Received: 25 June 2023 | Revised: 4 April 2024 | Accepted: 22 June 2024

DOI: 10.1002/jrsm.1734

RESEARCH ARTICLE

Research Synthesis Methods WILEY

A comparison of two models for detecting inconsistency in network meta-analysis

Lu Qin[1] | **Shishun Zhao**[1] | **Wenlai Guo**[2] | **Tiejun Tong**[3] | **Ke Yang**[4]

[1]Center for Applied Statistical Research and College of Mathematics, Jilin University, Changchun, China

[2]Department of Hand Surgery, The Second Hospital of Jilin University, Changchun, China

[3]Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

[4]Department of Statistics and Data Science, Beijing University of Technology, Beijing, China

Correspondence
Ke Yang, Department of Statistics and Data Science, Beijing University of Technology, Beijing, China.
Email: yangke@bjut.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 12071176, 12071305, 12371294; General Research Fund, Grant/Award Numbers: HKBU12300123, HKBU12303421

Abstract

The application of network meta-analysis is becoming increasingly widespread, and for a successful implementation, it requires that the direct comparison result and the indirect comparison result should be consistent. Because of this, a proper detection of inconsistency is often a key issue in network meta-analysis as whether the results can be reliably used as a clinical guidance. Among the existing methods for detecting inconsistency, two commonly used models are the design-by-treatment interaction model and the side-splitting models. While the original side-splitting model was initially estimated using a Bayesian approach, in this context, we employ the frequentist approach. In this paper, we review these two types of models comprehensively as well as explore their relationship by treating the data structure of network meta-analysis as missing data and parameterizing the potential complete data for each model. Through both analytical and numerical studies, we verify that the side-splitting models are specific instances of the design-by-treatment interaction model, incorporating additional assumptions or under certain data structure. Moreover, the design-by-treatment interaction model exhibits robust performance across different data structures on inconsistency detection compared to the side-splitting models. Finally, as a practical guidance for inconsistency detection, we recommend utilizing the design-by-treatment interaction model when there is a lack of information about the potential location of inconsistency. By contrast, the side-splitting models can serve as a supplementary method especially when the number of studies in each design is small, enabling a comprehensive assessment of inconsistency from both global and local perspectives.

KEYWORDS

design-by-treatment interaction model, heterogeneity, inconsistency detection, network meta-analysis, side-splitting models

Highlights

What is already known

- The application of network meta-analysis is becoming increasingly widespread, and detecting inconsistency is an important issue. Among the existing methods for inconsistency detection, the design-by-treatment interaction model and the side-splitting models are most commonly used.

What is new

- By treating the data structure of network meta-analysis as missing data and parameterizing the potential complete data for the two types of models, we investigate the relationship between the design-by-treatment interaction model and the side-splitting models both analytically and numerically. We further provide some practical guidance for detecting inconsistency in network meta-analysis.

Potential impact for *Research Synthesis Methods* readers

- For researchers interested in network meta-analysis, the analytical results based on potential complete data proposed in this paper provide new research perspectives. For practitioners of network meta-analysis, this paper offers some important practical guidance for detecting inconsistency.

1 | INTRODUCTION

The traditional meta-analysis, which combines evidence from multiple independent studies, involves direct comparisons between two interventions. The fixed-effect model and the random-effects model are the most commonly used models in traditional meta-analysis. The fixed-effect model assumes a common effect size for different studies, providing a straightforward interpretation. In the contrary, the random-effects model assumes that the effect sizes from different studies follow a certain distribution, making it particularly suitable for handling heterogeneous studies, albeit with a more complex interpretation. In some practical applications, there may be insufficient direct comparisons between the two interventions of interest. In such cases, researchers may need to include indirect evidence between the two interventions through a common comparator. This type of comparison is known as indirect comparison, with the method proposed by Bucher et al.[1] being among the most widely used.

Network meta-analysis (NMA), also known as mixed treatment comparisons or multiple treatments meta-analysis, consists of both the direct and indirect comparisons to compare and rank multiple interventions simultaneously. It overcomes the limitations of the traditional pairwise meta-analyses by incorporating all available evidence into a general statistical framework.[2] It has been over 20 years since the development of NMA, and its utilization in clinical literature has been steadily growing.[3–11] For a comprehensive understanding of concepts, statistical models, and methods employed in NMA, there have been numerous reviews in the literature.[12–21] In particular, NMA can be performed within either a frequentist or a Bayesian framework.[14,22–27] To achieve dependable and understandable results from an NMA, it is crucial to meet the condition of exchangeability, which can be thought of as qualitative homogeneity.[28] This implies similarity[29] concerning effect modifiers across all treatment comparisons. Exchangeability, in turn, ensures consistency. Consistency indicates that the effect sizes within predefined subsets of the data are alike. Consequently, it becomes feasible to synthesize the effect sizes from both direct and indirect evidence. When there is a lack of consistency, the terms "incoherence"[30,31] and "inconsistence"[13,32] are often used, and we also note that there are quite a few inconsistency detection methods in the literature. They include, for example, the estimation based on linear mixed models,[30] the Bayesian hierarchical models,[32] the multidimensional scaling model,[33] the side-splitting models,[34,35] the two-stage model,[36] the design-by treatment interaction model,[37] and the graph-theoretical model.[38]

In this paper, we concentrate on two popular models for detecting inconsistency in NMA, namely the design-by-treatment interaction model and the side-splitting models. The former, proposed by Higgins et al.[37] and White et al.,[39] considers studies with the same

interventions as the same design. It detects inconsistency globally by introducing inconsistency factors between different designs. The latter, including the original side-splitting model proposed by Dias et al.[34] and the symmetric side-splitting model proposed by White,[35] detect local inconsistency between the direct and indirect evidence for a specific intervention comparison. Although the original side-splitting model was proposed in a Bayesian framework, we adopt the frequentist approach in this paper. These are both contrast-based models. Tu[40] rewrote them as arm-based models based on the generalized linear mixed models and pointed out that the side-splitting models are special cases of the design-by-treatment interaction model. This paper aims to more deeply explore the relationship between these two types of models from a missing data perspective based on the original contrast-based models.

The remainder of the paper is organized as follows. In Section 2, we provide a comprehensive review of the design-by-treatment interaction model and the side-splitting models, followed by three examples of NMA with varying data structures. And for each example, we further discuss the parameterization forms of the potential complete data and the corresponding parameterization forms of the observed data, which allows us to explore the characteristics of the two types of models and examine their relationship. In Section 3, we present some interesting findings from our numerical studies, in which we evaluate the performance of the design-by-treatment interaction model and the original side-splitting model in terms of the estimation accuracy of the heterogeneity and the empirical size and power of the inconsistency test. In Section 4, we provide an in-depth analysis of the current study, highlight potential areas for future research in detecting inconsistency in NMA, and offer practical guidance based on our analysis.

2 | METHODS

We consider a network including a total of T treatments. For treatments that are not involved in a study, the outcomes can be regarded as missing. In this paper, we refer to the observed data together with the missing outcomes as potential complete data. Following Higgins et al.,[37] we denote the "design" of a study by the set of treatments compared within the study. Let $J = A, B, \ldots$ index treatments; $d = 1, \ldots, D$ index designs; and $i = 1, \ldots, n_d$ index studies within the dth design. Without loss of generality, let A be the reference (baseline) treatment. Contrast-based methods are used to establish statistical models for the observed effect size y_{di}^{AJ}, which represents the observed contrast between treatment J ($J = B, C, \ldots$) and treatment A for the ith trial in the dth design. For a treatment contrast JJ' that does not involve the reference treatment A, it can be expressed as $y_{di}^{JJ'} = y_{di}^{AJ'} - y_{di}^{AJ}$.

2.1 | Design-by-treatment interaction model

The design-by-treatment interaction model proposed by Higgins et al.[37] and White et al.[39] constructs the statistical model for y_{di}^{AJ} as

$$y_{di}^{AJ} = \mu^{AJ} + \omega_d^{AJ} + \beta_{di}^{AJ} + \epsilon_{di}^{AJ}, \quad (1)$$

where μ^{AJ} is the fixed effect of treatment J relative to A for the whole network, ω_d^{AJ} are the design-by-treatment interaction parameters that reflect inconsistency, β_{di}^{AJ} stand for the heterogeneity within each design, and ϵ_{di}^{AJ} represent the within-study errors. Model (1) explains the variation of y_{di}^{AJ} between studies by three components, namely the design-level variation referring the inconsistency, the heterogeneity between studies within a design, and the random error.

The design-level effect sizes, denoted as $d_d^{AJ} = \mu^{AJ} + \omega_d^{AJ}$, represent the combination of the fixed effect and the design-by-treatment interactions. By the fact that $y_{di}^{JJ'} = y_{di}^{AJ'} - y_{di}^{AJ}$ for $J, J' \neq A$, the design-level effect size is $d_d^{JJ'} = d_d^{AJ'} - d_d^{AJ}$. In the context of model (1), the inconsistency is defined as the difference between d_d^{AJ} from different designs, which is also referred to as "design inconsistency."[37] Through the analysis of design inconsistency, we can examine both the direct and indirect comparisons of a treatment contrast within a specific loop.[37] For example, consider the treatment contrast between B and C in the loop ABC. It is important to note that, for any pair of distinct designs d and d'—regardless of whether they involve the treatments A, B, and C being of interest for the loop—we consider the existence of design-level effect sizes d_d^{AB}, d_d^{AC}, d_d^{BC}, $d_{d'}^{AB}$, $d_{d'}^{AC}$, and $d_{d'}^{BC}$. This perspective aligns with the concept of the potential complete data in our paper. The direct comparison d_d^{BC} is equal to the indirect comparison $d_d^{AC} - d_d^{AB}$ within the same design d, indicating loop consistency. However, the direct comparison d_d^{BC} in design d may differ from the indirect comparison $d_{d'}^{AC} - d_{d'}^{AB}$ based on design d', resulting in inconsistency for the loop ABC. The design-by-treatment interaction model (1) takes into account the inconsistency for all treatment contrasts from a global perspective, allowing for the evaluation of both loop inconsistency and design inconsistency.[37,39] It extends the approach proposed by Lu and Ades,[32] which only addresses loop inconsistency.

Following White et al.,[39] for the ith study in the dth design, we regard ω_d^{AJ} as fixed effects in this paper. When $\omega_d^{AJ}=0$ for all d and J, model (1) reduces to a consistency model. The degrees of heterogeneity $\left(\beta_{di}^{AB},\beta_{di}^{AC},...\right)^{\mathrm{T}}$ are assumed to be random effects and follow a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}=(0.5+0.5\mathbf{1}\mathbf{1}^{\mathrm{T}})\tau^2$, where $\mathbf{1}\mathbf{1}^{\mathrm{T}}$ is a matrix with all elements being one. We further denote the study-level effect sizes by $d_{di}^{AJ}=\mu^{AJ}+\omega_d^{AJ}+\beta_{di}^{AJ}$. The random errors $\left(\epsilon_{di}^{AB},\epsilon_{di}^{AC},...\right)^{\mathrm{T}}$ are assumed to follow a multivariate normal distribution with mean vector $\mathbf{0}$ and a known covariance $\mathbf{S}_{di}$. The parameters for the potential complete data are μ^{AJ}, ω_d^{AJ}, and τ^2 with $J=B,C,...$ and $d=1,...,D$. Due to missing data, it is not possible to estimate all the parameters using only the observed data. In Section 2.3, we will provide three examples of NMA with different data structures and present the parameterization forms for both the potential complete data and the corresponding observed data.

In the design-by-treatment interaction model, different reference treatments can lead to different parameterizations, resulting in different interpretations of the estimated parameters. However, this does not affect the number of inconsistency parameters, the estimated design-level effect sizes, and the test results for inconsistency. In this paper, we adopt the frequentist approach[39] to estimate the parameters and test for inconsistency using the Wald test. The Stata commands *network meta consistency*, *network meta inconsistency*, and *mvmeta* can be used to implement the estimation and test procedures.[35]

2.2 | Side(node)-splitting models

The original node-splitting model[34] was initially proposed within a Bayesian framework to assess inconsistency between direct evidence and indirect evidence for a specific treatment contrast. This model treats each treatment contrast as a "node," and focuses on detecting inconsistency locally. Later, White[35] introduced the term "side-splitting" model to describe this approach since treatment contrasts are represented as sides in the network diagram, and moreover, he proposed a frequentist estimation approach for this model. In this paper, we adopt the terminology used in White[35] and refer to this model as the original side-splitting model. The main idea of the original side-splitting model is dividing the treatment contrast of interest into two distinct components: the direct comparison and the indirect comparison. The effect size of the treatment contrast based on the direct evidence is derived from all direct comparisons, while the effect size based on the indirect evidence is obtained from NMA of the remaining evidence. The consistency assumption between the direct and indirect evidence is then assessed.

Without loss of generality, assume that the treatment contrast of interest is side AB and the reference treatment is A. To detect whether there is a difference between the direct comparison and the indirect comparison for treatment contrast AB, we divided the designs in the network into two groups, the first group comprises designs containing both treatments A and B, denoted as S_{AB}, and the remaining designs belong to the second group. Then the original side-splitting model is given as

$$y_{di}^{AJ}=\begin{cases}\mu^{AJ}+\omega\cdot I(J=B)+\beta_{di}^{\mathrm{AJ}}+\epsilon_{di}^{AJ} & \text{for } d\in S_{AB},\\ \mu^{AJ}+\beta_{di}^{AJ}+\epsilon_{di}^{AJ} & \text{for } d\notin S_{AB},\end{cases} \tag{2}$$

where $I(\cdot)$ represents the indicator function.

In model (2), there is a single inconsistency parameter ω to represent inconsistency between the treatment contrast AB. Therefore, it focuses on detecting local inconsistency. Similarly, let d_d^{AJ} represent the design-level effect sizes. For a design d in the set S_{AB}, the design-level effect size for the treatment contrast AB is given by $d_d^{AB}=\mu^{AB}+\omega$. Since d_d^{AB} can be directly estimated using the observed data y_{di}^{AB}, it indicates a direct evidence. On the other hand, for design $d'\notin S_{AB}$ where $y_{d'i}^{AB}$ is not observed, the design-level effect size for the treatment contrast AB is $d_{d'}^{AB}=\mu^{AB}$. In this case, since $y_{d'i}^{AB}$ cannot be directly estimated, $d_{d'}^{AB}$ indicates an indirect evidence.

In model (2), for $J\neq A,B$ and $d\in S_{AB}$, $d'\notin S_{AB}$, we observe that $d_d^{AJ}=d_{d'}^{AJ}$ and $d_d^{AB}\neq d_{d'}^{AB}$. Then based on the equation $d_d^{JJ'}=d_d^{AJ'}-d_d^{AJ}$ for the same design d, it can be derived that $d_d^{BJ}\neq d_{d'}^{BJ}$. In summary, model (2) implies the assumption that the design-level effect sizes $d_d^{AJ}=d_{d'}^{AJ}$ but $d_d^{BJ}\neq d_{d'}^{BJ}$. This indicates that for the treatment contrast AB, there exists an unfairness between treatment A and treatment B. As a consequence, the original side-splitting model is highly sensitive to the choice of the baseline treatment. To address this limitation of the model, White[35] proposed the symmetric side-splitting model. With A being the baseline treatment and AB being the treatment contrast of interest, the symmetric side-splitting model can be expressed as

$$y_{di}^{AJ}=\begin{cases}\mu^{AJ}+0.5\omega+0.5\omega\cdot I(J=B)+\beta_{di}^{AJ}+\epsilon_{di}^{AJ} & \text{for } d\in S_{AB},\\ \mu^{AJ}+\beta_{di}^{AJ}+\epsilon_{di}^{AJ} & \text{for } d\notin S_{AB}.\end{cases} \tag{3}$$

In model (3), for $d\in S_{AB}$ and $d'\notin S_{AB}$, an inconsistency parameter ω is introduced to capture the difference between d_d^{AB} and $d_{d'}^{AB}$, represented as $d_d^{AB}=d_{d'}^{AB}+\omega$.

TABLE 1 Design-level effect sizes of the original side-splitting model on mixture data.

Side	Design	Potential complete data		Observed data		
		B versus *A*	*C* versus *A*	*B* versus *A*	*C* versus *A*	*C* versus *B*
AB	*AB*	$\mu^{AB}+\omega$	μ^{AC}	$\mu^{AB}+\omega$		
	ABC	$\mu^{AB}+\omega$	μ^{AC}	$\mu^{AB}+\omega$	μ^{AC}	
	AC	μ^{AB}	μ^{AC}		μ^{AC}	
	BC	μ^{AB}	μ^{AC}			$\mu^{AC}-\mu^{AB}$
AC	*AB*	μ^{AB}	μ^{AC}	μ^{AB}		
	ABC	μ^{AB}	$\mu^{AC}+\omega$	μ^{AB}	$\mu^{AC}+\omega$	
	AC	μ^{AB}	$\mu^{AC}+\omega$		$\mu^{AC}+\omega$	
	BC	μ^{AB}	μ^{AC}			$\mu^{AC}-\mu^{AB}$
BC	*AB*	μ^{AB}	μ^{AC}	μ^{AB}		
	ABC	μ^{AB}	$\mu^{AC}+\omega$	μ^{AB}	$\mu^{AC}+\omega$	
	AC	μ^{AB}	μ^{AC}		μ^{AC}	
	BC	μ^{AB}	$\mu^{AC}+\omega$			$\mu^{AC}-\mu^{AB}+\omega$

Additionally, for $J \neq A,B$, model (3) distributes the inconsistency parameter ω equally to the treatment contrasts AJ and BJ, resulting in $d_d^{AJ} = d_{d'}^{AJ} + 0.5\omega$ and $d_d^{BJ} = d_{d'}^{BJ} - 0.5\omega$. This equal distribution ensures a balanced allocation of the inconsistency parameter across relevant treatment contrasts.

In this paper, we employ a frequentist framework to estimate the parameters and test for inconsistency in the two side-splitting models. We assume that the inconsistency parameter ω is fixed, and that the degree of heterogeneity β_{di}^{AJ} and the random error ϵ_{di}^{AJ} are the same as those in the design-by-treatment interaction model. To carry out the heterogeneity estimation and the inconsistency test, we utilize the Stata commands *network sidesplit all nosymmetric* and *network sidesplit all*.

2.3 | Examples: Parameterization of the two types of models

To further illustrate the two types of models as well as their relationship, we provide three examples. These examples cover two types of NMA data structures, with each structure considering three treatments: *A*, *B*, and *C*. The first data structure is referred to as mixture data, which consists of four designs: *AB*, *AC*, *BC*, and *ABC*. The second data structure is referred to as two-arm data, which includes *AB*, *AC*, and *BC* designs. In practical applications, mixture data is more commonly encountered, while the networks with only two-arm trials are relatively rare.[41,42] In each example, we present the parameterization of the design-level effect sizes for both the potential complete data and the observed data. The parameterization of the potential complete data helps us understand the characteristics of the models, while the parameterization of the observed data enables us to figure out which parameters can be estimated and how they are estimated.

2.3.1 | Example I: The original side-splitting model on mixture data

In this example, our goal is to illustrate the characteristics of the side-splitting models, which detect inconsistency locally. We consider three treatments, and separately analyze the three sides: *AB*, *AC*, and *BC*. For the sake of simplicity, we focus on the original side-splitting model in this illustration. Similar results can be obtained for the symmetric side-splitting model.

For the mixture data scenario, Table 1 displays the parameters of the design-level effect sizes for both the potential complete data (based on model (2)) and the observed data. Each row in the table corresponds to a specific side that is being split, with the first treatment in each side assumed to be the baseline treatment. According to the parameterization of the observed data, there are three assumptions for the model splitting side *AB*. Firstly, the design-level effect size for the treatment comparison *AB* remains consistent between designs *AB* and *ABC*. Secondly, the design-level effect size for the treatment comparison *AC* also remains consistent between designs *AB* and *ABC*. Lastly, the design-level effect size for the treatment comparison *BC* in design *BC* may

TABLE 2 Design-level effect sizes of the design-by-treatment interaction model and the side-splitting models on mixture data.

Model	Design	Potential complete data		Observed data		
		B* versus *A	***C* versus *A***	***B* versus *A***	***C* versus *A***	***C* versus *B***
Model 1	*AB*	$\mu^{AB}+\omega_1^{AB}$	$\mu^{AC}+\omega_1^{AC}$	$\tilde{\mu}^{AB}$		
	ABC	$\mu^{AB}+\omega_2^{AB}$	$\mu^{AC}+\omega_2^{AC}$	$\tilde{\mu}^{AB}+\tilde{\omega}_1$	$\tilde{\mu}^{AC}$	
	AC	$\mu^{AB}+\omega_3^{AB}$	$\mu^{AC}+\omega_3^{AC}$		$\tilde{\mu}^{AC}+\tilde{\omega}_2$	
	BC	$\mu^{AB}+\omega_4^{AB}$	$\mu^{AC}+\omega_4^{AC}$			$\tilde{\mu}^{AC}-\tilde{\mu}^{AB}+\tilde{\omega}_3$
Model 2	Split side *AB*					
	AB	$\mu^{AB}+\omega$	μ^{AC}	$\mu^{AB}+\omega$		
	ABC	$\mu^{AB}+\omega$	μ^{AC}	$\mu^{AB}+\omega$	μ^{AC}	
	AC	μ^{AB}	μ^{AC}		μ^{AC}	
	BC	μ^{AB}	μ^{AC}			$\mu^{AC}-\mu^{AB}$
Model 3	Split side *AB*					
	AB	$\mu^{AB}+\omega$	$\mu^{AC}+0.5\omega$	$\mu^{AB}+\omega$		
	ABC	$\mu^{AB}+\omega$	$\mu^{AC}+0.5\omega$	$\mu^{AB}+\omega$	$\mu^{AC}+0.5\omega$	
	AC	μ^{AB}	μ^{AC}		μ^{AC}	
	BC	μ^{AB}	μ^{AC}			$\mu^{AC}-\mu^{AB}$

Note: Model 1, the design-by-treatment interaction model. Model 2, the original side-splitting model. Model 3, the symmetric side-splitting model.

exhibit inconsistency with the other designs. Interestingly, these assumptions coincide with those in the model splitting side *AC*. Consequently, in this scenario, the models for splitting sides *AB* and *AC* are identical. However, from the same perspective, the model splitting side *BC* differs.

2.3.2 | Example II: The design-by-treatment interaction model and the side-splitting models on mixture data

In this example, our objective is to illustrate the general relationship between the design-by-treatment interaction model and the side-splitting models using mixture data. We specifically focus on splitting side *AB* in the side-splitting models, while assuming that treatment *A* is the reference (baseline) treatment in all three models. Table 2 provides the parameters of the design-level effect sizes for both the potential complete data (based on models (1), (2), and (3)) and the observed data.

In the design-by-treatment interaction model, there are initially 10 parameters in the design-level effect sizes for the potential complete data. However, due to missingness and for the purpose of model identification, only five parameters can be estimated in the design-level effect sizes with the observed data. The corresponding relationships are as follows: $\tilde{\mu}^{AB}=\mu^{AB}+\omega_1^{AB}$, $\tilde{\mu}^{AC}=\mu^{AC}+\omega_2^{AC}$, $\tilde{\omega}_1=\omega_2^{AB}-\omega_1^{AB}$, $\tilde{\omega}_2=\omega_3^{AC}-\omega_2^{AC}$, and $\tilde{\omega}_3=\omega_4^{AC}-\omega_2^{AC}-\omega_4^{AB}+\omega_1^{AB}$.

Upon comparing the design-level effect sizes for the potential complete data of the three models, we observe that certain parameter restrictions can lead to the design-by-treatment interaction model reducing to either the original side-splitting model or the symmetric side-splitting model for a specific side. For instance, when $\omega_1^{AB}=\omega_2^{AB}$ and $\omega_3^{AB}=\omega_4^{AB}=0$, along with $\omega_1^{AC}=\omega_2^{AC}=\omega_3^{AC}=\omega_4^{AC}=0$, the design-by-treatment interaction model reduces to the original side-splitting model for the side *AB*. Similarly, when $\omega_1^{AB}=\omega_2^{AB}$, $\omega_3^{AB}=\omega_4^{AB}=0$, and $\omega_1^{AC}=\omega_2^{AC}$, $\omega_3^{AC}=\omega_4^{AC}=0$, the design-by-treatment interaction model reduces to the symmetric side-splitting model for the side *AB*. Analogously, similar parameter restrictions can lead to the reduction of the design-by-treatment interaction model to the original side-splitting model and the symmetric side-splitting model for sides *AC* and *BC*. From this perspective, it is evident that the side-splitting models can be considered as special cases of the design-by-treatment interaction model with additional assumptions.[37,40]

2.3.3 | Example III: The design-by-treatment interaction model and the side-splitting models on two-arm data

In this example, we present a special case of a network with three treatments and two-arm trials, where both the design-by-treatment interaction model and the

TABLE 3 Design-level effect sizes of the design-by-treatment interaction model and the side-splitting models on two-arm data.

Model	Design	Potential complete data		Observed data		
		B* versus *A	***C* versus *A***	***B* versus *A***	***C* versus *A***	***C* versus *B***
Model 1	*AB*	$\mu^{AB}+\omega_1^{AB}$	$\mu^{AC}+\omega_1^{AC}$	$\tilde{\mu}^{AB}$		
	AC	$\mu^{AB}+\omega_2^{AB}$	$\mu^{AC}+\omega_2^{AC}$		$\tilde{\mu}^{AC}$	
	BC	$\mu^{AB}+\omega_3^{AB}$	$\mu^{AC}+\omega_3^{AC}$			$\tilde{\mu}^{AC}-\tilde{\mu}^{AB}+\tilde{\omega}$
Model 2	*AB*	μ^{AB}	μ^{AC}	μ^{AB}		
	AC	μ^{AB}	μ^{AC}		μ^{AC}	
	BC	μ^{AB}	$\mu^{AC}+\omega$			$\mu^{AC}-\mu^{AB}+\omega$
Model 3	AB	μ^{AB}	μ^{AC}	μ^{AB}		
	AC	μ^{AB}	μ^{AC}		μ^{AC}	
	BC	$\mu^{AB}-0.5\omega$	$\mu^{AC}+0.5\omega$			$\mu^{AC}-\mu^{AB}+\omega$

Note: Model 1, the design-by-treatment interaction model. Model 2, the original side-splitting model. Model 3, the symmetric side-splitting model.

side-splitting models exhibit identical parameterizations. As shown in Table 3, the design-by-treatment interaction model involves eight parameters for the design-level effect sizes with the potential complete data. However, for the observed data structure, this number reduces to 3, including only 1 inconsistency parameter. Interestingly, in both the original side-splitting model and the symmetric side-splitting model, the allocation of the inconsistency parameter ω to treatments A and B cannot be discerned based on the estimable parameters with the observed data. Consequently, the two types of models become indistinguishable. In conclusion, for NMA that only includes two-arm trials with three treatments, the design-by-treatment interaction model and the two side-splitting models are equivalent to each other.

3 | RESULTS

This section compares the empirical performance of the design-by-treatment interaction model and the side-splitting models through simulations. For simplicity, the original side-splitting model is chosen as a representative of the side-splitting models. Given that clinical trials commonly utilize dichotomous data, our analysis focuses on dichotomous outcome data and measures the effect sizes using log odds ratio ($\ln \mathrm{OR}$). In our simulations, we systematically investigate networks consisting of 3, 4, and 7 treatments. More specifically, the networks with 3 treatments are used to scrutinize and validate the relationship between the design-by-treatment interaction model and the original side-splitting model. Networks involving 4 treatments are employed to assess how the number of studies affects the detection of inconsistency. Furthermore, the network with 7 treatments are crafted to more accurately reflect the inherent complexity in real-world network meta-analyses. We represent the treatments with capital letters $A, B, \ldots$. The investigation of these treatment quantities is illustrated across five network structures, as depicted in Figure 1. The simulated data for the five networks were generated based on the potential complete data mechanism using the following procedure:

a. *Set up the design-level effect sizes for the potential complete data*. We used treatment A as the reference (baseline) treatment and set the design-level effect sizes $\left(d_d^{AB}, d_d^{AC}, \ldots\right)^{\mathrm{T}}$ for the dth design based on Case 1 to Case 5. The inconsistency parameter ω varied in equally spaced increments from 0 to 1, with intervals of 0.1, representing different levels of inconsistency.
b. *Simulate the study-level effect sizes*. For the ith study in the dth design, we generated the study-level effect sizes $\left(d_{di}^{AB}, d_{di}^{AC}, \ldots\right)^{\mathrm{T}}$ from a normal distribution with mean vector $\left(d_d^{AB}, d_d^{AC}, \ldots\right)^{\mathrm{T}}$ and covariance matrix $\boldsymbol{\Sigma} = (0.5 + 0.5\mathbf{1}\mathbf{1}^{\mathrm{T}})\tau^2$, where $\mathbf{1}$ denotes a vector with all elements equal to one and $\tau^2 = 0.1$.
c. *Simulate the event rate of the reference (baseline) treatment and compute the event rate for each arm in each study*. For each study, we generated the event rate of treatment A from a uniform distribution: $p_{di}^A \sim U(0.25, 0.75)$. Further by the definition of the odds ratio, we computed the event rates for other treatment arms as follows:

$$p_{di}^J = \frac{p_{di}^A \exp\left(d_{di}^{AJ}\right)}{1 - p_{di}^A + p_{di}^A \exp\left(d_{di}^{AJ}\right)},$$

where J represents $B, C, \ldots$.

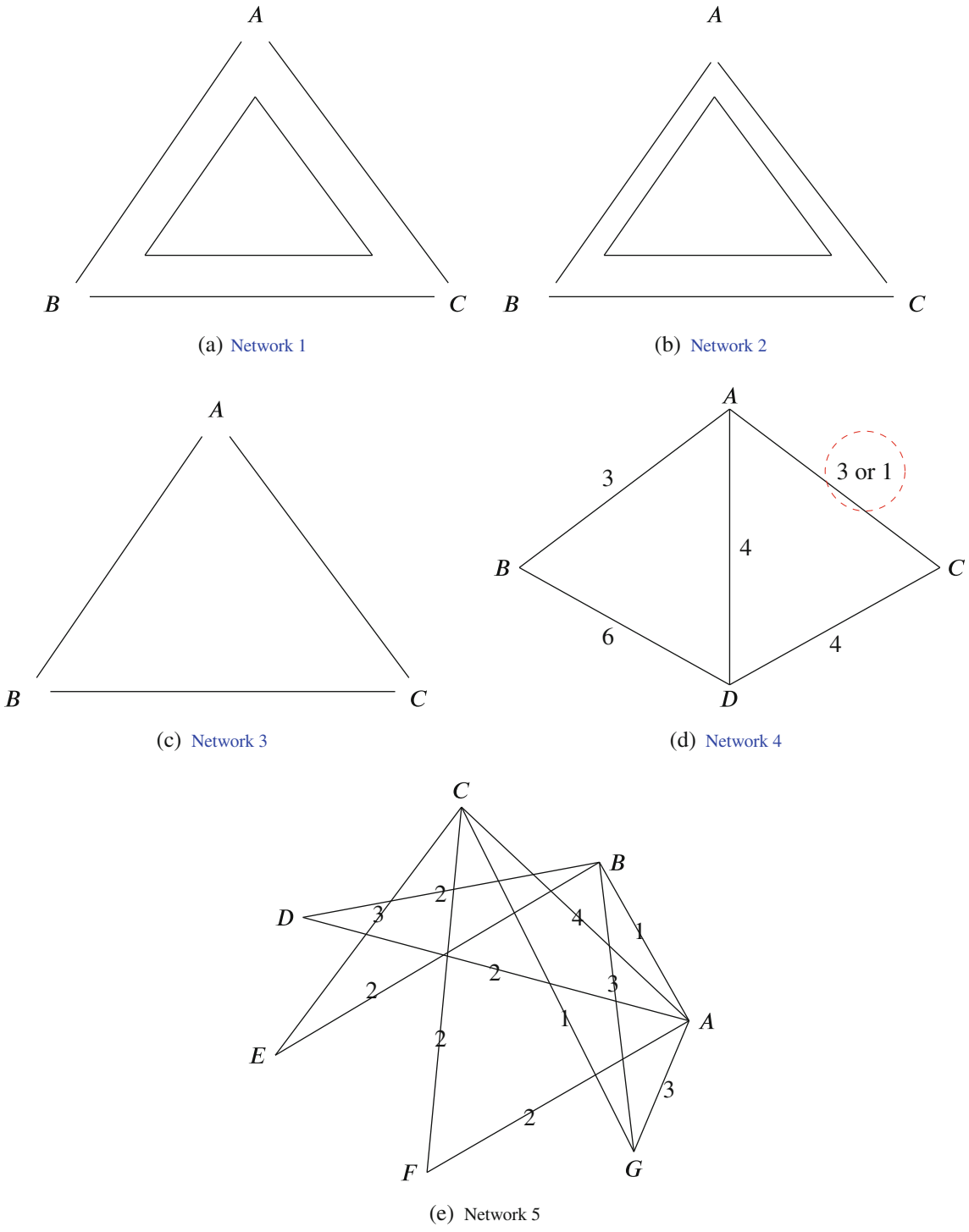


FIGURE 1 Structures of the networks. The number represents the number of studies for direct comparison between each pair of treatments. In Networks 1, 2, and 3, the number of studies for each design is the same, either 3 or 15.

d. *Generate the arm data for each study*. Considering balanced studies with an equal number of sample size for each arm, we generated the sample size for each study from a uniform distribution: $n_{di} \sim U(50,150)$. We further simulated the observed event number for each arm using a binomial distribution: $m_{di}^{J} \sim \text{Bin}\left(n_{di}, p_{di}^{J}\right)$, where J represents $A, B, \ldots$.

e. *In each study, omit the arm data for the treatments that are not included in the design*.

Based on the observed data, we can calculate the observed lnOR as the observed effect sizes, as well as the covariance of the observed effect sizes within each study. We then perform the parameter estimation and

TABLE 4 Design-level effect sizes for the networks with 3 treatments.

		Potential complete data		Observed data		
	Design	***B* versus *A***	***C* versus *A***	***B* versus *A***	***C* versus *A***	***C* versus *B***
Case 1	*AB*	0.25	0.5	0.25		
	ABC	$0.25 + \omega$	0.5	$0.25 + \omega$	0.5	
	AC	0.25	$0.5 + \omega$		$0.5 + \omega$	
	BC	0.25	$0.5 + \omega$			$0.25 + \omega$
Case 2	*AB*	0.25	0.5	0.25		
	ABC	0.25	0.5	0.25	0.5	
	AC	0.25	0.5		0.5	
	BC	0.25	$0.5 + \omega$			$0.25 + \omega$
Case 3	*AB*	0.25	0.5	0.25		
	AC	0.25	0.5		0.5	
	BC	0.25	$0.5 + \omega$			$0.25 + \omega$

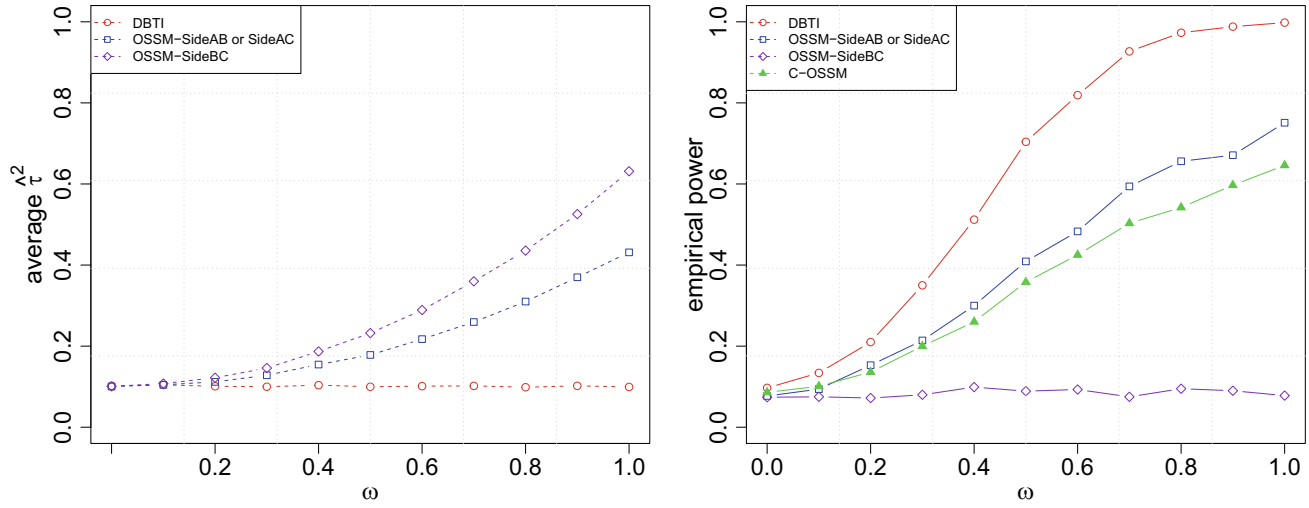
the inconsistency test with the significance level $\alpha = 0.05$ separately using the design-by-treatment interaction model (1) and the original side-splitting model (2). This allows us to obtain the estimate of the heterogeneity ($\widehat{\tau}^2$) and the inconsistency test results. In the original side-splitting model, we conduct a local inconsistency test by splitting all possible sides. Additionally, we perform a global inconsistency test by simultaneously evaluating all splits, incorporating the Bonferroni correction for multiple comparisons. To elucidate, in the original side-splitting model, if there are m distinct tests, the null hypothesis suggesting no inconsistency across all splits will be rejected if any individual test is rejected at the significance level of α/m. We repeat the process of data generation, estimation, and inconsistency test for $M = 1000$ times, considering each case and each value of ω, and then calculate the average $\widehat{\tau}^2$ and the probability of rejecting the null hypothesis in the inconsistency test for each model.

3.1 | Networks 1–3 with three treatments

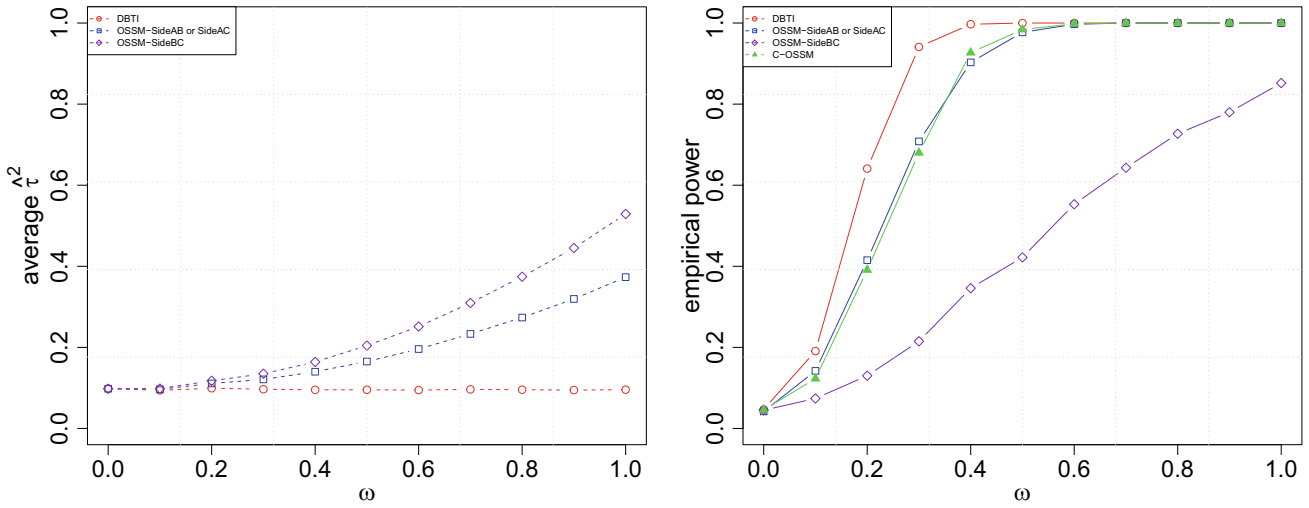
In this section, we focus on networks consisting of three treatments: *A*, *B*, and *C*, and we present three cases with distinct network structures. The network structures for the first two cases are illustrated in Figure 1a,b, while the network structure for the third case is displayed in Figure 1c. The main distinction between Case 1 and Case 2 lies in the consideration of design inconsistency that appears only in Case 1. To be more specific, Case 1 assumes that the effect sizes associated with *AB* differ between design *AB* and design *ABC*, and the effect sizes associated with *AC* differ between design *AC* and design *ABC*, whereas Case 2 regards these effect sizes as identical. The design-level effect sizes for the observed data are summarized in Table 4. We consider a balanced network with equal number of studies for each design. To be precise, we set the number of studies (NS) for each design as NS = 3 or NS = 15. The simulation results for the three cases are presented in Figures 2–4, respectively.

3.1.1 | Case 1

In this case, both the estimation and inconsistency test results are identical for the original side-splitting models for sides *AB* and *AC*, as demonstrated in Example I. Therefore, $m = 2$ for the Bonferroni correction in the original side-splitting model. From the left panels of Figure 2, it is evident that when $\omega = 0$, both the design-by-treatment interaction model and the original side-splitting models for all three sides accurately estimate τ^2. When the inconsistency parameter ω increases, the design-by-treatment interaction model continues to provide accurate estimates of τ^2, while the original side-splitting models tend to overestimate τ^2 for all three sides. In the right panels of Figure 2, we observe the behavior of the models in terms of the empirical type I error rate and the empirical power. As ω increases, the empirical power of the design-by-treatment interaction model increases more rapidly compared to the original side-splitting models, indicating its greater capability for detecting inconsistency. In contrast, the empirical power of the original side-splitting model for side *BC* remains around 0.05, suggesting its failure in inconsistency test. Comparing the right panels of Figure 2a,b, we find that



(a) Simulation results of Case 1 with NS=3.

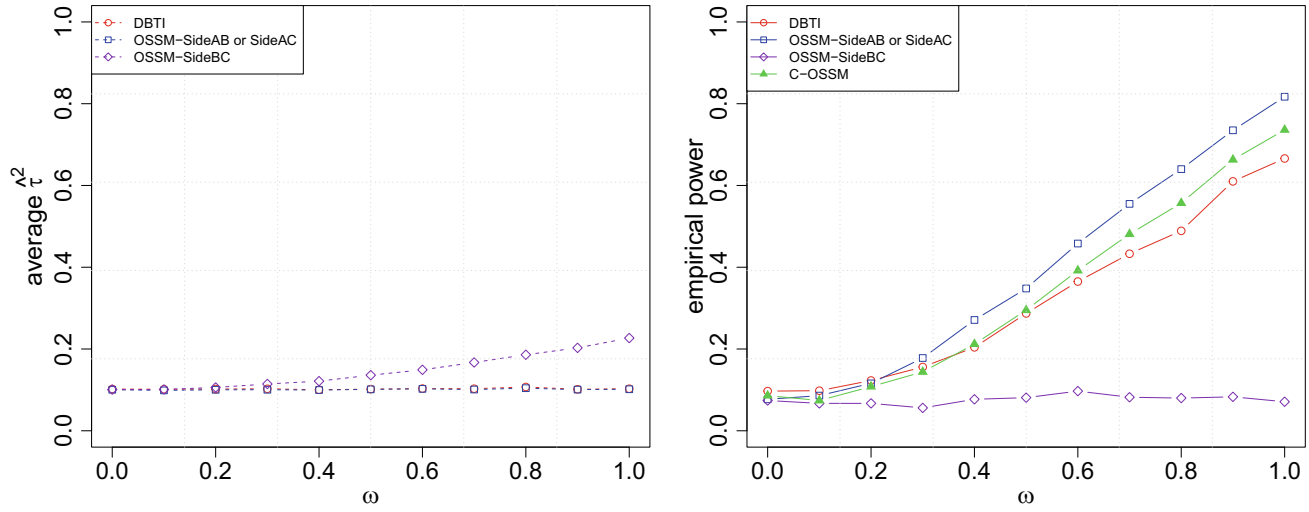


(b) Simulation results of Case 1 with NS=15.

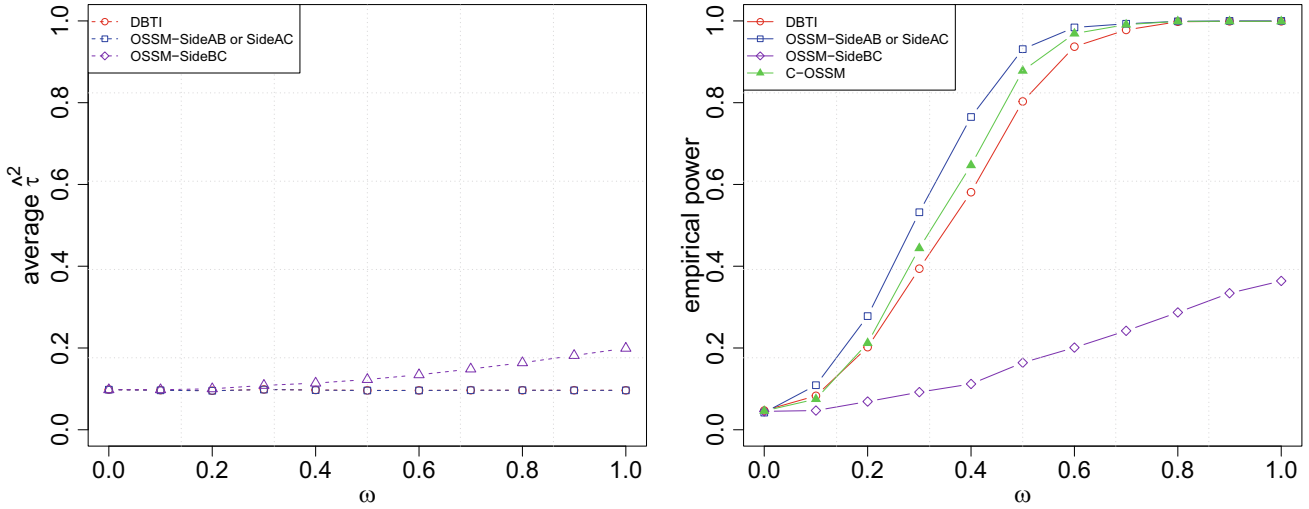
FIGURE 2 Simulation results of Case 1. "DBTI" represents the design-by-treatment interaction model, "OSSM" represents the original side-splitting model, and "C-OSSM" represents the Bonferroni correction of the original side-splitting model.

when the number of studies in each design is small, the design-by-treatment interaction model cannot control the type I error rate. Additionally, the empirical power of the global test for the original side-splitting models is lower than the design-by-treatment interaction model.

To further explain the observed results, we compare the true effect sizes at the design-level (presented in Table 4) in the simulation with the assumptions made by each model. Regarding the design-by-treatment interaction model, when comparing the model assumptions for the observed data in Table 2 with the true effect sizes in Table 4, we find that the design-by-treatment interaction model accurately captures the true design-level effect sizes by setting $\tilde{\mu}^{AB} = 0.25$, $\tilde{\mu}^{AC} = 0.5$, and $\tilde{\omega}_1 = \tilde{\omega}_2 = \tilde{\omega}_3 = \omega$. On the other hand, the assumptions made by the original side-splitting models for the three sides (Table 1) do not align with the simulation settings in Case 1. For instance, the original side-splitting model that splits side *AB* assumes that the design-level effect size for the treatment comparison *AB* is the same in the *AB* and *ABC* designs. However, in our simulation settings for Case 1, these effect sizes differ. This difference represents inconsistency between different designs, but the original side-splitting model estimates it as heterogeneity. The assumption for the treatment comparison *AC* follows a similar pattern. This difference leads to an overestimation of heterogeneity τ^2 and incomplete detection of inconsistency. Similar reasons apply to the original



(a) Simulation results of Case 2 with NS=3.

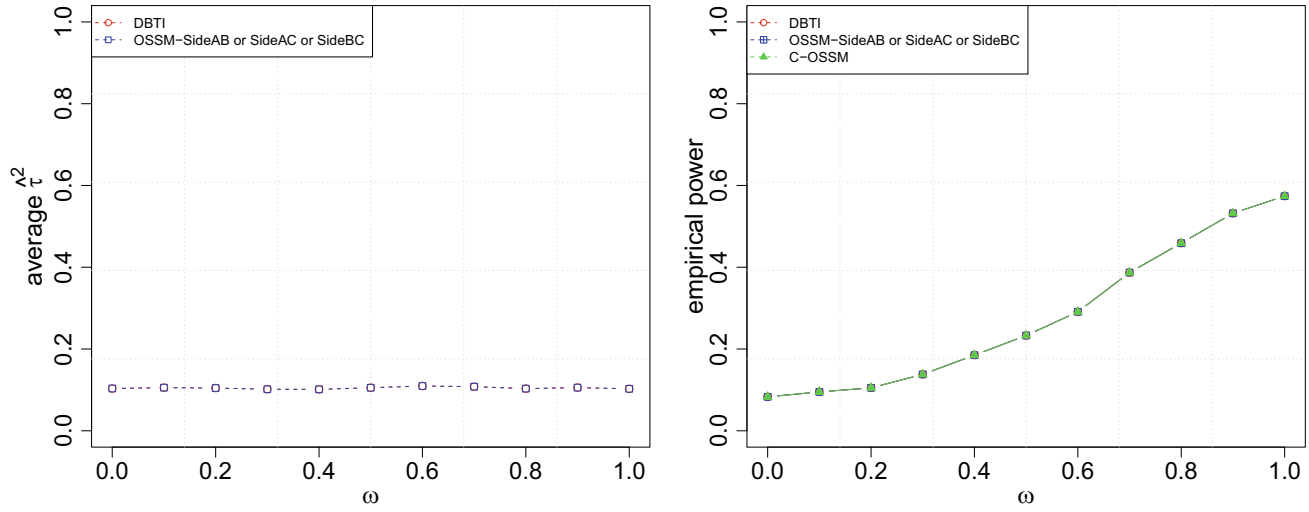


(b) Simulation results of Case 2 with NS=15.

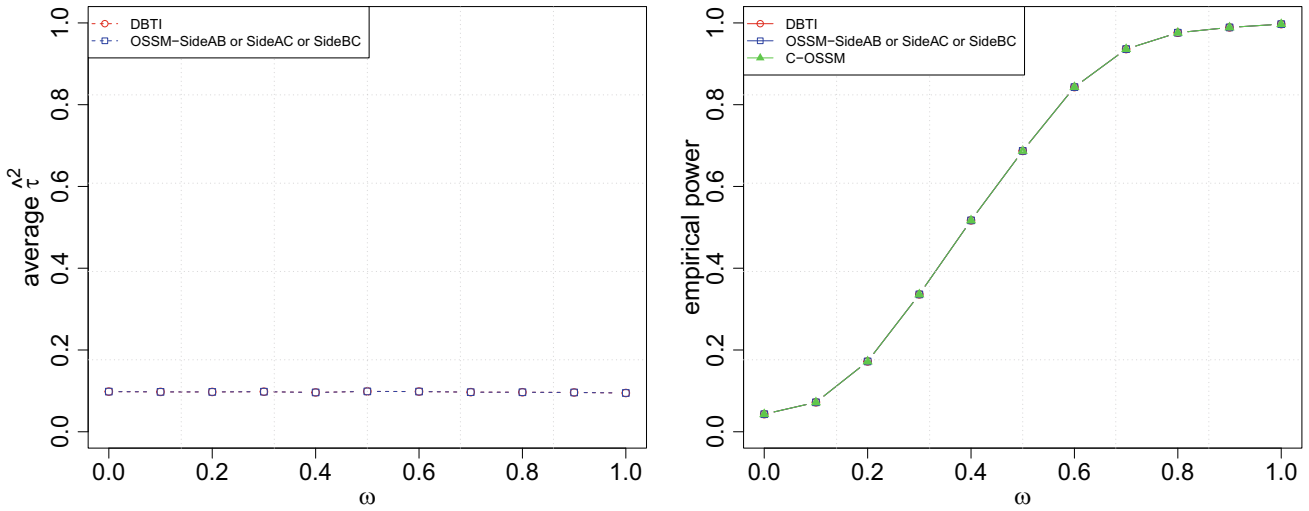
FIGURE 3 Simulation results of Case 2. "DBTI" represents the design-by-treatment interaction model, "OSSM" represents the original side-splitting model, and "C-OSSM" represents the Bonferroni correction of the original side-splitting model.

side-splitting models that splits side *AC* and *BC*. When the assumptions fail to capture the simulation settings, τ^2 cannot be accurately estimated, which consequently affects the power of inconsistency test. Moreover, the assumption made by the original side-splitting model that splits the side *BC* deviates significantly from the actual settings in Case 1, resulting in a complete failure of the inconsistency test.

In this case, there is only one inconsistency factor in the true effect sizes at the design-level. The design-by-treatment interaction model assumes three inconsistency factors, which is higher than the actual number. From the estimation of τ^2, it can be seen that although the model is redundant, it is still able to estimate the parameters correctly and provides a higher power in inconsistency test. On the other hand, the original side-splitting model assumes one inconsistency factor, but the assumptions in this model do not match the actual inconsistency factor, leading to inaccurate parameter estimation and the risk of complete failure in inconsistency test. Furthermore, due to the relatively small number of studies with NS = 3 compared to the three inconsistency factors in the design-by-treatment interaction model, the Wald test is unable to adequately control the type I error rate when $\omega = 0$. In contrast, when considering the same number of studies but with only one inconsistency factor,



(a) Simulation results of Case 3 with NS=3.



(b) Simulation results of Case 3 with NS=15.

FIGURE 4 Simulation results of Case 3. "DBTI" represents the design-by-treatment interaction model, "OSSM" represents the original side-splitting model, and "C-OSSM" represents the Bonferroni correction of the original side-splitting model.

the Wald test of the original side-splitting model demonstrates a better control over the empirical type I error rate. Lastly, when the number of studies for each design increases to NS = 15, the empirical type I error rates of all models approach their nominal levels.

3.1.2 | Case 2

In this case, both the estimation and inconsistency test results are identical for the original side-splitting models for sides *AB* and *AC*. Therefore, $m = 2$ for the Bonferroni correction in the original side-splitting model. From the left panels of Figure 3, we can observe that the design-by-treatment interaction model and the original side-splitting models that split sides *AB* and *AC* can accurately estimate the heterogeneity τ^2, while the original side-splitting model that splits side *BC* overestimates τ^2. From the right panels of Figure 3, we can see that the original side-splitting models that split sides *AB*, *AC*, and the Bonferroni correction have slightly higher power in inconsistency test compared to the design-by-treatment interaction model, while the original side-splitting model that splits side *BC* completely fails in inconsistency test.

The interpretation of the results for the design-by-treatment interaction model and the original

TABLE 5 Design-level effect sizes for the network with 4 treatments.

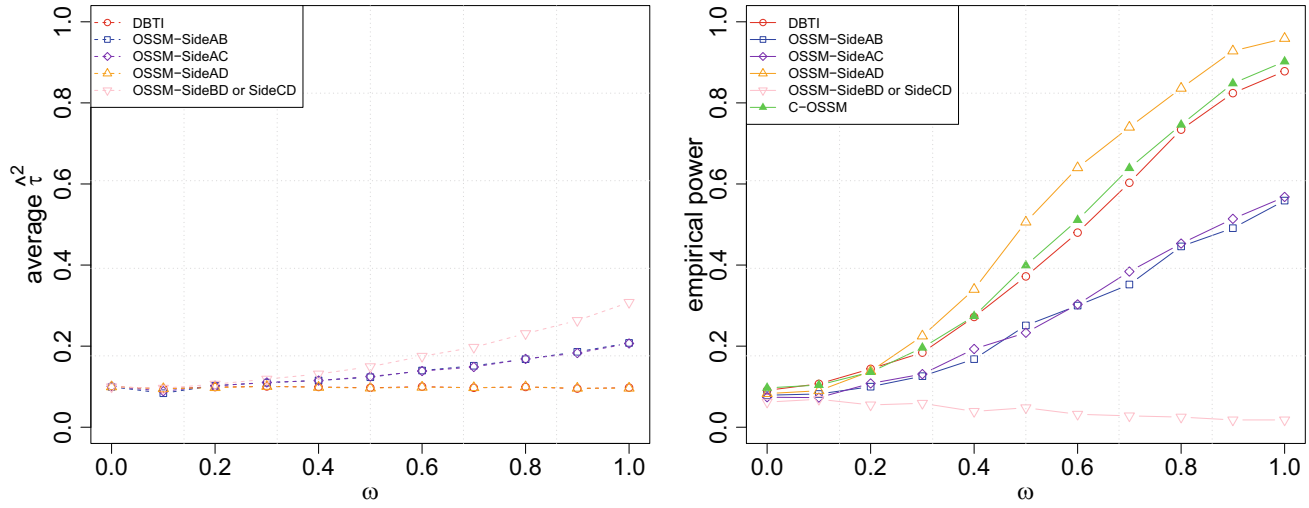
		Potential complete data			Observed data				
	Design	***B*** **versus** ***A***	***C*** **versus** ***A***	***D*** **versus** ***A***	***B*** **versus** ***A***	***C*** **versus** ***A***	***D*** **versus** ***A***	***D* versus *B***	***D* versus *C***
Design-level effect sizes									
Case 4	*ABD*	$0.25+\omega$	$0.25+\omega$	0.5	$0.25+\omega$		0.5		
	ACD	$0.25+\omega$	$0.25+\omega$	0.5		$0.25+\omega$	0.5		
	BD	0.25	0.25	0.5				0.25	
	CD	0.25	0.25	0.5					0.25
Parameterization									
Model 1	*ABD*	$\mu^{AB}+\omega_1^{AB}$	$\mu^{AC}+\omega_1^{AC}$	$\mu^{AD}+\omega_1^{AD}$	$\tilde{\mu}^{AB}$		$\tilde{\mu}^{AD}$		
	ACD	$\mu^{AB}+\omega_2^{AB}$	$\mu^{AC}+\omega_2^{AC}$	$\mu^{AD}+\omega_2^{AD}$		$\tilde{\mu}^{AC}$	$\tilde{\mu}^{AD}+\tilde{\omega}_1$		
	BD	$\mu^{AB}+\omega_3^{AB}$	$\mu^{AC}+\omega_3^{AC}$	$\mu^{AD}+\omega_3^{AD}$				$\tilde{\mu}^{AD}-\tilde{\mu}^{AB}+\tilde{\omega}_2$	
	CD	$\mu^{AB}+\omega_4^{AB}$	$\mu^{AC}+\omega_4^{AC}$	$\mu^{AD}+\omega_4^{AD}$					$\tilde{\mu}^{AD}-\tilde{\mu}^{AC}+\tilde{\omega}_3$
Model 2	Split side *AD*								
	ABD	μ^{AB}	μ^{AC}	$\mu^{AD}+\omega$	μ^{AB}		$\mu^{AD}+\omega$		
	ACD	μ^{AB}	μ^{AC}	$\mu^{AD}+\omega$		μ^{AC}	$\mu^{AD}+\omega$		
	BD	μ^{AB}	μ^{AC}	μ^{AD}				$\mu^{AD}-\mu^{AB}$	
	CD	μ^{AB}	μ^{AC}	μ^{AD}					$\mu^{AD}-\mu^{AC}$
Model 2	Split side *BD*								
	ABD	μ^{AB}	μ^{AC}	$\mu^{AD}+\omega$	μ^{AB}		$\mu^{AD}+\omega$		
	ACD	μ^{AB}	μ^{AC}	μ^{AD}		μ^{AC}	μ^{AD}		
	BD	μ^{AB}	μ^{AC}	$\mu^{AD}+\omega$				$\mu^{AD}-\mu^{AB}+\omega$	
	CD	μ^{AB}	μ^{AC}	μ^{AD}					$\mu^{AD}-\mu^{AC}$
Model 2	Split side *CD*								
	ABD	μ^{AB}	μ^{AC}	μ^{AD}	μ^{AB}		μ^{AD}		
	ACD	μ^{AB}	μ^{AC}	$\mu^{AD}+\omega$		μ^{AC}	$\mu^{AD}+\omega$		
	BD	μ^{AB}	μ^{AC}	μ^{AD}				$\mu^{AD}-\mu^{AB}$	
	CD	μ^{AB}	μ^{AC}	$\mu^{AD}+\omega$					$\mu^{AD}-\mu^{AC}+\omega$

Note: Model 1, the design-by-treatment interaction model. Model 2, the original side-splitting model.

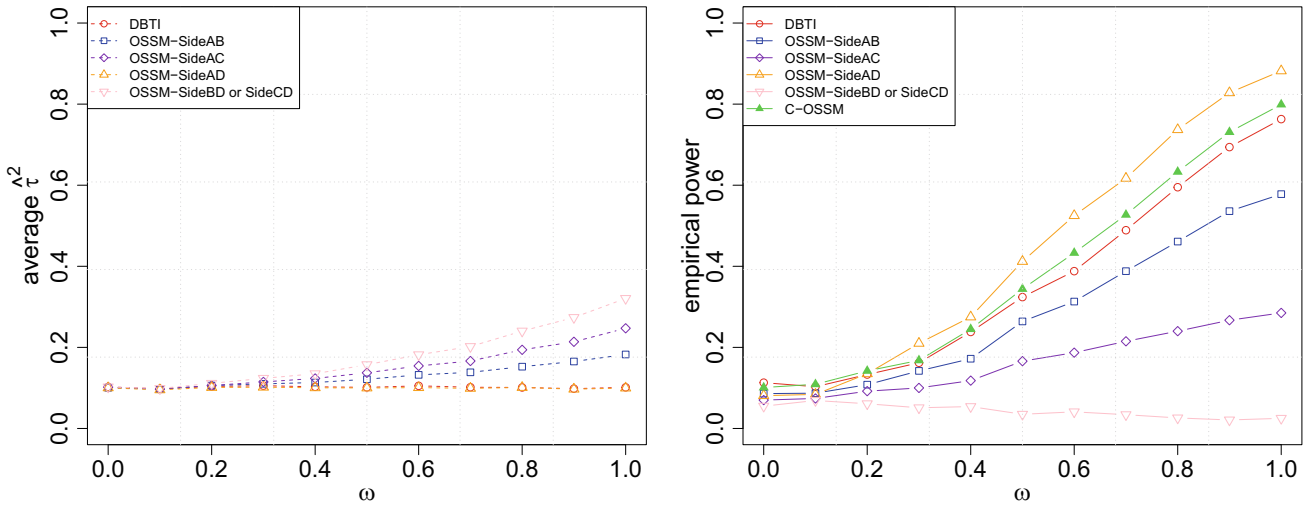
side-splitting model that splits side *BC* is similar to the previous case. For the original side-splitting model that splits side *AB*, by comparing the model assumptions for the observed data in Table 1 and the simulation settings for the design-level effect sizes in Table 4, we can see that the model assumptions align with the actual settings by setting $\mu^{AB}=0.25-\omega$ and $\mu^{AC}=0.5$. This model has only one inconsistency factor, which perfectly matches the true data. In this case, it exhibits higher power in inconsistency test compared to the design-by-treatment interaction model with redundant parameters. Similarly to Case 1, the design-by-treatment interaction model exhibits poor control over the empirical type I error rate when the number of studies for each design is only NS = 3. And as expected based on the large sample theory, the Wald test performs better when the number of studies increases to NS = 15. The same logic applies to the original side-splitting model that splits side *AC*.

3.1.3 | Case 3

In this case, we focus on networks with three interventions and two-arm trials. From Table 3, it is evident that the design-by-treatment interaction model and the original side-splitting approach (which splits side *AB*) are



(a) Simulation results of Case 4 with NS=3 (*ACD*).



(b) Simulation results of Case 4 with NS=1 (*ACD*).

FIGURE 5 Simulation results of Case 4. "DBTI" represents the design-by-treatment interaction model, "OSSM" represents the original side-splitting model, and "C-OSSM" represents the Bonferroni correction of the original side-splitting model.

equivalent and effectively capture the characteristics of the design-level effect sizes for Case 3 in Table 4. Although not explicitly shown, the original side-splitting models that split sides *AC* and *BC* also accurately depict the data. In this specific scenario, all four models' assumptions align with the actual simulation settings. The findings depicted in Figure 4a,b further validate that, in this case, the design-by-treatment interaction model and the original side-splitting models (splitting sides *AB*, *AC*, and *BC*) yield identical results. Thus, $m = 1$ for the Bonferroni correction in the original side-splitting model. Similarly to the previous cases, when NS = 15, the empirical type I error rates of all models approach their nominal levels. However, it is important to note that the data structure presented here represents a specific case, and real network data tends to be more intricate and complex.

3.2 | Network 4 with four treatments

In this section, we examine a network comprising four treatments: *A*, *B*, *C*, and *D*. The designs and design-level effect sizes in the network are detailed in Table 5. To investigate the impact of the number of studies on inconsistency detection, we consider two network structures. Both structures feature three studies for each design, with the exception of design *ACD*, which has three studies in

TABLE 6 Design-level effect sizes for the network with 7 treatments.

		Observed data									
Design	**Number of Studies**	***B* versus *A***	***C* versus *A***	***D* versus *A***	***F* versus *A***	***G* versus *A***	***D* versus *B***	***E* versus *B***	***G* versus *B***	***E* versus *C***	***F* versus *C***
ABD	1	0.1		0.3							
AC	3		0.2								
ACG	1		$0.2+\omega_1$			0.6					
AD	1			$0.3+\omega_2$							
AF	2				0.5						
AG	2					$0.6+\omega_3$					
BD	1						$0.2+\omega_4$				
BE	2							0.3			
BG	3								$0.5+\omega_5$		
CE	3									$0.2+\omega_6$	
CF	2										$0.3+\omega_7$

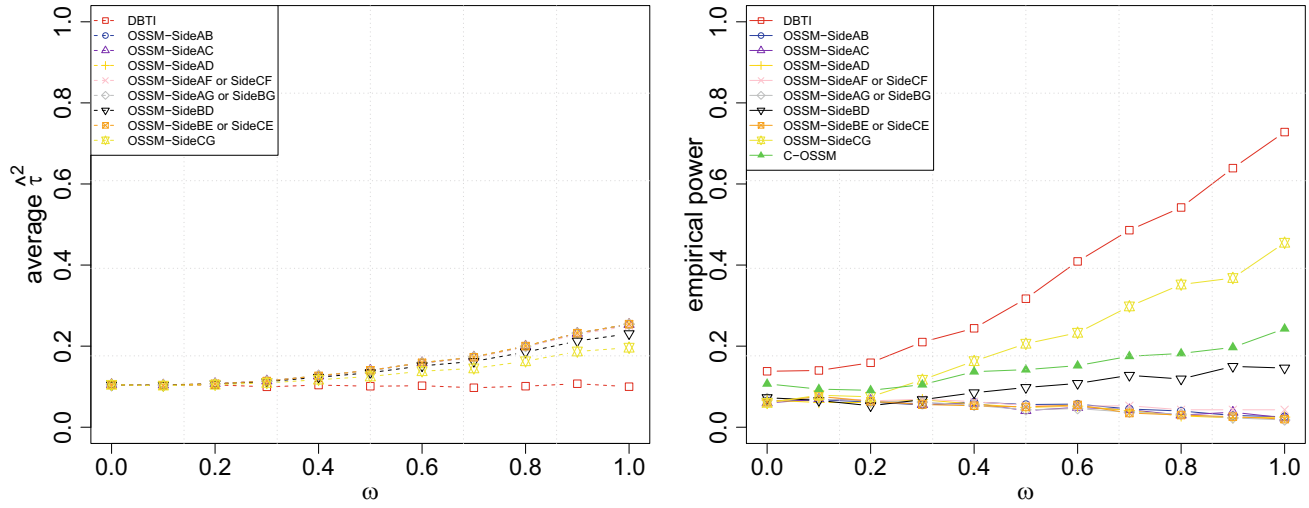
the first structure and a single study in the second. In this network structure, the number of studies for direct comparisons between each pair of interventions is shown in Figure 1d. Therefore, the primary distinction between two network structures lies in the number of studies for design *ACD*, which is highlighted in bold in Table 5.

In our simulations, we compared the design-by-treatment interaction model with all possible splits of sides for the original side-splitting model. Table 5 outlines the parametric form for the design-level parameters of the design-by-treatment interaction model. In the case of the original side-splitting model, Table 5 specifically provides the parametric forms for the design-level parameters associated with splitting sides *AD*, *BD*, and *CD*. From these parametric forms, it is evident that splitting sides *BD* and *CD* yield identical results for parameter estimation and inconsistency detection. In addition, there is no direct comparison between treatments *B* and *C*, the total number of distinct tests for the Bonferroni correction in the original side-splitting model is $m=4$. The simulation results are depicted in Figure 5.

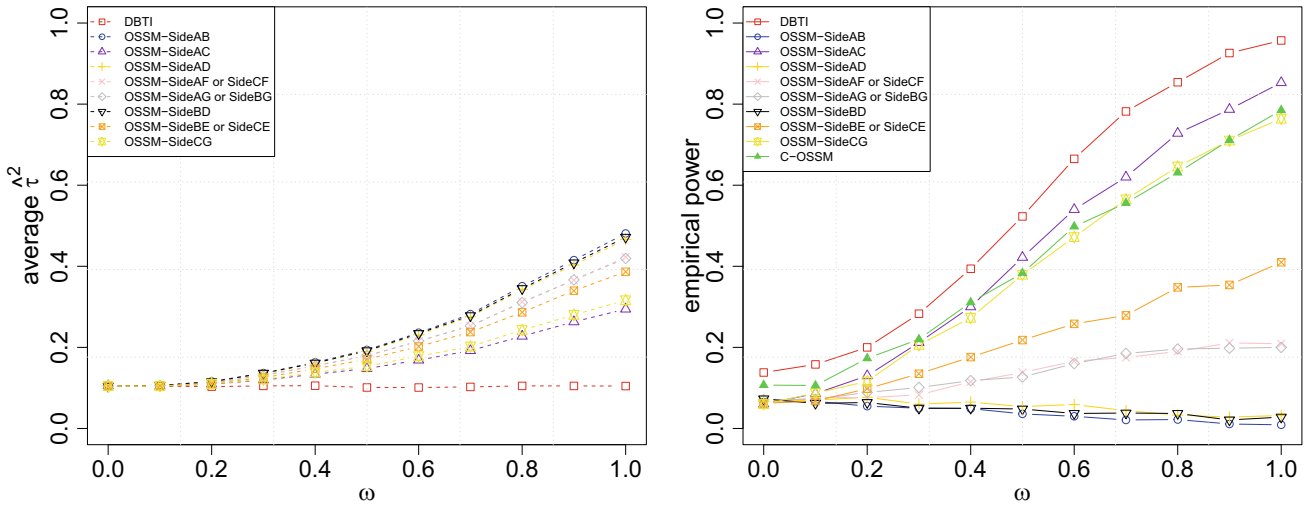
From the simulation results, two main phenomena are observed. Firstly, as depicted in Figure 5, no matter what the inconsistency parameter ω is, both the design-by-treatment interaction model and the original side-splitting model that splits side *AD* can accurately estimate τ^2. This capability stems from their ability to capture the design-level effect sizes for Case 4. Moreover, considering that the original side-splitting model that splits side *AD* only has one inconsistency parameter, it demonstrates higher power than the design-by-treatment interaction model in detecting inconsistency. Benefiting from the robust performance of the original side-splitting model that splits side *AD*, the overall inconsistency detection power of this model is slightly higher than that of the design-by-treatment interaction model. Secondly, in the balanced case where all designs involve three studies, treatments *B* and *C* exhibit complete symmetry. Consequently, simulation results of the original side-splitting models for sides *AB* and *AC* are similar, as shown in Figure 5a. However, as the number of studies for design *ACD* decreases from 3 to 1, the power for inconsistency detection of the original side-splitting model that splits side *AC*, experiences a significant reduction, as illustrated in Figure 5b.

3.3 | Network 5 with seven treatments

In this section, we explore a network with seven treatments: *A*, *B*, *C*, *D*, *E*, *F*, and *G*. The designs and their corresponding numbers of studies are detailed in Table 6. The number of studies for direct comparisons between each pair of treatments is illustrated in Figure 1e. This intricate network structure more closely aligns with the network encountered in real-world data analyses. For the settings of the design-level effect sizes, we begin by considering a consistent network for the potential complete data, setting the basic parameters $\left(\ln \mathrm{OR}^{AB}, \ln \mathrm{OR}^{AC}, \cdots, \ln \mathrm{OR}^{AG}\right)^{\mathrm{T}} = (0.1, 0.2, ..., 0.6)^{\mathrm{T}}$. Subsequently, we introduce seven inconsistency factors for different designs, which is the maximum number permitted by the network. Finally, based on the design-level effect sizes for the potential complete data, we obtain the



(a) Simulation results of Case 5.1.



(b) Simulation results of Case 5.2.

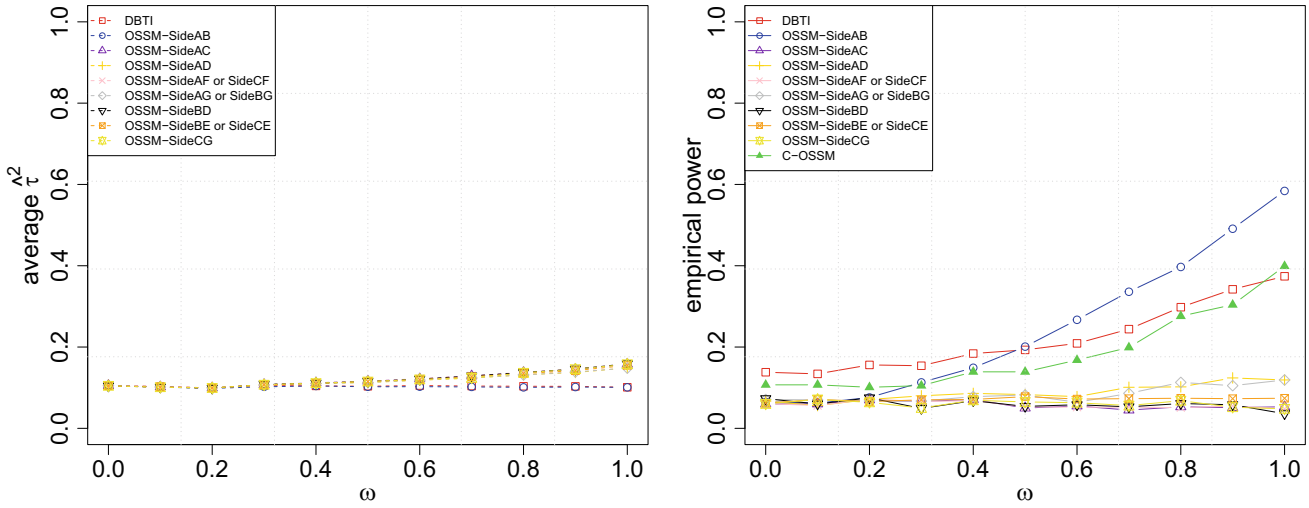
FIGURE 6 Simulation results of Case 5. "DBTI" represents the design-by-treatment interaction model, "OSSM" represents the original side-splitting model, and "C-OSSM" represents the correction of the original side-splitting model.

design-level effect sizes for the observed data, as detailed in Table 6.

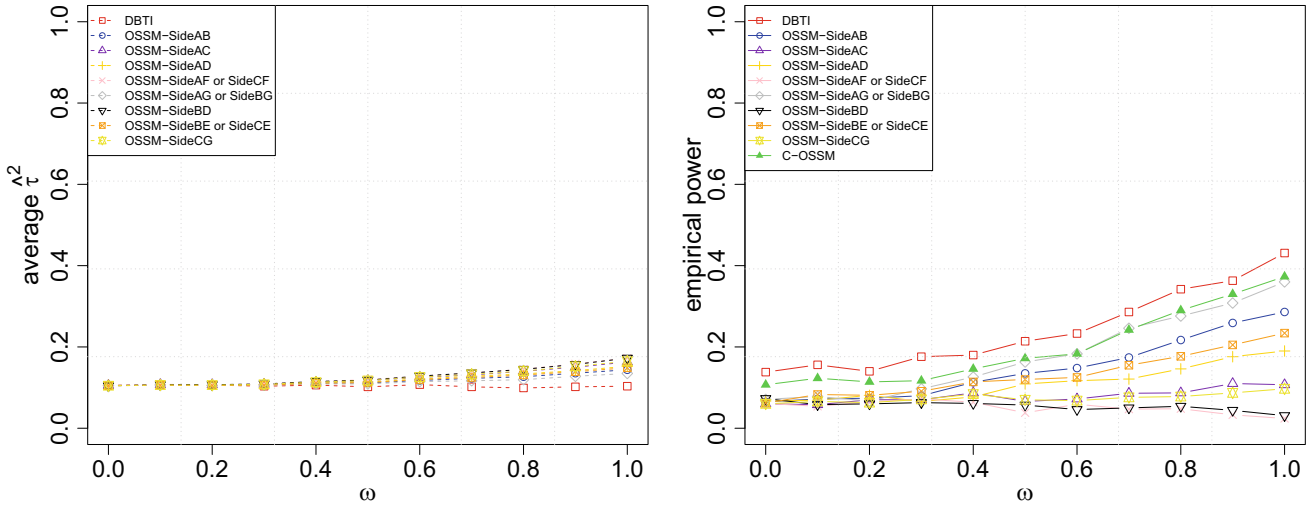
In this network, with a total of seven inconsistency factors, the design-by-treatment interaction model can fully capture the design-level effect sizes. From Figure 1e, it is apparent that among the 7 treatments, there are direct comparisons between 11 pairs of treatments, resulting in 11 splits for the original side-splitting model. In addition, splitting sides *AF* and *CF* yield the same model, splitting sides *AG* and *BG* yield the same model, and splitting sides *BE* and *CE* yield the same model. Consequently, for the original side-splitting model's Bonferroni correction, the number of distinct tests to consider is $m = 8$. For the seven inconsistency factors, we consider the following four settings:

a. *Case 5.1*: $\omega_1 = \omega_2 = \omega_3 = \omega$, and $\omega_4 = \omega_5 = \omega_6 = \omega_7 = 0$,
b. *Case 5.2*: $\omega_1 = \omega_2 = \omega_3 = \omega_4 = \omega_5 = \omega_6 = \omega_7 = \omega$,
c. *Case 5.3*: $\omega_4 = \omega_5 = -\omega_6 = \omega$, and $\omega_1 = \omega_2 = \omega_3 = \omega_7 = 0$,
d. *Case 5.4*: $\omega_4 = \omega_5 = \omega$, $\omega_1 = \omega_2 = \omega_3 = \omega_6 = \omega_7 = 0$,

where the inconsistency parameter ω varied in equally spaced increments from 0 to 1, with intervals of 0.1, representing different levels of inconsistency. In Case 5.1, there are 3 non-zero inconsistency factors. In Case 5.2, this number increases to 7. By comparing these two



(c) Simulation results of Case 5.3.



(d) Simulation results of Case 5.4.

FIGURE 6 (Continued)

cases, we can observe how the number of inconsistency factors influences inconsistency detection. For the design-level effect sizes in Case 5.3, the original side-splitting model splitting side *AB* effectively captures them. The only difference between Cases 5.4 and 5.3 lies in the setting for ω_6. The simulation results are presented in Figure 6.

From Figure 6, it is evident that the design-by-treatment interaction model consistently provides accurate estimates of τ^2. However, it cannot control the type I error rate when $\omega = 0$. When comparing the results for Cases 5.1 and 5.2, the power of the design-by-treatment interaction model increases with the number of inconsistency factors. In the third setting for the inconsistency factors, the original side-splitting model that splits side *AB* significantly outperforms the design-by-treatment interaction model, as it effectively captures the data in Case 5.3. It demonstrates superior control of the type I error rate when $\omega = 0$ and exhibits higher power when $\omega > 0.5$. This suggests that if prior knowledge about the location of the inconsistency factor is available, the original side-splitting model is a preferable choice over the design-by-treatment interaction model. However, if there is a misidentification of the source of inconsistency, as demonstrated in Case 5.4 if the inconsistency is erroneously attributed to side *AB*, and the true

value for ω_6 is zero, the original side-splitting model can perform worse than the design-by-treatment interaction model.

4 | DISCUSSION

Detecting inconsistency accurately in NMA is essential for ensuring the reliability of the meta-analytic results. Previous literature has conducted several simulation studies in this field. Among them, the studies by Glenny et al.[43] and Mills et al.[44] relied on the indirect comparison methods to estimate the effect sizes under a frequentist framework, which may not be suitable for complex networks. Song et al.[45] and Veroniki et al.[46] were among the first to investigate the statistical power for testing inconsistency in NMA through simulations, yet their studies focused only on closed loops formed by two-arm trials. To be more specific, Song et al.[45] compared the Bucher method (indirect treatment comparison), the frequentist mixed treatment comparisons, and the Bayesian mixed treatment comparisons. They concluded that the statistical power of these methods was influenced by the degree of heterogeneity. Veroniki et al.[46] evaluated the properties of the Z-test for detecting inconsistency and examined various factors that can impact the inconsistency test, including pairwise comparisons with or without heterogeneity and different heterogeneity estimation methods. Kiefer et al.[47] conducted a simulation study involving five treatments, but it only included two-arm trials. They found that the assessment of inconsistency was dependent on the degrees of heterogeneity. Apart from the heterogeneity, other factors may also influence inconsistency, such as the choice of measurement scale[48,49] and the data sparsity. Furthermore, inconsistency and heterogeneity are related concepts. When there is considerable heterogeneity among studies, it is necessary to reassess the included evidence, as the results will lack interpretability.[28]

In this paper, we examine the two types of common models for detecting inconsistency, both analytically and numerically. We observe that the performance of these models depends on the extent to which the data conform to the underlying assumptions of the models. Apart from the random errors, the variability in the observed effect sizes can arise from either heterogeneity or inconsistency. Consequently, the estimation and detection of inconsistency are heavily influenced by the assumptions made and the heterogeneity that is assumed and estimated. In the case of the side-splitting models, improper model assumptions can lead to incorrect estimation, mistakenly treating partial inconsistencies as heterogeneity, resulting in low power for inconsistency test. Moreover, it can distort the measurements of both heterogeneity and inconsistency, ultimately leading to a complete failure in detecting inconsistency. Besides, numerical results show that the design-by-treatment interaction model cannot adequately control the type I error rate when the number of studies in each design is small.

Through this study, we have a few important observations for the design-by-treatment interaction model and the side-splitting models within a frequentist framework.

- The design-by-treatment interaction model offers the highest degree of freedom for modeling inconsistency. The side-splitting models, on the other hand, are special cases that require additional assumptions for the inconsistency factors.
- The design-by-treatment interaction model demonstrates robustness in both estimating the heterogeneity and detecting the inconsistency.
- When the side-splitting models align with the data, these models exhibit greater statistical power for detecting the inconsistency compared to the design-by-treatment interaction model. By contrast, if the data does not adhere to the inconsistency assumptions of the side-splitting models, the estimation of heterogeneity will tend to be inaccurate, consequently affecting the inconsistency detection.
- In networks with three treatments and two-arm trials, the design-by-treatment interaction model is equivalent to the side-splitting models.

Recall that the design-by-treatment interaction model is utilized for detecting global inconsistency, whereas the side-splitting models are employed for detecting local inconsistency, along with an assessment of the statistical performance of these two models. We provide practical guidance for inconsistency detection in NMA as follows.

- For detecting inconsistency on a global scale, it is recommended to apply the design-by-treatment interaction model. For detecting inconsistency on a local scale, the side-splitting models can be the suitable choice.
- In cases where there is no specific purpose for inconsistency detection or a lack of information about the potential location of inconsistency, it is advisable to use the design-by-treatment interaction model. The side-splitting models can be employed as supplementary methods, particularly when the number of studies in each design is small. This allows for a comprehensive assessment of inconsistency from both global and local perspectives.

In the current research, our objective is to analyze the design-by-treatment interaction model and the side-splitting models from a statistical perspective. It is crucial to note that, to obtain reliable and interpretable results from an NMA, careful consideration is required for various aspects such as the method for selecting the included studies, the definitions of the network nodes and outcomes, among others. This ensures the construction of a prior consistent network, which should also be subsequently checked statistically.[50] Besides, it is worth mentioning that our research has some limitations that should be acknowledged. Firstly, the determination of whether inconsistency exists relies solely on the *p*-value. It is recommended, however, to take into account the model fit and changes to the estimated between-study heterogeneity in the side-splitting models.[34] A substantial reduction in heterogeneity may also indicate the presence of inconsistency. Secondly, we did not investigate the potential impact of the proportion of direct and indirect comparisons on the inconsistency test, leaving unanswered questions about the extent to which this proportion influences the detection of inconsistency. Thirdly, we did not extensively explore the effects of indirect comparison on the inconsistency test when different numbers of comparators are involved. Fourthly, we did not delve into the exploration of how different heterogeneity estimation methods may affect the estimation of inconsistency. And lastly, given that many real NMAs involve designs with a limited number of studies, there is a need to enhance the performance of the Wald test within the design-by-treatment interaction model. These limitations also highlight the need for further research.

AUTHOR CONTRIBUTIONS

Lu Qin: Software; writing – original draft; formal analysis. **Shishun Zhao:** Validation; methodology; supervision. **Wenlai Guo:** Validation; conceptualization. **Tiejun Tong:** Conceptualization; writing – review and editing; methodology. **Ke Yang:** Writing – original draft; software; investigation.

ACKNOWLEDGMENTS

The authors sincerely thank the Editor, the Associate Editor, and the Reviewer for their constructive comments that have led to a substantial improvement of this paper. Shishun Zhao's research was supported in part by National Natural Science Foundation of China (12071176). Tiejun Tong's research was supported in part by General Research Fund (HKBU12300123 and HKBU12303421) and National Natural Science Foundation of China (12071305). Ke Yang's research was supported in part by National Natural Science Foundation of China (12371294).

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Tiejun Tong https://orcid.org/0000-0003-0947-3990
Ke Yang https://orcid.org/0000-0002-3065-5062

REFERENCES

1. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*. 1997;50(6):683-691.
2. Tu YK. Use of generalized linear mixed models for network meta-analysis. *Med Decis Making*. 2014;34(7):911-918.
3. Psaty BM, Lumley T, Furberg CD, et al. Health outcomes associated with various antihypertensive therapies used as first-line agents: a network meta-analysis. *JAMA*. 2003;289(19):2534-2544.
4. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *Br Med J*. 2005;331(7521):897-900.
5. Cooper NJ, Sutton AJ, Lu G, Khunti K. Mixed comparison of stroke prevention treatments in individuals with nonrheumatic atrial fibrillation. *Arch Intern Med*. 2006;166(12):1269-1275.
6. Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. *Lancet*. 2007;369(9557):201-207.
7. Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *Br Med J*. 2009;338:b1147.
8. Donegan S, Williamson P, Gamble C, Tudur-Smith C. Indirect comparisons: a review of reporting and methodological quality. *PLoS One*. 2010;5(11):e11054.
9. Karyotaki E, Efthimiou O, Miguel C, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA Psychiatry*. 2021;78(4):361-371.
10. Rotshild V, Hirsh-Raccah B, Miskin I, Muszkat M, Matok I. Comparing the clinical efficacy of COVID-19 vaccines: a systematic review and network meta-analysis. *Sci Rep*. 2021;11(1):22777.
11. Sbidian E, Chaimani A, Garcia-Doval I, et al. Systemic pharmacological treatments for chronic plaque psoriasis: a network meta-analysis. *Cochrane Database Syst Rev*. 2023;7(7):CD011535.
12. Baker SG, Kramer BS. The transitive fallacy for randomized trials: if A bests B and B bests C in separate trials, is A better than C? *BMC Med Res Methodol*. 2002;2(13):1-5.
13. Salanti G, Higgins JP, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. *Stat Methods Med Res*. 2008;17(3):279-301.
14. Madden L, Paul P. Assessing heterogeneity in the relationship between wheat yield and fusarium head blight intensity using

random-coefficient mixed models. *Phytopathology*. 2009;99(7): 850-860.
15. Caldwell DM, Welton NJ, Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. *J Clin Epidemiol*. 2010;63(8):875-882.
16. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods*. 2012;3(2):80-97.
17. Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med*. 2013;159(2):130-137.
18. Caldwell DM. An overview of conducting systematic reviews with network meta-analysis. *Syst Rev*. 2014;3(1):109.
19. Hutton B, Salanti G, Chaimani A, et al. The quality of reporting methods and results in network meta-analyses: an overview of reviews and suggestions for improvement. *PLoS One*. 2014;9(3): e92508.
20. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med*. 2015;162(11): 777-784.
21. Tonin FS, Rotta I, Mendes AM, Pontarolo R. Network meta-analysis: a technique to gather evidence from direct and indirect comparisons. *Pharm Pract (Granada)*. 2017;15(1):943.
22. Mavridis D, Sutton A, Cipriani A, Salanti G. A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis. *Stat Med*. 2013;32(1):51-66.
23. Hong H, Carlin BP, Shamliyan TA, et al. Comparing Bayesian and frequentist approaches for multiple outcome mixed treatment comparisons. *Med Decis Making*. 2013;33(5):702-714.
24. Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol*. 2015;15(58):1-9.
25. Greco T, Edefonti V, Biondi-Zoccai G, et al. A multilevel approach to network meta-analysis within a frequentist framework. *Contemp Clin Trials*. 2015;42:51-59.
26. Hu D, O'Connor AM, Wang C, Sargeant JM, Winder CB. How to conduct a Bayesian network meta-analysis. *Front Vet Sci*. 2020;7:271.
27. Seide SE, Jensen K, Kieser M. A comparison of Bayesian and frequentist methods in random-effects network meta-analysis of binary data. *Res Synth Methods*. 2020;11(3):363-378.
28. Ades AE, Welton NJ, Dias S, Phillippo DM, Caldwell DM. Twenty years of network meta-analysis: continuing controversies and recent developments. *Res Synth Methods*. 2024. doi:10.1002/jrsm.1700
29. Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *Br Med J*. 2003;326(7387):472.
30. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med*. 2002;21(16):2313-2324.
31. Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *Can Med Assoc J*. 2009; 181(8):488-493.
32. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc*. 2006;101(474):447-459.
33. Chung H, Lumley T. Graphical exploration of network meta-analysis data: the use of multidimensional scaling. *Clin Trials*. 2008;5(4):301-307.
34. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med*. 2010;29(7–8):932-944.
35. White IR. Network meta-analysis. *Stata J*. 2015;15(4):951-985.
36. Lu G, Welton NJ, Higgins JP, White IR, Ades AE. Linear inference for mixed treatment comparison meta-analysis: a two-stage approach. *Res Synth Methods*. 2011;2(1):43-60.
37. Higgins JP, Jackson D, Barrett J, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods*. 2012;3(2):98-110.
38. Krahn U, Binder H, König J. A graphical tool for locating inconsistency in network meta-analyses. *BMC Med Res Methodol*. 2013;13(35):1-18.
39. White IR, Barrett JK, Jackson D, Higgins JP. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Res Synth Methods*. 2012; 3(2):111-125.
40. Tu YK. Node-splitting generalized linear mixed models for evaluation of inconsistency in network meta-analysis. *Value Health*. 2016;19(8):957-963.
41. Nikolakopoulou A, Chaimani A, Veroniki AA, Vasiliadis HS, Schmid CH, Salanti G. Characteristics of networks of interventions: a description of a database of 186 published networks. *PLoS One*. 2014;9(1):e86754.
42. Veroniki AA, Vasiliadis HS, Higgins JP, Salanti G. Evaluation of inconsistency in networks of interventions. *Int J Epidemiol*. 2013;42(1):332-345.
43. Glenny AM, Altman DG, Song F, et al. Indirect comparisons of competing interventions. *Health Technol Assess*. 2005;9(26): 1-134.
44. Mills EJ, Ghement I, O'Regan C, Thorlund K. Estimating the power of indirect comparisons: a simulation study. *PLoS One*. 2011;6(1):e16237.
45. Song F, Clark A, Bachmann MO, Maas J. Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons. *BMC Med Res Methodol*. 2012;12(138): 1-14.
46. Veroniki AA, Mavridis D, Higgins JP, Salanti G. Characteristics of a loop of evidence that affect detection and estimation of inconsistency: a simulation study. *BMC Med Res Methodol*. 2014;14(106):1-12.
47. Kiefer C, Sturtz S, Bender R. A simulation study to compare different estimation approaches for network meta-analysis and corresponding methods to evaluate the consistency assumption. *BMC Med Res Methodol*. 2020;20(36):1-13.
48. Caldwell DM, Welton NJ, Dias S, Ades AE. Selecting the best scale for measuring treatment effect in a network meta-analysis: a case study in childhood nocturnal enuresis. *Res Synth Methods*. 2012;3(2):126-141.

49. Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med*. 2002;21(11):1575-1600.
50. Cope S, Zhang J, Saletan S, Smiechowski B, Jansen JP, Schmid P. A process for assessing the feasibility of a network meta-analysis: a case study of everolimus in combination with hormonal therapy versus chemotherapy for advanced breast cancer. *BMC Med*. 2014;12(93):1-17.

How to cite this article: Qin L, Zhao S, Guo W, Tong T, Yang K. A comparison of two models for detecting inconsistency in network meta-analysis. *Res Syn Meth*. 2024;15(6):851-871. doi:10.1002/jrsm.1734