CASE STUDY

Component network meta-analysis compared to a matching method in a disconnected network: A case study

Gerta Rücker1 | Susanne Schmitz2 | Guido Schwarzer1

1 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg, Freiburg, Germany
2 Competence Center for Methodology and Statistics, Department of Population Health, Luxembourg Institute of Health, Strassen, Luxembourg

Abstract

Network meta-analysis is a method to combine evidence from randomized controlled trials (RCTs) that compare a number of different interventions for a given clinical condition. Usually, this requires a connected network. A possible approach to link a disconnected network is to add evidence from nonrandomized comparisons, using propensity score or matching-adjusted indirect comparisons methods. However, nonrandomized comparisons may be associated with an unclear risk of bias. Schmitz et al. used single-arm observational studies for bridging the gap between two disconnected networks of treatments for multiple myeloma. We present a reanalysis of these data using component network meta-analysis (CNMA) models entirely based on RCTs, utilizing the fact that many of the treatments consisted of common treatment components occurring in both networks. We discuss forward and backward strategies for selecting appropriate CNMA models and compare the results to those obtained by Schmitz et al. using their matching method. CNMA models provided a good fit to the data and led to treatment rankings that were similar, though not fully equal to that obtained by Schmitz et al. We conclude that researchers encountering a disconnected network with treatments in different subnets having common components should consider a CNMA model. Such models, exclusively based on evidence from RCTs, are a promising alternative to matching approaches that require additional evidence from observational studies. CNMA models are implemented in the R package netmeta.

KEYWORDS

component network meta-analysis, disconnected network, network meta-analysis, matching

1 INTRODUCTION

Network meta-analysis (NMA) is a method to combine evidence from randomized controlled trials (RCTs) that compare a number of different interventions for a given clinical condition to obtain estimates of relative effects between all interventions (Efthimiou et al., 2015). One typical requirement is that the network of interventions is connected. However,
studies included in an NMA may not provide a connected network, but rather a decomposition into two or more separate networks. Such a network is called disconnected. For example, Schmitz et al. (2018) investigated 25 treatments and treatment combinations for multiple myeloma in an NMA of 25 trials. They encountered two separate networks with 15 and 10 treatments (16 and 9 studies).

To analyze connected or disconnected networks, arm-based methods of NMA were suggested (Hawkins, Scott, & Woods, 2016; Hong, Chu, Zhang, & Carlin, 2016; Piepho, Madden, Roger, Payne, & Williams, 2018), however controversially discussed (Dias & Ades, 2016; Hong, Chu, Zhang, & Carlin, 2016b) and recently theoretically and empirically investigated and compared to the contrast-based approach (Karahalios et al., 2017; White, Turner, Karahalios, & Salanti, 2019). Goring et al. (2016) and Béliveau, Goring, Platt, and Gustafson (2017) discuss Bayesian approaches to link disconnected networks. Another possible approach is to add evidence from nonrandomized trials to a contrast-based network, using propensity score or matching-adjusted indirect comparisons (MAIC) methods (Petto, Kadziola, Brnabic, Saure, & Belger, 2019; Phillippo et al., 2017; Signorovitch et al., 2012; Veroniki, Straus, Soobiah, Elliott, & Tricco, 2016). However, nonrandomized comparisons may be associated with an unclear risk of bias, potentially higher than for RCTs.

Schmitz et al. (2018) used an alternative approach based on single-arm observational studies to fill the gap between disconnected networks. They were able to construct five matches between the networks, based on a distance metric derived from covariate information (treatment history, age, baseline stage, and gender), thus connecting the two separate networks. We noticed that many of these treatments were combination treatments, consisting of common treatment components occurring in both networks.

Component network meta-analysis (CNMA) models represent a generalization of NMA models which can be utilized in disconnected networks (Rücker, Petropoulou, & Schwarzer, 2020; Welton, Caldwell, Adamopoulos, & Vedhara, 2009). They can be applied if treatments are combinations of one or more components and if treatments in separate networks share at least one common component. An example of an application of CNMA to a disconnected network (though this is not explicitly mentioned) is given in Pompoli et al. (2018). The purely additive version of the CNMA model assumes that the effect of a treatment consisting of several components is the additive sum of the effects of its components. This model can be further generalized by adding interactions to allow synergistic or antagonistic effects between components.

We present a reanalysis of the data by Schmitz et al. (2018) using suitable CNMA models, utilizing the component character of most of the treatments in this NMA and thus bridging the gap between the two separate networks, and compare the methodological basics and the results of the different approaches. The objectives of this paper are twofold. First, we want to identify CNMA models that provide a good fit to the data at hand. Secondly, we want to compare the results of the CNMA-based approach with those of the matching approach by Schmitz et al. (2018).

The structure of the paper is as follows. We present the data in Section 2, explain the methods in Section 3, show the results in Section 4 and interpret and discuss them in Section 5. Finally, in Section 6 we give recommendations for analyzing such data in practice.

2 | DATA

Multiple myeloma represents the second most common form of blood cancer with an annual incidence of approximately 6 per 100,000 individuals in the United States of America and Europe (Röllig, Knop, & Bornhäuser, 2015). There is no cure and prognosis for relapsed and refractory multiple myeloma remains poor despite improvements through new treatment options in recent years. There is a lack of comparative data between existing treatment options and an unmet need for effective and evidence-based treatment pathways.

The data of the aggregate NMA were provided by the second author (SS) and were partly published along with WinBUGS code in Schmitz et al. (2018, additional file 2 therein). The outcome was progression-free survival, and the relative effects were measured as hazard ratios (HRs). The authors found that the network of treatments was not connected, but split into two separate networks, see Figure 1 which corresponds to Schmitz et al. (2018, fig. 2 therein). Following Schmitz et al. (2018), we call the two parts of the network the “white network” (15 treatments, evidence from 16 RCTs) and the “black network” (10 treatments, 9 RCTs). The data are available from the Supporting Information of this paper.
METHODS

In our paper, we reanalyze the data and compare the results of three approaches to analyzing this disconnected network: (a) The separate analysis of the white and black network, (b) the matching method applied by Schmitz et al. (2018) in the original analysis, and (c) a number of CNMA models with and without considering different treatment interaction structures. We also address issues of model selection. All approaches are briefly described in the following sections, with technical details of the CNMA method provided in the Appendix. For additional details on the methodology, see Schmitz et al. (2018) and Rücker, Petropoulou et al. (2020).

3.1 Separate analyses

In a first step, we analyzed both networks separately with the common effects model (also termed fixed effects model) using the frequentist NMA method implemented in R package \texttt{netmeta} (R Core Team, 2020). Results were compared with the outcomes of the standard Bayesian analysis of the separate networks in Schmitz et al. (2018).

3.2 Matching method

Evidence from observational studies was sought to connect the disconnected network. In the absence of comparative observational studies, single-arm studies were considered for matching, where a full covariate profile allowed for the assessment of similarity between studies using a distance metric based on covariate profiles. Single-arm studies with a small distance were thought sufficiently similar for matching. Of the eight identified pairs (called matches), three referred to comparisons within one network (one for the white and two for the black network); while five matches compared a treatment from the white and a treatment from the black network, thus connecting both networks (Schmitz et al., 2018, fig. 4 therein). The scenario considered for comparison in this paper refers to the outcomes of a Bayesian network analysis incorporating RCT data from both networks as well as the five matches connecting both networks. Models were fitted in \texttt{WinBUGS} using the \texttt{R2WinBUGS} package in R (R Core Team, 2020; Sturtz, Ligges, & Gelman, 2005). Further details can be found in Schmitz et al. (2018).

3.3 CNMA models

We follow the notation in Rücker, Petropoulou et al. (2020).
3.3.1 Additive CNMA model

The additive model assumes that the effect of each combined treatment is the additive sum of its components, that is, equal components cancel out in pairwise comparisons. Let the data consist of $m$ pairwise comparisons of $n$ treatments some of which are multicomponent treatments. Let the number of components be $c$ with $c \leq n$. Each comparison $j = 1, \ldots, m$ is represented by an observed (relative) treatment effect $d_j$ with standard error SE($d_j$). The standard errors of multiarm studies are adjusted as described in Rücker and Schwarzer (2014).

The design matrix of the additive model is the $n \times c$ matrix

$$X = BC. \tag{1}$$

Matrix $X$ combines information on the structure of the network ($B$ with $m$ rows, representing the pairwise comparisons, and $n$ columns, representing the treatments) and information how the $n$ treatments are composed of the $c$ components ($n \times c$ matrix $C$). The elements of $X$ are 0, 1, or $-1$ and represent the components (columns) of the comparison in the row.

The additive model (common effects version) is given by

$$d = X\beta + \epsilon,$$

where $d \in \mathbb{R}^m$ is the vector of observed relative effects (differences) from the studies, $X$ the design matrix given in (1), $\beta \in \mathbb{R}^c$ a parameter vector of length $c$, representing the components, and $\epsilon \in \mathbb{R}^m$ is a multivariate normally distributed error. $\beta$ is estimated using weighted least squares regression. For details of the estimation, see Appendix; more information including a generalization to a random effects model is found in Rücker, Petropoulou et al. (2020).

3.3.2 Interaction CNMA models

The additive CNMA model can be extended to a two-way interaction CNMA model (Welton et al., 2009, Model 3) where the combination of components may provide larger or smaller effects than the sum of their effects (Welton et al., 2009). Interaction CNMA models are implemented by adding a further column to the combination matrix $C$ for each interaction of interest. In principle, for each treatment combination that exists in a data set an interaction term can be added. If all existing combinations are added, this leads to the standard NMA model. Nested models can be compared using the multivariate version of Cochran’s $Q$ as described in Rücker, Petropoulou et al. (2020).

3.3.3 CNMA models for disconnected networks

Under certain circumstances, CNMA models allow reconnecting a disconnected network. For this to work, at least some treatments must consist of components, and the subnets must have a sufficient number of common components. A simple hypothetical example was given in Rücker, Petropoulou et al. (2020).

3.3.4 Model selection

In our application, we conducted a model selection in two directions.

Forward selection means starting from the purely additive model and systematically adding in turn each two-way and three-way interaction that was observed in the data to the additive model. We also tested all models that combined two or more of these interaction terms. Among these models, the final model was selected based on Cochran’s $Q$ statistic.

Alternatively, we investigated backward selection. Ideally, this would mean starting from a standard NMA and replacing some, but not all interactions with an additive assumption about their components. As a standard NMA is impossible in a disconnected network, we started by assuming additivity for just one component that is common to both parts of the network and whose separation is sufficient to connect the subnetworks. We will discuss this approach including its drawbacks.
TABLE 1  P-scores for the white and the black network

| White network      | P-score | Black network      | P-score |
|--------------------|---------|--------------------|---------|
| dara + dex + len   | 1.0000  | dara + bor + dex   | 0.9784  |
| carf + dex + len   | 0.8794  | carf + dex         | 0.8434  |
| ixa + dex + len    | 0.8510  | thal + bor + dex   | 0.7418  |
| elo + dex + len    | 0.7908  | bor + dex + pan    | 0.6261  |
| bor + PLD          | 0.6546  | elo + bor + dex    | 0.5395  |
| dex + len          | 0.6098  | thal + dex         | 0.4713  |
| bor + bev          | 0.5516  | bor + dex + sil    | 0.3367  |
| pom + dex          | 0.3378  | bor + dex          | 0.2685  |
| bor + vor          | 0.4820  | bor + dex + cyc    | 0.1532  |
| bor                | 0.3873  | bor + dex + peri   | 0.0412  |
| pom                | 0.2842  |                    |         |
| thal               | 0.2334  |                    |         |
| dex                | 0.1383  |                    |         |
| ob + dex           | 0.0981  |                    |         |
| thal + IFN         | 0.0017  |                    |         |

3.3.5  Implementation

CNMA models for connected networks are implemented in function netcomb() of the R package netmeta for NMA (R Core Team, 2020; Rücker, Krahn, König, Efthimiou, & Schwarzer, 2020). The analysis of additive and interaction CNMA models for disconnected networks is implemented in function discomb() of the R package netmeta (Rücker, Krahn et al., 2020). R code for the analyses in this paper is available from the Supporting Information.

4  RESULTS

For our case study, the multiple myeloma data, we observed 18 treatment components in total, six of which occurred in both subnets (bor, carf, dara, dex, elo, thal), eight occurred in the white network only (bev, IFN, ixa, len, ob, PLD, pom, vor), and four components occurred in the black network only (cyc, pan, peri, sil). The large number of common components motivated us to apply CNMA models for disconnected networks to these data.

4.1  Separate analyses

We first analyzed the two networks separately using netmeta and obtained results that were very similar to those obtained by Schmitz et al. (2018). Our results are shown as forest plots in Figures S1 and S2. We note that the results did not depend on whether the common or the random effects model was used, because the heterogeneity variance parameter \( \tau \) was estimated to be zero in the white network and could not be estimated in the black network because there was only one study for each comparison.

Table 1 shows a within-network ranking using P-scores (Rücker & Schwarzer, 2015). P-scores are mean ranks, scaled such that 1 is best and 0 is worst and thus corresponding to the Bayesian "surface under the cumulative ranking curve" (SUCRA) values (Salanti, Ades, & Ioannidis, 2011).

For the white network, treatments were compared to thal+IFN, which was the worst treatment with respect to the P-score (0.0017). Treatment dara+dex+len was superior to all others (P-score 1.00). For the black network, the reference treatment was bor+dex+peri (P-score 0.0412); the best treatment here was dara+bor+dex (P-score 0.9784).

4.2  Matching method

Figure 2 shows the results of the matching method by Schmitz et al. (2018) as a forest plot (random effects model). At the top, they found the best treatment from the white network (dara+dex+len), followed by the top treatment...
of the black network (dara+bor+dex). The bottoms were dex, ob+dex, and thal+IFN, which serves as the reference treatment.

### 4.3 Additive CNMA model (Model A)

The result of the additive model (random effects model, model A) is shown in Supplementary Figure S3. The test of heterogeneity or inconsistency provided $Q = 27.78$ ($df = 7$, $p = .0002$). At the top, we found the best two treatments of the white network (dara+dex+len, carf+dex+len), followed by the top treatment of the black network, dara+bor+dex, and the next two treatments from the white network, ixa+dex+len and elo+dex+len. The bottoms were ob+dex (white), bor+dex+peri (black), and thal+IFN (white), as expected. The order of treatments within each network was only slightly changed for the black network (only one transposition between bor+dex+pan and thal+bor+dex), but considerably for the white network. We investigate this in more detail in the next subsection.

### 4.4 Interaction models (forward selection)

Starting from the additive model, we added interaction terms in a systematic model selection procedure.

#### 4.4.1 Adding single two-way interactions

We had data from 10 combinations of two treatments, corresponding to 10 models, each defined as the additive model plus one two-way interaction (bor*bev, bor*dex, bor*PLD, bor*vor, carf*dex, dex*len, ob*dex, pom*dex, thal*dex, and thal*IFN). Of these models, briefly denoted by the included interaction term, only three (bor*dex, pom*dex, thal*dex) led to a reduction of the degrees of freedom and a reduction of $Q$. This was most remarkable when adding bor*dex ($Q = 9.40$, $p = .002$). The best two treatments were dara+bor+dex and carf+dex+len. The bottom was ob+dex (white), bor+dex+peri (black), and thal+IFN (white), as expected. The order of treatments within each network was only slightly changed for the black network (only one transposition between bor+dex+pan and thal+bor+dex), but considerably for the white network. We investigate this in more detail in the next subsection.
RÜCKER ET AL.

**Table 2** Separate NMA models, additive CNMA model, and a number of interaction CNMA models

| Model            | Interaction terms                                | Q   | df | p-value |
|------------------|--------------------------------------------------|-----|----|---------|
| White network    | —                                                | 0.30| 2  | .8595   |
| Black network    | —                                                | 0.00| 0  | —       |
| A                | —                                                | 27.78| 7  | .0002   |
| B                | bor*dex                                          | 9.40| 6  | .1523   |
|                  | pom*dex                                          | 11.72| 6  | .0686   |
|                  | thal*dex                                         | 26.12| 6  | .0002   |
| C                | carf*dex*len                                     | 9.47| 6  | .1487   |
|                  | dara*bor*dex                                     | 27.24| 6  | .0001   |
|                  | dara*dex*len                                     | 27.24| 6  | .0001   |
|                  | elo*bor*dex                                      | 27.76| 6  | .0001   |
|                  | elo*dex*len                                      | 27.76| 6  | .0001   |
|                  | thal*bor*dex                                     | 26.40| 6  | .0002   |
|                  | thal*dex + pom*dex                               | 7.91| 5  | .1612   |
|                  | thal*dex + thal*dex                              | 7.91| 5  | .1612   |
| D                | bor*dex + carf*dex*len                           | 0.91| 5  | .9699   |
|                  | bor*dex + thal*bor*dex                           | 4.09| 5  | .5370   |
|                  | pom*dex + carf*dex*len                           | 4.67| 5  | .4577   |
|                  | pom*dex + thal*bor*dex                           | 4.67| 5  | .4577   |
|                  | carf*dex*len + thal*bor*dex                      | 4.67| 5  | .4577   |
|                  | thal*dex + carf*dex*len                          | 6.82| 5  | .2341   |
|                  | thal*dex + thal*bor*dex                          | 24.89| 5  | .0001   |
| E                | bor*dex + carf*dex*len + thal*bor*dex            | 0.86| 4  | .9301   |
|                  | pom*dex + carf*dex*len + thal*bor*dex            | 4.67| 5  | .4577   |
| F                | bor*dex + carf*dex*len + thal*bor*dex + dara*bor*dex | 0.33| 3  | .9553   |
| S                | bor*dex + carf*dex*len + thal*bor*dex + dara*bor*dex + elo*bor*dex | 0.30| 2  | .8595   |
|                  | Separating component thal or carf                | 0.30| 2  | .8595   |

$df = 6, p = .1523$), followed by the model with pom*dex ($Q = 11.72, df = 6, p = .0686$) (see Table 2). The addition of thal*dex alone did not improve the model fit. Figure S4 shows the forest plot for model bor*dex which we call model B.

Depending on the network structure, adding an interaction does not always lead to a new model. For example, adding the interaction dex*len to the additive model just results in the additive model—the number of parameters has seemingly increased, but the degrees of freedom has not been reduced. Looking at the network graphs reveals why: the combination dex+len occurs only either in studies where it is given in both arms, such that its effect cancels out, or in studies where it is compared to dex alone, such that they estimate the effect of len (for the additive model) or the effect of len + dex*len (for the interaction model). Because there are no other studies investigating len, the interaction model cannot separate dex*len from len alone, and the model is overparameterized. Formally, the design matrix has not full rank: the column that represents the added interaction (here dex*len) happens to be a linear combination of other columns (here it is simply equal to column len). We point to Appendix for mathematical details.

### 4.4.2 Adding single three-way interactions

We observed 11 combinations of three treatments, each corresponding to a model with one three-way interaction added to the additive model (bor*dex*cyc, bor*dex*pan, bor*dex*peri, bor*dex*sil, carf*dex*len, dara*bor*dex, dara*dex*len, elo*bor*dex, elo*dex*len, ixa*dex*len, thal*bor*dex). Of these models, six led to a reduction of the degrees of freedom,
but only one (carf*dex*len, referred to as model C) considerably improved the model fit ($Q = 9.47, df = 6, p = .1487$) (Table 2). Figure S5 shows the forest plot for this model.

### 4.4.3 Combining two or more interaction terms

Additional models were defined by combining two interaction terms (see Table 2). The best model fit with two interaction terms was given by Model D, which adds bor*dex + carf*dex*len ($Q = 0.91, df = 5, p = .9699$). Figure S6 shows the forest plot for this model. Combining three interaction terms, as in model E (bor*dex + carf*dex*len + thal*bor*dex) did not further significantly improve the model. Based on Q, we interpret model D as our final model.

Nevertheless, it was possible to reduce Q further by adding more interaction terms (one or two out of dara*bor*dex, elo*bor*dex, dara*dex*len or elo*dex*len) until it reached its lower bound, given by the sum of the Q values of the separate networks, here $Q = 0.30 (df = 2, p = .8595$, see, for example, Table 2, Models F and S). We call model S the saturated model because Q cannot fall below this limit, as the separate NMAs constitute the weighted least squares solutions and thus the best fit for each network alone which cannot be amended. The forest plot is shown in Figure 3. Results for models E and F are given in Figures S7 and S8.

### 4.5 Interaction models (backward selection)

Backward selection means to start with a model that is as rich as possible and to omit interactions to achieve more parsimonious models. For a connected network, the standard NMA would be a natural starting point, but for a disconnected network this does not exist. However, we may connect the separate networks by making an additive assumption for just one component that is observed in all subnetworks. For our example, thal is such a component. It occurs in the white network as a single treatment and in combination with IFN, in the black network in the combinations thal+dex and...
### Table 3: Pairwise comparisons between selected models

| Model | Compared to model | Q | df | p-Value |
|-------|-------------------|---|----|---------|
| A     |                   |   |    |         |
| B bor*dex |                   | 18.38 | 1 | .0000   |
| C carf*dex*len |               | 18.31 | 1 | .0000   |
| D bor*dex + carf*dex*len |             | 26.87 | 2 | .0000   |
| E bor*dex + carf*dex*len + thal*bor*dex |       | 26.92 | 3 | .0000   |
| F bor*dex + carf*dex*len + thal*bor*dex + dara*bor*dex |   | 27.45 | 4 | .0000   |
| S bor*dex + carf*dex*len + thal*bor*dex + dara*bor*dex + elo*bor*dex | | 27.47 | 5 | .0000   |
| B     |                   |   |    |         |
| D bor*dex + carf*dex*len |             | 8.50 | 1 | .0036   |
| E bor*dex + carf*dex*len + thal*bor*dex |       | 8.54 | 2 | .0140   |
| F bor*dex + carf*dex*len + thal*bor*dex + dara*bor*dex |       | 9.08 | 3 | .0283   |
| S bor*dex + carf*dex*len + thal*bor*dex + dara*bor*dex + elo*bor*dex |   | 9.10 | 4 | .0587   |
| C     |                   |   |    |         |
| D bor*dex + carf*dex*len |             | 8.57 | 1 | .0034   |
| E bor*dex + carf*dex*len + thal*bor*dex |       | 8.61 | 2 | .0135   |
| F bor*dex + carf*dex*len + thal*bor*dex + dara*bor*dex |       | 9.15 | 3 | .0274   |
| S bor*dex + carf*dex*len + thal*bor*dex + dara*bor*dex + elo*bor*dex |   | 9.17 | 4 | .0570   |
| D     |                   |   |    |         |
| E bor*dex + carf*dex*len + thal*bor*dex |       | 0.04 | 1 | .8335   |
| F bor*dex + carf*dex*len + thal*bor*dex + dara*bor*dex |       | 0.58 | 2 | .7483   |
| S bor*dex + carf*dex*len + thal*bor*dex + dara*bor*dex + elo*bor*dex |   | 0.60 | 3 | .8959   |
| D     |                   |   |    |         |
| E bor*dex + carf*dex*len + thal*bor*dex |       | 0.54 | 1 | .4642   |
| F bor*dex + carf*dex*len + thal*bor*dex + dara*bor*dex |       | 0.56 | 2 | .7565   |
| S bor*dex + carf*dex*len + thal*bor*dex + dara*bor*dex + elo*bor*dex |   | 0.02 | 1 | .8815   |

The idea is to separate the component thal from its compositions. Then IFN remains as another component, while dex (white network) and bor+dex (black network) are already present. Figure S9 shows the result of this model which is similar, but not equal to the saturated model S (Figure 3). It also minimizes Q. Alternatively, we tried a different model by separating component carf, which is present in combination carf+dex+len (white network) and carf+dex (black network). This model also minimized Q, but it provided a very different result, see Figure S10.

### 4.6 Comparing all models

Table 3 shows the results of Q tests (corresponding to likelihood ratio tests) comparing the best models with one, two, or more interactions to the additive model and the models they are nested in. We see that all interaction models in Table 3 fit the data better than the additive model and that model bor*dex + carf*dex*len further improves the fit, whereas the model with three interactions does not significantly improve the fit. If we prefer limiting the number of parameters, model D (bor*dex + carf*dex*len) is a good candidate. Models E, F, S, and T provide even better fit in terms of Q.

Figure 4 shows the effects of all treatments, compared to the reference of the white network (thal + IFN) for nine different models (from right to left): the matching method by Schmitz et al. (2018), the purely additive model A, models B, C, D, E, F, the saturated model S, and the NMA of the white network. A forest plot comparing all models, including 95% confidence intervals, is provided in the supplementary material (Figure S11). Note that for both plots, we use thal+IFN from the white network as reference treatment, and for the NMA of the white network only comparisons within this network can be shown.
Figure 5 shows the effects of all treatments, compared to the reference of the black network (bor+dex+peri) for nine different models: the NMA of the black network and the same models as in Figure 4 except the NMA of the white network.

5 | DISCUSSION

In this paper, we have shown how CNMA can be used to reconnect a disconnected network, provided the interventions in the separate parts consist of sufficiently many common components. We also compared the CNMA models to a matching approach based on single-arm observational studies.

The differences between the results were not disturbingly large. Particularly, all models identified dara+dex+len as the most effective and thal+IFN as the least effective intervention. The matching method tended to effect estimates with less uncertainty and farther from 1, compared to the purely additive CNMA model A and the interaction model C, while models B to F, S, and T led to results similar to the matching method.

5.1 | Assumptions of the approaches

In our case study, we do not know the truth. All approaches rely on certain assumptions. The additive CNMA model assumes additivity for the component effects in multicomponent treatments, while CNMA interaction models relax the additivity assumption for some or almost all of these interventions. The matching method assumes that the added
observational studies are comparable with respect to the covariate distribution in the study populations and thus makes an assumption of no unmeasured confounders. Moreover, all models rely on the typical assumptions of all NMA approaches, such as transitivity of effect moderators across different studies.

5.2 Matching approach

The matching approach seeks to connect the subnets by finding additional evidence from single-arm observational studies that are suitable to fill the gap. Schmitz et al. (2018, table 5 and fig. 4 therein) found a number of such connections. Thus they created loops in the otherwise tree-shaped network and could use standard Bayesian NMA to analyze the enriched data set. They did not make any additivity assumptions, in contrast to our CNMA approach. In other words, their model included all observed interactions.

5.3 CNMA forward selection

We applied a systematic model selection method where we added single interaction terms one by one, identifying two interactions (bor*dex, carf*dex*len) that notably decreased heterogeneity, that is, these models (models B, C) fit the data better than the additive model A. We then looked at all combinations of two interaction terms and identified model D (bor*dex + carf*dex*len) as an even better fit. While adding more interaction terms did not significantly increase the fit, it was possible to reduce Q to its minimum by adding three further terms (models E, F, S).
As in general, the model fit improved as more interaction terms were included in the model. For a connected network, the standard NMA model often corresponds to a CNMA model that includes all observed interactions. An exception would be if some components occur only in a fixed combination. In this case, a CNMA model that separates these components, and also includes all interactions, can have more parameters than the standard NMA model and may be not identifiable. A good fit for an interaction model (measured in terms of the likelihood ratio statistic $Q$) is associated with large similarity of the estimates from this model to those from the separate analyses of each subnetwork alone. This is seen for the white network in Figures 4 and S11 and for the black network in Figure 5. In this sense, the separate analyses act as a benchmark, the more so as they rely on RCT evidence without additive assumptions on components. Note, however, that in our particular example we have only little information on inconsistency due to the tree structure of the separate networks.

In general, in a disconnected network we have less information about direct comparisons than usually in NMA, because for any pair of treatments from different connectivity components (subnets), direct comparisons are not observed, and indirect comparisons can be estimated only based on a CNMA model. Notably, comparisons between networks cannot contribute to $Q$. We are not aware of a discussion of the special challenge of assessing inconsistency in the situation of a disconnected network.

It does not make sense to add interactions that do not occur as combinations in the network, as there is no information to estimate them. We note that this holds for standard NMA as well. On the other hand, additive models, if plausible, have the potential to make predictions on how unseen combinations may work (Pompoli et al., 2018).

### 5.4 CNMA backward selection

The backward selection approach can be seen as an attempt to mimic a standard NMA. In our example, it was possible to connect the network by separating just a single component (thal). However, model selection involves a trade-off between model fit and sparseness. While in many applications sparse models are preferred to avoid overfitting, in CNMA of a disconnected network this at first seems to be less an issue: one would have used standard NMA, had the network been connected, and, of note, standard NMA would have automatically included all observed interactions. In principle, backward selection offers a way to do “almost” standard NMA without making too many additivity assumptions.

### 5.5 Limitations

Backward selection has important drawbacks in CNMA. When separating out a single component, the connection between the networks is still loose. Therefore different models, even when providing the same $Q$, can lead to very different results, as we have seen when separating thal and carf. The explanation is that models of this type may reproduce the separate NMA within each network (thus minimizing $Q$), but provide little information for estimating the contrasts across the networks (that do not contribute to $Q$, as they are not observed). These unknown contrasts are typically informed by only few studies. For example, in the model that separates carf, there is only one connection between the white and the black subnetwork, the comparison carf+dex+len (white) versus carf+dex (black) which provides another estimate of dex+len vs dex. Thus, all across-network estimations depend on just two studies. In other words, all across-network information flows through this bottleneck.

This points to a limitation of the $Q$ criterion: A small value of $Q$ means a good fit within the subnetworks, but does not necessarily indicate valid effect estimates between the networks. The actual trade-off is between goodness of fit (measured by $Q$, only measurable within the subnetworks) and connectedness (maximized by the CNMA model with the largest degrees of freedom, which is the additive model). For this reason, we prefer forward selection, based on a stepwise reduction of $Q$ only as long as the model fit substantially improves. In our example, this strategy led to model D as the final model (Table 3). For the future, as a next step, we plan a comprehensive simulation study that may shed more light on this.

### 6 CONCLUSION

CNMA models allow estimating effects of treatment components of multicomponent interventions. This means they can borrow strength from studies with common components. Particularly, they allow bridging gaps between disconnected networks, if these have common components. Though the critical assumption of additivity should be explored, we
recommend that CNMA models be used more commonly. Specifically, they should be considered when encountering disconnected networks. Where necessary and appropriate, interaction terms may be introduced in a principled way.

Though backward selection seems a particularly attractive way to mimic a standard NMA, we recommend to start with an additive model and add some interactions until a satisfactory fit is reached, preferably driven by subject-matter knowledge. To apply CNMA in clinical practice, we recommend to always consult medical experts, such that clinical knowledge impacts decision making on whether additivity can be assumed, or which interactions should be included.

ACKNOWLEDGMENT
G.R. was funded by DFG (German Research Foundation), Grant number RU1747/1-2.

CONFLICT OF INTEREST
The authors have declared no conflict of interest.

OPEN RESEARCH BADGES
This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge “Reproducible Research” for making publicly available the code necessary to reproduce the reported results. The results reported in this article were reproduced partially due to their computational complexity.

REFERENCES
Albert, A. E. (1972). *Regression and the Moore-Penrose pseudoinverse*. Mathematics in Science and Engineering. New York: Academic Press. ISBN: 0-12-048450-1.

Béliveau, A., Goring, S., Platt, R. W., & Gustafson, P. (2017). Network meta-analysis of disconnected networks: How dangerous are random baseline treatment effects? *Research Synthesis Methods, 8*(4), 465–474. https://doi.org/10.1002/jrsm.1256

Dias, S., & Ades, A. E. (2016). Absolute or relative effects? Arm-based synthesis of trial data. *Research Synthesis Methods, 7*(1), 23–28. https://doi.org/10.1002/jrsm.1184. No abstract available.

Efthimiou, O., Debray, T. P. A., van Valkengoed, G., Trelle, S., Panayidou, K., Moons, K. G. M., … GetReal Methods Review Group (2015). GetReal in network meta-analysis: A review of the methodology. *Research Synthesis Methods, 7*(3), 236–263. https://doi.org/10.1002/jrsm.1195

Goring, S., Gustafson, P., Liu, Y., Saab, S., Cline, S., & Platt, R. (2016). Disconnected by design: analytic approach in treatment networks having no common comparator. *Research Synthesis Methods, 7*(4), 420–432. https://doi.org/10.1002/jrsm.1204

Hawkins, N., Scott, D., & Woods, B. (2016). “Arm-based” parameterization for network meta-analysis. *Research Synthesis Methods, 7*(3), 306–313. https://doi.org/10.1002/jrsm.1187. Epub 2015 Nov 27.

Hong, H., Chu, H., Zhang, J., & Carlin, B. P. (2016a). A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Research Synthesis Methods, 7*(1), 6–22. https://doi.org/10.1002/jrsm.1153

Hong, H., Chu, H., Zhang, J., & Carlin, B. P. (2016b). Rejoinder to the discussion of “A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons,” by S. Dias and A. E. Ades. *Research Synthesis Methods, 7*(1), 29–33. https://doi.org/10.1002/jrsm.1186

Karalhalios, A., Salanti, G., Turner, S., Herbison, G., White, I., Veroniki, A., … Mckenzie, J. (2017). An investigation of the impact of using different methods for network meta-analysis: A protocol for an empirical evaluation. *Systematic Reviews, 6*(1), 119. https://doi.org/10.1186/s13643-017-0511-x

Petto, H., Kadziola, Z., Brnabic, A., Saure, D., & Belger, M. (2019). Alternative weighting approaches for anchored matching-adjusted indirect comparisons via a common comparator. *Value in Health, 22*(1), 85–91. Retrieved from http://www.sciencedirect.com/science/article/pii/S1098301518322708.

Phillippo, D. M., Ades, A. E., Dias, S., Palmer, S., Abrams, K., & Welton, N. J. (2017). Methods for population-adjusted indirect comparisons in health technology appraisal. *Medical Decision Making, 38*(2), 200–211.

Piepho, H. P., Madden, L., Roger, J., Payne, R., & Williams, E. (2018). Estimating the variance for heterogeneity in arm-based network meta-analysis. *Pharmaceutical Statistics, 17*, 264–277. https://doi.org/10.1002/pst.1857

Pompoli, A., Furukawa, T. A., Efthimiou, O., Imai, H., Tajika, A., & Salanti, G. (2018). Dismantling cognitive-behaviour therapy for panic disorder: A systematic review and component network meta-analysis. *Psychological Medicine, 48*(12), 1945–1953. https://doi.org/10.1017/S0033291717003919

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Röllig, C., Knop, S., & Bornhäuser, M. (2015). Multiple myeloma. *Lancet, 385*(9983), 2197–2208.

Rücker, G., Krahne, U., König, J., Efthimiou, O., & Schwarz, G. (2020). netmeta: Network meta-analysis using frequentist methods. R package version 1.2-0. Retrieved from https://cran.r-project.org/web/packages/netmeta/
Rücker, G., Petropoulou, M., & Schwarzer, G. (2020). Network meta-analysis of multicomponent interventions. *Biometrical Journal, 62*(3), 808–821. https://doi.org/10.1002/bimj.201800167

Rücker, G., & Schwarzer, G. (2014). Reduce dimension or reduce weights? Comparing two approaches to multi-arm studies in network meta-analysis. *Statistics in Medicine, 33*, 4353–4369. https://doi.org/10.1002/sim.6236

Rücker, G., & Schwarzer, G. (2015). Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Medical Research Methodology, 15*, 58. https://doi.org/10.1186/s12874-015-0060-8

Salanti, G., Ades, A. E., & Ioannidis, J. P. (2011). Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: An overview and tutorial. *Journal of Clinical Epidemiology, 64*(2), 163–171. https://doi.org/10.1016/j.jclinepi.2010.03.016

Schmitz, S., Maguire, Á., Morris, J., Ruggeri, K., Haller, E., Kuhn, I., … Walsh, C. (2018). The use of single arm observational data to closing the gap in otherwise disconnected evidence networks: A network meta-analysis in multiple myeloma. *BMC Medical Research Methodology, 18*(1), 66. https://doi.org/10.1186/s12874-018-0509-7

Signorovitch, J. E., Sikirica, V., Erder, M. H., Xie, J., Lu, M., Hodgkins, P. S., … Wu, E. Q. (2012). Matching-adjusted indirect comparisons: A new tool for timely comparative effectiveness research. *Value in Health, 15*(6), 940–947. Retrieved from http://www.sciencedirect.com/science/article/pii/S1098301512016105

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from *R. Journal of Statistical Software, 12*(3), 1–16.

Veroniki, A. A., Straus, S. E., Soobiah, C., Elliott, M. J., & Tricco, A. C. (2016). A scoping review of indirect comparison methods and applications using individual patient data. *BMC Medical Research Methodology, 16*(1), 47. https://doi.org/10.1186/s12874-016-0146-y

Welton, N. J., Caldwell, D. M., Adamopoulos, E., & Vedhara, K. (2009). Mixed treatment comparison meta-analysis of complex interventions: Psychological interventions in coronary heart disease. *American Journal of Epidemiology, 169*(9), 1158–1165. https://doi.org/10.1093/aje/kwp014

White, I. R., Turner, R. M., Karahalios, A., & Salanti, G. (2019). A comparison of arm-based and contrast-based models for network meta-analysis. *Statistics in Medicine, 38*(7), 5197–5213. https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8360

**SUPPORTING INFORMATION**

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

**How to cite this article:** Rücker G, Schmitz S, Schwarzer G. Component network meta-analysis compared to a matching method in a disconnected network: A case study. *Biometrical Journal*. 2021;63:447–461. https://doi.org/10.1002/bimj.201900339

**APPENDIX**

We briefly provide some details of the estimation method where we closely follow Rücker, Petropoulou et al. (2020). The common (fixed) effect model is

\[ \mathbf{d} = \mathbf{X}\beta + \epsilon, \]

where \( \mathbf{d} \in \mathbb{R}^m \) is the vector of observed relative effects (differences) from the studies, \( \mathbf{X} \) the design matrix given in (1), \( \beta \in \mathbb{R}^c \) a parameter vector of length \( c \), representing the components, and \( \epsilon \in \mathbb{R}^m \) is a vector of multivariate normally distributed errors.

Based on the inverse variance weights \( s_i^{-1} \) (\( i = 1, \ldots, m \)) of the observed effects (potentially adjusted for multiarm studies), we define a diagonal weight matrix \( \mathbf{W} = \text{diag}(s_1^{-1}, \ldots, s_m^{-1}) \) of dimension \( m \times m \). A weighted least squares estimate of \( \beta \) is then obtained by

\[ \hat{\beta} = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^+ \mathbf{X}^\top \mathbf{W} \mathbf{d} \]

with estimated covariance matrix

\[ \hat{\text{Cov}}(\hat{\beta}) = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^+. \]

Here, \((\mathbf{X}^\top \mathbf{W} \mathbf{X})^+\) denotes the Moore–Penrose pseudoinverse of the matrix \( \mathbf{X}^\top \mathbf{W} \mathbf{X} \) (Albert, 1972).
The heterogeneity statistic for the component network meta-analysis (CNMA) model compares the model-based effect estimates with the observed effects. It is defined as

\[ Q = (d - X\hat{\beta})^\top W (d - X\hat{\beta}), \]

where \(X\hat{\beta}\) denotes the vector of effect estimates based on the CNMA model as given above. Under the usual assumption of known and fixed standard errors, \(Q\) approximately follows a Chi-squared distribution with

\[ df = n_a - k - \text{rank}(X) \]

degrees of freedom, where \(n_a\) is the total number of arms in the network and \(k\) is the number of studies. Particularly, if there are only two-arm trials, \(Q\) has \(k - \text{rank}(X)\) degrees of freedom. We note that differences of \(Q\) statistics from nested CNMA models are likewise Chi-square distributed, with the degrees of freedom being the difference of the \(df\)s of the models. This allows to compare nested CNMA models to each other and particularly to the additive model (as the most parsimonious model) on the one hand and, in the case of a connected network, to a standard network meta-analysis (NMA) model on the other hand.

Remarks. The rank of the design matrix \(X\) is crucial for identification of the model. Given a number \(c\) of components, the components can be uniquely identified only if the rank of \(X\) is \(c\). If the rank is less than \(c\), the model is not fully identifiable. Such models can simplified without loss by deleting some components that are linear combinations of others until the number of components corresponds to the rank.

We note that \(Q\) is a sum over observed contrasts within the subnetworks. It is not affected by any between-subnetwork effects estimated by the model. Therefore, there can be different models with the same \(Q\). For example, \(Q\) is minimized by all models that reduce to the standard NMAs for all subnetworks.

In a standard network graph, the number of degrees of freedom \(df\) corresponds to the number of closed loops (multiple studies with the same design included with their multiplicity). With respect to a model, \(df\) can be interpreted as a measure of connectedness.