Received: 30 October 2023 | Revised: 22 March 2024 | Accepted: 22 March 2024

DOI: 10.1002/jrsm.1718

RESEARCH ARTICLE

Research Synthesis Methods WILEY

Advancing unanchored simulated treatment comparisons: A novel implementation and simulation study

Shijie Ren[1] | **Sa Ren**[1] | **Nicky J. Welton**[2] | **Mark Strong**[1]

[1]School of Medicine and Population Health, University of Sheffield, Sheffield, UK

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

Correspondence
Shijie Ren, Sheffield Centre for Health and Related Research (SCHARR), School of Medicine and Population Health, University of Sheffield, Sheffield S1 4DA, UK.
Email: s.ren@sheffield.ac.uk

Funding information
National Institute for Health and Care Research; NIHR Advanced Fellowship, Grant/Award Number: NIHR300590

Abstract

Population-adjusted indirect comparisons, developed in the 2010s, enable comparisons between two treatments in different studies by balancing patient characteristics in the case where individual patient-level data (IPD) are available for only one study. Health technology assessment (HTA) bodies increasingly rely on these methods to inform funding decisions, typically using unanchored indirect comparisons (i.e., without a common comparator), due to the need to evaluate comparative efficacy and safety for single-arm trials. Unanchored matching-adjusted indirect comparison (MAIC) and unanchored simulated treatment comparison (STC) are currently the only two approaches available for population-adjusted indirect comparisons based on single-arm trials. However, there is a notable underutilisation of unanchored STC in HTA, largely due to a lack of understanding of its implementation. We therefore develop a novel way to implement unanchored STC by incorporating standardisation/marginalisation and the NORmal To Anything (NORTA) algorithm for sampling covariates. This methodology aims to derive a suitable marginal treatment effect without aggregation bias for HTA evaluations. We use a non-parametric bootstrap and propose separately calculating the standard error for the IPD study and the comparator study to ensure the appropriate quantification of the uncertainty associated with the estimated treatment effect. The performance of our proposed unanchored STC approach is evaluated through a comprehensive simulation study focused on binary outcomes. Our findings demonstrate that the proposed approach is asymptotically unbiased. We argue that unanchored STC should be considered when conducting unanchored indirect comparisons with single-arm studies, presenting a robust approach for HTA decision-making.

KEYWORDS

indirect treatment comparison, marginal treatment effect, population adjustment, unanchored simulated treatment comparison

Highlights

What is already known

- Population-adjusted indirect comparisons are increasingly used to account for differences in patient characteristics between two treatments evaluated in different trials.
- Unanchored population-adjusted indirect comparisons such as matching-adjusted indirect comparison (MAIC) and simulated treatment comparison (STC), are more commonly used compared to anchored indirect comparisons due to evidence involving single-arm trials.
- There is a notable underutilisation of unanchored STC in HTA, largely because of challenges in understanding and implementing of such method.

What is new

- We developed a novel way to implement unanchored STC using standardisation/marginalisation and the NORTA algorithm for sampling covariates. This methodology ensures the derivation of a suitable marginal treatment effect without aggregation bias for health technology assessment evaluations.
- We conducted the first simulation study evaluating the performance of unanchored STC.
- Our comprehensive simulation study establishes proof-of-principle, demonstrating that our proposed unanchored STC approach is asymptotically unbiased and provides good coverage.

Potential impact for *Research Synthesis Methods* readers

- The unanchored STC approach should be considered for population-adjusted indirect comparisons. Care needs to be taken in the implementation of such method to ensure the derivation of unbiased estimate for the marginal treatment effect and appropriately quantified uncertainty associated with it.

1 | INTRODUCTION

Randomised controlled trials (RCTs) are considered to be the gold standard for evaluating a treatment or intervention because the randomisation eliminates confounding bias. However, an RCT may not be possible due to practical or ethical concerns which leads to the existence of 'single-arm trials' in which all patients receive the same treatment. A comparison is then made with (possibly aggregate) outcome data from a separate study that tested the comparator treatment. A review showed that 76 unique indications were granted approval by the European Medicines Agency (EMA) and the United States Food and Drug administration (FDA) without RCT evidence during 1999–2014, and the majority were for haematological malignancies, oncology and metabolic conditions.[1]

Given randomised evidence with a common comparator, anchored indirect treatment comparison could be performed. However, in the case that evidence is from single-arm trials, there is no common comparator. The simplest method to incorporate single-arm trial data to obtain a relative treatment effect is to perform a naïve unanchored indirect comparison (the indirect comparison without a common comparator), where individual arms of different studies are naively compared with each other. This method is prone to bias due to a lack of randomisation, which could lead to different patient populations in different studies and no control for confounding. When individual patient-level data (IPD) are available for both individual arms of different studies, methods such as regression adjustment, propensity score and doubly robust methods have been proposed to reduce the bias caused by selection on observables (i.e., incorrectly omitting observed variables that determine both the treatment and the outcome).[2] Instrumental variable methods and panel data models could be used to reduce bias caused by selection on unobservables.[2]

Health technology assessment (HTA) is a multidisciplinary process to evaluate the clinical, economic and broader impact of the use of a health technology. In a technology appraisal, it is common that the evidence for the efficacy of the comparator treatment is available only as a set of published summary statistics (i.e., as aggregate data), and that IPD for the treatment of interest are available only to the pharmaceutical company that developed the treatment. The lack of individual-level data for the comparator treatment restricts the types of statistical analysis which could be used to estimate the relative treatment effect.

Two main approaches that have been used in HTA to attempt to adjust for confounders in the case of only aggregate data from the comparator studies are the matching-adjusted indirect comparison (MAIC)[3,4] and the simulated treatment comparison (STC).[4,5] MAIC is based on propensity score weighting and STC is based on outcome regression adjustment. Both approaches adjust imbalance in observed covariates and provide an estimate of relative treatment effect which reflects the effect in the trial population with aggregate data (i.e., the comparator study). In the case where the evidence was from single-arm trials without a common comparator, an unanchored comparison would be made. An unanchored MAIC and STC assume that all prognostic factors (variables that affect outcome) and effect modifiers (variables that alters the effect of treatment on outcomes) are accounted for, which is largely considered impossible to meet.[6] Neither method is able to adjust for bias due to residual confounding, which is the distortion that remains after controlling for imbalance in observed covariates.

Another limitation of MAIC and STC is that both methods are limited to pairwise indirect treatment comparisons (i.e., two treatments [or three treatments in anchored comparisons] from two studies) and limited to predict the treatment effect for the comparator study population. Phillippo et al. (2020) proposed a new method (multilevel network meta-regression [ML-NMR]) to overcome these limitations.[7] ML-NMR can be applied to treatment networks of any size and can provide the treatment effect in any target population given covariates. However, ML-NMR is currently only applicable for anchored indirect treatment comparisons.

As MAIC and STC are based on different statistical approaches to adjust for population differences between the two studies, they could lead to different treatment estimations due to different estimands (marginal vs. conditional) used in the analysis. The marginal estimands can be estimated from an unadjusted analysis by not including any covariates in the regression model. The conditional estimands can be estimated from an adjusted analysis by including covariates in the regression model. However, 'marginal'/'conditional' and 'unadjusted'/'adjusted' should not be used interchangeably because marginal treatment effects could also be derived from the covariate-adjusted analyses.[8]

MAIC is a propensity score weighting approach, hence always estimates a marginal or population-average treatment effect. That is the average effect at the population level. Note that this is the case for both anchored and unanchored MAIC. STC is an outcome regression-based modelling approach, which relies on regression models with covariates included. For the anchored STC, the conventional use estimates a conditional treatment effect. That is the average effect conditional on the effects of the covariates included in the regression model. A marginalisation method was proposed to ensure estimating marginal effect.[9] For unanchored STC, it always estimates a marginal effect because the regression model used would not have treatment as a covariate since the study is single-arm by design. The average treatment effect estimated is for the entire population in the comparator study.

Since the publication of the NICE Decision Support Unit (DSU) Technical Support Document (TSD) on population-adjusted indirect comparisons methods,[6] several simulation studies have been conducted to evaluate the different population-adjusted indirect comparison methods.[10–15] However, no simulation study evaluated the performance of unanchored STC. In addition, not all the simulation studies investigated the performance of MAIC and STC simultaneously and there was a lack of consensus on the superiority of performance where they were both included in the simulation study. The discrepancy in the conclusions could be due to the different scenario settings used between the studies. However, all the simulation studies conclude that covariates selection is important, and all adjustment methods would lead to bias if not all covariates need to be adjusted for are included in the model.

Reviews show that indirect comparisons in HTA were mostly unanchored and MAIC was used way more frequently than STC.[16–18] We believe that the more frequently use of MAIC is not because the performance of unanchored STC is worse than unanchored MAIC, but purely due to the lack of awareness on how to conduct unanchored STC. MAIC's reweighting procedure to balance population differences is intuitive and given the reweighted data standard statistical analysis methods can be used to obtain the indirect treatment effect. On the other hand, it is unclear how to interpret and then use the results from the outcome regression model in the STC approach. The above discussion on marginal vs. conditional estimands also highlights that depending

on whether it is anchored or unanchored STC, the interpretation could be different.

Under correct model specification, STC approach would be more efficient than MAIC approach because regression-based approaches give more precise estimates. This advantage becomes more notable where overlap is poor between the studies, and MAIC suffers from a large reduction in the effective sample size after weighting. In an extreme case of a lack of covariate overlap (large number of covariates and small sample size) MAIC may fail to produce feasible weights; whereas STC would still be feasible due to its ability to extrapolate beyond the observed covariate space in the IPD study. The STC approach also has the advantage that the model assumptions could be checked explicitly.[19]

In this paper, we aim to present a new methodology to perform unanchored STC, which avoids aggregation bias. We also aim to make the unanchored STC method more accessible to users so that this approach can readily be used in practice. We propose to use the NORmal To Anything (NORTA) algorithm to simulate multivariate, non-normal covariates to address the challenges in simulating categorical variables raised by Ishak et al. (2015).[4] We also present a non-parametric bootstrapping procedure to obtain the correct standard error for the estimated relative treatment effect. We conduct a comprehensive simulation study to evaluate the performance of the proposed unanchored STC method to fill the evidence gap in the literature.

2 | METHODS

We assume that we are interested in estimating the treatment effect of treatment B versus treatment A, where we have IPD from a single-arm study for treatment B (the B study) and aggregate data from another single-arm study for treatment A (the A study). STC is a form of outcome regression approach, where a regression model is fitted to the IPD in the B study:

$$g\left(\theta_{i(B)}(\boldsymbol{X_i})\right) = \beta_0 + \boldsymbol{\beta}_1^{\top}\boldsymbol{X_i}, \tag{1}$$

where $\theta_{i(B)}(\boldsymbol{X_i})$ is the expected outcome for individual i with covariate values $\boldsymbol{X_i}$ in the B study (e.g., the probability for binary outcomes); the subscript (B) indicates the population; g is an appropriate link function (e.g., the logit function for binary outcomes); β_0 is the intercept, $\boldsymbol{\beta_1}$ is a vector of coefficients for prognostic factors and effect modifiers; and $\boldsymbol{X_i}$ is the full covariate vector including prognostic factors and effect modifiers for individual i. This formula is a simplified version of the standard STC formula presented in Phillippo et al. (2018)[6] as the regression model does not have the treatment group as a covariate in an unanchored STC.

Once the regression model coefficients in Equation (1) have been estimated, the second step in the STC is to predict the outcome for the population in the A study where only aggregate data are available. The predicted average effect of treatment B in the A study population $\widehat{d}_{B(A)} = g\left(\widehat{\theta}_{B(A)}\right)$ is obtained via marginalisation/standardisation of the predicted conditional estimates for the sampled individuals in the A study. This calculation is explained in detail in the next section. Given the predicted average effect of treatment B in the A study population and the reported treatment effect of treatment A in the A study population $\widehat{d}_{A(A)}$, an unanchored STC produces an estimate of the relative treatment effect of B versus A in the A study population:

$$\widehat{d}_{AB(A)} = \widehat{d}_{B(A)} - \widehat{d}_{A(A)}. \tag{2}$$

Note that the indirect comparison in Equation (2) must be formed on the linear predictor scale because the impact of an effect modifier is scale dependent.[6] For example, for a binary outcome $\widehat{d}_{B(A)}$ is the estimated log odds of receiving treatment B and $\widehat{d}_{A(A)}$ is the estimated log odds of receiving treatment A, both for the A study population. Note that the A study may not report $\widehat{d}_{A(A)}$ on the linear predictor scale direct and transformation is required. For example, for a binary outcome, the A study may have reported probability of experiencing the outcome $\widehat{\theta}_{B(A)}$. This needs to be transformed to the linear predictor scale (log odds) before using it in Equation (2).

2.1 | The prediction step

For linear regression models with an identity link function for continuous outcome, $\widehat{d}_{B(A)}$ can be obtained by plugging in the mean values of the $\boldsymbol{X}$ reported from the A study in Equation (1), $\overline{\boldsymbol{X}}$. The estimate of the treatment effect of B versus A in the comparator study population for Equation (2) becomes

$$\begin{aligned}\widehat{d}_{\mathrm{AB}(A)} &= \widehat{d}_{B(A)} - \widehat{d}_{A(A)} \\ &= \left(\widehat{\beta}_0 + \hat{\boldsymbol{\beta}}_1^{\top}\overline{\boldsymbol{X}}_{(A)}\right) - \widehat{\theta}_{A(A)}.\end{aligned} \tag{3}$$

This 'plugging-in' approach would lead to aggregation bias for models with non-identity link function.[7,20] This is because there is non-linearity between the outcome and covariates $\boldsymbol{X}$ (i.e., the mean of predicted outcome is not the same as the predicted outcome calculated

at the mean of $\boldsymbol{X}$). By applying Bayes' theorem, Chang et al. (2000) shows that the size of the aggregation bias is $\left[\log f\left(\boldsymbol{x}_{ij}|y_{ij}=1\right)-\log f\left(\boldsymbol{x}_{ij}|y_{ij}=0\right)\right]$ for a logit link function, where $f\left(\boldsymbol{x}_{ij}|y_{ij}\right)$ is the probability density function for covariates $\boldsymbol{x}_{ij}$ given binary outcome y_{ij}.[21]

To deal with non-linearity, an alternative approach is to simulate individuals with a range of covariate values, $\boldsymbol{X}$, to match the population in the A study, and then obtain the adjusted absolute effect on the linear predictor scale (i.e., the log odds for binary outcomes) by averaging the predictions of these individuals. Ishak et al. (2015) suggests to simulate the same number of individuals as the sample size of the comparator study.[4] A simulation study from Zhang et al. (2023) shows that there is an under-coverage problem when maintaining the sample size in simulating the comparator study, and nominal coverage is achieved when increasing the sample size to infinity (100,000).[22] Remiro-Azócar et al. (2022) proposes that the sample size does not have to correspond to the sample size of the comparator study, but the size needs to be sufficiently large to ensure stability and minimise sampling variability.[9] We opted for choosing a sufficiently large sample size.

We now describe the procedure step-by-step using a binary outcome as an example. Let the outcome Y be binary. The expected outcome $\theta_{B(A)}$ in Equation (1) becomes the probability of experiencing the outcome Y when treated with treatment B in the A study population:

$$\begin{aligned}\theta_{B(A)} &= P(Y=1)\\ &= E[P(Y=1|\boldsymbol{X})]\\ &= \int P(Y=1|\boldsymbol{X}) f_{\boldsymbol{X}}(X) d\boldsymbol{X},\end{aligned} \tag{4}$$

where $f_{\boldsymbol{X}}(\boldsymbol{X})$ is the joint probability density function for X representing the A study population if $\boldsymbol{X}$ contains all continuous covariates; $f_{\boldsymbol{X}}(\boldsymbol{X})$ is the joint probability mass function if $f_{\boldsymbol{X}}(\boldsymbol{X})$ contains all discrete covariates; and is a joint density function with respect to an appropriate dominating measure if $\boldsymbol{X}$ is a mixture of continuous and discrete covariates. Equation (4) can be evaluated using Monte Carlo integration with random samples $\boldsymbol{X_j}$ from $f_{\boldsymbol{X}}(\boldsymbol{X})$:

$$\widehat{P}_{B(A)}(Y=1) = \frac{1}{N}\sum_{j=1}^{N} P\left(Y=1|\boldsymbol{X_j}\right). \tag{5}$$

The adjusted relative treatment effect in the A study population, $\widehat{d}_{AB(A)}$, can be obtained using Equation (2) given the predicted probability from Equation (5) and the reported probability when receiving treatment A from the A study ($\widehat{P}_{A(A)}(Y=1)$):

$$\begin{aligned}\widehat{d}_{\mathrm{AB}(A)} &= \widehat{d}_{B(A)} - \widehat{d}_{A(A)}\\ &= \mathrm{logit}\left(\widehat{p}_{B(A)}\right) - \mathrm{logit}\left(\widehat{p}_{A(A)}\right)\\ &= \log\left(\frac{\widehat{P}_{B(A)}(Y=1)}{1-\widehat{P}_{B(A)}(Y=1)}\right) - \log\left(\frac{\widehat{P}_{A(A)}(Y=1)}{1-\widehat{P}_{A(A)}(Y=1)}\right).\end{aligned} \tag{6}$$

The general formula for the estimator $\widehat{d}_{B(A)}$ is

$$\hat{d}_{B(A)} = g\left(\frac{1}{N}\sum_{j=1}^{N} g^{-1}\left(\hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1^{\top}\boldsymbol{X}_{\boldsymbol{j}(\boldsymbol{A})}\right)\right), \tag{7}$$

where we firstly obtain the predicted outcome on the natural scale for all N simulation samples sampled from the joint covariate distribution, $g^{-1}\left(\hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1^{\top}\boldsymbol{X}_{\boldsymbol{j}(\boldsymbol{A})}\right)$, (e.g., the probability for binary outcomes); then we obtain the average predicted outcome on the natural scale (e.g., the average probability for binary outcomes); finally we transform the predicted outcome to the linear predictor scale using g (e.g., the log odds for binary outcomes).

This is a similar approach to the method used to estimate the marginal causal log odds ratio from analyses that adjust for covariates proposed by Zhang (2008),[23] which was also illustrated by Daniel et al. (2019).[8] Remiro-Azócar et al. (2022) proposed using this approach for anchored STC so that the estimand is the marginal effect.[9] In the literature, this approach has been variously named standardisation, marginalisation or G-computation.

The difference between the standardisation method used for unanchored STC and G-computation is that for unanchored STC we only apply the standardisation for arm B and use the reported effect for the arm A as the marginal treatment effect for arm A. This is because the regression model in Equation (1) does not have treatment as a covariate as there is only a single arm, B.

When all covariates are continuous and normally distributed, the random samples of $\boldsymbol{X_j}$ needed in Equation (7) could be generated from a multivariate normal distribution with the observed means $\overline{\boldsymbol{X}}$ and the correlation structure observed in the B study where IPD are available as the comparator study would normally not report the correlation structure. If data are not normally distributed, then an appropriate transformation is required before sampling. When the covariates contain discrete variables, it would not be appropriate to sample from a multivariate normal distribution. We introduce the NORmal To Anything (NORTA) algorithm to sample the joint covariate distribution in Section 2.2, which allows us to handle the situation where covariates are discrete or a mixture of continuous and discrete.

2.2 | NORTA for sampling covariates

One of the most popular approaches to simulate multivariate, non-normal data in the life or social sciences is the NORmal To Anything (NORTA) approach proposed by Cario and Nelson (1997).[24] The idea is to sample from a multivariate normal distribution and then transform the sampled multivariate normal variables into variables with other marginal distributions. This approach has a long history in statistics and simulation and can be dated back in 1970s.[25,26] Cario and Nelson (1997) extended the idea to discrete/mixed marginal distributions.[24]

To simulate a random vector $\boldsymbol{X} = (X_1, ..., X_k)$ with the following properties

1. $X_i \sim F_{X_i}$, $i = 1, ..., k$ and F_{X_i} is the cumulative distribution function (CDF) for X_i; and
2. $\text{Corr}(\boldsymbol{X}) = \Sigma_X$.

The NORTA algorithm proceeds as follows:

1. Simulate from $\boldsymbol{Z} = (Z_1, ..., Z_k)$, where Z_i follows a multivariate normal distribution with mean 0 and correlation matrix Σ_Z, $i = 1, ..., k$, that is $Z_i \sim MVN(0, \Sigma_Z)$.
2. Apply the probability integral transformation to sampled Z_i, such that $U_i = \Phi(Z_i)$, where $\Phi(.)$ is the standard normal CDF and U_i follows a standard uniform distribution, $U_i \sim U[0,1]$.
3. $\boldsymbol{X}$ is obtained using $X_i = F_{X_i}^{-1}(U_i)$, where $F_{X_i}^{-1}(.)$ is the inverse CDF of X_i.

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_i \\ \vdots \\ Z_k \end{pmatrix} \underset{\rightarrow}{X_i = F_{X_i}^{-1}(\Phi(Z_i))} \begin{pmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_k. \end{pmatrix}$$

The NORTA method is also known as a Gaussian copula method.[27–29] This approach was used to model the joint distribution for covariates in ML-NMR and parametric G-computation approach for population-adjusted treatment comparisons in anchored indirect comparisons.[7,9]

One key point to note is that for $i \neq j$, the correlation $\rho_X(i,j)$ depends exclusively on $\rho_Z(i,j)$, and in general $|\rho_X(i,j)| \leq |\rho_Z(i,j)|$.[24,30] When applying NORTA to simulate covariates for the comparator study with aggregate data, we need to specify the marginal distributions given the summary statistics (such as mean and standard deviation for a continuous covariate and percentage for a binary covariate) reported from the comparator study and the correlation matrix Σ_X. We assume that the correlation structure is the same between the comparator study and the company's study with IPD, that is $\Sigma_X^{AgD} = \Sigma_X^{IPD}$, and we choose the form of the marginal distribution for the covariates in the comparator study based on assumptions about the sampling distribution of the covariate (e.g., a Bernoulli distribution for a binary covariate).

If we use Σ_X^{IPD} in step 1 of NORTA algorithm, the sampled covariates will not have the desired correlation matrix as Σ_X^{IPD} and instead the correlations would be smaller for the comparator study than the study with IPD, $|\rho_X^{AgD}(i,j)| \leq |\rho_X^{IPD}(i,j)|$. In both ML-NMR and parametric G-computation approach for anchored indirect comparisons, the impact of not simulating covariates for the comparator with the desired correlation matrix was not discussed explicitly. A simulation study by Phillippo et al. (2020) investigated the impact of correlation between covariates for anchored STC and concluded that the performance of anchored STC is not affected by the correlation between covariates.[10] We also propose to use NORTA/Gaussian copula to sample covariates for the comparator study given its convenience in obtaining the samples for multiple data types. We will evaluate the impact of not obtaining the desired correlation for the covariates for the comparator study in unanchored case in a simulation study (Section 3).

2.3 | Bootstrap for estimating standard error

We propose to use the non-parametric bootstrap method[31] to compute the variance of $\widehat{d}_{AB(A)}$ in Equation (7) due to the lack of a closed-form expression for this variance. Because only the IPD from the B study with treatment B can be resampled, we propose to calculate the variance of $\widehat{d}_{AB(A)}$ by decomposing this variance into two parts (the variance for $\widehat{d}_{B(A)}$ and the variance for $\widehat{d}_{A(A)}$), where the variance for $\widehat{d}_{B(A)}$ is computed using the bootstrap approach and the variance of $\widehat{d}_{A(A)}$ is computed using the reported summary statistics from the study with aggregate data (for example, with binary data there is a closed-form formula of the variance for the log odds).

$$\begin{aligned} Var\left(\widehat{d}_{AB(A)}\right) &= Var\left(\widehat{d}_{B(A)}\right) + Var\left(\widehat{d}_{A(A)}\right) \\ &= \widehat{V}\left(\widehat{d}_{B(A)}\right) + Var\left(\widehat{d}_{A(A)}\right), \end{aligned}$$

where $\widehat{V}\left(\widehat{d}_{B(A)}\right)$ is the variance for $\widehat{d}_{B(A)}$ obtained using the bootstrap approach. The bootstrap procedure proceeds as follows:

1. Draw a bootstrap sample $P_1^*,...,P_n^*$ from the IPD study. Fit an appropriate regression model to the bootstrap sample $P_1^*,...,P_n^*$. Predict the outcome for the study A population. Compute $\widehat{d}_{B(A)}^*$ using Equation (7).
2. Repeat step 1 for M times, which yields to $\widehat{d}_{B(A),1}^*,...,\widehat{d}_{B(A),M}^*$.
3. Compute the bootstrap mean: $\overline{d}_{B(A)}^* = \frac{1}{M}\sum_{j=1}^{M}\widehat{d}_{B(A),j}^*$.
4. Compute the bootstrap variance: $Var\left(\widehat{d}_{B(A)}\right) = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(\widehat{d}_{B(A),j}^* - \overline{d}_{B(A)}^*\right)}$.

2.4 | Sparse data bias in generalised linear model

The maximum likelihood estimates (MLEs) for generalised linear models may be biased when there are few or no study participants at key combinations of the outcome and covariates. The bias, sometime called sparse data bias tends to move away from the null (i.e., a downward bias when the estimate is below 1 and an upward bias when the estimate is above 1).[32] This bias is often ignored in practice under the assumption that the bias is negligible compared with the standard error of the estimate.[33] However, a simulation study showed that when the number of events is relatively small and there is large imbalance in the levels of a covariate, an odds ratio estimated from a logistic regression model using the maximum likelihood method would have a *large* bias.[34]

Firth-type penalisation has been used in practice to reduce the small sample bias of the coefficients derived using the maximum likelihood method.[35] Puhr et al. (2016) has shown that while bias in the estimates of the coefficients is reduced using a penalisation approach, it comes at the cost of introducing bias in the prediction of probabilities.[36] They also illustrate that the maximum likelihood method provides an average predicted probability equal to the observed proportion of events observed in a logistic regression.[36] Because the STC method involves predicting the probability for the comparator study, we propose to use the maximum likelihood method without penalisation. Our simulation study also investigates the impact of sparse data bias on the performance of the STC approach when using a generalised linear model.

3 | SIMULATION STUDY

The design of this simulation study follows a structured approach proposed by Morris et al. (2019),[37] which involves specifying aims, data-generating mechanisms, methods, estimands and performance measures.

All simulations and analyses were performed using R software version 4.2.2. Program code file for data simulation and analyses can be found at https://github.com/SRenScharr/unanchored-simulated-treatment-comparison. The GitHub page also contains an example analysis to demonstrate the use of the proposed method in practice.

3.1 | Aims

This simulation study aims to (i) assess the impact of simulated covariates for the comparator study using the NORTA algorithm do not have the desired correlation matrix in unanchored STC, (ii) assess the impact of spare data bias on the performance of the unanchored STC when the covariates included in the model have large imbalance within the strata. For completeness, the 'plugging-in' of the mean covariates approach is also performed.

3.2 | Data-generating mechanisms

In this simulation study, we consider a binary outcome with two covariates representing the full set of known prognostic factors and effect modifiers and an additional covariate indicating treatments (A and B). Data are simulated for two studies from a logit model: the IPD study and the aggregate data study, where each study has two treatment groups. We compute the summary statistics for the aggregate data study and only take the A arm of the aggregate data study and the B arm of the IPD study to the analysis.

The logit model with interaction follows the form

$$y_{ij} \sim \text{Bern}\left(\theta_{ij}\right)$$

$$\text{logit}\left(\theta_{ij}\right) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 \times trt_{ij} + \beta_4 \times trt_{ij} \times x_{ij2}\mathbb{I}$$

where y_{ij} denotes individual i in study j which is generated from a Bernoulli distribution with probability θ_{ij}; x_{ij1} and x_{ij2} denote the two covariates; trt_{ij} is the treatment indicator which is 1 if treated with treatment B and 0 if treated with treatment A; $\mathbb{I}$ is the indicator function which is 1 when there is an interaction between covariate x_2 and treatment and is 0 otherwise. We set $\beta_0 = -0.25$, $\beta_1 = 0.09$ and $\beta_2 = 0.15$.

We consider the situation where both covariates are binary and also consider a series of scenario analysis to explore if the unanchored STC approach is sensitive to the magnitude of the difference in the marginal means of

covariates within each study and between studies, the strength of the correlation between the covariates, the strength of overlap between studies, the strength of treatment and covariate interaction β_4 (i.e., effect modification), the strength of the treatment effect β_3, and the sample size per arm.

In total, 120 scenario analyses were explored in this simulation study with the parameters varied in the following way:

1. Magnitude of the marginal means and strength of overlap between studies:
   i. small for both studies (0.1 for aggregate data study vs. 0.2 for IPD study, large overlap);
   ii. large for both studies (0.9 for aggregate data study vs. 0.8 for IPD study, large overlap);
   iii. moderate for both studies (0.5 for aggregate data study vs. 0.6 for IPD study, large overlap);
   iv. small for one and large for the other (0.2 for aggregate data study vs. 0.8 for IPD study, small overlap).

A crude measure based on the difference between the two binary variables is used to indicate strength of overlap: a small difference of 0.1 for large overlap and a large difference of 0.6 for small overlap.

2. Strength of correlation between the covariates (ρ_Z is the correlation for the multivariate normal variables in step 1 of NORTA. The actual correlation between the two covariates for each scenario can be found in the online supporting material Appendix 1 Table 1.):
   i. weak in both studies ($\rho_Z = 0.2$);
   ii. strong in both studies ($\rho_Z = 0.9$);
   iii. weak in one and strong in the other ($\rho_Z = 0.2$ for aggregate data study and $\rho_Z = 0.9$ for IPD study).
3. Strength of effect modification:
   i. no effect modification $\beta_4 = 0$;
   ii. small effect modification $\beta_4 = -0.1$;
   iii. large effect modification $\beta_4 = -0.3$.
4. Strength of treatment effect
   i. large treatment effect $\beta_3 = -0.45$;
   ii. small treatment effect $\beta_3 = -0.1$.
5. Sample size:
   i. 200;
   ii. 1000.

The sample size of 200 reflects the typical size of studies in submissions used population-adjustment approaches to HTA authorities.[38] The sample size of 1000 is used to test the asymptotic properties of the proposed STC approach.

3.3 | Estimands

The estimand of interest is the marginal log-odds ratio in the aggregate data study population. Because of the non-collapsibility issue with the odds ratio, there is no closed form for calculating the true marginal logodds ratio. In the simulation study, we calculate the true log-odds ratio by conducting a simple logistic regression with treatment as the covariate using the simulated data for both arms (setting the sample size to be 5 million per arm) in the aggregate data study.

3.4 | Methods

Each simulated dataset is analysed using the proposed unanchored STC approach described in Section 2 where covariate distribution for the aggregate data study is simulated using the NORTA algorithm given the marginal means from the aggregate data study and assuming the correlation structure of the covariates in the aggregate data study is the same as the IPD study. 1000 bootstrap samples were used to derive the appropriate standard error of the treatment effect. 10,000 individual profiles were simulated in the prediction step.

3.5 | Performance measure

2000 Monte Carlo replicates were generated for each scenario. The performance of the methods is evaluated using the bias, empirical standard error, model standard error and coverage probability. The bias provides a measure of the accuracy of the unanchored STC method and is computed as the average difference between the estimate from each repetition and the truth. The empirical standard error measures the true variability of the estimate and is computed as the standard error of the repeated estimates. The model standard error provides an estimate of the empirical standard error and is computed as the square root of the average estimate variance of the treatment estimate from each repetition. The coverage probability measures the probability that the confidence intervals contain the true value and is computed as the proportion of repetitions with its 95% confidence interval (CI) contains the truth. We consider that a good method would be unbiased (the bias is close to zero) with small empirical standard error, is well estimated by the model empirical error, and the coverage probability is close to 95%.

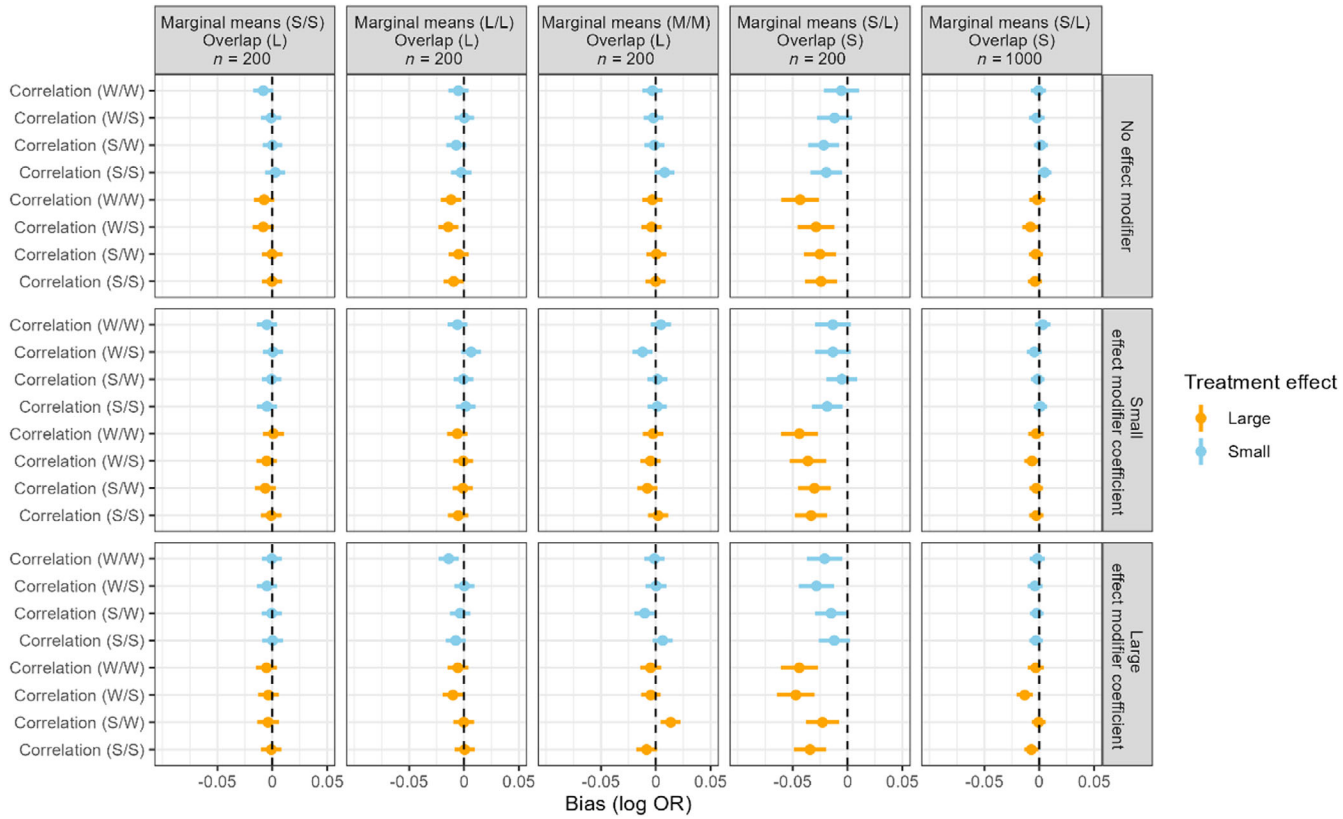

FIGURE 1 Bias along with 95% Monte Carlo confidence intervals for different scenarios. (S, M and L for marginal means indicate small, moderate and large marginal mean. S and L for overlap indicate small and large overlap. W and S for correlation indicate weak and strong correlation.)

3.6 | Results

Figures 1–3 show the simulation results for all 120 scenarios. The results are also tabulated in Appendix 1 Tables 1–5. Figure 1 shows that the size of the bias is affected by overlap, size of the study and the strength of the relative treatment effect. The bias is close to 0 in scenarios where there is large overlap in covariates between the two studies (i.e., the difference in the covariate marginal means between studies is small). The bias in the scenarios with small overlap in covariates between the studies and small sample size is noticeably higher than other scenarios. Within the scenarios with small overlap and small sample size, the size of the bias seems associated with the magnitude of the relative treatment effect. The bias is larger in the case when the treatment effect is large than the case when the treatment effect is small. The bias reduces to close to 0 in the scenarios with small overlap when the sample size increases from 200 to 1000 per arm.

The size of the bias is not influenced by the strength of the correlation between the two covariates or the strength of the correlation across the two studies. It is also not affected by the fact that we do not obtain the desired correlation matrix using the correlation from the IPD study in step 1 of the NORTA algorithm.

Figure 2 shows that across all scenarios the standard errors estimate the empirical standard errors well in general. Increasing the sample size may be associated with a reduction in the discrepancy between the model standard error and the empirical standard error. Coverage is at the nominal level across all scenarios with no obvious patterns observed (Figure 3).

The results from using the 'plugging-in' of the mean covariates approach are presented in the online supporting material Appendix 2. In summary, the 'plugging-in' approach performs similarly to the simulation-based approach when assessing bias. However, the coverage was below the nominal level (between 70% and 80%) for the 'plugging-in' approach in the case there are small overlap between the two studies. In the low coverage cases, the model standard error was much smaller than the empirical standard error.

In summary, this simulation demonstrates that the unanchored STC approach using the NORTA algorithm to simulate covariates provides an asymptotically unbiased estimate for the population-adjusted indirect comparison for binary outcomes. Within the unanchored

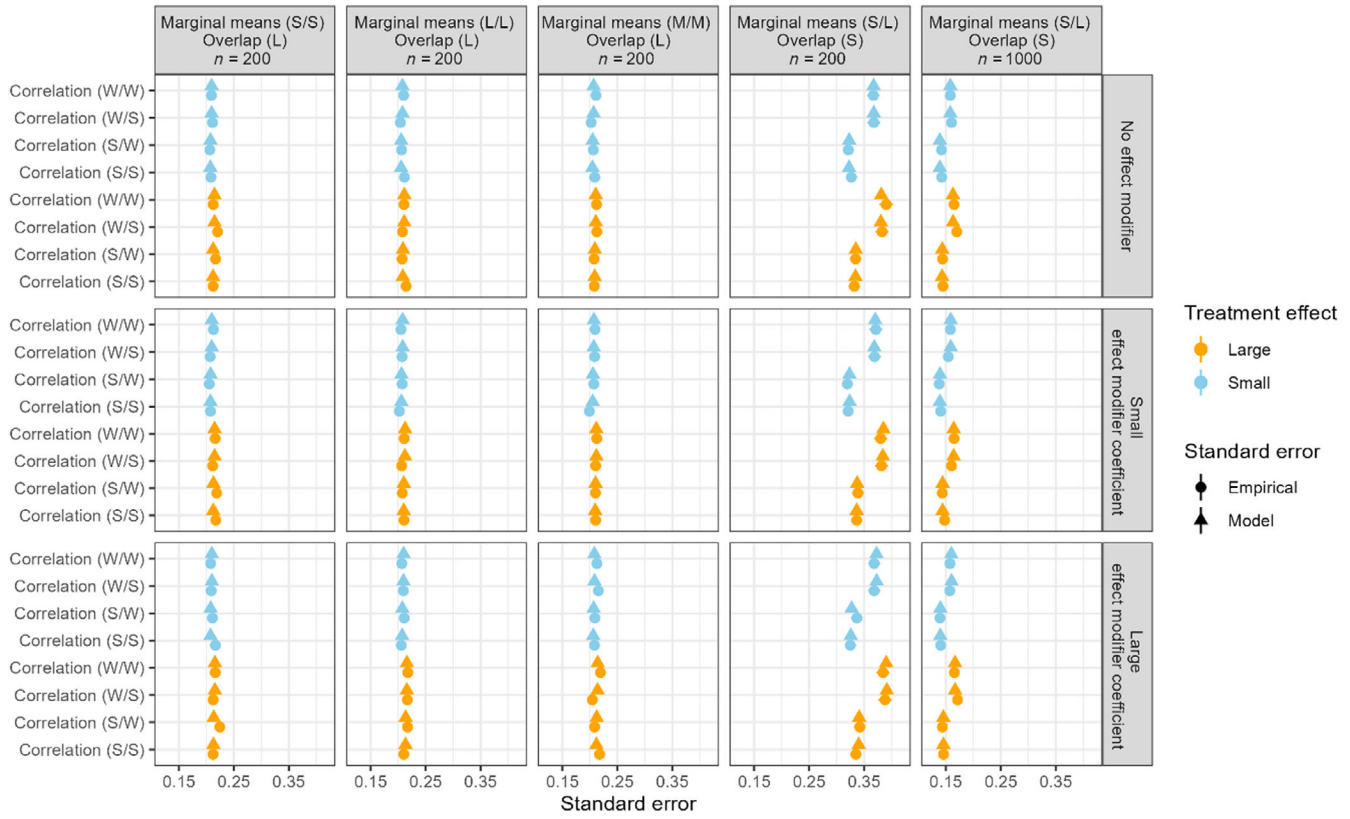

FIGURE 2 Empirical and model error along with 95% Monte Carlo confidence intervals for different scenarios. (S, M and L for marginal means indicate small, moderate and large marginal mean. S and L for overlap indicate small and large overlap. W and S for correlation indicate weak and strong correlation.)

STC, our proposed way of using the NORTA algorithm to sample covariates would result the correlation matrix from the sampled covariates not matching the input correlation matrix from the IPD study. The simulation study shows that unanchored STC is unaffected by this issue, and still provides an asymptotically unbiased estimate. In the case of sparse data, (i.e., when the covariates have large imbalance within the strata), unanchored STC is associated with the sparse data bias only in the situation where the overlap in covariates between the studies are small and the studies have small sample size. In the situation where there is sufficient overlap or limited overlap but large sample size, unanchored STC still provides an unbiased estimate.

4 | DISCUSSION

Unanchored MAIC or STC is required to adjust for the population differences when deriving the indirect treatment effect using single-arm studies with only aggregate data are available for the comparator arm. It is unclear why unanchored MAIC is used far more frequently than unanchored STC in HTA. This may relate to the lack of clarity on how unanchored STC could be performed. We proposed to use the NORTA/Gaussian copula approach to simulate covariates with any desired marginal distributions. We also illustrated how to predict the treatment effect and obtain the appropriate standard errors using bootstrap so that the estimate is unbiased and has good coverage rates. We aim to make unanchored STC more accessible to users in practice by providing example R code for data simulation and analysis on GitHub.

Our proposed way of using the NORTA algorithm to simulate covariates from the comparator study is to incorporate the correlation in the IPD study without any transformation. This approach is easy to implement, however the drawback is that the simulated covariates for the comparator study will not have the desired correlation matrix (i.e., the correlation for the sampled covariates does not match the correlation in the IPD study).

We emphasise that the use of marginalisation and predicting on the natural scale to obtain the average predicted outcome on the natural scale before transforming to the linear predictor scale to obtain the marginal treatment effect is important. This procedure ensures that the estimated marginal treatment effect is unbiased.

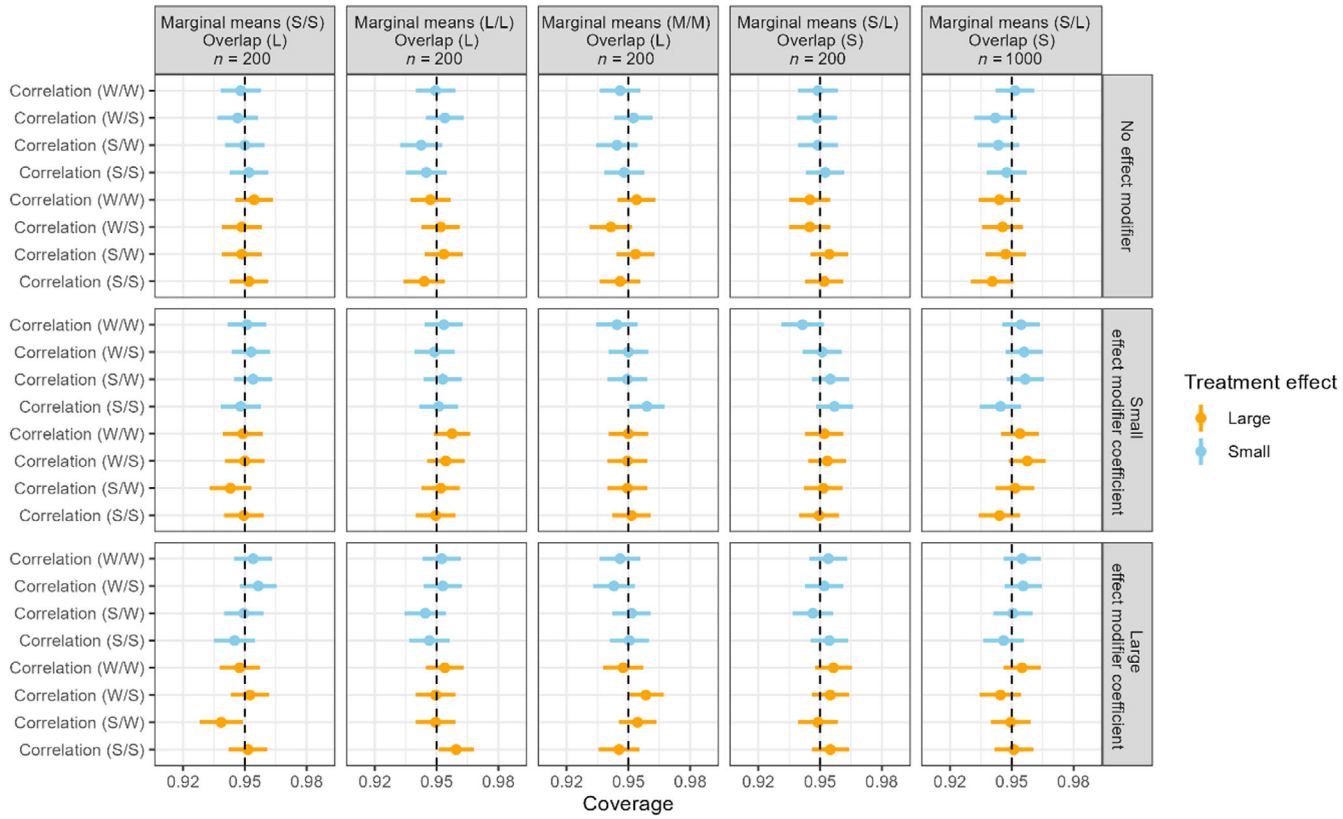

FIGURE 3 Coverage along with 95% Monte Carlo confidence intervals for different scenarios. (S, M and L for marginal means indicate small, moderate and large marginal mean. S and L for overlap indicate small and large overlap. W and S for correlation indicate weak and strong correlation.)

How much bias would be introduced by the 'plugging-in' approach depends on how non-linear the function is and whether the density mass of the input parameters for the non-linear function would be concentrated on the non-linearity part of the function. Although in our simulation settings, the 'plugging-in' approach does not appear to introduce additional bias we argue that the simulation-based approach should always be used in the case of a non-linear link function in practice to eliminate aggregation bias.

Bootstrap need to be applied to obtain the correct standard error for the treatment effect. Because only the IPD arm could be resampled, the variance calculation should be separately for the treatment and control arm in the comparator study. This is relatively easy to achieve for either binary or continuous outcome as there is a closed formula for the variance for the control arm in the comparator study. However, this is not the case for time-to-event outcome. We used 1000 bootstrap samples due to computational constraints in the simulation study. In practice, a larger number of bootstrap samples may be required depending on the input data. We suggest users determine the appropriate number of bootstrap samples by increasing the number of bootstrap samples until the results from multiple analyses using different random seeds stabilise.[9]

Our simulation study is the first simulation study for unanchored STC. It shows the issue relating to using the correlation from the IPD study directly in the NORTA algorithm does not have an impact on the performance of the unanchored STC method. We varied the strength of the correlation between the two covariates within each study and across the two studies, and observed similar magnitude of bias and coverage across all scenarios. This finding is consistent with the findings from Phillippo et al. (2020)[10] that the assumed correlation structure of the aggregate data trial has negligible effect on the results in the anchored STC.

We found that sparse data bias together with the size of overlap in covariates between the two studies influences the size of the bias the most. When overlap is large or the sample size is big, a large imbalance within the strata of a covariate does not have an impact on the performance of the unanchored STC method. The magnitude of bias and coverage were similar in the case with either small or large marginal means for the two covariates compared to the case with moderate size marginal means. However, when there is one covariate with a

small marginal mean and other covariate with a large marginal mean, there is a noticeable increase in the size of the bias. As the overlap reduces, the bias increases; however, the bias reduces with an increasing of the sample size even in the case of large proportion of non-overlap between covariates. This shows that the unanchored STC method provides an asymptotically unbiased estimate of the treatment effect for binary outcomes. This property is as expected because the maximum likelihood estimate for a logistic regression is asymptotically unbiased.

We also found that whether a covariate is an effect modifier or the strength of the effect modification does not have an impact on the performance of the unanchored STC method. This finding is in line with the findings from Phillippo et al. (2020)[10] in the anchored STC. When the true relative treatment effect is large, the lack of overlap between the two studies has the most profound effect on the bias and this scenario is associated with the largest bias across all scenarios varying other factors.

Our simulation only investigated the performance of the unanchored STC with binary data. Single-arm studies most frequently occur in oncology trials, where time-to-event endpoint such as overall survival or progression-free survival would be the primary outcome of interest. The use of marginalisation and the NORTA algorithm to sample covariates should still be applied because the link function is not identical in the outcome regression model. However, it remains unclear what model may be the best to use in the outcome regression step, for example, a cox regression model vs. a parametric survival model, as this would have an impact on the extrapolation which is an important area for consideration in HTA due to the limited follow-up time within a clinical study. Another challenge with time-to-event outcome is how to obtain the marginal treatment effect and uncertainty associated with this estimation, for example, a constant hazard vs. time-varying hazards; and the relationship between this marginal treatment effect to the outcome regression used for prediction. Further research is required to develop an unbiased unanchored STC approach for time-to-event data.

We only included two covariates (both binary) in our simulation study whereas in practice it is likely a larger number of covariates and other types of covariates would be determined as the potential prognostic factors and effect modifiers. Our simulation study shows that the correlation structure does not influence the performance of the method. We fully expect this property to be maintained with a higher dimensional correlation matrix, although further research is required to confirm this. In terms of implementing the NORTA/Gaussian copula approach to simulate more than 2 covariates and other types of covariates, the code available on GitHub could be easily extended [e.g., the normalCopula() function] by choosing the appropriate marginal distribution given the summary statistics for the covariates.

We did not investigate the scenario where not all prognostic factor and effect modifiers are included in the regression model as it is clear from previous work[10–15] that not adjusting for all important covariates lead to bias. We also did not investigate the scenario where the covariate-outcome relationship is nonlinear as this would obviously lead to biased results because the STC approach relies on the outcome regression model correctly model the relationship between covariates and outcome, and this has been demonstrated in the anchored STC case from previous work.[10] We also did not compare the performance of unanchored MAIC with our proposed unanchored STC approach. MAIC approach has been extensively investigated in the literature and its properties are well understood.[10–15] However, there is discrepancy in the conclusions on the superiority of performance between MAIC and STC in the anchored case, which could be due to the different scenario settings used between the studies. A further comprehensive simulation study comparing unanchored MAIC and unanchored STC could be useful.

To conclude, we provided a step-by-step guide on how to conduct an unanchored STC approach and performed a simulation study to evaluate the performance of the proposed procedure using a binary data. The simulation study shows that our proposed implementation of unanchored STC performs well across all scenarios apart from in the case of small overlap between the studies with small sample size. In this worst-case scenario, the absolute bias was still less than 0.05 on the log-odds ratio scale in the simulation scenarios that we explored. We encourage analysts to consider using unanchored STC when conducting unanchored indirect comparisons with single-arm studies.

AUTHOR CONTRIBUTIONS

Shijie Ren: Conceptualization; methodology; software; investigation; writing – original draft; writing – review and editing; visualization; validation; project administration; formal analysis; data curation; funding acquisition; supervision. **Sa Ren:** Software; formal analysis; visualization; writing – review and editing; data curation; methodology. **Nicky J. Welton:** Writing – review and editing; methodology. **Mark Strong:** Writing – review and editing; methodology.

ACKNOWLEDGEMENTS

Shijie Ren is supported by funding from NIHR Advanced Fellowship (Grant number: NIHR300590). The authors

would like to thank David M Phillippo for the helpful discussion on marginal and conditional treatment effect in unanchored indirect treatment comparisons.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

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

ORCID

Shijie Ren https://orcid.org/0000-0003-3568-7124
Nicky J. Welton https://orcid.org/0000-0003-2198-3205

REFERENCES

1. Hatswell AJ, Baio G, Berlin JA, Irs A, Freemantle N. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999–2014. *BMJ Open*. 2016;6(6):e011666. doi:10.1136/bmjopen-2016-011666
2. Faria R, Hernandez Alava M, Manca A, Wailoo AJ. NICE DSU technical support document 17: The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: Methods for comparative individual patient data. 2015.
3. Signorovitch JE, Sikirica V, Erder MH, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health*. 2012;15(6):940-947. doi:10.1016/j.jval.2012.05.004
4. Ishak KJ, Proskorovsky I, Benedict A. Simulation and matching-based approaches for indirect comparison of treatments. *Pharmacoeconomics*. 2015;33(6):537-549. doi:10.1007/s40273-015-0271-1
5. Caro JJ, Ishak KJ. No head-to-head trial? Simulate the missing arms. *Pharmacoeconomics*. 2010;28(10):957-967. doi:10.2165/11537420-000000000-00000
6. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for population-adjusted indirect comparisons in health technology appraisal. *Med Decis Making*. 2018;38(2):200-211. doi:10.1177/0272989X17725740
7. Phillippo DM, Dias S, Ades AE, et al. Multilevel network meta-regression for population-adjusted treatment comparisons. *J R Stat Soc Ser A Stat Soc*. 2020;183(3):1189-1210. doi:10.1111/rssa.12579
8. Daniel R, Zhang J, Farewell D. Making apples from oranges: comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biom J*. 2021;63(3):528-557. doi:10.1002/bimj.201900297
9. Remiro-Azocar A, Heath A, Baio G. Parametric G-computation for compatible indirect treatment comparisons with limited individual patient data. *Res Synth Methods*. 2022;13(6):716-744. doi:10.1002/jrsm.1565
10. Phillippo DM, Dias S, Ades AE, Welton NJ. Assessing the performance of population adjustment methods for anchored indirect comparisons: a simulation study. *Stat Med*. 2020;39(30):4885-4911. doi:10.1002/sim.8759
11. Weber D, Jensen K, Kieser M. Comparison of methods for estimating therapy effects by indirect comparisons: a simulation study. *Med Decis Making*. 2020;40(5):644-654. doi:10.1177/0272989X20929309
12. Jiang Y, Ni W. Performance of unanchored matching-adjusted indirect comparison (MAIC) for the evidence synthesis of single-arm trials with time-to-event outcomes. *BMC Med Res Methodol*. 2020;20:241. doi:10.1186/s12874-020-01124-6
13. Hatswell AJ, Freemantle N, Baio G. The effects of model misspecification in unanchored matching-adjusted indirect comparison: results of a simulation study. *Value Health*. 2020;23(6):751-759. doi:10.1016/j.jval.2020.02.008
14. Cheng D, Ayyagari R, Signorovitch J. The statistical performance of matching-adjusted indirect comparisons: estimating treatment effects with aggregate external control data. *Ann Appl Stat*. 2020;14(4):1806-1833. doi:10.1214/20-aoas1359
15. Remiro-Azocar A, Heath A, Baio G. Methods for population adjustment with limited access to individual patient data: a review and simulation study. *Res Synth Methods*. 2021;12(6):750-775. doi:10.1002/jrsm.1511
16. Sultana N, Ren S. Review of methods used to estimate treatment effects against relevant comparators using evidence from single-arm studies in NICE single technology appraisals. *Value Health*. 2022;25(12):S10.
17. Serret-Larmande A, Zenati B, Dechartres A, Lambert J, Hajage D. A methodological review of population-adjusted indirect comparisons reveals inconsistent reporting and suggests publication bias. *J Clin Epidemiol*. 2023;163:1-10. doi:10.1016/j.jclinepi.2023.09.004
18. Truong B, Tran LT, Le TA, Pham TT, Vo TT. Population adjusted-indirect comparisons in health technology assessment: a methodological systematic review. *Res Synth Methods*. 2023;14(5):660-670. doi:10.1002/jrsm.1653
19. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submission to NICE. 2016.
20. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient-versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med*. 2002;21(3):371-387. doi:10.1002/sim.1023
21. Chang B-H, Lipsitz S, Waternaux C. Logistic regression in meta-analysis using aggregate data. *J Appl Stat*. 2000;27(4):411-424. doi:10.1080/02664760050003605
22. Zhang L, Bujkiewicz S, Jackson D. Four alternative methodologies for simulated treatment comparison: how could the use of simulation be re-invigorated? *Res Synth Methods*. 2023;15(2):227-241. doi:10.1002/jrsm.1681
23. Zhang Z. Estimating a marginal causal odds ratio subject to confounding. *Commun Stat Theory Methods*. 2008;38(3):309-321. doi:10.1080/03610920802200076
24. Cario MC, Nelson BL. *Modeling and Generating Random Vectors with Arbitrary Marginal Distributions and Correlation Matrix*. Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL;1997.
25. Mardia KV. A translation family of bivariate distributions and Fréchet's bounds. *Sankhyā*. 1970;32(1):119-122.

26. Li ST, Hammond JL. Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Trans Syst Man Cybern*. 1975;5(5):557-561. doi:10.1109/TSMC.1975.5408380
27. Xiao Q. Generating correlated random vector involving discrete variables. *Commun Stat Theory Methods*. 2017;46(4):1594-1605. doi:10.1080/03610926.2015.1024860
28. Astivia OOL. Issues, problems and potential solutions when simulating continuous, non-normal data in the social sciences. *Meta-Psychol*. 2020;4. doi:10.15626/mp.2019.2117
29. Bonofiglio F, Schumacher M, Binder H. Recovery of original individual person data (IPD) inferences from empirical IPD summaries only: applications to distributed computing under disclosure constraints. *Stat Med*. 2020;39(8):1183-1198. doi:10.1002/sim.8470
30. Astivia OOL, Kroc E, Zumbo BD. Simultaneous estimation of the intermediate correlation matrix for arbitrary marginal densities. *Behav Res Methods*. 2024;56:1852-1862. doi:10.3758/s13428-023-02123-3
31. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci*. 1986;1(1):54-75.
32. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *BMJ*. 2016;352:i1981. doi:10.1136/bmj.i1981
33. Cordeiro GM, McCullagh P. Bias correction in generalized linear models. *J R Stat Soc B Methodol*. 1991;53(3):629-643. doi:10.1111/j.2517-6161.1991.tb01852.x
34. Gosho M, Ohigashi T, Nagashima K, Ito Y, Maruo K. Bias in odds ratios from logistic regression methods with sparse data sets. *J Epidemiol*. 2023;33(6):265-275. doi:10.2188/jea.JE20210089
35. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993;80(1):27-38. doi:10.2307/2336755
36. Puhr R, Heinze G, Nold M, Lusa L, Geroldinger A. Firth's logistic regression with rare events: accurate effect estimates and predictions? *Stat Med*. 2017;36(14):2302-2317. doi:10.1002/sim.7273
37. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019;38(11):2074-2102. doi:10.1002/sim.8086
38. Phillippo DM, Dias S, Elsada A, Ades AE, Welton NJ. Population adjustment methods for indirect comparisons: a review of National Institute for health and care excellence technology appraisals. *Int J Technol Assess Health Care*. 2019;35(3):221-228. doi:10.1017/S0266462319000333

SUPPORTING INFORMATION

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

How to cite this article: Ren S, Ren S, Welton NJ, Strong M. Advancing unanchored simulated treatment comparisons: A novel implementation and simulation study. *Res Syn Meth*. 2024;1-14. doi:10.1002/jrsm.1718