Aus dem Institut für Medizinische Biometrie
Universitätsklinikum Heidelberg
Geschäftsführender Direktor: Prof. Dr. sc. hum. Meinhard Kieser

Network meta-analysis for the integrated evaluation of targeted therapies

Inauguraldissertation
zur Erlangung des
Doctor scientarium humanarum
an der Medizinischen Fakultät Heidelberg
der Ruprecht-Karls-Universität

vorgelegt von
Tanja Proctor
aus
Freudenstadt

2022

Dekan:	Prof. Dr. med. Hans-Georg Kräusslich
Doktorvater:	Prof. Dr. sc. hum. Meinhard Kieser

Contents

Abbreviations and Symbols

- AD: Aggregate data
- b: Baseline treatment
- β: Regression coefficient
- BM_+: Biomarker-positive
- BM_-: Biomarker-negative
- C: Control treatment
- d: Treatment effect
- $(d)^+$: Treatment effect for biomarker-positive patients
- δ_j: Study-specific treatment effect
- DGM: Data generation model
- E_+: Targeted treatment
- EGFR: Epidermal growth factor receptor
- FDA: Food and drug administration
- HN: Half-normal distribution
- IPD: Individualised patient data
- MAR: Missing at random
- MCMC: Markov Chain Monte Carlo
- μ_{jk}: Log odds of an event for treatment k in study j
- $\mathcal{N}$: Normal distribution
- NMA: Network Meta-Analysis
- NSCLC: Non-small-cell lung cancer
- n: Number of studies
- LOR: Log Odds Ratio
- ORR: Overall Response Rate
- p_{jk}: Event probabilities in study j for treatment k
- r_{jk}: Number of observed events in study j for treatment k
- RMSE: Root Mean Squared Error
- S: Standard treatment

- τ: Between-study heterogeneity
- TE: Treatment Effect
- TEE: Treatment Effect Estimate
- TKI: Tyrosine kinase inhibitors
- $\mathcal{U}$: Uniform distribution
- w_j: Study-specific weight
- x_j: Proportion of biomarker-positive patients in study j

List of Tables

List of Figures

Chapter 1

Introduction

Comment: Parts of the following Chapter have already been published in the article "Integrated evaluation of targeted and non-targeted therapies in a network meta-analysis" (Proctor et al., 2020) and in the article "A comparison of methods for enriching network meta-analyses in the absence of individual patient data" (Proctor et al., 2022).

1.1 Background

Recent development in clinical research is showing a shift from a generalised medical approach to more personalized medicine, where genetic or other biomarker information is used to make treatment decisions. The FDA (Food and Drug Administration) defines a biomarker as a "characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions" (FDA-NIH Biomarker Working Group, 2016). During drug development clinical research accordingly focuses on the question if a drug seems to be more promising in a specific subgroup of patients (i.e., in patients with a particular biomarker) compared to an originally defined population (Renfro et al., 2016). Sometimes this specific subgroup of patients is only identified retrospectively in subgroup analyses because the molecular characteristics that drives a patient's response to a specific treatment are unknown during drug development and might only be discovered later. The advent of personalized medicine also has implications for clinical trial designs. These designs need to be adapted in order to optimize the usage of new therapies in biomarker-defined subgroup of patients (Buyse et al., 2011). To adapt clinical studies to this situation of personalized medicine in clinical development several clinical trial designs, such as the (partial) enrichment design, the biomarker-stratified design, the biomarker strategy-design, or more recently, specific basket designs have been developed (Freidlin et al., 2010; Yin et al., 2021). These new designs are readily applied in today's research.

In this work the focus is on synthesizing evidence from non-targeted studies together with evidence from targeted or enriched designs. In the latter designs patients are tested for a specific biomarker before they enter the trial. Only patients who are tested positive for this respective biomarker are randomized to an experimental or control treatment (Maitournam & Simon, 2005; Simon & Maitournam, 2004).

One very prominent field of individualized medicine is cancer research. The application example, which is introduced in detail in Chapter 2 in this work also stems from this area of medical research. In this example the focus is on non-targeted treatments and targeted treatments with tyrosine kinase inhibitors (TKI) for tumors harboring the epidermal growth factor receptor (EGFR) mutation in non-small-cell lung cancer (NSCLC). When there are multiple treatment options available for a specific disease, the difficulty for clinicians is to make a considered decision which therapy option is the best with regards to specific endpoints (e.g.: survival, response). In this case a systematic review,

with at best a (network) meta-analysis is needed in order to combine all evidence available and calculate a pooled treatment estimate of all studies. Network meta-analyses, which combine direct and indirect evidence, provide useful evidence for judiciously selecting the best treatment if more than one therapy option is available but each included trial only compares a subset of these treatment options (Dias et al., 2018; Hoaglin et al., 2011). This pooling of studies with different study populations where targeted and non-targeted therapies are investigated however is not trivial. For the situation described above, specific network meta-analysis methods are needed to synthesize evidence for targeted therapies on patient subgroups (patients with a specific biomarker) and non-targeted therapies on a mixed patient population with regard to a special biomarker status. With other words network meta-analysis (NMA) methods need to be investigated as to whether the integrated evaluation of new targeted therapies is feasible. In broader context, these NMA methods could then also be applied for synthesizing studies with an overall population and studies with a subpopulation regarding a certain covariate. The main aim in this work is to evaluate different approaches on how to best conduct these network meta-analyses

The classical approach for conducting a NMA in the situation of different study population only includes studies with this specific subpopulation. This can be studies on this subpopulation or targeted studies only (e.g. the EGFR positive patients in Liang et al. (2014) and Lin et al. (2018)) or studies where retrospective subgroup results on this subpopulation are available (e.g. the EGFR example by Greenhalgh et al. (2021)). However, this classical approach results in a huge loss of information (and therefore statistical power) as the information from the subpopulation that is included in studies with the overall population will be ignored when no subgroup results are available and the studies are therefore excluded from the analysis. Lumping the studies together irrespective of the distribution of the effect modifier (e.g. the biomarker status) though, results in a systematically increased between-trial heterogeneity and, regarding the population, a non-conclusive treatment effect.

There is therefore the need for a network meta-analysis model which includes all evidence available but also accounts for the fact that the studies consist of heterogeneous patient populations. Recent publications, (Ishak & Benedict, 2015; Petto et al., 2019; Phillippo et al., 2020; Remiro-Azócar et al., 2021; Weber et al., 2020) evaluated the performance of methods to combine evidence of studies with different population groups in one network meta-analysis model and adjusting for population heterogeneity. These methods such as the matching-adjusted comparison approach (Signorovitch et al., 2012) or the simulated treatment comparison approach (Caro & Ishak, 2010) however rely on the availability of individual patient data (IPD) when synthesizing evidence from mixed populations and targeted populations. In this present work the focus lies on the situation where only aggregated data (AD) for all studies, especially the studies with

mixed patient population is available, because IPD is usually not easy to obtain and most study data is available only as AD. Up to now a detailed evaluation of methods for network meta-analyses including studies on AD with overall and subpopulations regarding a specific attribute or biomarker covariate is missing, and this work will provide such an evaluation.

1.2 Aim and structure of the thesis

The aim of this thesis is to evaluate the performance of different methods to synthesise all evidence available in a network meta-analysis and taking at the same time into account that the patient population of the included studies differs.
In the following, the focus lies on the setting for targeted and non-targeted therapies, which is also the situation of the motivational example in this work. However, as discussed above, the methods could also be applied in other settings where the overall patient population as well as subpopulations are available. In this work, it is assumed that there are three different treatment options available for a certain medical condition and the particular interest is in comparing a newly developed targeted treatment E_+ for patients with a certain biomarker (biomarker-positive (BM_+) patients) with the standard treatment S. Treatment S is assumed to be initially developed for the whole patient population. For simplification it is further assumed that a patient population can be divided in biomarker-positive (BM_+) and biomarker-negative (BM_-) patients and that the group of BM_- patients is the complement of BM_+ patients. No further differentiation with regards to the biomarker status are made and it is further assumed that there are no patients in the study population with an unclear biomarker status.
Furthermore it is assumed that there is only little direct evidence, one study, available for comparing treatment E_+ versus treatment S (see Figure 1.1). For the indirect comparison via a control treatment C there a varying number of non-targeted studies for C versus S (CS) available from very little evidence (n = 2) to a relatively high evidence (n = 20). Additionally three studies for C versus E_+ (CE_+) are assumed to be available. The number of available studies for CE_+ and S versus E_+ (SE_+) are based on the real world example introduced in Section 2.8. Regarding the percentage of biomarker-positive patients included in the studies comparing treatment S and treatment C, two different scenarios, 'A' and 'B', displayed in Figure 1.1 will be evaluated.
In setting 'A' it is assumed that no CS studies with 100 % biomarker-positive patients exist but only studies with a mixed patient population. The percentage x of biomarker-positive patients per study is assumed to be known and for all studies smaller than 100%. Setting 'B' is a further development from setting 'A' which is based on real world example. Here it is assumed that for CS there are additionally two studies with 100% biomarker-positive patients available ($x \leq 100$).
The aim in this work is to use all evidence available (direct and indirect) to determine a

treatment effect estimate for biomarker-positive patients comparing the targeted treatment E_+ and non-targeted therapies S based on treatment comparisons in the targeted as well as in the whole patient population. To arrive at the treatment effect estimate SE_+, the evidence of non-targeted CS studies need to be integrated where a varying proportion of biomarker-positive patients is assumed to be part of the mixed patient population, with the biomarker-positive population in the CE_+ studies. The network model in this work is assumed to include only two-arm trials and to be consistent, i.e. direct evidence agrees with indirect evidence on each treatment comparison (Salanti, 2012). The network setting described above is displayed in Figure 1.1. The focus of the evaluation of network meta-analysis methods lies in the following on this specific network.

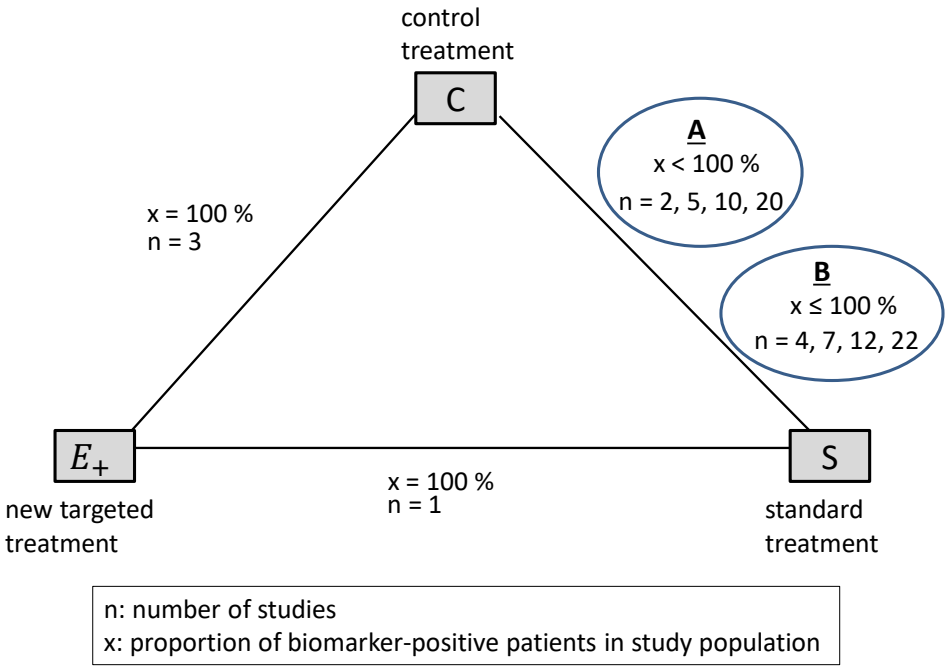


Figure 1.1: *Network model displaying the number of studies available and the proportion of biomarker patients depending on the setting.*

This work is structured as follows: In Chapter 2, different network meta-analysis models are outlined in the subsequent sections of the chapter to determine a treatment effect estimate for biomarker-positive patients. These methods range from a naive network meta-analysis model and a stand-alone network meta-analysis model (Section 2.3) over a missing data approach (Section 2.4) and a network meta-regression model (Section 2.5) to models adapted from Efthimiou et al. (2017) such as the enriching-through-weighting model (Section 2.6) and the informative prior model (Section 2.7). A motivational

example is introduced in Section 2.8, which will be used for an application in the results. To evaluate the performance of the varying approaches, a simulation study is conducted (Section 2.9) focusing on binary outcomes. Methods will be compared with respect to bias, Root-Mean-Squared Error (RMSE), coverage, precision, and power of the estimate for SE_+ and the results will be displayed in Chapter 3. The different models will be exemplified using a data set from a real clinical example (Section 3.1). A discussion is given in Chapter 4 and the thesis ends with a summary in Chapter 5.

Chapter 2

Methods

Comment: Parts of the following Chapter have already been published in the article "Integrated evaluation of targeted and non-targeted therapies in a network meta-analysis" (Proctor et al., 2020) and in the article "A comparison of methods for enriching network meta-analyses in the absence of individual patient data" Proctor et al. (2022). The original manuscripts were written by myself, but contain also comments and corrections from the co-authors.

2.1 General methodological aspects

Systematic reviews are positioned at the top of the evidence-based practice hierarchy (Tonin et al., 2017) and become increasingly more important. In the Cochrane handbook the benefit of systematic reviews are described as follows (Chandler et al., 2021): "Systematic reviews seek to collate evidence that fits pre-specified eligibility criteria in order to answer a specific research question. They aim to minimize bias by using explicit, systematic methods documented in advance with a protocol." The statistical synthesis of the data, the *meta-analysis*, is a key element in most systematic reviews. Systematic reviews and meta-analyses are of great importance for medicine when it needs to be decided which treatments are best based on the available study data (Borenstein et al., 2021). In a meta-analysis, the estimated treatment effects of two or more studies are combined with statistical methods in order to form a pooled estimate. The decision if studies can be pooled or not depends also on other criteria, e.g. the quality of studies or the reported measure of estimates, strongly on the study population of interest. The study population in the different studies needs to be considered homogeneous enough in order to be pooled. If the study population is considered to be too heterogeneous it cannot be pooled without adjusting for this heterogeneity. In this thesis methods are evaluated on how to adjust for potentially heterogeneous study populations and still use all evidence available.
The advantage of conducting a meta-analysis as part of a systematic review is that it provides a transparent, objective, and replicable framework on how to pool the single studies. Weights are assigned to each study based on mathematical criteria that are specified in advance (Borenstein et al., 2021) and then a pooled estimate is calculated. Since the publication of the first meta-analysis model by DerSimonian and Laird (1986), methods and models have been further developed. One particular advancement is the extension of the pairwise meta-analysis to an analysis of a network of treatments, a *network meta-analysis* (NMA). In a NMA, three or more treatments for a given medical condition are compared, based on combining information from multiple existing comparisons among subsets of the treatments (Bucher et al., 1997; Higgings & Whitehead, 1996; Lu & Ades, 2006, 2009; Lu et al., 2011; Lumley, 2002; Salanti, 2012). Caldwell et al. (2005) first introduced a NMA model in order to perform a simultaneous analysis of multiple treatments. The use of a NMA has since become more and more popular

because it allows to compare more than two treatments simultaneously and to use all evidence available in order to make a treatment recommendation. As discussed in the introduction, Chapter 1, the focus in this current work lies on the evaluation of NMA methods for the setting illustrated in Figure 1.1.

In general, a (network) meta-analysis can be performed as a fixed-effect or a random-effect analysis. For a fixed-effect (network) meta-analysis, it is assumed that the participants in the included studies are similar and that the same true effect size underlies the different studies. In the case of this work however, it is assumed that the participants in studies differ and that therefore different true effects exist withon the different studies, which are assumed to have been sampled from a distribution of these true effects (Borenstein et al., 2021). This so called random-effect model takes, with other words, two sources of variance into account, the within-study error in estimating the effect in each study, and the between-study error (Borenstein et al., 2021). This between-study error or heterogeneity τ, quantifies the variance of true effects across the studies (Harrer et al., 2021). If the heterogeneity is too high, the trust in the pooled estimate can be challenged.

(Network) meta-analysis models are based on either a frequentist or a Bayesian framework. One well known frequentist NMA framework is described in detail in Rücker et al. (2020). For the NMA model which is applied in this present work, a Bayesian framework is used to synthesize the data using Markov Chain Monte Carlo (MCMC) methods (Lunn et al., 2013) implemented in the software JAGS (Just Another Gibbs Sampler) (Plummer, 2003). In this work, the Bayesian approach is preferable to the frequentist approach because it offers more flexibility in the modification of the model and can therefore be used for e.g. network meta-regression or models where prior information is implemented. Another advantage is that the MCMC implementation of the Bayesian approach automatically takes into account uncertainty about the between-study heterogeneity parameter τ (Dias et al., 2018), and therefore better reflects the overall uncertainty in the treatment effect estimates compared to the frequentist approach.

In the following, the properties of this Bayesian approach, also including the so-called 'core model' of this approach (see Dias et al. (2018)) are explained in more detail. In Bayesian statistics, probability distributions are attached to the parameters of interest. To each unknown quantity, a prior probability distribution is assigned to. This prior then describes an *a priori* uncertainty about the quantity before seeing the data. The Bayesian analysis itself combines the prior distribution with the data, turning it into a posterior probability distribution for the unknown quantity. One widely discussed aspect of Bayesian (network) meta-analysis is the choice of the prior distribution for the mean treatment effect, the standard deviation and the heterogeneity. The choice needs to be made between "informative priors" who can contribute information from external

sources, e.g. historical trials, and so called "non-informative" or "vague" priors which are chosen such that they do not impact to the posterior probability distribution. In the present work, mainly non-informative or vague priors were chosen in order to keep the setting as generic as possible. These priors are specified in the following sections for the different models. According to Röver et al. (2021) even a so-called non-informative prior for heterogeneity could become informative, i.e. the prior could be influential, when the data provided is insufficient (e.g.: a small number of studies) and therefore the result of the (network) meta-analysis has to be interpreted with caution. This is why the choice of the heterogeneity prior is to be evaluated in the simulation study as well.
In contrast to frequentist approaches, a Bayesian analysis produces credible intervals instead of confidence intervals. A 95%-credible interval from a Bayesian analysis is a summary of the posterior distribution, such that the probability is equal to 95% that the true (unobserved) estimate is contained within the interval, given the observed evidence (Borenstein et al., 2021). The precision measures in this work are therefore based on credible intervals.
For different endpoints in a clinical study there are different scale levels (e.g.: nominal or interval scale). Dependent on the effect size of interest there are a number of different methods to perform a (network) meta-analysis (Borenstein et al., 2021). In this current work, only methods on binary outcome data are considered, and the odds ratio as an effect measure is accordingly used as suggested in Borenstein et al. (2021).
In the following, different approaches are introduced to synthesize the study data, always assuming that some studies are enriched in their design and exclusively include a targeted population while others are not (see study availability in Figure 1.1).
All study data is assumed to be available as binary aggregated data contributing to the network introduced in Figure 1.1 in Chapter 1 and is analysed with a Bayesian random-effect network meta-analysis model. Vague priors are used for the treatment effect estimate and the heterogeneity unless explicitly otherwise stated. The treatment effect estimate of main interest it the comparison of SE_+.

2.2 Standard pairwise meta-analysis model

The setting used in all following approaches is a network of treatments as displayed in Figure 1.1. In order to introduce the baseline model for synthesizing study data, a meta-analysis model is described, which is also relevant for the informative prior approach (see Section 2.7). This standard meta-analysis random effects model used in this work is based on the Bayesian MCMC approach for pairwise meta-analysis on binomial data as introduced first in (Smith et al., 1995). The notation used in the following is based on the notation in the Technical Support Document 2 in Dias et al. (2011). For a better illustration, the treatments in this model are explicitly stated as the comparison of C versus S in Figure 1.1. This also holds for other comparisons such as C versus

E_+ or E_+ versus S. In a random-effects meta-analysis model, it is assumed that each j^{th} study provides an estimate of the study-specific treatment effect δ_{jCS} (Dias et al., 2018). These study-specific estimates δ_{jCS} are assumed not as equal but exchangeable, therefore not dependent on the single study j. With other words, this assumption of exchangeability, also called heterogeneity, is equivalent to stating that the trial-specific treatment effects originate from a common distribution with mean d_{CS} and variance τ^2_{CS} (Dias et al., 2018). This common distribution is chosen to be a normal distribution, with $\delta_{jCS} \sim N\left(d_{CS}, \tau^2_{CS}\right)$. It is further assumed that the data generation process follows a Binomial likelihood with the logit link function that maps the probabilities into a continuous measure between plus and minus infinity:

$$\begin{gathered} r_{jC} \sim \text{Bin}\left(p_{jC}, n_{jC}\right), \quad r_{jS} \sim \text{Bin}\left(p_{jS}, n_{jS}\right) \\ \text{logit}(p_{jk}) = \begin{cases} \mu_{jC} & \text{for } k = C, \\ \mu_{jC} + \delta_{jCS} & \text{for } k = S, \end{cases} \\ \delta_{jCS} \sim N\left(d_{CS}, \tau^2_{CS}\right) \end{gathered} \tag{2.1}$$

Treatment k with $k \in \{C, S\}$ and study j are denoted by their respective indices. The expression r_{jk} denotes the number of observed events and n_{jk} the total number of individuals per study-arm. The event probabilities are represented by p_{jk}. μ_{jC} is the log odds of an event for the baseline treatment C. The log odds ratio δ_{jCS} for the treatment S relative to the baseline treatment C is assumed to be normally distributed with a common mean d_{CS} and between-study heterogeneity variance τ^2_{CS}.
Prior distributions need to be specified for μ_{jC}, d_{CS} and τ^2_{CS}. As recommended in Dias et al. (2011) flat or vague priors are chosen for $\mu_{jC} \sim N(0, 10^6)$, $d_{CS} \sim N(0, 10^6)$ and $\tau_{CS} \sim \mathcal{U}\left(0, 2\right)$. These normal density priors have an extremely large variance and give similar prior values over a large range of parameter values. Therefore they are considered as flat. This prior distribution for τ assigns equal likelihood on all possible values of the parameter and is therefore regarded as non-informative. For τ, the upper limit of 2 represents a large range of trial-specific treatment effects. In the simulation study, the use of a half-normal prior ($\tau \sim \mathcal{HN}\left(0.5\right)$) was therefore investigated as well as suggested in Röver et al. (2021) (see Section 2.9). The half-normal distribution is a zero-mean normal distribution that is restricted to take only positive values: $\mathcal{HN}(\sigma) = \sqrt{\frac{2}{\pi\sigma^2}}\exp(-\frac{x^2}{2\sigma^2})$ for $x \in \mathbb{R}_{\geq 0}$ Using the half-normal distribution as heterogeneity prior leads to a monotonically decaying tail when the heterogeneity increases and therefore implies decreasing probability for increasing τ values, whereas a uniform distribution implies the same probability for increasing τ values.

2.3 Network meta-analysis

The meta-analysis model as introduced above can be regarded as a special case of a NMA where there are only studies available for one arm of the network. Regarding the situation displayed in the network in Figure 1.1, three pairwise comparisons could be possible, C versus S, C versus E_+, and E_+ versus S. The situation of interest in this present work lies in the comparison of treatment E_+ compared to treatment S for a subgroup of patients. When there are studies available comparing C with S and C with E_+, but no study is available comparing directly treatment S with E_+ then an indirect comparison could be performed as illustrated in more detail in Bucher et al. (1997). However when the availability of studies form a closed loop as in Figure 1.1, i.e. study data for S versus E_+ are available as well, then two sources of evidence exist for the comparison of S versus E_+, namely direct evidence from the trials E_+ versus S and indirect evidence from trials C versus S and C versus E_+. As before with the random-effect meta-analysis, exchangeability of the treatment effects δ_{jCS} for the trials comparing C and S, δ_{jCE_+} for trials comparing C and E_+, and δ_{jSE_+} for trials comparing S and E_+ is assumed. Now, it is further assumed that the within-trial transitivity relation for $\delta_{jSE_+} = \delta_{jCE_+} - \delta_{jCS}$ is valid. Then the exchangeability of all treatment effects δ_{jbk} is implied (see Dias et al. (2018)) and according to Lu and Ades (2006, 2009), the consistency of the mean of the treatment effects is fulfilled as well with $d_{SE_+} = d_{CE_+} - d_{CS}$. According to Dias et al. (2018), exchangeability is the only assumption for NMA and exchangeability implies consistency. However, one must be aware that a higher degree of patient population heterogeneity might make it difficult to know what relevance the pooled estimate of the trial data has to a certain population, and therefore methods to adjust for the heterogeneity are investigated in this work.
The terms exchangeability and consistency are not always defined in a consistent way. For example in Salanti (2012), the term transitivity as an alternative to consistency has been introduced. But according to Salanti (2012) both terms are equivalent and here the notation as in Dias et al. (2018) is used. With this assumption of exchangeability, the NMA model generalizes the indirect comparison approach of Bucher et al. (1997). The assumption of exchangeability of all true trial-specific treatment effects needs to hold for the entire set of trials. For the setting in this work, only studies with two-arm trials are considered where three treatments are compared in total. Here it needs to be assumed that some of the treatment arms are missing at random (MAR), i.e. "that the missingness is without regard for the presence of effect modifier" (Dias et al., 2018) or the missingness needs to be unrelated to relative efficacy. In this work, it is assumed that all effect modifiers apart from the biomarker status are distributed equally across the trials and that the MAR assumption does not depend on other effect modifiers. This MAR assumption concerning the postulation that the biomarker status is an effect modifier however depends on the specific treatments which are compared in a study (targeted

or non-targeted studies for biomarker) as well as on the type of the included studies (mixed studies or only targeted studies) dependent on the methods discussed below. This is why, in the following sections, the validity of the exchangeability assumption is discussed separately for each method.

The consistency implied through exchangeability can be formally noted as: $\delta_{jbk} \sim N\left(d_{bk}, \tau^2\right) \sim \mathcal{N}\left(d_{Ck} - d_{Cb}, \tau^2\right)$. As in Dias et al. (2018), equal between-study variances in all models are assumed with $\tau^2_{CK} = \tau^2_{Cb} \cdots = \tau^2$. This leads to the following base network meta-analysis model which is based on the model in Equation 2.1 and following the notation of Dias et al. (2018). The difference to the meta-analysis model above is that the baseline treatment is now denoted with $b \in \{C, S\}$ dependent on the single study. $k \in \{C, S, E_+\}$ denotes the comparator treatment.

$$\begin{aligned} r_{jk} &\sim \text{Bin}\left(p_{jk}, n_{jk}\right), \\ \text{logit}(p_{jk}) &= \begin{cases} \mu_{jb} & \text{for } k = b, \\ \mu_{jb} + \delta_{jbk} & \text{for } k \neq b, \end{cases} \\ \delta_{jbk} &\sim N\left(d_{bk}, \tau^2\right) \end{aligned} \tag{2.2}$$

All parameters and their priors are specified as above in Equation 2.1.
As discussed before, the aim of this work is to evaluate models for using all evidence available and to conduct a network meta-analysis even when the study populations of the single studies to be included are heterogeneous. A first option for pooling such data as introduced in Figure 1.1 would be using this classical random-effects network meta-analysis (Equation 2.2). This model includes between-trial heterogeneity by employing a random-effect but does not further adjust for the presence of possible heterogeneous study populations (see Figure 1.1). Two different specifications of the patient population could now be used for estimation: all studies (irrespective of the biomarker status of the patient population of the study) or the studies including solely biomarker-positive patients. The second specification only works for setting "B" (see Chapter 1), when for all treatment comparisons there are study arms available with only biomarker-positive patients or at least subgroup results of biomarker-positive patients. Both approaches are evaluated and refer to the estimation ignoring biomarker status as '**naive analysis**' and to the estimation concentrating on the biomarker status as '**stand-alone analysis**'. A naive analysis can be justified if it is assumed that biomarker-positive patients perform similarly to biomarker-negative patients when treated with treatment S compared to treatment C, i.e. the population is not truly separated by the suggested biomarker or the biomarker status is no effect modifier. However, if this assumption is wrong and the patient population was truly separated by the biomarker, this approach potentially increases the between-trial heterogeneity through the heterogeneous study population and results in a treatment effect estimate for the mixed population (mix of biomarker-positive

and -negative patients) instead of the treatment effect estimate for a biomarker-positive population. Another limitation of the naive analysis is that it might also violate the exchangeability assumption, e.g. in the situation where the biomarker status is an effect modifier and the MAR assumption is not independent of the effect modifier anymore. Assuming the biomarker status is an effect modifier for BM_+ but not BM_- patients then the treatment E_+ would not be included in a trial with BM_- patients because of lower relative efficacy in this population. E_+ can then not be considered to be missing at random in the CS trials with a mixed patient population because it was excluded due to lower relative efficacy. Additionally, if the effect modifier is not distributed equally in both arms because of inclusion criteria, it is unclear how to interprete the pooled effect estimate Dias et al. (2018), as its target population is unclear. With the stand-alone approach it is assumed that biomarker-positive patients have outcomes different to biomarker-negative patients, implying that the biomarker truly separates the population. Using only studies with 100%-biomarker-positive patients avoids to pool a too heterogeneous population and delivers an estimate for the targeted patient population. But this approach can only be applied if there are studies with 100% biomarker-positive patients comparing the treatments C versus S available (setting B). The advantage of the stand-alone approach is that the MAR assumption holds, since it can be assumed that the missingness is unrelated to the relative effectiveness and that the effect modifier biomarker status across the different treatments does not differ because all studies include 100% biomarker-positive patients. A disadvantage of the stand-alone analysis is that the information from studies with mixed population cannot be used for treatment S and therefore it might result in a loss in precision if only few or in an extreme case no biomarker-positive studies are available for comparing C versus S (setting A).
Conducting both, the naive analysis and the stand-alone analysis, can help to evaluate the agreement of treatment effects for biomarker-positive and -negative patients, as the analyses are both extreme in their assumptions.
All estimation approaches that are included in the following are based upon the model introduced in Eq. 2.2 and will modify different aspects and assumptions to take the heterogeneity of patient populations into account more explicitly.

2.4 Missing data approach

In the presented network in Figure 1.1, it is assumed that the patient population can differ between the comparisons C versus S and C versus E_+ when the biomarker status acts as an effect modifier, because studies comparing the treatment C versus S included different proportions of biomarker-positive patients. If all data available should be used and the stand-alone analysis is not feasible because there a no studies with $x = 100\%$ (e.g. setting "A"), one first "crude" and alternative analysis approach to the naive approach could be a kind of a "**missing data approach**". Here the missing information on

the biomarker status for 'mixed' studies comparing C and S is treated as a missing data problem and imputes the biomarker status by allocating the response rates in different ways to the biomarker-positive patients. It is assumed that only the treatment effect for the entire study population as well as the percentage of biomarker-positive patients is known but not the treatment effect for the subgroup of biomarker-positive patients. With the following different specifications of missing data approaches, the response rates of the subgroup of biomarker-positive patients in both treatment groups are determined. Then these response rates for the treatments C and S are used to estimate treatment effects for CS, CE_+, E_+S with the network meta-analysis model introduced in Section 2.3, Eq.2.2. For studies including 100 % biomarker-positive patients, i.e. the studies comparing C versus E_+ and E_+ versus S, the response rates of the biomarker-positive patients are not missing and can be directly implemented in the network meta-analysis model.
Five different scenarios for a synthesis model based on the missing data approach are described in the following:

1. Best-case scenario: In this scenario, the maximum possible number of biomarker-positive patients are assigned as having a response in the control treatment group C as well as in the standard group S given the overall success rate and the proportion of the biomarker-positive patients.

2. Best-case scenario only for biomarker-positive patients in the standard treatment group: The best-case scenario is only applied to the biomarker-positive patients in the treatment group S. For the biomarker-positive patients in the control group C, it is assumed that the response rate is equal to the response rate of the overall population in the control group, with other words, that there is no difference in the performance between the biomarker-positive subgroup and the biomarker-negative subgroup in the control group.

3. Worst-case scenario: In contrast to the best-case scenario, the worst-case scenario assumes that the minimum possible number of biomarker-positive patients show a response in the both treatment groups C and S.

4. Worst-case scenario only for biomarker-positive patients in the standard treatment group: In this scenario, only the biomarker-positive patients in the treatment group S perform in the worst possible way, whereas biomarker-positive patients in the control group C perform equivalently to the overall population. With other words, biomarker-positive patients in the control group have the same response rate as the overall population in the control group.

5. Biomarker status not predictive for response rate: This scenario assumes that the biomarker status of patients is not predictive for the treatment effect, i.e. the overall response rate in the subgroup of the biomarker-positive patients is the same as the overall response rate of the overall population for both treatment groups.
In contrast to the naive approach only the subgroup of biomarker-positive patients is used in the analysis.
The aim of these different scenarios introduced above is to define the limits of the best and worst possible treatment effect of the biomarker-positive subgroup, to use all evidence available and to compare the performance of these simple approaches with the other approaches, especially the naive approach. One limitation here is that exchangeability assumption needed to conduct a NMA could be violated in this approach dependent on the assumption of the biomarker-status as effect modifier. This needs to be discussed when evaluating the results using the missing data approach.

2.5 Network meta-regression

The aforementioned missing data approach is a first crude approach in order to combine evidence of possible heterogeneous population in a network meta-analysis and to define the crude limits of the treatment effect estimate. As a second approach a **network meta-regression model** is introduced at this point as an attempt to take the heterogeneity of the population into account by explicitly modeling it. The network meta-regression allows to investigate the effect of continuous as well as categorical characteristics, and in principle allows the effects of multiple factors to be investigated simultaneously (although this is rarely possible due to an inadequate numbers of studies) (Thompson & Higgins, 2002). The network meta-analysis model of Eq. 2.2 is extended to a network meta-regression model by including a covariate that codes the proportion of biomarker-positive patients per study, x_j, and is denoted following the notation of Saramago et al. (2012). The motivation for this network meta-regression model can be derived from the assumption that the treatment effect estimate for a mixed patient population in study j comparing k with b (θ_{jbk}) is composed of the sum of the treatment effect estimate for BM_+ patients (θ_j^+) times the proportion of BM_+ patients (x_j) and the treatment effect estimate of BM_- patients (θ_j^-) times the proportion of BM_- patients ($1 - x_j$). It is further assumed that: $\theta_{jbk}^- = \delta_{jbk}$ and $\theta_{jbk}^+ = \delta_{jbk} + \beta$.
Then it follows:

$$\begin{aligned}\theta_{jbk} &= (1 - x_j) \cdot \theta_{jbk}^- + x_j \cdot \theta_{jbk}^+ \\ &= \theta_{jbk}^- + x_j \cdot (\theta_{jbk}^+ - \theta_{jbk}^-) \\ &= \delta_{jbk} + x_j \cdot \beta_{bk}.\end{aligned}$$

So β_{bk} is equivalent to the difference in the treatment effect for the biomarker-positive and negative population ($\beta_{bk} = \theta^{+}_{jbk} - \theta^{-}_{jbk}$). In the model below, binary endpoints with the logit link function are considered and therefore θ_{jbk} is given as $\text{logit}(p_{jk})$.

$$\begin{gathered} r_{jk} \sim \text{Bin}\left(p_{jk}, n_{jk}\right) \\ \text{logit}(p_{jk}) = \begin{cases} \mu_{jb} & \text{for } k = b \\ \mu_{jb} + \delta_{jbk} + \beta_{bk} \cdot x_j & \text{for } k \neq b, \end{cases} \\ \delta_{jbk} \sim N\left(d_{bk}, \tau^2\right) \sim \mathcal{N}\left(d_{Ck} - d_{Cb} + \beta_{bk} \cdot x_j, \tau^2\right), \\ \beta_{bk} \text{ regression coefficient,} \\ x_j \text{ study-level specific covariate.} \end{gathered} \tag{2.3}$$

Here the term $\beta_{bk} \cdot x_j$ represents a study-level specific covariate regression term for k relative to b for each study j. The covariate x_j represents the mean proportion of biomarker-positive patients of study j, while all other parameters are specified as in Equation 2.2. Consistency now requires the additional assumption that β is expressed as $\beta_{bk} \sim \mathcal{N}\left(\beta_{Ck} - \beta_{Cb}, \tau^2_B\right)$ where β_{bk} is assumed to be different but exchangeable and τ^2_B denotes the variance for the regression parameter.
Prior distributions are chosen for $\beta \sim \mathcal{N}\left(0, 10^6\right)$ and $\tau_B \sim \mathcal{U}\left(0, 2\right)$ following Saramago et al. (2012). The prior for β is deliberately specified with a wide variance in order to arrive at a vague prior. The heterogeneity prior is chosen as a vague prior with a relatively high upper limit of 2.

As with the naive and missing data approach, the exchangeability assumption within the network meta-regression approach is not easy to argue for. In this specific setting it could be assumed that treatment C and S are available to biomarker-positive as well as biomarker-negative patients because these treatments are not targeted, but E_+ might only be available for the targeted group (the biomarker-positive patients). Therefore, the treatment arm E_+ might not be missing at random and, most importantly, the biomarker status is taken into consideration as effect modifier in the network meta-regression. On the other hand however network meta-regression has been discussed as a way to adjust for differences in the patient population and therefore to take into account the underlying heterogeneity (Salanti, 2012). According to Dias et al. (2018), limiting the degree of heterogeneity is one approach to make the assumption of exchangeability more valid. This limitation will be discussed later in the results section.

2.6 Enriching-through-weighting model

Another limitation of the network meta-regression model is that ecological bias could be introduced due to regression based on aggregated patient-level data. With other words

results from making a causal inference about a covariate on patient-level data based on aggregated data could lead to bias in the effect estimation (Morgenstern, 1982). Additionally, according to Higgins and Green (2011), a regression approach is in general not recommended when the number of data points, studies in the present case, is lower than 10. In this current section, an alternative approach to the network meta-regression is therefore introduced.

This model is a modification of the 'design-adjusted' approach presented in Efthimiou et al. (2017) for the integration of randomized and non-randomized evidence in network meta-analysis. This approach translates to the presented setting as a random-effects network meta-analysis model where studies with mixed patient populations are given less weight compared to studies which only included the targeted population. The aim is to estimate a treatment effect for biomarker-positive patients and therefore studies with mixed patient-population might be biased in giving a treatment effect estimate for only biomarker-positive patients. However, since they contain information about biomarker-positive patients they should not be completely ignored. This down-weighting of the mixed population studies is achieved in inflating the variance of the mean effect of these studies. The degree of the down-weighting can be chosen according to the confidence in the mixed population studies and is given by the down-weighting factor w_j.
The general network meta-analysis model (Eq. 2.2) is now extended in the following by including w_j in the variance of the estimated treatment effect of study j.

$$\begin{aligned} r_{jk} &\sim \text{Bin}\left(p_{jk}, n_{jk}\right), \\ \text{logit}(p_{jk}) &= \begin{cases} \mu_{jb} & \text{for } k = b, \\ \mu_{jb} + \delta_{jbk} & \text{for } k \neq b, \end{cases} \\ \delta_{jbk} &\sim \mathcal{N}\left(d_{bk}, \frac{\tau^2}{w_j}\right) \sim N\left(d_{Ck} - d_{Cb}, \frac{\tau^2}{w_j}\right) \end{aligned} \tag{2.4}$$

For $0 < w_j < 1$, the variance of the treatment effect is inflated so that the weight of the j^{th} study in the network meta-analysis is decreased.
Two different choices for w_j are investigated:

1. The proportion of biomarker-positive patients per study, x_j, is assigned to w_j. By doing so, studies with a higher proportion of biomarker-positive patients are contributing more evidence than studies with only a few biomarker-positive patients and targeted studies (x_j =1) are included without down-weighting ($w_j = 1$).

2. The weight of studies with $x_j < 1$ is decreased in assigning w_j a hyper-prior in form of a prior uniform distribution which depends on the assumed "general trust" in the evidence of the mixed population studies. In this current work, three different distributions were investigated to evaluate the different levels of trust in the mixed

population data:
1. $w_j \sim \mathcal{U}(0, 0.3)$; 2. $w_j \sim \mathcal{U}(0.3, 0.7)$; 3. $w_j \sim \mathcal{U}(0.7, 1)$.

One limitation of the enriching-through-weighting approach is the possible violation of the missing at random assumption for the exchangeability assumption because all studies are included with an assumed different distribution of effect modifiers. As discussed before, a network meta-analysis could still be legitimate because the down-weighting of mixed-population studies can be seen as an adjustment for the heterogeneous study population.

2.7 Informative prior model

So far the naive approach, the missing data approach, the network meta-regression approach, and the enriching through-weighting approach included all studies available comparing C versus S in the network meta-analysis model independent of the biomarker-status of the population, i.e. all available evidence was used. The disadvantage here is that the possibly heterogeneous study populations included in network meta-analysis might lead to a biased estimate, despite the efforts to adjust for it. Furthermore, the exchangeability assumption might be violated when conducting a network meta-analysis with mixed population where the biomarker status is seen as an effect modifier. With the stand-alone approach however, only 100% biomarker-positive studies are included and the evidence of the mixed studies is not considered at all which comes at the price of precision because the number of available studies to be included in the network meta-analysis might be lower. Therefore as a last alternative approach to conduct a network meta-analysis with the data in the introduced network (Fig. 1.1) in this current work, an adaption of the informative prior approach of Efthimiou et al. (2017) is introduced here. This model is an extension of the stand-alone network meta-analysis model and allows to include the evidence available from the studies with a mixed patient population as a prior.
The informative prior approach consists of two steps:

1. A meta-analysis of studies with a mixed patient-population ($x_j \in (0, 1)$) is conducted in order to calculate a mean relative treatment effect d_{CS}^{mix} with variance ${\tau_{Ck}^{mix}}^2$. This meta-analysis here only includes studies comparing C and S, because for C versus E_+ and E_+ versus S no mixed studies are assumed to exist. The

meta-analysis model is defined analogue to Equation 2.1:

$$\begin{aligned} r_{jk} &\sim \text{Bin}\left(p_{jk}, n_{jk}\right) \\ \text{logit}(p_{jk}) &= \begin{cases} \mu_{jC} & \text{for } k = C \\ \mu_{jC} + \delta_{jCk}^{mix} & \text{for } k = S \end{cases} \\ \delta_{jCk}^{mix} &\sim \mathcal{N}\left(d_{Ck}^{mix}, {\tau_{Ck}^{mix}}^2\right) \end{aligned} \tag{2.5}$$

The distribution of δ_{jCk}^{mix} is defined analogously to δ_{jCk} before. Vague prior distributions are again used for μ_{jC}, d, τ.

2. The mean treatment effect and the related variance from the first step is used to define a prior distribution for the treatment effect d_{CS} in a network meta-analysis with only biomarker-positive patients. Here a network meta-analysis of studies with 100% biomarker-positive patients is conducted. The resulting posterior from step 1, $\mathcal{N}\left(d_{CS}^{mix}, {\sigma_{CS}^{mix}}^2\right)$, is used as informative prior for the comparison S vs. C in this model. To account for the fact that the populations from step 1 and step 2 can differ systematically, the variance of this prior distribution can be additionally inflated by a pre-defined factor w, such that the variance of the prior is wider than the posterior from step 1 if systematically differing populations are expected.
The network meta-analysis model is then defined as:

$$\begin{aligned} r_{jk} &\sim \text{Bin}\left(p_{jk}, n_{jk}\right) \\ \text{logit}(p_{jk}) &= \begin{cases} \mu_{jb} & \text{for } k = b \\ \mu_{jb} + \delta_{jbk} & \text{for } k \neq b \end{cases} \\ \delta_{jbk} &\sim \mathcal{N}\left(d_{bk}, \sigma^2\right) \sim N\left(d_{Ck} - d_{Cb}, \sigma^2\right) \\ d_{CS} &\sim \mathcal{N}\left(d_{CS}^{mix}, \frac{{\sigma^{mix}}^2}{w_{CS}}\right) \end{aligned} \tag{2.6}$$

The important difference in this network meta-analysis model to the models before is that the prior distribution for d_{CS} is now informative, i.e. defined by the prior distribution calculated before and down-weighted by an inflation factor $0 < w_{CS} < 1$. This down-weighting is conducted to incorporate the confidence placed in the estimated treatment effect d_{CS}^{mix}. With other words the choice of w_{CS} depends on the assumed agreement between evidence from only biomarker-positive studies and evidence from studies with mixed population comparing the treatments C and S. In this current work it was assumed that there was moderate agreement between the available evidence, with $w_{CS} \sim \mathcal{U}(0.3, 0.7)$. All other prior distributions are chosen as before.

As with the stand-alone analysis, the exchangeability assumption is assumed to be valid here with regards to the effect modifier biomarker status, because only studies with 100% biomarker-positive patients are included in the NMA and so the missing at random assumption can be considered as valid.

2.8 Application

The application of the methods introduced above will be illustrated by a clinical example for targeted therapies in non-small cell lung cancer, which has been applied as well in Proctor et al., 2020 and Proctor et al., 2022.
The category non-small-cell lung carcinoma (NSCLC), which includes adenocarcinoma and squamous cell carcinoma according to classification of tumors in the lung of the World Health Organization (WHO) account for 85% of all lung cancer cases (Didkowska et al., 2016; Inamura, 2017). EGFR (Epidermal Growth Factor Receptor) mutations have a 48% incidence in Asian patients with lung adenocarcinoma (compared to 19% in Western patients) and the mutations appear predominantly in nonsmokers, women and patients of a younger age (Chapman et al., 2016; Dearden et al., 2013). The treatment of EGFR-mutated lung cancers with TKIs (tyrosine kinase inhibitors) achieves an increased tumor response rate, longer progression-free survival and less toxicity compared to chemotherapy (Greenhalgh et al., 2021).
Two examples for these TKIs are the treatments Gefitinib and Erlotinib (Cheng et al., 2012). The efficacy of Gefitinib compared to a chemotherapeutical treatment was first tested in a mixed patient population where patients where included in the study independently of their biomarker status (see e.g. Fukuoka et al., 2003; Han et al., 2012; Kim et al., 2008; Mok et al., 2009; Vansteenkiste, 2004; J.-H. Yang et al., 2014; Q. Zhou et al., 2014). With other words, non-targeted trial designs were used. One well-known study of the efficacy of the treatment Gefitinib is the non-targeted study IPASS (Mok et al., 2009) which also included patients independently of their respective biomarker status at first. In this study however, planned subgroup analyses showed that the presence of an activating EGFR mutation provides the best response for the EGFR TKI therapy, Gefinitib in this case (Mok et al., 2009; Vansteenkiste, 2018). Based on the results of this trial among others, targeted or enriched studies such as Maemondo et al., 2010; Mitsudomi et al., 2010 were conducted which only included EGFR-positive patients. For the drug Gefitinib as a therapy for NSCLC patients, studies including a mixed patient population exist alongside studies with a solely biomarker-positive patient population. Furthermore another EGFR TKI, namely Erlotinib, is used for the treatment of NSCLC. The efficacy of Erlotinib compared to a chemotherapeutical agent was only investigated in targeted studies for the biomarker EGFR (e.g. Rosell et al., 2012; Wu et al., 2015; C. Zhou et al., 2011). In further trials, targeted and non-targeted studies approaches were pursued such as in Urata et al., 2016; Xie et al., 2015; J.-J. Yang et al., 2017, where the treatments Erlotinib and Gefitinib were compared directly with each other.
The main interest in this example lies in estimating the contrast between the targeted-therapy Erlotinib to the standard therapy Gefitinib. Therefore direct and indirect evidence via the chemotherapeutical control treatments is used analogously to the scenario displayed in Fig.1.1.

In order to be able to synthesize direct and indirect evidence, the exchangeability assumption needs to hold as discussed above in Section 2.3. The validity of the exchangeability assumption for the different methods is one aspect explained above. For methods only including studies with 100% biomarker-positive patients, the exchangeability assumption is assumed to be fulfilled. For methods including mixed patient population as well as only biomarker-positive patients, the exchangeability assumption might be challenged. The chemotherapeutical treatments C are not assumed to differ systematically between the trials of targeted E_+ and non-targeted S treatments and therefore treatment C can be represented by a single node. Thus this aspect of exchangeability is given so that within-trial transitivity can be assumed. It needs to be further assumed that trials with the treatment Gefitinib are missing at random in studies comparing Erlotinib with the chemotherapeutical agent or the other way round. With other words the assumption that Erlotinib would also be given to BM_- patients and that these studies are not missing systematically linked to the presence or non-presence of effect modifier needs to hold. One could assume that the treatment Gefitinib would have been included in studies on biomarker-positive patients comparing a chemotherapeutical treatment with Erlotinib, because biomarker-positive patients have been included in Gefitinib studies. Also earlier phase II studies, such as Giaccone et al., 2006 or Jackman et al., 2007 show that Erlotinib could also in principle be given to biomarker-negative patients. So it could be argued that the missing at random assumption is also valid for the treatment Erlotinib in studies comparing Gefitinib to a chemotherapeutical agent for a mixed patient population. However this assumption could be challenged if one considers the fact that Erlotinib has later on only been given to biomarker-positive patients due to a higher efficacy in this patient group. This aspect needs to be considered when interpreting the results later.

The binary outcome used to exemplify the evaluation is the overall response rate (ORR). Biomarker-positive and biomarker-negative patients are in the following defined according to the original publications. In general, however, several (inconsistent) definitions of the EGFR-status exist, which might result in a slightly varying assignment of patients that is not considered here. As noted above, some studies such as e.g. the IPASS study (Mok et al., 2009) and the WJTOG study (Mitsudomi et al., 2010) included the patients independently of the biomarker status and only retrospectively assessed the biomarker status for some patients. In this current work the analysis is restricted to studies which report patients with biomarker re-assessment for these trials in order to be able to evaluate the performance of the methods compared to a gold standard, 'the retrospective analysis'. For simplicity, different chemotherapeutical agents are summarized in multi-arm trials into one control arm, which increases its sample size and precision but could also lead to an increased heterogeneity in the conducted analysis.

The response rates and the proportion of biomarker-positive patients out of the patients

analyzed for biomarker status in each study used for this motivating clinical example are listed in Table 2.1 along with the citation of the original publications.

Table 2.1: *Studies used in the motivational example*

Study	**Intervention**	**ORR***	**Control**	**ORR***	**Prop. BM_+^{**}**
ENSURE (Wu et al., 2015)	Erlotinib (E_+)	62.7%	Gemcitabine + Cisplatin (C)	33.6%	100%
EURTAC (Rosell et al., 2012)	Erlotinib (E_+)	58.1%	Cisplatin + Docetaxel/Gemcitabine (C)	14.9%	100%
OPTIMAL (C. Zhou et al., 2011)	Erlotinib (E_+)	82.9%	Gemcitabine + Carboplatin (C)	36.1%	100%
IPASS (Mok et al., 2009)	Gefitinib (S)	42.6%	Carboplatin + Paclitaxel (C)	37.9%	59.7%**
NEJGSG (Maemondo et al., 2010)	Gefitinib (S)	73.7%	Carboplatin + Paclitaxel (C)	30.7%	100%
First Signal (Han et al., 2012)	Gefitinib (S)	54.7%	Gemcitabine + Cisplatin (C)	47.6%	44%**
WJTOG (Mitsudomi et al., 2010)	Gefitinib (S)	62.1%	Cisplatin + Docetaxel (C)	32.2%	100%
Yang (J.-H. Yang et al., 2014)	Gefitinib (S)	48.6%	Pemetrexed+Cisplatin (C)	51.3%	67.7 %
V-15-32† (Maruyama et al., 2008)	Gefitinib (S)	66.7%	Docetaxel (C)	45.5%	100%
INTEREST† (Kim et al., 2008)	Gefitinib (S)	12.0%	Docetaxel (C)	11.3%	14.23%
Zhou† (Q. Zhou et al., 2014)	Gefitinib (S)	11.1%	Pemetrexed (C)	11.1%	71%
Yang† (J.-J. Yang et al., 2017)	Erlotinib (E_+)	59.4%	Gefitinib (S)	52.3%	100%
Xie (Xie et al., 2015)	Erlotinib (E_+)	60.9%	Gefitinib (S)	55.6%	100%
WJOG5108L† (Urata et al., 2016)	Erlotinib (E_+)	46.3%	Gefitinib (S)	47.7%	71.7%

*ORR denotes the overall response rate of the patients analyzed for biomarker status

**Prop. BM_+ denotes the proportion of biomarker-positive patients of the patients analyzed for biomarker status.

† this studies only included previously treated patients (2^{nd} line studies)

The application of the network meta-analysis methods to these real data examples was conducted with the software `R` (Version 4.1.1) (R Core Team, 2022) and `JAGS` Plummer, 2003. The R-code for reproducing the results of the application example and the simulation study is available in the Appendix B.

2.9 Simulation study

The models outlined above were implemented in the software `R` (Version 4.1.1) (R Core Team, 2022) and the software `JAGS` in order to be evaluated in a simulation study. The associated R-code is printed in the Appendix B. The main interest of the simulation study was to determine which parameter settings in combination with which approaches delivered the best estimate of the treatment effect comparing E_+ with S for biomarker-positive patients (Figure 1.1). Throughout the whole simulation study, it was assumed that studies comparing treatment C with treatment S provided only the number of proportion of biomarker-positive patients and the overall response rate. In all simulation scenarios, the models described in the sections above (2.3, 2.4, 2.5, 2.6 and 2.7) were used in order to perform a network meta-analysis or -regression and estimate a treatment effect for E_+ versus S for biomarker-positive patients. For the simulation of the studies to be included in the NMA model, a modification of the data-generating mechanism (DGM) published in Seide et al. (2019) was used. In Seide et al. (2019) the DGM is used to simulate data in the setting of multi-arm trials in a random-effects network meta-analysis. For the simulation of the data for the treatment arm S, the arm needs to be split in biomarker-positive and biomarker-negative patients. Therefore, it is assumed here that the network analysis model consists of n three-arm studies, where one arm for treatment S only includes biomarker-positive patients, the second arm considers the same treatment S but only includes biomarker-negative patients, and the control arm C, the third arm, includes both populations. Additionally, it is assumed that there exist three two-arm studies in the network comparing the control treatment C with the treatment for the targeted therapy E_+. A third comparison is included in the network model with one study comparing E_+ with S. The network model is assumed to be consistent and the single response rates in this DGM are chosen in a way to support this assumption. For the missing data approach described in Section 2.4, the numbers of positive responses (success) for the biomarker-positive subgroup are then calculated according to the assumptions outlined in Section 2.4 using the previously generated response rate for the single studies.

The combined response rates generated by the DGM for biomarker-positive (and -negative patients) for treatments C, E_+ and S as well as the information of the proportion of BM_+ patients were used to investigate the performance of the approaches for biomarker-positive and biomarker-negative patients in the different treatment arms for different settings. These settings varied for given treatment effects, proportion of BM_+ patients in a study, number of studies, and values of between-trial heterogeneity.

2.9.1 Simulation scenarios

The investigated scenarios are displayed in Table 2.2 and are all based on binary outcome data.

Table 2.2: *Different settings used in the simulation study*

Parameter	Scenario	Arms compared			
		CS		$(CE_+)^+$	$(SE_+)^+$
J (number of trials)	A	2; 5; 10; 20		3	1
	B	4; 7; 12; 22		3	1
x_j (proportion of biomarker-positive patients in study j)	A	$x_j \in [x_1, .., x_J]$; $x_1 = 0.3$, $x_J = 0.7$		1	1
		$x_j \in [x_1, ..., x_J]$; x_1=0.4, x_J=0.5			
	B	$x_j \in [x_1, x_{J-2}, 1, 1]$; x_1=0.3, x_{J-2}=0.7		1	1
		$x_j \in [x_1, x_{J-2}, 1, 1]$; x_1=0.4, x_{J-2}=0.5			
d (treatment effect)	A & B	BM_+ :	BM_-:		
		-0.21	0.2	1.79	2
		0.2	0.2	1.79	1.59
		0.4	0.2	1.79	1.39
		1.25	0.2	1.79	0.54
n (sample size)	A & B	400			
τ (heterogeneity)	A & B	0.001; 0.5; 1; 2			
prior for τ	A	$\tau \sim \mathcal{U}(0, 2)$			
	B	$\tau \sim \mathcal{U}(0, 2)$, $\tau \sim \mathcal{HN}(0, 0.5)$			

Contrasts with $(...)^+$ only included studies with biomarker-positive patients.
Setting A has a varying number of trials and the proportion of biomarker-positive patients $x_j \in (0, 1)$. In setting B ($x_j \in (0, 1]$) there are additionally 2 trials with $x_j = 1$ included. x_j is set to J equidistant values for 2 ranges of x_j with minimum x_1 and maximum x_J. The treatment effects d are given as log odds ratios and the sample size n is equally distributed between the two treatments. For setting B two different prior distributions are chosen.

As mentioned before, two different settings were assumed for the availability of trials comparing C versus S.
In setting 'A', it is assumed that only studies with a mixed patient population $x < 100\%$ are available for C versus S. For the comparison CS in setting 'B', it is assumed, inspired by the application example, that in each scenario there are two studies available comparing C versus S with 100% biomarker-positive patients and a varying number of studies with mixed patient population. The number of trials for the comparisons CE_+ and SE_+ remained the same in both settings and were chosen according to a real world data example, which is later used for the application as well.
The range of the study specific proportion of biomarker-positive patients, noted as x_j, varies in the mixed patient population between a wide and low range and is set to J equidistant values as displayed in Table 2.2. In setting B, there are additionally two studies with 100 % biomarker-positive patients added to the number of studies in setting A.
The treatment effects are given as log odds ratios (LORs). The theoretical LORs for CE_+ are based on the estimated LORs of the applicational data. For the comparison CS, the LORs for biomarker-positive patients varies whereas the LORs for biomarker-negative patients remains fixed. The varying log odds ratios illustrate low (same or different direction of treatment effect), moderate, and high agreement with the biomarker-negative patients. The treatment effect values originate from assumed response rates for the standard treatment ${p_S}^+ = (0.2, 0.4, 0.5, 0.7)$ for biomarker-positive patients and ${p_S}^- = 0.45$ for biomarker-negative patients. The response rate in the control treatment is set to $p_C = 0.4$. In order to keep the network consistent, the treatment effect SE_+ varies accordingly. The sample size n is the same for every study and equally distributed between the two treatments. Different heterogeneity values τ were chosen corresponding to values reflecting a range of relatively low to massive between-trial heterogeneity on a log-odds-ratio scale. For programming reasons, the lowest value for τ was chosen to be 0.0001 instead of 0.

Two different between-study heterogeneity priors were considered in the simulation study. As stated in the models above, the non-informative prior for the between-study heterogeneity, $\tau \sim \mathcal{U}(0, 2)$, was used. However, if the number of studies is small, it is suggested by Friede et al. (2017) to use weakly informative priors such as e.g. half normal priors. This might help to capture heterogeneity values typically seen in meta-analyses of heterogeneous studies. The use of a second prior was only investigated in setting 'B'. Here, a vague prior $\tau \sim \mathcal{U}(0, 2)$ as well as a weakly informative half-normal prior $\tau \sim HN(0.5)$ was applied for the between-trial heterogeneity as suggested in Friede et al. (2017). This second specification still includes relatively high values of heterogeneity and includes the range of values used in the simulation study by Spiegelhalter et al. (2004).

Table 2.3: *Overview of different performance measures used in the evaluation of the simulation study*

Performance measure	Definition
Mean Bias	$\frac{1}{m}\sum_{i=1}^{m}((d_{SE_+,i})^+ - (\widehat{d_{SE_+,i}})^+)$
Root-Mean-Squared-Error	$\sqrt{\frac{1}{m}\sum_{i=1}^{m}((d_{SE_+,i})^+ - (\widehat{d_{SE_+,i}})^+)^2}$
Coverage	$\frac{1}{m}\sum_{i=1}^{m}\mathbf{I}((\widehat{d_{SE_{+,low_i}}})^+ \leq (d_{SE_+,i})^+ \leq (\widehat{d_{SE_{+,upp_i}}})^+)$
Mean Precision	$\frac{1}{m}\sum_{i=1}^{m}((\widehat{d_{SE_{+,upp_i}}})^+ - (\widehat{d_{SE_{+,low_i}}})^+)$
Power	$\frac{1}{m}\sum_{i=1}^{m}\mathbf{I}((\widehat{d_{SE_{+,upp_i}}})^+ < 0 \mid (\widehat{d_{SE_{+,low_i}}})^+ > 0)$

m: number of iterations, $d_{SE_{+,low_i}}$: lower credible interval limit,
$d_{SE_{+,upp_i}}$: upper credible interval limit.

All together, 2560 different scenarios (512 for setting A and 2048 for setting B) were evaluated. (4 different values of J (A&B), 2 different values for x_j (A&B), 4 different values for d_CS, 4 different values for τ (A&B), 2 different heterogeneity prior (B), 4 different methods for setting A and 8 for setting B (naive (A&B), network meta-regression(A&B), enriching-through-weighting with x (A&B), missing data (A), stand-alone (B), 3 times enriching-through-weighting with prior weights (B), informative prior(B)).
Each scenario was simulated 1000 times. All simulations were performed in `R` version 4.1.1 (R Core Team, 2022). The program codes are provided in the Appendix B. The `R` package `R2jags` (Su & Yajima, 2015) was used in order to conduct the Bayesian analysis in `JAGS` (Plummer, 2003). The number of burn-ins in `JAGS` was set to 20.000 with 50.000 iterations using 3 chains and a thinning interval of 2. Convergence checks were conducted using Gelman and Rubins diagnostics.

2.9.2 Evaluation of results using performance measures

The focus in this evaluation lies on the treatment effect estimate for biomarker-positive patients comparing the treatments E_+ with S, denoted by $(\widehat{d_{SE_+}})^+$, calculated by the models. The treatment effect estimate obtained by the network meta-regression approach is obtained as a sum of the estimated regression parameter $\widehat{\beta_{SE_+}}$ and the 'intercept' $(\widehat{d_{SE_+}})^-$ which is the treatment effect for biomarker-negative patients. The estimated treatment effect $(\widehat{d_{SE_+}})^+$ which has been derived from data of the mixed population as well as the data from biomarker-positive population is compared to the true treatment effect $(d_{SE_+})^+$ (see Table 2.2).
Five different performance measures, displayed in Table 2.3 were used to evaluate the different models.

The first performance measure to evaluate the different models was the mean bias, which was calculated as the mean difference between the true treatment $(d_{SE_+})^+$ and the estimated treatment effect for biomarker-positive patients $(\widehat{d_{SE_+}})^+$ over 1000 simulations per scenario. Secondly, the Root-Mean-Squared-Error (RMSE) was used as a performance measure, which represents the square root of the quadratic mean of the differences, per scenario, between $(d_{SE_+})^+$ and $(\widehat{d_{SE_+}})^+$. Thirdly, the coverage of the different models was calculated as the percentage of the simulated credible intervals which included the true treatment effect. Fourthly, the average length of the credible intervals of the simulated treatment effect estimate was applied to evaluate the precision. As the last performance measure the resulting power of the different models was evaluated by calculating the percentage of iterations, with a significant estimated treatment effect. In other words, it is the percentage of simulations where either the upper bounds of the credible intervals ($d_{SE_{+,upp_i}}$) were smaller than zero or the lower bounds of the credible intervals ($d_{SE_{+,low_i}}$) were greater than zero.

Chapter 3

Results

Comment: The application results presented below differ in some instances slightly from those already published in Proctor et al., 2022 (mainly regarding the informative prior and enriching-through-weighting approach). This is due to recently optimized calculations implemented in the programming code as well as a correction of the used down-weighting factor w. However these changes in the application do not affect the general conclusions made in this publication.

3.1 Application

The different approaches introduced in the chapter above were applied to the data example illustrated in Section 2.8 in order to conduct a network meta-analysis with these 14 studies. Eight studies were available for the comparison of the standard therapy Gefitinib S with the control therapy C, thereof two studies which included 100%-biomarker-positive patients and six studies with a mixed patient population. This application example corresponds thus to setting 'B' described in the Introduction (Figure 1.1). Three studies with 100%-biomarker-positive patients were available for the comparison of the targeted therapy Erlotinib E_+ with the control therapy C. Further three studies comparing E_+ versus S were available. Two of those studies included 100%-biomarker-positive patients, while one study investigated a mixed patient population with 72% biomarker-positive patients.
The given treatment effect estimates (TEEs) were derived from the single publications using the reported response rates for the overall population, the responses rate for the subpopulation as well as the number of patients in the overall population and the subpopulation. For some studies, information on the subpopulation were only available from retrospective analyses and therefore these numbers were used. In Table 3.1, the treatment effects for the overall population, which are denoted as d, and for the biomarker-positive patients, which are denoted as $(d)^+$ of the single studies are given as log odds ratios. Additionally the table displays the proportions of biomarker-positive patients per study x.

Table 3.1: *Listing of the treatment effect estimates of the single studies based on the reported overall response rates in the publications.*

	Study	**n**	**Treatment**	**Control**	**x**	**d**	**d$^+$**
						d_{CE_+}	$(d_{CE_+})^+$
1	ENSURE (Wu et al., 2015)	217	Erlotinib	Chemo	1.00	1.20	1.20
2	EURTAC (Rosell et al., 2012)	173	Erlotinib	Chemo	1.00	2.07	2.07
3	OPTIMAL (C. Zhou et al., 2011)	154	Erlotinib	Chemo	1.00	2.15	2.15
						d_{CS}	$(d_{CS})^+$
4	IPASS (Mok et al., 2009)	437	Gefitinib	Chemo	0.60	0.20	1.01
5	NEJGSG (Maemondo et al., 2010)	228	Gefitinib	Chemo	1.00	1.84	1.84
6	First Signal (Han et al., 2012)	95	Gefitinib	Chemo	0.44	0.28	2.22
7	WJTOG (Mitsudomi et al., 2010)	117	Gefitinib	Chemo	1.00	1.24	1.24
8	Yang14 (J.-H. Yang et al., 2014)	74	Gefitinib	Chemo	0.68	-0.11	0.25
9	V-15-32† (Maruyama et al., 2008)	20	Gefitinib	Chemo	1.00	0.88	0.88
10	INTEREST† (Kim et al., 2008)	267	Gefitinib	Chemo	0.14	0.07	1.00
11	Zhou† (Q. Zhou et al., 2014)	108	Gefitinib	Chemo	0.71	0.00	1.67
						d_{SE_+}	$(d_{SE_+})^+$
12	Yang17† (J.-J. Yang et al., 2017)	256	Erlotinib	Gefitinib	1.00	0.29	0.29
13	Xie (Xie et al., 2015)	50	Erlotinib	Gefitinib	1.00	0.22	0.22
14	WJOG5108L† (Urata et al., 2016)	419	Erlotinib	Gefitinib	0.72	-0.06	-0.16

$^\dagger 2^{nd}$ line study

The study size "n" is given, which denotes the number of patients (retrospectively) assessed for their biomarker (EGFR) status. "x" is given as proportion of biomarker-positive patients in the respective the study population. "d" is the respective TEE (LOR) per study obtained by using the reported overall response rates (ORR) for the whole population analysed for their EGFR status. "d$^+$" is the respective TEE (LOR) per study of the biomarker-positive patients, obtained by using the reported ORR of the respective study (partly from a retrospective analysis).

In the evaluation of this application, it was first assumed that the treatment effect estimates obtained with the subgroup analyses were unknown and that only the treatment effect estimates of the overall population for the respective single studies (d) in Tab. 2.1 were available. This data was used in a network meta-analysis using the different methods (naive, regression, enriching-through-weighting, informative prior, standalone) introduced above in order to calculate a pooled treatment effect estimate for the biomarker-positive patients comparing Erlotinib versus Gefitinib. Additionally, a standard network meta-analysis was performed using the single study results of the aforementioned (retrospective) conducted subgroup analyses (d^+). The results of this network meta-analysis with data from retrospective analyses was used as a 'gold - standard' to compare the TEEs resulting from the different methods with the overall population.

Figure 3.1 displays the resulting pooled treatment effect estimate for Erlotinib versus Gefitinib for different network meta-analysis methods used on the data of the application example.

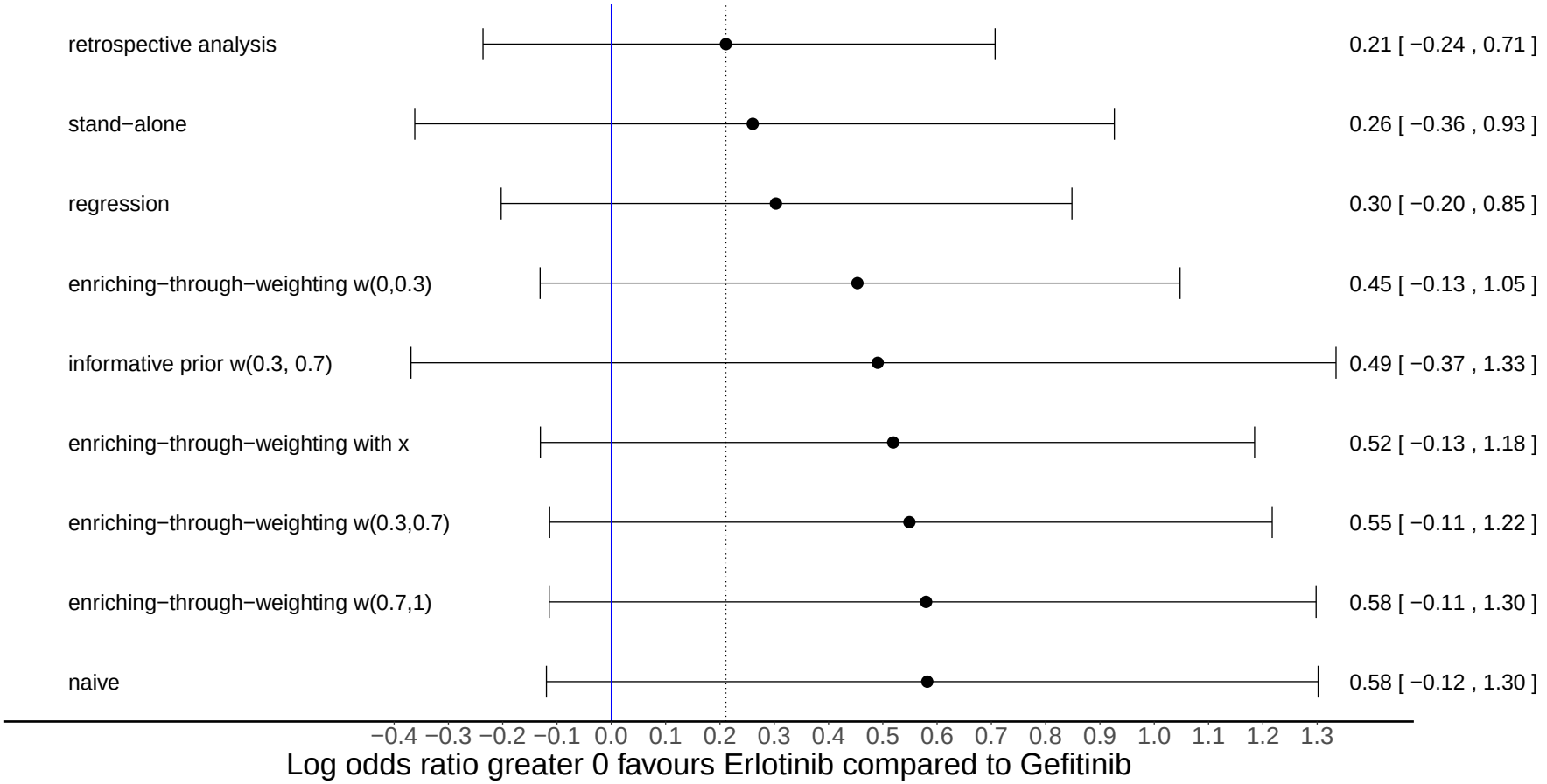


Figure 3.1: *Mean treatment effect estimates for Erlotinib vs. Gefitinib for biomarker-positive patients given as log odds ratios with their corresponding 95% - credible intervals.*

Overall it can be seen that all TEEs showed a treatment benefit for Erlotinib compared to Gefitinib. In this example, the biomarker status seemed to be an effect modifier for the treatment Gefitinib as well. This can be seen in Table 3.1 where the TEEs are higher for the BM_+ population compared to the overall population analysed for EGFR. Therefore the treatment effect estimate for Erlotinib versus Gefitinib was higher when using evidence on the whole population, compared to the retrospective analysis.
The treatment effect estimate for Erlotinib versus Gefitinib was closest to the one of the retrospective analysis (the 'gold-standard') when the stand-alone model was used. But this comparatively low deviation of the TEE from the retrospective analysis comes at the price of a higher variance. The TEE obtained using the regression model as well the enriching-through-weighting model with $w \sim \mathcal{U}(0, 0.3)$ result in a slightly higher deviation from the retrospective analysis compared to the stand-alone approach but less variance. Using the informative prior approach, the enriching-through-weighting model with $w \sim \mathcal{U}(0.3, 0.7)$ or with the respective study covariate x_j as down-weighting factor results in an even higher deviation from the retrospective result. Especially the informative prior approach has also a high variance. Nevertheless, the aforementioned models are still preferable to the models including all evidence with little (enriching-through-weighting with $w \sim \mathcal{U}(0.7, 1)$) or no adjusting (naive analysis).
For all approaches it can be seen that the treatment effect estimate of the retrospective analysis was included in the respective 95% - credible intervals. In general it can be noticed that the stand-alone approach in the network meta-analysis seems to result in the least deviation from the retrospective approach but comes at the price of low precision. The informative prior approach, which does not pool the mixed studies directly, seems also to result in a huge variance for the TEE when there is not much evidence available. The approaches which performs best considering deviation from assumed gold-standard and variance in this example seems to be the regression approach and the enriching-through-weighting model with $w \sim \mathcal{U}(0, 0.3)$. Both have a relatively low deviation but also an acceptable precision. When using the regression approach, caution has to be taken, because regressing on a patient-level covariate on a study-level can introduce the risk of ecological bias. Therefore this approach is only recommended as a sensitivity analysis to the enriching-through-weighting approach with $w \sim \mathcal{U}(0, 0.3)$.

Comment: Part of the results below have been already published in Proctor et al. (2020) (setting 'A') and in Proctor et al. (2022) (setting 'B'). For setting 'A', the results presented below differ slightly from those already published. This is due to an improved simulation algorithm and the calculation of the bias. For the bias, the results of the estimated treatment effect are in this work compared to the theoretical values and not to the retrospective subgroup values as in the publication. However, this does not affect the conclusions made in this respective publication.

3.2 Simulation study: general remarks

In the following, the results of the simulation study are presented in detail. In the first part, results from the simulation study with setting 'A' are illustrated. In this setting, it is assumed that only studies with a mixed patient population (biomarker proportion $x \in (0, 1)$) comparing treatment C with treatment S were available for the network meta-analysis (see Figure 1.1). For this setting, the *naive analysis*, (Section 2.3) *network meta-regression* (Section 2.5), *missing data* (Section 2.4) and *enriching-through-weighting* (Section 2.6) approaches were applied to perform a network meta-analysis for the simulated studies as explained in Section 2.9 and the results are evaluated.
In the second part of this section the results from the simulation study with setting 'B' are reported. Setting 'B' is an extension of setting 'A' in assuming that for the comparison of the treatments S vs. C, two studies additional are existing which included only biomarker-positive patients ($x \in (0, 1]$). In this setting, the simulated studies were analysed in a network meta-analysis with the *naive analysis*, the *stand-alone analysis* (Section 2.3), the *network meta-regression* (Section 2.5), the *enriching-through-weighting model* (Section 2.6) and the *informative prior* model (Section 2.7) and the results are evaluated. For this latter setting, the missing data approach was not included in the simulation study. When there are studies available with 100%-biomarker-positive patients, more methods to compare the network meta-regression approach against exist, and the missing data approach, a relatively crude approach, was not considered in the comparison as an alternative to the naive approach.

3.3 Simulation study: proportion of biomarker-positive patients $x \in (0, 1)$ (Setting A)

3.3.1 Bias

The mean bias of the treatment effect estimates given different scenarios and models was calculated as illustrated in Table 2.3 in Section 2.9. In the simulation study the use of different ranges of BM_+-patient proportions, [0.3,0.7] and [0.4,0.5] was evaluated. For the missing data approach, it can be noticed that, in general, a smaller range [0.4,

0.5] leads to an increased bias when the 'worst-case' and 'worst-case only for S_+' (worst-case only for biomarker-positive patients in the standard treatment group) approach are applied. For the other scenarios and approaches the range of BM_+ proportion does not seem to have an effect on the mean bias of the estimated treatment effect. This is why, in the following, only the results of the wider range ($x_j \in [0.3, 0.7]$) displayed in Figure 3.2 are evaluated in more detail. The simulation study results using the range $x_j \in [0.4, 0.5]$ are displayed in Figure A.1 in the Appendix. The different methods are illustrated in the following with different shapes. The columns of the trellis plot show the variation of the theoretical treatment effect estimates $(d_{CS})^+$, while and the rows illustrate the variation in the theoretical heterogeneity values τ (see Table 2.2). Underestimating the treatment effect for E_+ versus S for BM_+ patients ($(d_{SE_+})^+$) results in mean bias > 0, whereas overestimating the treatment effect results in a mean bias < 0.

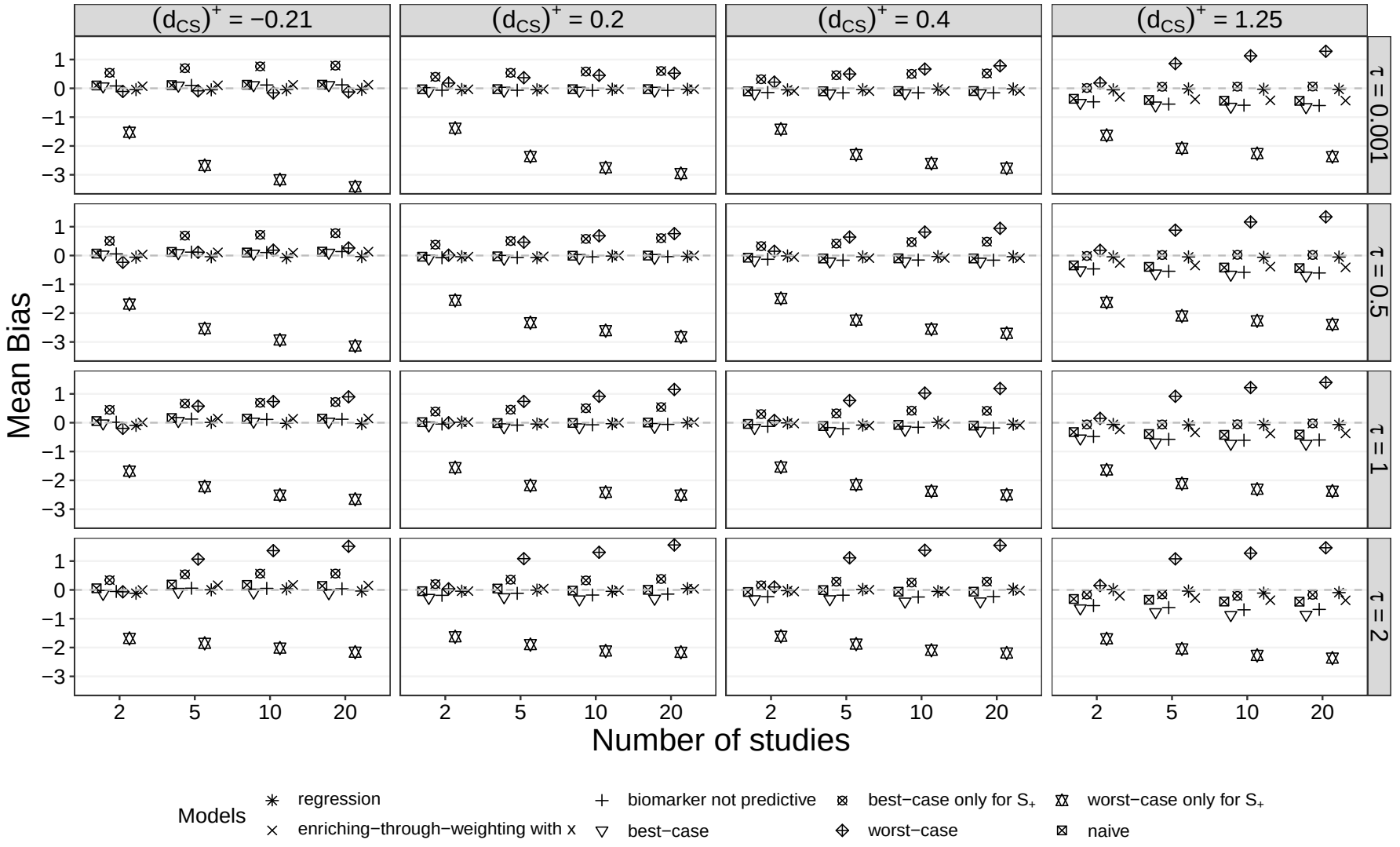


Figure 3.2: *Mean bias of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A)*

Overall it can be noticed that the 'worst case only for S_+' approach (worst case only for BM_+ patients in the treatment group S) quite substantially underestimates the treatment effect E_+ versus S for all scenarios.
When the 'worst-case' approach (worst-case for BM_+ patients in both treatment groups C and S) is applied, the TEE results in a mean bias close to zero in the setting when a treatment disadvantage is assumed for biomarker-positive patients in treatment S compared to C ($(d_{CS})^+ = -0.21$) and low heterogeneity ($\tau \in \{0.001, 0.5\}$). The TEE obtained by the 'worst-case' approach has a higher mean bias in all other settings, which increases with an increasing number of studies, i.e. an increasing proportion of mixed patient-population studies. The estimation based on the 'best-case for S_+' approach (best case only for biomarker-positive patients in the standard treatment group (S)) results in larger mean bias compared to the 'best-case' approach when no large treatment benefit or a treatment disadvantage for biomarker-positive patients is assumed for treatment S compared to C ($(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$). When a treatment benefit for biomarker-positive patients in treatment S compared to C is assumed ($(d_{CS})^+ = 1.25$), the 'best-case for S_+' approach results in a mean bias close to zero and especially a lower bias compared to all other approaches.
The naive, regression, 'biomarker not predictive' (Biomarker status not predictive for response rate) approaches, and enriching-through-weighting with x, in the following referred to as "enriching approach", perform comparably well for $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$ and low heterogeneity ($\tau \in \{0.001, 0.5\}$). For $(d_{CS})^+ = -0.21$, the treatment effect seems to be slightly underestimated by these approach whereas for $(d_{CS})^+ = 0.4$ the treatment effect seems to be slightly overestimated. With higher heterogeneity, the missing data approaches result in a higher bias compared to the naive, regression, and enriching approaches. When biomarker-positive patients are assumed to have a treatment effect benefit for S compared to C) ($(d_{CS})^+ = 1.25$), the 'best-case only for S_+' and regression approaches result in a mean bias close to zero, whereas the naive, the 'biomarker not predictive', enriching, and 'best-case' approaches overestimate the treatment effect for E_+ compared to S.
In general, it can be observed that the mean bias increases with an increasing assumed heterogeneity value and that the regression approach leads overall to the least mean bias.

3.3.2 RMSE

As a second performance measure, the Root-Mean-Square Error (RMSE) of the estimated treatment effects was calculated. Compared to the mean bias, it gives more weight to larger deviations from the true treatment effect in the different simulation runs. When the proportion of biomarker-positive patients lies in the range between 0.4 and 0.5 (see Figure A.2 in the Appendix), it can be noticed that the RMSE is increased

for the 'worst-case' and the 'worst-case only for S_+' approach compared to the wider range of x ($x_j \in [0.3, 0.7]$). When the assumed heterogeneity is high ($\tau = 2$), then the 'worst-case' combined with a small range of x_j leads to an even higher RMSE than the 'worst-case only for S_+' approach. For the regression approach it can also be seen that a higher heterogeneity ($\tau \in \{1, 2\}$) combined with smaller range of x_j seems to lead to higher RMSEs. In the following, the focus lies on the proportion range of $x_j \in [0.3, 0.7]$ and the results of the RMSE by the different models (shapes) and scenarios are displayed in 3.3.

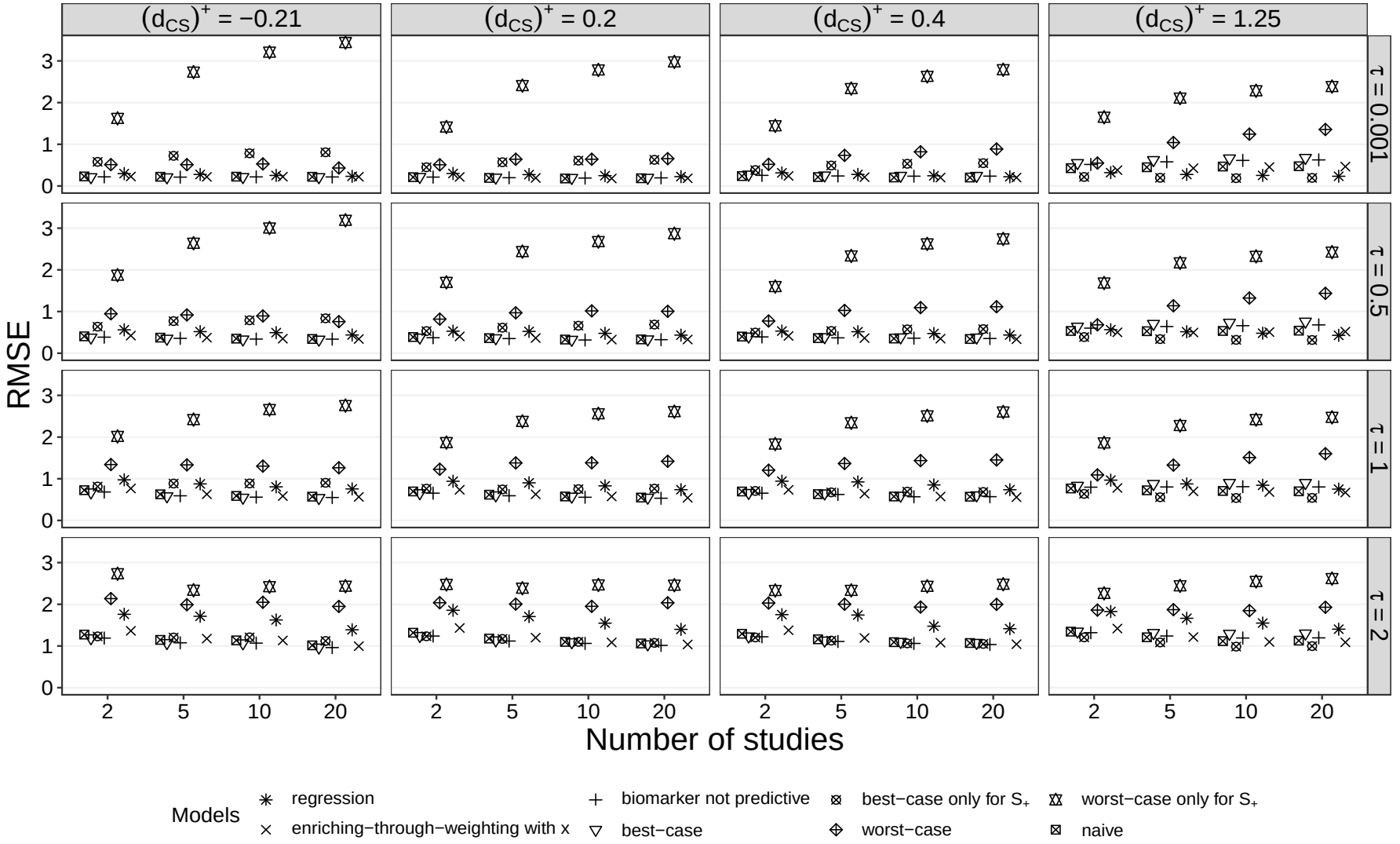


Figure 3.3: *RMSE of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A)*

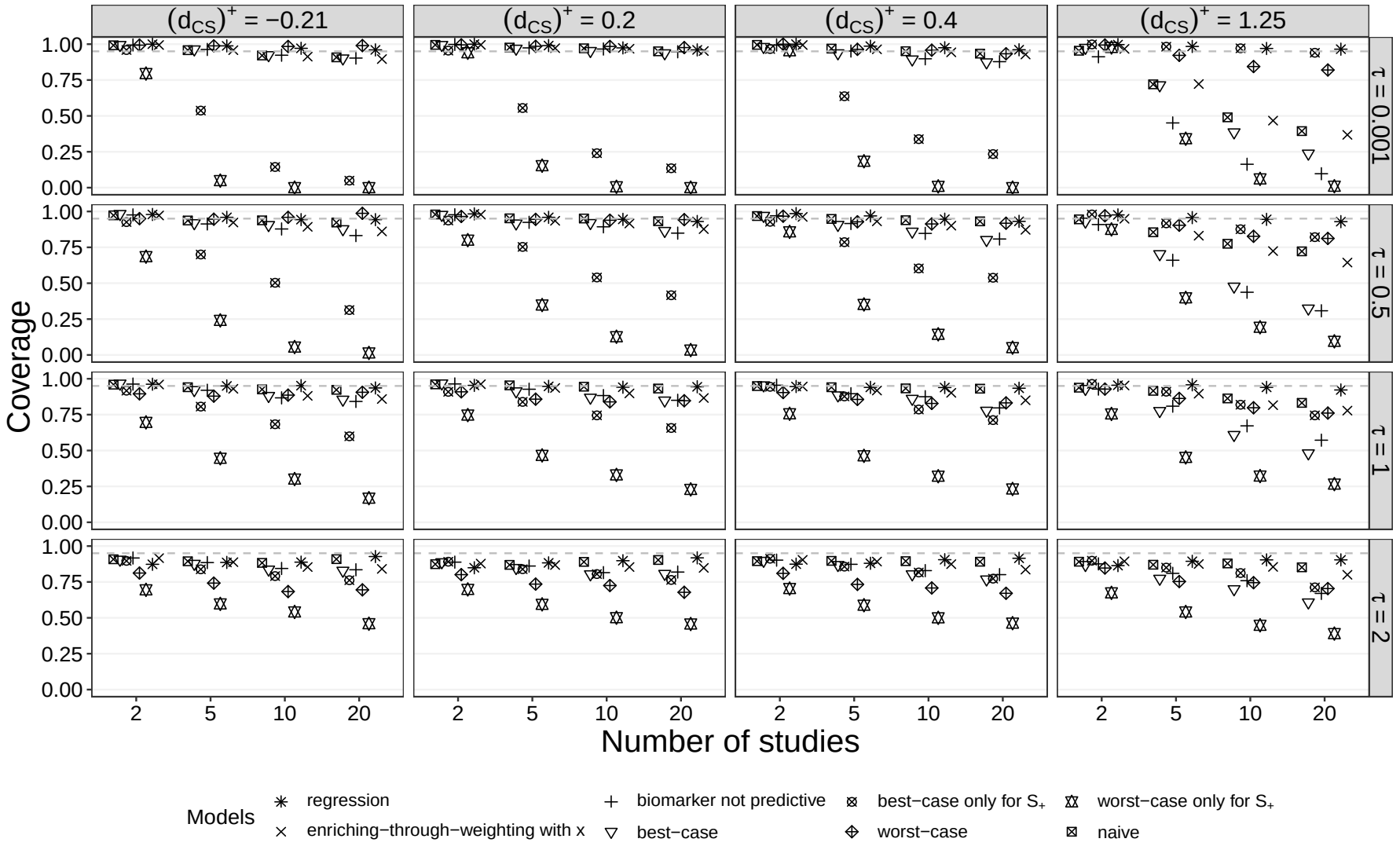
For all values of $(d_{CS})^+$, the estimation of τ with the 'worst-case only for S_+' approach leads to a noticeably higher RMSE compared to all other approaches.

Figure 3.3 illustrates that, for most scenarios apart from $(d_{CS})^+ = -0.21$ and $\tau = 0.001$, the application of the 'worst-case' model leads to the highest RMSE compared to all other models apart from the 'worst-case for S_+' approach. Comparing these other approaches for the scenarios $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$ and low to medium heterogeneity ($\tau \in \{0.001, 0.5, 1\}$), it can be seen that the 'best-case only for S_+' approach leads to a higher RMSE compared to the naive, best-case, biomarker not predictive, regression and enriching approaches. When there is medium to high heterogeneity assumed ($\tau \in \{1, 2\}$), using the regression approach results in a higher RMSE compared to the naive, 'best-case', 'best-case only for S_+', biomarker not predictive, and enriching models. For $(d_{CS})^+ = 1.25$, the estimation with the 'best-case only for S_+' approach leads to the lowest RMSE. In general (apart from the 'worst-case' and 'worst-case only for S_+'approaches) it can be noticed that the RMSE decreases with an increasing number of studies and increases with an increasing value of heterogeneity, resulting for $\tau = 2$ in a lower bound of the RMSE close to 1.

3.3.3 Coverage

Comparing the results of the wider proportion range ($x_j \in [0.3, 0.7]$) illustrated in Figure 3.4 to the smaller range ($x_j \in [0.4, 0.5]$) (see Figure A.3 in the appendix), it can be seen that the coverage of the approaches 'worst-case', 'worst-case only S_+', 'best-case only for S_+' is lower for a small proportion range compared to a wider proportion range. For the 'worst-case only for S_+' and 'best-case only for S_+' approaches with the range of $x_j \in [0.4, 0.5]$, low heterogeneity ($\tau \in \{0.001, 0.5\}$, and $n \geq 5$, the coverage is close to zero. Looking at the coverage of the other models, the covariate range does not seem to make a difference in the performance. The focus in the following is therefore lies on the wider proportion range ($x_j \in [0.3, 0.7]$).

In general, it can be seen that the 'worst-case only for S_+' and the 'best-case only for S_+ approach result in the lowest coverage compared to the other approaches.



A grey line indicates the nominal level of 0.95 in each panel.

Figure 3.4: *Coverage of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A)*

In the situation of a very low heterogeneity ($\tau = 0.001$) and $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$, almost all models, apart from the 'best-case only for S_+' and the 'worst-case only for S_+' approaches, have a coverage around 0.95. Throughout all scenarios, the regression approach results in a consistently high coverage close to 0.95. For $(d_{CS})^+ = 1.25$, the coverage of the biomarker not predictive, best-case, naive, and enriching approaches decreases when the number of studies increases, but their coverage increases with increasing heterogeneity. For $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$, the coverage of the 'best-case' approach increases with an increasing heterogeneity whereas the coverage of the other approaches slightly decreases. When the number of study increases, the coverage of the missing data approaches decreases. In case of high heterogeneity ($\tau = 2$), no approach results in a coverage over 0.95, but the lowest coverage is at 0.55 for all approaches apart from the 'worst-case only for S_+' approach.
Overall it can be noted, that the coverage decreases with increasing $(d_{CS})^+$, increasing τ and an increasing number of studies. The regression approach has overall the highest coverage.

3.3.4 Precision

The precision of the estimated treatment effect, given as the mean credible interval width is displayed in Figure 3.5 for a proportion range of [0.3, 0.7]. The results of the mean credible interval widths for the wider proportion range of [0.4, 0.5] are only displayed in the appendix in Figure A.4.

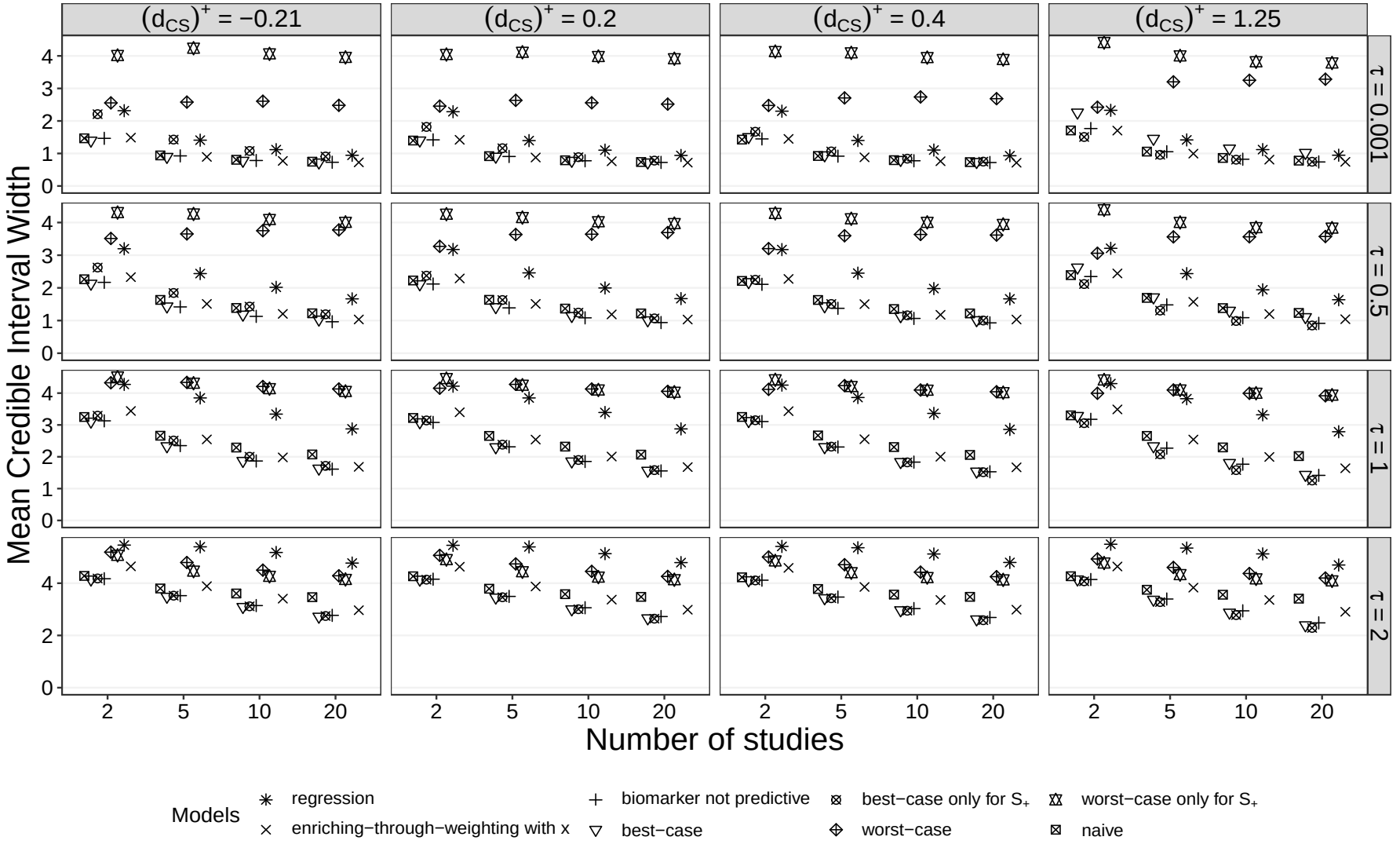


Figure 3.5: *Mean 95% credible width of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A)*

For the approaches 'worst-case', 'worst-case only for S_+' and regression, the mean precision with $x_j \in [0.4,\ 0.5]$ is lower compared to the mean precision of the respective approaches when a wider range for x_j ($x_j \in [0.3,\ 0.7]$) is given. For the other models, the proportion of BM_+ patients per study (x_j) does not visibly impact the mean credible interval widths. Therefore, the following evaluation focuses on $x_j \in [0.3,\ 0.7]$.

The network meta-analyses using the worst-case or the 'worst-case only for S_+' approaches result in a low precision throughout all scenarios. When the heterogeneity value is low to medium ($\tau \in \{0.001, 0.5, 1\}$) the precision is lower compared to the other approaches. The network meta-regression approach has the lowest precision compared to the other approaches when the heterogeneity is very high ($\tau = 2$). Conducting a network meta-analysis with the naive, best-case, best-case only for S_+, biomarker not predictive, and enriching approaches results in similar precision values which are higher compared to the regression, worst-case and 'worst-case only for $_+$' approaches. In general the precision increases with an increasing number of studies and decreases with increasing heterogeneity.

3.3.5 Power

The power of the different approaches and scenarios has been evaluated in the simulation study based on the mean treatment effect significance measured as the percentage of treatment effect estimates not including zero in the respective credible intervals. The results are displayed in Figure 3.6 for $x_j \in [0.3,\ 0.7]$.

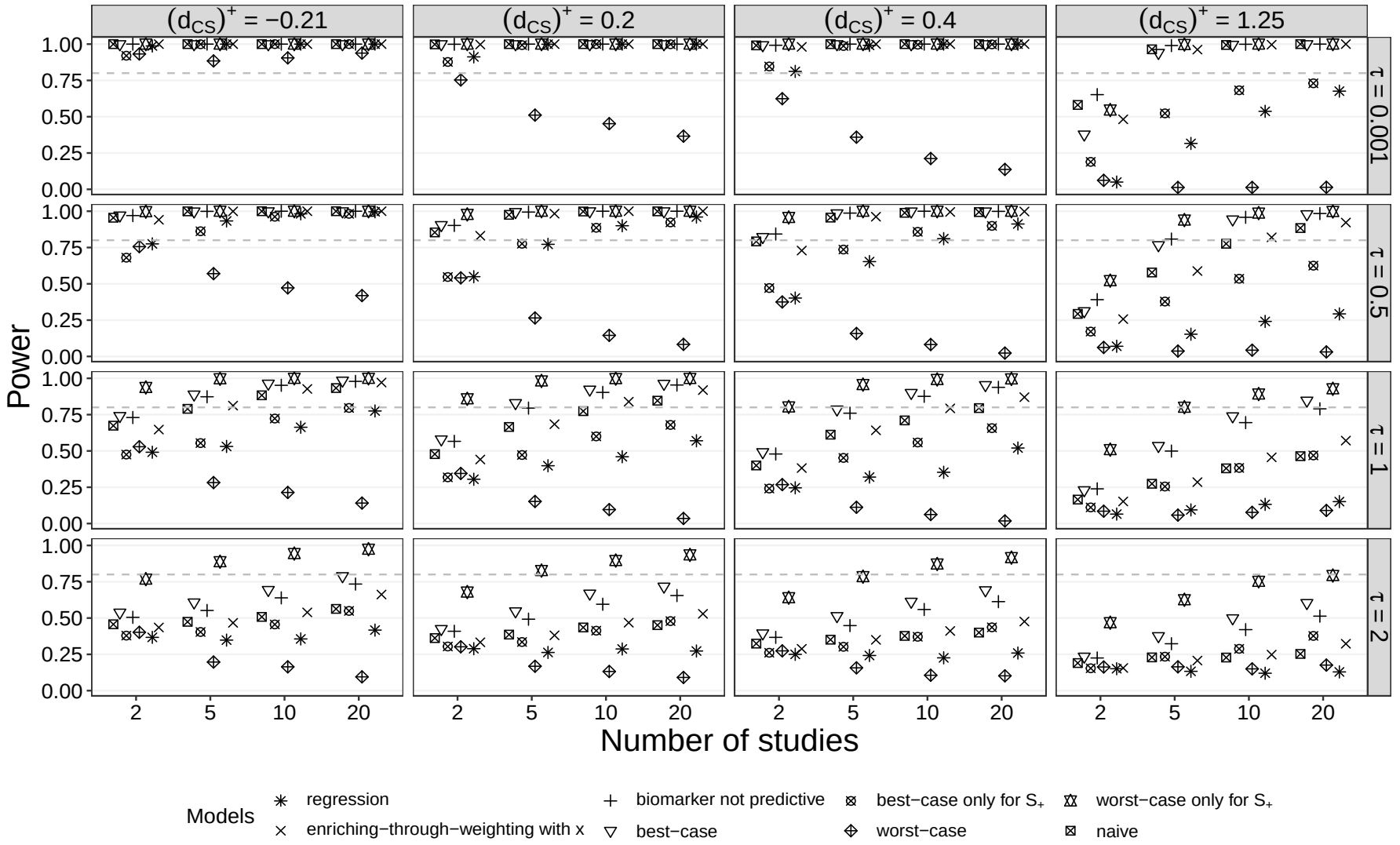


Figure 3.6: *Power of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A)*

In the setting of an assumed treatment effect with $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$ and a low heterogeneity $\tau \in \{0.001, 0.5\}$, the proportion of BM_+ patients per study does not seem to influence the power of most methods (see Figure A.5 in the Appendix). Only for the worst-case and regression approaches it can be noticed that a smaller proportion range ($x_j \in [0.4, 0.5]$) leads to an increased power. For $(d_{CS})^+ = 1.25$, a smaller proportion range leads to a decreased power for the regression approach, and an increased power for the other approaches compared to a wider range ($x_j \in [0.4, 0.5]$). With an increasing heterogeneity, the different proportion ranges lead to more similar power results. The results discussed in the following are based on the proportion range of BM_+ 0.3 to 0.7 as displayed in Figure 3.6.

Given the situation of a very low heterogeneity ($\tau = 0.001$) and $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$ or a slightly increased heterogeneity ($\tau = 0.5$) and more than five studies, the power is over 80% for nearly all approaches apart from the worst-case approach. When $(d_{CS})^+ = 1.25$ and $n > 5$, the naive, worst-case only for S_+, the biomarker not predictive, the best-case, and the enriching approach lead to a power of over 80%. The worst-case approach leads to the lowest power throughout all scenarios. The regression approach results in better power as the worst-case approach but lower power than the other approaches. The highest power throughout all settings is reached with the worst-case only for S_+ approach. The power increases with the number of studies increasing and decreases with increasing heterogeneity.

3.4 Simulation study: proportion of biomarker-positive patients $x \in (0, 1]$ (setting B)

The results evaluated in this section are based on the assumptions of setting B, as introduced in Chapter 1. The main difference to setting A is that for the comparison of C versus S two studies with only 100% biomarker-positive patients where included additionally (see Table 2.2). In this setting the missing data approach was not evaluated, because there was no great advantage compared to other methods available to be expected. Instead, the enriching-through-weighting approach with different weights, as well as the informative prior approach and the stand-alone approach were evaluated as alternative approaches to the naive, network meta-regression and enriching-through-weighting with x approach. The same performance measures as above (Table 2.3) were applied.

Using different ranges for the proportion of biomarker-positive studies in the simulation study did for a small proportion range not result in notable differences in the performance measures for this setting. This is why only results for $x_j \in [0.3, 0.7]$ are evaluated in the following. For the sake of completeness the results of the smaller range [0.4, 0.5] in the different simulation settings are presented in the appendix (Chapter A, Figures

A.6, A.7, A.8, A.9, A.10).
In this present setting, a weakly informative heterogeneity prior $\tau \sim \mathcal{HN}(0.5)$ was evaluated additionally to the non-informative prior $\tau \sim \mathcal{U}(0, 2)$ as already discussed in Chapter 2. For both, the results of the simulation study are discussed in the following. The Figures for the results of the half-normal prior are displayed in the Appendix A.2.

As above, the columns vary with the assumed theoretic treatment effects for C versus S, while the rows display different levels of between-trial heterogeneity. The x-axis displays the number of studies included in the comparison of S versus C whereas the y-axis displays the performance measures. The different shapes mark the different methods used to estimate the treatment effect of comparing E_+ with S.

3.4.1 Bias

Figure 3.7 displays the mean bias of the estimated treatment effect $(d_{CS})^+$ for biomarker-positive patients using the different models described in the chapter above and different simulation scenarios.

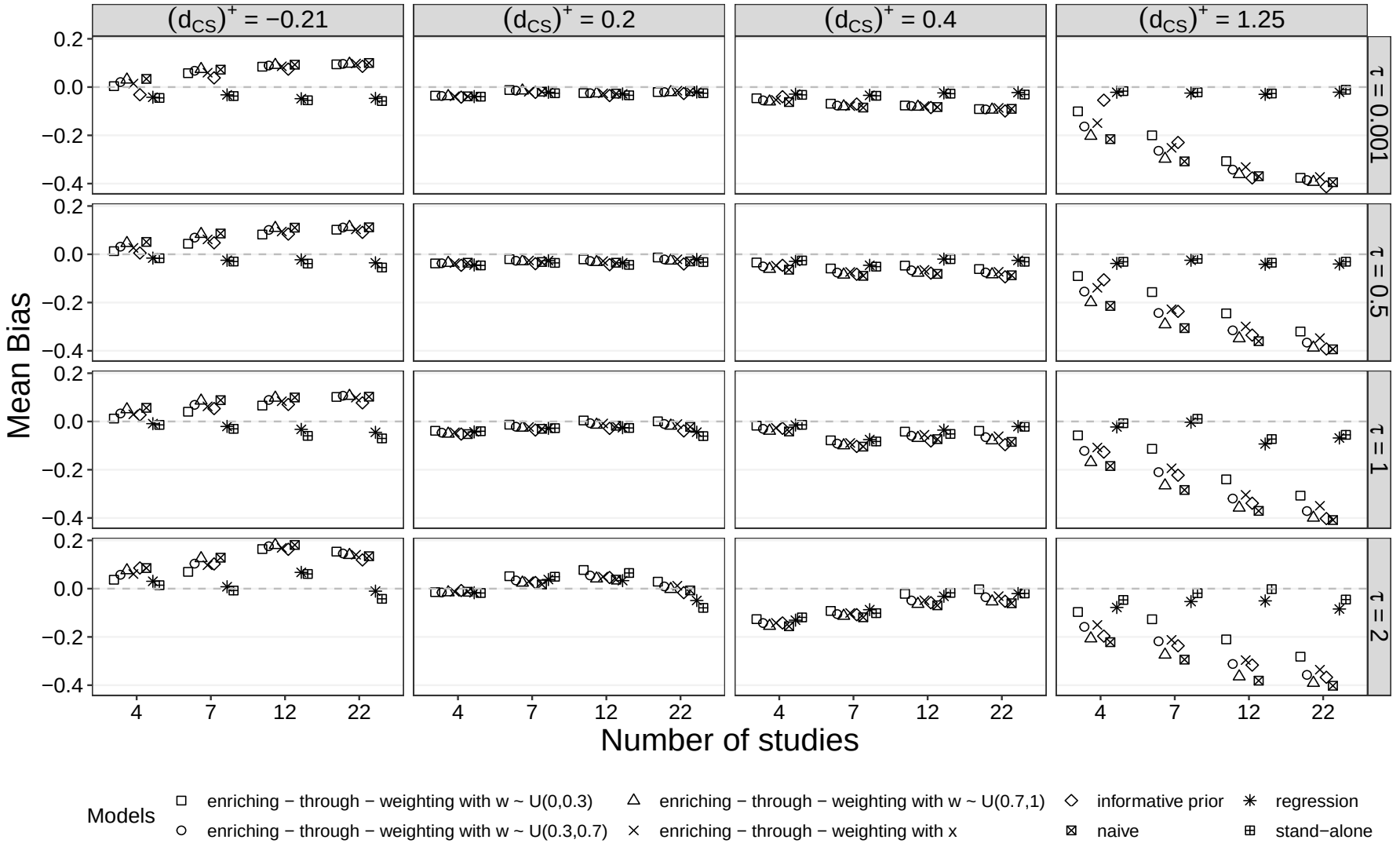


Figure 3.7: *Mean bias of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B)*

It can be seen that the treatment effect estimates obtained with regression as well as with the stand-alone approach have the lowest absolute mean bias throughout all investigated parameter settings. For $(d_{CS})^+ = -0.21$, assuming that BM_+ - patients perform worse than BM_- - patients, the estimation of the treatment effect leads to an overestimation of the true treatment effect for most approaches except for the regression and stand-alone analysis, which seem to underestimate the true treatment effect. The mean bias increases with an increasing number of studies i.e., an increasing proportion of mixed studies comparing S versus C and an increasing heterogeneity. For $(d_{CS})^+ = 0.2$, assuming that BM_+-patients have the same treatment effect as BM_--patients, a mean bias close to 0 can be noted independently of the number of studies in CS. This mean bias increases slightly with an increasing heterogeneity. For the treatment effect estimate $(d_{CS})^+ = 0.4$ which corresponds to the assumption of a slightly better treatment advantage for BM_+-patients compared to BM_- - patients, all models lead to an underestimation of the true treatment effect. Assuming a greater treatment advantage in BM_+ - patients compared to BM_- - patients, $(d_{CS})^+ = 1.25$, leads to an underestimation of the true treatment effect for BM_+ patients which is more pronounced for increasing proportion of mixed studies (higher n) for most approaches. Comparing specification of the enriching-through-weighting approaches, it can be seen that the approach with $w \sim \mathcal{U}(0, 0.3)$, and therefore the approach with the least trust in the mixed data, leads to a slightly decreased bias compared to the other enriching-through-weighting approaches.
The estimated treatment effects using a half-normal heterogeneity prior (see Figure A.11) instead of a uniform prior result in a similar mean bias compared to the use of a uniform prior.

3.4.2 RMSE

Figure 3.8 displays the Root-Mean-Squared-Error (RMSE) of estimated treatment effects comparing E_+ with S for the above-discussed scenarios. Care has to be taken when interpreting the plot because the scales on the y-axis are expanding with increasing heterogeneity to allow the illustration of the differences with increasing between-trial heterogeneity.

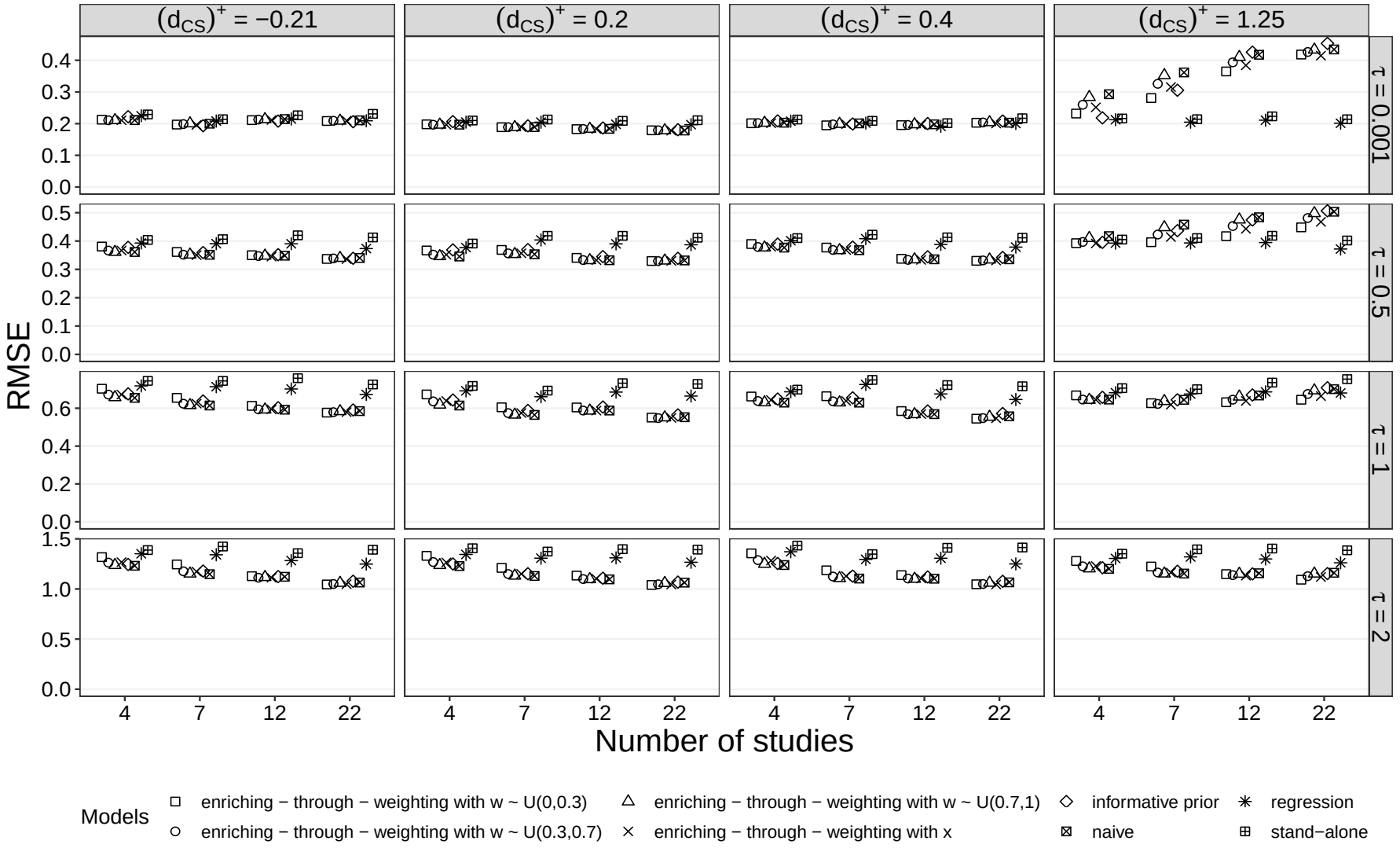
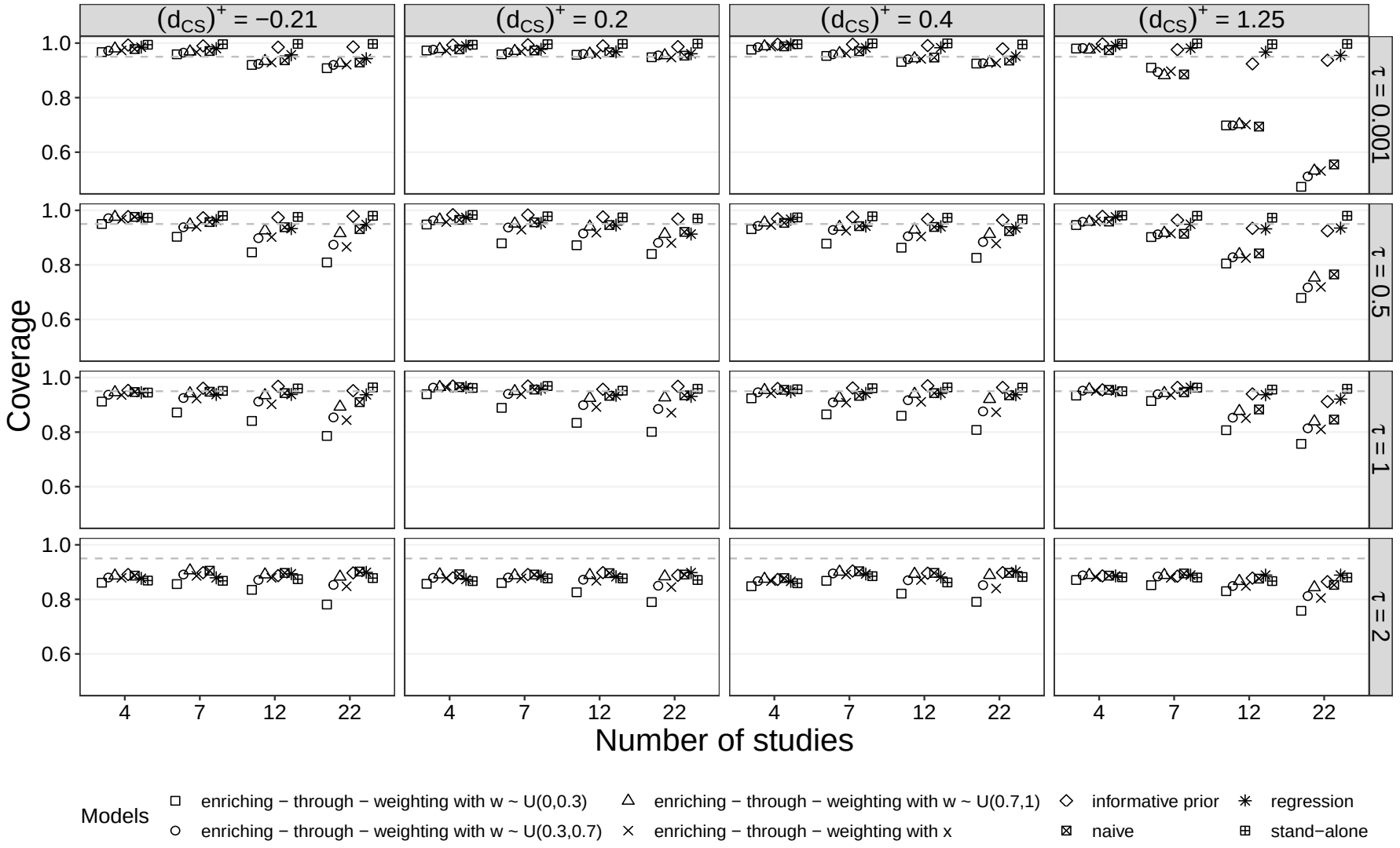


Figure 3.8: *RMSE of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B)*

When a treatment effect of $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$ and a heterogeneity of $\tau = 0.001$ is given, there is only a slight difference visible between the different models. For $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$ and at least a moderate heterogeneity, i.e., $\tau = 0.5$, it can be seen that the regression and the stand-alone approaches perform worse than the other approaches and even more so for increasing study numbers and increasing heterogeneity. In the setting where the treatment effect for biomarker-positive patients is higher than the treatment effect for biomarker-negative patients $(d_{CS})^+ = 1.25$, the performance of the regression and stand-alone approach is dependent on the heterogeneity. For low heterogeneity ($\tau \in \{0.001, 0.5\}$), these approaches perform better than the naive, informative, or enriching-through-weighting models. With increasing heterogeneity ($\tau \in \{1, 2\}$) however the RMSE of these approaches is higher compared to the RMSE of the other approaches.
The estimated treatment effects using a half-normal heterogeneity prior result in similar RMSE compared to the use of a uniform prior (see Figure A.12).

3.4.3 Coverage

Figure 3.9 displays the coverage for the TEE of the comparison E_+ versus S in biomarker-positive patients. A grey line indicates the nominal level of 95% in each panel.



A grey line indicates the nominal level of 0.95 in each panel.

Figure 3.9: *Coverage of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B)*

In general, it can be noticed that the coverage of all models are equal to or above 95% for $\tau = 0.001$ under the treatment effect $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$. For $\tau = 0.5$, the informative prior, stand-alone, and regression approach result in a coverage close to 95%. The other approaches result for $n \geq 7$ in a coverage smaller than 95% and the coverage decreases with an increasing number of studies. For $(d_{CS})^+ = 1.25$ and $\tau \in \{0.001, 0.5, 1\}$, the informative prior, regression, and stand-alone approaches still result in a coverage close to 95%. The enriching-through-weighting approaches only leads to the desired coverage of 95% for $n \in \{4, 7\}$. For $n > 7$ these models lead to coverage well below the 95%. The coverage of these approaches decreases with an increasing number of mixed studies to a coverage below 65% for $n = 22$ when the heterogeneity is low. For $\tau = 1$ and $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$, the coverage for the approaches informative prior, stand-alone regression, and enriching-through-weighting with $w \sim \mathcal{U}(0.7, 1)$ is still >90%. For $(d_{CS})^+ = 1.25$ and $n \geq 12$ it can be seen that the enriching-through-weighting approaches and the naive approaches have clearly lower coverages. The enriching-through-weighting approach with $w \sim \mathcal{U}(0, 0.3)$ leads to the lowest coverage for moderate and higher heterogeneity ($\tau \in \{0.5, 1, 2\}$). For a high between-trial heterogeneity ($\tau = 2$), the coverage of all approaches is below 95%.
The resulting coverage when using a half-normal prior $\tau \sim \mathcal{HN}(0.5)$ (see Figure A.13 in the Appendix) is lower when $\tau = 0.001$ and $(d_{CS})^+ = 1.25$ for most models compared to the usage of the uniform prior. For the regression and stand-alone approach there is no difference visible. For the same low τ and $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$, there is no considerable difference when using these respective heterogeneity priors. For intermediate to high heterogeneity values, $\tau \in \{0.5, 1, 2\}$ the coverage is lower for all treatment effects for all settings when a half-normal heterogeneity prior ($\mathcal{HN}(0.5)$) is used compared to the uniform prior. It can be noted that for $\tau = 0.5$ the stand-alone and regression approach still reach a coverage close to 95% when using the half-normal prior.

3.4.4 Precision

The mean precision displayed in Figure 3.10 is measured as the mean 95% credible interval width of the estimated treatment effects. Care has to be taken, when interpreting the illustrated results, because the y-axis scale increases with an increasing heterogeneity for a clearer visualization of results.

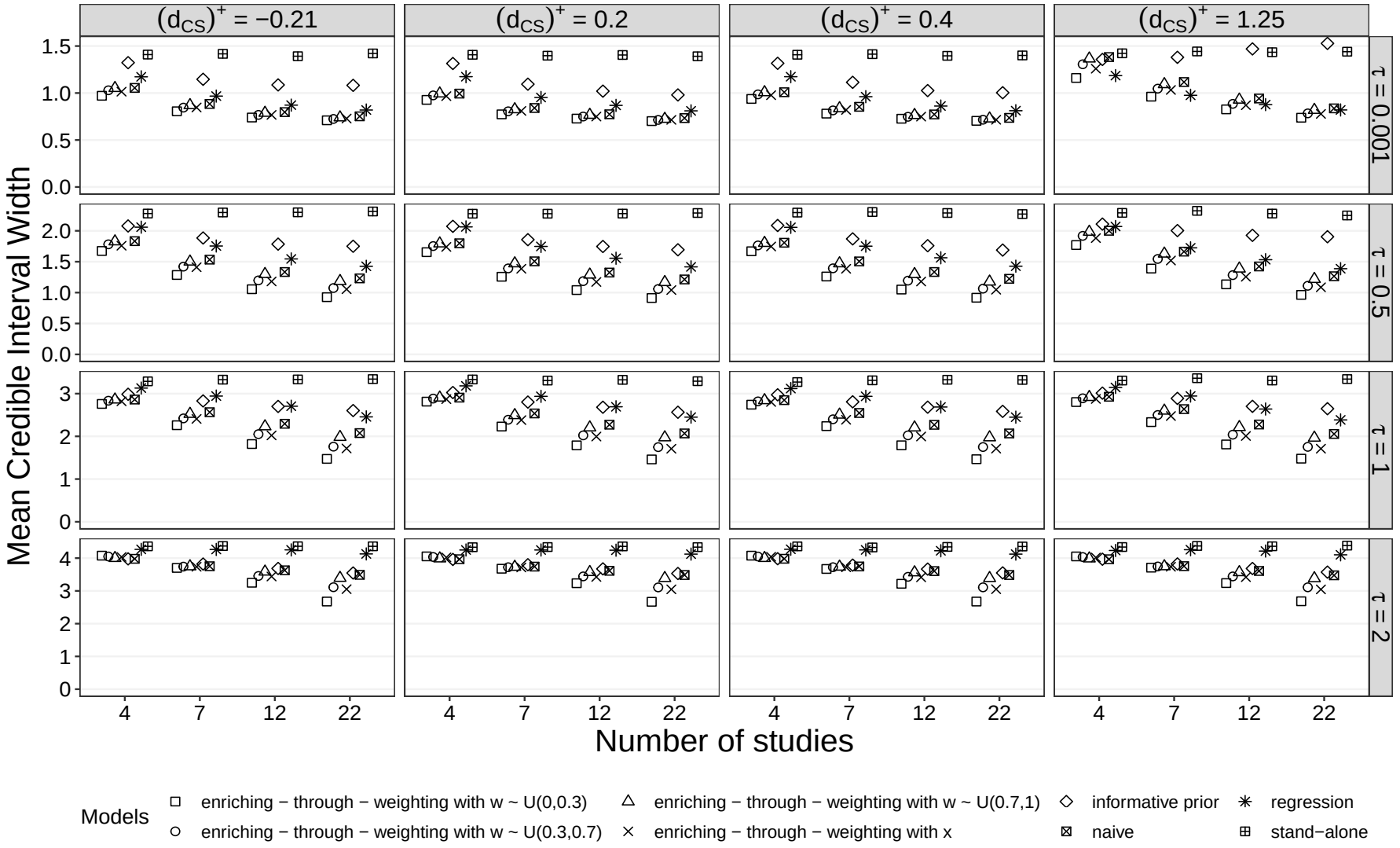


Figure 3.10: *Mean 95% credible width of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B)*

Using the stand-alone analysis leads to a low precision throughout all parameter settings. This was to be expected, because only two studies were included in the contrast C versus S. When the informative prior model is used to estimate the treatment effects, the resulting precision is still quite low but increases with a rising number of studies apart from the setting $(d_{CS})^+ = 1.25$, and $\tau = 0.001$. The highest precision for the treatment effect estimate was achieved with the enriching-through-weighting approaches, especially with $w \sim \mathcal{U}(0, 0.3)$. An increase in the heterogeneity parameter τ leads to a decrease in precision, which can especially be seen in the increase of the scaling parameter on the y-axis. The regression approach results in a decreased precision compared to the enriching-through-weighting and the stand-alone models especially for higher between-trial heterogeneity values. It can also be noticed that the enriching-through-weighting models and the naive approach lead to the highest precision values for all settings.
The use of a half-normal prior in the setting of a very low heterogeneity $\tau = 0.001$ results in higher precision for the naive and informative prior model compared to the uniform prior. When the network meta-analysis is conducted with the other methods, there is no visible difference when different heterogeneity priors are used (see Figure A.14 in the Appendix). The resulting precision for $\tau \in \{0.5, 1, 2\}$ in all settings is higher when a half-normal heterogeneity prior was used instead of a uniform heterogeneity prior.

3.4.5 Power

The power of the single approaches, measured as the percentage of the estimated credible interval not including zero, is displayed in Figure 3.11. The grey line indicates the desired power of 80%.

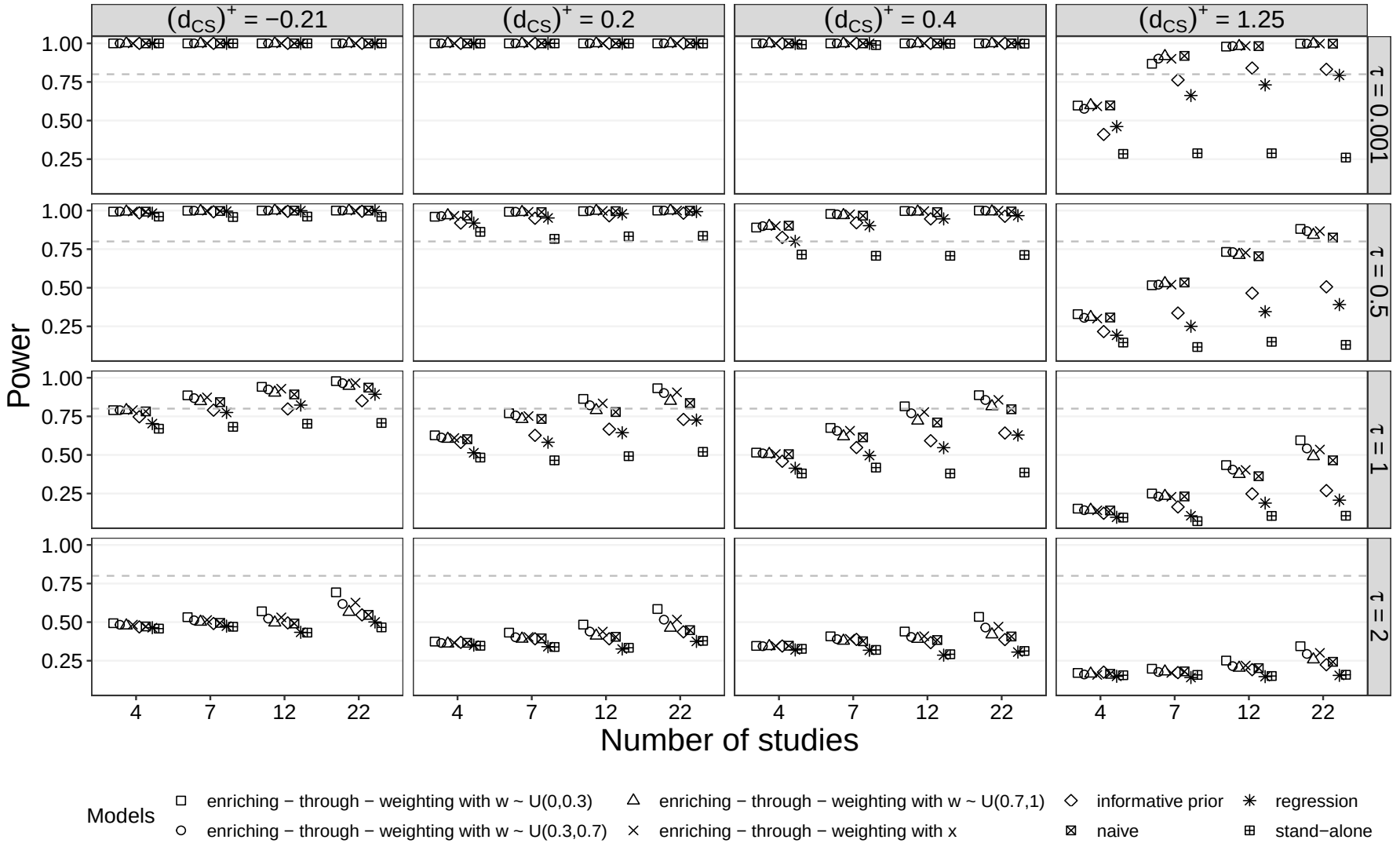


Figure 3.11: *Power of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B)*

The power is close to 100% for most approaches with low and moderate heterogeneity ($\tau \in \{0.001, 0.5\}$) and $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$. Only the stand-alone approach has a lower power for $(d_{CS})^+ \in \{0.2, 0.4\}$ and $\tau = 0.5$ The enriching-through-weighting approaches perform better or equal to the other approaches through all parameter settings with regards to the power. The power decreases with increasing heterogeneity and decreasing (d_{CS}^+) and increases with the number of studies, i.e. proportion of mixed studies for CS. For $\tau = 1$ and $(d_{CS})^+ = 2$ or $\tau = 2$, the power of 80% is not achieved by any of the models.
The use of the half-normal heterogeneity prior does not lead to a difference in the resulting power of all models when $\tau = 0.001$ and $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$ (see Appendix Figure A.15). For $(d_{CS})^+ = 1.25$, the power for all given heterogeneity values τ is higher when using a half-normal prior compared to a uniform prior. For the stand-alone model with $\tau = 0.5$ and $(d_{CS})^+ \in \{0.2, 0.4\}$, it can be noted that the resulting power is higher when a half-normal prior is used compared to a uniform prior. Especially for the treatment effect $(d_{CS})^+ = 0.4$, the difference is considerable. A power of over 80% is still reached when using a half-normal prior compared to less than 80% power when using a uniform prior. For intermediate or high heterogeneity values $\tau \in \{1, 2\}$, the resulting power is slightly lower for all models when using a uniform prior compared to the resulting power using a half-normal heterogeneity prior.

3.5 Setting A versus setting B

In the two sections above, the results of the different methods given different scenarios for the two respective settings (A and B) have been evaluated. In the following the performances of the naive, regression and enriching - through - weighting with x approaches are compared in more detail between setting A and setting B, because only these three approaches have been used in both settings. The proportion of BM_+ patients for this comparison is chosen to be between 0.3 and 0.7, and the heterogeneity prior is chosen to be uniform ($\tau \sim \mathcal{U}(0, 2)$).
For the evaluated settings of the simulation study, it can be noted that adding two studies with 100% biomarker-positive patients results in a less biased estimation for the treatment effect comparing E_+ versus S. The RMSE is also smaller when these two studies are added in a NMA. It leads to better coverage, precision, and higher power. In the following, the results of the comparison of setting A versus setting B are illustrated and explained in more detail with the focus of this evaluation on the comparison of the performance of the models in these different settings.

3.5.1 Bias

In Figure 3.12, the mean bias of the estimated treatment effects are displayed for the setting A and setting B marked in black and grey and the different models marked with different shapes.

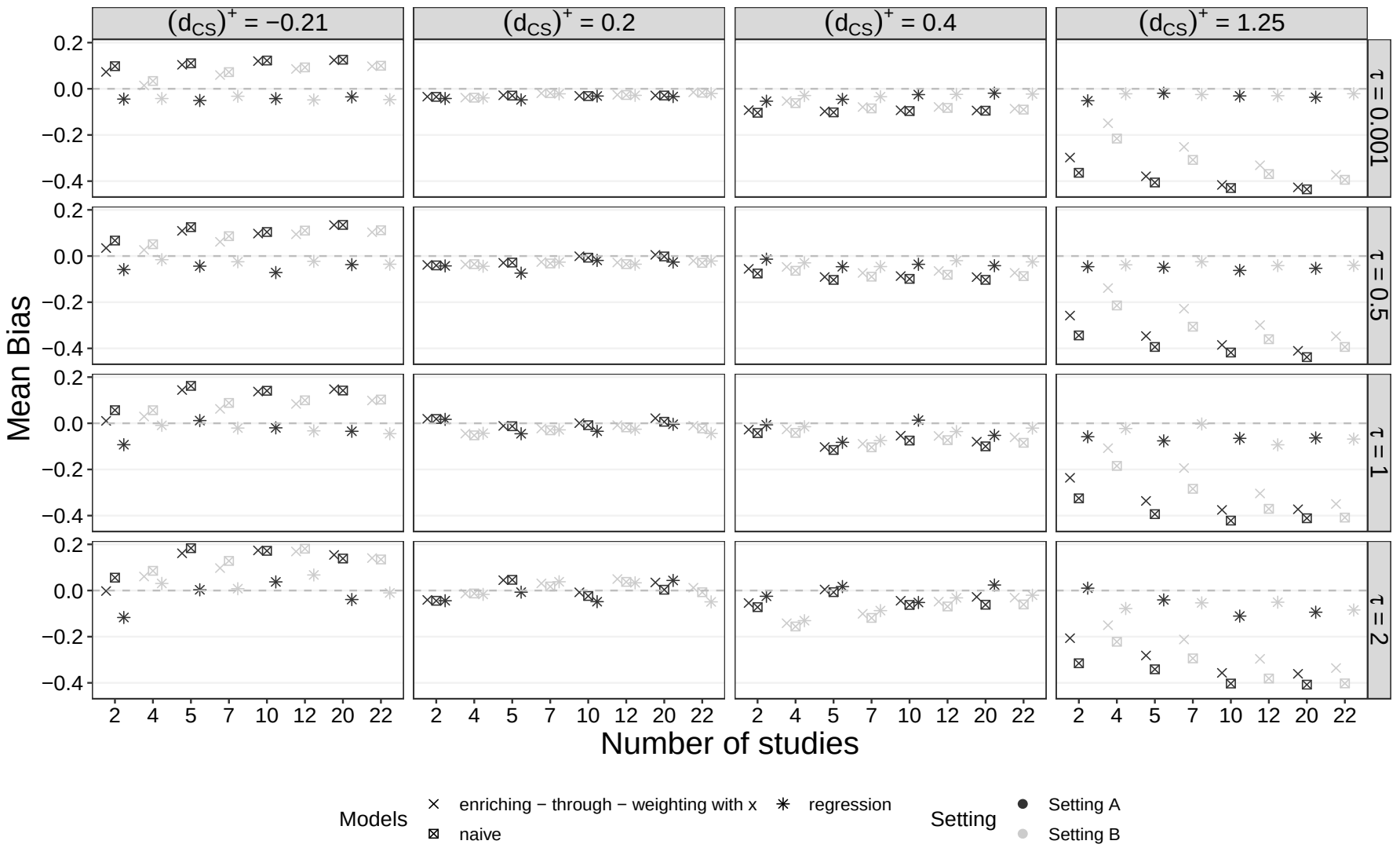


Figure 3.12: *Mean bias of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A vs. setting B)*

In almost all scenarios (apart from $(d_{CS})^+ = 0.4$, $n < 5$ and $\tau = 2$) the absolute value of the mean bias is smaller when two studies with 100% biomarker-positive patients are added for the comparison of S versus C (setting B) compared to having only mixed patient-population studies (setting A). The difference is more pronounced for the enriching-through-weighting approach and the naive approach compared to the regression approach. For $(d_{CS})^+ = 1.25$, the difference in the mean bias between the two settings is especially distinct for a small number of studies $n < 5$. When comparing the mean bias between the two settings for a different number of studies, it can be noticed that adding two 100% biomarker-positive studies results in a lower mean bias than simply adding two mixed population studies.

3.5.2 RMSE

The mean RMSE of the estimated treatment effects by different methods is displayed in Figure 3.13 for the two settings marked in black and grey. For clarity, the y-scales differ for different values of τ.

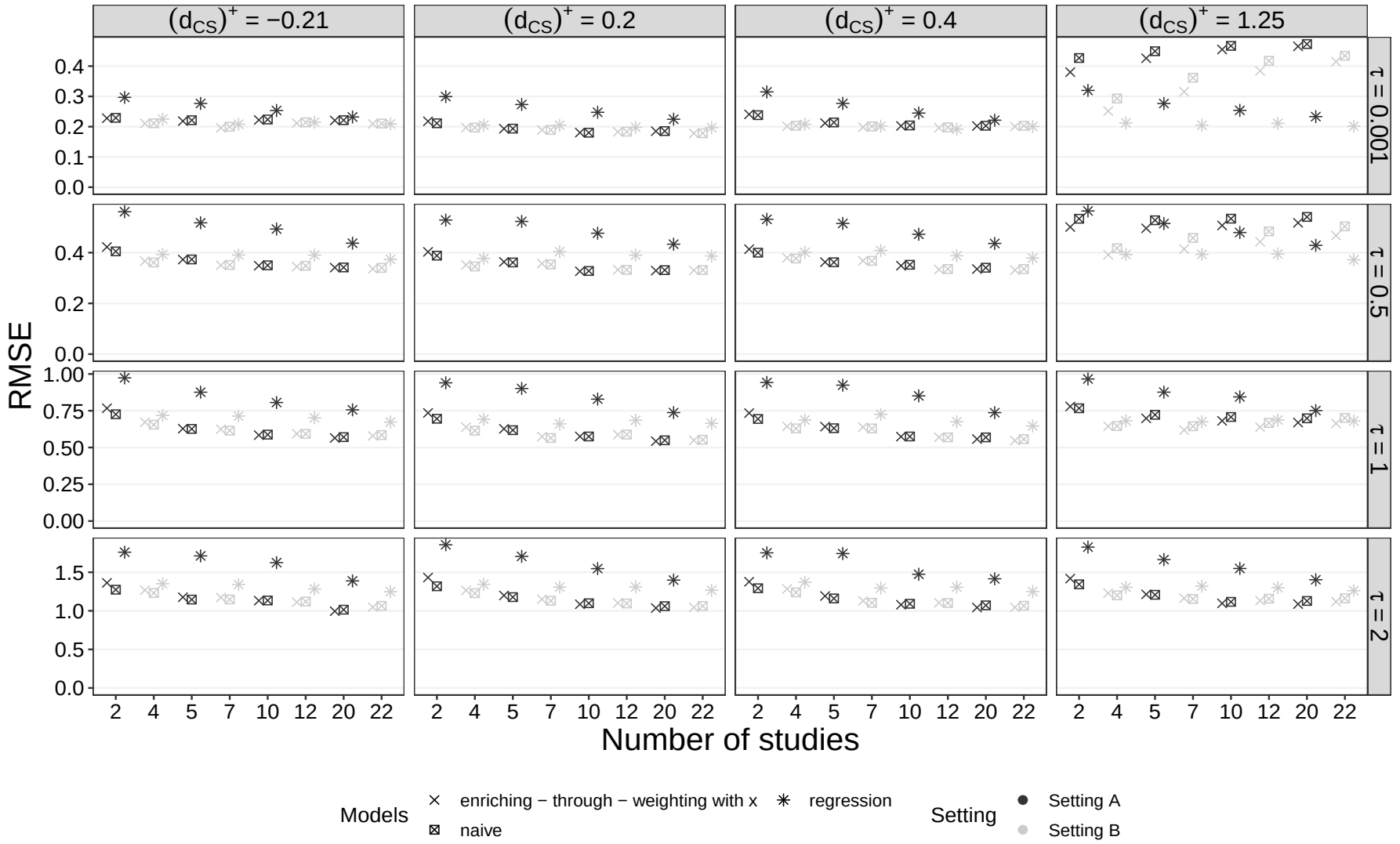
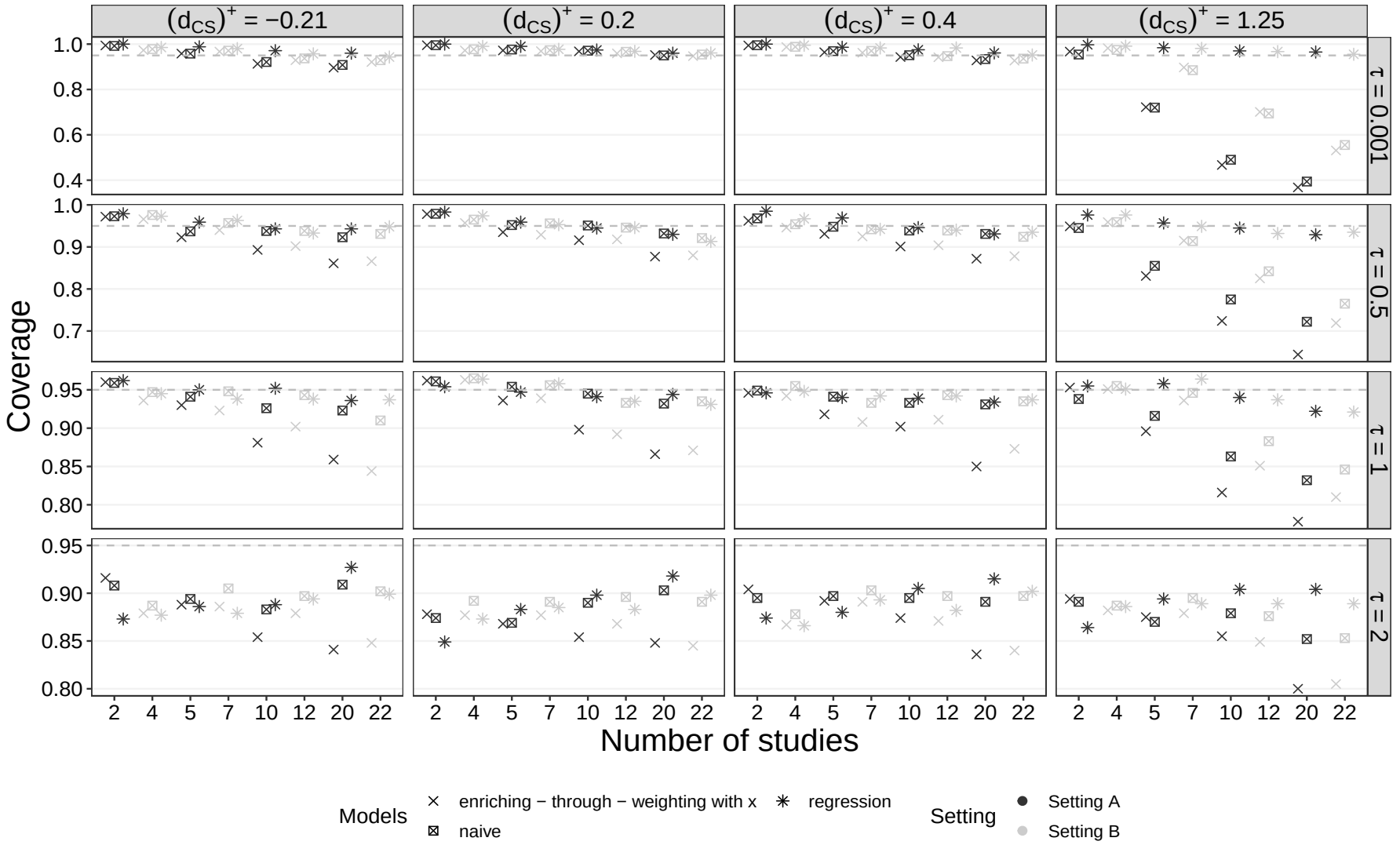


Figure 3.13: *RMSE of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A vs. setting B)*

For the regression approach, the mean RMSE of the estimated treatment effects throughout all scenarios is higher when there are no 100-% biomarker-positive studies included compared to the inclusion of two 100-% biomarker positive studies. It can also be seen that, with an increasing number of studies, the RMSE is decreasing. For the naive and the enriching-through-weighting with x approach, there is no big difference between the two settings visible when $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$ and the number of studies is larger than 2 or 4. A treatment effect of $(d_{CS})^+ = 1.25$ results in a clearly higher RMSE for all models in setting A compared to setting B.

3.5.3 Coverage

The coverage of the estimated treatment effect by using different methods is displayed in Figure 3.14 for different methods and settings. The y-scales are changing with varying τ.



A grey line indicates the nominal level of 0.95 in each panel.

Figure 3.14: *Coverage of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A vs. setting B)*

When the heterogeneity is low ($\tau \in \{0.001, 0.5\}$) and the true treatment effects $(d_{CS})^+$ are $\in \{-0.21, 0.2, 0.4\}$ there is no visible difference between the different settings. For $(d_{CS})^+ = 1.25$ and $\tau \in \{0.001, 0.5, 1\}$, the resulting coverage of the approaches enriching-through-weighting with x and naive is clearly lower for $n > 2$ in setting A compared to setting B. For a very high heterogeneity value ($\tau = 2$), setting A leads to a higher coverage compared to setting B but all methods and settings result in a coverage below the desired 95%-level.

3.5.4 Precision

The resulting mean precision given by the mean credible interval widths of the different methods is displayed in Figure 3.15. The scale of the y-axis changes with varying values of τ.

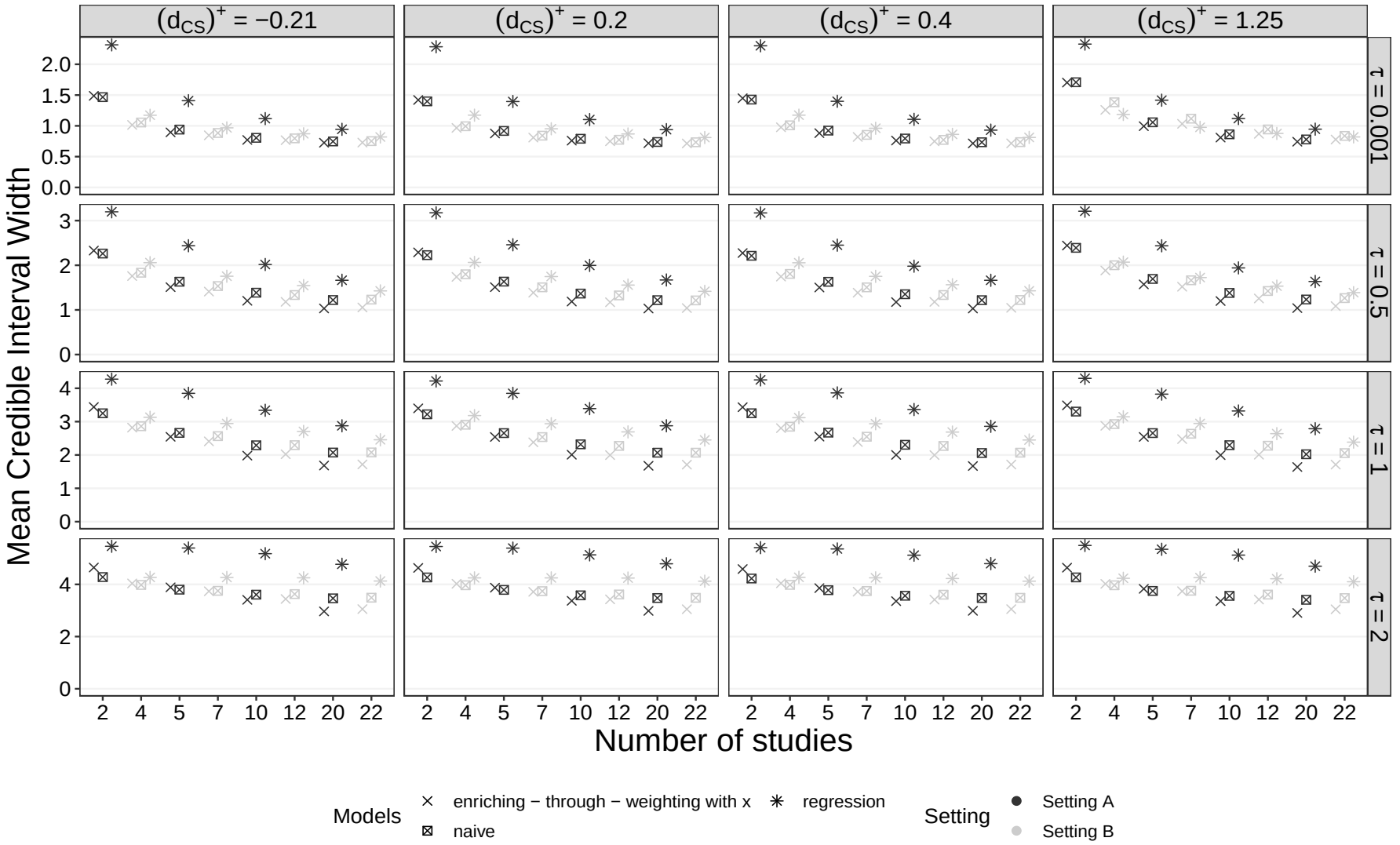


Figure 3.15: *Mean 95% credible width of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A vs. setting B)*

Overall, it can be noticed that the precision is lower, i.e. the mean credible interval width higher in setting A compared to setting B. The difference is especially distinct for the regression approach and a low number of studies. When the proportion of 100%-biomarker-positive studies decreases, the difference in precision between the settings decreases as well.

3.5.5 Power

Figure 3.16 displays the resulting power of the different methods in estimating a treatment effect comparing E_+ versus S. The power is given as the percentage of significant treatment effects.

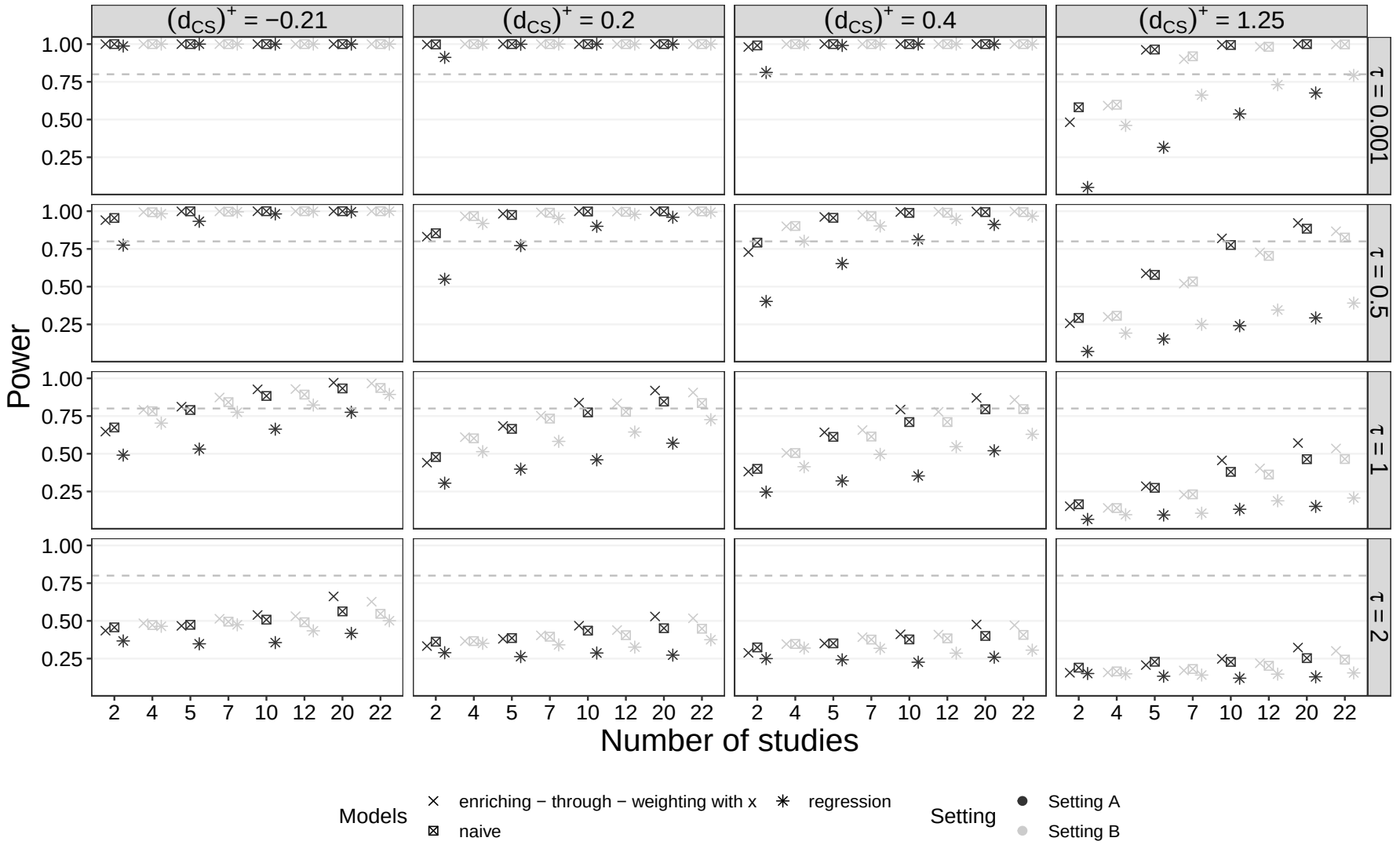


Figure 3.16: *Power of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A vs. setting B)*

The regression approach has a higher power when there are studies with 100% - biomarker-positive patients included in the comparison S versus C for $\tau \in \{0.001, 0.5, 1\}$) and the true treatment effects $(d_{CS})^+ \in \{-0.21, 0.2, 0.4\}$. For the treatment effect $(d_{CS})^+ = 1.25$, all three models result in a higher power for setting A compared to setting B.

Chapter 4

Discussion

This chapter summarizes the contributions of this thesis to research. Moreover, limitations and directions for future research are presented.

4.1 Contributions to research and discussion

The aim of this thesis was to evaluate methods to estimate a pooled treatment effect for a specific patient subgroup in a network meta-analysis based on direct and indirect evidence. In this setting, targeted as well as non-targeted therapies where included. Evidence on non-targeted therapies was characterized by the response rates for an overall population consisting of biomarker-positive and -negative patients and the percentage of biomarker-positive patients. Two settings were evaluated in a simulation study. In setting 'A', it was assumed that no studies with 100% biomarker-positive patients existed for the comparison of a standard treatment S with a control treatment C. For setting 'B' as an extension to setting 'A', it was assumed that additionally two studies with 100% biomarker-positive patients existed for the comparison of S versus C. This setting corresponds to the application example.
In setting 'A', the performance of four approaches, namely the missing data, the network meta-regression, the enriching-through-weighting model with the proportion of BM_+-patients x, and the naive model were investigated.
The performance of the missing data approach is linked to the underlying true treatment effect. With other words, an assumption of the performance of BM_+-patients is needed. The performance of the 'best-case' and 'worst-case' approach are favorable in any of the scenarios. It can be seen that the 'worst-case only for S_+' approach highly overestimates the treatment effect of S for biomarker-positive patients in all scenarios. It seems to be a too crude approach to use and also suffers from the problem of rare events because the minimum of events are assigned to biomarker-positive patients. When it is assumed that the biomarker-positive patients benefit from the standard treatment compared to the control treatment, the 'best-case only for S_+' approach performs equally well as the enriching-through-weighting with x approach in terms of mean bias and RMSE and outperforms this approach in terms of coverage and precision. However, for a small number of studies, it has less power compared to the enriching-through-weighting approach. If there is the assumption that the biomarker is no effect modifier and therefore the treatment effect of biomarker-positive patients is equal to the treatment effect of the biomarker-negative patients in the standard treatment S, then either the 'biomarker not predictive' or the 'naive' approach can be used. The advantage of using the 'biomarker not predictive' approach is that it reflects the size of the patients in the biomarker-positive subgroup more adequatly. In general the missing data approach does not seem to outperform the other methods and, as a really crude approach, would only be recommended if there substantial reason to believe that the assumptions are fulfilled. The 'best-case for only S_+' approach could then be used as a sensitivity analysis.

Using the enriching-through-weighting approach as well as the network meta-regression approach lead to low mean bias and low RMSE. When higher between-study heterogeneity is present, the network meta-regression approach leads to lower precision compared to the analysis with the enriching-through-weighting approach. However the estimation with the network meta-regression approach results in a coverage close to 90% throughout all scenarios. The enriching-through-weighting approach suffers from low coverage in the setting which assumes that the biomarker-positive patients have a significant treatment effect advantage compared to the biomarker-negative patients undergoing non-targeted therapies. In terms of power, the enriching-through weighting approach has a higher power in the aforementioned scenario compared to the meta-regression approach.
The advantage of the network meta-regression and the 'enriching-through-weighting with x' approaches compared to the missing data approach is that no assumption of the treatment performance on biomarker-positive or biomarker-negative patients is needed. It can be seen that the network meta-regression with only a few studies shows less precision and, in the presence of heterogeneity, a slightly increased RMSE. According to the Cochrane Handbook 9.6.4 (Higgins & Green, 2011), a network meta-regression should not be considered when the number of studies included is smaller than ten. According to Borenstein et al. (2021), small studies, few studies, or a small covariate effect could also lead to problems with low power in network meta-regression. For only two studies in one arm and a small covariate effect, this can be seen in the simulation results of the network meta-regression displayed in the Appendix (Figure A.1). Another limitation of conducting a (network) meta-regression based on patient specific data on a study-level is the risk of ecological bias. Ecological bias means that causal inference about a covariate on patient-level data based on aggregated data of this covariate could lead to a biased effect estimation (Morgenstern, 1982; Thompson & Higgins, 2002). However, in the simulation study of this work, the mean bias of the network meta-regression model was relatively small throughout all scenarios. Because of this aforementioned limitation of the network meta-regression approach, the enriching-through-weighting approach with x seems to be favorable to use, especially for small study numbers. This approach performs similar to the naive approach in most scenarios but results in less bias when a treatment effect benefit for biomarker-positive patients compared to biomarker-negative patients in treatment S versus C is assumed or when high between-study heterogeneity is present. Therefore it seems to be a good choice if the effect of a covariate e.g. the biomarker status, is unknown.

The inclusion of studies with only biomarker-positive patients for the comparison of C versus S in the network as in setting 'B' offers more choices for network meta-analysis methods. In this situation, also a stand-alone network meta-analysis including only studies with biomarker-positive patients and an informative prior approach using the

treatment effect of the mixed patient population as prior are feasible. Another option is using the enriching-through weighting approach not only with the value 'proportion of BM_+ patients' as a weighting factor but with specific weights for the trust in the evidence from a mixed patient population compared to the 100% biomarker-positive patient population. These additional methods were compared in a simulation study to the network meta-regression approach, the naive approach, and the enriching-through-weighting approach with x. The missing data approach was not evaluated in setting 'B' because the assumption of knowing the performance of biomarker-positive patients seemed too strong and none of the approaches seemed to perform exceptionally well compared to the enriching-through-weighting and the network meta-regression approaches. The network meta-regression and stand-alone approaches lead throughout all scenarios to the lowest bias and high coverage. The enriching-through-weighting approaches result in a decreased RMSE and higher precision, especially when there is some between-study heterogeneity. In general, it can be noticed that adding two studies with 100%-biomarker positive patients leads to a decreased mean bias, decreased RMSE, more precision and better coverage in most scenarios for the methods used already in setting 'A'. The improved performance of the methods is not only caused by the fact that the number of studies included is always increased by two, but also by the fact that more weight is given to studies with 100-% biomarker-positive patients in the enriching-through-weighting approach with x and the network meta-regression approach. As discussed before, using the network meta-regression approach implies the potential risk of ecological bias, because the regression covariate biomarker status is based on patient-level data. If studies of 100% biomarker-positive patients are available, the informative approach could be advisable to use because it does not have the risk of ecological bias. Especially if there is some heterogeneity assumed to be present, it can lead to less variance and a more precise estimate without loss of coverage and is therefore favorable to the stand-alone approach. The enriching-through weighting approaches are recommended to be used for sensitivity analyses. These results might give hints on how the studies with overall populations differ from studies with only biomarker-positive patients.

In this thesis, the use of a half-normal prior for the between-study heterogeneity compared to a uniform prior was also investigated. The half-normal prior ($\mathcal{HN}(0.5)$) resulted in higher precision values compared to the uniform prior ($\mathcal{U}(0, 2)$) especially in the presence of low between study-heterogeneity. This was to be expected because the prior has more weight in low heterogeneity values (Röver et al., 2021). For intermediate or high heterogeneity values, the resulting power increased when using a half-normal prior.

Another parameter investigated in this thesis is the impact of the proportion range of BM_+ patients on the results. This range influences especially the results of the network meta-regression approach. The resulting RMSEs in the simulation study are always higher for smaller proportion ranges compared to larger proportion ranges in the net-

work meta-regression. This was to be expected because variance increases with a smaller range of independent values and is inversely related to the spread of covariate variables, in this case the proportion of BM_+ patients.
The least biased estimate in the application was generated using the stand-alone approach. However, this small bias comes at the cost of a high variability or wider credible intervals. The enriching-through-weighting approach with weight $w \sim \mathcal{U}(0, 0.3)$ and the network meta-regression approach both resulted in a tolerable bias and better precision. In general, it can be noticed that the treatment effect estimates of the approaches which give more weight to mixed studies shift towards a higher treatment effect for Erlotinib. This might be caused by the fact that the treatment effect for Gefitinib is decreased in a mixed patient populations compared to the Gefitinib studies where exclusively biomarker-positive patients were included (see Table 3.1). In this example, all approaches indicate a treatment effect advantage for Erlotinib compared to Gefitinib with regards to the response rate.

4.2 Limitations and directions for future research

The limitation of the missing data approach, with possibly too crude assumptions or the possibility of introducing ecological bias in the network meta-regression have already been discussed above.
The ranges of respective weights chosen for the enriching-through-weighting approach were based on the work of (Efthimiou et al., 2017). In further research, the choice of these weights could be changed and the trust in the evidence of mixed patient population (weights) could, for example, be based on some historical evidence. The performance of an analysis with stand-alone approach is quite good. This might be caused by the fact that the two single biomarker studies have been simulated from the same true mean treatment effect as the part of the population which is biomarker-positive. A variation on this part might also change the performance of this approach.
As in every work, the assumptions chosen for specific setting restrict the extent of scenarios to be evaluated. One limitation in this work is the very specific setting of a network meta-analysis with targeted and non-targeted therapies. In order to choose the evaluation scenarios, the situation for given parameters had to be restricted. But despite of this very specific scenario chosen in the simulation study, an indication can be given based on the simulation results on how the aforementioned methods would perform in a more general setting in a network with certain subpopulations and overall populations.
In this work, it was assumed that only aggregate data was available for the studies with an overall population. In Proctor et al. (2020), it was assumed that IPD was available for the targeted arm but, in this specific setting with binary endpoints, no advantage of the IPD could be seen compared to AD. This is why here all available study data

was assumed to be on only aggregated data due to reasons of simplicity. If individual patient data were available for some studies investigating an overall population, another choice of methods would be available. These could be compared to the ones discussed in this work in further research. In Ishak and Benedict (2015) the MAIC or STC approach were discussed which rely on AD and IPD availiability and seem to address a similar problem. In the situation of this thesis however, these approaches would not work, because it was considered that only studies with 100% biomarker-positive patients were available for one arm of the network and this TEE could not be adjusted for a proportion of biomarker patients in the other arm which is less than 100%. (Phillippo et al., 2020) suggested another method for analysing IPD and AD together and conducting a network meta-regression with the multilevel network meta-regression approach. This approach also requires to have IPD available in all arms of the network which is not the case in the setting of this work. These aforementioned approaches show as well that methods for network meta-analyses for heterogeneous populations are a current field of research.

As mentioned before, the endpoint regarded in this work was chosen to be binary and all methods were compared for binary data. In further research, the models could be extended for continuous and time-to-event data, and might also include multi-arm trials.

Another rather strong assumption in this work was the assumption of consistency of the network. Consistency follows according to Dias et al. (2018) from the exchangeability assumption. This assumption might be challenged depending on the specific setting as explained in the chapters before. Therefore the consistency assumption might not be always fulfilled. Further research is needed on how to address this issue adequately.

One other limiting factor is the assumption that the percentage of biomarker-positive patients is the same in both treatment groups which implies a perfectly balanced randomization. In real clinical trials, this might not always be given.

Some of the used methods rely on the fact that the percentage of biomarker-positive patients is known. In a real clinical trial, it could be the case that this percentage is unknown, e.g. when studies are conducted without considering the biomarker status in the analyses. On the other side however, if biomarker tests are conducted, then subgroup analyses are performed quite often (see e.g. Mok et al. (2009) and Han et al. (2012)). In this case, the subgroup specific treatment effect is known and the adjustment methods are not needed. However often only a small amount of patients are tested retrospectively and using only the subgroup results might lead to loss of information.

All approaches used were based on Bayesian methods and required the choice of priors for certain parameters. In the simulation study, most priors were chosen to be vague. But especially the choice of the heterogeneity prior needs to be discussed in more detail. According to Röver et al. (2021), it might be the case that, when only few studies are considered, the data have little information to add and therefore the analysis results

are largely determined by the prior settings, even if the prior is chosen to be vague. According to Lambert et al. (2005), precision can vary additionally with the choice of the prior distribution leading to varying coverage intervals and, potentially, to different statistical inferences. The width of credible intervals might be increased when using a flatter prior for between-trial heterogeneity (Seide et al., 2020), especially when only a few studies are included in the network meta-analysis. A between-study heterogeneity close to zero might lead to an upward biased variance estimate. This is why sensitivity analyses are recommended with different priors (Lambert et al., 2005). In this work, only the use of a half-normal prior $\mathcal{HN}(0.5)$ was evaluated as sensitivity analysis. In further research, the use of other vague, weakly informative, or even informative priors in justified settings could be investigated.
The values chosen for the theoretical treatment effects in the simulation studies were based on specific response rates of the application example. In other settings, it might be interesting to evaluate how other theoretical treatment effects would influence the performance of the methods. One might e.g. decide for a proportional distance between the chosen treatment effects.
The results in this work could only be calculated using simulation studies due to the complexity of the models. Given the limited computing capacity, only a small number of repetitions (1000) could be performed and thus there may be random variations in the results.

4.3 Conclusion

In this thesis, different methods were evaluated to conduct a network meta-analysis including evidence for targeted and non-targeted therapies. In the simulation study, it was shown that in the case of only non-targeted studies in one arm of the network, the enriching-through-weighting approach with x (setting A) leads in most settings to less bias and an acceptable variance compared to the other methods. The network meta-regression could be used for sensitivity analysis. If there are targeted studies for all arms of the network available (setting B), the informative prior approach might be advisable. Sensitivity analyses with the enriching-through-weighting approaches with different weights and the network meta-regression approach are recommended. These aforementioned methods give the possibility to obtain a treatment effect estimate for a specific population in the network using all evidence available and therefore help to come to a treatment recommendation for this specific subpopulation.

Chapter 5

Summary

5.1 Summary (English)

The aim of this thesis is to evaluate network meta-analysis methods including studies with an overall patient population and as well as studies on only subpopulations regarding a specific covariate.
In this thesis the situation is motivated by an actual network meta-analysis problem. Here a biomarker is identified as separating a patient population into patients with a higher and a lesser expected benefit from evaluated treatments. Consequently, later studies might be conducted in a targeted way, while earlier ones are performed in mixed patient populations. Additionally, it is assumed that there are only targeted studies of a newer treatment available. In order to compare these different treatments with each other and use all evidence available, a network meta-analysis needs to be conducted. Because the patient population is heterogeneous with regard to this biomarker, conducting a network meta-analysis of all available treatments (targeted and non-targeted) poses a challenge in evidence synthesis, especially if only aggregated study data are available. Currently existing methods are either assuming that individual patient data is available or do not include all evidence available. Therefore in this work, methods are discussed and evaluated regarding the following setting:
The network meta-analysis setting is a "triangular" network involving (little) direct and (more) indirect evidence on a particular comparison. Two settings (A and B), which differ regarding the proportion of biomarker-positive patients in the single studies, are evaluated. The first setting (A) assumes that in the non-targeted therapy arm only studies with mixed patient population regarding a specific biomarker are available. Setting B is an extension of setting A which assumes that additionally two studies with only biomarker-positive patients populations in the non-targeted study arm are available.
Based on these two scenarios, three commonly used network meta-analytic estimation methods, the naive estimation approach where the heterogeneity in the patient population is ignored, the stand-alone analysis which includes studies with only biomarker-positive patients, and the network meta-regression are investigated for the analysis of this "triangular network". For setting A, a missing data approach is introduced as a further solution. Additionally, the enriching through weighting approach, a method developed in evidence synthesis for combining randomized and non-randomized data, is modified and adapted for the setting of this triangular network. For setting B, further modification of the enriching-through-weighting approach as well as an informative prior approach, where the results of the mixed patient population are used as a prior, are investigated additionally to the naive, stand-alone, and network meta-regression approach.
The performance of these different methods is evaluated in a simulation study in calculating the mean bias, Root-Mean-Squared Error, precision, coverage and the power with regards to the estimated treatment effect comparing the targeted and non-targeted

therapies. Additionally, an actual clinical data set including targeted and non-targeted therapies is analysed and discussed.
The results of the simulation study for setting A as well as for setting B show that none of the methods are observed to be clearly favorable over all investigated scenarios. However, the missing data approach, the stand-alone analysis, and the naive estimation perform comparably or worse than the other methods in all evaluated performance measures and simulation scenarios and are therefore not recommended. While substantial between-trial heterogeneity is challenging for all estimation approaches, the performance of the network meta-regression, the enriching-through-weighting approach, and the informative prior approach are dependent on the simulation scenario and the performance measure of interest. Furthermore, as these estimation methods are based on slightly different assumptions, some of which require the presence of additional information for estimation, sensitivity-analyses are recommended wherever possible.
For setting A, the enriching-through weighting approach is recommended as the most favorable method, while the network meta-regression can additionally be recommended as a sensitivity analysis. If biomarker-positive studies are available for the non-targeted study arm (setting B) the informative prior or the enriching-through-weighting approach are recommended. The network meta-regression approach is additionally recommended as a sensitivity analysis for setting B.
All approaches are based on Bayesian Models. In this work one alternative heterogeneity prior, a half-normal prior with scale 0.5, was additionally analysed in setting B alongside the standard uniform heterogeneity prior. Using this more informative prior does not change the performance of the single approaches in terms of bias or root-mean-squared error but leads for most approaches to lower coverage and higher precision and for some scenarios to an even higher power. Therefore a sensitivity analysis with this heterogeneity prior is recommended, when conducting a Bayesian network meta-analysis.
This work provides an overview of methods to conduct a network meta-analysis with different patient populations and makes a contribution to the field of population-adjusted network meta-analyses. Although no method clearly performed best in all investigated scenarios, the enriching-through-weighting model, the network meta-regression model and informative prior model are recommendations to conduct a network meta-analysis and use all evidence available.

5.2 Zusammenfassung (Deutsch)

Das Ziel dieser Arbeit ist es, Methoden der Netzwerk-Metaanalyse zu evaluieren, die sowohl Studien mit einer allgemeinen Patientenpopulation sowie Studien, die nur eine Subpopulation, die anhand eines Biomarker identifiziert wird, einschließen.
In dieser Arbeit wird die betrachtete Situation durch ein reales Netzwerk-Metaanalyse-Problem motiviert. Bei diesem Beispiel wird ein Biomarker identifiziert, der eine Pa-

tientenpopulation in Patienten mit größerem und solche mit geringerem Nutzen von speziellen Behandlungen unterteilt. In späteren Studien wurden daher nur Teilpopulationen eingeschlossen, die von einer solchen zielgerichteten Behandlung profitieren, während in früheren Studien gemischte Patientenpopulationen, ungeachtet des Biomarkerstatus eingeschlossen wurden. Zudem wird davon ausgegangen, dass nur zielgerichtete Studien zu einer weiteren neueren Behandlung vorliegen. Um diese verschiedenen Therapien miteinander zu vergleichen und alle verfügbaren Studien miteinzuschließen, sollte eine Netzwerk-Metaanalyse durchgeführt werden. Da die Patientenpopulationen der Studien in Bezug auf diesen Biomarker heterogen sind, stellt die Durchführung einer Netzwerk-Metaanalyse aller verfügbaren Studien und Therapien (zielgerichtet und nicht zielgerichtet) eine Herausforderung bei der Evidenzsynthese dar, insbesondere wenn nur aggregierte Studiendaten verfügbar sind. Derzeit existierende Methoden gehen entweder davon aus, dass individuelle Patientendaten der Studien verfügbar sind, oder beinhalten nicht alle verfügbaren Studien. Daher werden in dieser Arbeit Methoden diskutiert, um solch eine Netzwerk-Metaanalyse durchzuführen und diese im Hinblick auf das im Folgenden dargestellte Setting bewertet. Das Netzwerk-Metaanalyse-Setting ist ein "Dreiecks-" Netzwerk, das (wenige) direkte und (viele) indirekte Beweise für einen bestimmten Vergleich enthält. Zwei Settings (A und B) werden untersucht, die sich hinsichtlich des Anteils an biomarker-positiven Patienten in den einzelnen Studien unterscheiden. Das erste Setting (A) geht davon aus, dass im nicht zielgerichteten Therapiearm nur Studien mit einem gemischtem Patientenkollektiv im Hinblick zu einem bestimmten Biomarker vorliegen. Für Setting B wird Setting A erweitert, indem angenommen wird, dass zusätzlich zwei Studien mit ausschließlich biomarker-positiven Patientenpopulationen im nicht zielgerichteten Studienarm verfügbar sind.
Basierend auf diesen beiden Szenarien werden drei häufig verwendete netzwerkmetaanalytische Methoden, der naive Ansatz, bei dem die Heterogenität in der Patientenpopulation ignoriert wird, die "Stand-alone-" Analyse, die Studien mit nur biomarker-positiven Patienten umfasst, und die Netzwerk Meta-regression zur Analyse dieses "Dreiecksnetzwerks" untersucht. Für Setting A wird auch als mögliche Lösung eine Art "Missing-data"-Ansatz eingeführt. Zusätzlich wird der "Enriching-through weighting" Ansatz, eine in der Evidenzsynthese entwickelte Methode zur Kombination randomisierter und nicht-randomisierter Daten, modifiziert und für das Setting dieses Dreiecksnetzwerks angepasst. Für Setting B werden zusätzlich zu den naiven, Stand-alone- und Netzwerk Meta-Regression-Ansätzen weitere Modifikationen des Enriching-through-weighting-Ansatzes, sowie ein "Informative-Prior"-Ansatz untersucht, bei dem die Ergebnisse der gemischten Patientenpopulation als Priorinformation verwendet werden.
Die Performance dieser verschiedenen Methoden wird in einer Simulationsstudie bei der Berechnung des mittleren Bias, des Root-Mean-Squared Error, der Präzision, der Abdeckungswahrscheinlichkeit und der Aussagekraft in Bezug auf den geschätzten Be-

handlungseffekt im Vergleich zwischen zielgerichteten und nicht zielgerichteten Therapien bewertet. Zusätzlich wird ein aktueller klinischer Datensatz mit zielgerichteten und nicht zielgerichteten Therapien analysiert und diskutiert.

Die Ergebnisse der Simulationsstudie sowohl für Setting A als auch für Setting B zeigen, dass keine der Methoden über alle untersuchten Szenarien als eindeutig günstig zu beobachten ist. Der Missing-data-Ansatz, die Stand-alone-Analyse und die naive Schätzung schneiden jedoch in allen bewerteten Leistungskennzahlen und Simulationsszenarien vergleichbar oder schlechter ab als die anderen Methoden und werden daher nicht empfohlen. Während eine erhebliche Heterogenität zwischen den Studien für alle Ansätze eine Herausforderung darstellt, hängen die Leistung der Netzwerk Meta-Regression, der Enriching-through-weighting-Ansatz und der Informative-Prior-Ansatz vor allem Simulationsszenario und des einzelnen Bewertungsmaßes ab. Da diesen Schätzmethoden außerdem leicht unterschiedliche Annahmen zugrunde legen, von denen einige das Vorhandensein zusätzlicher Informationen für die Schätzung erfordern, werden Sensitivitätsanalysen empfohlen, wo immer dies möglich ist.

Für Setting A empfiehlt sich der Enriching-through-weighting-Ansatz als günstigste Lösung und zusätzlich empfiehlt sich der Netzwerk Meta-Regressions-Ansatz als Sensitivitätsanalyse. Liegen biomarker-positive Studien für den nicht zielgerichteten Studienarm (Setting B) vor, sind der Informative-prior- oder der Enriching-through-weighting-Ansatz im Vorteil. Als Sensitivitätsanalyse wird zusätzlich der Ansatz der Netzwerk-Meta-Regression empfohlen.

Alle Ansätze basieren auf Bayesianischen Modellen. In dieser Arbeit wurde ein alternativer Heterogenitätsprior, ein halbnormaler Prior mit Skala 0.5, für das Setting B zusätzlich zum Heterogenitätsprior mit Gleichverteilung analysiert. Die Verwendung dieses informativeren Priors ändert nichts am Abschneiden der einzelnen Ansätze in Bezug auf Bias oder Root-Mean-Square Error, führt jedoch für die meisten Ansätze zu einer geringeren Abdeckungswahrscheinlichkeit und höherer Präzision und für einige Szenarien zu einer noch höheren Power. Daher wird eine Sensitivitätsanalyse mit diesem Heterogenitätsprior empfohlen, wenn eine Bayesianische Netzwerk-Metaanalyse durchgeführt wird.

Diese Arbeit gibt einen Überblick über Methoden zur Durchführung einer Netzwerk-Metaanalyse mit unterschiedlichen Patientenpopulationen und leistet einen Beitrag auf dem Gebiet der populationsadjustierten Netzwerk-Metaanalysen. Obwohl keine Methode in allen untersuchten Szenarien eindeutig die beste Leistung erbrachte, sind der Enriching-through-weighting-Ansatz, das Netzwerk-Meta-Regressionsmodell und das Informative-Prior-Modell zu empfehlen, um eine Netzwerk-Metaanalyse durchzuführen und alle verfügbare Evidenz zu nutzen.

Bibliography

Borenstein, M., Hedges, L., Higgings, J. H., and Rothstein, H. (2021). *Introduction to meta-analysis 2nd edition*. John Wiley; Sons, Ltd.

Bucher, H. C., Guyatt, G. H., Griffith, L. E., and Walter, S. D. (1997). The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology*, *50*, 683–691.

Buyse, M., Michiels, S., Sargent, D., Grothey, A., Matheson, A., and de Gramont, A. (2011). Integrating biomarkers in clinical trials. *Expert Review of Molecular Diagnostics*, *11*, 171–182.

Caldwell, D., Ades, A., and Higgins, J. (2005). Simultaneous comparison of multiple treatments: Combining direct and indirect evidence. *British Medical Journal*, *331*(7521), 897–900.

Caro, J., and Ishak, K. (2010). No head-to-head trial? simulate the missing arms. *Pharmacoeconomics*, *28*(10), 957–967.

Chandler, J., Cumpston, M., Thomas, J., Higgins, J., Deeks, J., and Clarke, M. (2021). Chapter I: Introduction. In J. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. Page, & V. Welch (Eds.), *Cochrane Handbook of Systematic Reviews of Interventions version (6.2), updated February 2021*. Cochrane.

Chapman, A., Sun, K., Ruestow, P., Cowan, D., and Madl, A. (2016). Lung cancer mutation profile of EGFR, ALK, and KRAS: Meta-analysis and comparison of never and ever smokers. *Lung cancer (Amsterdam, Netherlands)*, *102*, 122–134.

Cheng, L., Alexander, R., MacLennan, G., Cummings, O., Montironi, R., Lopez-Beltran, A., Cramer, H., Davidson, D., and Zhang, S. (2012). Molecular pathology of lung cancer: Key to personalized medicine. *Modern Pathology*, *25*, 347–369.

Dearden, S., Stevens, J., Wu, Y., and Blowers, D. (2013). Mutation incidence and coincidence in non small-cell lung cancer: Meta-analyses by ethnicity and histology (mutMap). *Annals of Oncology*, *24*(9), 2371–2376.

DerSimonian, R., and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188.

Dias, S., Ades, A., Welton, N., Jansen, J., and Sutton, A. (2018). *Network Meta-Analysis for Decision-Making* (Vol. 1). Wiley.

Dias, S., Sutton, A., Welton, N., and Ades, A. (2011). *NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials*. National Institute for Health; Clinical Excellence.

Didkowska, J., Wojciechowska, U., Marta Mańczuk, M., and Łobaszewski, J. (2016). Lung cancer epidemiology: Contemporary and future challenges worldwide. *Annals of Translational Medicine*, *4*(8). https://atm.amegroups.com/article/view/9532

Efthimiou, O., Mavridis, D., Debray, T., Samara, M., Belger, M., Siontis, G., Leucht, S., and Salanti, G. (2017). Combining randomized and non-randomized evidence in network meta-analysis. *Statistics in Medicine*, *36*(8), 1210–1226.

FDA-NIH Biomarker Working Group. (2016). BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Silver Spring (MD): Food and Drug Administration (US); Available from: https://www.ncbi.nlm.nih.gov/books/NBK326791/ Co-published by National Institutes of Health (US), Bethesda (MD) [Accessed at 23.08.2022].

Freidlin, B., McShane, L., and Korn, E. (2010). Randomized Clinical Trials With Biomarkers: Design Issues. *JNCI: Journal of the National Cancer Institute*, *102*(3), 152–160.

Friede, T., Röver, C., Wandel, S., and Neuenschwander, B. (2017). Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases. *Biometrical Journal*, *59*(4), 658–671.

Fukuoka, M., Yano, S., Giaccone, G., Tamura, T., Nakagawa, K., Douillard, J.-Y., Nishiwaki, Y., Vansteenkiste, J., Kudoh, S., Rischin, D., Eek, R., Horai, T., Noda, K., Takata, I., Smit, E., Averbuch, S., Macleod, A., Feyereislova, A., Dong, R.-P., and Baselga, J. (2003). Multi-Institutional Randomized Phase II Trial of Gefitinib for Previously Treated Patients With Advanced Non–Small-Cell Lung Cancer. *Journal of Clinical Oncology*, *21*(12), 2237–2246.

Giaccone, G., Ruiz, M., Le Chevalier, T., Thatcher, N., Smit, E., Rodriguez, J., Janne, P., Oulid-Aissa, D., and Soria, J.-C. (2006). Erlotinib for Frontline Treatment of Advanced Non–Small Cell Lung Cancer: a Phase II Study. *Clinical Cancer Research*, *12*(20), 6049–6055.

Greenhalgh, J., Boland, A., Bates, V., Vecchio, F., Dundar, Y., Chaplin, M., and Green, J. (2021). First-line treatment of advanced epidermal growth factor receptor (EGFR) mutation positive non-squamous non-small cell lung cancer. *Cochrane Database of Systematic Reviews*, *3*(3).

Han, J.-H., Park, K., Kim, S.-W., Lee, D., Kim, H., Kim, H., Ahn, M., Yun, T., Ahn, J., Suh, C., Lee, J.-S., Yoon, S., Han, J., Lee, J., Jo, S., and Lee, J. (2012). First-SIGNAL: First-Line Single-Agent Iressa Versus Gemcitabine and Cisplatin Trial in Never-Smokers With Adenocarcinoma of the Lung. *Journal of Clinical Oncology*, *30*, 1122–1128.

Harrer, M., Cuijpers, P., T.A., F., and Ebert, D. (2021). *Doing Meta-Analysis With R: A Hands-On Guide* (1st). Chapman; Hall/CRC Press.

Higgings, J., and Whitehead, A. (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine*, *15*(24), 2733–2749.

Higgins, J., and Green, S. (2011). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available at www.handbook.cochrane.org. www.handbook.cochrane.org

Hoaglin, D., Hawkins, N., Jansen, J., Scott, A., Itzler, R., Cappelleri, J., Boersma, C., Thompson, D., Larholt, K., Diaz, M., and Barrett, A. (2011). Conducting indirect-treatment-comparison and network-meta-analysis studies: Report of the ISPOR task force on indirect treatment comparisons good research practices – Part II. *Value in Health*, *14*, 429–437.

Inamura, K. (2017). Lung Cancer: Understanding Its Molecular Pathology and the 2015 WHO Classification. *Frontiers in oncology*, *7*, 193.

Ishak, I., J.K.and Proskorovsky, and Benedict, A. (2015). Simulation and Matching-Based Approaches for Indirect Comparison of Treatments. *PharmacoEconomics*, *33*, 537–549.

Jackman, D., Yeap, B., Lindeman, N., Fidias, P., Rabin, M., Temel, J., Skarin, A., Meyerson, M., Holmes, A., Borras, A., Freidlin, B., Ostler, P., Lucca, J., Lynch, T., Johnson, B., and Jänne, P. (2007). Phase II Clinical Trial of Chemotherapy-Naive Patients $\geq$ 70 Years of Age Treated With Erlotinib for Advanced Non–Small-Cell Lung Cancer. *Journal of Clinical Oncology*, *25*(7), 760–766.

Kim, E., Hirsh, V., T.Mok, Socinski, M., R.Gervais, Wu, Y.-L., Li, L.-Y., Watkins, C., Sellers, M., Lowe, E., Sun, Y., Liao, M.-L., Østerlind, K., M.Reck, Armour, A., Shepherd, F., Lippman, S., and Douillard, J.-Y. (2008). Gefitinib versus docetaxel in previously treated non-small-cell lung cancer (INTEREST): a randomised phase III trial. *The Lancet*, *372*(9652), 1809–1818.

Lambert, P., Sutton, A., Burton, P., Abrams, K., and Jones, D. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine*, *24*(15), 2401–2428.

Liang, W., Wu, X., Fang, W., Zhao, Y., Yang, Y., Hu, Z., Xue, C., Zhang, J., J.AND Zhang, Ma, Y., Zhou, T., Yan, Y., Hou, X., Qin, T., Dinglin, X., Tian, Y., Huang, P., Huang, Y., Zhao, H., and Zhang, L. (2014). Network Meta-Analysis of Erlotinib, Gefitinib, Afatinib and Icotinib in Patients with Advanced Non-Small-Cell Lung Cancer Harboring EGFR Mutations. *PLOS ONE*, *9*(2), 1–11.

Lin, J.-Z., Ma, S.-K., Wu, S.-X., Yu, S.-H., and Li, X.-Y. (2018). A network meta-analysis of nonsmall-cell lung cancer patients with an activating EGFR mutation: Should osimertinib be the first-line treatment? *Medicine (Baltimore)*, *97*, e11569.

Lu, G., and Ades, A. (2006). Assessing Evidence Inconsistency in Mixed Treatment Comparisons. *Journal of the American Statistical Association*, *101*(474), 447–459.

Lu, G., and Ades, A. (2009). Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics*, *10*(4), 792–805.

Lu, G., Welton, N., Higgins, J., White, I., and Ades, A. (2011). Linear inference for mixed treatment comparison meta-analysis: A two-stage approach. *Research Synthesis Methods*, *2*(1), 43–60.

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine*, *21*(16), 2313–2324.

Lunn, D., Barrett, J., Sweeting, M., and Thompson, S. (2013). Fully bayesian hierarchical modelling in two stages, with application to meta-analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*(4), 551–572.

Maemondo, M., Inoue, A., Kobayashi, K., Sugawara, S., Oizumi, S., Isobe, H., Gemma, A., Harada, M., Yoshizawa, H., Kinoshita, I., Fujita, Y., Okinaga, S., Hirano, H., Yoshimori, K., Harada, T., Ogura, T., Ando, M., Miyazawa, H., Tanaka, T., Saijo, Y., Hagiwara, K., Morita, S., and Nukiwa, T. (2010). Gefitinib or Chemotherapy for Non–Small-Cell Lung Cancer with Mutated EGFR. *New England Journal of Medicine*, *362*, 2380–2388.

Maitournam, A., and Simon, R. (2005). On the efficiency of targeted clinical trials. *Statistics in Medicine*, *24*(3), 329–339.

Maruyama, R., Nishiwaki, Y., Tamura, T., Yamamoto, N., Tsuboi, M., Nakagawa, K., Shinkai, T., Negoro, S., Imamura, F., Eguchi, K., Takeda, K., Inoue, A., Tomii, K., Harada, M., Masuda, N., Jiang, H., Itoh, Y., Ichinose, Y., Saijo, N., and Fukuoka, M. (2008). Phase III Study, V-15-32, of Gefitinib Versus Docetaxel in Previously Treated Japanese Patients With Non–Small-Cell Lung Cancer. *Journal of Clinical Oncology*, *26*(26), 4244–4252.

Mitsudomi, T., Morita, S., Yatabe, Y., Negoro, S., Okamoto, I., Tsurutani, J., Seto, T., Satouchi, M., Tada, H., Hirashima, T., K.Asami, Katakami, N., Takada, M., Yoshioka, H., Shibata, K., Kudoh, S., Shimizu, E., Saito, H., Toyooka, S., Nakagawa, K., and Fukuoka, M. (2010). Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. *The Lancet Oncology*, *11*, 121–128.

Mok, T., Wu, Y.-L., Thongprasert, S., Yang, C.-H., Chu, D.-T., Saijo, N., Sunpaweravong, P., Han, B., Margono, B., Ichinose, Y., Nishiwaki, Y., Ohe, Y., Yang, J.-J., Chewaskulyong, B., Jiang, H., Duffield, E., Watkins, C., Armour, A., and

Fukuoka, M. (2009). Gefitinib or Carboplatin-Paclitaxel in Pulmonary Adenocarcinoma. *New England Journal of Medicine*, *361*, 947–957.

Morgenstern, H. (1982). Uses of ecologic analysis in epidemiologic research. *American Journal of Public Health*, *72*(12), 1336–1344.

Petto, H., Kadziola, Z., Brnabic, A., D.Saure, and Belger, M. (2019). Alternative Weighting Approaches for Anchored Matching-Adjusted Indirect Comparisons via a Common Comparator. *Value in Health*, *22*(1), 85–91.

Phillippo, D., Dias, S., Ades, A., and Welton, N. (2020). Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study. *Statistics in Medicine*, *39*(30), 4885–4911.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.

Proctor, T., Jensen, K., and Kieser, M. (2020). Integrated evaluation of targeted and non-targeted therapies in a network meta-analysis. *Biometrical Journal*, *62*(3), 777–789.

Proctor, T., Zimmermann, S., Seide, S., and Kieser, M. (2022). A comparison of methods for enriching network meta-analyses in the absence of individual patient data. *Research Synthesis Methods*. https://doi.org/doi.org/10.1002/jrsm.1568

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Remiro-Azócar, A., A.Heath, and Baio, G. (2021). Methods for Population Adjustment with Limited Access to Individual Patient Data: A Review and Simulation Study. *Research Synthesis Methods*, *12*(6), 750–775.

Renfro, L., Mallick, H., An, M., Sargent, D., and Mandrekar, S. (2016). Clinical trial designs incorporating predictive biomarkers. *Cancer treatment reviews*, *43*, 74–82.

Rosell, R., Carcereny, E., Gervais, R., Vergnenegre, A., Massuti, B., Felip, E., Palmero, R., R.Garcia-Gomez, Pallares, C., Sanchez, J., Porta, R., Cobo, M., Garrido, P.,

Longo, F., Moran, T., Insa, A., Marinis, F. D., Corre, R., Bover, I., Illiano, A., Dansin, E., de Castro, J., Milella, M., Reguart, N., Altavilla, G., Jimenez, U., Provencio, M., Moreno, M., Terrasa, J., Muñoz-Langa, J., Valdivia, J., Isla, D., Domine, M., Molinier, O., Mazieres, J., Baize, N., Garcia-Campelo, R., Robinet, G., Rodriguez-Abreu, D., Lopez-Vivanco, G., Gebbia, V., Ferrera-Delgado, L., Bombaron, P., Bernabe, R., Bearz, A., Artal, A., Cortesi, E., Rolfo, C., Sanchez-Ronco, M., Drozdowskyj, A., Queralt, C., de Aguirre, I., Ramirez, J., Sanchez, J., Molina, M., Taron, M., and Paz-Ares, L. (2012). Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced {EGFR} mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *The Lancet Oncology*, *13*, 239–246.

Röver, C., Bender, R., Dias, S., Schmid, C., Schmidli, H., Sturtz, S., Weber, S., and Friede, T. (2021). On weakly informative prior distributions for the heterogeneity parameter in bayesian random-effects meta-analysis. *Research Synthesis Methods*, *12*(4), 448–474.

Rücker, G., Petropoulou, M., and Schwarzer, G. (2020). Network meta-analysis of multicomponent interventions. *Biometrical Journal*, *62*(3), 808–821.

Salanti, G. (2012). Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: Many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods*, *3*.

Saramago, P., Sutton, A. J., Cooper, N. J., and Manca, A. (2012). Mixed treatment comparisons using aggregate and individual participant level data. *Statistics in Medicine*, *31*, 3516–3536.

Seide, S., Jensen, K., and Kieser, M. (2019). Simulation and data-generation for random-effects network meta-analysis of binary outcome. *Statistics in Medicine*, *38*(17), 3288–3303.

Seide, S., Jensen, K., and Kieser, M. (2020). A comparison of Bayesian and frequentist methods in random-effects network meta-analysis of binary data. *Research Synthesis Methods*, *11*(3), 363–378.

Signorovitch, J. E., Sikirica, V., Erder, M. H., Xie, J., Lu, M., Hodgkins, P. S., Betts, K. A., and Wu, E. Q. (2012). Matching-adjusted indirect comparisons: A new

tool for timely comparative effectiveness research. *Value In Health*, *15*(6), 940–947.

Simon, R., and Maitournam, A. (2004). Evaluating the Efficiency of Targeted Designs for Randomized Clinical Trials. *Clinical Cancer Research*, *10*(20), 6759–6763.

Smith, T., Spiegelhalter, D., and Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine*, *14*(24), 2685–2699.

Spiegelhalter, D., Abrams, K., and Myles, J. (2004). *Bayesian approaches to clinical trials and health-care evaluation*. Wiley.

Su, Y.-S., and Yajima, M. (2015). *R2jags: Using R to Run 'JAGS'* [R package version 0.5-7]. https://CRAN.R-project.org/package=R2jags

Thompson, S., and Higgins, J. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, *21*(11), 1559–1573.

Tonin, F., Rotta, I., Mendes, A., and Pontarolo, R. (2017). Network meta-analysis: A technique to gather evidence from direct and indirect comparisons. *Pharm Pract (Granada)*.

Urata, Y., Katakami, N., Morita, S., Kaji, R., Yoshioka, H., Seto, T., Satouchi, M., Iwamoto, Y., Kanehara, M., Fujimoto, D., Ikeda, N., Murakami, H., Daga, H., Oguri, T., Goto, I., Imamura, F., Sugawara, S., Saka, H., Nogami, N., Negoro, S., Nakagawa, K., and Nakanishi, Y. (2016). Randomized Phase III Study Comparing Gefitinib With Erlotinib in Patients With Previously Treated Advanced Lung Adenocarcinoma: WJOG 5108L. *Journal of Clinical Oncology*, *34*(27), 3248–3257.

Vansteenkiste, J. (2004). Gefitinib (Iressa®): a novel treatment for non-small cell lung cancer. *Expert Review of Anticancer Therapy*, *4*(1), 5–17.

Vansteenkiste, J. (2018). Tyrosine kinase inhibition of EGFR: a successful history of targeted therapy for NSCLC since 20 years. *Annals of Oncology*, *29*, i1–i2.

Weber, D., Jensen, K., and Kieser, M. (2020). Comparison of Methods for Estimating Therapy Effects by Indirect Comparisons: A Simulation Study. *Medical Decision Making*, *40*(5), 644–654.

Wu, Y.-L., Zhou, C., Liam, C.-K., Wu, G., Liu, X., Zhong, Z., Lu, S., Cheng, Y., Han, B., Chen, L., Huang, C., Qin, S., Zhu, Y., Pan, H., Liang, H., Li, E., Jiang, G., How, S. H., Fernando, M. C. L., Zhang, Y., Xia, F., and Zuo, Y. (2015). First-line erlotinib versus gemcitabine/cisplatin in patients with advanced EGFR mutation-positive non-small-cell lung cancer: analyses from the phase III, randomized, open-label, ENSURE study. *Annals of Oncology*, *26*, 1883.

Xie, Y., Liang, J., and Su, N. (2015). Gefitinib versus Erlotinib as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer. *Nan Fang Yi Ke Da Xue Xue Bao*, *35*, 446–449.

Yang, J.-H., Kang, J., Mok, T., Ahn, M.-J., Srimuninnimit, V., Lin, C.-C., Kim, D.-W., Tsai, C.-M., Barraclough, H., Altug, S., M.Orlando, and Park, K. (2014). First-line pemetrexed plus cisplatin followed by gefitinib maintenance therapy versus gefitinib monotherapy in east asian patients with locally advanced or metastatic non-squamous non-small cell lung cancer: A randomised, phase 3 trial. *European Journal of Cancer*, *50*(13), 2219–2230.

Yang, J.-J., Zhou, Q., Yan, H.-H., Zhang, X.-C., Chen, H.-J., Tu, H.-Y., Wang, Z., Xu, C.-R., Su, J., Wang, B.-C., Jiang, B.-Y., Bai, X.-Y., Zhong, W.-Z., Yang, X.-N., and Wu, Y.-L. (2017). A phase III randomised controlled trial of erlotinib vs gefitinib in advanced non-small cell lung cancer with EGFR mutations. *British journal of cancer*, *116*, 568–574.

Yin, G., Yang, Z., Odani, M., and Fukimbara, S. (2021). Bayesian Hierarchical Modeling and Biomarker Cutoff Identification in Basket Trials. *Statistics in Biopharmaceutical Research*, *13*(2), 248–258.

Zhou, C., Wu, Y.-L., Chen, G., Feng, J., Liu, X.-Q., Wang, C., Zhang, S., Wang, J., Zhou, S., Ren, S., Lu, S., Zhang, L., Hu, C., Hu, C., Luo, Y., Chen, L., Ye, M., Huang, J., Zhi, X., Zhang, Y., Xiu, Q., Ma, J., Zhang, L., and You, C. (2011). Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncology*, *12*, 735–742.

Zhou, Q., Cheng, Y., Yang, J.-J., Zhao, M.-F., Zhang, L., Zhang, X.-C., Chen, Z.-H., Yan, H.-H., Song, Y., Chen, H.-J., Feng, W.-N., Xu, C.-R., Wang, Z., Chen, J.-H., Zhong, W.-Z., Liu, Y.-P., and Wu, Y.-L. (2014). Pemetrexed versus gefitinib as a second-line treatment in advanced nonsquamous nonsmall-cell lung cancer patients harboring wild-type EGFR (CTONG0806): a multicenter randomized trial. *Annals of Oncology*, *25*(12), 2385–2391.

Appendices

Appendix A

Additional simulation results

A.1 Results setting A

The results for the simulation study for setting A and covariate $x_j \in [0.4, 0.5]$ are displayed in the following.

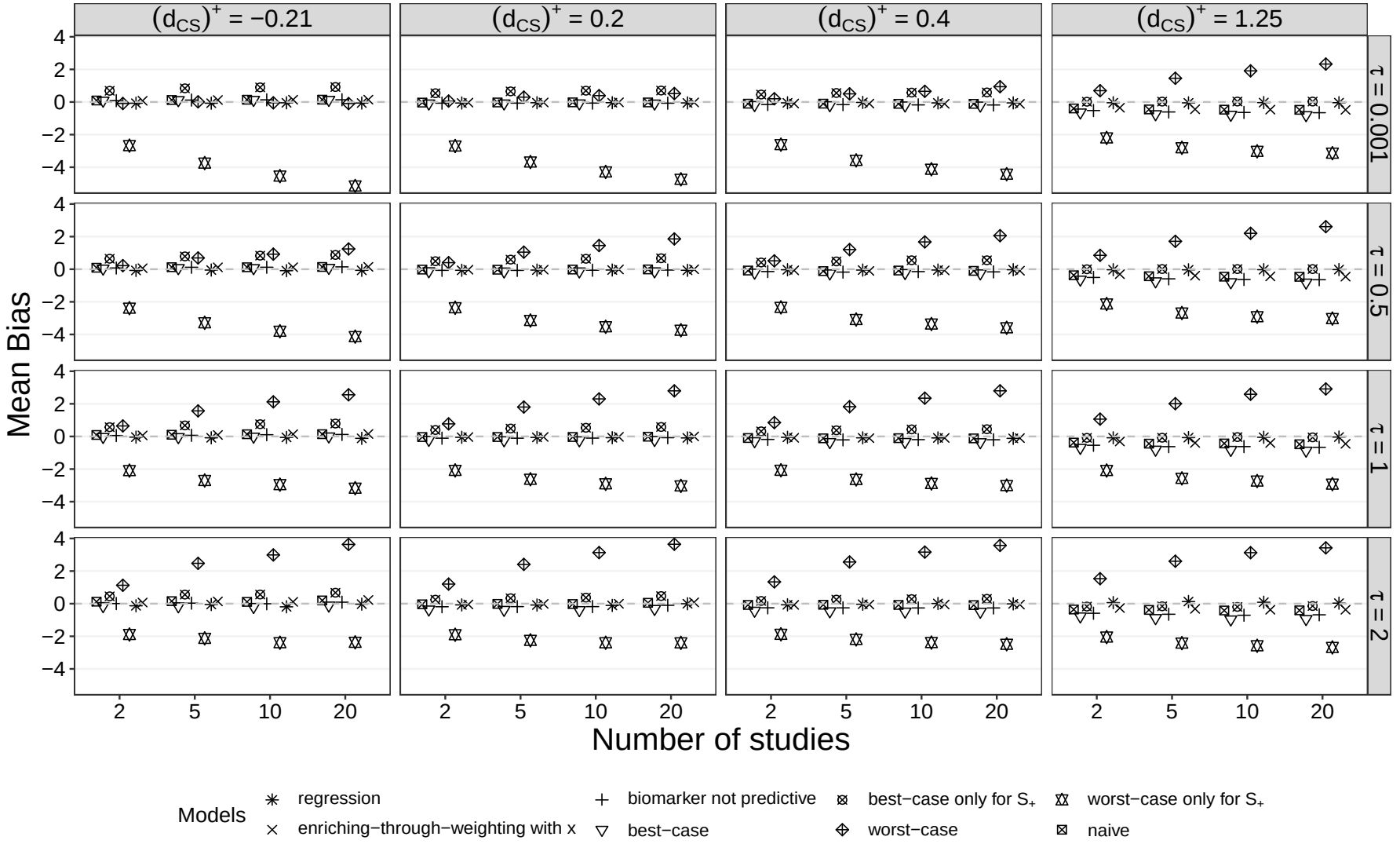


Figure A.1: *Mean bias of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A, $x_j \in [0.4, 0.5]$)*

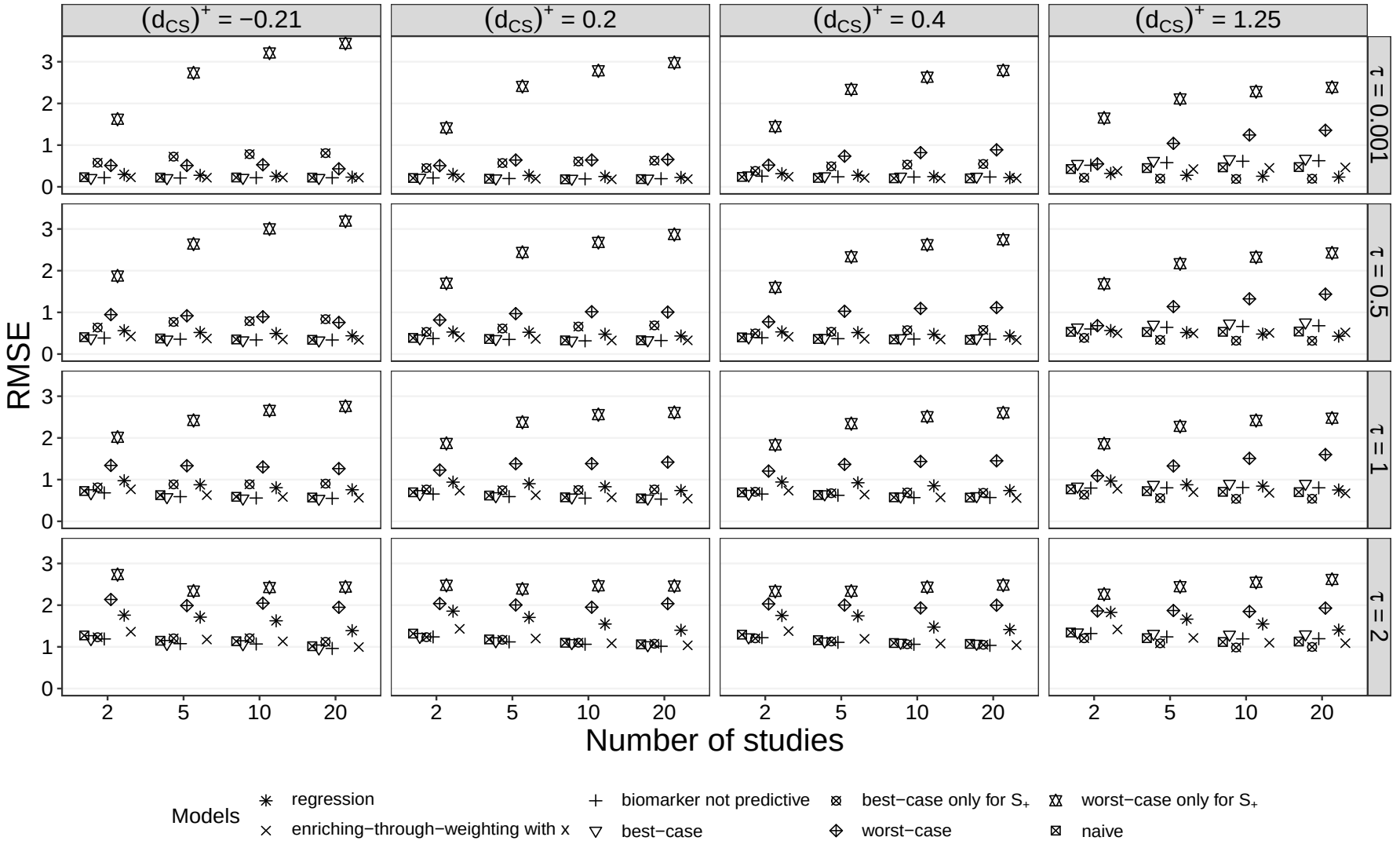


Figure A.2: *RMSE of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A, $x_j \in [0.4, 0.5]$)*

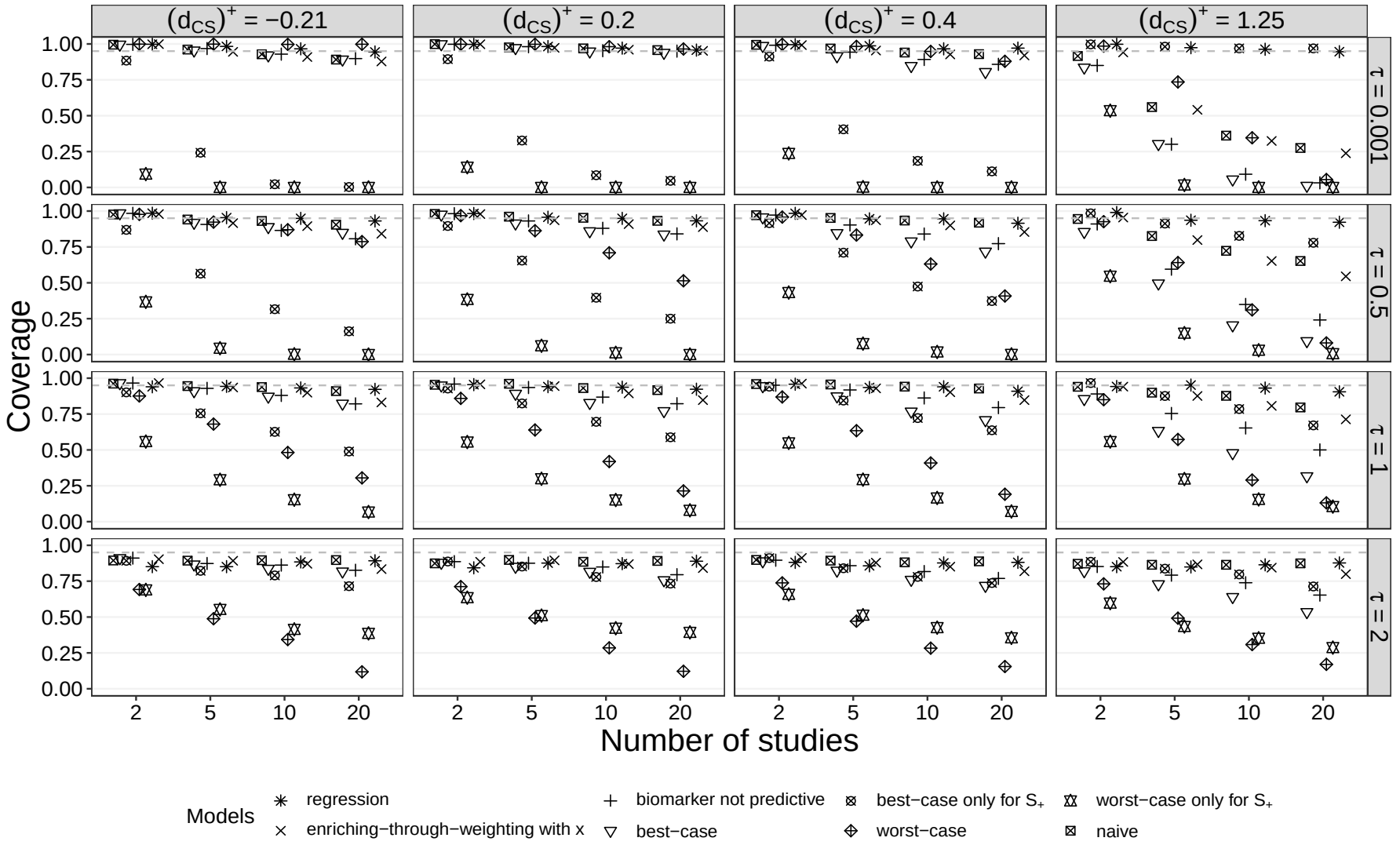


Figure A.3: *Coverage of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A, $x_j \in [0.4, 0.5]$)*

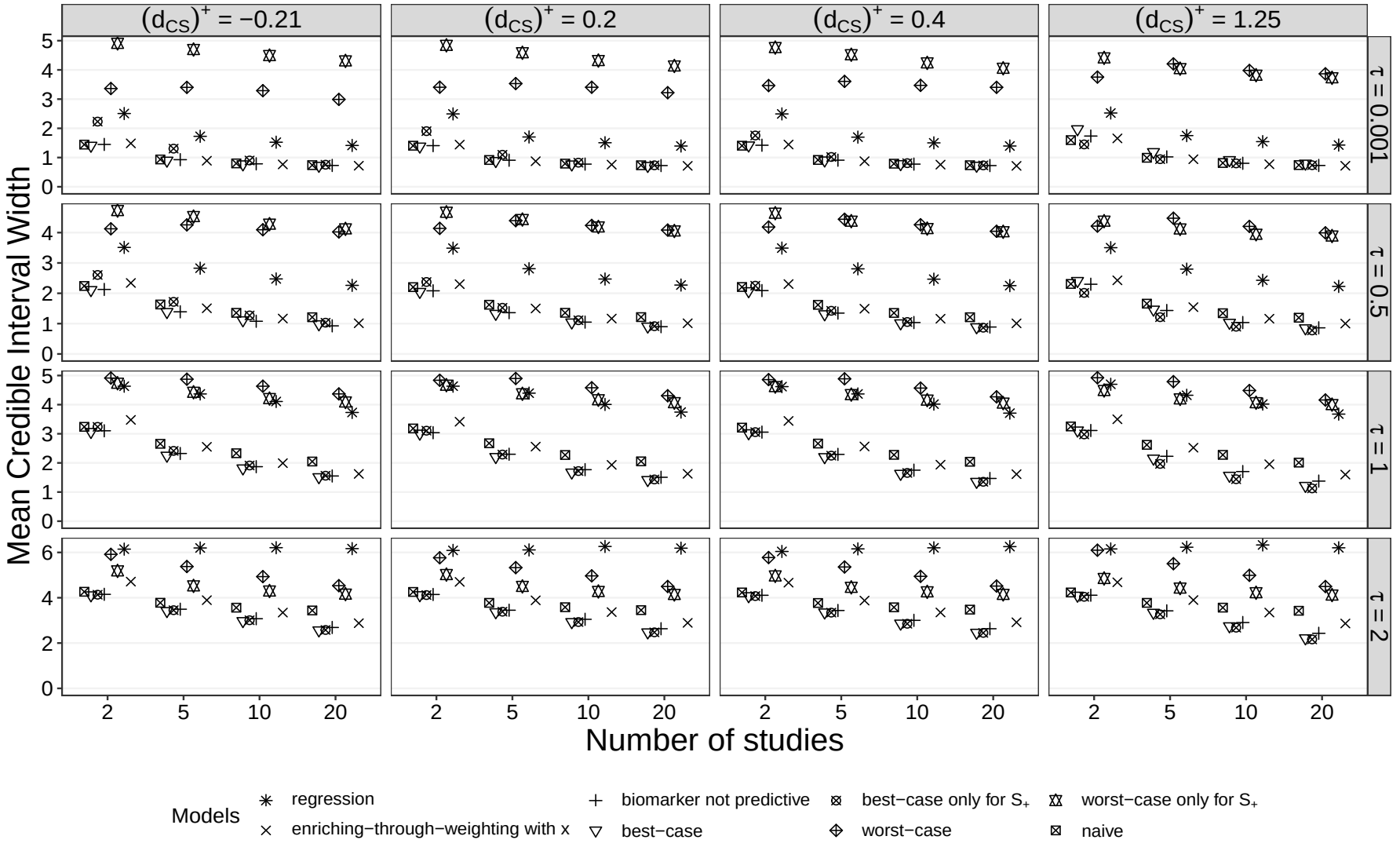


Figure A.4: *Mean 95% credible width of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A, $x_j \in$ [0.4, 0.5])*

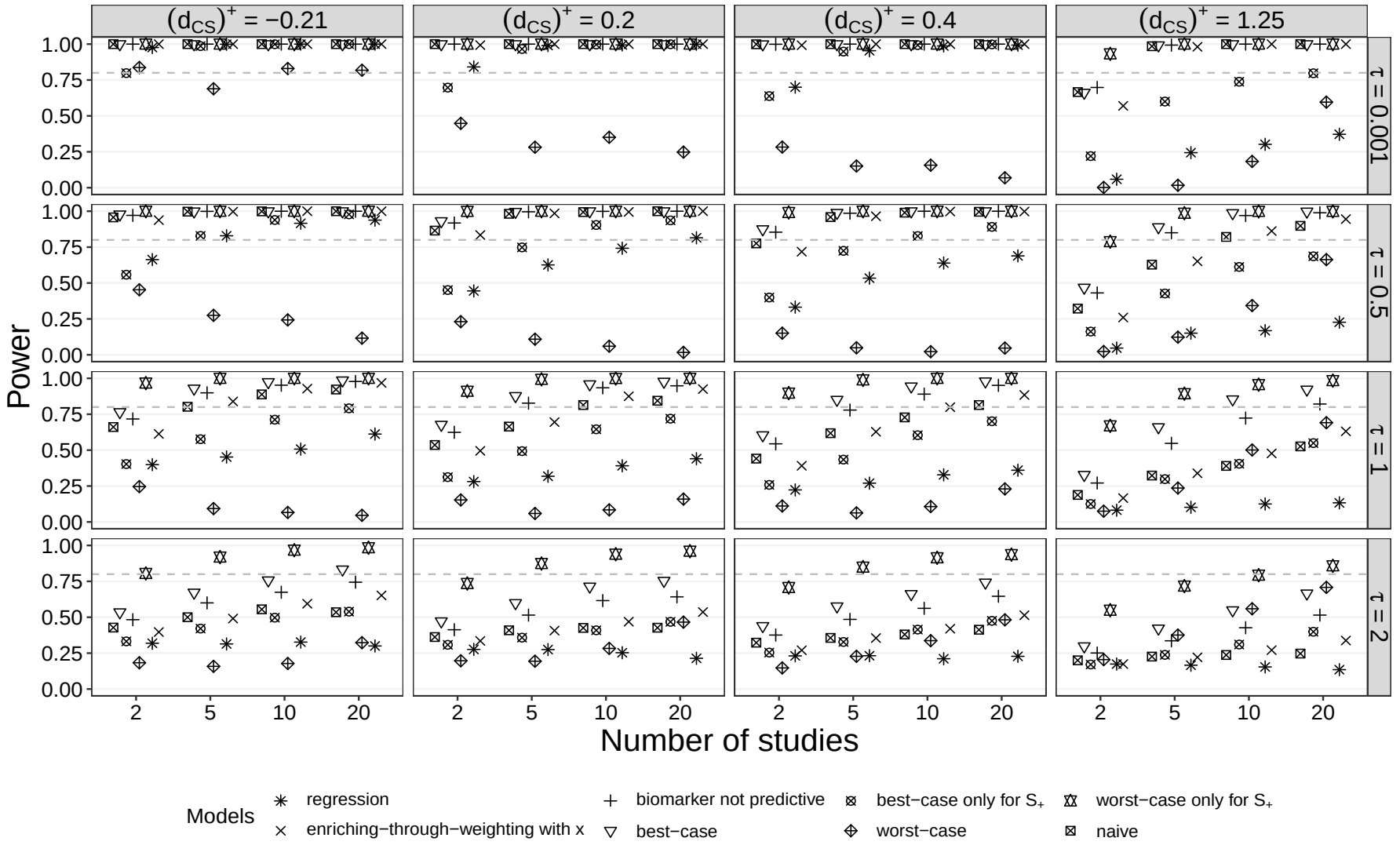


Figure A.5: *Power of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting A, $x_j \in [0.4, 0.5]$)*

A.2 Results setting B

A.2.1 Results of the simulation study with a uniform heterogeneity prior

The results for the simulation study for setting B and covariate $x_j \in [0.4, 0.5]$ are displayed in the following.

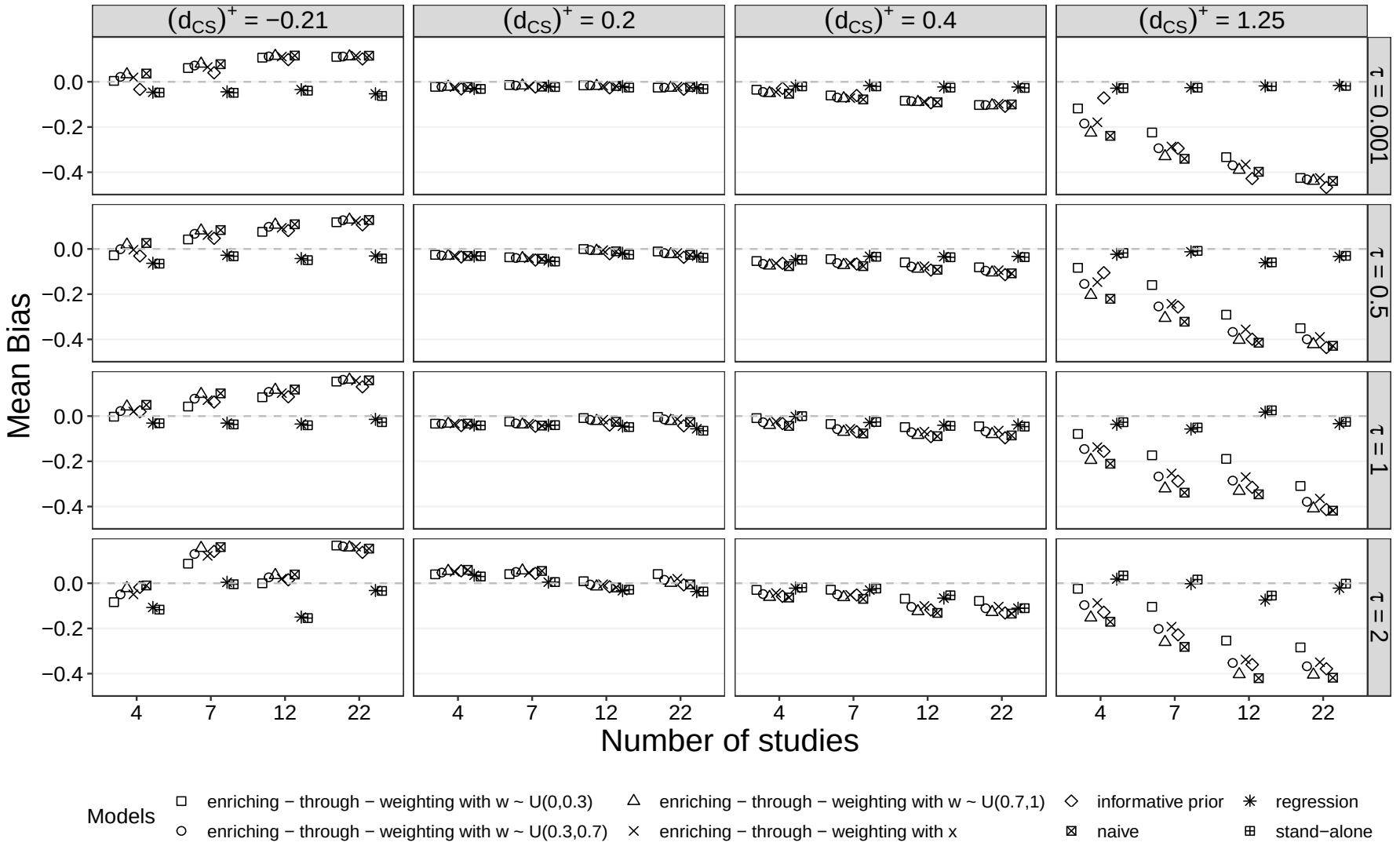


Figure A.6: *Mean bias of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B, $x_j \in [0.4, 0.5]$).*

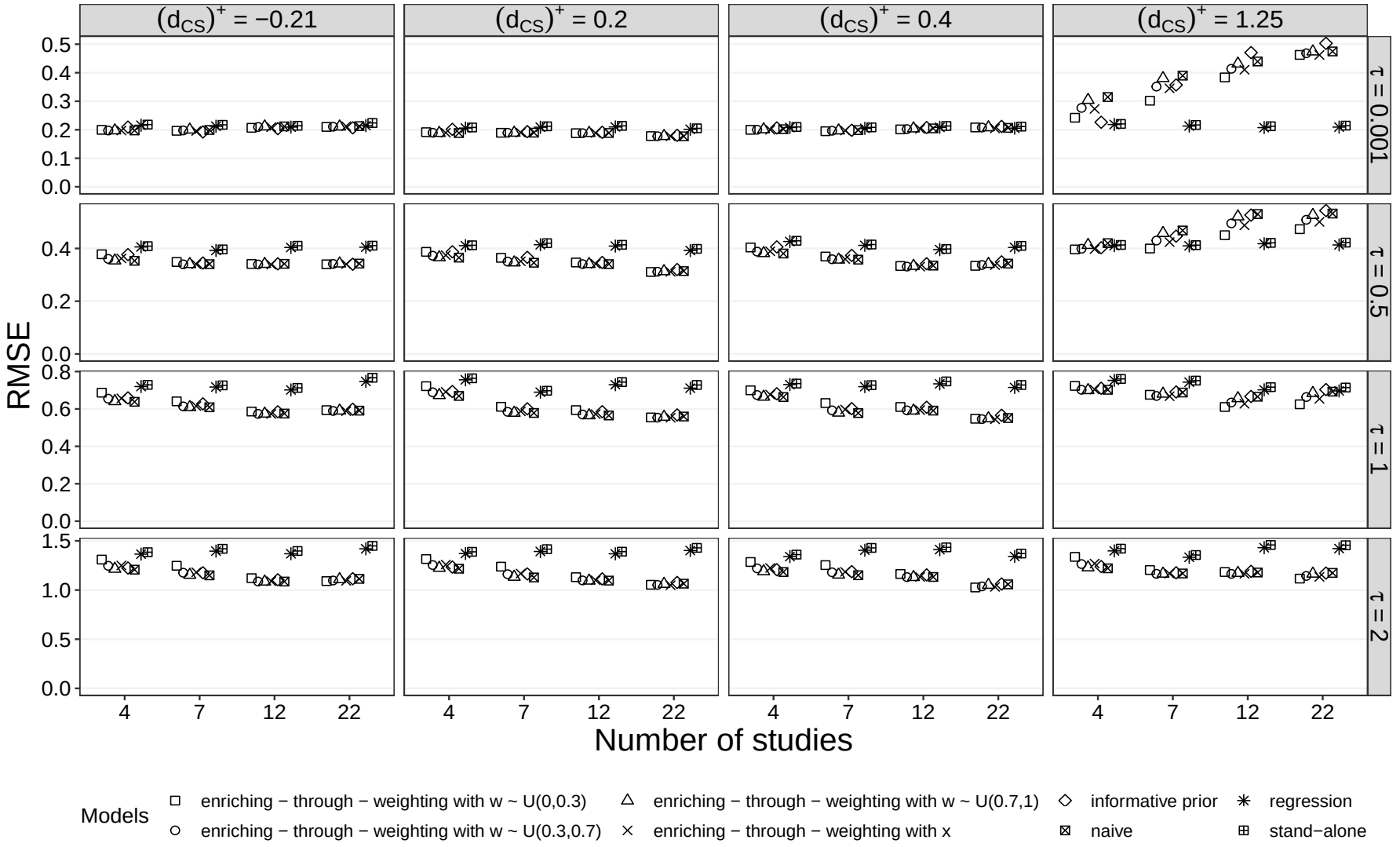


Figure A.7: *RMSE of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B, $x_j \in [0.4, 0.5]$).*

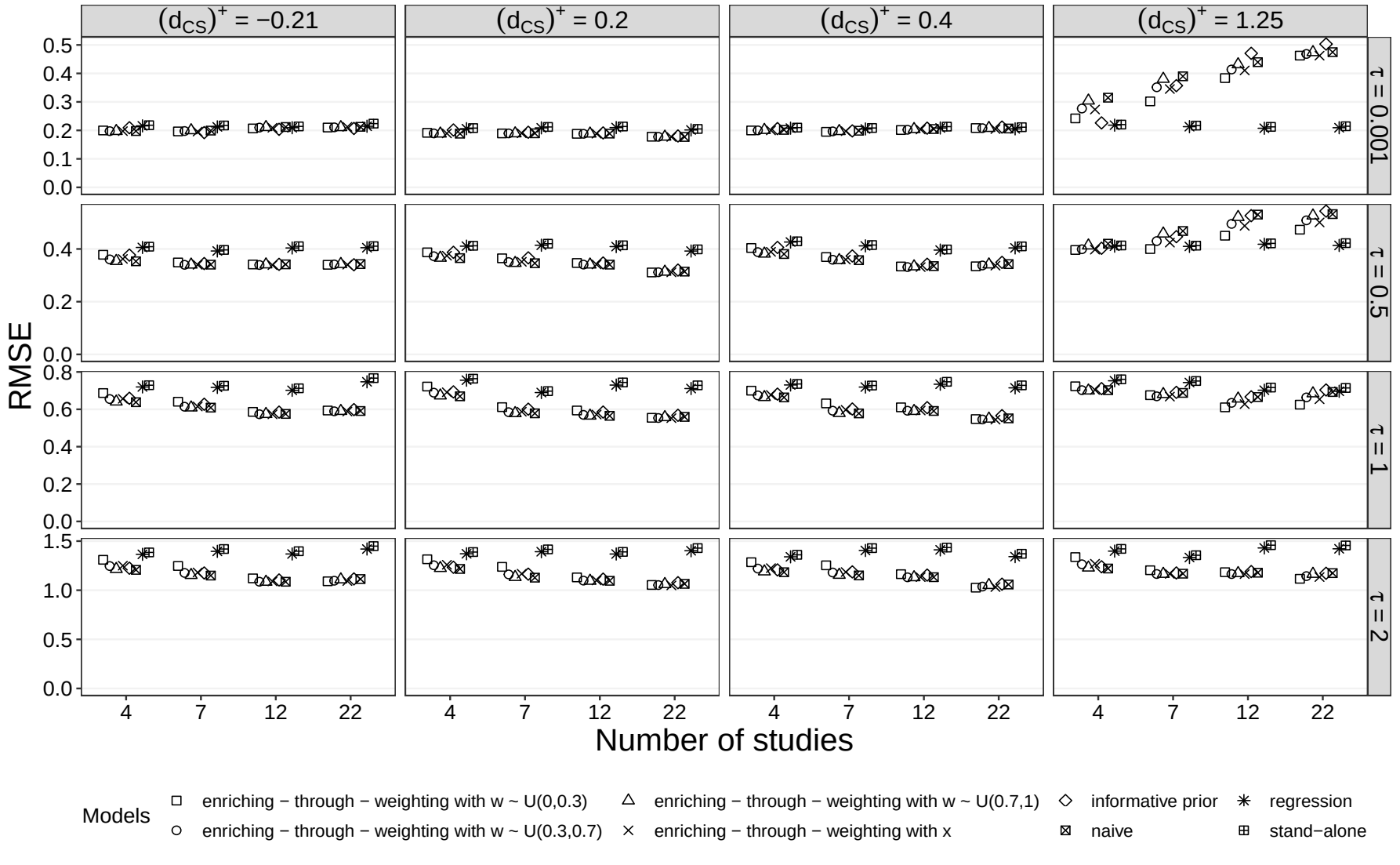


Figure A.8: *Coverage of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B, $x_j \in [0.4, 0.5]$).*

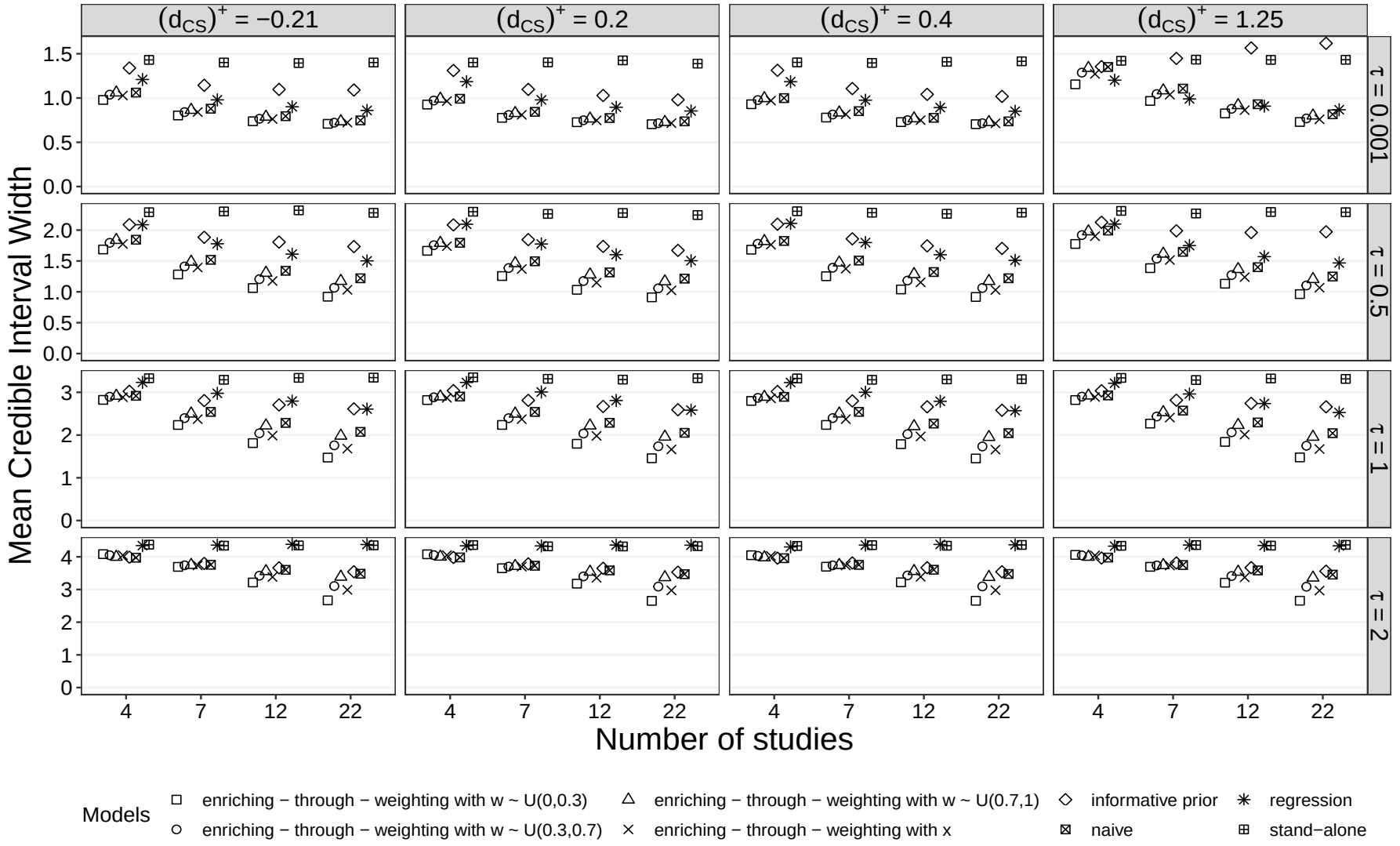


Figure A.9: *Mean 95% credible width of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B, $x_j \in$ [0.4, 0.5]).*

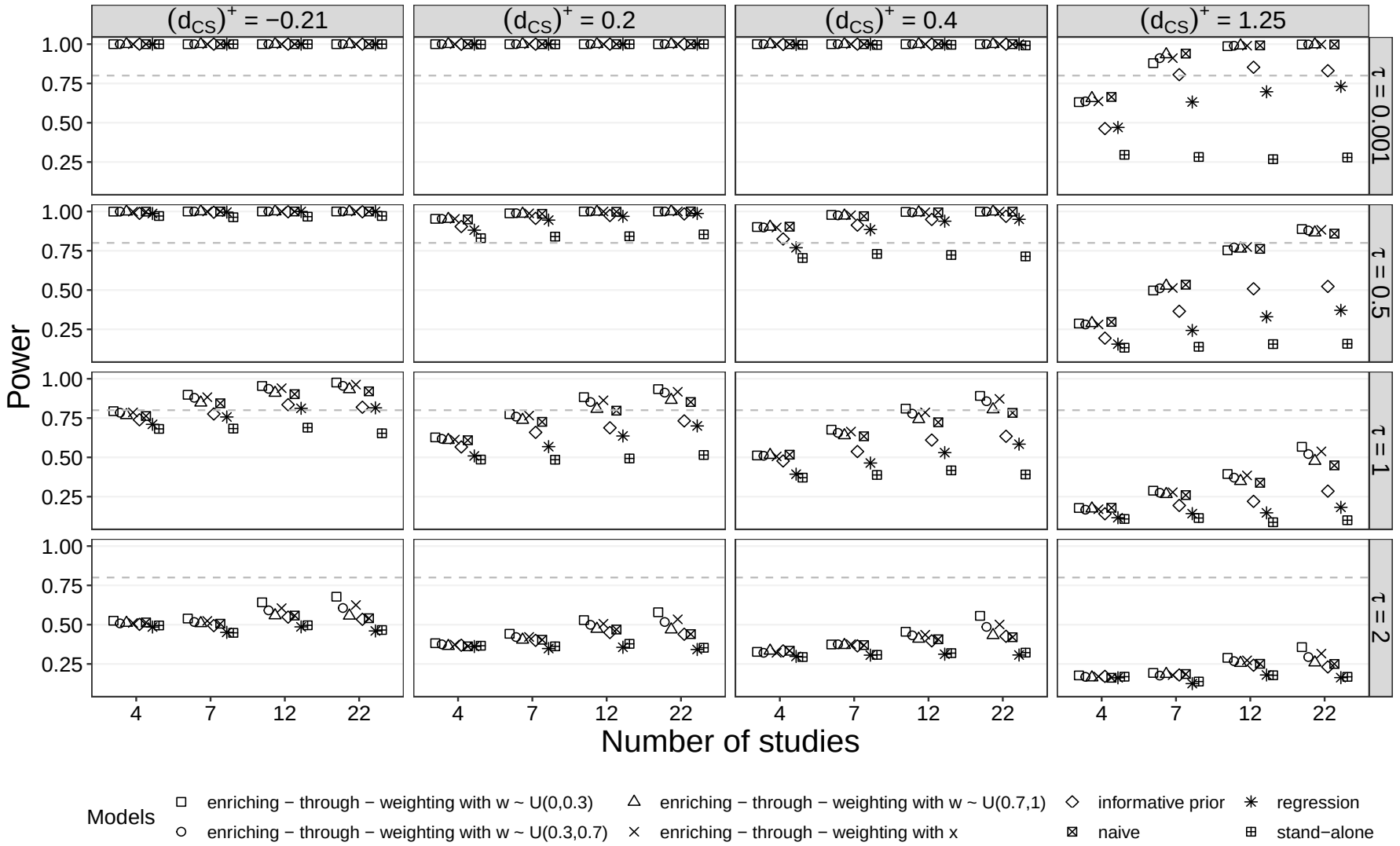


Figure A.10: *Power of the estimated treatment effect comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B, $x_j \in [0.4, 0.5]$).*

A.2.2 Results of the simulation study using a half-normal heterogeneity prior

The results for the simulation study for setting B and covariate $x_j \in [0.3, 0.7]$ are displayed in the following.

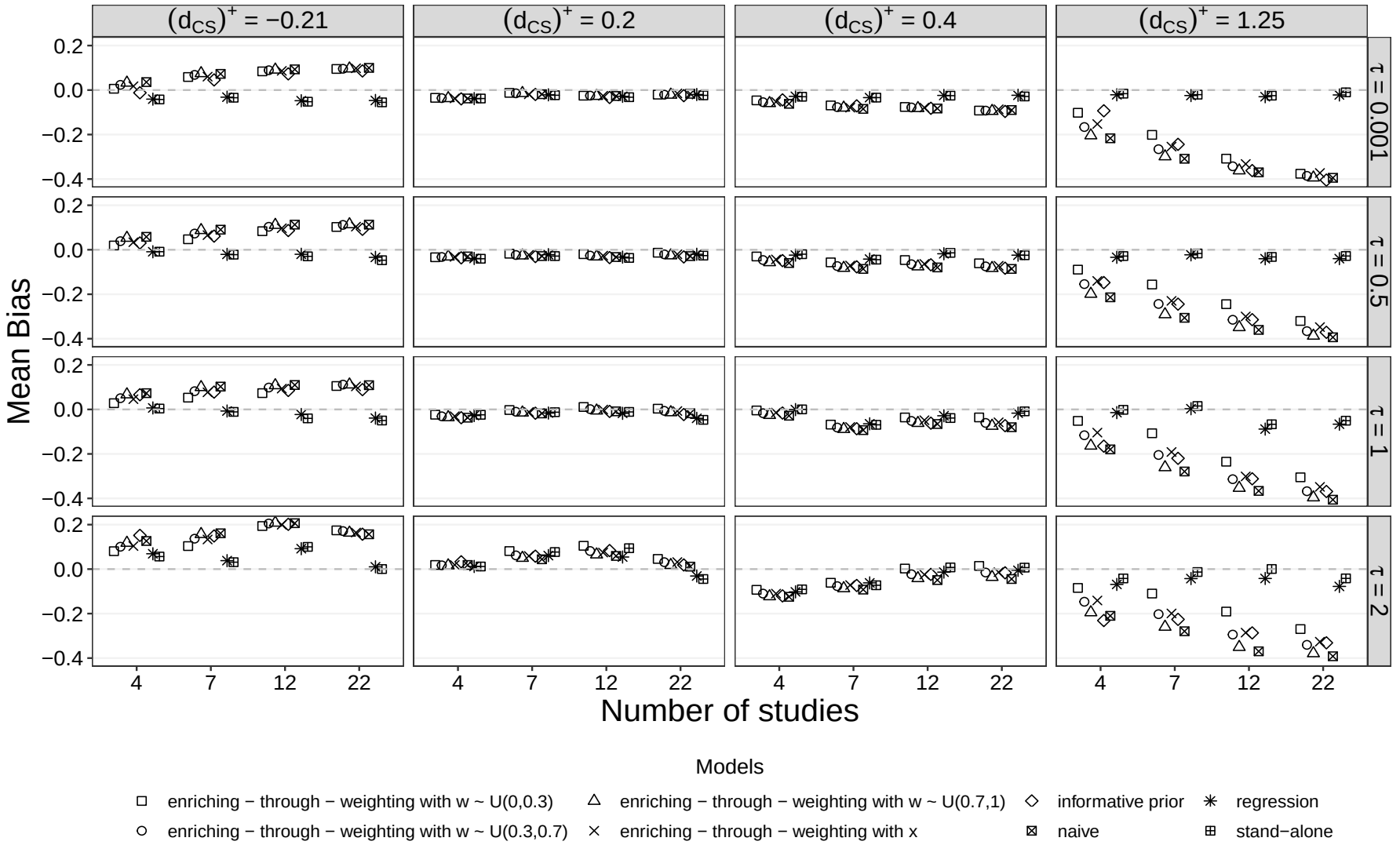


Figure A.11: *Mean bias of the estimated treatment effect with a half-normal heterogeneity prior comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B, half-normal prior)*

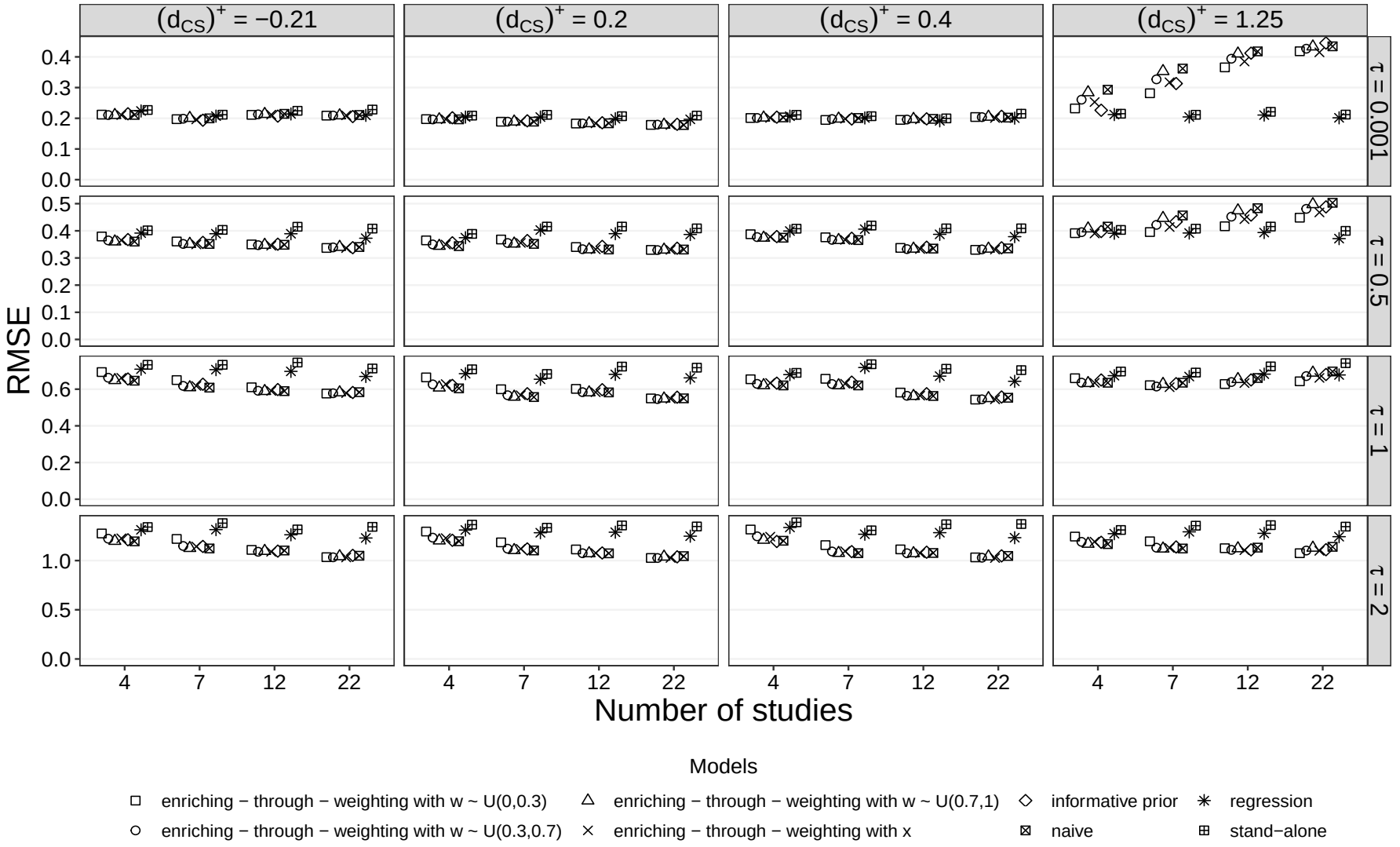


Figure A.12: *RMSE of the estimated treatment effect with a half-normal heterogeneity prior comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B, half-normal prior)*

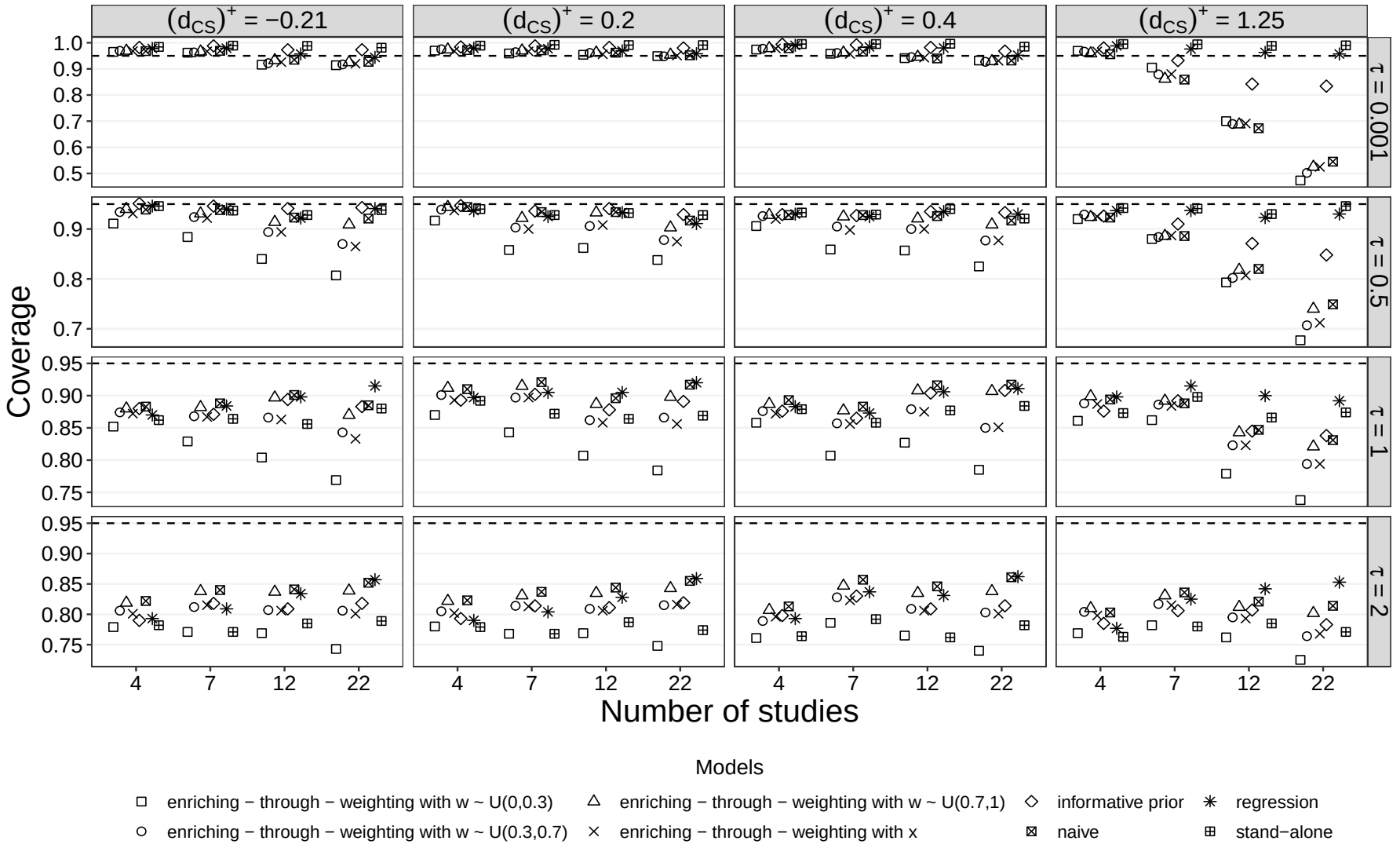


Figure A.13: *Coverage of the estimated treatment effect with a half-normal heterogeneity prior comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B, half-normal prior)*

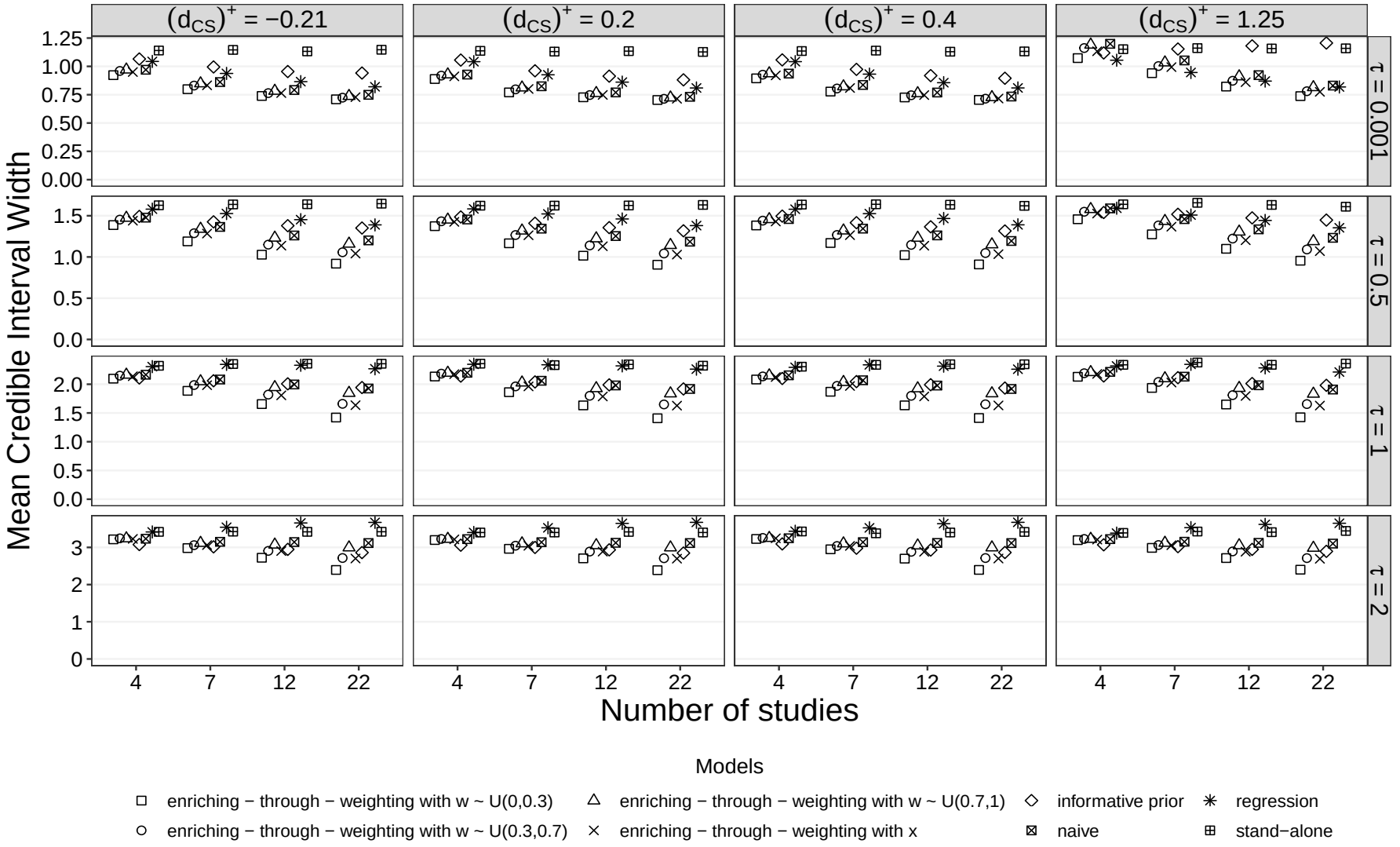


Figure A.14: *Mean 95% credible width of the estimated treatment effect with a half-normal heterogeneity prior comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B, half-normal prior).*

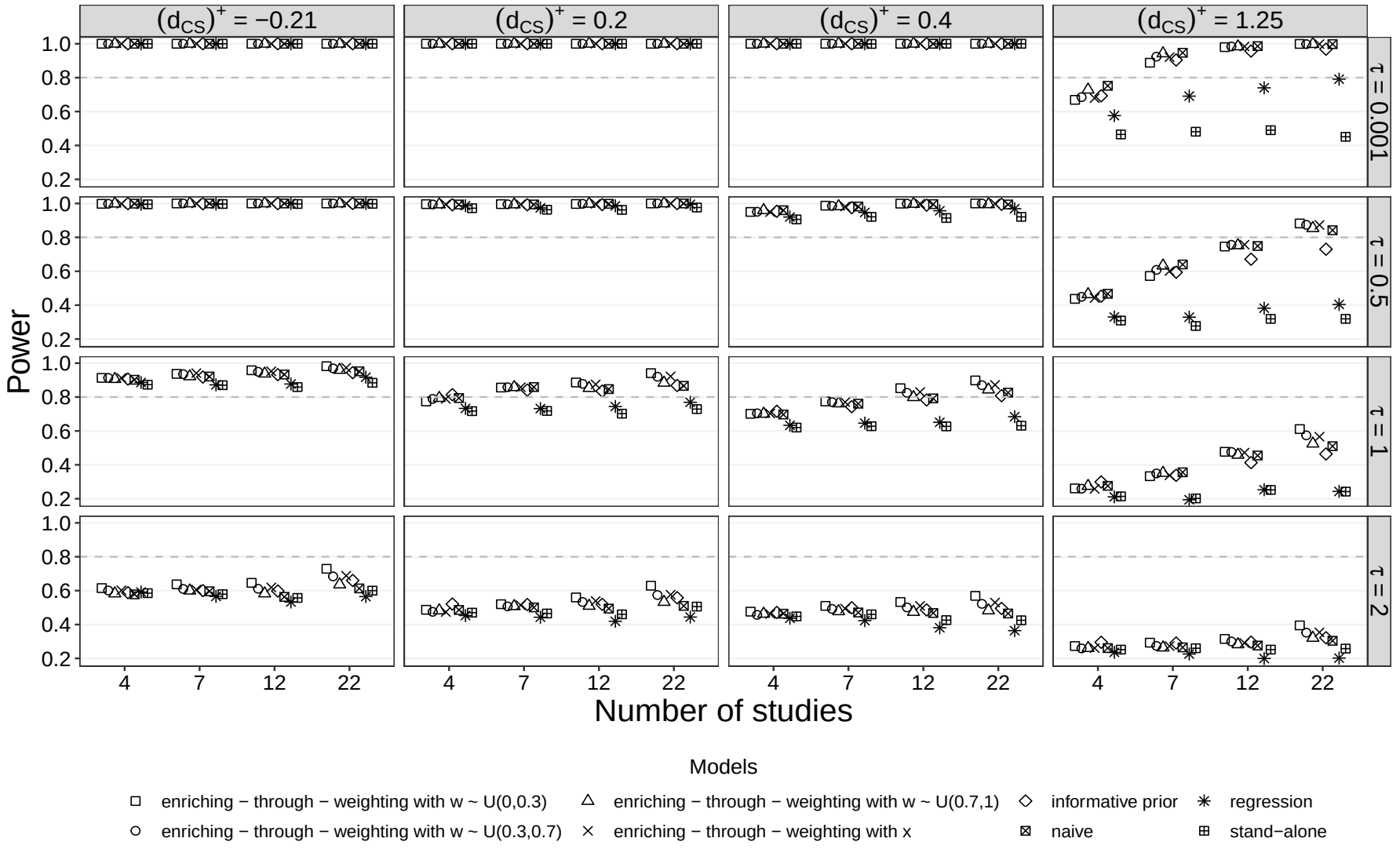


Figure A.15: *Power of the estimated treatment effect with a half-normal heterogeneity prior comparing treatment E_+ with S in a network meta-analysis in the simulation study for different methods (setting B, half-normal prior)*

Appendix B

R Code

B.1 Application

Data for application

study	treatment	n	response	cov	n_BM+	resp_BM+	n_all	resp_all
ENSURE	Erlotinib	110	69	1	110	69	110	69
ENSURE	Chemo	107	36	1	107	36	107	36
EURTAC	Erlotinib	86	50	1	86	50	86	50
EURTAC	Chemo	87	13	1	87	13	87	13
OPTIMAL	Erlotinib	82	68	1	82	68	82	68
OPTIMAL	Chemo	72	26	1	72	26	72	26
IPASS	Gefitinib	223	95	0.597	132	94	609	262
IPASS	Chemo	214	81	0.597	129	61	608	196
NEJGSG	Gefitinib	114	84	1	114	84	114	84
NEJGSG	Chemo	114	35	1	114	35	114	35
First_Signal	Gefitinib	53	29	0.44	26	22	159	88
First_Signal	Chemo	42	20	0.44	16	6	150	69
WJTOG	Gefitinib	58	36	1	58	36	58	36
WJTOG	Chemo	59	19	1	59	19	59	19
Yang14	Gefitinib	35	17	0.677	24	17	118	56
Yang15	Chemo	39	20	0.677	26	17	118	49
V1532	Gefitinib	9	6	1	9	6	9	6
V1533	Chemo	11	5	1	11	5	11	5
INTEREST	Gefitinib	125	15	0.1423	19	8	659	60
INTEREST	Chemo	142	16	0.1423	19	4	657	50
Zhou	Gefitinib	54	6	0.71	13	5	81	11
Zhou	Chemo	54	6	0.71	19	2	76	10
Yang17	Erlotinib	128	76	1	128	76	128	76
Yang18	Chemo	128	67	1	128	67	128	67
Xie	Erlotinib	23	14	1	23	14	23	14
Xie	Chemo	27	15	1	27	15	27	15
WJOG5108L	Erlotinib	201	93	0.717	160	88	227	100
WJOG5108L	Chemo	218	104	0.717	175	103	244	112

Code for application

```
# Calculate data for application study

# Packages
library(rjags)
library(plyr)
library(R2jags)
library(tidyverse)
library(readxl)

data <- read_excel("Daten_Beispiel.xlsx")

# prepare data for model
# subgroup - retrospective approach
n.agg.arms <- 28
n.agg.trials <- 14
a.study <- c(rep(1:14,times = 1, each = 2))
a.treat1 <- c(data$treatment)
a.treat <- as.numeric(revalue(a.treat1, c("Chemo" = 1,
                                          "Erlotinib" = 3,
                                          "Gefitinib" = 2)))

#Number of objective responses biomarker-positive patients
outcome.ad <- c(data$resp_bmpos)

#Number of biomarker-positive patients
n <- c(data$n_bmpos)
a.base <- c(rep(1,22),rep(2,6))

data.s <- list(max.treat = 3,
               n.agg.trials = n.agg.trials,
               n.agg.arms = n.agg.arms,
               a.study = a.study,
               a.treat = a.treat,
               outcome.ad = outcome.ad,
               n = n,
               a.base = a.base)

# set seed
seeds <- sample.int(4, n = .Machine$integer.max)
inits.s <- list(
  list(.RNG.name="base::Wichmann-Hill", .RNG.seed = seeds[1]),
  list(.RNG.name="base::Marsaglia-Multicarry", .RNG.seed = seeds[2]),
  list(.RNG.name="base::Super-Duper", .RNG.seed = seeds[3]),
```

```
  list(.RNG.name="base::Mersenne-Twister", .RNG.seed = seeds[4]))


parameters <- c("d[2]", "d[3]", "lor[2,3]","tau")
true.jags <- jags(data.sim.s,
                  inits.s,
                  parameters ,
                  model.file = "app_naive_U02.txt",
                  n.chains = 4,
                  n.iter = 50000,
                  n.burnin = 10000,
                  n.thin = 1, DIC = FALSE)

jags.true.sum <- true.jags$BUGSoutput$summary

LOR_frame_true <- data.frame(t(as.data.frame(true.jags$BUGSoutput$
    summary[3,
                                                 c(1, 2,
                                                   3,4,5,6,
                                                   7)])),
                             scenario = "true")

colnames(LOR_frame_true)[1] <- "mean_LOR_2_3"


# Enrichment-through-weighting
## with covariate cov[i]

# Create datasets
a.study <- c(rep(1:16,times = 1, each = 2))
a.treat1 <- c(rbind(as.vector(data$treament),
                    as.vector(data$control)))
a.treat <- as.numeric(revalue(a.treat1, c("Chemo" = 1,
                                          "Erlotinib"= 3,
                                          "Gefitinib" = 2)))

#Number of objective responses Biomarker-positive patients
outcome.ad <- c(rbind(as.vector(data$response_treat),
                      as.vector(data$response_cont)))

#Number of Biomarker positive patients
n <- c(rbind(as.vector(data$n_treat),
             as.vector(data$n_cont)))
a.base <- c(rep(1,22),rep(2,6))
a.cov <- rep(data$cov, each=2)
```

```
data.enrich <- list(max.treat = 3,
                    n.agg.trials = n.agg.trials,
                    n.agg.arms = n.agg.arms,
                    a.study = a.study,
                    a.treat = a.treat,
                    outcome.ad = outcome.ad,
                    n = n,
                    a.base = a.base,
                    cov = a.cov)

parameters <- c("d[2]","d[3]", "lor[2,3]","tau")
true.jags <- jags(data.enrich,
                  inits.s,
                  parameters ,
                  model.file = "app_enrich_cov_U02.txt",
                  n.chains = 4,
                  n.iter = 50000,
                  n.burnin = 10000,
                  n.thin = 1,
                  DIC = FALSE)


jags.enrich_cov <- true.jags$BUGSoutput$summary

LOR_enrich_cov <- data.frame(t(as.data.frame(true.jags$BUGSoutput$
    summary["lor[2,3]",
                        c(1, 2, 3,4,5,6, 7)])), scenario = "design_
                            adjust_cov")

colnames(LOR_enrich_cov)[1] <- "mean_LOR_2_3"
parameters <- c("d[2]","d[3]", "lor[2,3]","tau")

## enrichment U(0,0.3)
true.jags <- jags(data.enrich,
                  inits.s,
                  parameters,
                  model.file = "app_enrich_003_U02.txt",
                  n.chains = 4,
                  n.iter = 50000,
                  n.burnin = 10000,
                  n.thin = 2, DIC = FALSE)

jags.enrich_prior_003 <- true.jags$BUGSoutput$summary
```

```
LOR_enrich_prior_003 <- data.frame(t(as.data.frame(true.jags$
    BUGSoutput$summary["lor[2,3]",
                       c(1, 2, 3,4,5,6, 7)])), scenario = "
                           design_adjust_commonw003")

colnames(LOR_enrich_prior_003)[1] <- "mean_LOR_2_3"



## enrichment U(0.3,0.7)
parameters <- c("d[2]","d[3]", "lor[2,3]","tau")
true.jags <- jags(data.enrich,
                  inits.s,
                  parameters ,
                  model.file = "app_enrich_0307_U02.txt",
                  n.chains = 4,
                  n.iter = 50000,
                  n.burnin = 10000,
                  n.thin = 2, DIC = FALSE)

jags.enrich_prior_0307 <- true.jags$BUGSoutput$summary

LOR_enrich_prior_0307 <- data.frame(t(as.data.frame(true.jags$
    BUGSoutput$summary["lor[2,3]",
                       c(1, 2, 3,4,5,6, 7)])), scenario = "
                           design_adjust_commonw0307")

colnames(LOR_enrich_prior_0307)[1] <- "mean_LOR_2_3"

## enrichment U(0.7,1)

parameters <- c("d[2]","d[3]", "lor[2,3]","tau")
true.jags <- jags(data.enrich,
                  inits.s,
                  parameters ,
                  model.file = "app_enrich_071_U02.txt",
                  n.chains = 4,
                  n.iter = 50000,
                  n.burnin = 10000,
                  n.thin = 2, DIC = FALSE)

jags.enrich_prior_071 <- true.jags$BUGSoutput$summary

LOR_enrich_prior_071 <- data.frame(t(as.data.frame(true.jags$
    BUGSoutput$summary["lor[2,3]",
```

```
                                c(1, 2, 3,4,5,6, 7)])), scenario = "design_
                                    adjust_commonw071")

colnames(LOR_enrich_prior_071)[1] <- "mean_LOR_2_3"

# naive approach
data.naive <- list(max.treat = 3,
                   n.agg.trials = n.agg.trials,
                   n.agg.arms = n.agg.arms,
                   a.study = a.study,
                   a.treat = a.treat,
                   outcome.ad = outcome.ad,
                   n = n,
                   a.base = a.base)

parameters <- c("d[2]","d[3]", "lor[2,3]", "tau")

true.jags <- jags(data.naive,
                  inits.s ,
                  parameters ,
                  model.file = "app_naive_U02.txt",
                  n.chains = 4, n.iter = 50000,
                  n.burnin = 20000,
                  n.thin = 1,
                  DIC = FALSE)


jags.naive<- true.jags$BUGSoutput$summary

LOR_naive <- data.frame(t(as.data.frame(true.jags$BUGSoutput$summary[
    "lor[2,3]",
                                c(1, 2, 3,4,5,6, 7)])), scenario = "
                                    naive")

colnames(LOR_naive)[1] <- "mean_LOR_2_3"


#Filter studies with only mixed population
mix_pat_data <- data %>%
  filter(cov!= 1)
n.agg.arms <- 2*nrow(mix_pat_data)
n.agg.trials <- nrow(mix_pat_data)
a.study <- c(rep(1:nrow(mix_pat_data),times=1, each=2))
a.treat1 <- c(treatment)
a.treat <- as.numeric(revalue(a.treat1, c("Chemo" = 1,
```

```
                                              "Erlotinib"= 3,
                                              "Gefitinib" = 2)))

outcome.ad <- c(as.vector(mix_pat_data$resp_all))
n <- c(as.vector(mix_pat_data$n_all))
a.base <- c(rep(1,10),rep(2,2))
studyname <- rep(mix_pat_data$study,each=2)
treatname <- a.treat1

data.mix <- list(max.treat = 3,
                 n.agg.trials = n.agg.trials,
                 n.agg.arms = n.agg.arms,
                 a.study = a.study,
                 a.treat = a.treat,
                 outcome.ad = outcome.ad,
                 n = n,
                 a.base = a.base)

parameters <- c("d[2]","d[3]","tau")
true.jags <- jags(data.mix,
                  inits.s,
                  parameters,
                  model.file = "app_mix_U02.txt",
                  n.chains = 4,
                  n.iter = 50000,
                  n.burnin = 10000,
                  n.thin = 2, DIC = FALSE)


jags.only.mixed <- true.jags$BUGSoutput$summary

# Stand-alone approach
only_bmp_data <- data %>%
  filter(cov == 1)
n.agg.arms <- 2*nrow(only_bmp_data)
n.agg.trials <- nrow(only_bmp_data)
a.study <- c(rep(1:nrow(only_bmp_data),times = 1, each = 2))
a.treat1 <- c(rbind(as.vector(only_bmp_data$treament),
                    as.vector(only_bmp_data$control)))
a.treat <- as.numeric(revalue(a.treat1, c("Chemo"=1,
                                          "Erlotinib"= 3,
                                          "Gefitinib" = 2)))

outcome.ad <- c(rbind(as.vector(only_bmp_data$response_treat),
                      as.vector(only_bmp_data$response_cont)))
```

```
n <- c(rbind(as.vector(only_bmp_data$n_treat),
             as.vector(only_bmp_data$n_cont)))

a.base <- c(rep(1,12),rep(2,4))

data.bm <- list(max.treat = 3,
                n.agg.trials = n.agg.trials,
                n.agg.arms = n.agg.arms,
                a.study = a.study,
                a.treat = a.treat,
                outcome.ad = outcome.ad,
                n = n,
                a.base = a.base)

parameters <- c("d[2]","d[3]", "lor[2,3]", "tau")

true.jags <- jags(data.bm,
                  inits.s ,
                  parameters ,
                  model.file = "app_stand_alone_U02.txt",
                  n.chains = 4, n.iter = 50000,
                  n.burnin = 20000,
                  n.thin = 1,
                  DIC = FALSE)

jags.stand_alone <- true.jags$BUGSoutput$summary

LOR_standalone <-data.frame(t(as.data.frame(true.jags$BUGSoutput$
    summary["lor[2,3]",
                        c(1, 2, 3,4,5,6, 7)])), scenario = "stand-
                            alone")

colnames(LOR_standalone)[1] <- "mean_LOR_2_3"

# Informative prior
data.inf <- list(n.agg.trials = n.agg.trials,
                n.agg.arms = n.agg.arms,
                a.study = a.study,
                a.treat = a.treat,
                outcome.ad = outcome.ad,
                n = n,
                a.base = a.base)

parameters <- c("d[2]","d[3]", "lor[2,3]", "tau")
```

```
true.jags <- jags(data.inf,
                  inits.s ,
                  parameters ,
                  model.file = "app_inf_prior_U02.txt",
                  n.chains = 4, n.iter = 50000,
                  n.burnin = 20000,
                  n.thin = 1,
                  DIC = FALSE)

jags.inf <- true.jags$BUGSoutput$summary

LOR_inf <- data.frame(t(as.data.frame(true.jags$BUGSoutput$summary["
    lor[2,3]",
                c(1, 2, 3,4,5,6, 7)])), scenario = "informative_prior
                    ")

colnames(LOR_inf)[1] <- "mean_LOR_2_3"

# Informative prior with different weight
data.inf <- list(n.agg.trials = n.agg.trials,
                 n.agg.arms = n.agg.arms,
                 a.study = a.study,
                 a.treat = a.treat,
                 outcome.ad = outcome.ad,
                 n = n,
                 a.base = a.base)

parameters <- c("d[2]","d[3]", "lor[2,3]", "tau")
true.jags <- jags(data.inf,
                  inits.s ,
                  parameters ,
                  model.file = "app_inf_prior_U02_ww003.txt",
                  n.chains = 4, n.iter = 50000,
                  n.burnin = 20000,
                  n.thin = 1,
                  DIC = FALSE)

jags.inf1 <- true.jags$BUGSoutput$summary

LOR_inf1 <- data.frame(t(as.data.frame(true.jags$BUGSoutput$summary["
    lor[2,3]",
                      c(1, 2, 3,4,5,6, 7)])), scenario = "informative
                          _prior_ww003")

colnames(LOR_inf1)[1] <- "mean_LOR_2_3"
```

```
# Regression model
parameters <- c("dz[2]","dz[3]","lorz[2,3]", "tau")

regression.jags <- jags(data.enrich,
                        inits.s ,
                        parameters,
                        model.file = "app_regression_U02.txt",
                        n.chains = 4, n.iter = 50000,
                        n.burnin = 20000,
                        n.thin = 1,
                        DIC = FALSE)

jags.regression <- regression.jags$BUGSoutput$summary

LOR_frame_reg <- data.frame(t(as.data.frame(regression.jags$
    BUGSoutput$summary["lorz[2,3]",
                       c(1, 2, 3,4,5,6, 7)])),scenario="regression
                           ")

colnames(LOR_frame_reg)[1] <- "mean_LOR_2_3"

## Combine all forest data
forest_data_8 <- rbind(LOR_frame_true,
                       LOR_enrich_cov,
                       LOR_enrich_prior_003,
                       LOR_enrich_prior_0307,
                       LOR_enrich_prior_071,
                       LOR_inf,
                       LOR_inf1,
                       LOR_naive,
                       LOR_standalone,
                       LOR_frame_reg)

save(forest_data_8, file = "forest_data_diss.Rda")
```

Code for application plot

```
# Set working directory
setwd("~/application")

# Packages
library(tidyverse)
library(xtable)
library(ggplot2)
```

```
# Load data
load("forest_data_diss_U02.Rda")
latextab<-xtable(forest_data_8)

forest_data_CI<-mutate(forest_data_8,
                       Precision=X97.5.-X2.5.)
latextabCI<-xtable(forest_data_CI)

forest_data <- forest_data_8 %>% mutate %>% add_column(y=0) %>%
               filter(!scenario %in% c("informative_prior_ww003"))
forest_data$scenario1 <- plyr::revalue(forest_data$scenario,
                                c("design_adjust_cov"="enriching-
                                   through-weighting␣with␣x",
                                  "design_adjust_commonw003"= "
                                     enriching-through-weighting␣w
                                     (0,0.3)",
                                  "design_adjust_commonw0307"= "
                                     enriching-through-weighting␣w
                                     (0.3,0.7)",
                                  "design_adjust_commonw071"= "
                                     enriching-through-weighting␣w
                                     (0.7,1)",
                                  "informative_prior"="informative␣
                                     prior␣w(0.3,␣0.7)",
                                  "true"="retrospective␣analysis"))
forest_data2<-forest_data %>%
  arrange(desc(mean_LOR_2_3)) %>%
  mutate(scenario1 = factor(scenario1, levels=scenario1))


pd <- position_dodge(0.3)

ggplot(forest_data2, aes(y,mean_LOR_2_3, shape = scenario1),
       label = sprintf("%0.2f", round(mean_LOR_2_3, digits = 2)))+
  geom_point(size = 5, position = position_dodge(0.1)) +
  geom_errorbar(aes(ymax = X2.5.,
                    ymin = X97.5.), width = .04, position=position_
                       dodge(0.1))+
  geom_text(aes(label = paste( sprintf("%0.2f", mean_LOR_2_3),"␣[␣",
                               sprintf("%0.2f", X2.5.), "␣,␣",
                               sprintf("%0.2f", X97.5.),"␣]␣",sep = "
                                  "),
                y=1.36),
            size=7,
```

```
            position=position_dodge(0.1),
            hjust=0) +
  geom_text(aes(label= scenario1,
                y=-1),
            position=position_dodge(0.1),
            hjust=0,
            size = 7)+
  coord_flip(clip = 'off')+
  ylab("Log odds ratio greater 0 favours Erlotinib compared to
      Gefitinib")+
  geom_hline(aes(yintercept = 0.2108953), linetype="dotted")+ #value
      true
  geom_hline(aes(yintercept=0),colour="blue")+
  scale_shape_manual(values = c(rep(16,9)))+
  scale_x_discrete(expand = c(0,0))+
  scale_y_continuous(breaks=seq(-0.4,1.3,0.1))+
  theme_bw()+
  theme(axis.line.x = element_line(color="black", size = 1),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank(),
        axis.title.x = element_text(size = rel(2.5), angle = 00),
        axis.text.x = element_text(angle = 00, size=rel(2.5)),
        axis.ticks.y=element_blank(),
        axis.text.y=element_blank(),
        axis.title.y = element_blank(),
        legend.position = "none",
        plot.margin = margin(0.1, 5, 0.1, 0.05, "cm"))
  ggsave("application_plot_diss_uniform.pdf", width=20, height=10)
```

B.2 Simulation study

Code for data simulation:

```
#######################################################################
# Code to simulate the data for the network meta-analysis,
# this code is based on the approach of Svenja Seide et al (2021)
#######################################################################

########################################
# packages needed
########################################
library("tidyverse")
```

```
library("mvtnorm")

# Set working directory
setwd("~/dgm_11052021")

# Create theoretical network

## function for odd ratio
OR <- function(x,y){
  (x/(1-x))/(y/(1-y))
}

nc <- 200
nt <- 200
repetitions <- 1000
# these are the study numbers for the mixed patient population
n_study <- c(2,5,10,20)
tau2 <- c(0.001, 0.5, 1, 2)^2
ptpos.th <- c(0.7,0.5,0.45,0.35)
pc.th <- 0.4
ptneg.th <- 0.45
pe.th <- 0.8


set.seed(1234)

simulated_scenarios_response_rate_long_overall <- list()
simulated_scenarios_response_rate_long_subgroup <- list()

for(u in 1:length(ptpos.th)){
  simulated_scenarios_study_long_overall <- list()
  simulated_scenarios_study_long_subgroup <- list()

  for (c in 1:length(n_study)){
    #Setting A
    cov03 <- c(seq(0.3,0.7, length = n_study[c]))
    cov04 <- c(seq(0.4,0.5, length = n_study[c]))

    #Setting B
    #cov03 <- c(seq(0.3,0.7, length = n_study[c]),1,1)
    #cov04 <- c(seq(0.4,0.5, length = n_study[c]),1,1)

    cov <- c()

    simulated_scenarios_cov_long_overall <- list()
```

```
simulated_scenarios_cov_long_subgroup <- list()

for (d in 1:2){
  if (d == 1){
    cov <- cov03}
  else {
    cov <- cov04}

  # Set treat_1 (1 = control all; 4 = E_+)
  # Set treat_2 (2 = BM_+ patients, 3 = BM_- patients )
  theo_network <- data.frame(
                    treat_1 = c(1,1,1,4),
                    treat_2 = c(2,3,4,2),
                    oddsratio = c(OR(ptpos.th[u],pc.th),
                                  OR(ptneg.th,pc.th),
                                  OR(pe.th,pc.th),
                                  OR(ptpos.th[u],pe.th)),
                    theta = c(log(OR(ptpos.th[u],pc.th)),
                              log(OR(ptneg.th,pc.th)),
                              log(OR(pe.th,pc.th)),
                              log(OR(ptpos.th[u],pe.th))),
                  # Setting A
                  n_trial = c(n_study[c],
                              n_study[c],3,1))
                  # Setting B
                  # 2 studies with only BM_+ patients are added
                  # to the number of study with mixed pat.pop.
                    # n_trial = c((n_study[c]+2),
                    #             (n_study[c]+2),3,1))

  ### spread into the network meta-analysis
  theo_nma <- theo_network %>%
    rowwise() %>%
    slice(rep(1:n(), times = n_trial)) %>%
    data.frame() %>%
    select(-n_trial)

  theo_nma <- theo_nma %>%
    mutate(# 3-arm studies
      # Setting A
      study.id = c(rep(c(1:(n_study[c])),2),
      # Setting B
      #study.id = c(rep(c(1:(n_study[c]+2)),2),
                   # 2-arm studies
                   (n_study[c]+3),(n_study[c]+4),
```

```
                    # 2-arm studies
                    (n_study[c]+5),(n_study[c]+6)),
    narm_1 = c(
      round(c(nc*cov,nc*(1-cov))),107,87,72,86),
    narm_2 = c(
      round(c(nt*cov,nt*(1-cov))),110,86,82,82),
    #Setting A
    cov_giv = c(cov,cov),
    #Setting B
    #cov_giv = c(cov,cov,1,1,1,1),
    rate_th_treat=c(
      # Setting A
      rep(ptpos.th[u],(n_study[c])),
      rep(ptneg.th,(n_study[c])),
      # Setting B
      #rep(ptpos.th[u],(n_study[c]+2)),
      #rep(ptneg.th,(n_study[c]+2)),
      rep(pe.th,3),ptpos.th[u]),
    rate_th_cont=c(
      # Setting A
      rep(pc.th,(2*(n_study[c]))),
      # Setting B
      # rep(pc.th,(2*(n_study[c]+2))),
      rep(pc.th,3),pe.th)
  ) %>%
  select(study.id, treat_1, treat_2, theta,
         narm_1, narm_2,oddsratio,cov_giv,
         rate_th_treat, rate_th_cont)

# narm_1=c(nc,nc,89,86),
# 3rd pos: n_c := mean from standard in sim before;
# 4rd pos: n_e_+ := mean from before
# narm_2=c(nt/2,nt/2,93,82),
# 3rd n_e+:= mean from e_+ in sim before;
# 4rd n_s:=mean from before

simulated_scenarios_tau_long_overall <- list()
simulated_scenarios_tau_long_subgroup <- list()

for(v in 1:length(tau2)){

  simulated_data_long_overall <- list()
  simulated_data_long_subgroup <- list()

  for(l in 1:repetitions){
```

```
simulated_nma <- c()
for(j in 1:length(unique(theo_nma$study.id))){

  current.study.id <-  unique(theo_nma$study.id)[j]
  current.study <- theo_nma[which(
    theo_nma$study.id == current.study.id),]


  current.Sigma <- matrix(0.5*tau2[v],
                          ncol = nrow(current.study),
                          nrow = nrow(current.study)) +
    diag(0.5*tau2[v], nrow(current.study))

  theta.i <- mvtnorm::rmvnorm(n = 1,
                              mean = current.study$theta,
                              sigma = current.Sigma)

  current.study <- current.study %>%
    mutate(theta.i = as.vector(theta.i))

 ### for DGM fixed modified
  if(nrow(current.study) == 1){ #2-arm studies
    fr <- function(x) {
      (((current.study$rate_th_treat[1]+
           current.study$rate_th_cont[1])/2) - x)^2 +
        ((x*exp(current.study$theta.i[1]))/
                (1 - x + x*exp(current.study$theta.i[1]))
                   -
                 ((current.study$rate_th_treat[1]+
                    current.study$rate_th_cont[1])/2))^2
    }
  }

  if(nrow(current.study) == 2){#3-arm studies
  fr <- function(x) {
    (((current.study$rate_th_treat[1]+
         current.study$rate_th_cont[1])/2)  - x)^2 +
      ((x*exp(current.study$theta.i[1]))/
         (1 - x + x*exp(current.study$theta.i[1])) -
         ((current.study$rate_th_treat[1]+
             current.study$rate_th_cont[1])/2))^2 +
      ((x*exp(current.study$theta.i[2]))/
         (1 - x + x*exp(current.study$theta.i[2])) -
         ((current.study$rate_th_treat[2]+
```

```
                current.study$rate_th_cont[2])/2) )^2
    }
    }
    current.pi.arm1 <- optimize(f = fr, interval = c(0, 1),
                                maximum = FALSE)$minimum

    current.study$pi.arm1.DGMmf <- rep(current.pi.arm1,
                                       nrow(current.study))

    current.study$pi.arm2.DGMmf <- (current.study$pi.arm1.
        DGMmf *
                                      exp(current.study$theta
                                          .i))/
      (1 - current.study$pi.arm1.DGMmf +
         current.study$pi.arm1.DGMmf *
         exp(current.study$theta.i))


    simulated_nma <- rbind(simulated_nma, current.study)

  }

  simulated_nma1 <- c()
  for(j in 1:length(unique(simulated_nma$study.id))){
    current.study.id <-  unique(simulated_nma$study.id)[j]
    current.study <- simulated_nma[which(
      simulated_nma$study.id == current.study.id),]
    if(current.study$treat_1[1]==1){#here treat 1 is base
        treatment
    current.pi.arm1 <- runif(n = 1, min = 0.3, max = 0.5)
    current.study$pi.arm1.DGMf <- rep(current.pi.arm1,
                                      nrow(current.study))


    current.study$x.arm1.DGMf <- rbinom(n = nrow(current.
        study) ,
                                        prob = current.pi.
                                            arm1,
                                        size = current.study$
                                            narm_1)

    current.study$pi.arm2.DGMf <- (current.study$pi.arm1.DGMf
        *
                                     exp(current.study$theta.
                                         i))/
```

```
      (1 - current.study$pi.arm1.DGMf +
         current.study$pi.arm1.DGMf *
         exp(current.study$theta.i))

    current.study$x.arm2.DGMf <- rbinom(
                          n = nrow(current.study),
                          prob = current.study$pi.arm2.DGMf,
                          size = current.study$narm_2)
    } else { #here E_+ (4) is base for comparison E_+ versus
       S
      current.pi.arm1 <- runif(n = 1, min = 0.7, max = 0.9)
      current.study$pi.arm1.DGMf <- rep(current.pi.arm1,
                                        nrow(current.study))


      current.study$x.arm1.DGMf <- rbinom(n = nrow(current.
          study) ,
                                          prob = current.pi.
                                             arm1,
                                          size = current.
                                             study$narm_1)

      current.study$pi.arm2.DGMf <- (current.study$pi.arm1.
          DGMf *
                                       exp(current.study$
                                          theta.i))/
        (1 - current.study$pi.arm1.DGMf +
           current.study$pi.arm1.DGMf *
           exp(current.study$theta.i))

      current.study$x.arm2.DGMf <- rbinom(
                                n = nrow(current.study),
                                prob = current.study$pi.arm2.
                                   DGMf,
                                size = current.study$narm_2)
                                        }

    simulated_nma1 <- rbind(simulated_nma1, current.study)

  }

  simulated_nma <- cbind(simulated_nma, simulated_nma1$pi.
      arm1.DGMf,
                         simulated_nma1$x.arm1.DGMf,
                         simulated_nma1$pi.arm2.DGMf,
```

```
                        simulated_nma1$x.arm2.DGMf)

  simulated_data_long_subgroup1 <- simulated_nma %>%
  mutate(outcomearm_1 = simulated_nma1$x.arm1.DGMf,
         outcomearm_2 = simulated_nma1$x.arm2.DGMf)%>%
  filter(treat_2!= 3)

simulated_data_long_subgroup[[1]] <- reshape(
  as.data.frame(simulated_data_long_subgroup1),
  varying = c("treat_1","treat_2",
              "narm_1","narm_2",
              "outcomearm_1","outcomearm_2"),
  direction = "long",
  idvar = "study.id",
   sep = "_",
  timevar = "order")

#Add number of arms up in order to get numbers for overall
    population;
#Arm 1 number BM_+ patients,Arm 2 number of BM_- patients
simulated_data_long_overall_1 <- simulated_nma %>%
  mutate(narmaddup_1 = ave(narm_1,study.id,FUN=sum),
         narmaddup_2 = ave(narm_2,study.id,FUN=sum),
         outcomeaddup_1 = ave(simulated_nma1$x.arm1.DGMf,
             study.id,FUN=sum),
         outcomeaddup_2 = ave(simulated_nma1$x.arm2.DGMf,
             study.id,FUN=sum))%>%
  group_by(study.id) %>%
  slice(1)

simulated_data_long_overall[[1]] <- reshape(
  as.data.frame(simulated_data_long_overall_1),
                varying = c("treat_1",
                            "treat_2",
                           "narm_1",
                           "narm_2",
                           "narmaddup_1",
                           "narmaddup_2",
                           "outcomeaddup_1",
                           "outcomeaddup_2"),
                direction = "long",
                    idvar = "study.id",
                      sep = "_",
                  timevar = "order")
```

```
        }
        simulated_scenarios_tau_long_overall[[v]] <- simulated_data_
            long_overall
        simulated_scenarios_tau_long_subgroup[[v]] <-simulated_data_
            long_subgroup
      }
      simulated_scenarios_cov_long_overall[[d]] <- simulated_
          scenarios_tau_long_overall
      simulated_scenarios_cov_long_subgroup[[d]] <- simulated_
          scenarios_tau_long_subgroup
    }
    simulated_scenarios_study_long_overall[[c]] <- simulated_
        scenarios_cov_long_overall
    simulated_scenarios_study_long_subgroup[[c]] <- simulated_
        scenarios_cov_long_subgroup
  }
  simulated_scenarios_response_rate_long_overall[[u]] <- simulated_
      scenarios_study_long_overall
  simulated_scenarios_response_rate_long_subgroup[[u]] <- simulated_
      scenarios_study_long_subgroup
}

save(simulated_scenarios_response_rate_long_overall,
     file ="simulated_data_overall.RData")
save(simulated_scenarios_response_rate_long_subgroup,
     file="simulated_data_subgroup.RData")
```

Code for simulation study

```
####################################
#
# Simulate network meta-analysis
# R. Version 4.0.5 (2021-03-31)"
# "Shake and Throw"
#
########################################

# set working directory
setwd("~/simulation_study")

# load packages
library(dplyr)
library(tictoc)
```

```
# Get distributions
set.seed(1234)
pc.th <- 0.4
ptneg.th <- 0.45
nc <- 200
nt <- 200
repetitions <- 1000
tau2 <- c(0.001, 0.5, 1, 2)^2
ptpos.th <- c(0.7,0.5,0.45,0.35)
n_study <- c(2,5,10,20)

count <-0
load("~/simulated_data.RData")

results_response_rate<- list()

for(u in 1:length(ptpos.th)){

  # set logodds
  logodds<-function(x){
    log(x/(1-x)) }

  odds<-function(x){
    x/(1-x)
  }

 results_nstudy <- list()
 for (c in 1:length(n_study)){
   # Setting A
   cov03 <- c(seq(0.3,0.7, length = n_study[c]))
   cov04 <- c(seq(0.4,0.5, length = n_study[c]))
   # Setting B
   # cov03 <- c(seq(0.3,0.7, length = n_study[c]),1,1)
   # cov04 <- c(seq(0.4,0.5, length = n_study[c]),1,1)
   cov<-c()

   results_cov <- list()

   for (d in 1:2){
     if (d==1){
       cov <- cov03
      }else {
       cov <- cov04
      }
```

```
 results_tau <- list()

for(v in 1:length(tau2)){

   require(doSNOW)
   require(foreach)
   source("cl.R")

   cl1 <- makeCluster(cl) # set cernels

   registerDoSNOW(cl1)

   results_data <- list()

   results_data <- foreach(j = 1:repetitions) %dopar% {
     library(R2jags)
     library(dplyr)

     df_ad <- purrr::pluck(simulated_data,u,c,d,v,j)
     df_ad <- df_ad_overall %>%
       mutate(treat_new = ifelse(treat == 4,3,treat)) #E_+ is 3
     outcome.ad <- df_ad$outcomeaddup
     a.treat <- df_ad$treat_new
     n <- df_ad$narmaddup
     a.cov <- df_ad$cov_giv
     a.study <- df_ad$study.id
     # Setting A
     a.base <- c(rep(c(rep(1,(n_study[c]+3)),3),2))
     n.agg.trials <- n_study[c] + 4
     n.agg.arms <- 2*(n_study[c] + 4)

     # Setting B
     a.base <- c(rep(c(rep(1,((n_study[c] + 2) + 3)),3),2))
     n.agg.trials <- (n_study[c] + 2) + 4
     n.agg.arms <- 2*((n_study[c] + 2) + 4)


     data.sim <- list(max.treat = 3,
                      n.agg.trials = n.agg.trials,
                      n.agg.arms = n.agg.arms,
                      a.study = a.study,
                      a.treat = a.treat,
                      outcome.ad = outcome.ad,
                      n = n,
                      a.base = a.base,
```

```
                    cov = a.cov)


inits <- list(list(d = c(NA,0,0),
                   # Setting A
                   delta = c(rep(0,(2*(n_study[c]+4)))),
                   mu.ad = c(rep(0,n_study[c]+4)),
                   # Setting B
                   #delta = c(rep(0,(2*(n_study[c]+2+4)))),
                   #mu.ad = c(rep(0,n_study[c]+2+4)),
                   tau = 1),
                   #Alternative heterogeneity prior
                   # tau = 0.15),
              list(d = c(NA,0.1,0),
                   # Setting A
                   delta = c(0,0,0.25,rep(0,(2*(n_study[c]+4)
                       )-3)),
                   mu.ad = c(0,0.25,rep(0,(n_study[c]-2+4))),
                   # Setting B
                   # delta = c(0,0,0.25,rep(0,(2*(n_study[c
                       ]+2+4))-3)),
                   # mu.ad = c(0,0.25,rep(0,(n_study[c
                       ]+2-2+4))),
                   tau = 0.80),
                   #Alternative heterogeneity prior
                   #tau = 0.3),
              list(d = c(NA,0,0.1),
                   #Setting A
                   delta = c(0,0.25,rep(0,(2*(n_study[c]+4))
                       -2)),
                   mu.ad = c(0.25,rep(0,(n_study[c]-1+4))),
                   # Setting B
                   # delta = c(0,0.25,rep(0,(2*(n_study[c
                       ]+2+4))-2)),
                   # mu.ad = c(0.25,rep(0,(n_study[c]+2-1+4))
                       ),
                   tau = 0.90))
                   #Alternative heterogeneity prior
                   #tau = 0.45))

# Set parameter to monitor
parameter <- c("d[2]","d[3]","tau.sq",
               "lor[2,3]","lor[3,2]",
               "lor[2,3]","lor[1,2]",
               "lor[1,3]")
```

```
          set.seed(1234)
          tryCatch({
          overall.jags <- jags(data.sim, inits , parameter ,
                               # replace here with fitting model
                                  model.file="model.txt",
                                  n.chains = 3, n.iter = 50000,
                                  n.burnin = 20000, n.thin = 2,
                                  DIC = FALSE)

          jags.sum <- overall.jags$BUGSoutput$summary
          jags.sum}, error = function(e){NULL})
        }

        stopCluster(cl1)

        results_tau[[v]] <- results_data
      }

      results_cov[[d]] <- results_tau
    }

    results_nstudy[[c]] <- results_cov
  }
  save(results_cov, file = paste0("tau2",v,"_cov",d,"_nstudy",c,"
     ptpos_",u,"_results.RData"))
  results_response_rate[[u]] <- results_nstudy
}
save(results_response_rate, file = paste0("u_",u,"v_",v,"_c",c,"_d",d
   ,"results_response_rate.RData"))
```

Model functions

In the following the code of the different models used in the simulation study and in the application are displayed:

Naive, subgroup, stand-alone model

```
model{
  for(i in 1:n.agg.arms){
       outcome.ad[i] ~ dbin(pa[i],n[i])
       logit(pa[i]) <- mu.ad[a.study[i]] + delta[i]*
            (1-equals(a.treat[i],a.base[i]))
       delta[i] ~ dnorm(md.ad[i],prec)
       md.ad[i] <- d[a.treat[i]] - d[a.base[i]] }
```

```
    # Prior
     for(j in 1:n.agg.trials){
         mu.ad[j] ~ dnorm(0,1.0E-6)}

    # Vague priors for basic paramters
     d[1] <- 0
     for(k in 2:max.treat){
        d[k] ~ dnorm(0,1.0E-6)}

    # pairwise odds ratios
       for(c in 1:3){
         for(k in 2:3){
           lor[c,k] <- (d[k] - d[c])}}

     # Heterogeneity prior
      # Uniform prior
        tau ~ dunif(0,2)T(0,)
      # Alternative half-normal prior scale 0.5
       #tau ~ dnorm(0,4)T(0,)

        tau.sq <- tau*tau
        prec <- 1/(tau.sq)
         }
```

Regression model

```
model{
  for(i in 1:n.agg.arms){
     outcome.ad[i] ~ dbin(pa[i],n[i])
     logit(pa[i]) <- mu.ad[a.study[i]] + delta[i]*
          (1-equals(a.treat[i],a.base[i]))
     delta[i]~dnorm(md.ad[i],prec)
     md.ad[i] <- d[a.treat[i]] - d[a.base[i]]+
            (bb[a.treat[i]]-bb[a.base[i]])*cov[i]

   # Prior
    for(j in 1:n.agg.trials){
      mu.ad[j] ~ dnorm(0,1.0E-6)}

   # Vague priors for basic paramters
     bb[1] <- 0
      d[1] <- 0
     for(k in 2:max.treat){
       bb[k] ~ dnorm(m.betab,prec.betab)
```

```
      d[k] ~ dnorm(0,1.0E-6) }

   for(k in 1:max.treat){
      dz[k] <- d[k] + bb[k] }

  # pairwise odds ratios
   for (c in 1:3){
   for (k in 2:3){
    lorz[c,k] <- (dz[k]-dz[c])}}

   # Vague priors for random effects
    m.betab ~ dnorm(0,1.0E-2)
    tau.betab ~ dunif(0,2)
    tau.sq.betab <- (tau.betab*tau.betab)
    prec.betab <- 1/(tau.sq.betab)



   # Uniform prior
    tau ~ dunif(0,2)
   # Alternative heterogeneity prior
   # Half-normal with scale 0.5
    #tau ~ dnorm(0,4)T(0,)
    tau.sq <- tau*tau
    prec <- 1/(tau.sq)}
```

Enriching-through-weighting model with covariate

```
model{
  for(i in 1:n.agg.arms){
     outcome.ad[i] ~ dbin(pa[i],n[i])
     logit(pa[i]) <- mu.ad[a.study[i]] + delta[i]*
          (1-equals(a.treat[i],a.base[i]))
     delta[i] ~ dnorm(md.ad[i],prec*cov[i])
     md.ad[i] <- d[a.treat[i]]-d[a.base[i]]}

  # Prior
   for(j in 1:n.agg.trials){
     mu.ad[j] ~ dnorm(0,1.0E-6)}

  # Vague priors for basic paramters
   d[1] <- 0
   for(k in 2:max.treat){
      d[k] ~ dnorm(0,1.0E-6)}

  # pairwise odds ratios
```

```
    for(c in 1:3){
    fo (k in 2:3){
      lor[c,k] <- (d[k] - d[c])}}

    # Vague priors for random effects
    # Uniform prior
     tau ~ dunif(0,2)
    # Alternative heterogeneity prior
    # Half-normal with scale 0.5
     #tau ~ dnorm(0,4)T(0,)
     tau.sq <- tau*tau
     prec <- 1/(tau.sq) }
```

Enriching-through-weighting model with weights

```
model{
  for(i in 1:n.agg.arms){
     outcome.ad[i] ~ dbin(pa[i],n[i])
     logit(pa[i]) <- mu.ad[a.study[i]] + delta[i]*
          (1-equals(a.treat[i],a.base[i]))
     ind[i] <- ifelse(cov[i] == 1,prec,prec*ww)
     delta[i] ~ dnorm(md.ad[i],ind[i])
     md.ad[i] <- d[a.treat[i]]-d[a.base[i]]}

   # Prior
    for(j in 1:n.agg.trials){
     mu.ad[j] ~ dnorm(0,1.0E-6)}

   # Vague priors for basic paramters
    d[1] <- 0
    for(k in 2:max.treat){
       d[k] ~ dnorm(0,1.0E-6)}

   # pairwise odds ratios
    for (c in 1:3){
    for (k in 2:3){
      lor[c,k] <-(d[k]-d[c])}}

    # Vague priors for random effects
    # Uniform prior
     tau ~ dunif(0,2)
    # Alternative heterogeneity prior
    # Half-normal with scale 0.5
     #tau ~ dnorm(0,4)T(0,)
     tau.sq <- tau*tau
```

```
    prec <- 1/(tau.sq)

  # Weighting factor
  ww ~ dunif(0, 0.3)

  #Alternative weighting factors
  #ww ~ dunif(0.3, 0.7)
  #ww ~ dunif(0.7, 1)  }
```

Summary of simulation results

```
#########################################################################
#
#    Summarize results for different scenarios compared to fix values
#
#########################################################################
# Set working directory
setwd("~/simulation_results")

# Packages
library(tidyverse)

# Create data frame which need to be filled with results of 1000
    repetitions
x <- c("scenarios", "covariate", "studies","tau_2","pt.pos",
       "tau","bias_d2","bias_d3", "bias_d2d3",
       "rel_bias_d2","rel_bias_d3", "rel_bias_d2d3",
       "rmse_d2","rmse_d3","rmse_d2d3","d2d3","d2","d3",
       "mae_d2","mae_d3","mae_d2d3",
       "coverage.d2","coverage.d3","coverage.d2d3",
       "width_CI.d2","width_CI.d3","width_CI.d2d3",
       "TE_significant.d2","TE_significant.d3","TE_significant.d2d3")

summary_results<- data.frame(matrix(ncol = length(x),
                                    nrow = 96))
colnames(summary_results) <- x
levels(summary_results$covariate) = c("0.3-0.7",
                                      "0.4-0.5")

repetitions <- 1000
tau2 <- c(0.001, 0.5, 1, 2)
ptpos.th <- c(0.7,0.5,0.45,0.35)
n_study <- c(2,5,10,20)
pc.th <- 0.4
ptneg.th <- 0.45
```

```
dSE <- c(0.54, 1.39, 1.59, 2)
dCS <- c(1.25, 0.41, 0.2, -0.21)
dCE <- 1.79
i = 1

#define Odds Ratio
OR<-function(x,y){
  (x/(1-x))/(y/(1-y))
}

#### regression ----
load("~/u_4v_4_c4_d2results_regression.RData")
for (u in 1:4){
  for (c in 1:length(n_study)){
    # Setting A
    cov03 <- seq(0.3,0.7, length = n_study[c])
    cov04 <- seq(0.4,0.5, length = n_study[c])
    # Setting B
    # cov03 <- c(seq(0.3,0.7, length = n_study[c]),1,1)
    # cov04 <- c(seq(0.4,0.5, length = n_study[c]),1,1)
    cov <- c()
    for (d in 1:2){
      if (d == 1){
        cov <- cov03}
      else {
        cov <- cov04}
      for(v in 1:length(tau2)){

        res_d2d3 <- c()
        res_tau <- c()

        coverage_d2d3 <- c()
        width_CI_d2d3 <- c()
        TE.significant.d2d3 <- c()

        for(l in 1:repetitions){
          res_overall <- purrr::pluck(results_response_rate,u,c,d,v,l
              )

          res_d2d3[l] <- res_overall["lorz[2,3]",1]
          res_tau[l] <- res_overall["tau.sq",1]

          coverage_d2d3[l] <- ifelse((dSE[u] < res_overall["lorz[2,3]
              ",3]) |
```

```
                                        (dSE[u] > res_overall["lorz
                                            [2,3]",7]),0,1)

          width_CI_d2d3[l] <- res_overall["lorz[2,3]",7] -
                                                res_overall["lorz[2,3]"
                                                    ,3]

          TE.significant.d2d3[l] <- ifelse((res_overall["lorz[2,3]"
              ,3]> 0|
                                                res_overall["lorz[2,3]"
                                                    ,]< 0),1,0)
        }

        summary_results[i,]<- summary_results[i,] %>%
          mutate(scenarios = "regression",
                 covariate = ifelse(d==1, "0.3-0.7",
                                    "0.4-0.5"),
                 # Setting A
                 studies = n_study[c],
                 # Setting B
                 # studies = n_study[c]+2,
                 tau_2 = tau2[v],
                 pt.pos = ptpos.th[u],
                 tau = mean(res_tau),
                 bias_d2d3 = mean(dSE[u]-res_d2d3),
                 rmse_d2d3 = sqrt(mean((dSE[u]-res_d2d3)^2)),
                 d2d3 = mean(res_d2d3),
                 coverage.d2d3 = mean(coverage_d2d3),
                 width_CI.d2d3 = mean(width_CI_d2d3),
                 TE_significant.d2d3 = mean(TE.significant.d2d3))

        i <- i+1
      }
    }
  }
}

# Enriching-through weigthing with cov
load("u_4v_4_c4_d2results_response_rate_enrich_cov.RData")
for (u in 1:4){
  for (c in 1:length(n_study)){
    #Setting A
    cov03 <- seq(0.3,0.7, length = n_study[c])
    cov04 <- seq(0.4,0.5, length = n_study[c])
    # Setting B
```

```
# cov03 <- c(seq(0.3,0.7, length = n_study[c]),1,1)
# cov04 <- c(seq(0.4,0.5, length = n_study[c]),1,1)
cov <- c()

for (d in 1:2){
  if (d == 1){
    cov <- cov03}
  else {
    cov <- cov04}
  for(v in 1:length(tau2)){
    res_d2d3 <- c()
    res_tau <- c()

    coverage_d2d3 <- c()
    width_CI_d2d3 <- c()
    TE.significant.d2d3 <- c()

    for (l in 1:repetitions){

      res_overall <- purrr::pluck(results_response_rate_enrich_
          cov,u,c,d,v,l)
      res_d2d3[l] <- res_overall[5,1] # lor([2,3])
      res_tau[l] <- res_overall[11,1]

      coverage_d2d3[l] <- ifelse((dSE[u] < res_overall[5,3]) |
                                   (dSE[u] > res_overall[5,7])
                                      ,0,1)

      width_CI_d2d3[l] <- res_overall[5,7] -res_overall[5,3]

      TE.significant.d2d3[l] <- ifelse((res_overall[5,3]> 0|
                                          res_overall[5,7]< 0)
                                         ,1,0)
    }

    summary_results[i,]<- summary_results[i,] %>%
      mutate(scenarios = "enrichment-through weighting with cov",
             covariate = ifelse(d == 1, "0.3-0.7",
                                "0.4-0.5"),
             studies = n_study[c],
             tau_2 = tau2[v],
             pt.pos = ptpos.th[u],

             tau = mean(res_tau),
             bias_d2d3 = mean(dSE[u]-res_d2d3),
```

```
                    rmse_d2d3 = sqrt(mean((dSE[u]-res_d2d3)^2)),
                    d2d3 = mean(res_d2d3),
                    coverage.d2d3 = mean(coverage_d2d3),
                    width_CI.d2d3 = mean(width_CI_d2d3),
                    TE_significant.d2d3 = mean(TE.significant.d2d3))
        i <- i + 1
      }
    }
  }
}

------# equivalent code as above for all other models ----------

## Results ----

save(summary_results, file="summary_results.RData")
```

Own publications

Partial results of this thesis were published in the following publications:

1. **Proctor, T.**, Jensen, K., Kieser, M. (2020). Integrated evaluation of targeted and non-targeted therapies in a network meta-analysis. *Biometrical Journal, 62(3)*, 777–789.

2. **Proctor, T.**, Zimmermann, S., Seide, S., Kieser, M. (2022): A comparison of methods for enriching network meta-analyses in the absence of individual patient data. *Research Synthesis Methods.* doi: 10.1002/jrsm.1568. Online ahead of print.

Publication 1 is based on the results from Chapter 3.3. Consequently, also the described methods in Chapter 2.3 -2.6, as well as some content of the introduction and discussion was described in this publication. My contribution to this publication was programming the simulation study, creating graphics, the description and interpretation of results, and preparation of the manuscript draft.

Publication 2 is based on the results from Chapter 3.4 Also the described methods in Chapter 2.3-2.9, the application in Chapter 2.8, the simulation study in Chapter 2.9, as well as some content of the introduction and discussion were described in this publication. Generating the data, programming the simulation study, creating graphics, the description and interpretation of results, and preparation of the manuscript draft were part of my work on the manuscript.

Further own publications

Kowalewski K. F., Garrow C. R., **Proctor T.**, Preukschas A. A., Friedrich M., Müller P. C., Kenngott H. G., Fischer L., Müller-Stich B. P., Nickel F. (2018). LapTrain: multi-modality training curriculum for laparoscopic cholecystectomy—results of a randomized controlled trial.*Surgical Endoscopy, 32(9)*,3830-3838.

Kowalewski K. F., Schmidt M. W., **Proctor T.**, Pohl M., Wennberg E., Karadza E., Romero P., Kenngott H. G., Mueller-Stich B. P., Nickel F. (2018). Skills in minimally invasive and open surgery show limited transferability to robotic surgery: results from a prospective study. *Surgical Endoscopy, 32(4)*, 1656-1667.

Kowalewski K.F, Hendrie J.D, Schmidt M.W, **Proctor T**, Paul S, Garrow C.R, Kenngott H, Mueller-Stich B.P, Nickel F (2017). Validation of the mobile serious game application Touch Surgery for cognitive training and assessment of laparoscopic cholecystectomy. *Surgical endoscopy, 31(10)*, 4058-4066.

Kowalewski KF, Hendrie JD, Schmidt MW, Garrow CR, Bruckner T, **Proctor T**, Paul S, Adigüzel D, Bodenstedt S, Erben A, Kenngott H, Erben Y, Speidel S, Müller-Stich BP, Nickel F.(2016) Development and validation of a sensor- and expert model-based training system for laparoscopic surgery: the iSurgeon. *Surgical Endoscopy, 31(5)*, 2155-2165.

Nikas IP, Seide S, **Proctor T**, Kleinaki Z, Kleinaki M, Reynolds JP.(2022). The Paris System for Reporting Urinary Cytology: A Meta-Analysis. *Journal of Personalized Medicine, 12(2)*, 170.

Nikas IP, **Proctor T**, Seide S, Chatziioannou SS, Reynolds JP, Ntourakis D.(2022). Diagnostic Performance of Pancreatic Cytology with the Papanicolaou Society of Cytopathology System: A Systematic Review, before Shifting into the Upcoming WHO International System. *International Journal of Molecular Sciences, 23(3)*, 1650.

Pinart M, Kranz J, Jensen K, **Proctor T**, Naber K, Kunath F, Wagenlehner F, Schmidt S (2017). Optimal dosage and duration of pivmecillinam treatment for uncomplicated lower urinary tract infections: a systematic review and meta-analysis. *International Journal of Infectious Diseases, 58*, 96-109.

Proctor T., Schumacher M. (2016): Analysing adverse events by time-to-event models: the CLEOPATRA study. *Pharmaceutical Statistics, 15(4)*, 306-314.

Sedaghat-Hamedani F., Kayvanpour E., Tugrul O. F., Lai A., Amr A., Haas J., **Proctor T.**, Ehlermann P., Jensen K., Katus H. A., Meder B. (2018). Clinical outcomes associated with sarcomere mutations in hypertrophic cardiomyopathy: a meta-

analysis on 7675 individuals. *Clinical Research in Cardiology, 107(1)*, 30-41

Streit M. R., Haeussler D., Bruckner T., **Proctor T.**, Innmann M. M., Merle C., Gotterbarm T., Weiss S. (2016). Early Migration Predicts Aseptic Loosening of Cementless Femoral Stems: A Long-term Study. *Clinical Orthopaedics & Related Research, 474(7)*, 1697-1706.

Unkel S, Amiri M, Benda N, Beyersmann J, Knoerzer D, Kupas K, Langer F, Leverkus F, Loos A, Ose C, **Proctor T**, Schmoor C, Schwenke C, Skipka G, Unnebrink K, Voss F, Friede T.(2019). On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. *Pharmaceutical Statistics, 18(2)*, 166-183

Zylla, M.M., Hillmann, H.A.K., **Proctor**, T. et al. Heart Vessels (2018). Use of the wearable cardioverter-defibrillator (WCD) and WCD-based remote rhythm monitoring in a real-life patient cohort. *Heart Vessels, 33(11)* 1390-1402.

Curriculum Vitae

Personal information

Name:	Tanja Proctor (née Zinßer)
Date of birth:	27.05.1988
Place of birth:	Freudenstadt

Education

Since 10/2015	Doctoral student at the University of Heidelberg
04/2013 – 09/2015	Study of mathematics at Albert-Ludwigs-Universität Freiburg, degree: Bachelor of Science
09/2012 – 02/2013	Study of mathematics and geography at Université Bordeaux
10/2009 – 09/2015	Study of mathematics and geography for 'Lehramt an Gymnasien', Albert-Ludwigs-Universität Freiburg, degree: 1. Staatsexamen
10/2008 – 09/2009	'Sonderschulpädagogik' at Pädagogische Hochschule Ludwigsburg
09/1999 – 06/2008	Kepler-Gymnasium Freudenstadt, degree: Abitur

Professional experience

since 10/2015	Research fellow at Institute of Medical Biometry , University hospital Heidelberg
09/2012 – 07/2015	Student assistant at Center of Medical Biometry and Medical Informatics, University hospital Freiburg

Acknowledgements

Acknowledgements

First of all, I would like to thank my supervisor Prof. Dr. Meinhard Kieser for giving me the opportunity to write this thesis at the Institute of Medical Biometry. I greatly appreciate the guidance, encouragement, constructive suggestions and continued support he has provided throughout my time of research and writing of this thesis.

I also would like to thank my late colleague and second supervisor Dr. Katrin Jensen, who initiated the whole project and contributed greatly to the development of this work. I miss you Katrin.

I also gratefully acknowledge the support for this work by the program „Methoden und Werkzeuge für die individualisierte Medizin" of the German Federal Ministry of Education and Research (BMBF) under grant 01EK1504.

Furthermore, I would like to thank my colleague Dr. Svenja Elisabeth Seide for her joint in-depth discussions and constructive critique. Her support has contributed greatly to improve this work.

In addition I thank my current and former colleagues at the Institute of Medical Biometry for creating a friendly atmosphere, for inspiring discussion, moral support and valuable feedback. Special thanks goes to Dr. Lorenz Uhlmann and Dr. Johannes Krisam for the time they invested to discuss my work.

I also would like to thank my family and family-in-law for their support in the last seven years, for always believing in me, making sure I had the freedom to work on it. A special thanks goes to my sister-in-law for her valuable feedback.

A final and special thank you goes to my husband Iain, who has supported me a lot, always believed in me and motivated me when it got difficult.

Eidesstaatliche Versicherung

1. Bei der eingereichten Dissertation zu dem Thema "Network meta-analysis of targeted therapies" handelt es sich um meine eigenständig erbrachte Leistung.
2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.
3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.
4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.
5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Ort und Datum Unterschrift