Received: 21 August 2018 | Revised: 19 February 2019 | Accepted: 14 March 2019

DOI: 10.1002/sim.8158

RESEARCH ARTICLE

WILEY Statistics in Medicine

Network meta-analysis of rare events using the Mantel-Haenszel method

Orestis Efthimiou[1] | Gerta Rücker[2] | Guido Schwarzer[2] | Julian P.T. Higgins[3] | Matthias Egger[1] | Georgia Salanti[1]

[1]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

[2]Institute of Medical Biometry and Statistics, Medical Faculty and Medical Center, University of Freiburg, Freiburg, Germany

[3]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

Correspondence
Orestis Efthimiou, Institute of Social and Preventive Medicine, University of Bern, 3012 Bern, Switzerland.
Email: oremiou@gmail.com

Funding information
Swiss National Science Foundation (SNSF), Grant/Award Number: 166656 and 174281

The Mantel-Haenszel (MH) method has been used for decades to synthesize data obtained from studies that compare two interventions with respect to a binary outcome. It has been shown to perform better than the inverse-variance method or Peto's odds ratio when data is sparse. Network meta-analysis (NMA) is increasingly used to compare the safety of medical interventions, synthesizing, eg, data on mortality or serious adverse events. In this setting, sparse data occur often and yet there is to-date, no extension of the MH method for the case of NMA. In this paper, we fill this gap by presenting a MH-NMA method for odds ratios. Similarly to the pairwise MH method, we assume common treatment effects. We implement our approach in R, and we provide freely available easy-to-use routines. We illustrate our approach using data from two previously published networks. We compare our results to those obtained from three other approaches to NMA, namely, NMA with noncentral hypergeometric likelihood, an inverse-variance NMA, and a Bayesian NMA with a binomial likelihood. We also perform simulations to assess the performance of our method and compare it with alternative methods. We conclude that our MH-NMA method offers a reliable approach to the NMA of binary outcomes, especially in the case or sparse data, and when the assumption of methodological and clinical homogeneity is justifiable.

KEYWORDS

adverse events, mixed treatment comparison, multiple treatments meta-analysis, rare events, rare outcomes

1 | INTRODUCTION

Meta-analysing studies with rare binary outcomes can be methodologically challenging, especially when some of the studies report zero events in one or both treatment arms. The inverse-variance (IV) method, assuming common effects ("fixed-effects") or random effects,[1] is the most widely used approach to pairwise meta-analysis.[2] The method requires estimates of a relative treatment effect from each study along with a standard error. For the case of binary outcomes, the approach employs approximations that do not perform well when event rates are low and/or the sample sizes of the included studies are small. If studies report zero events in one of their arms, the basic formulae cannot be used to estimate odds ratios, risk ratios, and their standard errors because the calculations involve division by zero. A simple way to overcome this problem is to add a fixed number (eg, 0.5) to the number of events and nonevents of all treatment

arms in studies that report zero events in one of their arms; studies with zero events in both arms are usually excluded from the analysis. This so-called "continuity correction" bypasses the problem of zero events, and allows the use of the IV method. However, it has been shown that this approach leads to bias in estimated effects.[3] Sweeting et al suggested a flexible approach, where the continuity correction is adapted to each study, and showed that such corrections performed better than fixed ones.[4] However, the use of any type of continuity correction has been criticized because data are imputed and because the, ie, essentially arbitrary, choice of the correction may bias results.[5,6]

Another approach to address this problem is to use the risk difference instead of odds or risk ratios. The risk difference can readily be estimated in the presence of zero events. Unfortunately, as shown in simulations by Bradburn et al,[3] *"all risk difference methods yield very conservative confidence interval coverage when events are rare, and have associated poor statistical power, which make them unsuitable for meta-analysis of rare events."* Alternative methods have been proposed, including the use of the arcsine difference,[7] an exact method combining confidence intervals,[8] beta-binomial models,[6] etc; for a short review, see a recent paper by Efthimiou.[9]

The Mantel-Haenszel (MH) method is a popular approach to meta-analysing binary outcomes.[10] It has been formulated for the case of odds ratios, risk ratios and risk differences. Estimating MH odds-ratios does not require a continuity correction for the case of studies with zero events in one of their arms, unless all studies in the dataset have zero events for the same treatment. Meta-analysis using the MH odds-ratio has been shown to outperform the IV method when events are rare.[3] Note that, using the MH method, we exclude studies with zero events in all treatment arms. This has been criticised, because such studies may carry information through their sample size.[6] Another popular approach to meta-analysing rare binary outcomes is Peto's odds ratio.[11] This method has been shown to work well when specific conditions are met (event rates are less than 1%, treatment groups are balanced, and relative effects are not very large).[3,4]

Network meta-analysis (NMA) is an extension of pairwise meta-analysis for the case when studies compare multiple treatments.[12-16] The frequentist approaches that are usually employed for fitting a NMA follow the IV method. Thus, they are expected to perform poorly when event rates are low. Bayesian approaches to NMA that utilize the exact binomial likelihood of the data[17] are also widely used.[18] However, when data are sparse, the choice of prior distributions for a Bayesian NMA becomes very important, and distributions that are thought to be "uninformative" or "vague" may strongly influence results.[19] This problem may be even more pronounced for the case of priors for heterogeneity in a random-effects meta-analysis.[20] Including informative priors in the analysis may help overcoming such issues. Stijnen et al proposed an alternative approach to NMA of sparse data, based on a noncentral hypergeometric (NHC) function.[19] Higgins and Whitehead proposed an extension of Peto's method for NMA.[21] This method, however, will have the same limitations as Peto's pairwise meta-analysis.

In this paper, we introduce an MH-NMA method for odds ratios, which can be of particular value when event rates are low. We implement our method in the **netmeta** package in R. In order to illustrate our approach, we reanalyze data from two previously published NMAs. We also perform simulations to assess the performance of our method in comparison with alternative NMA methods.

2 | ILLUSTRATIVE DATA FROM PUBLISHED NETWORKS

In this section, we briefly describe two data sets that we use to illustrate the methods we present in this paper.

2.1 | Inhaled medications for patients with chronic obstructive pulmonary disease

The first dataset comes from a review that compared the safety of inhaled medications in patients with chronic obstructive pulmonary disease.[22] The outcome we focus on is mortality. The available data included 41 randomized trials, with a total of 52 462 patients. Mortality was low, with 2408 deaths (4.6%) reported across all studies. There were nine studies that reported zero events in at least one of the treatment arms and three additional studies had zero events in all treatment arms ("all-zero studies"). The network is depicted in the left panel of Figure 1. The data is given in Section 1 of the Appendix.

2.2 | Methods to decrease blood loss and blood transfusion requirements for patients with liver transplantation

The second data set comes from a review that compared methods for decreasing blood loss and blood transfusion requirements during liver transplantation.[23] The outcome we analyse is mortality at 60 days posttransplantation. The network

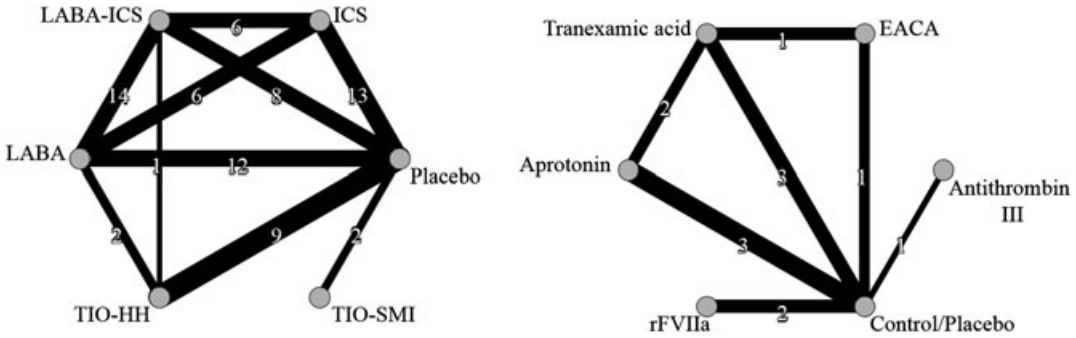

FIGURE 1 The thickness of the lines connecting the treatments is proportional to the inverse standard error of the corresponding treatment effects, based on a Mantel-Haenszel network meta-analysis. The numbers on the edges show the number of studies per comparison. A, Inhaled medications in patients with chronic obstructive pulmonary disease; B, Methods to prevent blood loss in liver transfusion. EACA, epsilon amino caproic acid; ICS, inhaled corticosteroid; LABA, long-acting $\beta 2$ agonist; rFVIIa, recombinant factor VIIa; TIO-HH, tiotropium dry powder; TIO-SMI, tiotropium solution

compared seven alternative methods. Fourteen studies reported mortality at 60 days, in 1002 patients. Forty-five deaths were reported across all studies (4.5%). Six studies observed deaths in all treatment arms while three studies did not observe any deaths. The network is depicted in the right panel of Figure 1. Note that one of the treatments in the original dataset (solvent detergent plasma) was only included in one study with zero events in all treatment arms ("all-zero study"). This treatment was excluded from the network graph.

3 | METHODS

3.1 | Mantel-Haenszel NMA method

In this section, we describe our method for NMA using MH odds ratios. The analysis is performed in three stages.

Stage 1: Setting up the data

In the first stage, we bring the data in an appropriate format for MH-NMA in five steps.

i. We remove all-zero studies from the dataset. These studies contribute no information to the calculation of MH odds ratios.

ii. We group studies by design. Here, the word "design" is used to denote the treatments being compared in each study.[24] For example, a study comparing treatments X and Y is of an XY design, a study comparing X, Y, and Z is of an XYZ design, etc.

iii. Within each design, we search for treatments for which all studies in the design reported zero events, and remove the treatments from the design. For example, if in all XYZW studies there were no events observed in arm Y, we remove the Y arms from these studies. These studies now include information only on X, Z, and W. This is required because MH odds-ratios cannot be estimated; however, the design will still be labelled as XYZW. We follow this approach aiming to be consistent with the original idea about how study design may interact with treatment effects.[24] Quoting from that paper, "*… we implicitly assume that different designs (ie, different sets of included treatments) may serve as a proxy for one or more important effect modifiers.*" Note that all these relate to the way we later assess inconsistency; see Section 3.2 as follows. Also note that instead of removing these treatment arms, we could use a continuity correction, but as discussed above, this strategy has important limitations. In our software implementation, we allow both choices (see Section 4.1).

iv. If designs are left with only one treatment arm after step *(iii)*, we completely remove these designs from the data.

v. We check the network connectivity. After steps *(iii)* and *(iv)* above, there is the chance that the network becomes disconnected (or completely disappear, if all designs are affected). In such cases, a NMA cannot be performed in the whole data set, but only in connected subnetworks.

At the end of this first stage, we have a new data set, typically a subset of the original data, in which all studies are grouped by design.

TABLE 1 Data available from study i, comparing treatments X, Y, Z, etc

Study arm	Events	Non-events	Total
Treatment X	a_{Xi}	b_{Xi}	t_{Xi}
Treatment Y	a_{Yi}	b_{Yi}	t_{Yi}
Treatment Z	a_{Zi}	b_{Zi}	t_{Zi}
…	…	…	…
Total	a_{+i}	b_{+i}	t_{+i}

Stage 2: Direct MH meta-analyses per design

At the second stage, we synthesise data within each design using the MH meta-analysis method. Note that a total of $T_d(T_d-1)/2$ different log-odds ratios (logORs) can be estimated from each design d, where T_d is the number of treatments in this design. For the NMA, however, we only need T_d-1 parameters to be estimated per design.[25] These parameters can be chosen to be the logORs of all treatments versus an arbitrary treatment in this design. Thus, at the end of this stage, for each design d, we need to obtain a vector $\widehat{\boldsymbol{\theta}}_{(d)}$ of MH summary logORs with dimensions $1\times(T_d-1)$ and the corresponding variance-covariance matrix $\boldsymbol{V}_{(d)}$. $\boldsymbol{V}_{(d)}$is a symmetric matrix, with dimensionality $(T_d-1)\times(T_d-1)$. Following the work of Lu et al,[25] the summary information of this first stage of the analysis can be then written compactly as a vector $\boldsymbol{\theta}=(\widehat{\boldsymbol{\theta}}_{(1)},\widehat{\boldsymbol{\theta}}_{(2)},\dots,\widehat{\boldsymbol{\theta}}_{(N_d)})^T$ and a matrix $\boldsymbol{V}=\text{diag}(\boldsymbol{V}_{(1)},\boldsymbol{V}_{(2)},\dots,\boldsymbol{V}_{(N_d)})$, where N_d is the number of different designs in the network. Thus, the dimensions of $\boldsymbol{\theta}$ are $\sum_d(T_d-1)\times 1$ and the dimensions of $\boldsymbol{V}$ are $\sum_d(T_d-1)\times\sum_d(T_d-1)$. Estimation of $\widehat{\boldsymbol{\theta}}_{(d)}$ and $\boldsymbol{V}_{(d)}$ is standard when the design has only two arms ($T_d=2$) and can be performed using already available software (eg, **metan**[26] in Stata or **meta**[27] in R). For designs with more than two arms ($T_d>2$), we employ a generalized MH estimator, as proposed by Greenland.[28] The details of this estimator are presented as follows.

Let us assume that, in design d, there are S_d studies comparing T_d different treatments. Assume that study i provides data in the form of a $(T_d\times 2)$ table, as shown in Table 1. Following Greenland's notation, we define $c_{XYi}=a_{Xi}b_{Yi}/t_{+i}$ and $C_{XY}=\sum_{i=1}^{S_d}c_{XYi}$. In this notation, C_{XY}/C_{YX} corresponds to the usual MH estimator for the comparison X versus Y, when $T_d=2$. Mickey and Elashoff suggested that this expression can also be used to estimate odds ratios for $T_d>2$.[29] Greenland provided an alternative estimator which he showed to have an efficiency advantage,[28] and this is what we will use here.

For simplicity, and without loss of generality, let us consider treatment X to be the reference treatment for this design. Then, our goal is to estimate the T_d-1 summary MH logORs $\widehat{\boldsymbol{\theta}}_{(d)}=(\widehat{\theta}_{d,XY},\widehat{\theta}_{d,XZ},\dots)$ and their variance-covariance matrix $\boldsymbol{V}_{(d)}$. Let us define $L_{XY}=\ln(C_{XY}/C_{YX})$, and $w_{XYi}=(a_{Xi}+b_{Yi})/t_{+i}$. Greenland's estimator is defined as

$$\widehat{\theta}_{d,XY}=\frac{L_{X+}-L_{Y+}}{T_d}, \tag{1}$$

where $L_{X+}=\sum_{J=1}^{T_d}L_{XJ}$. This estimator incorporates all data from each $(T_d\times 2)$ table. As an example, for the case of three treatments XYZ, $\widehat{\theta}_{d,XY}=(2L_{XY}+(L_{XZ}-L_{YZ}))/3$.

The variance of L_{XY} is $\text{var}(L_{XY})=U_{XYY}$, where

$$U_{XYY}=\frac{\sum_{i=1}^{S_d}c_{XYi}w_{XYi}}{2C_{XY}^2}+\frac{\sum_{i=1}^{S_d}c_{XYi}w_{YXi}+c_{YXi}w_{XYi}}{2C_{XY}C_{YX}}+\frac{\sum_{i=1}^{S_d}c_{YXi}w_{YXi}}{2C_{YX}^2}. \tag{2}$$

Note that $U_{XYY}=U_{YXX}$, and thus $\text{var}(L_{XY})=\text{var}(L_{YX})$. The covariance of L_{XY} and L_{XZ}, when $X\neq Y\neq Z\neq X$, is given by

$$\begin{aligned}U_{XYZ}=&\frac{\sum_{i=1}^{S_d}a_{Xi}b_{Yi}b_{Zi}/t_{+i}^2}{3C_{XY}C_{XZ}}+\frac{\sum_{i=1}^{S_d}t_{Xi}b_{Yi}a_{Zi}/t_{+i}^2}{3C_{XY}C_{ZX}}\\&+\frac{\sum_{i=1}^{S_d}t_{Xi}a_{Yi}b_{Zi}/t_{+i}^2}{3C_{YX}C_{XZ}}+\frac{\sum_{i=1}^{S_d}b_{Xi}a_{Yi}a_{Zi}/t_{+i}^2}{3C_{YX}C_{ZX}}.\end{aligned} \tag{3}$$

All elements of U_{XYZ} for which $X=Y$ or $X=Z$ are set to zero. Following Greenland's notation, let us also define $U_{XX}^{+}=U_{X++}=\sum_{Y,Z}U_{XYZ}$ and $U_{XY}^{+}=\sum_J(U_{JXY}-U_{XYJ}-U_{YXJ})+U_{XYY}, X\neq Y$. Using these two definitions, the variance of $\widehat{\theta}_{d,XY}$ is

$$\text{var}(\widehat{\theta}_{d,XY})=\frac{U_{X++}-2U_{XY}^{+}+U_{Y++}}{T_d^2}. \tag{4}$$

The covariance between two estimates $\widehat{\theta}_{d,XY}, \widehat{\theta}_{d,XZ}$ is

$$\operatorname{cov}\left(\widehat{\theta}_{d,XY}, \widehat{\theta}_{d,XZ}\right) = \frac{U^{+}_{XX} - U^{+}_{XZ} - U^{+}_{YX} + U^{+}_{YZ}}{T^{2}_{d}}. \tag{5}$$

More generally, the covariance between $\widehat{\theta}_{d,XY}$ and $\widehat{\theta}_{d,WZ}$ is

$$\operatorname{cov}\left(\widehat{\theta}_{d,XY}, \widehat{\theta}_{d,WZ}\right) = \frac{U^{+}_{XW} - U^{+}_{XZ} - U^{+}_{YW} + U^{+}_{YZ}}{T^{2}_{d}}. \tag{6}$$

Equations (1), (4), and (5) can be used to estimate $\widehat{\boldsymbol{\theta}}_{\boldsymbol{d}}$ and $\boldsymbol{V}_{(d)}$, and consequently $\boldsymbol{\theta}$ and $\boldsymbol{V}$.

Stage 3: Synthesis of direct MH odds-ratios across designs assuming consistency

In this stage, we start by arbitrarily defining a reference treatment for the network. All treatment contrasts versus this reference treatment are the *basic parameters* of the NMA model. Without loss in generality, let us define X to be the reference treatment in the network. Then, the relative effects δ_{XY}, $\delta_{XZ}, \ldots$ are the basic parameters of the model. All other relative effects, the *functional parameters*, can be calculated as linear combinations of the basic parameters, eg, $\delta_{YZ} = \delta_{XZ} - \delta_{XY}$. If the data set includes a total of T treatments, then there are $T(T-1)/2$ different treatment contrasts, which can be grouped in a vector $\boldsymbol{\delta}$. From these contrasts, $T-1$ are the basic parameters ($\boldsymbol{\delta}_{\text{basic}}$) and the rest are functional. Following the work of Lu et al,[25] we can write $\boldsymbol{\delta} = \boldsymbol{H}\boldsymbol{\delta}_{\text{basic}}$. Matrix $\boldsymbol{H}$ has elements 1, 0, and -1 and maps the basic parameters into all possible treatment comparisons in the network. $\boldsymbol{H}$ has dimensions $\frac{T(T-1)}{2} \times (T-1)$.

Next, we need to define the design matrix $\boldsymbol{X}$. This is a matrix with dimensions $\sum_d (T_d - 1) \times (T-1)$, which describes which treatments are being compared in each design and maps the corresponding comparison into the basic parameters. For additional details on how to setup $\boldsymbol{X}$, we refer our readers to section 7.1 in the work of Lu et al.[25]

The weighted least squares NMA estimates for the basic parameters are given by $\widehat{\boldsymbol{\delta}}_{\mathbf{basic}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{V}^{-\mathbf{1}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{V}^{-\mathbf{1}}\boldsymbol{\theta}$.[30] The corresponding variance-covariance matrix is given by $\operatorname{Cov}(\widehat{\boldsymbol{\delta}}_{\mathbf{basic}}) = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{V}^{-\mathbf{1}}\boldsymbol{X})^{-\mathbf{1}}$. Finally, the NMA estimates for all treatment effects are given by[25]

$$\widehat{\boldsymbol{\delta}} = \boldsymbol{H}\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{V}^{-\mathbf{1}}\boldsymbol{X}\right)^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{V}^{-\mathbf{1}}\boldsymbol{\theta}, \tag{7}$$

with a variance-covariance matrix equal to

$$\operatorname{Cov}(\widehat{\boldsymbol{\delta}}) = \boldsymbol{H}\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{V}^{-\mathbf{1}}\boldsymbol{X}\right)^{-\mathbf{1}}\boldsymbol{H}^{\boldsymbol{T}}. \tag{8}$$

3.2 | Statistical evaluation of the consistency assumption in MH-NMA

Consistency refers to the (statistical) agreement between the various sources of information in NMA.[24,31] We employ two approaches for assessing consistency, ie, the first is a global approach (in the whole network) and the second is local (corresponding to each design).

The global method is based on a generalized Cochran's Q statistic, calculated for the whole network as $Q_{\text{inc}} = (\boldsymbol{\theta} - \boldsymbol{X}\,\widehat{\boldsymbol{\delta}}_{\mathbf{basic}})'\boldsymbol{V}^{-\mathbf{1}}(\boldsymbol{\theta} - \boldsymbol{X}\,\widehat{\boldsymbol{\delta}}_{\mathbf{basic}})$.[32] Under the null hypothesis of consistency, the Q_{inc} statistic follows a chi-squared distribution with $\sum_d (T_d - 1) - T + 1$ degrees of freedom, where T corresponds to the total number of treatments in the network. This statistic can be used to assess inconsistency in the whole network.

In order to identify local inconsistency in the network, we propose a new approach, which shares similarities with the so-called "separate indirect from direct evidence" (SIDE) or "node-splitting" approach.[33] In the SIDE approach, the focus is on one pairwise comparison at a time, eg, XY. Following this approach, if there are only two-arm (and no multiarm) studies comparing X and Y, we remove them from the network, and we use them to perform a pairwise meta-analysis. This provides a direct estimate of the relative treatment effects of X versus Y. We then perform a NMA on the remaining studies, which provides an indirect estimate for X versus Y. The indirect and direct relative treatments effects are subsequently compared. If there are multiarm studies in the network that compare X and Y; this approach splits only the corresponding direct estimate. For example, if there is a three-arm study XYZ, this approach uses the XY estimate in the pairwise meta-analysis of direct effects, and the XZ and YZ estimates in the NMA that estimates the indirect effects.

Although this method works well when all studies are two-arm, it runs into problems for the case when there are multiarm studies in the network. In the example, the removed XY direct estimate from the multiarm study will always agree with the indirect estimate obtained from the XZ and ZY estimates of the same study. This might dilute the evidence

of inconsistency from the rest of the network. In addition, different choices of the baseline treatment in each study may lead to different estimates regarding the difference between direct and indirect evidence and some practical approaches have been suggested, eg, see the documentation in the **network**[34] command in Stata.

An adaptation that can overcome this problem is to "separate indirect from direct design evidence" (SIDDE). Focusing again on XY, we remove all studies that compare these two treatments, both two-arm and multiarm. Thus, all XY, XYZ, XYZW, etc, studies are excluded from the network. The rest of the estimating procedure is the same as in the standard SIDE approach, ie, we use the excluded studies to estimate the XY effects directly, and the rest of the network to estimate them indirectly. Note that the SIDDE approach shares some similarities with the net heat plot for detecting inconsistency, proposed by Krahn et al.[32] Also note that, using this approach, we can only estimate inconsistency for treatment comparisons for which there is both direct and indirect evidence.

3.3 | Software for fitting MH-NMA

We developed a function, `netmetabin`, which is included in the **netmeta** package[35] in R.[36] This function can be used to fit the MH-NMA method presented above. It can also fit NMA with a NHC likelihood using the Breslow approximation, as proposed by the work of Stijnen et al.[19] This approximation is valid when the total number of events is small relative to the group sizes. In that case, the NHC distribution can be approximated by a binomial distribution.[37] We will refer to this approach as NCH-NMA. We have also implemented the SIDDE approach to inconsistency in the (existing) `netsplit` function of R package **netmeta**.

On a technical note, following Section 3.1, and more specifically the discussion in step *(iii)* of Stage 1, the default of `netmetabin` when fitting MH-NMA does not perform any continuity correction when there are designs in which some of the treatment arms had no events. Instead, the arms are excluded from the network. However, the user can override this default by setting argument *cc.pooled* to TRUE and specify a fixed value for the continuity correction (argument *incr*). Note that NCH-NMA can be used in these scenarios without having to remove treatment arms or to use continuity corrections, as long as there are events for all the treatments in the network, irrespective of design.

4 | CLINICAL EXAMPLES

4.1 | Fitting details

We compared the results from our MH-NMA method to three alternative approaches. First, we fitted NCH-NMA, using the R function `netmetabin`. Second, we fitted a common-effects NMA with the usual, IV approach,[38] also using `netmetabin`. We will term this method IV-NMA. In order to use this approach, we employed a 0.5 continuity correction in studies with zero events in one or more treatment arms. All-zero studies were removed from the dataset. Third, we fitted a common-effects Bayesian NMA model with a binomial likelihood. The Bayesian model was fitted in OpenBUGS.[39] For all model parameters, ie, the baseline risk in each study and the true logORs, we used "vague" prior distributions, $N(0, \sigma^2 = 100)$. We ran 2 chains in parallel, performed 100 000 iterations, and discarded the first 20 000 samples of each chain. We checked convergence using the Brooks-Gelman-Rubin criterion.[40] The code we used for fitting all methods is provided in the Appendix.

4.2 | Inhaled medications for patients with chronic obstructive pulmonary disease

Results are shown in the upper part of Table 2. We also fitted random-effects IV-NMA, but between-study variance (τ^2) was estimated to be zero, and thus, results were identical to the common-effects IV-NMA.

It is clear that all methods give almost identical results. This is due to the fact that, even though the event was relatively rare, there were many large studies in the network, ie, the average sample size in the studies was 1280 patients. This resulted in many of the studies having enough events to adequately allow all methods to estimate relative treatment effects between the treatments in the network.

Regarding inconsistency, the Q_{inc} statistic was found to be 8.35 (9 degrees of freedom), corresponding to a p-value of 0.50, thus showing no evidence of global network inconsistency. The SIDDE approach to local inconsistency identified three treatment comparisons that showed some disagreement between direct and indirect estimates; long-acting $\beta 2$ agonist (LABA) versus LABA–inhaled corticosteroid (ICS), LABA versus tiotropium dry powder (TIO-HH), and LABA-ICS versus

TABLE 2 Comparison of four common-effects methods to estimate summary odds ratios in two previously published networks.[22,23] Treatment abbreviations as given in Figure 1. An odds ratio larger than 1 favours placebo in the COPD network, and control/placebo in the liver transplantation network

	Treatment	MH-NMA	NCH-NMA	IV-NMA	Bayesian NMA binomial likelihood
Inhaled medications for patients with chronic obstructive pulmonary disease (COPD)[22] *Mortality odds ratios compared to placebo [95% Confidence/ Credible Intervals]*	ICS	1.03 [0.88; 1.21]	1.02 [0.88; 1.19]	1.02 [0.86; 1.19]	1.03 [0.88, 1.21]
	LABA	0.93 [0.79; 1.08]	0.94 [0.81; 1.09]	0.93 [0.79; 1.08]	0.93 [0.79; 1.09]
	LABA-ICS	0.79 [0.67; 0.94]	0.81 [0.69, 0.95]	0.78 [0.66; 0.93]	0.80 [0.67, 0.95]
	TIO-HH	0.92 [0.81; 1.05]	0.93 [0.82; 1.05]	0.92 [0.81; 1.04]	0.92 [0.81; 1.04]
	TIO-SMI	1.52 [1.05; 2.19]	1.50 [1.05; 2.14]	1.50 [1.04; 2.17]	1.52 [1.06; 2.19]
Methods to decrease blood loss and blood transfusion requirements for patients with liver transplantation[23] *Mortality odds ratios compared to control/placebo [95% Confidence/ Credible Intervals]*	Aprotonin	0.36 [0.13; 0.99]	0.36 [0.14; 0.96]	0.40 [0.15; 1.08]	0.32 [0.11; 0.87]
	EACA	0.74 [0.14; 3.89]	0.77 [0.16; 3.77]	0.81 [0.16; 4.26]	0.71 [0.12; 3.92]
	rFVIIa	1.54 [0.31; 7.53]	1.51 [0.32; 7.17]	1.54 [0.31; 7.53]	1.58 [0.37; 10.42]
	Tranexamic acid	0.77 [0.23; 2.61]	0.81 [0.27; 2.43]	0.90 [0.28; 2.96]	0.76 [0.24; 2.54]
	Antithrombin III	-	-	0.21 [0.01; 4.89]	0.0003 [0.00; 0.72]

Abbreviations: IV-NMA, inverse-variance network meta-analysis; MH-NMA, Mantel-Haenszel network meta-analysis; NCH-NMA, non-central hypergeometric network meta-analysis.

TIO-HH (p-values 0.02, 0.06, and 0.08, respectively). This, in turn, might call for a closer examination of the studies that contribute to these particular comparisons, to check for breaches in the transitivity assumption or the appropriateness of the assumption of homogeneity. More specifically, there are two studies comparing LABA versus TIO-HH, which also contribute to the indirect evidence for LABA versus LABA-ICS and LABA versus TIO-HH. These studies may warrant further investigation.

4.3 | Methods to decrease blood loss and blood transfusion requirements for patients with liver transplantation

Results are shown in the lower part of Table 2. As in the previous example, we also fitted a random-effects IV-NMA, but heterogeneity (τ^2) was estimated to be zero. Thus, results from this method were identical to the common-effects IV-NMA shown in Table 2.

In contrast to the first example, there are differences between the approaches. The data that we imputed using the continuity correction accounted for almost 20% of the total events included in the IV-NMA. Thus, results will be strongly influenced by the imputed data, and the method cannot be trusted to give reliable results.[3] The Bayesian model might also be problematic, because the prior distributions used for the models' parameters, although chosen to be "uninformative" or "vague," might have a strong effect on results.[20] In this example, if we switch the prior distributions for the model parameters to $U(-5,5)$, we get quite different estimates. For example, for recombinant factor VIIa versus placebo, the point estimate [95% credible interval] changes from 1.58 [0.37; 10.42] to 1.11 [0.32; 3.75]. Thus, in this example, given that there is no available external information to feed in the Bayesian models in the form of informative priors, our MH-NMA or the NCH-NMA might be the best option for analysis.

Note that Antithrombin III does not feature in the results of MH-NMA and NCH-NMA in Table 2. This is because this treatment was only included in one two-arm study with zero events in one of each arms, and such designs are removed from these two methods (unless a continuity correction is used). This highlights one of the potential disadvantages of the frequentist approaches, as opposed to the Bayesian ones. Based on the findings of the Bayesian model with the binomial likelihood presented in Table 2, one might argue that there is enough evidence that Antithrombin III is safer than

TABLE 3 Results from the assessment of inconsistency in the two clinical examples

Inhaled medications for patients with chronic obstructive pulmonary disease			
	Q	**df**	**p-value**
Global inconsistency in the network	8.35	9	0.50
Inconsistency estimated using the SIDDE approach	**Ratio of odd ratios (direct over indirect odds ratios)**		**p-value**
ICS:LABA	1.29 [0.77; 2.17]		0.34
ICS: LABA-ICS	1.03 [0.58; 1.85]		0.92
LABA: LABA-ICS	0.47 [0.24; 0.89]		0.02
LABA: Placebo	0.88 [0.60; 1.29]		0.52
LABA:TIO-HH	1.45 [0.98; 2.16]		0.06
LABA-ICS:Placebo	1.10 [0.69; 1.75]		0.78
LABA-ICS:TIO-HH	0.59 [0.33; 1.06]		0.08
Placebo:TIO-HH	1.05 [0.74; 1.50]		0.79
Methods to decrease blood loss and blood transfusion requirements for patients with liver transplantation			
	Q	**df**	**p-value**
Global inconsistency in the network	1.88	2	0.39
Inconsistency estimated using the SIDDE approach	**Ratio of odd ratios (direct over indirect odds ratios)**		**p-value**
Aprotonin:Control/Placebo	5.97 [0.36; 98.99]		0.21
Aprotonin:Tranexamic acid	0.17 [0.01; 2.78]		0.21
Control/Placebo:Tranexamic acid	0.16 [0.01; 2.62]		0.20

Abbreviations: ICS, inhaled corticosteroid; LABA, long-acting $\beta 2$ agonist; TIO-HH, tiotropium dry powder.

control/placebo (estimated odds ratio 0.0003, with a 95% credible interval 0.00 to 0.72). However, as discussed above, such results are heavily dependent on the prior. Using a $U(-5,5)$ prior, the odds ratio estimate is 0.02 [0.0002; 1.36], shedding doubts about the effectiveness of this intervention.

No evidence of global inconsistency was found for this network. Q_{inc} was 1.88 (2 degrees of freedom, for a p-value of 0.39). The SIDDE approach did not provide any evidence for inconsistency either. Results are shown in Table 3.

5 | SIMULATIONS

In this section, we describe a simulation study to compare our MH-NMA method with other methods for NMA. Our aim was to assess the performance of the competing approaches under different scenarios, by varying factors regarding data availability, heterogeneity, network structure, event rates, etc. Our simulation study follows the structure proposed by Morris et al.[41]

5.1 | Data generating mechanisms

We explored a total of 20 scenarios. In all scenarios we generated only fully connected networks, with at least two studies for every treatment comparison, in order to avoid having simulated data sets that result in disconnected networks when the assumed event rate is very low. For each scenario, we independently generated 1000 datasets. The simulated a number of studies varied across scenarios (see details as follows). In all studies, we assumed equal number of patients per treatment arm. The number of patients per treatment arm was generated separately for each study.

In scenarios 1 to 16, we generated only two-arm studies. The number of patients per treatment arm was simulated by drawing from a uniform distribution $U(30,60)$, and then rounding to a whole number. For each study, we generated a study baseline risk, ie, the risk of event of patients receiving a reference treatment (treatment 1) irrespective of whether this was evaluated in the study. The study baseline risk was generated from a uniform distribution with parameters that varied by scenario (details as follows). The underlying true logORs of each treatment versus treatment 1 were fixed at equal intervals between 0 and 1. For example, in a scenario with five treatments, the true logORs for treatments 2, 3, 4, and 5 versus treatment 1 were set to 0.25, 0.50, 0.75, and 1.00, respectively. For scenarios assuming homogeneity, we used these logORs together with the simulated, study-specific baseline risk to calculate the probability of an event in each study treatment arm. The number of events in a study arm was generated from a binomial distribution, using the

TABLE 4 Overview of the scenarios we explored in our simulations. For each scenario we generated 1000 independent data sets. Rows shaded grey correspond to scenarios where we assumed heterogeneity

#	Treatments In the network	Patients per treatment arm	Number or studies per comparison	Heterogeneity standard deviation (τ)	Baseline risk
1	5	30 to 60	2	0	3% to 5%
2	5	30 to 60	2	0.1	3% to 5%
3	5	30 to 60	2	0	5% to 10%
4	5	30 to 60	2	0.1	5% to 10%
5	5	30 to 60	4	0	3% to 5%
6	5	30 to 60	4	0.1	3% to 5%
7	5	30 to 60	4	0	5% to 10%
8	5	30 to 60	4	0.1	5% to 10%
9	8	30 to 60	2	0	3% to 5%
10	8	30 to 60	2	0.1	3% to 5%
11	8	30 to 60	2	0	5% to 10%
12	8	30 to 60	2	0.1	5% to 10%
13	8	30 to 60	4	0	3% to 5%
14	8	30 to 60	4	0.1	3% to 5%
15	8	30 to 60	4	0	5% to 10%
16	8	30 to 60	4	0.1	5% to 10%
17	3	30 to 60	8	0	1% to 10%
18	5	100 to 200	2	0	30% to 50%
19	5	100 to 200	2	0.1	30% to 50%
20	3	30 to 60	8	0	1% to 2%

(study- and treatment-specific) probability of an event, as well as the study-specific number of patients per treatment arm. For scenarios that assume heterogeneity, we set a common standard deviation of the random effects (τ) for all treatment comparisons in the network. For these scenarios, we simulated study-specific logORs, accounting for τ, ie, by drawing from a normal distribution with a mean determined by the comparison, and standard deviation equal to τ; otherwise, the procedure was unchanged. In scenarios 1 to 16, we explored the following settings.

- number of treatments in the network: 5 or 8.
- number per studies per treatment comparison: 2 or 4.
- standard deviation of random effects: $\tau = 0$ (homogeneous treatment effects) and $\tau = 0.1$ (heterogeneous treatment effects).
- baseline event rate in each study: generated after drawing from $U(0.03,0.05)$ or $U(0.05,0.10)$

In scenarios 17 to 20, we considered more complicated situations that included multiarm studies, large sample sizes and event rates, or very small event rates. More specifically:

- In scenario 17, we assumed three treatments in the network populated by three-arm studies and no heterogeneity.
- In scenario 18, we assumed more patients per treatment arm ($U(100,200)$) and large event rates (baseline risk $U(0.30,0.50)$), only two-arm studies and no heterogeneity.
- Scenario 19 was equivalent with scenario 18 but with heterogeneity $\tau = 0.1$.
- In scenario 20, we assumed three treatments in the network, and very low baseline event rates from $U(0.01,0.02)$). We assumed only two-arm studies and no heterogeneity.

Table 4 provides an overview of the data generating mechanism in each scenario.

5.2 | Methods compared, estimands, and measures of performance

Each of the 1000 data sets of each scenario was analysed using four different approaches, ie, (1) common-effects IV-NMA with 0.5 continuity correction, (2) random effects IV-NMA with 0.5 continuity correction, (3) MH-NMA without

TABLE 5 Overview of simulation results. Methods' abbreviations as per Table 2. Rows shaded grey correspond to scenarios where we assumed heterogeneity

Scenario	IV - FIXED			IV - RANDOM			MH-NMA			NCH-NMA		
	Mean bias	Mean absolute bias	Coverage	Mean bias	Mean absolute bias	Coverage	Mean bias	Mean absolute bias	Coverage	Mean bias	Mean absolute bias	Coverage
1	−0.08	0.30	97.8%	−0.07	0.30	97.9%	−0.05	0.31	97.4%	−0.05	0.31	96.4%
2	−0.07	0.31	97.6%	−0.07	0.31	97.6%	−0.05	0.32	96.8%	−0.05	0.32	96.0%
3	−0.04	0.24	97.6%	−0.04	0.23	97.8%	−0.01	0.24	96.8%	−0.06	0.23	95.8%
4	−0.04	0.24	96.7%	−0.03	0.24	97.2%	−0.01	0.25	95.7%	−0.06	0.24	95.2%
5	−0.08	0.22	97.2%	−0.08	0.22	97.2%	0.00	0.23	96.0%	−0.02	0.22	95.1%
6	−0.09	0.22	97.4%	−0.08	0.22	97.4%	−0.01	0.23	96.2%	−0.03	0.22	95.5%
7	−0.04	0.17	96.5%	−0.04	0.17	96.8%	0.00	0.17	95.6%	−0.06	0.17	94.0%
8	−0.05	0.17	96.6%	−0.05	0.17	96.9%	−0.01	0.18	95.3%	−0.07	0.17	94.2%
9	−0.07	0.24	97.3%	−0.07	0.24	97.3%	−0.05	0.26	96.8%	−0.05	0.25	95.8%
10	−0.08	0.24	97.5%	−0.07	0.24	97.5%	−0.05	0.25	96.6%	−0.05	0.25	95.8%
11	−0.04	0.19	96.5%	−0.04	0.19	96.7%	−0.02	0.19	95.9%	−0.06	0.19	95.0%
12	−0.04	0.19	96.7%	−0.04	0.19	96.9%	−0.02	0.20	95.9%	−0.06	0.19	94.8%
13	−0.07	0.18	96.8%	−0.07	0.18	96.9%	0.00	0.18	96.1%	−0.02	0.18	95.3%
14	−0.08	0.18	97.1%	−0.08	0.18	97.1%	−0.02	0.18	95.8%	−0.04	0.18	95.0%
15	−0.04	0.14	96.5%	−0.04	0.14	96.6%	0.00	0.14	95.6%	−0.05	0.14	94.2%
16	−0.05	0.14	96.2%	−0.05	0.14	96.3%	−0.01	0.14	95.4%	−0.07	0.14	93.3%
17	−0.06	0.23	96.2%	−0.06	0.23	96.6%	0.03	0.25	94.3%	−0.05	0.23	94.5%
18	0.00	0.08	95.7%	0.00	0.08	96.6%	0.00	0.08	95.7%	−0.31	0.31	21.0%
19	0.00	0.09	92.5%	0.00	0.09	94.6%	0.00	0.09	92.4%	−0.31	0.31	21.0%
20	−0.18	0.26	97.1%	−0.18	0.26	97.1%	0.01	0.29	96.2%	0.00	0.29	95.5%

continuity correction, and (4) NHC-NMA without continuity correction. We did not consider Bayesian models, as for rare outcomes the performance of a Bayesian model might heavily depend on the prior distributions used (see Section 4.3 for an illustration).

The target estimands of our analyses were the logORs between treatments versus treatment 1 (ie, the basic parameters). The performance of each method was assessed by comparing the estimated logORs with their corresponding true values. We calculated the mean bias, defined as the mean difference between the estimated and the true logORs, the mean absolute bias, and the mean coverage in each scenario, as the percent of 95% confidence intervals that included the corresponding true logOR. All analyses were performed in R using the **netmeta** package. The codes used for our simulations can be found in https://github.com/esm-ispm-unibe-ch-REPRODUCIBLE/MH_NMA.

5.3 | Results from simulations

We present the results of the simulation study in Table 5.

The inverse variance method with fixed effects (IV-FE) and random effects (IV-RE) gave almost identical results for all scenarios. This is because when event rates are low, it is difficult to estimate τ, and the method of moments estimates $\hat{\tau} = 0$. In scenarios 1 to 16, MH-NMA always provided the least biased estimates, followed closely by NCH, which had identical mean bias in 4 of 16 scenarios. In most of these scenarios, NCH had a coverage slightly closer to the nominal level (95%) than MH. The IV method had the worst performance both in terms of bias and coverage; although, in some cases, differences were trivial. Moreover, simulating heterogeneity led to only a marginal increase of bias in the estimates of MH and NCH.

In scenario 17, with multiarm studies only, MH gave again the least biased results, with similar coverage with NCH. Scenarios 18 and 19 assumed higher frequencies of events and NCH performed very poorly due to using the Breslow approximation, which is only valid for rare events (see Section 3.3). In scenario 20, with very low event rates, the IV method showed large biases, while MH and NCH were practically unbiased. The NCH again provided a slightly better coverage, ie, slightly closer to the nominal level.

Note that, in Table 5, we show the values of mean bias, absolute bias, and coverage of all estimated logORs, ie, we average over all basic parameters of each scenario. Thus, the values presented in Table 5 could be further split, for each basic parameter in each scenario. For example, in scenario 1, where we assumed five treatments (four basic parameters), the true values of the logORs were (0.25, 0.50, 0.75, 1.00). The corresponding mean biases across the four basic parameters were (−0.04, −0.07, −0.09, and −0.11), while the mean bias for the common-effects IV-NMA across all basic parameters was −0.08 (as shown in Table 5). Similarly, for MH, the mean bias across all basic parameters was −0.05, and the biases of the basic parameter were (−0.04, −0.05, −0.05, and −0.06), respectively.

Overall, we conclude that IV-NMA with a continuity correction is a suboptimal choice, when events are rare. The MH-NMA and NCH-NMA with the Breslow approximation performed comparably in most scenarios. Due to the fact that this approximation is inadequate for large event rates or when there is a mixture of low and high event rates across the studies, we recommend the use of MH-NMA. For the case when all studies show low event rates, NCH-NMA might offer a slight advantage over MH-NMA in terms of coverage. Readers should keep in mind, however, that, in our simulations, we did not explore the performance of the full NCH-NMA method (ie, including random effects, without the Breslow approximation). See also the discussion section, regarding possible areas of future research.

6 | DISCUSSION

We presented a method for NMA of binary outcomes, using a generalized version of the MH approach to pairwise meta-analysis. The method synthesises odds ratios and it does not rely on the normal approximation to estimate study variances. It allows the inclusion of information from studies with zero events in some, but not all, treatment arms, without a continuity correction.

One limitation of our method is that it is limited to odds-ratios. An extension to risk ratios might be interesting to pursue in future work. However, for the case of sparse data, the difference between odds and risks becomes negligible, and thus odds ratios can be interpreted as risk ratios for practical purposes. Another limitation of our method is that it can only be used to perform a common-effects NMA. However, when the event rate is low, the estimation of heterogeneity in a frequentist setting can be difficult.[42] In both clinical applications we presented in this paper, the heterogeneity standard deviation was estimated to be zero using the method of moments,[43] and the IV-NMA analyses were performed

under a common-effects assumption. A similar picture was seen in our simulation study, ie, when events were rare, the random-effects IV-NMA gave on average almost identical results to the fixed-effects IV-NMA. Thus, although our MH-NMA method's inability to include random effects is in theory a disadvantage, for the case of rare events, the IV-NMA method might in practice be also limited to the common-effects approach. Note that the Cochrane Handbook suggests that incorporation of heterogeneity should be a secondary consideration when attempting to estimate treatment effects from sparse data.[44]

The fact that the MH-NMA method does not account for heterogeneity suggests that the results from the global test or the SIDDE method to assess inconsistency should be interpreted with caution. If heterogeneity is present in the form of within-design variability, then ignoring it when estimating the variance of the direct summary odds-ratios might contribute to inconsistency. Hence, important inconsistency in the data should challenge the assumption of the homogeneity that underlines the model. Researchers should also keep in mind that tests for inconsistency have in general low power.[45] Consequently, lack of evidence for inconsistency should not be interpreted as evidence for consistency. Careful consideration of the study inclusion criteria and evaluation of their similarity with respect to effect modifiers should always take place to ensure that the network has low risk of intransitivity.[16]

The majority of published NMA are fitted within a Bayesian framework.[18,46] In this case, the exact binomial likelihood for the data can be employed without requiring any "correction" for zero events in study arms. However, similar to other Bayesian analyses, the sparser the data, the bigger the influence of the prior distributions on the posterior estimates of the model. Different "uninformative" or "vague" prior distributions may lead to different estimated effects sizes. The impact is generally larger for scale parameters (eg, the variance of random effects) than location parameters (eg, the true underlying effect size).[20] The problem can be mitigated with the use of informative priors for at least some of the model parameters, and in particular for heterogeneity.[47] As illustrated in the liver transplant example in Section 4.3, the choice of priors for the location parameters can have a strong impact on some of the estimated effect sizes in a NMA. Thus, unless meta-analysts have at their disposal high-quality external information for all model parameters, results from applying Bayesian methods may not be robust.

As illustrated in the chronic obstructive pulmonary disease example, the IV-NMA method might be a reasonable frequentist alternative when there are several studies with relatively large numbers of events in the dataset. Conversely, in cases where the IV-NMA method requires the imputation of a large amount of data (as in the liver transplant example), MH-NMA might be a better option for a frequentist analysis. Stijnen et al proposed an alternative frequentist approach to NMA, which can be used when the outcome is rare.[19] This uses an exact conditional likelihood, ie, it models the likelihood of events in each study arm given the total number of events in the study using the NHC distribution. For both examples presented in this paper, MH-NMA and the common-effects NHC method gave similar results. However, the two methods make different distributional assumptions about the observed data (ie, the approach by Stijnen et al[19] implements a NHC distribution) and might provide different estimates in other cases, especially if the random effects version of that approach is used (but we have not yet implemented this option in **netmeta**).

In our simulation study, we compared the performance of MH-NMA with the IV-NMA method (both fixed- and random-effects IV-NMA, using a 0.5 continuity correction) and also with the fixed effects method proposed by Stijnen et al[19] (NCH-NMA), employing an approximation only valid for small event rates. We found that our method performed similarly to NCH-NMA in most scenarios, when events are rare; both methods performed better than IV-NMA. Consequently, we recommend that researchers perform sensitivity analyses using these two methods alongside Bayesian NMAs, to evaluate the robustness of conclusions.[5,9] The use of IV-NMA for binary outcomes should be restricted to the case when events are not rare.

In order to provide recommendations regarding the optimal approach in more variable scenarios, future work could focus on comparing in simulations our MH-NMA model with the random-effects method by Stijnen et al[19] (without approximation) and also Bayesian NMA models with different priors. The different approaches could also be empirically compared in large collections of meta-analyses, to see if, and in which cases, there are important differences in the estimates of the different methods in practice. Additional simulation studies would also be needed to assess the performance of the available methods for assessing inconsistency. One other area of future research could be to explore different approaches to MH-NMA. For example, one could perform a "data augmentation" to all studies,[48] ie, artificially impute treatment arms with zero events and zero nonevents for the missing treatments of all studies. This would lead to all studies having the same design (comparing all treatments), and would render stage 3 of Section 3.1 obsolete. However, there might be no way to check for inconsistency with this approach. Finally, other models currently available only for pairwise meta-analysis could also be extended for the case of NMA, such as beta-binomial models,[6,49] the simple average estimator,[42] and others.

In summary, our extension of the MH method to NMA offers a useful new approach for the synthesis of binary outcomes, especially when the events are rare, and/or the sample sizes of the included studies are small. Moreover, using the **netmeta** command in R, the application of our methods is straightforward in practice.

ACKNOWLEDGEMENT

This project was supported by the Swiss National Science Foundation (Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, SNF) through project grant 166656. Matthias Egger was supported by the SNSF through special project funding under grant 174281.

ORCID

Orestis Efthimiou https://orcid.org/0000-0002-0955-7572
Gerta Rücker https://orcid.org/0000-0002-2192-2560
Guido Schwarzer https://orcid.org/0000-0001-6214-9087
Georgia Salanti https://orcid.org/0000-0002-3830-8508

REFERENCES

1. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188.
2. Page MJ, Shamseer L, Altman DG, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med*. 2016;13(5):e1002028. https://doi.org/10.1371/journal.pmed.1002028
3. Bradburn MJ, Deeks JJ, Berlin JA, Russell Localio A. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Statist Med*. 2007;26(1):53-77. https://doi.org/10.1002/sim.2528
4. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statist Med*. 2004;23(9):1351-1375. https://doi.org/10.1002/sim.1761
5. Keus F, Wetterslev J, Gluud C, Gooszen HG, van Laarhoven CJHM. Robustness assessments are needed to reduce bias in meta-analyses that include zero-event randomized trials. *Am J Gastroenterol*. 2009;104(3):546-551. https://doi.org/10.1038/ajg.2008.22
6. Kuss O. Statistical methods for meta-analyses including information from studies without any events—add nothing to nothing and succeed nevertheless. *Statist Med*. 2015;34(7):1097-1116. https://doi.org/10.1002/sim.6383
7. Rücker G, Schwarzer G, Carpenter J, Olkin I. Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. *Statist Med*. 2009;28(5):721-738. https://doi.org/10.1002/sim.3511
8. Tian L, Cai T, Pfeffer MA, Piankov N, Cremieux P-Y, Wei LJ. Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent 2×2 tables with all available data but without artificial continuity correction. *Biostatistics*. 2009;10(2):275-281. https://doi.org/10.1093/biostatistics/kxn034
9. Efthimiou O. Practical guide to the meta-analysis of rare events. *Evid Based Ment Health*. 2018;21(2):72-76. https://doi.org/10.1136/eb-2018-102911
10. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*. 1959;22(4):719-748. https://doi.org/10.1093/jnci/22.4.719
11. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis*. 1985;27(5):335-371. https://doi.org/10.1016/S0033-0620(85)80003-7
12. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Statist Med*. 2004;23(20):3105-3124. https://doi.org/10.1002/sim.1875
13. Salanti G, Higgins JP, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. *Stat Methods Med Res*. 2008;17(3):279-301. https://doi.org/10.1177/0962280207080643
14. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Meth*. 2012;3(2):80-97.
15. Higgins JPT, Welton NJ. Network meta-analysis: a norm for comparative effectiveness? *Lancet*. 2015;386(9994):628-630. https://doi.org/10.1016/S0140-6736(15)61478-7
16. Efthimiou O, Debray TPA, van Valkenhoef G, et al. GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods*. 2016;7(3):236-263. https://doi.org/10.1002/jrsm.1195
17. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Mak*. 2013;33(5):607-617. https://doi.org/10.1177/0272989X12458724
18. Petropoulou M, Nikolakopoulou A, Veroniki A-A, et al. Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015. *J Clin Epidemiol*. 2017;82:20-28. https://doi.org/10.1016/j.jclinepi.2016.11.002
19. Stijnen T, Hamza TH, Ozdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med*. 2010;29(29):3046-3067. https://doi.org/10.1002/sim.4040
20. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statist Med*. 2005;24(15):2401-2428. https://doi.org/10.1002/sim.2112

21. Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med*. 1996;15(24):2733-2749. https://doi.org/10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0
22. Dong Y-H, Lin H-H, Shau W-Y, Wu Y-C, Chang C-H, Lai M-S. Comparative safety of inhaled medications in patients with chronic obstructive pulmonary disease: systematic review and mixed treatment comparison meta-analysis of randomised controlled trials. *Thorax*. 2013;68(1):48-56. https://doi.org/10.1136/thoraxjnl-2012-201926
23. Gurusamy KS, Pissanou T, Pikhart H, Vaughan J, Burroughs AK, Davidson BR. Methods to decrease blood loss and transfusion requirements for liver transplantation. *Cochrane Database Syst Rev*. 2011;12:CD009052. https://doi.org/10.1002/14651858.CD009052.pub2
24. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Meth*. 2012;3(2):98-110.
25. Lu G, Welton NJ, Higgins JPT, White IR, Ades AE. Linear inference for mixed treatment comparison meta-analysis: a two-stage approach. *Res Synth Meth*. 2011;2(1):43-60.
26. Harris R, Bradburn M, Deeks J, Harbord R, Altman D, Sterne J. Metan: fixed- and random-effects meta-analysis. *Stata J*. 2008;8(1):3-28.
27. Schwarzer G. Meta: General Package for Meta-Analysis. https://cran.r-project.org/web/packages/meta/index.html. Accessed April 12, 2017.
28. Greenland S. Generalized Mantel-Haenszel estimators for K 2 x J tables. *Biometrics*. 1989;45(1):183-191. https://doi.org/10.2307/2532044
29. Mickey RM, Elashoff RM. A generalization of the Mantel-Haenszel estimator of partial association for 2 x J x K tables. *Biometrics*. 1985;41(3):623-635. https://doi.org/10.2307/2531282
30. Aitken AC. On least squares and linear combination of observations. *Proc Royal Soc Edinb*. 1936;55:42-48. https://doi.org/10.1017/S0370164600014346
31. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc*. 2006;101(474):447-459.
32. Krahn U, Binder H, Konig J. A graphical tool for locating inconsistency in network meta-analyses. *BMC Med Res Methodol*. 2013;13(1);35. https://doi.org/10.1186/1471-2288-13-35
33. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Statist Med*. 2010;29(7-8);932-944. https://doi.org/10.1002/sim.3767
34. White IR. Network meta-analysis. *Stata J*. 2015;15(4):951-985.
35. Rücker G, Krahn U, König J, Efthimiou O, Schwarzer G. Netmeta: Network Meta-Analysis Using Frequentist Methods. 2019. https://CRAN.R-project.org/package=netmeta
36. R Core. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. http://www.R-project.org
37. Bakbergenuly I, Kulinskaya E. Meta-analysis of binary outcomes via generalized linear mixed models: a simulation study. *BMC Med Res Methodol*. 2018;18(1):70. https://doi.org/10.1186/s12874-018-0531-9
38. Rücker G. Network meta-analysis, electrical networks and graph theory. *Res Synth Meth*. 2012;3(4):312-324.
39. Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: evolution, critique and future directions. *Statist Med*. 2009;28(25):3049-3067. https://doi.org/10.1002/sim.3680
40. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat*. 1998;7(4):434-455. https://doi.org/10.1080/10618600.1998.10474787
41. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. 2017. http://arxiv.org/abs/1712.03198
42. Bhaumik DK, Amatya A, Normand S-L, et al. Meta-analysis of rare binary adverse event data. *J Am Stat Assoc*. 2012;107(498):555-567. https://doi.org/10.1080/01621459.2012.664484
43. Jackson D, White IR, Riley RD. Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. *Statist Med*. 2012;31(29):3805-3820. https://doi.org/10.1002/sim.5453
44. Higgins JPT, Deeks JJ, Atman DG. Special topics in statistics. In: Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. Chichester, UK: John Wiley & Sons Ltd; 2011;chap 16. *The Cochrane Collaboration*. www.cochrane-handbook.org
45. Song F, Clark A, Bachmann MO, Maas J. Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons. *BMC Med Res Methodol*. 2012;12(1):138. https://doi.org/10.1186/1471-2288-12-138
46. Nikolakopoulou A, Chaimani A, Veroniki AA, Vasiliadis HS, Schmid CH, Salanti G. Characteristics of networks of interventions: a description of a database of 186 published networks. *PLoS ONE*. 2014;9(1):e86754. https://doi.org/10.1371/journal.pone.0086754
47. Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JPT. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statist Med*. 2015;34(6):984-998. https://doi.org/10.1002/sim.6381
48. White IR, Barrett JK, Jackson D, Higgins JPT. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Res Synth Meth*. 2012;3(2):111-125.
49. Chen Y, Chu H, Luo S, Nie L, Chen S. Bayesian analysis on meta-analysis of case-control studies accounting for within-study correlation. *Stat Methods Med Res*. 2015;24(6):836-855. https://doi.org/10.1177/0962280211430889

How to cite this article: Efthimiou O, Rücker G, Schwarzer G, Higgins JPT, Egger M, Salanti G. Network meta-analysis of rare events using the Mantel-Haenszel method. *Statistics in Medicine*. 2019;1–21. https://doi.org/10.1002/sim.8158

APPENDIX A

A.1 | Example 1: Inhaled medications in patients with chronic obstructive pulmonary disease

Stydy ID	Treatment	Events	Randomized
1	TIO-SMI	52	1989
1	placebo	38	2002
2	TIO-SMI	35	1337
2	placebo	9	653
3	TIO-HH	381	2986
3	placebo	411	3006
4	TIO-HH	3	266
4	placebo	6	288
5	TIO-HH	0	117
5	placebo	0	117
6	TIO-HH	15	608
6	placebo	4	305
7	TIO-HH	1	69
7	placebo	2	73
8	TIO-HH	7	500
8	placebo	8	510
9	TIO-HH	1	55
9	placebo	0	53
10	TIO-HH	22	914
10	placebo	19	915
11	TIO-HH	7	550
11	placebo	7	317
12	TIO-HH	64	3707
12	LABA	78	3669
13	TIO-HH	1	402
13	LABA	6	405
13	placebo	5	400
14	TIO-HH	38	665
14	LABA-ICS	21	658
15	LABA	6	316
15	placebo	5	318
16	LABA	3	440
16	placebo	0	217
17	LABA	1	201
17	placebo	2	207
18	ICS	5	127
18	placebo	5	127
19	ICS	3	128
19	placebo	0	132
20	ICS	4	123
20	placebo	0	121
21	ICS	32	376
21	placebo	36	375
22	ICS	8	634
22	placebo	10	643
23	ICS	4	145
23	placebo	5	145

(Continues)

(Continued)

24	ICS	0	142
25	ICS	0	434
25	placebo	0	206
26	LABA-ICS	6	479
26	LABA	0	239
27	LABA-ICS	4	394
27	LABA	6	403
28	LABA-ICS	6	394
28	LABA	3	388
29	LABA-ICS	7	507
29	LABA	9	487
30	LABA-ICS	2	189
30	LABA	4	184
31	LABA-ICS	5	92
31	LABA	7	94
32	LABA-ICS	7	988
32	LABA	4	495
32	placebo	4	481
33	LABA-ICS	1	131
33	LABA	0	131
33	placebo	0	125
34	LABA-ICS	2	297
34	placebo	0	148
35	LABA-ICS	7	845
35	LABA	1	284
35	ICS	2	275
35	placebo	1	300
36	LABA-ICS	193	1546
36	LABA	205	1542
36	ICS	246	1552
36	placebo	231	1544
37	LABA-ICS	0	178
37	LABA	0	177
37	ICS	0	183
37	placebo	0	185
38	LABA-ICS	2	358
38	LABA	3	372
38	ICS	3	374
38	placebo	7	361
39	LABA-ICS	5	254
39	LABA	13	255
39	ICS	6	257
39	placebo	5	256
40	LABA-ICS	6	208
40	LABA	6	201
40	ICS	5	198
40	placebo	9	205

(Continues)

(Continued)

41	LABA-ICS	0	165
41	LABA	0	160
41	ICS	0	168
41	placebo	3	181
	Total	**2408**	**52462**

Abbreviations: ICS, inhaled corticosteroid; LABA, long-acting $\beta 2$ agonist; TIO-HH, tiotropium dry powder; TIO-SMI, tiotropium solution.

A.2 | Example 2: Methods to decrease blood loss and blood transfusion requirements for patients with liver transplantation

	Stydy ID	Treatment	Events	Randomized
1	Findlay 2001	Control/Placebo	1	30
2	Findlay 2001	Aprotonin	0	33
3	Garcia-Huete 1997	Control/Placebo	3	41
4	Garcia-Huete 1997	Aprotonin	1	39
5	Porte 2000	Control/Placebo	4	48
6	Porte 2000	Aprotonin	5	89
7	Boylan 1996	Control/Placebo	3	20
8	Boylan 1996	Tranexamic acid	0	25
9	Dalmau 2000	Control/Placebo	2	20
10	Dalmau 2000	Tranexamic acid	3	42
11	Dalmau 2000	EACA	3	42
12	Kaspar 1997	Control/Placebo	0	16
13	Kaspar 1997	Tranexamic acid	1	16
14	Baudo 1992	Control/Placebo	2	16
15	Baudo 1992	Antithrombin III	0	13
16	Lodge 2005	Control/Placebo	1	62
17	Lodge 2005	rFVIIa	3	121
18	Planinsic 2005	Control/Placebo	1	19
19	Planinsic 2005	rFVIIa	5	64
20	Pugliese 2007	Control/Placebo	0	10
21	Pugliese 2007	rFVIIa	0	10
22	Himmelreich 1992	Control/Placebo	0	10
23	Himmelreich 1992	Aprotonin	0	13
24	Williamson 1999	Control/Placebo	0	13
25	Williamson 1999	Solvent detergent plasma	0	12
26	Dalmau 2004	Aprotonin	1	63
27	Dalmau 2004	Tranexamic acid	4	64
28	Ickx 2006	Aprotonin	0	24
29	Ickx 2006	Tranexamic acid	2	27
		Total	**45**	**1002**

Abbreviations: EACA, epsilon amino caproic acid; rFVIIa, recombinant factor VIIa.

APPENDIX B

Fitting the models in R package netmeta

```
library(netmeta)

settings.meta(digits = 2, digits.pval = 2)
options(width = 120)

######################################################################

# Dataset of example 1
data(Dong2013)

p1 = pairwise(treatment, death, randomized, studlab = id,
              data = Dong2013, sm = "OR")

# Conduct Mantel-Haenszel network meta-analysis
MH1 = netmetabin(p1, ref = "Placebo")
# Network graph (Figure 1a)
netgraph(MH1, seq = "optimal", col = "black", plastic = FALSE,
         points = TRUE, pch = 21, cex.points = 3, col.points = "black",
         bg.points = "gray", thickness = "se.fixed",
         multiarm = FALSE, number.of.studies = TRUE)

# Results of network meta-analysis
summary(MH1)
netleague(MH1)
forest(MH1)
# Assess inconsistency
print(netsplit(MH1), show = "both", ci = TRUE, overall = FALSE)

# Conduct NCH network meta-analysis
NCH1 = netmetabin(p1, ref = "Placebo", method = "NCH")
# Results of network meta-analysis
summary(NCH1)
netleague(NCH1)
forest(NCH1)
# Assess inconsistency
print(netsplit(NCH1), show = "both", ci = TRUE, overall = FALSE)


# Conduct inverse variance network meta-analysis
IV1 = netmetabin(p1, ref = "Placebo", method = "Inverse",
                 comb.random = FALSE, allstudies = TRUE)
summary(IV1)


######################################################################

# Dataset of example 2
data(Gurusamy2011)

p2 = pairwise(treatment, death, n, studlab = study,
              data = Gurusamy2011, sm = "OR")
```

```
# Conduct Mantel-Haenszel network meta-analysis
MH2 = netmetabin(p2, ref = "Control/Placebo")
# Results of network meta-analysis
summary(MH2)
netleague(MH2)
forest(MH2)
# Assess inconsistency
print(netsplit(MH2), show = "both", ci = TRUE, overall = FALSE)
```

APPENDIX C

OpenBUGS code used in this paper

A fixed effects Bayesian NMA model can be fitted using the following code.

```
# Conduct NCH network meta-analysis
NCH2 = netmetabin(p2, ref = "Control/Placebo", method = "NCH")
# Results of network meta-analysis
summary(NCH2)
netleague(NCH2)
forest (NCH2)
# Assess inconsistency
print(netsplit(NCH2), show = "both", ci = TRUE, overall = FALSE)
```

```
# Conduct inverse variance network meta-analysis
IV2 = netmetabin(p2, ref = "Control/Placebo", method = "Inverse",
                 comb.random = FALSE)
# Network graph (Figure 1b)
netgraph(IV2, seq = "optimal", col = "black", plastic = FALSE,
         points = TRUE, pch = 21, cex.points = 3, col.points = "black",
         bg.points = "gray", thickness = "se.fixed",
         multiarm = FALSE, number.of.studies = TRUE)
# Results of network meta-analysis
summary(IV2)
```

Data for example 1

```
model {
  for (i in 1:ns) {
    # binomial likelihood of number of events for each arm k of study i

    for (k in 1:na[i]) {
      r[i, k] ~ dbin(p[i, k], n[i, k])
    }

    # parameterization of the 'true' effect of each comparison of arm
    # k vs. baseline arm (1) of study i
    logit(p[i,1])<- u[i]
    for (k in 2:na[i]) {
      logit(p[i, k]) <- u[i] + mean[i, k]
      # consistency equations
      mean[i, k] <- d[t[i, k]] - d[t[i, 1]]
    }
  }
```

```
  # prior distribution for log-odds in baseline arm of study i
  for (i in 1:ns) {
    u[i] ~ dnorm(0, 0.01)
  }

  # prior distribution for heterogeneity
  tau ~ dnorm(0, 1)I(0, )
  prec <- 1 / pow(tau, 2)
  tau.sq <- pow(tau, 2)

  # prior distribution for basic parameters
  for (k in 1:nt) {
    d[k] ~ dnorm(0, 0.01)
  }

  # OR for each comparison

    for(i in 1:(nt - 1)) {
      for (j in (i + 1):nt) {
        OR[j, i] <- exp(d[j] - d[i])
      }
    }
  }
```

Data for example 2

```
list(ns = 41, nt = 6,
     t = structure(.Data = c(
      2,1,NA,NA,2,1,NA,NA,2,1,NA,NA,2,1,NA,NA,3,1,NA,NA,
      4,1,NA,NA,2,1,NA,NA,5,4,2,1,3,4,1,NA,5,4,2,1,
      5,4,2,1,5,4,2,1,5,4,2,1,3,1,NA,NA,3,1,NA,NA,
      3,1,NA,NA,4,1,NA,NA,5,4,NA,NA,4,1,NA,NA,3,1,NA,NA,
      3,1,NA,NA,2,1,NA,NA,5,4,NA,NA,5,1,NA,NA,5,4,2,1,
      3,1,NA,NA,3,1,NA,NA,3,1,NA,NA,3,5,NA,NA,5,4,NA,NA,
      5,4,2,1,2,1,NA,NA,5,4,NA,NA,5,4,1,NA,5,4,NA,NA,
      6,1,NA,NA,3,4,NA,NA,6,1,NA,NA,2,1,NA,NA,5,4,NA,NA,5,4,1,NA),
      .Dim = c(41, 4)),
     na = c(2,2,2,2,2,2,2,4,3,4,4,4,4,2,2,2,2,2,2,2,
      2,2,2,2,4,2,2,2,2,2,4,2,2,3,2,2,2,2,2,2,3),
     n = structure(.Data = c(
      142,139,NA,NA,634,643,NA,NA,145,145,
      NA,NA,376,375,NA,NA,550,317,NA,NA,201,207,NA,NA,123,121,
      NA,NA,165,160,168,181,402,405,400,NA,178,177,183,185,358,
      372,374,361,254,255,257,256,208,201,198,205,500,510,NA,NA,
      55,53,NA,NA,914,915,NA,NA,440,217,NA,NA,189,184,
      NA,NA,316,318,NA,NA,608,305,NA,NA,69,73,NA,NA,128,132,NA,NA,507,487,
      NA,NA,297,148,NA,NA,1546,1542,1552,1544,2986,3006,NA,NA,266,
      288,NA,NA,117,117,NA,NA,665,658,NA,NA,394,388,NA,NA,845,
      284,275,300,127,127,NA,NA,394,403,NA,NA,988,495,481,NA,479,
      239,NA,NA,1989,2002,NA,NA,3707,3669,NA,NA,1337,653,
      NA,NA,434,206,NA,NA,92,94,NA,NA,131,131,125,NA),
      .Dim = c(41, 4)),
     r = structure(.Data = c(
      0,2,NA,NA,8,10,NA,NA,4,5,NA,NA,32,36,NA,NA,7,7,NA,NA,
      1,2,NA,NA,4,0,NA,NA,0,0,0,3,1,6,5,NA,0,0,0,0,2,3,3,7,5,13,
      6,5,6,6,5,9,7,8,NA,NA,1,0,NA,NA,22,19,NA,NA,3,0,NA,NA,2,4,
      NA,NA,6,5,NA,NA,15,4,NA,NA,1,2,NA,NA,3,0,NA,NA,7,9,NA,NA,
      2,0,NA,NA,193,205,246,231,381,411,NA,NA,3,6,NA,NA,0,0,NA,NA,
      38,21,NA,NA,6,3,NA,NA,7,1,2,1,5,5,NA,NA,4,6,NA,NA,7,4,4,NA,
      6,0,NA,NA,52,38,NA,NA,64,78,NA,NA,35,9,NA,NA,0,0,NA,NA,5,7,
      NA,NA,1,0,0,NA),
      .Dim = c(41, 4))
)
```

```
list(ns = 14, nt = 7, ref = 1,
     t = structure(.Data = c(
      1,2,NA,1,2,NA,1,2,NA,1,3,NA,1,3,4,1,3,NA,1,5,NA,1,6,NA,
      1,6,NA,1,6,NA,1,2,NA,1,7,NA,2,3,NA,2,3,NA),
      .Dim = c(14, 3)),
     na = c(2,2,2,2,3,2,2,2,2,2,2,2,2,2),
     n = structure(.Data = c(
      30,33,NA,41,39,NA,48,89,NA,20,25,NA,20,42,42,16,16,NA,
      16,13,NA,62,121,NA,19,64,NA,10,10,NA,10,13,NA,13,12,NA,
      63,64,NA,24,27,NA),
      .Dim = c(14, 3)),
     r = structure(.Data = c(
      1,0,NA,3,1,NA,4,5,NA,3,0,NA,2,3,3,0,1,NA,2,0,NA,1,3,
      NA,1,5,NA,0,0,NA,0,0,NA,0,0,NA,1,4,NA,0,2,NA),
      .Dim = c(14, 3)))
```