Inconsistency identification in network meta-analysis via stochastic search variable selection

Georgios Seitidis1 | Stavros Nikolakopoulos2,3 | Ioannis Ntzoufras4 | Dimitris Mavridis1

1Department of Primary Education, University of Ioannina, Ioannina, Greece
2Department of Psychology, University of Ioannina, Ioannina, Greece
3Department of Biostatistics, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
4Department of Statistics, Athens University of Economics and Business, Athens, Greece

Correspondence
Georgios Seitidis, Department of Primary Education, University of Ioannina, Ioannina, Greece.
Email: g.seitidis@uoi.gr

The reliability of the results of network meta-analysis (NMA) lies in the plausibility of the key assumption of transitivity. This assumption implies that the effect modifiers’ distribution is similar across treatment comparisons. Transitivity is statistically manifested through the consistency assumption which suggests that direct and indirect evidence are in agreement. Several methods have been suggested to evaluate consistency. A popular approach suggests adding inconsistency factors to the NMA model. We follow a different direction by describing each inconsistency factor with a candidate covariate whose choice relies on variable selection techniques. Our proposed method, stochastic search inconsistency factor selection (SSIFS), evaluates the consistency assumption both locally and globally, by applying the stochastic search variable selection method to determine whether the inconsistency factors should be included in the model. The posterior inclusion probability of each inconsistency factor quantifies how likely is a specific comparison to be inconsistent. We use posterior model odds or the median probability model to decide on the importance of inconsistency factors. Differences between direct and indirect evidence can be incorporated into the inconsistency detection process. A key point of our proposed approach is the construction of a reasonable “informative” prior concerning network consistency. The prior is based on the elicitation of information derived historical data from 201 published network meta-analyses. The performance of our proposed method is evaluated in two published network meta-analyses. The proposed methodology is publicly available in an R package called ssifs, published on CRAN and developed and maintained by the authors of this work.

KEYWORDS
consistency, NMA, SSVS, transitivity, variable selection
1 | INTRODUCTION

In evidence synthesis, network meta-analysis (NMA) is typically employed to evaluate the relative efficacy and/or safety of multiple available interventions.\(^1\)-\(^3\) A key assumption in NMA is that of transitivity, also termed in the literature as the similarity assumption, which refers to the ability to learn about a treatment effect indirectly.\(^3\) NMA synthesizes direct and indirect evidence and violations of transitivity may manifest themselves in large differences between these two sources of evidence, also known as inconsistency in the network.

Several methods have been suggested to evaluate the consistency assumption that can be broadly categorized into local and global methods. Local methods evaluate consistency either in every loop or comparison in the network for which we have both direct and indirect evidence.\(^4\)-\(^9\) Global methods include approaches where inconsistency factors are added to the NMA model and evaluated simultaneously,\(^10\)-\(^12\) employing the generalized Cochran’s Q statistic,\(^13\) fitting the NMA model using integrated nested Laplace approximations,\(^14\) and treating the comparisons as independent parameters for estimation.\(^15\) Local methods have an increased probability of false positives for large networks, while global methods suffer from lack of power.\(^10\) Moreover, local methods in which for each comparison we contrast direct evidence and indirect effect estimates (node-splitting, symmetric side-splitting, separate indirect from direct design evidence SIDDE), have been shown in the presence of multi-arm trials to disregard the principle of independence between direct and indirect evidence and as a result, may lead to false conclusions.\(^8\)

The Lu and Ades model is a popular method for testing the consistency assumption globally, which is evaluated by adding inconsistency factors to each independent closed-loop.\(^12\) Drawbacks of the method are (1) the model is not uniquely defined and different parameterizations (adding an inconsistency factor to different comparisons within a loop) may lead to different results, (2) it accounts only for loop inconsistencies, and (3) in the presence of multi-arms studies the implementation of the method is not straight-forward. Until now, the implementation of the method has been carried out manually, by specifying the inconsistency factors by hand. Although this may not be an issue in small networks with few loops, in complex networks, the specification of the inconsistency factors is extremely hard, making the implementation of the method in most cases not feasible.

When employing NMA in a Bayesian framework, inconsistency models have been used to relax the consistency assumption by treating inconsistency as an additional random effect.\(^10\)-\(^12\) Popular methods for exploring inconsistency include estimating the variance of the inconsistency random-effects, computing deviance contribution and comparing the deviance information criterion (DIC) for consistency and inconsistency models.\(^15\) It should be noted that both consistency and inconsistency models are essentially linear regression models. In the inconsistency model, extra inconsistency factors are added as covariates to account for any difference between direct and indirect evidence. Therefore, by employing variable selection techniques, the importance of inconsistency factors and the plausibility of the consistency assumption can be thoroughly assessed.

The scope of this article is two-fold: Firstly, we propose employing a Bayesian variable selection approach tailored to the NMA inconsistency model, aiming to evaluate the consistency assumption both locally and globally. Subsequently, we present a consistent algorithm for specifying the inconsistency factors in the Lu and Ades model. For the implementation of the suggested methodology, we developed the R package \texttt{ssifs}\(^16\) which is available on CRAN. The article is structured as follows. We first describe the NMA random-effects model and how it is formulated to include inconsistency factors. Next, we present our proposed method, and we illustrate it in two datasets from the field of medicine. The article concludes with a discussion about possible extensions of the method.

2 | METHODS

2.1 | The NMA consistency and inconsistency models

Suppose that we have considered trials that evaluate the efficacy of three treatments A, B, and C (ABC network). Further, assume that all studies under consideration provide information for the comparison between all three pairs of interventions forming a closed loop of evidence. The indirect estimate (denoted with an upper index “\(\text{Ind}\)”) for a comparison (say BC) would be \(\mu_{BC}^{\text{Ind}} = \mu_{AC}^{\text{Dir}} - \mu_{AB}^{\text{Dir}}\) where the upper index “\(\text{Dir}\)” denotes direct evidence. Consistency implies that parameters measuring direct and indirect evidence should be equal for all treatment comparisons. Hence, for the
In NMA, we have two types of inconsistency: (a) loop and (b) design inconsistency. 10

In order to specify the first, we need first to introduce the notion of loop and closed-loop in a network. The term “loop” is used to specify a path formed by a group of interventions. A loop is closed when a group of interventions is connected in
the network graph with a polygon. In practice, it means that we can calculate all pairwise treatment comparisons either directly or indirectly. Loop inconsistency is present, when in a closed-loop, direct and indirect evidence vary substantially. For example, in a three-treatment loop, if Equation (1) does not hold, we have loop inconsistency. Multi-arm trials are by definition consistent complicating the definition of loop inconsistency.

The second type, design inconsistency, was introduced by Higgins et al. The term design refers to the set of treatments compared within a study. Design inconsistency refers to differences in effect sizes between studies involving different sets of interventions. For example, suppose that in the previous ABC network, ABC three-arm studies are also present, and the effect estimate of comparison AB between the AB and ABC studies differs substantially.

2.2.1 Specification of inconsistency structure

Several models have been developed for the specification of $Z$ matrix, which is the design matrix specifying which inconsistency parameters should be added in the linear predictor (2). Lu and Ades proposed a model which accounts only for loop inconsistencies in the networks, and the consistency assumption is tested in every independent closed-loop. In complex networks, the implementation of the method is challenging since the specification of the inconsistency factors is sequentially performed manually. Thus, we have developed an algorithm that automatically specifies the comparisons where inconsistency factors should be added. Figure 1 presents the process flowchart of the algorithm. More details about the algorithm can be found in the appendix.

The algorithm identifies the Lu and Ades model by defining the functional parameters of the model and identifying the independent loops of the network. It should be noted that the treatment ordering is ignored, and the algorithm identifies only a subset of the potential Lu and Ades models. Hence, we suggest using the design-by-treatment model instead, which captures all the Lu and Ades models and addresses both sources of inconsistency. Moreover, the random-effects implementation of the design-by-treatment interaction model developed by Jackson et al can be used, in which consistency is evaluated at each design of the network.

FIGURE 1 Process flowchart of identifying the comparisons in which inconsistency factor should be added based on the Lu and Ades model. (Abbreviation FN: functional parameters).
2.3 | Stochastic search inconsistency factor selection

Concerning the evaluation of the network inconsistencies, we propose an implementation of the stochastic search variable selection (SSVS) algorithm\textsuperscript{17} which will be referred in the following as stochastic search inconsistency factors selection (SSIFS). SSIFS is a two-step method in which, at the first step, we add inconsistency factors in the NMA model based on the specification method of matrix $Z$ described in Section 2.2.1. At the second stage, we treat the detection of inconsistency factors as a variable selection problem and use SSVS to identify inconsistencies in a network under study.

After specifying $Z$, following the SSVS model specification, each element of $b$ is modeled as a realization from a mixture of two normal distributions with different variances. A latent vector $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_p)$ is associated with each element of $b$ which operates as an indicator variable (taking values 0 or 1) identifying whether the respective element of $b$ lies in a small area around zero or not. We represent the normal mixture by

$$b_\ell | \gamma_\ell \sim (1 - \gamma_\ell)N(0, \psi_\ell^2) + \gamma_\ell N(0, c^2\psi_\ell^2),$$

where $\ell = 1, 2, \ldots, p$. The first density (“spike”) denotes the effect of the inconsistency factor when it is minor (ie, close to zero) and therefore it should not be included in the model ($\gamma_\ell = 0$), while the second (“slab”), denotes the inconsistency factor effect when it away from zero and therefore it should be included in the model ($\gamma_\ell = 1$). Note that when an inconsistency factor is not included in the model, its effect is not exactly zero but shrunken toward zero. Therefore, it is important to specify parameters $c$ and $\psi_\ell$ with caution to assure a density similar to the depicted density of Figure 2. The variance of the spike ($\psi_\ell^2$) should be small, and $c^2$ should be sufficiently large in order the slab variance to support plausible values of the inconsistency factor when it is included in the model (see more details in Section 2.3.3).

The mixture of these two densities can be represented in a condensed matrix notation by

$$b | \gamma \sim N(0, D_\gamma R D_\gamma),$$

where $R$ is the prior correlation matrix of the elements in $b$, $D_\gamma = \text{diag}(a_1\psi_1, \ldots, a_p\psi_p)$ with $a_\ell = c^\ell$ (ie, $a_\ell = 1$ if $\gamma_\ell = 0$ and $a_\ell = c$ if $\gamma_1 = 1$); where $\psi_\ell$ and $c$ are tuning parameters controlling the mixing ability of the method, the penalty induced for each extra parameter included in the model and the interval of non-importance for each inconsistency.

2.3.1 | Prior setup

A convenient choice for matrix $R$ is to assume that is equal to the identity matrix $I_p$. This setting implies that the inconsistency factors are independent, and the inclusion of an inconsistency factor does not affect the inclusion of another. In a network of interventions where each effect estimate affects all other estimates, independence of inconsistency factors is questionable. A popular choice for prior setup in Bayesian variable selection is the Zellner’s $g$-prior.\textsuperscript{18} Under this approach, the correlation matrix $R$ is written by the following expression:

$$R = g(Z^TZ)^{-1}\sigma^2,$$

![Figure 2](image-url) Graphical representation of the SSVS spike and slab prior setup.
while we consider an improper prior for the error variance of the type

$$\pi(\sigma^2) \propto \frac{1}{\sigma^2}.$$  

The main advantages of the Zellner’s g-prior are that the general correlation of \(b\) is encapsulated by the matrix \((Z^TZ)^{-1}\) while we have only one remaining variance parameter to specify in comparison with the general formulation where we need to specify the whole matrix \(R\). Nevertheless, the specification of \(g\) should be cautious in order not to trigger the Jeffreys–Lindley–Bartlett paradox,\(^{19,21}\) in which for proper large values of \(g\) the posterior odds tend to favor the simplest model, which in our case is the consistency model. For this reason, several approaches have been proposed, including the use of hyper-prior\(^{22,23}\) or assuming \(g\) to be fixed. We use the common choice of unit-information criteria approach,\(^{24}\) which mathematically translates to \(g = N\). The prior parameter \(g\) has a straightforward interpretation in terms of how much data information we add in our procedure. Actually, the amount of information we add is equivalent to \(N/g\) data points so the standard choice is to select \(g = N\) in order to have the unit information prior.\(^{25}\) Also, the Zellner’s g-prior is directly connected with the Bayesian Information Criterion (BIC) in the case where \(g = N\). From this connection, we can derive that the posterior model probabilities can be considered as information criteria with penalty given by \(\log(g)\).\(^{26}\) Under this perspective, the results may be sensitive to the choice of \(g\) since it controls the penalty and the sparsity of the procedure. Nevertheless, the standard choice of setting \(g = N\) insures a well-behaved model selection procedure.\(^{25}\)

The error variance term \(\sigma^2\) is assumed to follow the Jeffreys scale-invariant prior. An alternative for the error variance is to assume a non-informative inverse-gamma prior with equal shape and scale parameters, typically set to \(10^{-3}\) or \(10^{-4}\).\(^{27}\) This choice has the advantage of being proper, and the results are not sensitive to the choice of this prior or parameters provided that these values are large enough.

### 2.3.2 Prior inclusion probabilities

Prior inclusion probabilities define the probability of including each inconsistency factor in the model, a priori. It is crucial to specify these probabilities appropriately to avoid issues with multiple comparisons. The standard but naïve non-informative choice of equal prior probabilities for all models does not account for multiplicity.\(^{28}\) A standard approach in Bayesian variable selection is to use a Beta-binomial distribution on the prior inclusion probabilities, which accounts for multiplicity by automatically introducing a penalty that handles multiple testing.\(^{28,29}\) This expressed mathematically as follows:

\[
\gamma_\ell \sim \text{Bernoulli}(\pi), \\
\pi \sim \text{Beta}(\alpha, \beta). \quad (3)
\]

In the simple case where \(\alpha = \beta = 1\), this is equivalent to assigning a uniform prior to \(\pi\). Even though this setting appears appealing, it cannot be applied to SSIFS as it will unrealistically favor inconsistent NMA models. Consider the simple case where we assume that each inconsistency factor is Bernoulli distributed with 0.50 inclusion probability

\[
\gamma_\ell \sim \text{Bernoulli}(0.50), \ell = 1, 2, \ldots, p.
\]

Let \(P(\gamma_\ell = 1)\) denotes the probability that the inconsistency factor is included in the model and \(1 - \pi_\ell = P(\gamma_\ell = 0)\) the probability not included. The prior probability \(\pi_{\text{cons}}\) of the consistency model will then be given by

\[
\pi_{\text{cons}} = \prod_{\ell=1}^{p} P(\gamma_\ell = 0) = 0.50^p.
\]

Therefore, the probability of the consistency model decreases exponentially as the number of inconsistency parameters increases. Also, this choice will tend to favor models with \(\frac{p}{2}\) inconsistency factors. Alternatively, inclusion probabilities should be specified in an such manner where the probability of the consistency model remains fixed. Thus, the common
inclusion probability $\pi$ for all inconsistency factors can be obtained by

$$\pi_{\text{cons}} = P(\gamma_1 = 0, \ldots, \gamma_p = 0) \Leftrightarrow \pi_{\text{cons}} = (1 - \pi)^p \Leftrightarrow$$

$$\pi = 1 - \pi_{\text{cons}}^{1/p}.$$  

SSIFS can also incorporate past knowledge concerning network consistency. Parameter $\pi_{\text{cons}}$ denotes the analyst’s prior belief of having a consistent network. If an expert’s opinion is available, then an informative prior could be used. Otherwise, we consider the value of 0.50 to express our prior ignorance. In a review of 201 NMA networks, 44 networks were found to be globally inconsistent. Thus, the probability of having a consistent network is estimated from this study to be equal to 0.78. This could be used as an informative prior for $\pi_{\text{cons}}$. Parameter $\psi$ is set in terms of model comparison. The setting of $\psi$ can be interpreted as the prior odds of excluding the inconsistency factor when its effect is close to zero. The larger the value of $c$, the stricter we are about including inconsistency factors with small effects. It has been observed that values between 10 and 100 usually perform well in most cases however caution may be needed depending on the data characteristics.

$$\begin{align*}
\widehat{f}(\gamma_\ell = 1|y) &= \frac{1}{M - B} \sum_{t=B+1}^M I(\gamma^{(t)}_\ell = 1),
\end{align*}$$

where $B$ is the number of iterations considered as burn-in period, and $\gamma^{(t)}_\ell$ is the inclusion value of $\ell$ inconsistency factor at iteration $t$.

### 2.3.3 Tuning

Tuning of SSIFS parameters is essential to ensure a good mixing of the MCMC method, but also to attain sensible results in terms of model comparison. The setting of $\psi$ and $c$ must be in such a way that $\psi^2$ be small and $c^2\psi^2$ large. Possible values of $\psi$ could be the inconsistency factor’s standard deviations obtained from a pilot MCMC run of the full NMA model. Parameter $c$ can be interpreted as the prior odds of excluding the inconsistency factor when its effect is close to zero. The larger the value of $c$, the stricter we are about including inconsistency factors with small effects. It has been observed that values between 10 and 100 usually perform well in most cases however caution may be needed depending on the data characteristics.

To specify the prior parameters of the variable selection part ($\psi$ and $c$), we can consider a priori assume a minimum accepted value (say $\omega$) for the inconsistency parameters $b$. This is referred as the level of practical significance of an inconsistency parameter. Thus, an inconsistency factor with a coefficient larger than $\omega$ in absolute values ($|b_\ell| > \omega$), should be a-priori included in the NMA model ($\gamma_\ell = 1$). This can be achieved by setting the parameters $\psi$ and $c$ according to the intersection of the two priors used under the consistency and inconsistency status of the model; that is $f(b_\ell|\gamma_\ell = 0)$ and $f(b_\ell|\gamma_\ell = 1)$, respectively. The intersection point can be found by solving the equation $f(b_\ell|\gamma_\ell = 0) = f(b_\ell|\gamma_\ell = 1)$ with regard to $b$; see Appendix for details. In the case where $R = I_p$, the density of $f(b_\ell|\gamma_\ell = 1)$ will be higher than the density of $f(b_\ell|\gamma_\ell = 0)$ when $|b_\ell| > \psi_\ell \sqrt{\xi(c)}$, where $\xi(c) = \frac{2c^2 \log(c)}{c^2 - 1}$ denotes how many standard deviations an inconsistency factor should be away from zero to be considered significant. For example, if a difference above 0.2 is considered important, one possible parameterization is to set $c = 10$. Then, we set $0.2 = \psi_\ell \sqrt{\xi(c)}$, obtaining $\psi_\ell = \frac{0.2}{\sqrt{\xi(10)}} \approx 0.1$. In the case where a correlation between the $p$ elements of the inconsistency parameter vector $b$ is assumed, tuning should be based on solving the following inequality

$$b'[D_{\ell=0}(Z'Z)^{-1}D_{\ell=0}]^{-1}b \leq gp\sigma^2 \xi(c), \xi(c) = \frac{2c^2 \log(c)}{c^2 - 1}.$$  

More details about tuning can be found in References 33 and 34.
2.3.4 Inconsistency detection

In an NMA network with \( p \) inconsistency factors there are \( 2^p \) possible models. The aim of SSIFS is to find the most probable model in terms of posterior probability, termed as maximum a-posteriori model. A common approach for model comparison in the Bayesian framework is to calculate the posterior model odds. When comparing two models \( m_1 \) and \( m_2 \), the posterior odds of \( m_1 \) over \( m_2 \) calculated as follows:

\[
PO_{m_1,m_2} = \frac{f(m_1|y)}{f(m_2|y)} = \frac{f(y|m_1) f(m_1)}{f(y|m_2) f(m_2)} = BF_{m_1,m_2} \cdot \frac{f(m_1)}{f(m_2)},
\]

where \( BF_{m_1,m_2} \) is the Bayes factor of \( m_1 \) over \( m_2 \) and \( f(m_1) \) and \( f(m_2) \) are the prior model probabilities. If the prior model probabilities are equal, then the posterior model odds and the Bayes factors coincide. Posterior model odds can be obtained as the ratio of the posterior model probabilities which in SSIFS are estimated by

\[
\hat{f}(m|y) = \frac{1}{M-B} \sum_{i=B+1}^M I(m^{(i)} = m),
\]

where \( m^{(i)} \) is the model indicator in \( t \) iteration which transform the \( y \) to a unique decimal number and calculated as follows:

\[
m(y) = \sum_{\ell=1}^{p} y^{(\ell)} 2^{\ell-1}.
\]

Posterior odds of the consistent NMA model \((m(y) = 0)\) over all the other observed inconsistent NMA models \((m(y) \neq 0)\) indicates if the NMA model is globally consistent.

An alternative strategy is to find the median probability model which includes only inconsistency factors with posterior inclusion probability larger than 0.50. For complex networks and/or too many inconsistency factors (e.g., more than 20), the SSIFS may not accurately calculate the posterior model probabilities.\(^{31}\) In such cases, a good strategy is to reduce the dimension by excluding the inconsistency factors that have posterior inclusion probability below 0.20.\(^{35}\)

3 EXAMPLES

To illustrate our method, we used two examples, one on smoking cessation reported by the AHCPR Smoking Cessation Guideline Panel,\(^{12}\) and the other on comparative effectiveness of oral phosphodiesterase type-5 inhibitors for erectile dysfunction.\(^{30,36}\) Matrix \( Z \) in SSIFS is specified using the Lu and Ades model, the design-by-treatment and the Jackson’s approach. For the specification of the prior correlation matrix \( R \), we took into consideration both approaches. Firstly, we assume that inconsistency factors are independent by setting \( R = I_p \). Secondly, that there is a dependency between inconsistency factors which is described by the Zellner’s g-prior \((R = g(\Sigma Z^T Z)^{-1} \sigma^2)\); we assumed in the main analysis that \( g = N \), and as a sensitivity analysis, we also considered the cases where \( g = N/2 \) and \( g = 2N \). Tuning was performed by assuming 0.2 as the minimum value of inconsistency that is of practical significance on the log scale. Regarding the probability of observing a consistent network, we used two different scenarios, where at the first \( \pi_{cons} = 0.50 \) and at the second \( \pi_{cons} \sim Beta(157,44) \). The first is the uniform prior on model space approach while the second is informative based on previous historical information. To evaluate further the robustness of our findings, we considered as a sensitivity analysis the extreme scenario where inconsistency is favored a-priori by 75% \((\pi_{cons} = 0.25)\). Also, for the erectile dysfunction network, we evaluated network consistency by randomly excluding 10% of the studies; the smoking cessation network was not used due to the relatively small number of studies. A burn-in period of 50K iterations, and a total of 300K iterations were employed, using two chains to monitor convergence.

SSIFS was performed using the R package ssifs\(^{16}\) developed by the authors of this work. Details on the installation and the usage of package can be found at https://georgiosseitidis.github.io/ssifs/.
3.1 Smoking cessation

The dataset consists of 24 studies (22 two-arm studies and two three-arm studies) comparing the relative effects of no contact (reference), self-help, individual counseling, and group counseling on smoking cessation. Network’s geometry is presented in Figure 3.

SSIFS indicates that the NMA model is consistent since, in all scenarios inconsistency factor’s effects are estimated as non-important a-posteriori, yielding posterior inclusion probabilities close to zero (for more details see Figure 4 and Table S1). Hence, the median probability model clearly suggests that no important design or loop inconsistencies are present in the network. Moreover, from Table 1, we can see that the consistent NMA model (model without inconsistency factors) is estimated as maximum a-posteriori one with posterior probability around 0.56 when $\pi_{cons} = 0.50$ and above 0.81 when $\pi_{cons} \sim \text{Beta}(157, 44)$. As a result, the posterior odds favor the consistent NMA model when compared with the inconsistent NMA models (for more details see Tables 1 and S2–S13 in the supplementary material). Even if all the inconsistent NMA models are treated as a single model, the posterior odds of consistency vs inconsistency still favors the consistency hypothesis, with posterior odds ranging from $0.55 / (1 - 0.55) = 1.22$ to $0.58 / (1 - 0.58) = 1.38$ (marginal evidence) when the uniform prior on model space is used, and from $0.81 / (1 - 0.81) = 4.26$ to $0.83 / (1 - 0.83) = 4.88$ when historical information is used; see Table 1 for more details. Even in the extreme case where $\pi_{cons} = 0.25$, the median probability model continues to suggest that there are no important inconsistencies and the consistent NMA model was estimated as the most dominant (see Figure 4 and Tables S14–S19 for more details). Moreover, the posterior odds continue to favor the consistent NMA model when compared with the second most dominant inconsistent NMA model (posterior odds ranging from 1.66 to 6.14) and the rest inconsistent NMA models (see Tables 1 and S14–S19). It should be noted that the global test for consistency indicates marginal evidence against the consistency hypothesis with posterior odds ranging from $0.30 / (1 - 0.30) = 0.43$ to $0.32 / (1 - 0.32) = 0.47$ (see Table 1), which was expected due to the high probability that it was a priori given to the inconsistent NMA models. Thus, we may safely conclude that the network is globally consistent without any significant local inconsistencies. Conclusions regarding network consistency are the same in the cases where $g = N/2$ and $g = 2N$ since the posterior inclusion/model probabilities are similar (see Figures S1 and S2).

Our findings agree with previous work, which also suggests that the network is consistent. In addition, Dias et al revealed no substantial differences between the random-effects NMA model and the node-split models, indicating the absence of inconsistency. Also, Lu and Ades revealed that the global model fit statistics nor the inconsistency P-value of 0.27 suggests the presence of inconsistency. Nevertheless, they point out that due to the high level of heterogeneity, we cannot confidently conclude that there is no inconsistency. This is also reflected in SSIFS in the case where we used an indifferent prior with respect to the choice of the consistency ($\pi_{cons} = 0.50$), in which marginal evidence is available that the network is consistent.
FIGURE 4  Posterior inclusion probabilities for the smoking cessation example for three different inconsistency models.

TABLE 1  Posterior model odds (PO) of the consistent NMA model vs the most probable inconsistent NMA model and the corresponding probability of the consistent NMA model ($\hat{m}_{cons}$) for the three different inconsistency modeling approaches.

| Inconsistency modeling approach | $\pi_{cons} = 0.50$ | $\pi_{cons} \sim \text{Beta}(157,44)$ | $\pi_{cons} = 0.25$ |
|--------------------------------|---------------------|--------------------------------------|---------------------|
|                                | $R = I_p$            | $R = g(Z'Z)^{-1}\sigma^2$            | $R = I_p$            |
| Smoking cessation              | $\hat{m}_{cons}$ PO | $\hat{m}_{cons}$ PO                  | $\hat{m}_{cons}$ PO |
| DBT                            | 0.56 7.34            | 0.56 9.81                            | 0.31 3.55            |
| Lu and Ades                    | 0.56 3.75            | 0.58 4.46                            | 0.31 1.66            |
| Jackson                        | 0.57 11.28           | 0.55 12.75                           | 0.32 5.21            |

| Erectile dysfunction           | $\pi_{cons}$ PO      |
| DBT, Lu and Ades               | 0.14 0.16             |

Note: Prior Setup: (a) $\pi_{cons} = 0.50$ and $R = I$, (b) $\pi_{cons} = 0.50$ and $R = g(Z'Z)^{-1}\sigma^2$, (c) $\pi_{cons} \sim \text{Beta}(157,44)$ and $R = I$, (d) $\pi_{cons} \sim \text{Beta}(157,44)$ and $R = g(Z'Z)^{-1}\sigma^2$, (e) $\pi_{cons} = 0.25$ and $R = I$, and (f) $\pi_{cons} = 0.25$ and $R = g(Z'Z)^{-1}\sigma^2$.

Abbreviations: DBT, design-by-treatment; PO, posterior model odds; $\hat{m}_{cons}$, posterior probability of the consistent NMA model.

Note that despite the variability of the posterior inclusion probabilities of each inconsistency factor across different methods, the posterior probability of the consistent NMA model is equivalent and robust, taking values $\sim 0.31$ when $\pi_{cons} = 0.25$, $\sim 0.56$ when $\pi_{cons} = 0.50$ and values $\sim 0.82$ when the historical based beta prior is used for $\pi_{cons}$. The posterior inclusion probabilities of the Lu and Ades model are larger than the equivalent probabilities of the other models. This is because the Lu and Ades model has fewer inconsistency factors, yielding to smaller prior inclusion probabilities, and the posterior inclusion probabilities in this example are primarily affected by the prior distribution. For example, in the case where $\pi_{cons} = 0.50$, the Lu and Ades model has three inconsistency factors yielding a prior inclusion probability equal to $\pi_c = 1 - 0.50^3 \approx 1 - 0.79 \approx 0.21$, while the design-by-treatment model has seven with a prior inclusion probability of $\pi_c = 1 - 0.50^7 \approx 1 - 0.91 \approx 0.09$. 

Abbreviations: A = no contact, B = self-help, C = individual counseling, D = group counseling.
3.2 Comparative effectiveness of oral phosphodiesterase type-5 inhibitors for erectile dysfunction

The dataset consists of 69 studies two-arm studies comparing the effectiveness of six type-5 inhibitors (Mirodenafil, Sildenafil, Avanafil, Tadalafil, Udenafil, and Vardenafil). Placebo was used as reference treatment for our analysis. Network’s geometry is presented in Figure 5.

From the network geometry, we can see that only one closed loop exists between Sildenafil, Vardenafil and Placebo. Because the network contains only two-arms studies, the specification of matrix $Z$ is identical for both design-by-treatment and the Lu and Ades models; only one inconsistency factor can be applied between Sildenafil and Vardenafil. Thus, SSIFS will be the same for these models. For the same reason, the Jackson’s model was not used for the specification of matrix $Z$.

Figure 6 indicates that the NMA model is inconsistent since loop inconsistency was identified between treatments Sildenafil, Vardenafil and Placebo. The estimated inconsistency factor’s effect is evident when $\pi_{\text{cons}} = 0.50$, yielding to a posterior inclusion probability above 0.86 (larger than 0.50, for more details see Table S20) and posterior odds of $0.86/(1 - 0.86) = 6.14$ indicating substantial/positive evidence in favor of the hypothesis of inconsistency according to Kass and Raftery. The evidence against consistency is more intense in the case where $\pi_{\text{cons}} = 0.25$, in which the posterior inclusion probability estimated as 0.95, yielding to a posterior odds of $0.95/(1 - 0.95) = 19$ against the consistency hypothesis (see Table 1). Since only one inconsistency factor is added in the NMA model, the posterior inclusion probability is equal to the posterior model probability of the inconsistent NMA model. Moreover, the posterior odds show strong evidence of the inconsistent NMA model over the consistent NMA model (for more details see Table 1). Hence, the SSIFS suggests that the network is globally inconsistent due to the loop inconsistency. The inconsistency hypothesis is also supported when the historical-based prior is used for $\pi_{\text{cons}}$ (see Tables 1 and S20). Nevertheless, the evidence is much weaker with posterior probability of inconsistency equal to $1 - 0.36 = 0.64$ and posterior odds equal to $0.64/(1 - 0.64) = 1.78$ indicating marginal evidence in favor of inconsistency. This weaker evidence in comparison with the other two scenarios ($\pi_{\text{cons}} = 0.25$ and $\pi_{\text{cons}} = 0.50$) is logical here since the historical-based prior for $\pi_{\text{cons}}$ a-priori supports the consistency hypothesis with prior mean probability equal to 0.78. Note also that the conclusion regarding network inconsistency is the same in the cases where $g = N/2$ and $g = 2N$, since the posterior inclusion/model probabilities are similar (see Figure S3). Therefore, we can say that our findings are in general agreement with the classical inconsistency tests and previous work which also suggests that the network is inconsistent. In addition, the Separate Indirect from Direct Design Evidence (SIDDE) method indicates that there is considerable difference between direct and indirect evidence (see Table S21), and the global Q test of inconsistency suggests the presence of global inconsistency ($X_1^2 = 8.44$, $P$-value = 0.004). Yuan et al also report that the Bucher test of inconsistency indicates the presence of inconsistency for the comparison Sildenafil vs Vardenafil ($P$-value <0.001).
Network inconsistency is also supported when 10% of the studies are randomly excluded. More specifically, we randomly excluded six studies comparing either Sildenafil, Vardenafil, or Tadalafil with Placebo; we were cautious not to exclude other studies to maintain the geometric integrity of the network. For each prior setup, the exclusion process was repeated 1000 times, and the inconsistent NMA model was estimated as the maximum a-posteriori model in all scenarios; only in the case where $\pi_{\text{cons}} \sim \text{Beta}(157, 44)$, they were observed few cases (21 out of 1000) in which the consistent NMA model was estimated as the most probable with a posterior model probability slightly above from 0.50 (see Figure 7 for more details).
4 | DISCUSSION

Consistency is a fundamental assumption of NMA, and it should be evaluated correctly to prevent resulting in false conclusions. In this article, we have viewed the NMA consistency problem as a variable selection problem. Under this perspective, we have implemented Bayesian variable selection via stochastic search variable selection algorithm in NMAs which was named SSIFS. SSIFS includes not only the implementation of the Gibbs based SSVS algorithm in NMAs, but also the appropriate prior specification of the model parameters and the use of two alternative prior approaches for the instancies in our NMA (non-informative and based on historical information).

A common practice for testing consistency in the Bayesian framework, is to compare the consistent NMA model with the fully inconsistent NMA model using the DIC. Although, DIC has been criticized for its limitations from several authors, its usage in NMA for model selection has been widely advocated. Limitations of DIC include concerns about discriminatory performance, invariant \( p_D \) to reparameterization, lack of consistency, not based on a proper predictive criterion and weak theoretical justification. Moreover, Millar noticed that DIC based on the conditional likelihood was invalid for hierarchical models.

A further limitation of the DIC is the absence of a threshold for judging important differences among models. Moreover, in our problem, DIC can be used only to compare two competing models at each time. Here, we may compare the consistent NMA with the full or a partially inconsistent NMA model. No overall evaluation metric for comparing consistent vs inconsistent models can be obtained in an obvious way. This is in contrast with our proposed approach in this article which (a) estimates the overall posterior weight of the hypothesis of consistency vs inconsistency and (b) important prior information from previous studies can be incorporated in the analysis. By this way, we can adjust our analysis for the multiple comparisons we consider here since many partially inconsistent models are included in our analysis and the evaluation of the consistency hypothesis.

So far, Bayesian tests for inconsistency evaluate the consistency assumption globally. SSIFS evaluates consistency assumption both globally and locally. If for example, we observe a posterior inclusion probability above 0.50 for an inconsistency factor or an inconsistent model with a large posterior model probability, then this is clear evidence in favor of the existence of local inconsistency in our network which indirectly causes global inconsistency in the network. Also, in SSIFS consistency could be evaluated by taking into consideration previous knowledge about how likely is to have a consistent network, by applying an informative prior for \( \pi_{cons} \). An informative prior for the probability of having a consistent network is proposed based on a review of 201 networks.

SSIFS also has the advantage that it gives us the option to evaluate the consistency assumption based on a priori difference between direct and indirect evidence that is of practical significance. This difference is depending on the outcome's nature. In continuous outcomes, for example, one unit difference between direct and indirect may be insignificant while in dichotomous outcomes a difference of this magnitude in log-scale is significant. This could be performed in SSIFS, by specifying properly the parameters \( c \) and \( \psi_r \).

In this work, inconsistency factors were specified using either the Lu and Ades model, the design-by-treatment interaction model, or Jackson’s model. It is important to note that matrix \( Z \) can also be specified manually. If for some reason one would like to place inconsistency factors in some comparisons and not in others, matrix \( Z \) can be adjusted appropriately. Another interesting feature of the SSIFS is that the Lu and Ades and the Jackson’s model can incorporate random inconsistency effects. This can be achieved in the absence of multi-arm studies, by setting \( R = I \) and assuming different \( \psi_r \) value for each inconsistency factor. However, when multi-arm studies are involved, formulating the spike and slab mixture of SSVS becomes challenging and requires further research.

The proposed method is shown to be robust in identifying consistency or inconsistency in the network when there is strong evidence for either case. This was demonstrated by two examples, where the results were consistent regardless of the prior choice, even in the extreme cases of favoring consistency or inconsistency. In the case where many studies are included in the network, as described in the erectile dysfunction example (see Section 3.2), SSIFS consistently identified the network as inconsistent regardless of the prior probability of consistency. Conclusions regarding network consistency remain the same even if 10% of the studies are excluded, as demonstrated by the sensitivity analysis. Similarly, in the smoking cessation example (see Section 3.1), using different approaches for specifying the inconsistency matrix \( Z \) yielded to similar conclusions. However, the sensitivity to the prior setting increases in cases where there are few studies or comparisons, such as in sparse networks, which is a common issue for any Bayesian method. In such cases, a highly informative prior may heavily affect the results of SSIFS. Furthermore, in our examples the results are not affected by the prior choice of parameter \( g \), since in all scenarios the posterior inclusion/model
probabilities are similar yielding to the same conclusions concerning network consistency (see Figures S1–S3). Overall, SSIFS performed well in both examples, but further research and a thorough simulation study are required to explore the performance of the method under various scenarios. We suggest using the prior distribution informed by previous NMAs $\pi_{\text{cons}} \sim \text{Beta}(157, 44)$ that assumes a 0.78 probability of consistency and we have set this as the default prior distribution in our R package. As sensitivity analyses, we suggest using prior distributions that give probabilities of consistency equal to 0.50.

SSIFS can be also used to evaluate consistency in individual participant data (IPD) NMA by using the two-stage approach. Extension of this work may include the use of SSIFS in the one-stage IPD-NMA. Moreover, it could be tailored to the network meta-regression framework or to multivariate NMA. However, in the one-stage IPD-NMA we expect the method would be computationally intensive and in the cases of network meta-regression the model would complicate considerably given that inconsistency would be imposed to the regression coefficients across treatment effects as well. Furthermore, other variable selection methods such as the Gibbs variable selection \cite{42,43} or the Bayesian LASSO \cite{44} could be utilized. Lastly, as a referee pointed out, instead of using traditional variable selection approaches and the spike and slab approach we recommend here, a shrinkage method like the horseshoe prior can be used instead. Specifically, the horseshoe mimics the spike and slab approach by a single continuous density.\cite{45} Horseshoe prior is used in cases where there few data points and many predictors.\cite{46} Therefore, it could provide valuable insight in sparse networks, where there are many inconsistency factors but few studies. Nevertheless, we lose the advantage of quantifying the probability of inconsistency or the inclusion of individual links in the network.

Our proposed method could be implemented through our R package \textit{ssifs},\cite{16} which is published on CRAN and can also be found on GitHub at \url{https://github.com/georgiosseitidis/ssifs}.

**DATA AVAILABILITY STATEMENT**

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

**ORCID**

\textit{Georgios Seitidis} \url{https://orcid.org/0000-0003-0856-1892}

\textit{Dimitris Mavridis} \url{https://orcid.org/0000-0003-1041-4592}

**REFERENCES**

1. Seitidis G, Nikolakopoulos S, Hennessy E, Tanner-Smith E, Mavridis D. Network meta-analysis techniques for synthesizing prevention science evidence. \textit{Prev Sci}. 2022;23(3):415-424. doi: 10.1007/s11121-021-01289-6

2. 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. Khanin R, ed. \textit{PLoS ONE}. 2014;9(1):e86754. doi: 10.1371/journal.pone.0086754

3. 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. \textit{Res Synth Methods}. 2012;3(2):80-97. doi: 10.1002/jrs.m.1037

4. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. \textit{J Clin Epidemiol}. 1997;50(6):683-691. doi: 10.1016/S0895-4356(97)00049-8

5. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. \textit{Stat Med}. 2010;29(7-8):932-944. doi: 10.1002/sim.3767

6. Efthimiou O, Rücker G, Schwarzer G, Higgins JPT, Egger M, Salanti G. Network meta-analysis of rare events using the mantel-Haenszel method. \textit{Stat Med}. 2019;38(16):2992-3012. doi: 10.1002/sim.8158

7. Krahn U, Binder H, König J. A graphical tool for locating inconsistency in network meta-analyses. \textit{BMC Med Res Methodol}. 2013;13(1):35. doi: 10.1186/1471-2288-13-13

8. Shih M, Tu Y. An evidence-splitting approach to evaluation of direct-indirect evidence inconsistency in network meta-analysis. \textit{Res Synth Methods}. 2021;12(2):226-238. doi: 10.1002/jrsm.1480

9. White IR. Network meta-analysis. \textit{Stata J}. 2015;15(4):951-985. doi: 10.1177/1536867X1501500403

10. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. \textit{Res Synth Methods}. 2012;3(2):98-110. doi: 10.1002/jrsm.1044

11. Jackson D, Barrett JK, Rice S, White IR, Higgins JPT. A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. \textit{Stat Med}. 2014;33(21):3639-3654. doi: 10.1002/sim.6188

12. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. \textit{J Am Stat Assoc}. 2006;101(474):447-459. doi: 10.1198/01621450500001302

13. Freeman SC, Fisher D, White IR, Auperin A, Carpenter JR. Identifying inconsistency in network meta-analysis: is the net heat plot a reliable method? \textit{Stat Med}. 2019;38(29):5547-5564. doi: 10.1002/sim.8383
APPENDIX A

A.1 Algorithm for the implementation of the lu and ades model

We take the following steps:

• Firstly, exclude the disconnected nodes (if they exist) so that we have a connected network.
• Assume that inconsistency factors could be added in all functional parameters.
• Specify all closed-loops of the network by taking all the possible paths that start and end in the same node.
• Check if multi-arm studies are present in the network.
• If multi-arm studies are present: Exclude inconsistency factors from functional parameters present only in one multi-arm trial (that are by definition consistent)
• Exclude the bridges of the network by checking which functional parameters are not included in any closed loop.
• Identify the independent loops; a loop is considered independent if it contains at least one edge which is not a part of any other independent loop.
• Keep only the functional parameters that belong to independent loops. Note: Each independent loop must contribute one inconsistency factor, thus for each independent loop we keep only one functional parameter.
• The remaining functional parameters are the comparisons in which inconsistency factor should be added.

A.2 Tuning based on assuming a priori the minimum value of inconsistency that is of practical significance

Tuning parameters $\psi_{\ell}$ and $c$ in order inconsistency factor to be included in the NMA model when $|b_{\ell}|$ is larger than a threshold. In the case where the prior correlation matrix is $R = I_p$, we have that

$$f(b_\ell | \psi_\ell = 0) = f(b_\ell | \psi_\ell = 1),$$

$$\frac{1}{\psi_\ell \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{b_\ell}{\psi_\ell} \right)^2} = \frac{1}{c\psi_\ell \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{b_\ell}{c\psi_\ell} \right)^2},$$

$$-\frac{1}{2} \left( \frac{b_\ell}{\psi_\ell} \right)^2 = -\log(c) - \frac{1}{2} \left( \frac{b_\ell}{c\psi_\ell} \right)^2,$$

$$c^2 b_\ell^2 = 2c^2 \psi_\ell^2 \log(c) + b_\ell^2,$$

$$b_\ell^2 (c^2 - 1) = 2c^2 \psi_\ell^2 \log(c),$$

$$b_\ell^2 = \frac{2c^2 \psi_\ell^2 \log(c)}{c^2 - 1},$$

$$b_\ell^2 = \psi_\ell^2 \xi(c), \xi(c) = \frac{2c^2 \log(c)}{c^2 - 1}. $$
Inconsistency factors with coefficient $|b_x| > \psi_x \sqrt{\xi(c)}$ will have larger density to be included in the NMA model. If we assume that the elements of $b$ are correlated then,

$$f(b|Y = 0) = f(b|Y = 1),$$

$$|2\pi \; D_{r=0}RD_{r=0}|^{-1/2} e^{-\frac{1}{2} b' (D_{r=0}RD_{r=0})^{-1} b} = |2\pi \; D_{r=1}RD_{r=1}|^{-1/2} e^{-\frac{1}{2} b' (D_{r=1}RD_{r=1})^{-1} b}.$$  

We have

$$D_{r=1} = \begin{pmatrix} c \psi_1 & \cdots & c \psi_p \\ \vdots & \ddots & \vdots \\ c \psi_p \\ \end{pmatrix} = cD_{r=0}. $$

Thus,

$$|D_{r=0}RD_{r=0}|^{-1/2} e^{-\frac{1}{2} b' (D_{r=0}RD_{r=0})^{-1} b} = \left( c^2 \right)^{-\frac{1}{2}} |D_{r=0}RD_{r=0}|^{-1/2} e^{-\frac{1}{2} b' (D_{r=0}RD_{r=0})^{-1} b},$$

$$e^{-\frac{1}{2} b' (D_{r=0}RD_{r=0})^{-1} b} = c^{-p} e^{-\frac{1}{2} b' (D_{r=0}RD_{r=0})^{-1} b},$$

$$-\frac{1}{2} b' (D_{r=0}RD_{r=0})^{-1} b = -p \log c - \frac{1}{2c^2} b' (D_{r=0}RD_{r=0})^{-1} b,$$

$$b' (D_{r=0}RD_{r=0})^{-1} b \left( \frac{1}{c^2} - 1 \right) = -2p \log (c),$$

$$b' (D_{r=0}RD_{r=0})^{-1} b = \frac{2pc^2 \log (c)}{c^2 - 1}.$$

Zellner $g$-prior is assumed for the prior correlation matrix $R$

$$R = g(Z'Z)^{-1} \sigma^2.$$  

Hence,

$$b' \left[ D_{r=0}(Z'Z)^{-1}D_{r=0} \right]^{-1} b = \frac{2g\sigma^2c^2 \log (c)}{c^2 - 1}.$$  

Thus, it is considered unsignificant when

$$b' \left[ D_{r=0}(Z'Z)^{-1}D_{r=0} \right]^{-1} b \leq g\sigma^2 \xi(c).$$