13											
Everson Ann Int Med 2015	56	48	76	73	-0.10	-0.20	0.00	-1.40	-2.78	-0.02	
12 vs 15											
Everson Ann Int Med 2015	56	48	55	49	-0.03	-0.16	0.09	-0.31	-1.44	0.82	
13 vs 14											
Curry N Eng J Med 2015 ASTRAL-4	90	75	90	79	-0.04	-0.15	0.06	-0.36	-1.20	0.48	
13 vs 15											
Everson Ann Int Med 2015	76	73	55	49	0.07	-0.02	0.16	1.09	-0.34	2.52	

Note: For each corresponding trial arm, two numbers are given: the total number of patients in the arm and the number them who attained SVR. Some trials had more than two arms, hence they appear in more than one pairwise comparison.

Abbreviations: PegIFN, Peginterferon- α ; Riba, Ribavirin; SOF, Sofosbuvir; SVR, sustained virologic response, a binary measure of successful treatment; VELP, Velpatasvir; VOX, Voxilaprevir.

^aThis example shows our analysis of the TherapySelector database as it was in November 2019. Subsequent changes and corrections to the database may have occurred since. Numbers may differ slightly from numbers in the original papers because of different selection criteria in the TherapySelector.