

RESEARCH ARTICLE

OPEN ACCESS



A New Frequentist Implementation of the Daniels and Hughes Bivariate Meta-Analysis Model for Surrogate Endpoint Evaluation

Dan Jackson¹ | Michael Sweeting¹ | Robbie C. M. van Aert² | Sylwia Bujkiewicz³ | Keith R. Abrams⁴ | Wolfgang Viechtbauer⁵

¹Statistical Innovation Group, AstraZeneca, Cambridge, UK | ²Tilburg University, Tilburg, the Netherlands | ³Biostatistics Research Group, Department of Population Health Sciences, University of Leicester, Leicester, UK | ⁴Department of Statistics and Warwick Medical School, University of Warwick, Coventry, UK | ⁵Maastricht University, Maastricht, the Netherlands

Correspondence: Dan Jackson (daniel.jackson1@astrazeneca.com)

Received: 29 May 2024 | **Revised:** 23 January 2025 | **Accepted:** 13 February 2025

Keywords: bias correction | maximum likelihood estimation | meta-regression | multivariate meta-analysis | surrogate endpoint

ABSTRACT

Surrogate endpoints are used when the primary outcome is difficult to measure accurately. Determining if a measure is suitable to use as a surrogate endpoint is a challenging task and a variety of meta-analysis models have been proposed for this purpose. The Daniels and Hughes bivariate model for trial-level surrogate endpoint evaluation is gaining traction but presents difficulties for frequentist estimation and hitherto only Bayesian solutions have been available. This is because the marginal model is not a conventional linear model and the number of unknown parameters increases at the same rate as the number of studies. This second property raises immediate concerns that the maximum likelihood estimator of the model's unknown variance component may be downwardly biased. We derive maximum likelihood estimating equations to motivate a bias adjusted estimator of this parameter. The bias correction terms in our proposed estimating equation are easily computed and have an intuitively appealing algebraic form. A simulation study is performed to illustrate how this estimator overcomes the difficulties associated with maximum likelihood estimation. We illustrate our methods using two contrasting examples from oncology.

1 | Introduction

Surrogate endpoints are used when the primary endpoint is difficult to measure accurately, for example when little information is accessible or information may only become available in the longer term. A common example in the context of oncology trials is the use of progression-free survival (PFS) as a surrogate endpoint for overall survival (OS; Belin et al. 2020). Determining if an outcome is suitable to use as a surrogate endpoint is a challenging task in practice, despite the criteria that have been proposed for this purpose (e.g., Baker 2018; Daniels and Hughes 1997; Prentice 1989).

We focus on the situation where evidence from multiple trials is available and a statistical assessment of trial-level surrogacy is required. A number of approaches have been developed that require individual patient data (IPD; Burzykowski et al. 2001; Burzykowski et al. 2021; Buyse 2000; Dai and Hughes, 2012; Molenberghs et al. 2008). We agree with Li and Meredith (2003) that models using IPD are to be preferred when these data are available because associations between patient outcomes can then be quantified. However, it is typically the case that only aggregate (or summary level) data are available to the analyst and we address this common situation. Meta-analytic approaches focus on the association between treatment effects

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Biometrical Journal* published by Wiley-VCH GmbH.

and so, in the framework of Taylor and Elston (2009), assess “level one” evidence for surrogacy. In general, other methods will be needed to assess evidence at their second (associations between outcomes) and third (biological plausibility) levels.

Bujkiewicz et al. (2017) provide an accessible overview of methods for meta-analysis, that can be used to assess if an outcome is a suitable surrogate endpoint, requiring only aggregate level study information. One of these methods is the model proposed by Daniels and Hughes (1997), where a linear regression is assumed for the true effects across studies. Advantages of the Daniels and Hughes (1997) model are that it does not require the estimation of a between-study correlation parameter, which can be challenging (Burke et al. 2018), and it does not assume that the surrogate treatment effects are exchangeable (Bujkiewicz et al. 2017). However this model is, in some respects, unusual. For example, the marginal model is not a conventional linear model. Furthermore, it uses a separate fixed effect for the candidate surrogate treatment effect from each study, so that the number of unknown parameters increases at the same rate as the number of studies. These two features make the Daniels and Hughes (1997) model difficult to satisfactorily fit in a frequentist framework, where the main challenge relates to estimating its unknown variance component. This is because the presence of a separate fixed effect for each study can result in notable downward bias in maximum likelihood estimates of variance parameters (Jackson, Law, et al. 2018), analogously to the well-known Neyman and Scott (1948) problem.

Extensions of the Daniels and Hughes (1997) model are possible, for example, where there is more than one potential surrogate (Pozzi et al. 2016) or multiple treatment classes (Papanikos et al. 2020), and we return to this in the discussion. There has therefore been considerable methodological development since Daniels and Hughes (1997) and this model is gaining traction in Health Technology Assessment (HTA). For example, it is described in detail in NICE Technical Support Document 20 (Bujkiewicz, Achana, et al. 2019). We therefore anticipate that the application of the Daniels and Hughes model to increase. Our position is that it is useful to have Bayesian and frequentist estimation methods available to us. Our work fills an important void by providing, to the best of our knowledge, the first frequentist implementation of the Daniels and Hughes model.

The main competitor to the Daniels and Hughes model is the bivariate random-effects meta-analysis model (Jackson et al. 2011; Bujkiewicz et al. 2017; Bujkiewicz, Jackson, et al. 2019) that has been extended to the network meta-analysis setting (Achana et al. 2014; Bujkiewicz, Achana, et al. 2019; Efthimiou et al., 2014; Jackson, Bujkiewicz, et al. 2018). This is a linear model where the number of parameters does not depend on the number of studies, so that this model is more amenable to frequentist methods. However it requires the estimation of the between-study correlation and assumes that the surrogate treatment effects are exchangeable. Hence, the Daniels and Hughes and bivariate random-effects meta-analysis models are diametrically opposites in terms of these advantages and disadvantages. However, both these models acknowledge that the surrogate treatment effects are estimated with uncertainty. A wide variety of estimation methods for the between-study covariance matrix are available for the bivariate random-effects meta-analysis model but they are

not immediately applicable to the Daniels and Hughes model. For example, the theory for the restricted maximum likelihood estimator (REML; Jennrich and Schluchter 1986), that helps to correct for the downward bias of maximum likelihood estimates of variance components (Jackson et al. 2011), assumes a linear model. Moment-based estimators for bivariate random-effects meta-analysis (Jackson et al. 2013) have not been developed for the Daniels and Hughes model, and estimators of this type possess no optimality properties. Conventional meta-regression (Thompson and Sharp 1999) is another competing method (Bujkiewicz et al. 2017) that instead includes the treatment effect on the surrogate endpoint as a covariate.

Daniels and Hughes (1997) propose fitting their model in the Bayesian framework, which circumvents the estimation difficulties encountered when using frequentist methods, but comes at the price of sensitivity to prior distribution specification. In particular, the conventional univariate random-effects model for meta-analysis has been found to be sensitive to the prior distribution used for the between-study variance parameter (Lambert et al. 2005). This can also be anticipated to be the case for the unknown variance component of the Daniels and Hughes model. The aim of this paper is to develop frequentist estimation methods for this model and so avoid concerns related to prior sensitivity. We achieve this by considering maximum likelihood estimation and then using the resulting estimating equations to derive a bias adjusted estimator of the unknown variance component. This estimate can then be used in a bias adjusted frequentist analysis using the Daniels and Hughes (1997) model that addresses the potential bias in the corresponding maximum likelihood estimator.

The rest of the paper is set out as follows. In Section 2, we present the Daniels and Hughes model. In Section 3, we develop maximum likelihood estimating equations and we derive a bias adjusted estimator of the unknown variance component in Section 4. In Section 5, we perform a small-scale simulation study that provides proof of concept that our adjusted estimator overcomes the problems associated with using maximum likelihood estimation in conjunction with the Daniels and Hughes model. In Section 6, we apply our methodology to two contrasting examples and we conclude with a short discussion.

2 | The Daniels and Hughes Model

The Daniels and Hughes (1997) bivariate meta-analysis model for surrogate endpoint evaluation assumes that each study provides two estimates (with corresponding standard errors and within-study correlation): the estimated treatment effect on the clinical outcome of interest and the surrogate treatment effect. For example, the estimated treatment effect on the clinical outcome of interest could be an estimated log hazard ratio comparing two treatment groups for OS and the surrogate treatment effect could be the corresponding log hazard ratio for PFS. The overall aim is to determine if the surrogate is viable (see Section 2.4 below).

Let n be the number of studies included in the analysis. For the i th study, let θ_i and γ_i , $i = 1, 2, \dots, n$, be the true treatment effect on the clinical outcome of interest, and on the surrogate, respectively. Let

$\hat{\theta}_i$ and $\hat{\gamma}_i$ be the corresponding estimates that are usually available or can be extracted from published papers or reports, and so are aggregate level information available to the analyst.

2.1 | The Within-Study Model

Daniels and Hughes (1997) assume the standard within-study model for bivariate meta-analysis (their eq. (1); Jackson et al. 2011), and so assume that

$$\begin{pmatrix} \hat{\theta}_i \\ \hat{\gamma}_i \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_i \\ \gamma_i \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & \rho_i \sigma_i \delta_i \\ \rho_i \sigma_i \delta_i & \delta_i^2 \end{pmatrix}\right). \quad (1)$$

Model (1) describes how the distributions of $\hat{\theta}_i$ and $\hat{\gamma}_i$ depend on the true θ_i and γ_i . We treat the within-study covariance matrix in model (1) as fixed and known. Typically the within-study variances, σ_i^2 and δ_i^2 , will be known but the within-study correlations ρ_i between $\hat{\theta}_i$ and $\hat{\gamma}_i$ are unlikely to be directly reported in aggregate level data. We agree with Riley (2009) that using appropriate values for the within-study correlations is an important consideration. A variety of approaches are available that produce suitable values of ρ_i (Bujkiewicz et al. 2013; Bujkiewicz, Achana, et al. 2019; Riley et al. 2008; 2015; Wei and Higgins 2013) and we explore two possibilities in our examples below.

2.2 | The Between-Studies Model

Between studies, Daniels and Hughes (1997) assume that

$$\theta_i | \gamma_i \sim N(\alpha + \beta \gamma_i, \tau^2). \quad (2)$$

Model (2) is the between-studies model because it states the assumptions made about the relationships between the true treatment effects on the two outcomes across trials. Equation (2) is different to the conventional between-study model for random-effects bivariate meta-analysis (Jackson et al. 2011; their eq. (2)) that instead assumes that (θ_i, γ_i) follows a bivariate normal distribution. It is therefore model (2) that distinguishes the Daniels and Hughes model from the bivariate random-effects meta-analysis model. The parameter τ^2 is the (conditional on γ_i) between-study variance of θ_i and is the focus of attention in much of the work that follows. Model (2) follows from the bivariate random-effects model but does not make the assumption of bivariate normality.

2.3 | The Marginal Model

The implied marginal model, from models (1) and (2), is

$$\begin{pmatrix} \hat{\theta}_i \\ \hat{\gamma}_i \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha + \beta \gamma_i \\ \gamma_i \end{pmatrix}, \begin{pmatrix} \sigma_i^2 + \tau^2 & \rho_i \sigma_i \delta_i \\ \rho_i \sigma_i \delta_i & \delta_i^2 \end{pmatrix}\right). \quad (3)$$

In the frequentist framework, inferences are usually based on the marginal model. Hence we take model (3) to be the Daniels and Hughes model. The Daniels and Hughes model makes normality assumptions both within (model 1) and between (model 2)

studies. Methods for meta-analysis that make fewer normality assumptions are in general available (Jackson and White 2018) and we return to the possibility of relaxing these assumptions in the discussion. We assume that each pair of $(\hat{\theta}_i, \hat{\gamma}_i)$ is independent.

2.4 | Assessing Surrogacy

All three parameters α , β , and τ^2 in model (3) are useful when determining if the putative surrogate endpoint is suitable as a good predictor of clinical benefit measured by the clinical outcome of interest, which is discussed by Daniels and Hughes (1997). Briefly, $\alpha = 0$, $\beta \neq 0$, and $\tau^2 = 0$ are taken to be indicative of a perfect surrogate relationship. If $\alpha = 0$, then no treatment effect for the surrogate corresponds to no treatment effect for the outcome of clinical interest. If $\beta \neq 0$, then a treatment effect for the surrogate is associated with a treatment effect for this outcome; if the same directional effects indicate treatment benefit for the surrogate and outcome of clinical interest, then $\beta > 0$ is suggestive of a surrogate relationship. Finally, if $\tau^2 = 0$, then the treatment effect on the final clinical outcome can be predicted perfectly given the effect on the surrogate endpoint. All three parameters in model (2) should be used to assess surrogacy. The γ_i are usually regarded as nuisance parameters.

Hence $\hat{\alpha} \approx 0$ and $\hat{\tau}^2 \approx 0$, estimated with small uncertainty, and a confidence or credible interval for β that does not include zero, indicate that the surrogate is considered suitable. However, the extent to which different departures from this are tolerable when declaring surrogacy is an open question. We return to this issue in the discussion.

2.5 | Challenges for Frequentist Inference

As explained in the introduction, model (3) presents two main challenges for frequentist estimation. First, there is a separate parameter γ_i for each study, so that the number of parameters increases at the same rate as the number of studies. Hence the maximum likelihood estimator of τ^2 may be badly downwardly biased (Jackson, Law, et al. 2018). Second, model (3) is a nonlinear model because it contains the term $\beta \gamma_i$ in the linear predictor for $\hat{\theta}_i$, which is not linear in terms of the model parameters. One implication of this is that the theory of restricted maximum likelihood (REML; Jennrich and Schluchter 1986) does not apply.

Despite concerns about downward bias for the maximum likelihood estimate of τ^2 , we will begin by deriving maximum likelihood estimating equations for all model parameters. These estimating equations will then be used to motivate a bias adjusted estimator of τ^2 .

3 | Maximum Likelihood Estimating Equations

The likelihood function is the product of densities of the bivariate normal probability density function from model (3), where each study contributes one such term. Maximum likelihood estimating equations can be derived using the properties of the normal distribution.

3.1 | Estimating Equation for γ_i

We use $\hat{\gamma}_{i,MLE}$, $i = 1, \dots, n$, to denote the *model-based* maximum likelihood estimate of γ_i . Conceptually, $\hat{\gamma}_{i,MLE}$ and $\hat{\gamma}_i$ are estimates of the same quantity γ_i . The difference however is that $\hat{\gamma}_{i,MLE}$ uses information from all studies, assuming the meta-analysis model (3), whereas $\hat{\gamma}_i$ is aggregate level data used in analysis and uses information only from the i th study.

The likelihood function is the product of probability density functions implied by model (3), so that the only terms in this function that contain γ_i are from the i th study. When differentiating the log-likelihood function with respect to γ_i , only the logarithm of the bivariate normal density contribution for the i th study remains. Reparameterizing so that $\hat{\theta}_i^* = \hat{\theta}_i - \alpha$, model (3) becomes

$$\begin{pmatrix} \hat{\theta}_i^* \\ \hat{\gamma}_i \end{pmatrix} \sim N\left(\mathbf{C}_{(\beta,1)}\gamma_i, \begin{pmatrix} \sigma_i^2 + \tau^2 & \rho_i\sigma_i\delta_i \\ \rho_i\sigma_i\delta_i & \delta_i^2 \end{pmatrix}\right), \quad (4)$$

where $\mathbf{C}_{(\beta,1)}$ is a column vector of length 2 where the first entry is β and the second is one. Noting that only the contribution from the i th study remains when differentiating the log-likelihood function with respect to γ_i , $\hat{\gamma}_{i,MLE}$ satisfies the maximum likelihood estimating equation for γ_i from model (4) where all other parameters are replaced by their maximum likelihood estimates. Model (4) is a weighted linear regression model, so that the standard theory for this type of model can be used to derive the resulting maximum likelihood estimating equation for γ_i .

Let $\hat{\mathbf{W}}_i$ be the inverse of the covariance matrix in (4) where τ^2 is replaced by its maximum likelihood estimate, $\hat{\tau}^2$. Standard theory for weighted linear regression models results in the estimating equation

$$\hat{\gamma}_{i,MLE} = (\mathbf{C}_{(\beta,1)}^T \hat{\mathbf{W}}_i \mathbf{C}_{(\beta,1)})^{-1} \mathbf{C}_{(\beta,1)}^T \hat{\mathbf{W}}_i \begin{pmatrix} \hat{\theta}_i^* \\ \hat{\gamma}_i \end{pmatrix}, \quad (5)$$

where all estimates are maximum likelihood estimates. Estimating Equation (5) could be used as part of an iterative scheme to estimate all parameters, where estimates at the previous step are used on the right-hand side to produce the maximum likelihood estimate of γ_i at the current step on the left. Equation (5) is the basis for n estimating equations, because there is one such equation for each γ_i .

3.2 | Estimating Equations for α and β

We factorize the contributions to the likelihood from the i th study resulting from the bivariate normal model (3) as contribution from the marginal density of $\hat{\gamma}_i \sim N(\gamma_i, \delta_i^2)$ and the contribution from the conditional density

$$\hat{\theta}_i | \hat{\gamma}_i \sim N(\alpha + \beta\gamma_i + (\rho_i\sigma_i/\delta_i)(\hat{\gamma}_i - \gamma_i), \sigma_i^2(1 - \rho_i^2) + \tau^2).$$

To simplify notation, let $\mu_i(\hat{\gamma}_i) = \alpha + \beta\gamma_i + (\rho_i\sigma_i/\delta_i)(\hat{\gamma}_i - \gamma_i)$ and $v_i^2 = \sigma_i^2(1 - \rho_i^2)$. Then the conditional distribution of $\hat{\theta}_i | \hat{\gamma}_i$ above becomes

$$\hat{\theta}_i | \hat{\gamma}_i \sim N(\mu_i(\hat{\gamma}_i), v_i^2 + \tau^2). \quad (6)$$

Model (6) is a weighted linear regression with an offset. Estimating equations are therefore easily derived, as explained in Appendix A.

3.3 | Estimating Equation for τ^2

The marginal model $\hat{\gamma}_i \sim N(\gamma_i, \delta_i^2)$ does not depend on τ^2 . Hence, when differentiating the log-likelihood function with respect to τ^2 , only terms from the conditional distributions of $\hat{\theta}_i | \hat{\gamma}_i$ from model (6) contribute. When omitting terms that do not include τ^2 , to within a constant the log-likelihood is

$$\ell(\tau^2) = \sum_{i=1}^n (\log(v_i^2 + \tau^2) + (\hat{\theta}_i - \mu_i(\hat{\gamma}_i))^2 / (v_i^2 + \tau^2)). \quad (7)$$

Differentiating Equation (7) with respect to τ^2 , setting the resulting expression to zero, and replacing all parameters with their maximum likelihood estimates gives the estimating equation

$$\sum_{i=1}^n (1/(v_i^2 + \hat{\tau}^2) - (\hat{\theta}_i - \hat{\mu}_i(\hat{\gamma}_i))^2 / ((v_i^2 + \hat{\tau}^2)^2)) = 0, \quad (8)$$

where $\hat{\mu}_i(\hat{\gamma}_i) = \hat{\alpha} + \hat{\beta}\hat{\gamma}_{i,MLE} + (\rho_i\sigma_i/\delta_i)(\hat{\gamma}_i - \hat{\gamma}_{i,MLE})$. Rearranging Equation (8) quickly provides the estimating equation

$$\hat{\tau}^2 = \frac{\sum_{i=1}^n ((\hat{\theta}_i - \hat{\mu}_i(\hat{\gamma}_i))^2 - v_i^2) / (v_i^2 + \hat{\tau}^2)^2}{\sum_{i=1}^n 1 / (v_i^2 + \hat{\tau}^2)^2}. \quad (9)$$

Equation (9) is analogous to eq. (3b) from Thompson and Sharp (1999). We truncate $\hat{\tau}^2 = 0$ if the estimate from Equation (9) is negative.

We have now derived maximum likelihood estimating Equations (5, 9, A2, A3), where two of these equations are in Appendix A, for all model parameters. Note that parameter estimates appear on the right-hand sides of these estimating equations, so that they may be used in iterative scheme to fit the model; they are not closed-form solutions that can be used to compute the maximum likelihood estimates.

3.3.1 | Assumptions Required for Estimating Equation (9) for τ^2 to Be Unbiased

Before considering bias correction, it is insightful to understand the conditions under which estimating Equation (9) is unbiased. Suppose that all model parameters, α , β , τ^2 , and all γ_i , were known and are used to evaluate the right-hand side of estimating Equation (9). Then we replace the estimates in Equation (9) with their corresponding true parameter values and this estimating equation becomes

$$\hat{\tau}^2 = \frac{\sum_{i=1}^n ((\hat{\theta}_i - \mu_i(\hat{\gamma}_i))^2 - v_i^2) / (v_i^2 + \tau^2)^2}{\sum_{i=1}^n 1 / (v_i^2 + \tau^2)^2}. \quad (10)$$

We can evaluate the expectation of (10) as

$$E[\hat{\tau}^2] = \frac{\sum_{i=1}^n (E[E[(\hat{\theta}_i - \mu_i(\hat{\gamma}_i))^2 | \hat{\gamma}_i]] - v_i^2) / (v_i^2 + \tau^2)^2}{\sum_{i=1}^n 1 / (v_i^2 + \tau^2)^2}. \quad (11)$$

From the conditional model (6), in Equation (11), we have $E[(\hat{\theta}_i - \mu_i(\hat{\gamma}_i))^2 | \hat{\gamma}_i] = v_i^2 + \tau^2$, where v_i^2 and τ^2 are both treated as constants. Hence $E(\hat{\tau}^2) = \tau^2$, so that $\hat{\tau}^2$ is then unbiased. If we knew the true parameter values, there would of course be no need to estimate τ^2 , but this does clarify that the form of the estimating Equation (9) is a good starting point.

4 | Bias Adjusted Estimating Equation for τ^2

The main concern is that the presence of an unknown fixed effect γ_i for each study may result in downward bias for the maximum likelihood estimate of τ^2 . We now repeat the analysis in Section 3.3.1 but where we take into account the fact that the estimates $\hat{\gamma}_{i,MLE}$, rather than the true values γ_i , are used in the estimating equation for τ^2 in Equation (9). However, we will treat $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\tau}^2$ as fixed constants when evaluating the right-hand side of Equation (9). This is because the main difficulties for the estimation of τ^2 are associated with the number of unknown γ_i increasing at the same rate as the number of studies. This approach deals with the most serious problem in an analytically tractable way.

Our proposed bias adjusted estimate of τ^2 is a modified version of the estimating Equation (9) given by

$$\hat{\tau}_{adjusted}^2 = \frac{\sum_{i=1}^n ((1 - b_i)^{-2} (\hat{\theta}_i - \hat{\mu}_i(\hat{\gamma}_i))^2 - v_i^2) / (v_i^2 + \hat{\tau}^2)^2}{\sum_{i=1}^n 1 / (v_i^2 + \hat{\tau}^2)^2}, \quad (12)$$

where $b_i = (\hat{\beta} - \rho_i \sigma_i / \delta_i) w_{1i}$; w_{1i} is the first entry of the row vector $(\mathbf{C}_{(\hat{\beta},1)}^T \hat{\mathbf{W}}_i \mathbf{C}_{(\hat{\beta},1)}^T)^{-1} \mathbf{C}_{(\hat{\beta},1)}^T \hat{\mathbf{W}}_i$.

The mathematical result required to derive the bias adjusted estimating Equation (12) is derived in Appendix B. We motivate this estimating equation by taking iterated expectations in (12), in the same way as in Section 3.3.1, but now using Equation (B4) from Appendix B to evaluate the inner expectations. This calculation shows that the expected value of $\hat{\tau}_{adjusted}^2$ in estimating Equation (12) is τ^2 , so that estimating Equation (12) removes bias in the maximum likelihood estimate of τ^2 associated with the need to estimate the γ_i . We truncate $\hat{\tau}_{adjusted}^2 = 0$ if the estimate from Equation (12) is negative.

4.1 | Evaluating the Bias Correction Terms

The key quantities are the “bias adjustment terms” $b_i = (\hat{\beta} - \rho_i \sigma_i / \delta_i) w_{1i}$ in estimating Equation (12). This is because if all $b_i = 0$, then no adjustment to the maximum likelihood estimate $\hat{\tau}^2$ from Equation (9) is made. We can explicitly compute w_{1i} by computing $(\mathbf{C}_{(\hat{\beta},1)}^T \hat{\mathbf{W}}_i \mathbf{C}_{(\hat{\beta},1)}^T)^{-1} \mathbf{C}_{(\hat{\beta},1)}^T \hat{\mathbf{W}}_i$ and extracting the first entry. This is straightforward because $\hat{\mathbf{W}}_i$ is the inverse of the 2×2 covariance matrix in (4); when computing this inverse using the adjoint method the determinant cancels in the calculation of $(\mathbf{C}_{(\hat{\beta},1)}^T \hat{\mathbf{W}}_i \mathbf{C}_{(\hat{\beta},1)}^T)^{-1} \mathbf{C}_{(\hat{\beta},1)}^T \hat{\mathbf{W}}_i$, which simplifies the calculation, and we obtain

$$w_{1i} = \frac{\hat{\beta} \delta_i^2 - \rho_i \sigma_i \delta_i}{\sigma_i^2 + \hat{\tau}^2 + \hat{\beta}^2 \delta_i^2 - 2 \hat{\beta} \rho_i \sigma_i \delta_i}.$$

Hence

$$b_i = \frac{(\hat{\beta} - \rho_i \sigma_i / \delta_i)(\hat{\beta} \delta_i^2 - \rho_i \sigma_i \delta_i)}{\sigma_i^2 + \hat{\tau}^2 + \hat{\beta}^2 \delta_i^2 - 2 \hat{\beta} \rho_i \sigma_i \delta_i} = \frac{(\hat{\beta} \delta_i - \rho_i \sigma_i)^2}{(\hat{\beta} \delta_i - \rho_i \sigma_i)^2 + \hat{\tau}^2 + v_i^2}. \quad (13)$$

Equation (13) shows that $0 \leq b_i < 1$, so that in Equation (12) we have $1 \leq (1 - b_i)^{-2} < \infty$. Hence positive scaling terms that cannot result in a smaller estimate of τ^2 are applied to the squared residuals when using the bias adjusted estimator (12).

If $\hat{\beta} = \rho_i \sigma_i / \delta_i$, then $b_i = 0$ and no adjustment is made for the i th study. From Equation (2), β is the gradient of the between-study regression of θ_i on γ_i . Furthermore, in the within-study model (1) for the i th study, $\rho_i \sigma_i / \delta_i$ is the gradient associated with $\hat{\gamma}_i$ in $E(\hat{\theta}_i | \hat{\gamma}_i)$. We will therefore refer to $\rho_i \sigma_i / \delta_i$ and β as the “within-study gradient,” and “between-study gradient,” respectively. Conceptually, if the estimated within and between-study gradients associated with the clinical outcome of interest on the surrogate are identical, no adjustment is made, via the bias adjustment terms b_i , when estimating τ^2 using maximum likelihood. Furthermore, the adjustment becomes larger as the magnitude of the difference between these two gradients increases. The bias adjustment terms therefore have a clear and intuitive interpretation. Our results indicate that maximum likelihood estimation of τ^2 can be expected to perform well when the dependence of the clinical outcome on the surrogate is in this sense similar within and between studies. The bias adjustment terms b_i are functions of parameter estimates, so that they may also be conceptualized as estimates, but we refrain from presenting them as \hat{b}_i because they do not estimate parameters of inferential interest.

5 | Simulation Study

We performed a small-scale simulation study to provide proof of concept that maximum likelihood estimates of τ^2 can be badly biased downwards, and further that the corresponding bias adjusted estimator addresses this problem.

5.1 | Simulation Study Design

We used $n = 20$ studies throughout, so that the Daniels and Hughes model can be adequately estimated while assuming a plausible sample size. We assumed that the true γ_i are uniformly distributed from -5 to 5 . We assumed that the within-study variances, $\sigma_i^2 = \delta_i^2 = 0.05, 0.1, \dots, 1.00$, are the same and fixed for each simulated data set. For simplicity, we took $\sigma_i^2 = \delta_i^2$ for all studies.

We varied the other four model parameters, using three values of $\beta = 0, 0.4, 0.8$ and similarly three different values of $\rho_i = 0, 0.4, 0.8$ (where the same ρ_i was assumed for all studies). From the theory derived in Section 4, maximum likelihood estimation of τ^2 should perform well in terms of bias when $\rho = \beta = 0, 0.4$, or 0.8 because then, in the absence of bias in $\hat{\beta}$, $\hat{\beta} \approx \rho_i \sigma_i / \delta_i$ so that $b_i \approx 0$. Otherwise the bias adjusted estimator should provide an improvement. We used three different values of $\tau^2 = 0, 0.5, 1$. If $\tau^2 = 0$, then, because $\hat{\tau}^2 \geq 0$, the conventional maximum likelihood estimator can only exhibit positive bias. Hence $\tau^2 =$

0 was considered, not because this is necessarily plausible, but because it was of interest to assess the extent that the bias adjusted estimator may exacerbate this. We used two different values of $\alpha = 0, 0.25$ but we did not expect this to have much impact on the estimation performance. We therefore examined $3 \times 3 \times 3 \times 2 = 54$ different scenarios. One thousand simulations were used in each setting, so that in total our methods were applied to 54,000 simulated data sets.

5.2 | Computational Details

Conventional maximum likelihood estimation was performed numerically where $\log(\hat{\tau}^2)$, and all other unknown parameters, may take any real value. The resulting $\log(\hat{\tau}^2)$ was then exponentiated to provide $\hat{\tau}^2 > 0$, where very large negative values of $\log(\hat{\tau}^2)$ correspond to $\hat{\tau}^2 = 0$. All numerical maximum likelihood estimates were then substituted into the estimating Equations (5, 9, A2, A3) to confirm that the correct solutions had been found. An advantage of using the estimating equations is that they can be used to remove all doubt that numerical optimizers have correctly converged; there is a separate γ_i for each study, so the dimension of the optimization problem can be high, and convergence may be difficult to correctly identify in practice.

Standard errors were computed as the inverse of the Hessian matrix, where this matrix was obtained numerically. Computing the standard error of $\log(\hat{\tau}^2)$ in this way however presents problems when $\hat{\tau}^2 = 0$, because small changes in very large and negative values of $\log(\hat{\tau}^2)$ result in negligible differences in the likelihood function. If this occurs, then all partial derivatives with respect to $\log(\tau^2)$ will be calculated to be zero, resulting in a Hessian that is not invertible. To avoid this difficulty, the generalized matrix inverse of the Hessian was used, which is numerically equivalent to constraining $\tau^2 = 0$ in the analysis when $\hat{\tau}^2 = 0$. A more computationally intensive alternative is to use methods based on the profile likelihood and we return to this possibility in the discussion. Wald-type 95% confidence intervals for α and β were computed using their point estimates, standard errors, and normal approximations.

An adjusted analysis was then performed constraining $\tau^2 = \hat{\tau}_{adjusted}^2$ and performing maximum likelihood estimation for the other model parameters, so that the bias adjusted estimate of τ^2 was used instead of the maximum likelihood estimator. Estimating Equations (5, A2, A3) were used, with this constraint, to confirm that numerical optimizers had correctly converged. Using the estimated between-study variance as the true value in this way is a common approximation in meta-analysis (Jackson and White 2018). Although the accuracy of this approximation may be questionable, for example, in situations where the number of studies is small, the greater concern is that a conventional maximum likelihood-based analysis may result in severe downward bias for τ^2 .

The main outcomes of interest were the biases of the estimates of α, β , and τ^2 , and the coverage probabilities of the 95% confidence intervals of α and β . Calculating confidence intervals for τ^2 is a complex issue even under the univariate random-effects model for meta-analysis and we return to this issue in the discussion.

We also compared our frequentist analyses to a more conventional Bayesian analysis. Here “vague,” and independent, prior distributions of $\gamma_i, \alpha, \beta \sim N(0, 1000)$ were used. In general the prior for τ^2 can be expected to be influential (Lambert et al. 2005). Here we followed Bujkiewicz, Achana, et al. (2019) and used the prior $\tau \sim U(0, 2)$, that is, a uniform prior for the between-study standard deviation. To make the Bayesian results more comparable with our frequentist analyses, we used the same values of ρ_i in our modeling, instead of using prior distributions for the ρ_i that may instead be used to allow for the possibility that these values are unknown. Our Bayesian analyses are merely intended to be representative, so that we can compare our frequentist methods to a typical type of Bayesian analysis. Posterior means, and coverage probabilities of 95% percentile-based credible intervals, can then be used to provide comparisons with frequentist biases and coverage probabilities. Posterior distributions of τ^2 can be skew and biases were also computed using the posterior median. Burn ins of 1000 Markov chain Monte Carlo iterations were used. A further 10,000 iterations were used to make inferences with a thinning frequency of two.

5.3 | Results

The full results are shown in the [Supporting Information](#). No concerns relating to bias when estimating α were encountered in any setting or using any estimation method (see the [Supporting Information](#)).

The largest biases when estimating β were produced by the Bayesian analysis, and the adjusted analysis also resulted in noticeably smaller biases than conventional maximum likelihood estimation (Figure 1, left). A closer examination (see the [Supporting Information](#)) revealed that all three methods performed more similarly in scenarios where $\beta = \rho_i \sigma_i / \delta_i = \rho_i$. From our analysis in Section 3, we know that this criterion has important implications for the maximum likelihood estimation of τ^2 . It appears that it may also have implications for Bayesian analyses that also use the likelihood function. We should interpret this finding cautiously, because simulation study results may not generalize to other settings, but the observation that our proposed adjusted frequentist method performed best in terms of bias when estimating β is an important finding.

The Bayesian analysis consistently resulted in positive bias when estimating τ^2 (Figure 1, right). This can be explained by the prior $\tau \sim U(0, 2)$ that gives considerable probability to larger values of τ^2 than explored in our simulation study. This bias was reduced by instead using the posterior median of τ^2 (see [Supporting Information](#)). The standard maximum likelihood analysis resulted in slight positive bias when estimating τ^2 when the true value was zero, as expected, and the bias adjusted method slightly increased this (Figure 1, right). However standard likelihood-based estimation resulted in large negative bias when the true $\tau^2 > 0$ and the adjusted method greatly reduced, but did not completely remove, this bias. A closer examination (see the [Supporting Information](#)) revealed that the bias adjustment made little difference in the scenarios where $\beta = \rho_i \sigma_i / \delta_i = \rho_i$, for which the downward bias of the maximum likelihood estimate of τ^2 was much less severe, as expected. We conclude that the estimation of τ^2 is challenging but our adjusted method greatly

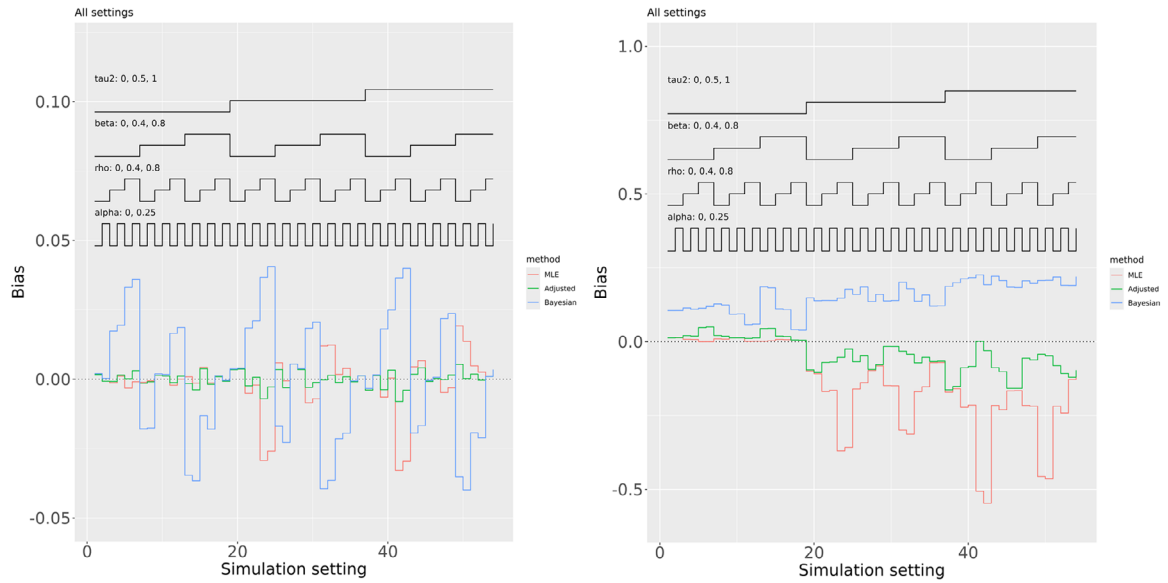


FIGURE 1 | Bias across all scenarios when estimating β (left) and τ^2 (right). Results for the maximum likelihood estimator (MLE) are shown in red, the bias adjusted (Adjusted) estimator are shown in green, and Bayesian (posterior mean) estimators are shown in blue.

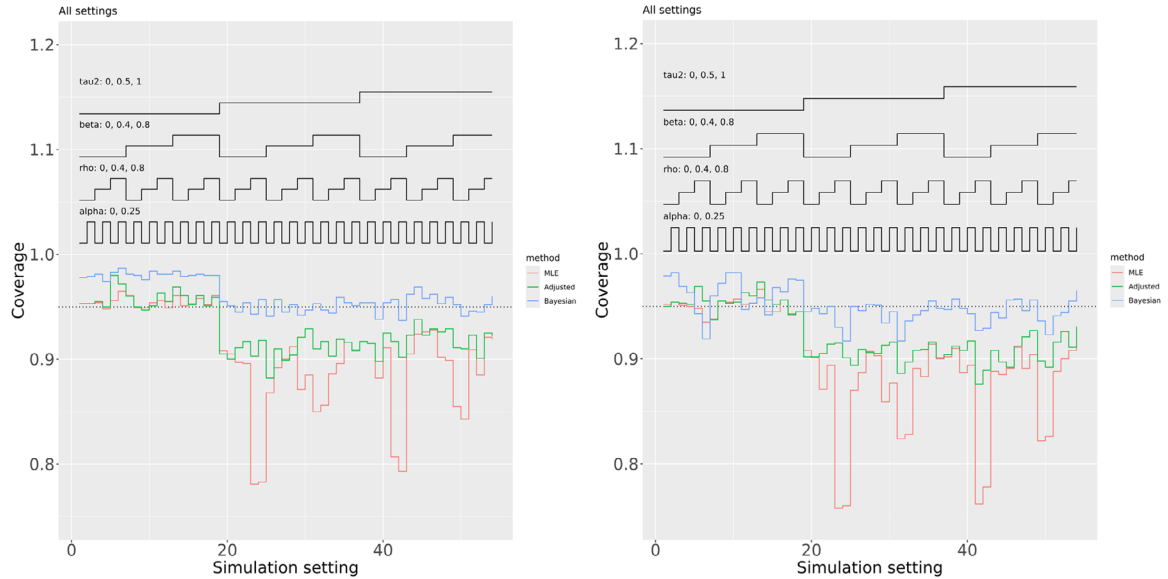


FIGURE 2 | Coverage probabilities for α (left) and β (right). Results for the maximum likelihood estimator (MLE) are shown in red, the bias adjusted (Adjusted) estimator are shown in green, and Bayesian (percentile based) estimators are shown in blue.

reduces the problems resulting from using maximum likelihood estimation and avoids concerns related to prior sensitivity in Bayesian analyses.

The results for the coverage probabilities of confidence and credible intervals are shown in Figure 2. The Bayesian analysis generally performed well in terms of the coverage probability of 95% credible intervals for α and β , where these probabilities were close to 0.95 when $\tau^2 > 0$ and overcoverage was observed when $\tau^2 = 0$. The conventional maximum likelihood analysis resulted in severe under coverage when $\tau^2 > 0$, for example, as low as under 0.8. The lowest adjusted analysis coverage probabilities were instead around 0.9, which is a considerable

improvement. The overall impression is that the bias adjustment greatly reduces the downward bias of the maximum likelihood estimation of τ^2 , and so avoids the severe under coverage of the confidence intervals for α and β , but does not entirely eliminate this. When $\tau^2 = 0$, the frequentist methods generally resulted in slight overcoverage for α but this was not evident for β . The two frequentist methods can only result in positive bias for τ^2 when this is zero, so that overcoverage for the other parameters might then be anticipated, but fully accommodating the uncertainty in many γ_i is likely to counteract this intuition. Ad hoc methods for increasing the coverage probability of confidence intervals, for example, by using the t distribution to compute them, might be worth considering

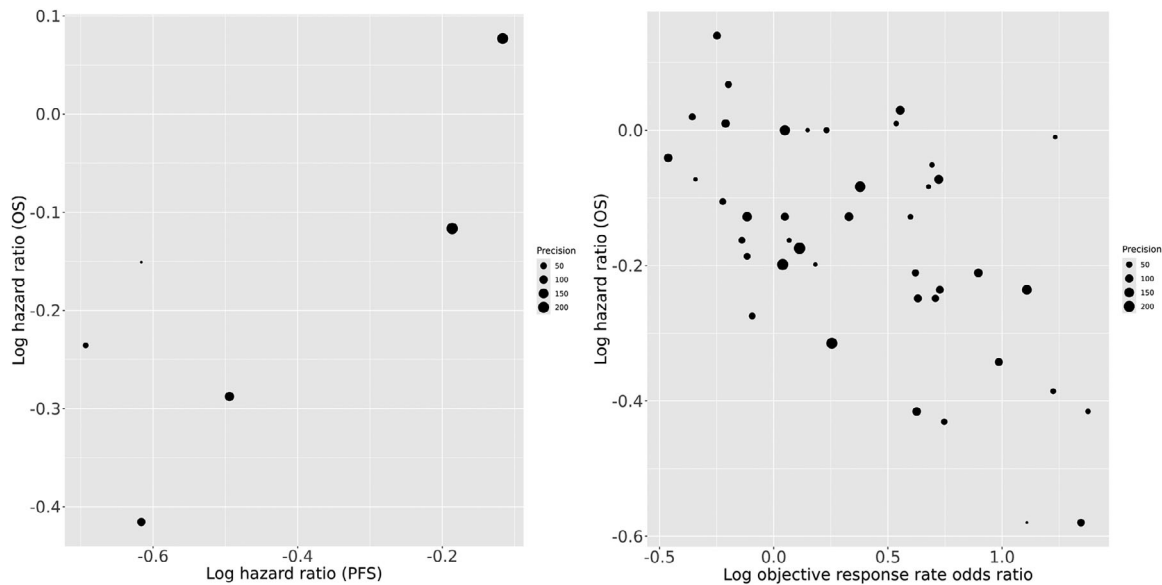


FIGURE 3 | Scatter plots for example 1 (left, Section 6.1) and example 2 (right, Section 6.2).

but any concrete proposal for this would need to be carefully justified.

To summarize, the proposed “bias adjusted method” performs better than conventional maximum likelihood estimation by greatly reducing its main problems. Bayesian estimation performed best in terms of coverage probabilities for α and β but raises concerns about sensitivity to prior distributions. The two frequentist methods, and the adjusted method in particular, performed better in terms of bias when estimating β . No method performed best in all respects but our position is that this simulation study indicates that our proposed bias adjusted method is, at the very least, a viable alternative to more conventional Bayesian estimation of the Daniels and Hughes model.

6 | Examples

6.1 | Example 1: Antiangiogenic Therapies for Metastatic Colorectal Cancer

The systematic review “Anti-angiogenic therapies for metastatic colorectal cancer” (Wagner et al. 2009) contains six studies that report hazard ratios for PFS and OS. A scatter plot of these six pairs of estimates (log hazard ratios) are shown in Figure 3 (left). A within-study correlation of $\rho_i = 0.513$ is available for one of the studies (Elia et al. 2020) but for the others this is unknown. We will perform an analysis assuming that the within-study correlations are similar across studies and so use $\rho_i = 0.513$ for all studies. Our analyses will explore if PFS is a potential surrogate endpoint for OS at the study level.

The results are shown in Table 1, where the same Bayesian method described in Section 5 was used for both this and the subsequent example. In addition to presenting posterior means and percentile based credible intervals (mean, %), as in the simulation study in Section 5, we also show posterior medians

and highest probability density (HPD)-based credible intervals. The adjusted and unadjusted frequentist estimates are $\hat{\tau}^2 = 0$ so that these two methods give the same results. This is because the six “within-study gradients,” $\rho_i \sigma_i / \delta_i$, take values in the interval $[0.453, 0.623]$. These values are not sufficiently numerically dissimilar to the $\hat{\beta} = 0.776$ for the bias adjustment to instead provide $\hat{\tau}_{adjusted}^2 > 0$. Although 95% credible intervals for τ^2 are shown in Table 1, the corresponding confidence intervals are omitted because the best way to compute them is left as further work. We return to this issue in the discussion. The Bayesian (mean %) and (median, HPD) results for α and β in Table 1 are similar but the results for τ^2 are notably different, and indicate that the posterior distribution of τ^2 is skew. This is not surprising because there are just six pairs of estimates contributing to the analysis.

Our frequentist analyses support the results from the Bayesian analysis that PFS may be a suitable surrogate endpoint for OS. In particular, they reinforce the conclusion that τ^2 is small, so that the variation in the true treatment effects on OS is explained by the corresponding treatment effects on PFS. Credible intervals for τ^2 are wide, however, so this finding should be treated with caution. Both types of the analysis provide $\hat{\alpha} \approx 0$, which is also supportive of surrogacy. The point estimates of β are also similar but the Bayesian analysis has resulted in much wider interval estimation. In an example such as this, where the number of studies is small, the Bayesian analysis can reasonably claim to better describe model uncertainty. The Bayesian credible intervals for β include zero, whereas the frequentist analyses exclude this.

The overall impression is that more evidence is required because confidence and, in particular, credible intervals are wide, but PFS may be a suitable surrogate endpoint for OS. Our adjusted analysis indicates that concerns relating to the downward bias of the maximum likelihood estimate of τ^2 is not an issue for this example.

TABLE 1 | Parameter estimates for example 1. Both frequentist analyses result in $\hat{\tau}^2 = 0$ and so give the same results. 95% confidence and credible intervals are shown in parentheses for α and β , and 95% credible intervals are also shown in parentheses for τ^2 .

Estimation method	α	β	τ^2
Frequentist (both)	0.102 (−0.025, 0.229)	0.776 (0.421, 1.130)	0
Bayesian (mean, %)	0.078 (−0.270, 0.401)	0.678 (−0.132, 1.369)	0.030 (0.000, 0.204)
Bayesian (median, HPD)	0.081 (−0.274, 0.395)	0.688 (−0.093, 1.405)	0.008 (0.000, 0.125)

6.2 | Example 2: Objective Response Rate and Survival-Based Endpoints in First-Line Advanced Non-Small Cell Lung Cancer

Goring et al. (2022) recently performed a systematic review and meta-analysis, “Correlations between objective response rate and survival-based endpoints in first-line advanced non-small cell lung Cancer: A systematic review and meta-analysis,” that includes 42 pairs of estimates. Here we reanalyze the data in their supplementary table A11 to explore whether the objective response rate is a potential surrogate endpoint for OS. A scatter plot of the 42 pairs of estimates (log objective response rate odds ratios and log OS hazard ratios) is shown in Figure 3 (right). The within-study correlations are unknown but can be expected to be negative, because we assume that increase in response rate implies a reduction in the mortality rate. Hence we will perform a sensitivity analysis where we take $\rho_i = -0.1, -0.3, -0.5, -0.7, -0.9$ for all studies. The values $\rho_i = -0.1$ and $\rho_i = -0.9$ are probably implausibly small and large, respectively, but it is of interest to assess the extent to which inferences are sensitive to the assumed within-study associations between the treatment effects on the potential surrogate endpoint and the final clinical outcome.

The results are shown in Table 2. The two frequentist methods give similar results for $\rho_i = -0.1, -0.3, -0.5$, where the bias adjusted estimates of τ^2 are only slightly larger than the maximum likelihood estimates. This is because the “within-study gradients,” $\rho_i \sigma_i / \delta_i$, take values in the intervals $[-0.073, -0.046]$, $[-0.220, -0.137]$, and $[-0.366, -0.228]$, for these three values of ρ_i respectively, that are not dissimilar to the corresponding maximum likelihood estimates of β (Table 2). However for $\rho_i = -0.7, -0.9$, the bias adjustment results in a much larger estimate of τ^2 and provides results that are in better agreement with the Bayesian analysis. This is especially evident for $\rho_i = -0.9$ where the maximum likelihood estimate of τ^2 is zero, whereas the corresponding bias adjusted estimate is $\hat{\tau}_{adjusted}^2 = 0.014$. The “within-study gradients,” $\rho_i \sigma_i / \delta_i$, take values in the intervals $[-0.513, -0.319]$ and $[-0.659, -0.410]$ for $\rho_i = -0.7$ and $\rho_i = -0.9$, respectively, that are notably different from the corresponding maximum likelihood estimates of β . Our proposed method has therefore had considerable impact for the last two values of ρ_i in our sensitivity analysis, where we observe better agreement with the Bayesian analysis when using our bias adjusted method. The Bayesian (mean, %) and (median, HPD) results are in much better agreement than the first example, as expected because many more pairs of estimates contribute to the analysis.

The overall impression is that our results support the conclusion from Goring et al. (2022) that objective response rate is associated with OS in this trial level meta-analysis ($\beta < 0$). Furthermore,

estimates of τ^2 are small. Estimates of α are not large but confidence and credible intervals do not include zero, which casts doubt on the use of the objective response rate as a surrogate for OS.

7 | Conclusions

We have developed a new frequentist estimation method for fitting the Daniels and Hughes bivariate meta-analysis model. The proposed estimation method is entirely based on this model but is otherwise fundamentally different to the current Bayesian implementation. Perhaps the main advantage of our approach is avoiding the need to specify prior distributions. Justifying particular priors for unknown variance components is challenging and in general inferences can be expected to be sensitive to this. The existing Bayesian approach, however, also has advantages. For example, it more naturally lends itself to decision analysis. We propose our estimation method as an alternative to, rather than a replacement of, a Bayesian analysis. A by-product of our work is that we are able to clarify the circumstances where a standard maximum likelihood based analysis is appropriate and so when more conventional frequentist analyses are feasible.

More sophisticated statistical modeling is possible, for example, by allowing more than one type of potential surrogate, multiple treatment classes, and/or considering models that make fewer normality assumptions. For example, alternative within-study models could be used to better describe binary outcome data and more sophisticated between-study regression models be used for the studies’ true treatment effects. Another important avenue for future research is to investigate the best methods for computing confidence intervals for the between-study variance, for which profile likelihood based confidence intervals may prove to be a good solution. A wide variety of methods are available for this under the univariate random-effects model (Veroniki et al. 2016). Extending some or all of these to the Daniels and Hughes model would be valuable. However this can be anticipated to be challenging because of the bias we have found in the corresponding maximum likelihood point estimate. More extensive simulation and empirical studies would also be highly worthwhile. This could include comparing the performance of methods based on different statistical models, for example the random-effects bivariate meta-analysis model and meta-regression.

Three model parameters have been used to make inferences about whether a potential surrogate endpoint is suitable but more guidance concerning how to assess this would be valuable. We have adopted the approach of considering each of these parameters in turn, but the development of alternative statistical procedures, for example, that consider these parameters simultaneously may

TABLE 2 | Parameter estimates for example 2. 95% confidence and credible intervals are shown in parentheses for α and β , and 95% credible intervals are also shown in parentheses for τ^2 .

Estimation method	ρ_i	α	β	τ^2
Frequentist (unadjusted)	-0.1	-0.088 (-0.137, -0.039)	-0.216 (-0.306, -0.126)	0.003
Frequentist (adjusted)	-0.1	-0.088 (-0.140, -0.036)	-0.214 (-0.308, -0.121)	0.004
Bayesian (mean, %)	-0.1	-0.097 (-0.148, -0.044)	-0.189 (-0.277, -0.103)	0.005 (0.000, 0.014)
Bayesian (median, HPD)	-0.1	-0.097 (-0.147, -0.044)	-0.188 (-0.278, -0.104)	0.004 (0.000, 0.012)
Frequentist (unadjusted)	-0.3	-0.094 (-0.143, -0.044)	-0.196 (-0.286, -0.107)	0.004
Frequentist (adjusted)	-0.3	-0.094 (-0.143, -0.044)	-0.196 (-0.286, -0.107)	0.005
Bayesian (mean, %)	-0.3	-0.095 (-0.149, -0.042)	-0.193 (-0.279, -0.102)	0.006 (0.001, 0.015)
Bayesian (median, HPD)	-0.3	-0.095 (-0.148, -0.041)	-0.193 (-0.279, -0.102)	0.005 (0.000, 0.013)
Frequentist (unadjusted)	-0.5	-0.100 (-0.150, -0.051)	-0.175 (-0.263, -0.086)	0.005
Frequentist (adjusted)	-0.5	-0.100 (-0.150, -0.050)	-0.176 (-0.265, -0.087)	0.006
Bayesian (mean, %)	-0.5	-0.093 (-0.143, -0.041)	-0.195 (-0.276, -0.110)	0.007 (0.001, 0.016)
Bayesian (median, HPD)	-0.5	-0.093 (-0.142, -0.040)	-0.195 (-0.276, -0.110)	0.006 (0.001, 0.014)
Frequentist (unadjusted)	-0.7	-0.112 (-0.160, -0.064)	-0.142 (-0.230, -0.054)	0.004
Frequentist (adjusted)	-0.7	-0.106 (-0.159, -0.052)	-0.157 (-0.249, -0.064)	0.009
Bayesian (mean, %)	-0.7	-0.092 (-0.143, -0.042)	-0.194 (-0.277, -0.111)	0.009 (0.003, 0.018)
Bayesian (median, HPD)	-0.7	-0.092 (-0.142, -0.041)	-0.194 (-0.275, -0.110)	0.008 (0.002, 0.017)
Frequentist (unadjusted)	-0.9	-0.151 (-0.193, -0.109)	-0.033 (-0.115, 0.048)	0
Frequentist (adjusted)	-0.9	-0.110 (-0.169, -0.050)	-0.144 (-0.243, -0.044)	0.014
Bayesian (mean, %)	-0.9	-0.091 (-0.141, -0.040)	-0.194 (-0.274, -0.114)	0.010 (0.004, 0.020)
Bayesian (median, HPD)	-0.9	-0.091 (-0.141, -0.040)	-0.194 (-0.272, -0.112)	0.010 (0.004, 0.018)

be worthwhile. For example, Daniels and Hughes (1997) used Bayes factors to test for associations. We have treated the γ_i as nuisance parameters but study specific effects may also be of interest (van Aert et al. 2021). Other statistical outputs are also often of interest, for example, prediction intervals (van Aert et al. 2021) and I^2 statistics (Higgins and Thompson 2002). We leave the best way to compute quantities such as these as another avenue for further work.

To summarize, our work provides a new way to fit the Daniels and Hughes bivariate meta-analysis model. We have applied our methodology, and a more conventional Bayesian analysis, to two examples. We have also performed a simulation study. Computing codes, in R, for both the examples and the simulation study are provided in the [Supporting Information](#). To the best of our knowledge, our methods provide the first frequentist implementation of the Daniels and Hughes model. This model is gaining traction and so our work fills an important void in the current state of statistical science.

Conflicts of Interest

Sylwia Bujkiewicz is a member of the NICE Decision Support Unit (DSU) and the NICE Guidelines Technical Support Unit (TSU). She has served as a paid consultant, providing methodological advice in relevant area, to NICE, Roche, RTI Health Solutions, and IQVIA and has received payments for educational events from Roche.

Data Availability Statement

The data that support the findings of this study are available in the [Supporting Information](#) of this article.

Open Research Badges



This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the [Supporting Information](#) section.

This article has earned an open data badge “**Reproducible Research**” for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

References

- Achana, F., N. Cooper, S. Bujkiewicz, et al. 2014. “Network Meta-Analysis of Multiple Outcome Measures Accounting for Borrowing of Information Across Outcomes.” *BMC Medical Research Methodology* 14: 92.
- Baker, S. 2018. “Five Criteria for Using a Surrogate Endpoint to Predict Treatment Effect Based on Data From Multiple Previous Trials.” *Statistics in Medicine* 37: 507–518.
- Belin, L., A. Tan, Y. De Rycke, and A. Dechartres. 2020. “Progression-Free Survival as a Surrogate for Overall Survival in Oncology Trials: A Methodological Systematic Review.” *British Journal of Cancer* 122: 1707–1714.
- Burke, D., S. Bujkiewicz, and R. Riley. 2018. “Bayesian Bivariate Meta-Analysis of Correlated Effects: Impact of the Prior Distributions on the

- Between-Study Correlation, Borrowing of Strength, and Joint Inferences.” *Statistical Methods in Medical Research* 27: 428–450.
- Burzykowski, T., M. Buyse, G. Molenberghs, H. Geys, and D. Renard. 2001. “Validation of Surrogate end Points in Multiple Randomized Clinical Trials With Failure Time end Points.” *Journal of the Royal Statistical Society, Series C* 50: 405–422.
- Burzykowski, T., M. Buyse, G. Molenberghs, A. Alonso, W. Van der Elst, and Z. Sckedy. 2021. “Meta-Analytic Approach to Evaluation of Surrogate Endpoints.” *Handbook of Meta-Analysis*, edited by C. Schmid, T. Stijnen, and I. White. Chapman and Hall, 457–477.
- Buyse, M., G. Molenberghs, T. Burzykowski, D. Renard, and H. Geys. 2000. “The Validation of Surrogate Endpoints in Meta-Analyses of Randomized Experiments.” *Biostatistics* 1: 49–67.
- Bujkiewicz, S., J. Thompson, A. Sutton, et al. 2013. “Multivariate Meta-Analysis of Mixed Outcomes: A Bayesian Approach.” *Statistics in Medicine* 32: 3926–3943.
- Bujkiewicz, S., J. Thompson, E. Spata, and K. Abrams. 2017. “Uncertainty in the Bayesian Meta-Analysis of Normally Distributed Surrogate Endpoints.” *Statistical Methods in Medical Research* 26: 2287–2318.
- Bujkiewicz, S., F. Achana, T. Papanikos, R. Riley, and K. Abrams. 2019. “Nice DSU Technical Support Document 20: Multivariate Meta-Analysis of Summary Data for Combining Treatment Effects on Correlated Outcomes and Evaluating Surrogate Endpoints.” Report by the Decision Support Unit.
- Bujkiewicz, S., D. Jackson, J. Thompson, et al. 2019. “Bivariate Network Meta-Analysis for Surrogate Endpoint Evaluation.” *Statistics in Medicine* 38: 3322–3341.
- Dai, J., and J. Hughes. 2012. “A Unified Procedure for Meta-Analytic Evaluation of Surrogate End Points in Randomized Clinical Trials.” *Biostatistics* 13: 609–624.
- Daniels, M., and M. Hughes. 1997. “Meta-Analysis for the Evaluation of Potential Surrogate Markers.” *Statistics in Medicine* 16: 1965–1982.
- Efthimiou, O., D. Mavradis, A. Cipriani, A. Leucht, P. Bagos, and G. Salanti. 2014. “An Approach for Modelling Multiple Correlated Outcomes in a Network of Interventions Using Odds Ratios.” *Statistics in Medicine* 33: 2275–2287.
- Elia, E., N. Stadler, O. Ciani, R. Taylor, and S. Bujkiewicz. 2020. “Combining Tumour Response and Progression Free Survival as Surrogate Endpoints for Overall Survival in Advanced Colorectal Cancer.” *Cancer Epidemiology* 64: 101665.
- Goring, S., N. Varol, N. Waser, et al. 2022. “Correlations Between Objective Response Rate and Survival-Based Endpoints in First-Line Advanced Non-Small Cell Lung Cancer: A Systematic Review and Meta-Analysis.” *Lung Cancer* 170: 122–132.
- Higgins, J., and S. Thompson. 2002. “Quantifying Heterogeneity in a Meta-Analysis.” *Statistics in Medicine* 21: 1539–1558.
- Li, Z., and M. Meredith. 2003. “Exploring the Relationship Between Surrogates and Clinical Outcomes: Analysis of Individual Patient Data vs. Meta-Regression on Group-Level Summary Statistics.” *Journal of Biopharmaceutical Statistics* 13: 777–792.
- Jackson, D., R. Riley, and I. White. 2011. “Multivariate Meta-Analysis: Potential and Promise.” *Statistics in Medicine* 30: 2481–2498.
- Jackson, D., I. White, and R. Riley. 2013. “A Matrix-Based Method of Moments for Fitting the Multivariate Random Effects Model for Meta-Analysis and Meta-Regression.” *Biometrical Journal* 55: 231–245.
- Jackson, D., and I. White. 2018. “When Should Meta-Analysis Avoid Making Hidden Normality Assumptions?” *Biometrical Journal* 60: 1040–1058.
- Jackson, D., M. Law, T. Stijnen, W. Viechtbauer, and I. White. 2018. “A Comparison of Seven Random-Effects Models Formeta-Analyses That Estimate the Summary Odds Ratio.” *Statistics in Medicine* 37: 1059–1085.
- Jackson, D., S. Bujkiewicz, M. Law, R. Riley, and I. White. 2018. “A Matrix-Based Method of Moments for Fitting the Multivariate Random Effects Model for Meta-Analysis and Meta-Regression.” *Biometrics* 74: 548–556.
- Jennrich, R., and M. Schluchter. 1986. “Unbalanced Repeated-Measures Models With Structured Covariance Matrices.” *Biometrics* 42: 805–820.
- Lambert, P., A. Sutton, P. Burton, K. Abrams, and D. Jones. 2005. “How Vague is Vague? A Simulation Study of the Impact of the Use of Vague Prior Distributions in MCMC Using WinBUGS.” *Statistics in Medicine* 24: 2401–2428.
- Molenberghs, G., T. Burzykowski, A. Alonso, P. Assam, A. Tilahun, and M. Buyse. 2008. “The Meta-Analytic Framework for the Evaluation of Surrogate Endpoints in Clinical Trials.” *Journal of Statistical Planning and Inference* 138: 432–449.
- Neyman, J., and E. Scott. 1948. “Consistent Estimates Based on Partially Consistent Observations.” *Econometrica* 16: 1–32.
- Pozzi, L., H. Schmidli, and D. Ohlssen. 2016. “A Bayesian Hierarchical Surrogate Outcome Model for Multiple Sclerosis.” *Pharmaceutical Statistics* 15: 341–348.
- Papanikos, T., J. Thompson, K. Abrams, et al. 2020. “Bayesian Hierarchical Meta-Analytic Methods for Modeling Surrogate Relationships That Vary Across Treatment Classes Using Aggregate Data.” *Statistics in Medicine* 39: 1103–1124.
- Prentice, R. 1989. “Surrogate Endpoints in Clinical Trials: Definition and Operational Criteria.” *Statistics in Medicine* 8: 431–440.
- Riley, R., J. Thompson, and K. Abrams. 2008. “An Alternative Model for Bivariate Random-Effects Meta-Analysis When the Within-Study Correlations are Unknown.” *Biostatistics* 9: 172–186.
- Riley, R. 2009. “Multivariate Meta-Analysis: The Effect of Ignoring Within-Study Correlation.” *Journal of the Royal Statistical Society, Series A* 172: 789–811.
- Riley, R., M. Price, D. Jackson, et al. 2015. “Multivariate Meta-Analysis Using Individual Participant Data.” *Research Synthesis Methods* 6: 157–174.
- Thompson, S., and S. Sharp. 1999. “Explaining Heterogeneity in Meta-Analysis: A Comparison of Methods.” *Statistics in Medicine* 18: 2693–2708.
- Taylor, R. S., and J. Elston. 2009. “The use of Surrogate Outcomes in Model-Based Cost-Effectiveness Analyses: A Survey of UK Health Technology Assessment Reports.” *Health Technology Assessment* 13: 8.
- van Aert, R., C. Schmid, D. Svensson, and D. Jackson. 2021. “Study Specific Prediction Intervals for Random-Effects Meta-Analysis: A Tutorial.” *Research Synthesis Methods* 12: 429–447.
- Veroniki, A., D. Jackson, W. Viechtbauer, et al. 2016. “Methods to Estimate the Between-Study Variance and Its Uncertainty in Meta-Analysis.” *Research Synthesis Methods* 7: 55–79.
- Wagner, A., D. Arnold, A. Grothey, J. Haerting, and S. Unverzagt. 2009. “Anti-Angiogenic Therapies Formetastatic Colorectal Cancer.” *Cochrane Database of Systematic Reviews* 3: CD005392.
- Wei, Y., and J. Higgins. 2013. “Estimating Within-Study Covariances in Multivariate Meta-Analysis With Multiple Outcomes.” *Statistics in Medicine* 32: 1191–1205.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Appendix A: Deriving Estimating Equations for α and β

Write $o_i(\hat{\gamma}_i) = (\rho_i \sigma_i / \delta_i)(\hat{\gamma}_i - \gamma_i)$ and $V_i^2 = v_i^2 + \tau^2$ in model (6), so that

$$\hat{\theta}_i | \hat{\gamma}_i \sim N(\alpha + \beta \gamma_i + o_i(\hat{\gamma}_i), V_i^2). \quad (\text{A1})$$

The marginal model $\hat{\gamma}_i \sim N(\gamma_i, \delta_i^2)$ does not depend on α or β . When differentiating the log-likelihood function with respect to either of these parameters, only terms from the n conditional densities $\hat{\theta}_i|\hat{\gamma}_i$ contribute. The maximum likelihood estimates of α and β therefore satisfy the maximum likelihood estimating equations from model (A1), where other parameters are replaced by their maximum likelihood estimates. Hence, if we define $\hat{o}_i(\hat{\gamma}_i) = (\rho_i \sigma_i / \delta_i)(\hat{\gamma}_i - \hat{\gamma}_{i,MLE})$, $\hat{\theta}_i^\dagger = \hat{\theta}_i - \hat{o}_i(\hat{\gamma}_i)$, $w_i = \hat{V}_i^{-2}$, $\bar{\theta}^\dagger$ and $\bar{\gamma}_{MLE}$ to be the *weighted* by w_i mean of the $\hat{\theta}_i^\dagger$ and $\hat{\gamma}_{i,MLE}$, respectively, the maximum likelihood estimates of α and β satisfy the estimating equations

$$\hat{\beta} = \frac{\sum_{i=1}^n w_i (\hat{\theta}_i^\dagger - \bar{\theta}^\dagger)(\hat{\gamma}_{i,MLE} - \bar{\gamma}_{MLE})}{\sum_{i=1}^n w_i (\hat{\gamma}_{i,MLE} - \bar{\gamma}_{MLE})^2} \quad (\text{A2})$$

and

$$\hat{\alpha} = \bar{\theta}^\dagger - \hat{\beta} \bar{\gamma}_{MLE}. \quad (\text{A3})$$

Estimating Equations (A2) and (A3) can be used to estimate α and β , similar to the way that estimating Equation (5) can be used to estimate $\hat{\gamma}_{i,MLE}$.

Appendix B: A Mathematical Result Required to Derive a Bias Adjusted Estimating Equation for τ^2

As explained in Section 4, we treat $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\tau}^2$ as fixed constants but this is not the case for the $\hat{\gamma}_{i,MLE}$. Let w_{1i} and w_{2i} be the first and second entries of the row vector $(\mathbf{C}_{(\hat{\beta},1)}^T \hat{\mathbf{W}}_i \mathbf{C}_{(\hat{\beta},1)})^{-1} \mathbf{C}_{(\hat{\beta},1)}^T \hat{\mathbf{W}}_i$, respectively. Then, from estimating Equation (5), we have

$$\hat{\gamma}_{i,MLE} = w_{1i} \hat{\theta}_i^* + w_{2i} \hat{\gamma}_i = w_{1i}(\hat{\theta}_i - \hat{\alpha}) + w_{2i} \hat{\gamma}_i \quad (\text{B1})$$

where in estimating Equation (9) we evaluate

$$\hat{\mu}_i(\hat{\gamma}_i) = \hat{\alpha} + \hat{\beta} \hat{\gamma}_{i,MLE} + (\rho_i \sigma_i / \delta_i)(\hat{\gamma}_i - \hat{\gamma}_{i,MLE}). \quad (\text{B2})$$

Substituting Equation (B1) into Equation (B2), and some algebra, gives

$$\hat{\mu}_i(\hat{\gamma}_i) = a_i + b_i \hat{\theta}_i + c_i \hat{\gamma}_i, \quad (\text{B3})$$

where $a_i = \hat{\alpha}(1 - (\hat{\beta} - \rho_i \sigma_i / \delta_i)w_{1i})$, $b_i = (\hat{\beta} - \rho_i \sigma_i / \delta_i)w_{1i}$, $c_i = (\hat{\beta} - \rho_i \sigma_i / \delta_i)w_{2i} + \rho_i \sigma_i / \delta_i$. All a_i , b_i , and c_i are treated as fixed constants, because $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\tau}^2$ are treated as such.

The term $\hat{\mu}_i(\hat{\gamma}_i) = a_i + b_i \hat{\theta}_i + c_i \hat{\gamma}_i$ is an estimate of the conditional mean of $\hat{\theta}_i|\hat{\gamma}_i$. We assume this estimate is unbiased, so that $E((\hat{\theta}_i - a_i - b_i \hat{\theta}_i - c_i \hat{\gamma}_i)|\hat{\gamma}_i) = 0$. Hence

$$\begin{aligned} E((\hat{\theta}_i - \hat{\mu}_i(\hat{\gamma}_i))^2|\hat{\gamma}_i) &= \text{Var}((1 - b_i)\hat{\theta}_i|\hat{\gamma}_i) = (1 - b_i)^2 \text{Var}(\hat{\theta}_i|\hat{\gamma}_i) \\ &= (1 - b_i)^2 (v_i^2 + \tau^2), \end{aligned}$$

so that

$$(1 - b_i)^{-2} E((\hat{\theta}_i - \hat{\mu}_i(\hat{\gamma}_i))^2|\hat{\gamma}_i) = v_i^2 + \tau^2. \quad (\text{B4})$$

Equation (B4) is the key mathematical result required to motivate our proposed bias adjusted estimator.