Title
Frequentist and Bayesian Meta-Analyses with Dynamic Variance

Author
Turitz, Michelle Rachel

Publication Date
2020

UNIVERSITY OF CALIFORNIA, MERCED

Frequentist and Bayesian Meta-Analyses with Dynamic Variance

A Thesis submitted in partial satisfaction of the requirements for the degree of Master of Arts

in

Quantitative Psychology

by

Michelle Turitz Mitchell

Committee in charge:

Professor Jack Vevea, Chair
Professor Sarah Depaoli
Professor Keke Lai

2020

The Thesis of Michelle Turitz Mitchell is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Sarah Depaoli

Keke Lai

Jack Vevea, Chair

University of California, Merced
2020

Contents

Figures

Abstract

Frequentist and Bayesian Meta-Analyses with Dynamic Variance
Michelle Turitz Mitchell
Masters, Psychological Sciences
University of California, Merced 2020
Committee Chair: Jack Vevea

When conducting a meta-analysis, the analyst uses the included studies' sampling variances for estimating the meta-analytic model. For some effect size measures, such as correlations and standardized mean differences, the variance depends on the population's true effect size, which is likely unknown. This study uses a new "dynamic method" for mixed-effects meta-analysis for correlations and standardized mean differences that takes this dependence into account. Comparing results using the dynamic method with other commonly used meta-analysis methods in the frequentist and Bayesian framework shows that the dynamic approach has improved estimation, particularly with correlation effect-size. Another finding of note is that the incorporation of Hedges bias correction (1981) standardized mean differences is shown to increase bias in conventional meta-analytic results.

General Introduction

Meta-analysis involves the quantitative synthesis and analysis of results obtained from multiple research studies. By combining the standardized effect-size estimates from various studies, one can obtain an estimate based on the results of multiple studies rather than a single study. However, no two studies are identical, and some studies will show greater precision than others. Most meta-analytic techniques (whether fixed effect or random effects) apply a weight to the studies with increased weight ascribed to studies with higher precision when combining effect-sizes.

The standard method is to weight studies by the inverse of their variances,

$$w_i = \frac{1}{\hat{\sigma}_i^2}, \tag{1}$$

where $\hat{\sigma}_i^2$ represents the total variance for study i. The exact calculation for the individual study weight differs based on the type of meta-analysis (i.e., random effects, fixed effect). For the fixed-effect meta-analytic model the total estimated variance would be equal to the within-study conditional variance (v_i). In random-effects meta-analysis $\hat{\sigma}_i^2$ would be the sum of the between-study variance component (τ^2) and v_i . Both the fixed-effect and random-effects models $\hat{\sigma}_i^2$ include v_i in the computation, meaning that the conditional variance influences the weight of studies for both fixed-effect and random-effects models.

This study focuses on one problem often overlooked within meta-analyses, combining effect size estimates when studies sampling variances are dependent on the actual parameter value of the effect size for the study. While for some measures of effect size (e.g., odds ratios, Fisher's *Z* transformed correlations) the approximation of the standard error is conditional solely on sample size, other common effect types (e.g., correlations, standardized mean differences) include a term in their function whereby the standard error is dependent on the "true" magnitude of the effect in the population. This true population magnitude is unknown; in actuality, it is what we are trying to estimate, and this dependence presents a problem.

One commonly used solution to address this dependence is to substitute the individual studies' effect sizes in place of the population effect size. In this case, the magnitude of the true population effect size remains unknown. The meta-analyst uses the individual study estimates in its place, treating them as if they are the known parameter value for the effect size. Another solution often used for correlation effect size is to transform the effect size estimates to the Fisher Z metric, where the variance does not depend on the effect-size parameter.

Meta-Analysis

Meta-analysis consists of methods for quantitatively describing, combining, and summarizing results across multiple studies. The meta-analytic mean estimate uses the synthesized results of multiple studies instead of just a single study. This summary not only provides an estimate of the measure of the parameter but can also address the inconsistency of results among studies.

In a meta-analysis, independent study effect sizes (T_i)[1] are distributed around a

[1] T_i represents a general study effect-size estimate and applies to multiple measures of effect sze(e.g. standardized mean difference or correlation)

mean population effect-size parameter. The meta-analyst must choose between fixed-effect and random-effects models. The decision between fixed- and random-effects meta-analysis should depend on the goal of inference (Hedges & Vevea, 1998). It is usually the case that people are interested in drawing inferences outside the available studies included in the literature, in which case random-effects meta-analysis is the most appropriate choice.

In a fixed-effect meta-analysis, the only source of variation comes from the error associated with the sampling of individual units (e.g., persons) into primary studies (v_i). All study effect sizes are distributed around a fixed parameter (θ)[2] for effect-size in the population. If T_i is the effect size for individual study i, $T_i \sim N(\theta, v_i)$ where v_i is the within-study sampling variance for study i. The model can also accommodate systematic heterogeneity by incorporating moderators of effect size, in which case $T_i \sim N(\boldsymbol{X\beta}, v_i)$, where $\mathbf{X}$ is a matrix of known study characteristics and $\boldsymbol{\beta}$ is a vector of regression coefficients, allowing the model to incorporate a conditional mean ($\mu_{\theta i} = \beta_0 + x_{i1}\beta_1 + \cdots x_{ip}\beta_p$).

In contrast to the fixed-effect model, random-effects meta-analysis does not assume a fixed mean parameter for all studies, allowing for different parameters for each study. Each study effect-size estimate has a corresponding effect size parameter (θ_i) representatively sampled from a distribution of population effect sizes with mean μ_θ. The random-effects model can be represented as a two-level model. At level one

$$T_i = \theta_i + e_i, \quad e_i \sim N(0, v_i), \tag{2}$$

where v_i is the within-study conditional variance of the individual study estimate (T_i) around the corresponding population parameter (θ_i). In the equation for the random-effects model, θ_i replaces θ from the fixed-effects meta-analyses indicating that instead of a single value, each study has a corresponding parameter estimate. The sampling error between the study estimate and its corresponding parameter (e_i) approximates a normal distribution with a mean of zero and a variance equal to the conditional variance of the study. At level two,

$$\theta_i = \mu_\theta + u_i, u_i \sim N(0, \tau^2), \tag{3}$$

where μ_θ is an overall mean effect-size parameter and u_i is the random variation between θ_i and μ_θ. The between-study variance component τ^2 represents the variation of the study level effect-size parameters ($\theta_1 \ldots . \theta_k$) around the overall mean effect-size parameter (μ_θ), as illustrated in Figure 1

Researchers can also incorporate study characteristics into the model with a mixed-effects meta-analysis. The model for a mixed-effects meta-analysis is an extension of the simple random-effects model described in Equations 2 and **Error! Reference source not found.**. Level one still represents the study estimate around the corresponding parameter **Error! Reference source not found.**. Level two is a regression model

$$\theta_i = \beta_0 + \beta_1 X_1 \ldots . \beta_p X_P + u_i, \tag{4}$$

where $X_i \ldots . X_p$ are study characteristics, β_0 is the intercept, and $\beta_1 \ldots \beta_p$ are regression coefficients indicating the predictive influence of study characteristics on effect size.

[2] θ here represent a general parameter value for the effect size. For specific notation ρ is often used to represent correlation, and δ standardized mean difference. θ indicates that the concept is more general and not specific to a single measure of effect size.

Effect Size Indices

Two commonly used measures of effect size that have the property of the conditional variance being dependent on the effect size parameter for the individual study θ_i are standardized mean differences and correlations[3]. There are various expressions for the sampling variance of the standardized mean difference; one commonly used one is

$$v_{\delta i} = \frac{N_{i1} + N_{i2}}{N_{i1}N_{i2}} + \frac{\delta_i^2}{2(N_{i1} + N_{i2} - 2)}, \tag{5}$$

where δ represents the corresponding population parameter, and N_1 and N_2 are the sample sizes for the conditions. The sampling variance for correlations is

$$v_{\rho i} = \frac{(1 - {\rho_i}^2)^2}{N_i - 1}, \tag{6}$$

where ρ and N are the corresponding effect size parameter and sample size for study i.[4]

The parameter, of course, is unknown, which presents a problem. One common solution for standardized mean difference and correlation estimates is to use the study estimate in place of the parameter. This estimate would be treated as a known value, although in actuality it is not:

$$\hat{v}_{di} = \frac{N_{1i} + N_{2i}}{N_{1i}N_{2i}} + \frac{d_i^2}{2(N_{i1} + N_{i2} - 2)}, \tag{7}$$

and

$$\hat{v}_{ri} = \frac{(1 - r_i^2)^2}{N_i - 1}. \tag{8}$$

Another potential solution when combining correlations effect sizes is to do Fisher's *r*-to-*Z* variance-stabilizing transformation, which has the benefit of the sampling variance not being dependent on the parameter. The approximate sampling variance of the Fishers *r*-to-*Z* transformed estimate is

$$\sigma_{zi}^2 = \frac{1}{N_i - 3}, \tag{9}$$

where *N* is the sample size. While some researchers recommend this variance-stabilizing transformation (Lipsey & Wilson, 2001; Shadish & Haddock, 2009), other research points to findings of substantial positive bias using Fishers *r*-to-*Z* when heterogeneity is high (Field, 2001; Hunter & Schmidt, 2004). Hafdahl (2009) suggested that the bias was due to the direct z-to-r transformation and introduced an integral z-to-r transformation where

$$\Psi_\zeta(\hat{\mu}_\zeta|\hat{\tau}_\zeta^2 = \int_{-\infty}^{\infty} \tanh(\zeta)\, f\left(\zeta|\hat{\mu}_\zeta, \hat{\tau}_\zeta^2\right) d\zeta \ , \tag{10}$$

and $f\left(\zeta|\hat{\mu}_\zeta, \hat{\tau}_\zeta^2\right)$ is the density function for $\zeta \sim N(\hat{\mu}_\zeta, \hat{\tau}_\zeta^2)$. This integral z-to-r

[3] θ_i suggests that each individual study has is associated with a different effect size parameter. This is true of random-effects meta-analysis. However, in fixed-effect meta-analysis all studies are assumed to come have the same population parameter (θ_1=θ_2=….=θ_i).

[4] In equations 5 and 6 we are now using δ to represent the effect size parameter for standardized mean difference and ρ for correlation. These equations are specific to a single measure of effect size, and the general notation θ will no longer work.

transformation is less biased than the direct z-to-r transformation (Hafdahl,2010; Hafdahl, 2009; Hafdahl & Williams, 2009).

The magnitudes of the study effect sizes for standardized mean difference and correlations are directly related to the calculation of the conditional variance (See Equations 7 and 8) and subsequently play an important role in determining the meta-analytic weight. For correlation effect-size estimates, higher values of effect magnitude lead to decreased variances and hence larger weights. For analysis of correlations, if the meta-analyst substitutes the study effects, studies that found an extremely high correlation have greater weight, even in cases where extreme values are due to the vicissitudes of sampling. In contrast, for standardized mean differences, more substantial study effects lead to lower weights in the analysis. In some cases, the influence of a single study with an extreme value can result in biased estimates of the mean effect and variance components. With the dynamic method v_i continues to be dynamically dependent on the estimate of μ_θ. The dynamic dependence of the conditional variance on μ_θ instead of individual study effect size estimates reduces the influence of extreme effect-size estimates obtained from individual studies.

The Dynamic Method

In a previous simulation study, Mitchell and Vevea (unpublished, 2016) proposed a new "*Dynamic*" method for meta-analysis of effect sizes for which sampling variances are dependent on the effect-size parameter. Results compared dynamic and conventional estimates of simple random-effects meta-analyses with correlation and standardized mean difference, with the dynamic method found to perform at least as well as or better than the other methods in almost all conditions.

Mixed- and random-effects models estimated by maximum likelihood involve iterative procedures. If we assume that effect sizes are independent and normally distributed at both levels of the random-effects model, the joint density function is

$$f\left(\mathbf{T}\middle|\mu_{\theta,}\tau^2;\mathbf{v}\right) = \prod_{i=1}^{k}\left[\frac{1}{\sqrt{2\pi(\mathbf{v}+\tau^2)}} * \exp\left(-\frac{1}{2} * \frac{(\mathbf{T}-\mu_\theta)^2}{\mathbf{v}+\tau^2}\right)\right], \qquad (1)$$

where $\mathbf{T}$ is a vector of effect-size estimates, $\mathbf{V}$ is a vector of conditional variances, k is the number of effects in the analysis, μ_θ is the population mean of the random-effects distribution, and τ^2 is the variance component that describes the dispersion of true effects in the random-effects distribution. The log-likelihood function is proportional to

$$logL\left(\mu_{\theta,}\tau^2\middle|\mathbf{T};\mathbf{v}\right) \propto \sum\left[\log(\mathbf{v}+\tau^2) + \frac{(\mathbf{T}-\mu_\theta)^2}{\mathbf{v}+\tau^2}\right]. \qquad (2)$$

The analyst can obtain the maximum-likelihood estimate using an iterative quasi-Newton optimizer. For the current study, the nlminb optimizer in the base R package is used (R Core Team, 2013). The model is converged when the partial derivatives are close to 0, and the maximum-likelihood estimates of the mean and variance component change between iterations by less than a specified predefined criterion set to be virtually zero.

As noted before, the within-study sampling variance v_i is conditional on the effect-size parameter θ_i. Conventional estimates of v_i substitute the study effect size T_i for θ_i, with the estimates treated as if they are a known parameter by the analyst. With the dynamic method, estimation of the within-study variances incorporates the current maximum-likelihood estimate of the true effect (for fixed-effect models) or mean effect (for random-effects models) at every iteration

Current Research

The purpose of this study is to compare methods of estimation for the dynamic method with conventional estimates for mixed-effects meta-analysis for two measures of effect size, correlation and standardized mean difference. Both correlation and standardized mean difference effect-size estimates have the property of dependency of the conditional variance on the true mean effect. Including study-level predictors in the model can account for differences that exist due to study characteristics and analyzing how the dynamic method estimates mixed-effects models will be useful for assessing the utility of the dynamic method.

Another area of focus in the current study is the use of the dynamic method within the Bayesian framework. A particular strength of the Bayesian approach is the ability to incorporate prior knowledge into the analysis, which is particularly useful when τ^2 is difficult to estimate (Higgins & Thompson, 2002). The dynamic method is easy to integrate into the Bayesian framework, which allows for the comparison of Bayesian *Conventional* and *Dynamic* estimates.

Correlation

For correlation effect-size, the current study compares the results using four methods of estimation: the conventional approach (*Conv*)[5], the dynamic method (*Dynamic*), Fisher's r-to-Z back-transformed correlations (*DirZtoR*), and an integral z-to-r transformation (*IntZtoR*) (Hafdahl, 2009). The integral back-transformed r-to-z developed by Hafdahl (2009) has been found to have less bias then estimation using the direct r-to-z method using the hyperbolic tangent[6] (Field, 2001; Hafdahl, 2009; Hunter & Schmidt, 2004).

Standardized Mean Difference

For standardized mean differences, Cohen's *d* is the conventional effect-size estimate. A bias-corrected estimate referred to as Hedges' *g* was developed by Hedges (1981), to correct for a positive small-sample bias in Cohen's *d*. Both the biased and unbiased forms of estimation can be applied using both the conventional and dynamic methods. For the current research, there are four methods of comparison for standardized mean difference effect-size estimates: conventional D (Conv*D)*, bias-corrected conventional G *(ConvG),* the uncorrected dynamic estimate (*DynD*), and the bias-corrected dynamic estimate (*DynG*). For the conventional estimated (*ConvD, ConvG*), the individual study effect size is substituted for the true effect size when computing the conditional variance, as shown in **Error! Reference source not found.**. In contrast, the dynamic methods (*DynD, DynG*) leave the v_i dependent on the current estimate of the true mean effect μ_θ.

Priors for Bayesian Analysis

A strength of the Bayesian approach is the ability to incorporate prior knowledge, but the question remains of where to obtain information about appropriate priors. In this study, we investigate the influence of different priors on the τ^2, in addition to assessing how the dynamic method translates into the Bayesian context. Priors on the variance

---

[5] where the study estimate is used in the formula for the sampling variance

[6] The hyperbolic tangent is what is commonly used to back transform Fisher's-z estimates to the r metric.

component are difficult to estimate.

Priors represent the prior probability assigned to the results and are determined based on one's current knowledge and beliefs about the phenomenon of interest. When there is no previous knowledge to incorporate into the model, vague priors are used, letting the data speak for itself. This study mainly focuses on manipulating priors for the variance component (precision), which can be challenging to define[7]. The accuracy of priors for the regression coefficients will also impact the outcome, so to alleviate the influence of priors for the regression coefficients, these priors are fixed as vague normally distributed priors for all conditions ($\beta_i \sim N(0, 1000^2)$), where β_i represents the regression coefficient, 0 is the mean, and 1000^2 is the variance.

In the current study, I utilize two different priors on precision for comparison of conventional and *Dynamic* conditions. The first prior used on precision is a flat vague uniform prior:

$$\frac{1}{\tau^2} \sim U(0,10). \tag{3}$$

A conjugate prior is one for which the prior distribution is in the same family as the posterior distribution. An inverse gamma prior is a conjugate prior for the variance of a normal distribution (Cooper, Hedges, & Valentine, 2009; Gelman, 2006). The second prior included uses estimates of the variance components reported in the literature of published meta-analyses. In a 2017 study, van Erp, Verhagen, Grasman, and Wagenmakesr collected data from meta-analyses published in Psychological Bulletin between 1990 and 2013, including data on heterogeneity, which is used to create a somewhat informative conjugate literature-based prior. To determine the prior, I first computed precision from the included published values for correlation and standardized mean difference. I then fit a gamma distribution to those values to determine the prior (See Appendix 2) and resulted in the following priors:

Correlation: $\frac{1}{\tau^2} = \Gamma(0.2372422, 0.0001740984),$ *(4)*

Standardized Mean Difference: $\frac{1}{\tau^2} = \Gamma 0.2509973, 0.0004352425).$ *(5)*

[7] The variance component is τ^2. However, Jags software uses precision which is $\frac{1}{\tau^2}$.

Methods

Variables in Current Research

The parameters and variables manipulated in the simulation are the same across all methods. Variables manipulated include the five levels for the number of studies, four levels for heterogeneity, and five levels for the magnitude of the intercept (5 X 4 X 5) for a total of 100 cells each for correlation and standardized mean difference.

The Number of Studies (k)

The simulation considers five different levels for the number of studies: $k = 10$, 20, 50, 80, and 150 (where k denotes the number of study effects in the analysis). These levels are approximated based on data provided by Coburn, Vevea, & Orrey (personal communication) on the number of studies included in published meta-analyses from two journals: British Medical Journal and Psychological Bulletin. The quartiles for the British Medical Journal ($Q_1 = 11$, $Q_2 = 21$, $Q_3 = 46$) and Psychological Bulletin ($Q_1 = 47.75$, $Q_2 = 84.5$, $Q_3 = 152$). The third quartile from the British Medical Journal and the first quartile from Psychological Bulletin both contain values close to k=50, resulting in five levels, approximating the values of the first, second, and third quartiles from these two journals.

Heterogeneity (τ)

Heterogeneity is the amount of between-study variation among studies included in a meta-analysis that is not associated with sampling individuals in the original studies. There are different ways to describe heterogeneity. Different statistics exist for characterizing the amount of heterogeneity in a meta-analytic data set. For the current study, I use different levels of τ, defined as the square root of the variance component (τ^2) and representing the standard deviation of the true underlying effects. Levels of τ for correlation simulations are 0, .05, .1 and .2. For standardized mean difference the levels for τ are 0, .1, .2, and .3.

Conditional Means

It is important to note that the mixed-effects model used in this simulation includes two dichotomous categorical study characteristics as predictors (e.g., male/female, yes/no). The coding for predictors consists of one or more repetitions of the following matrix:

$$
\begin{array}{ccc}
X_0 & X_1 & X_2 \\
\end{array}
$$
$$
\left|\begin{array}{ccc}
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
1 & 0 & 0 \\
1 & 1 & 1 \\
1 & 1 & 0 \\
1 & 1 & 1 \\
1 & 1 & 0 \\
1 & 1 & 1
\end{array}\right|
$$

For all cells in the simulation, the number of studies is a multiple of 10. Because of this property, the design matrix above is easily adaptable to all conditions by repeating the matrix the required number of times for the condition. By using the same base design matrix in every condition, this assures that the correlation between the predictors remains constant for every condition in the simulation. The correlation between the predictors also remains fixed at r =.2 across all conditions. Experimental studies will often have predictors that are correlated, and the inclusion of correlated predictors reflects this in the simulation.

Since there are two dichotomous predictors, there are also four different subgroups based on study characteristics: Group 1$(X_1 = 0, X_2 = 0)$, with the conditional mean equal to β_0 . Group 2$(X_1 = 0, X_2 = 1)$, with the conditional mean equal to $\beta_0 + \beta_2$. Group 3 $(X_1 = 1, X_2 = 0)$, with the conditional mean equal to $\beta_0 + \beta_1$. Group 4 $(X_1 = 1, X_2 = 1)$ with the conditional mean equal to $\beta_0 + \beta_1 + \beta_2$. (See Figure 2).

The intercept (β_0)

Values of β_0 for correlations are -0.3, - 0.2, 0,0.1, and 0.4. Values of β_0 for standardized mean difference are -0.3, -0.1, 0.2, 0.5, and 0.9. For both correlations and standardized mean difference, the slopes for the two predictors β_1 and β_2 are fixed at 0.2 and 0.1, respectively. These values result in effect-size magnitudes for the fourth conditional meant, which correspond approximately to effect sizes of no effect, small, medium, large, and extreme from the recommended guidelines for power analysis provided by (Cohen, 1988). The use of these guidelines for simulation should not be construed as condoning the pervasive use of Cohen's guidelines for interpretation of effect sizes. Cohen constructed those definitions to guide power analysis, and the problem of selecting effect sizes for a simulation is analogous to conducting a power analysis. The inclusion of an extreme case condition is to address that the normality assumption is more likely to be violated when the magnitude of the effect for correlations is high due to the restricted range of values (±1).

Sample size (n)

Sample sizes are not manipulated in this study but are an integral variable in the simulation. Sample sizes in the simulation are randomly generated from an empirical distribution (See Appendix 1) modeled from data on sample size collected from published meta-analyses in psychological journals.

Effect Sizes (r,d)

Study effect sizes use raw data simulated at the individual level. The simulation of individual raw data involves a multistep process. The effect size parameter for each study is generated from a distribution based on the set hyperparameters, $\theta_i \sim N(\mu_{\theta i}, \tau^2)$, where θ_i is the effect size parameter for study i (either ρ or δ) and $\mu_{\theta i}$ is the true conditional mean for study i. In each replication, there are k sample sizes generated from the empirical n distribution. For each study, I then generate raw data for n_k individuals based on the effect size parameter θ_i. The calculation of the study effect sizes (T_i) uses this simulates raw data. For code on simulating raw data for correlation and standardized mean differences, see appendices 2 and 3.

This process of simulating raw data is atypical of simulating effect sizes for meta-analytic studies. Often T_i is directly simulated from a distribution based on set parameters for θ and σ^2. However, simulating data in this manner misses much of the nuances of sampling. For example, effect sizes based on *d* is known to be positively biased, particularly when sample sizes are small. Hedges (1981) developed a commonly used bias correction for *d,* sometimes referred to as Hedges *g*. In a mini simulation where I generated 1,000,000 effect sizes based on the raw data with δ=0 and τ=0.[8] Sample sizes for this mini simulation come from the empirical n distribution used for the main study. Bias for the raw data matched expected values with estimates based on *d* having upward bias ($Bias_d$= 0.0085) and the application of Hedges bias correction results in less overall bias ($Bias_g$ -0.0001). However, for data simulated directly from the distribution, the trend of upward bias for *d* is not found in the estimates ($Bias_d$= 0.0002). Since Hedges bias correction reduces the estimate of *d*, simulating data this way leads to an increased bias in the negative direction when the correction is applied ($Bias_g$ -0.0082).

[8] v_i for data simulated directly from a distribution is based on the same sample size used to generate the raw data. Setting τ=0 makes this model a fixed-effect model.

Results

The results for the current study are divided into three parts for ease of interpretation. Part one assesses the performance of dynamic and conventional approaches for correlation and standardized mean difference in the context of frequentist estimation. In part two, using the same simulated data, I compare dynamic and conventional approaches within the Bayesian framework. Part three combines the results from parts 1 and 2 to compare frequentist and Bayesian modes of estimation.

Frequentist Mixed-Effects

I use R statistical software (R Core Team, 2013) to run all statistical analyses. Analytic results of the conventional approach and direct z-to-r back-transformed correlations utilized the metafor package (Viechtbauer, 2010). For the dynamic method and the integral z-to-r transformation (See Appendix 4), functions for estimation had to be specially coded.

Assessment of the performance of the different methods of estimation included bias, root-mean-squared error (RMSE), and coverage rate of confidence intervals. Bias is the average difference between the estimate and the true value of a parameter. This statistic also provides information regarding the direction of the difference, with negative bias indicating a tendency to underestimate the parameter and positive bias showing a tendency to overestimate.

$$Bias = \frac{\Sigma(T_i - \theta)}{reps} \ , \qquad (16)$$

$$RMSE = \sqrt{\frac{\Sigma[(T_i - \theta)^2]}{reps}} \ . \qquad (6)$$

where T_i represents an estimate of the parameter θ for replication *i* and *reps* is the total number of replications in the cell. I also report relative bias, which is bias divided by the true value of the estimated parameter. RMSE represents a total variation of estimates about the true parameter value. It is analogous to a standard deviation and defined as the square root of the average squared deviation between the estimate obtained from the meta-analysis and the true parameter value, calculated as
This definition contrasts with the empirical standard error of the estimate (i.e., the standard deviation of the parameter estimates) in that it incorporates both the standard error and the bias:

$$RMSE = \sqrt{variance + \ bias^2} \ . \qquad (7)$$

Correlation

For correlation effect size, dynamic, conventional (*Conv)*, direct-z-to-r transformation *(DirZtoR*), and integral z-to-r- transformation (*IntZtoR*) are all included for comparison. Estimation results for the four conditional means (See Figures 3, 4, 5, & 6) indicate that overall *Conv, DirZtoR, and IntZtoR* have a trend of increased positive bias as effect size and heterogeneity increased, while *Dynamic* trends toward negative bias effect size and heterogeneity increase. This trend holds in almost every condition except for when the correlation is highest and heterogeneity large. In these conditions, *Conv*, *DirZtoR,* and *IntZtoR* would shift downward changing direction. This shift is likely due to a border condition, as correlations cannot exceed 1.0.

Conv has a more substantial bias than *IntZtoR* and *Dynamic* across all conditions for correlation. When compared with *DirZtoR, Conv* estimates continue to have greater bias overall. However, when heterogeneity and effect sizes are extreme, as can be seen in the lower right corner of Figures 4, 5, and 6, *DirZtoR* has the most bias for all the conditional means except the first conditional mean (See Figure 3). The value of the first conditional mean is equal to the intercept (β_0). Because the values for β_1 and β_2 are positive, the first conditional mean will always contain the smallest correlation magnitude among the conditional means for each cell.[9] Taken together, the results comparing *Conv* and *DirZtoR* indicate that *DirZtoR* performs better than *Conv* but begins to break down when effect-size is high.

Results of estimation with *DirZtoR* are worse than *IntZtoR*. When heterogeneity is low, *DirZtoR* and *IntZtoR* have similar results for bias. However, as heterogeneity increased, *DirZtoR* does worse than the other two methods of estimation.

Bias for *IntZtoR* and *dynamic* estimates are similar when effect sizes are smaller. When effect sizes are larger, *Dynamic* estimates are less biased than *IntZtoR* for all four conditional means. However, when both heterogeneity and effect size are most extreme, *IntZtoR* estimates are less biased then the *Dynamic* (See the lower right corner of Figures 4, 5, & 6).

Results comparing RMSE for the four conditional means (See Figures 7, 8, 9, & 10), indicated that *Conv* estimates are worse than the other three methods. Results indicate that the performance of *DirZtoR, IntZtoR,* and *Dynamic* are similar when assessed with RMSE. Results for all methods have a trend of decreasing RMSE as the sample size increased.

Coverage for all methods is evaluated at the 95% nominal rate. Coverage rates for *Conv* are worse than all other methods across the board but are particularly low as sample size increased, and heterogeneity is low (See Figures 13, 14, 15, 16). *IntZtoR* has better coverage among the methods compared, followed by *Dynamic*. Similar coverage rates are seen for *DirZtoR* with *IntZtoR* when heterogeneity and effect size is smaller but has worse coverage in conditions where heterogeneity and effect size are largest (See Figures 14, 15, 16).

Both *Conv* and *Dynamic* trend toward increased positive bias and RMSE of τ as the number of studies increased (See Figures 9 & 16). Bias and RMSE of τ decreased as heterogeneity and effect size increased for both *Conv* and *Dynamic*. When the number of studies is small and heterogeneity high, *Conv* has less bias and lower RMSE than *Dynamic.* However, when heterogeneity is higher, *Dynamic* has less bias and RMSE than *Conv.* Fisher's *Z*-estimates of the variance component is not easily back-transformed, and *DirZtoR* and *IntZtoR* and excluded in the results of the analysis of the variance component[10].

Standardized Mean Difference

[9] This is due to the nature of the predictors in this simulation where both β1 and β2 are fixed positive predictors.

[10] *DirZtoR* and *IntZtoR* are on the z-metric while *Conv* and *Dynamic* are in the correlation metric

Conventional bias-corrected F*(ConvG)* and uncorrected estimates *(ConvD)* as well as Dynamic corrected *(DynG)* and un-corrected estimates *(DynD)*. For all methods, bias tended to increase as effect size magnitude increased (See Figures 23, 24, 25 & 26). The results of the simulation indicate that the bias-corrected dynamic method *(DynG)* is the least biased, followed by *ConvD.* Results indicated that *ConvG* has the most bias of the four methods compared.

When the magnitude is low, the conventional methods (*ConvD* and *ConvG*) both have increased negative bias as the magnitude of the effect size increased. The dynamic methods *(DynD* and *DynG)* trend in the opposite direction, with bias increasing in the positive direction as effect-size magnitude and heterogeneity increases. All methods perform similarly based on RMSE (See Figures 30, 31, 32, & 33).

The results of bias and RMSE for the standardized mean difference effect-size indicate that while all methods of estimation have similar precision, *DynG* obtains the same amount of precision with the least amount of bias.

Coverage rates for the four conditional means, assessed at the 95% nominal rate, are similar for all methods (See Figures 39, 40, 41, & 42). When the sample size is small and heterogeneity high, coverage rates are below the nominal rate but improved as the sample size increased.

Results for comparing the meta-analytic results for the different methods on τ indicated that bias (see Figure 25) and RMSE (See Figure 36) are similar for the four methods. Bias increased as the heterogeneity for the studies increased. Conditions with an increased number of studies also have decreased bias. The magnitude of the effect-size does not appear to influence bias for estimates of τ.

Bayesian Mixed-Effects

In part one of this study, there are 15,000 simulated replications in each cell for the frequentist analysis. However, due to the nature of MCMC sampling, Bayesian analysis can be time-consuming. Due to time constraints, the Bayesian analysis includes only a subset of 1000 of the replications from part one in each cell. However, the parameters for the different cells are identical to the parameters from study one, so we can draw comparisons between the frequentist and Bayesian methods.

Analysis of the Bayesian methods is done with JAGS in R using the 'R2Jags' package (Su & Yajima, 2015) and the CODA package (Plummer, Best, Cowles, & Vines, 2006). For each prior, the analysis is conducted with and without the inclusion of the dynamic method. The Bayesian analysis for each replication includes three chains of length 60,000, after a burn-in of 30,000.

The three chains each have different starting values. Chain one uses starting values below the parameter, chain two starts at the true value of the parameter, and chain three has starting values above the parameter value.

Convergence diagnostics include the Brooks, Gelman, Rubin diagnostic (Gelman & Rubin, 1992; 1997) for assessing convergence among parallel chains. The Brooks, Gelman, Rubin diagnostic uses a potential scale reduction factor (PSRF) to determine convergence within the chains. The chain is converged if the PSRF is near 1.0 (+/- .05).

Correlation

Convergence for the Brooks, Gelman, Rubin diagnostic across all cells ranged between 86-100%. Convergence for the literature-based gamma prior is 99-100% for the conventional model and 97-100% for the dynamic model. For the vague uniform prior, convergence ranged from 95-100% for the conventional model and 86%-100% for the dynamic model.

Both dynamic methods have similar bias and RMSE regardless of the prior precision (See Figures 31, 32, 33, & 34). Conventional models also have similar bias RMSE regardless of the prior set on precision (See Figures 35, 36, 37, & 38). This finding makes sense, given that the priors for the conditional means are not manipulated, therefore identical in all the models, and indicates that the prior put on precision did not have a large influence on other estimates. Bayesian analysis incorporating the dynamic method has better estimates than those using the dynamic model in almost all conditions. However, when heterogeneity and effect size is large, the dynamic methods have more bias than conventional methods (See Figure 34).

For estimation of τ overall, estimates from dynamic methods are better in comparison to conventional counterparts, with conventional methods having more bias in the positive direction. The estimates incorporating the dynamic method with the literature-based gamma prior are better than the conventional approach using the same prior, and the dynamic estimates with the uniform prior are better than the conventional method using the same prior. However, when the number of studies is small and heterogeneity high, estimated from conventional method utilizing the gamma prior is better than the dynamic, with the dynamic method showing negative bias (See Figure 39). When comparing results on RMSE, the dynamic method performs at least as well or better than the conventional method (See Figure 40). The addition of the literature-based prior improves the estimation of τ when sample sizes are small but has less influence on the estimate as sample sizes increased.

Standardized Mean Difference

Convergence for the Brooks, Gelman, Rubin diagnostic across all cells ranged between 80-100%. Convergence for the literature-based gamma prior is 100% for the conventional model and 99-100% for the dynamic model. For the vague uniform, prior convergence ranged from 85-100% for the conventional model and 80%-100% for the dynamic model.

In contrast to the results based on correlation effect-sizes, for standardized mean difference, dynamic estimates for all four conditional means are more biased than conventional estimates across all conditions (See Figures 41,42,43, & 44). RMSE for the conditional means has a similar pattern, with RMSE being higher for dynamic estimates (See Figures 45, 46,47, & 48).

Results based on the estimation of τ show that the dynamic and conventional methods perform similarly regarding bias (See Figure 49) and RMSE (see Figure 50). The literature-based gamma prior has improved estimation when the number of studies included is small and heterogeneity low.

Frequentist and Bayesian Mixed Effects

Correlation

Comparisons of results for frequentists and Bayesian estimation on correlation show that the frequentist approach resulted in less bias than Bayesian methods for estimation of the conditional means. The finding of less bias in the frequentist estimates is true for both the dynamic and conventional methods (see Figures 51 & 53). For the

Dynamic methods, RMSE is similar for *Dynamic, Gamma Dyn,* and *Unif Dyn* (see Figure 52). For the conventional methods, *IntZtoR* followed by *DirZtoR*, although *DirZtoR* does worse than the other methods when heterogeneity and effect size is larger (See Figures 52 & 54).

Results on τ also indicate that the dynamic frequentist method does better than the dynamic methods in the Bayesian framework (*Gamma Dyn* and *Unif Dyn)*. However, when the number of studies is small, dynamic methods became increasingly downward biased (See Figures 55 & 56).

Bias and RMSE of estimates of τ for *DirZtoR* and *IntZtoR* are excluded from the comparison because the estimates of the variance component are on a different metric. The results comparing *Conv, Gamma Conv,* and *Unif Conv* indicate that *Conv* estimates are best out of the three methods compared (See Figures 57 & 58).

Standardized Mean Difference

For standardized mean difference, frequentist methods continued to do better than Bayesian methods of effect-size estimation. Among the dynamic methods for standardized mean difference, the method utilizing the bias-corrected effect size estimate (*DynG)* has decreased bias compared to the other dynamic methods *DynD, Gamma Dyn,* and *Unif Dyn* (See Figure 59). For the conventional methods, *ConvD* has less bias than *ConvG, Gamma Conv,* and *Gamma Unif* (See Figure 61). RMSE results for effect-size estimates are similar for both dynamic and conventional methods (see Figures 60 & 62).

Dynamic estimates of τ based on frequentist methods are less biased than Bayesian methods when the number of studies is larger and the variance component smaller. When the number of studies is small (K=10, K=20) and the model contains heterogeneity, Bayesian estimates are less biased then dynamic. When the number of studies is large and heterogeneity high, all methods have a similar bias and RMSE for their estimates (See Figures 63 & 64). Conventional methods have a similar pattern to dynamic for estimates of τ (See Figures 65 & 66). Bayesian estimates have less bias and RMSE for their estimates of τ when the number of studies is low, and heterogeneity is high. When heterogeneity is low, and the number of studies high frequentist methods have less bias and RMSE then Bayesian. When heterogeneity is higher, and studies are large all method have a comparable amount of bias

Discussion

The results of this study provide support for the use of the dynamic model improving estimation in comparison to other methods, particularly for correlation, where the dynamic method produced less biased estimates than other methods for correlation across all conditions. In the Bayesian framework, for correlation, the dynamic method tended to do better than estimates with identical priors that did not incorporate the dynamic model in most conditions.

The usefulness of the dynamic method for standardized mean difference effect sizes is less clear than with correlation. The two methods with the least biased estimates are the bias-corrected dynamic method *DynG* and the uncorrected *ConvD* estimate. The finding that *ConvD* is less biased than *ConvG* is particularly worth noting. This estimate introduced by Hedges (1981) is often used because it reduces bias in individual effect-size estimates. However, a reduced bias of individual effect-size estimation does not necessarily translate to less bias in the meta-analytic estimate. Bias-corrected estimates will be smaller than initial estimates of *d*. As mentioned in the introduction for standardized mean differences, less substantial study effects lead to lower weights given to the study in the meta-analysis (See Equation 7). In instances where bias is low, bias-corrected estimates may lead to downward bias and inflation of the weight of the individual study in the meta-analysis. Since the dynamic method retains the dependence of the estimate of the conditional variance on μ_θ *DynG* has the benefit of incorporating the unbiased estimates while reducing the influence of biased estimates on the meta-analysis. In the Bayesian framework, the dynamic method did not appear to offer any additional benefit to estimation with standardized mean difference, and at times led to more biased estimates.

This study contained several limitations. For frequentist methods, only fixed categorical predictors are included in the model. In addition, although the predictors are slightly correlated, the interaction between the predictors is not taken into account in the model. Future research should look at the influence of varying predictors as well as how taking into account the relationship between predictors improves estimation.

Limitations in the Bayesian framework involved the use of priors. Priors for the effect-size estimates are fixed, but incorporating prior knowledge in these estimates may offer additional benefits. Also, the literature-based prior did not differentiate between areas of research. It is reasonable to assume that the distribution of precision will not be the same for all areas of research. Despite this, the general literature-based prior does appear to be useful as within this study, since improved estimation is still found for this prior when sample sizes are small is found in multiple conditions for heterogeneity. Future research should also look more closely at developing priors looking at specific areas of study. It is likely that while they would be less useful as a general prior, they would probably be of greater benefit to the particular area of study.

The pattern that the dynamic method consistently works better when heterogeneity is low makes sense. The dynamic method uses the current estimate of μ_θ for determining v_i and adjusting the weights. Using the μ_θ to represent θ_i instead of the study effect-size estimate lowers the influence of arbitrary effect-size estimates overly weighted in the model, leading to biased results. However, when heterogeneity is high μ_θ will be less representative of the true parameter θ_i for all studies. Based on the study results, I would recommend using the dynamic method when heterogeneity in the data is not high.

References

Cohen, J. (1988). Statistical power analysis for the behavior science. *Lawrance Eribaum Association*.

Cooper, H., Hedges, L. V, & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.

Field, A. P. (2001). Meta-Analysis of Correlation Coefficients : A Monte Carlo Comparison of Fixed- and Random-Effects Methods, *6*(2), 161–180. https://doi.org/10.1037//1082-989X.6.2.161

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models, (3), 515–533.

Hafdahl, A. R. (2009). Improved Fisher z estimators for univariate random-effects meta-analysis of correlations. *British Journal of Mathematical and Statistical Psychology*, *62*, 233–261. https://doi.org/10.1348/000711008X281633

Hafdahl, A. R., & Williams, M. A. (2009). Meta-Analysis of Correlations Revisited : Attempted Replication and Extension of Field ' s (2001) Simulation Studies, *14*(1), 24–42. https://doi.org/10.1037/a0014697

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics*, *6*(2), 107–128.

Hedges, L. V, & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486–504.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis, *1558*, 1539–1558. https://doi.org/10.1002/sim.1186

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of Meta-Analysis : Correcting Error and Bias in Research Findings.* (2nd ed.). SAGE Publications Inc. (US).

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-Analysis*. SAGE Publications.

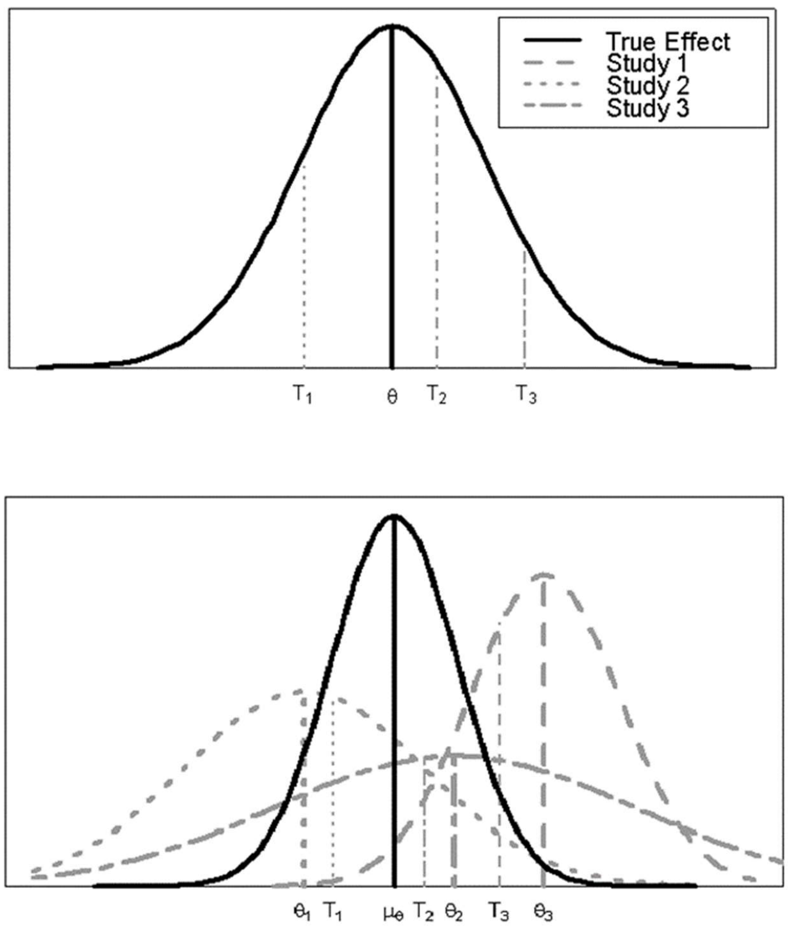
R Core Team. (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from http://www.r-project.org/

Shadish, W. R., & Haddock, C. K. (2009). Combining estimates of effect size. In H. Cooper, L. V Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 257–277). Russell Sage Foundation.

Figures

Figure 1

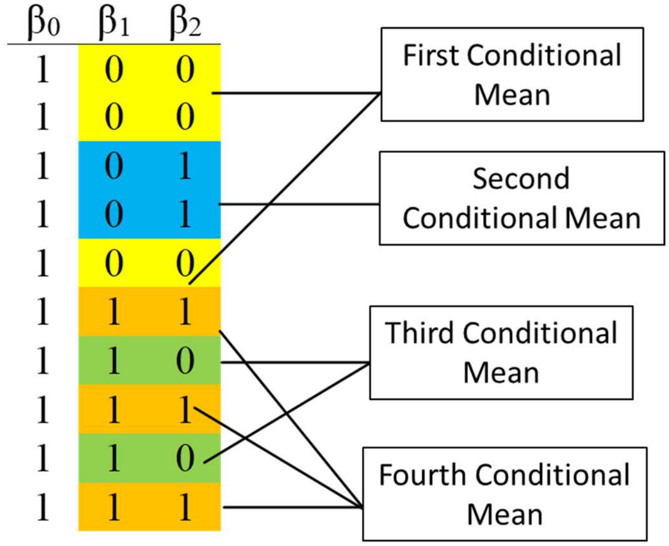
Illustration of fixed-effect and random-effects meta-analyses



Note. Illustration of fixed-effect and random-effects meta-analyses. The diagram on top represents a fixed-effect meta-analysis. The bold line represents the distribution of the true effect size, and all studies come from this "fixed" distribution. The bottom diagram represents a random-effects meta-analysis, where the bold line represents the distribution of the true mean effect. Each study has an associated population effect size (θ), which varies around the true mean effect. Study effect sizes estimates are taken from a distribution around θi

Figure 2

The Design Matrix



Note. The design matrix with the four conditional means indicated.

Figure 3

Bias for the First Conditional Mean for Correlation Effect Size Estimates: Frequentist Analysis

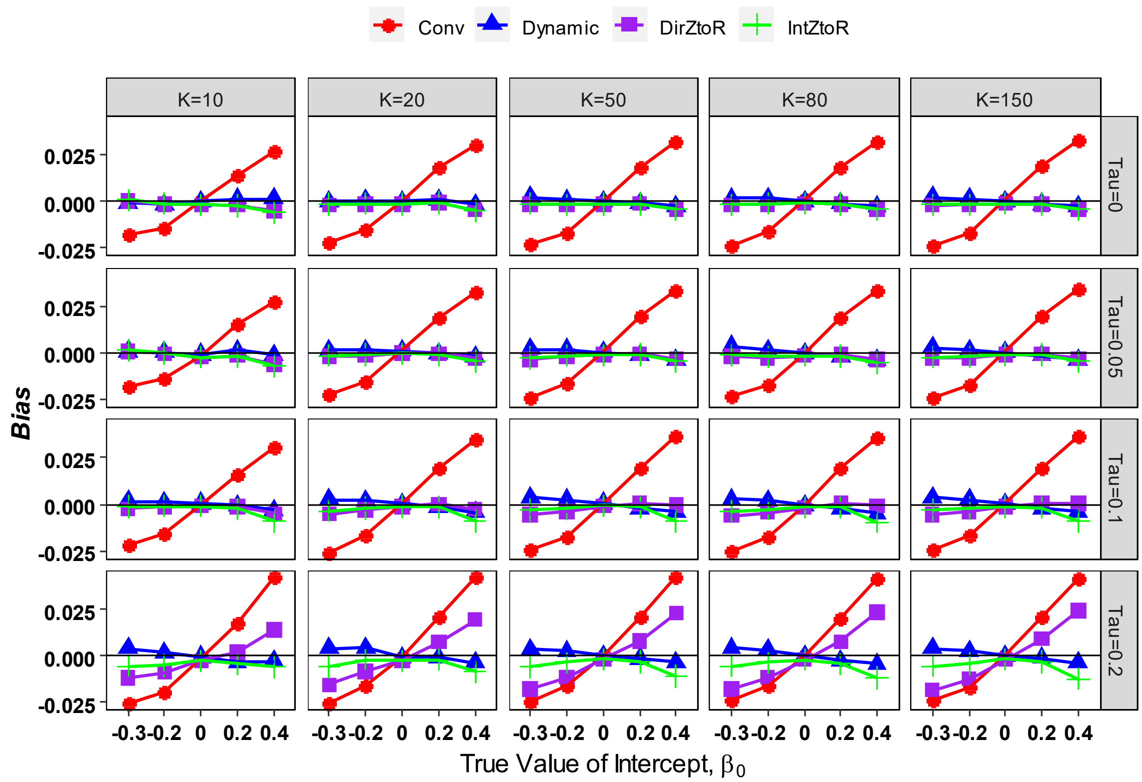
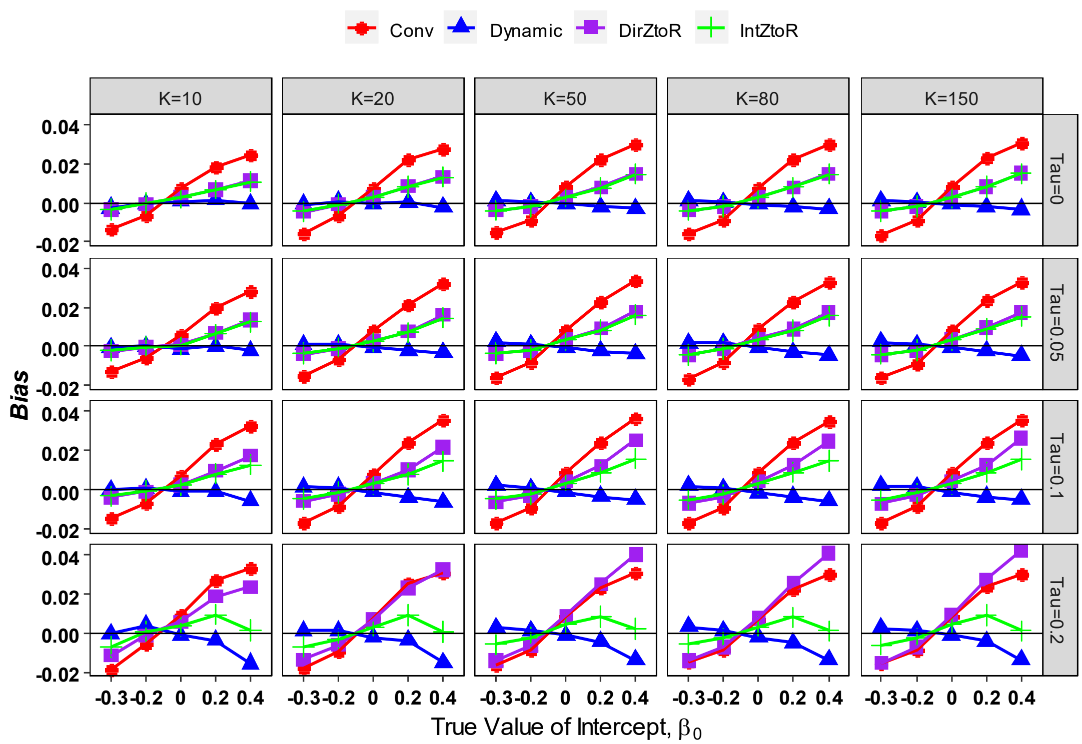


Figure 4

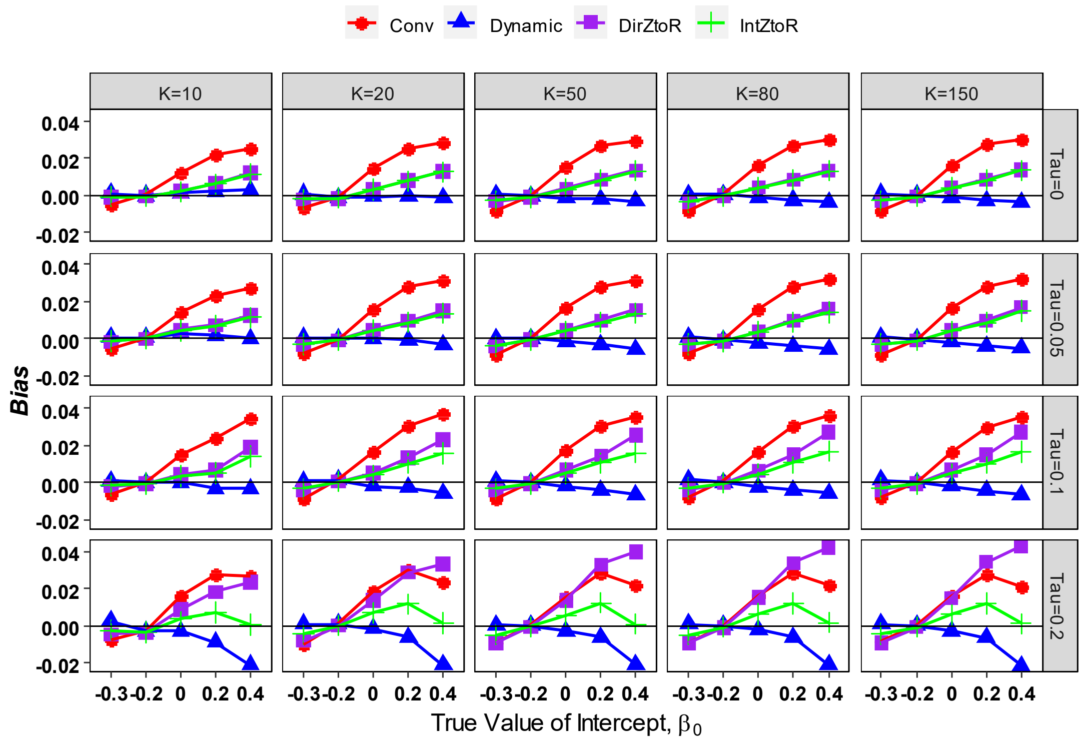
Bias for the Second Conditional Mean for Correlation Effect Size Estimates: Frequentist Analysis



Note. The second conditional mean is the sum of intercept and the regression coefficient of the second predictor $(\beta_2 = .1)$.

Figure 5

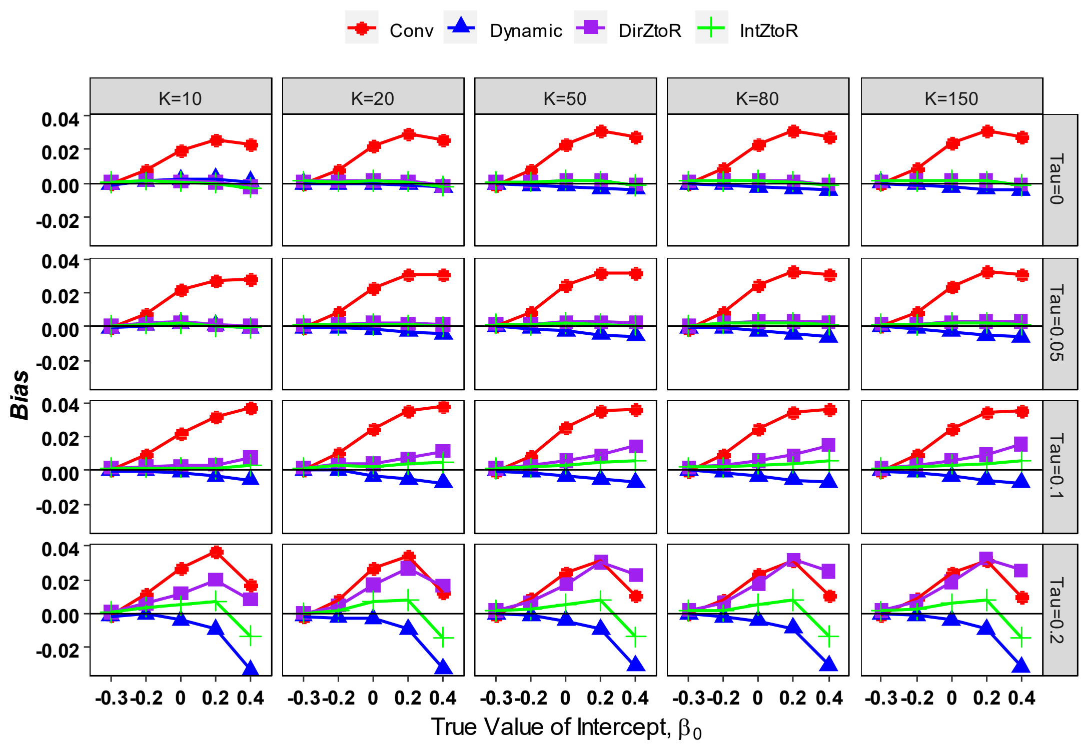
Bias for the Third Conditional Mean for Correlation Effect Size Estimates: Frequentist Analysis



Note. The Third conditional mean is the sum of intercept and the regression coefficient of the first predictor $(\beta_1 = .2)$.

Figure 6

Bias for the Fourth Conditional Mean for Correlation Effect Size Estimates: Frequentist Analysis



Note. The Fourth conditional mean is the sum of intercept and the regression coefficients of both predictors $(\beta_1 + \beta_2 = .3)$.

Figure 7

RMSE for the First Conditional Mean for Correlation Effect Size Estimates: Frequentist Analysis

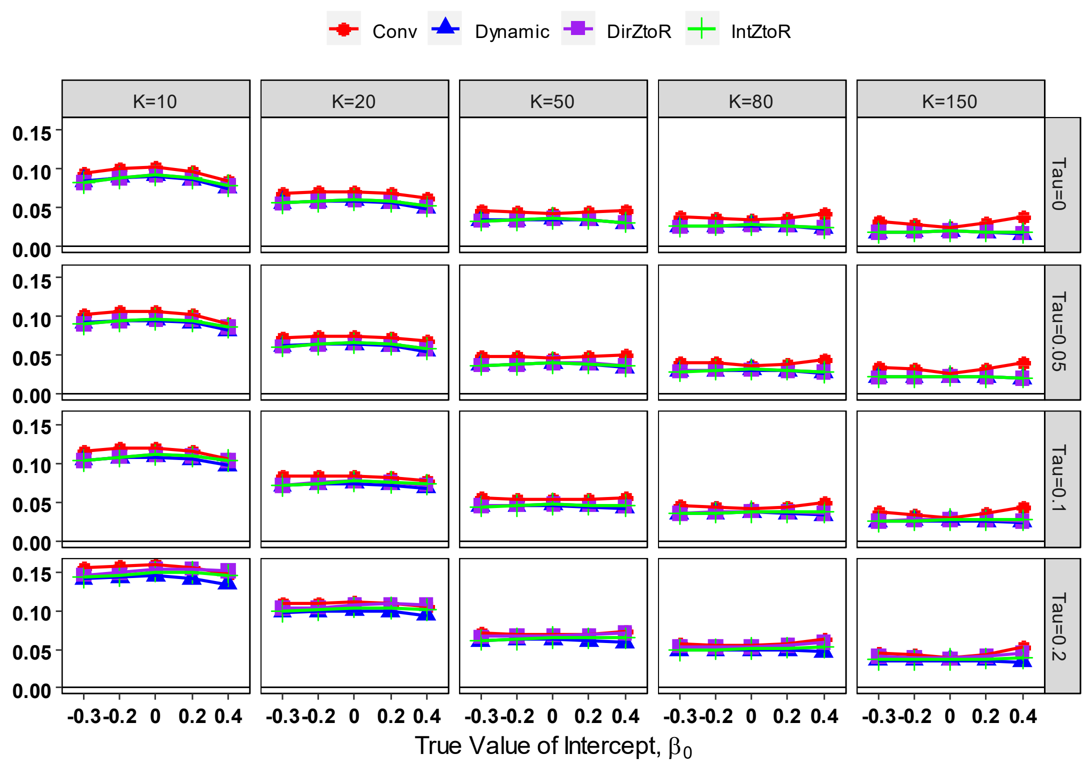
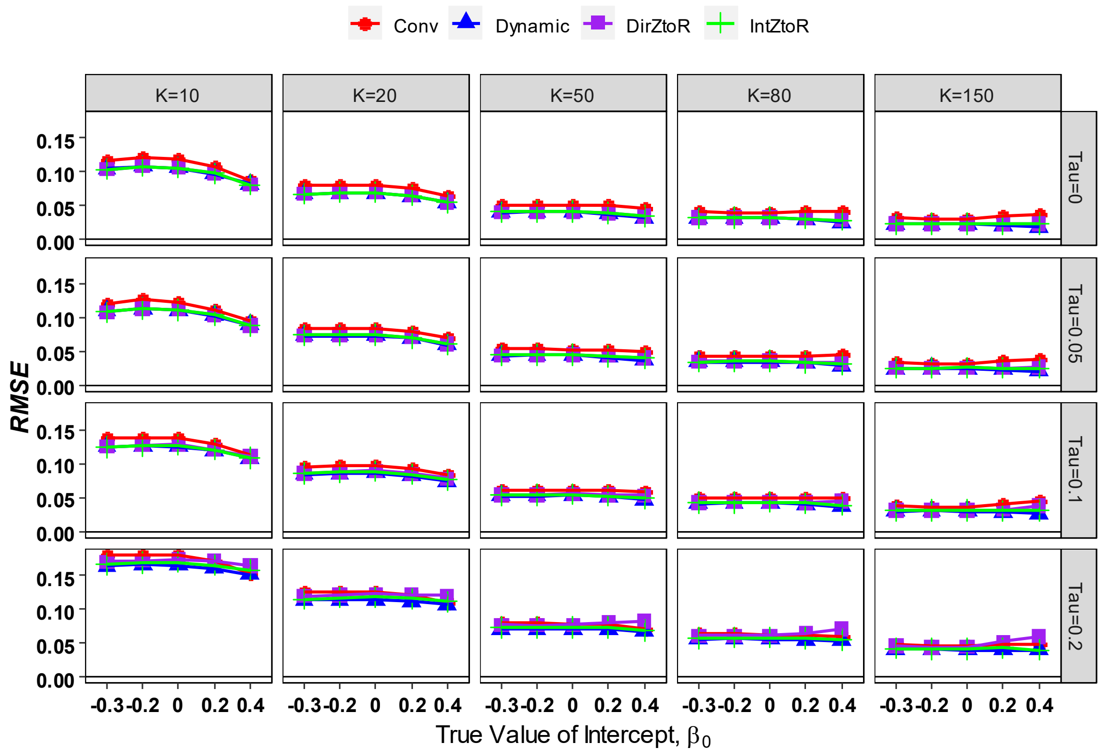


Figure 8

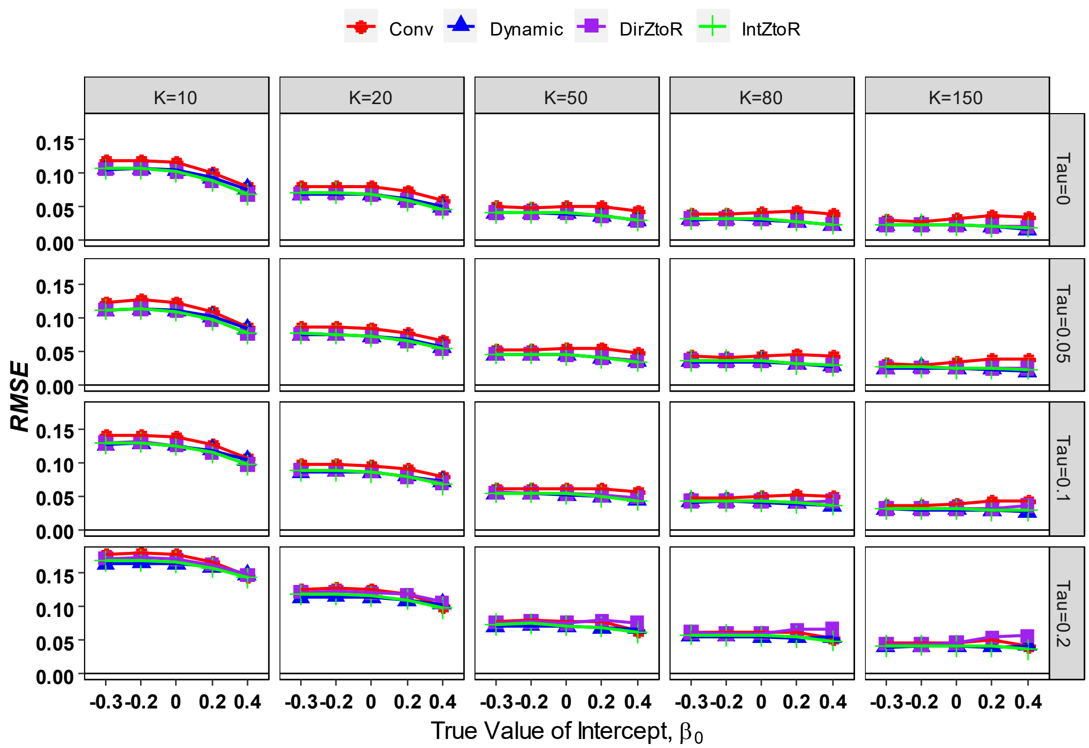
RMSE for the Second Conditional Mean for Correlation Effect Size Estimates: Frequentist Analysis



Note. The second conditional mean is the sum of intercept and the regression coefficient of the second predictor ($\beta_2 = .1$)

Figure 9

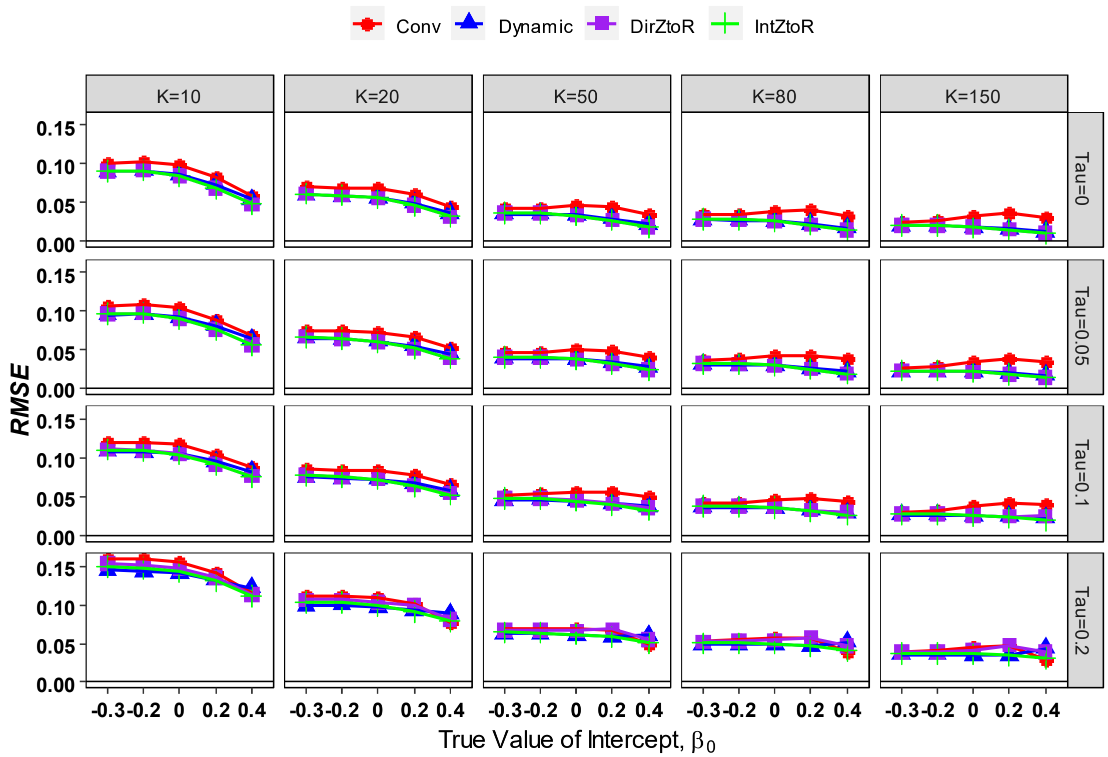
RMSE for the Third Conditional Mean for Correlation Effect Size Estimates: Frequentist Analysis



Note. Third conditional mean is the sum of intercept and the regression coefficient of the first predictor $(\beta_1 = .2)$.

Figure 10

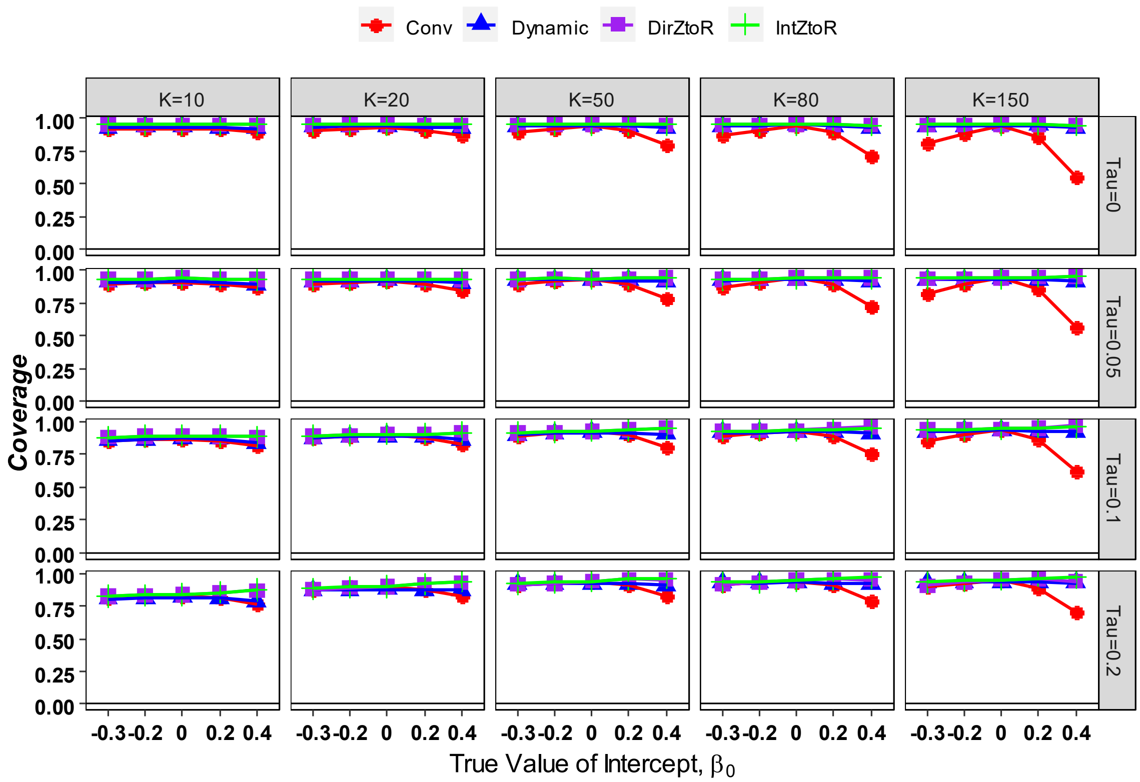
RMSE for the Fourth Conditional Mean for Correlation Effect Size Estimates: Frequentist Analysis



Note. Fourth conditional mean is the sum of intercept and the regression coefficients of both predictors $(\beta_1 + \beta_2 = .3)$.

Figure 11

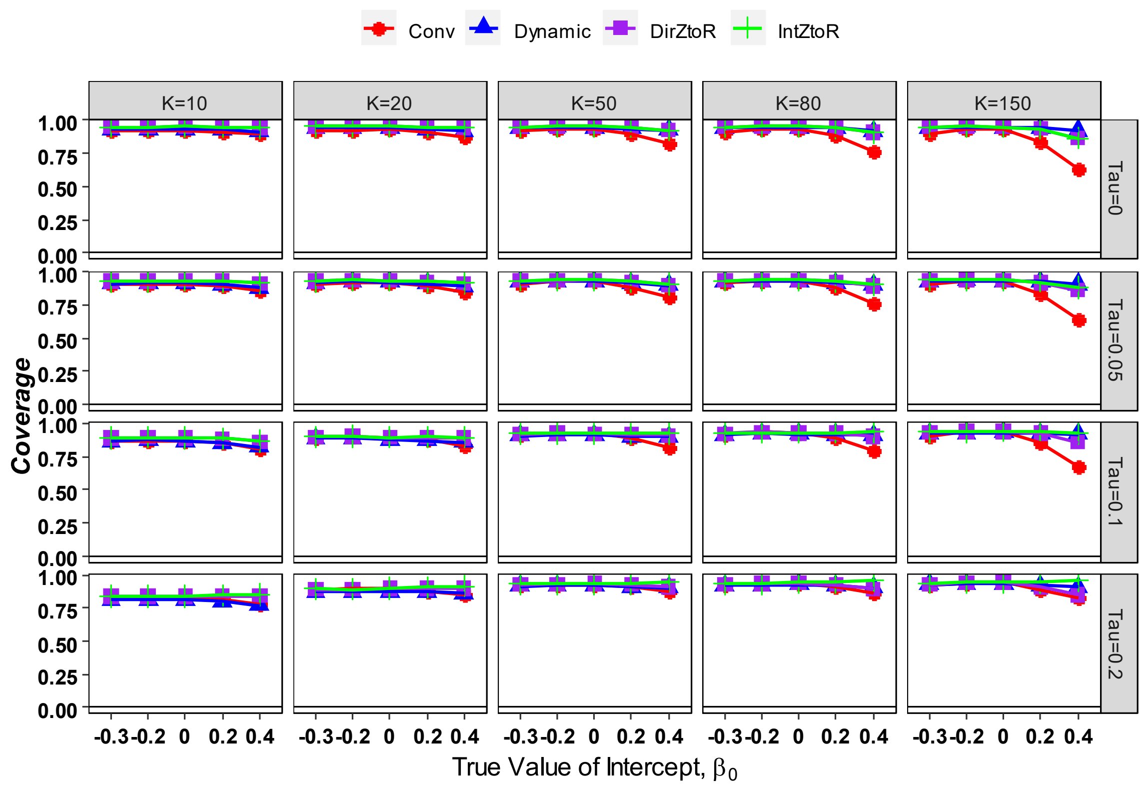
Coverage for the First Conditional Mean for Correlation Effect Size Estimates: Frequentist Analysis



Note. Coverage rate with 95% Nominal probability.

Figure 12

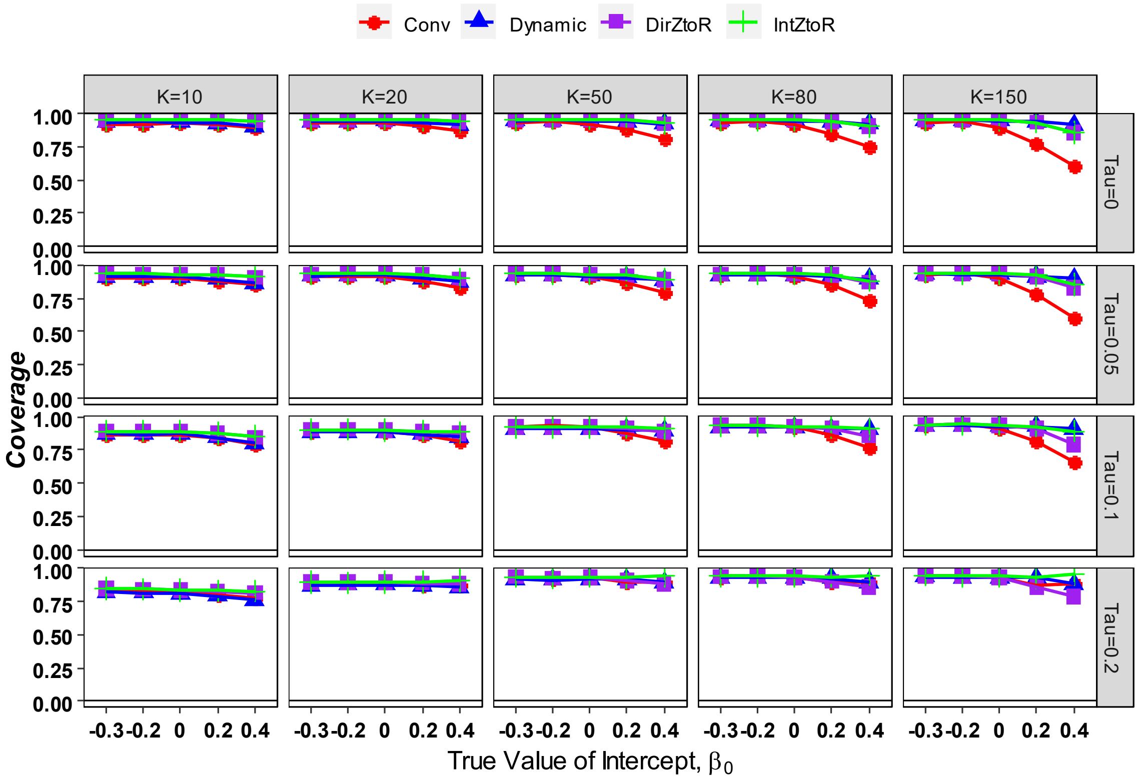
Coverage for the Second Conditional Mean for Correlation Effect Size Estimates: Frequentist Analysis



Note. Coverage rate with 95% Nominal probability. The second conditional mean is the sum of intercept and the regression coefficient of the second predictor $(\beta_2 = .1)$

Figure 13

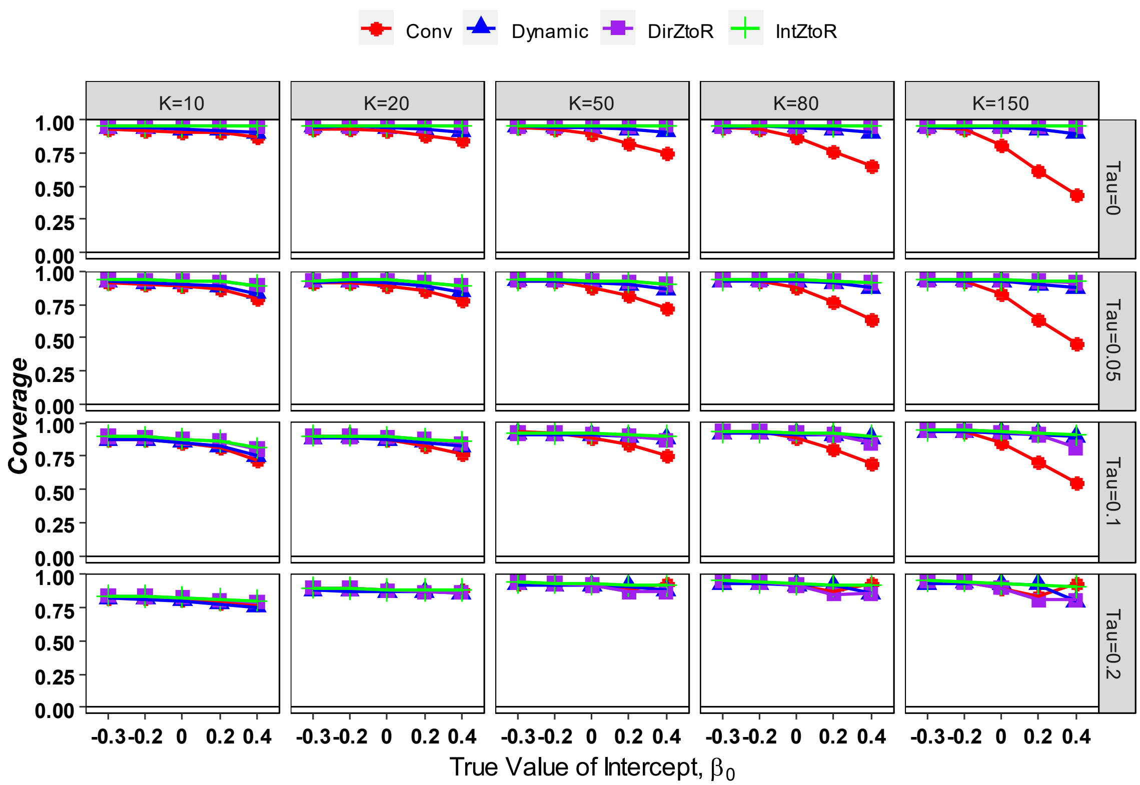
Coverage for the Third Conditional Mean for Correlation Effect Size Estimates: Frequentist Analysis



Note. Coverage rate with 95% Nominal probability. The Third conditional mean is the sum of intercept and the regression coefficient of the first predictor $(\beta_1 = .2)$.

Figure 14

Coverage for the Fourth Conditional Mean for Correlation Effect Size Estimates: Frequentist Analysis



Note. Coverage rate with 95% Nominal probability. The Fourth conditional mean is the sum of intercept and the regression coefficients of both predictors $(\beta_1 + \beta_2 = .3)$.

Figure 15

Bias for Tau for Correlation Effect Size Estimates: Frequentist Analysis

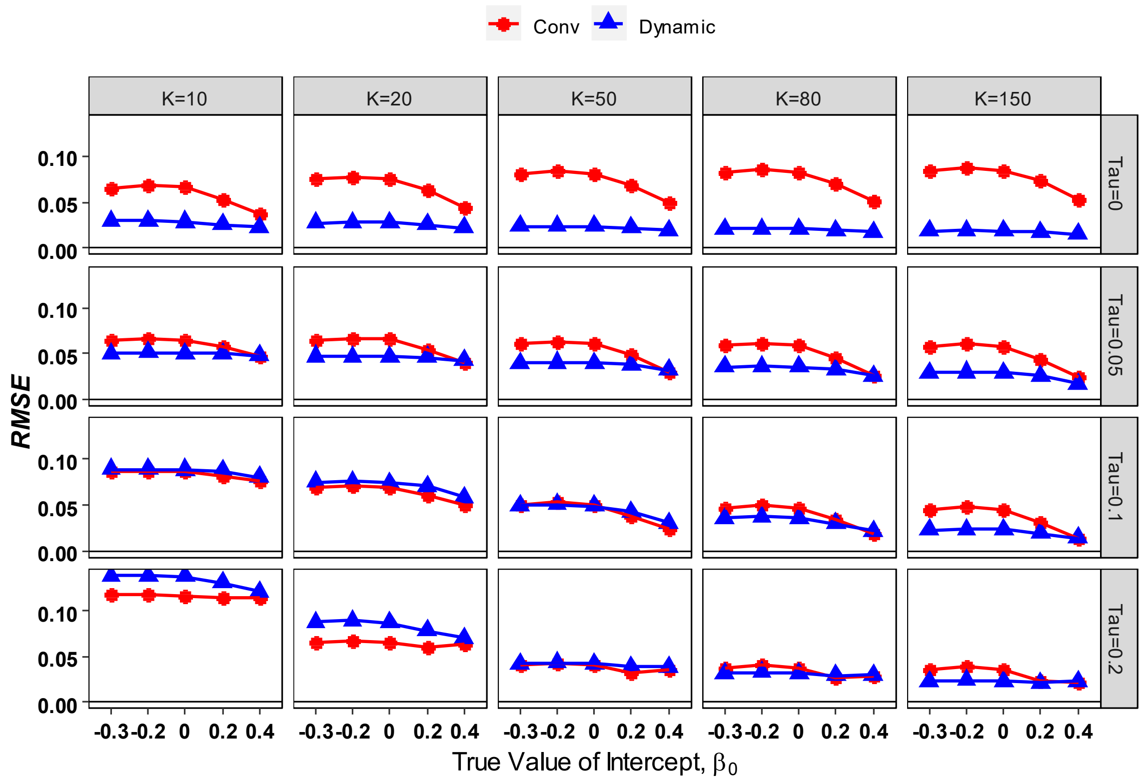


Figure 16

RMSE for Tau for Correlation Effect Size Estimates: Frequentist Analysis

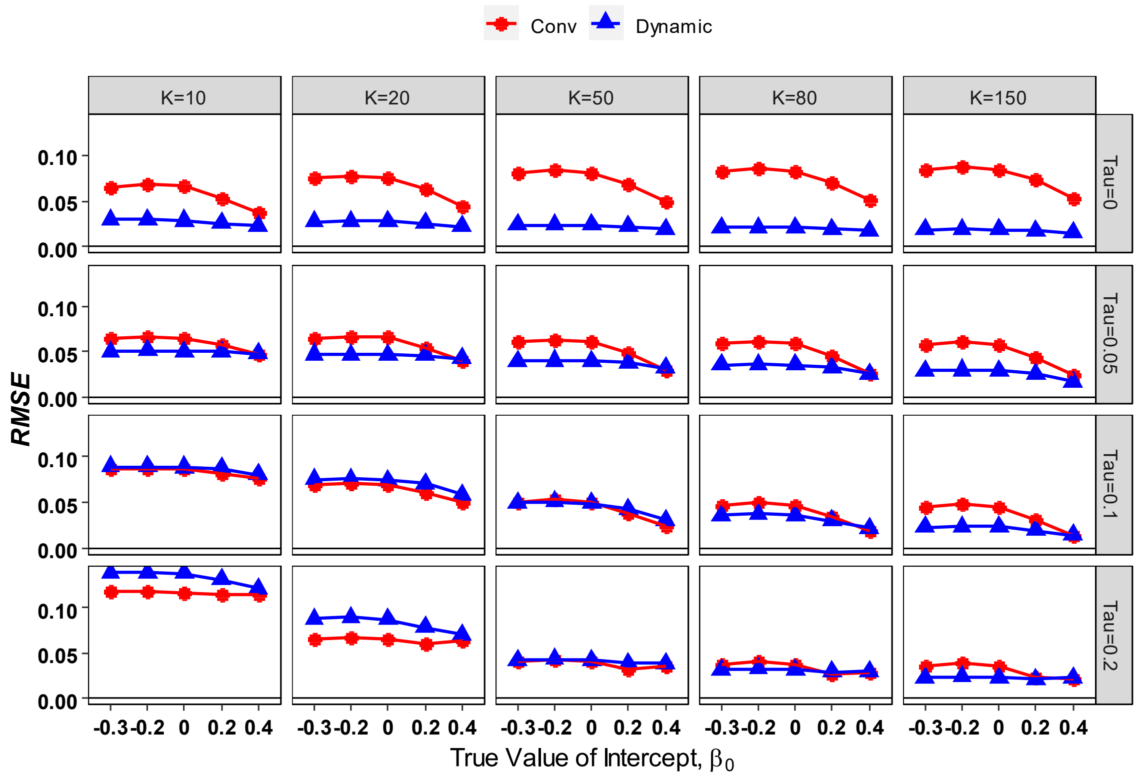


Figure 17

Bias for the First Conditional Mean for Standardized Mean Difference Effect Size Estimates: Frequentist Analysis

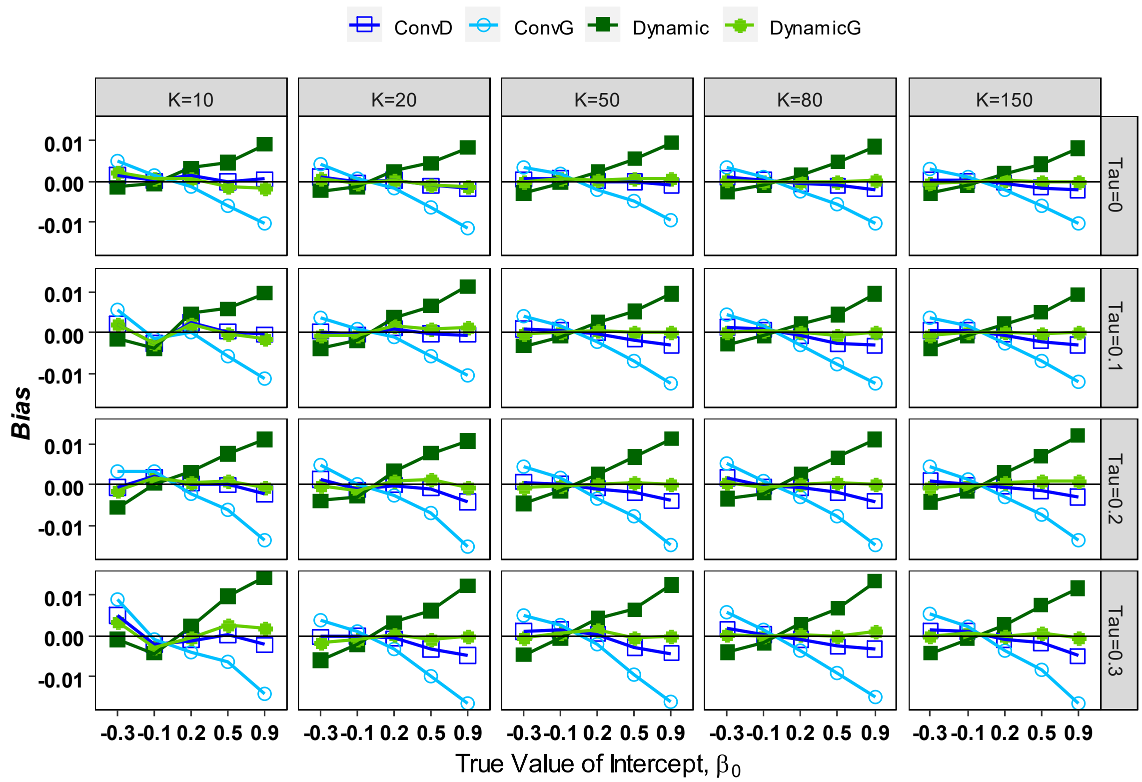
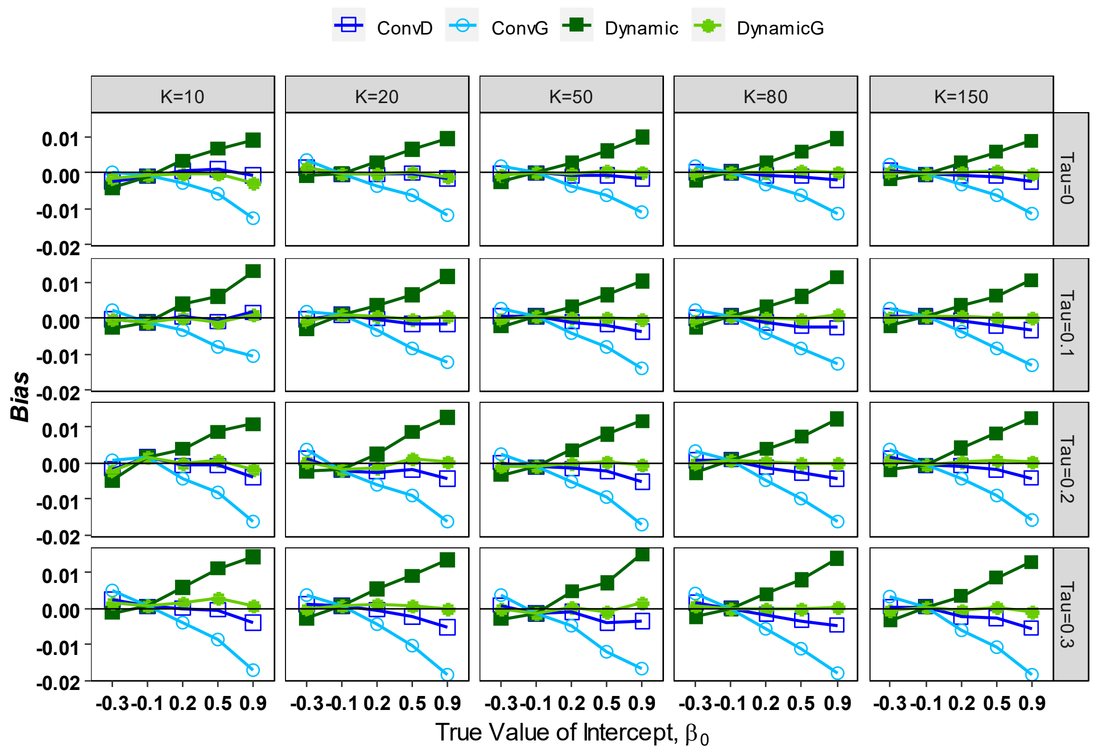
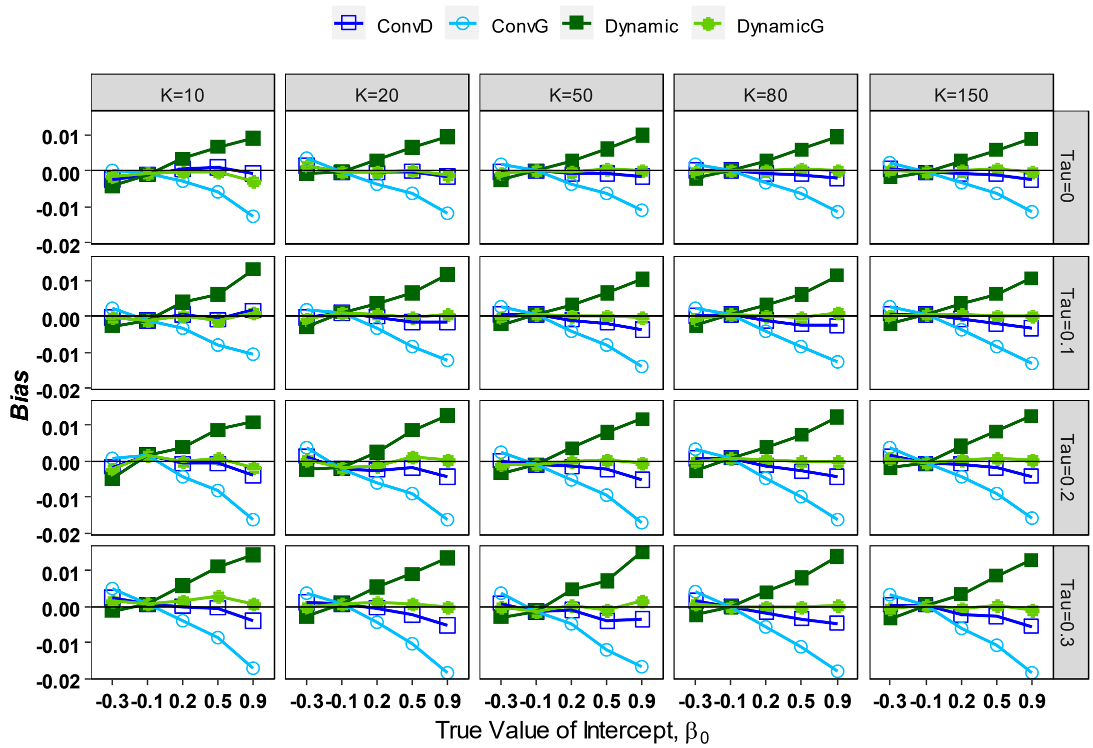


Figure 18

Bias for the Second Conditional Mean for Standardized Mean Difference Effect Size Estimates: Frequentist Analysis

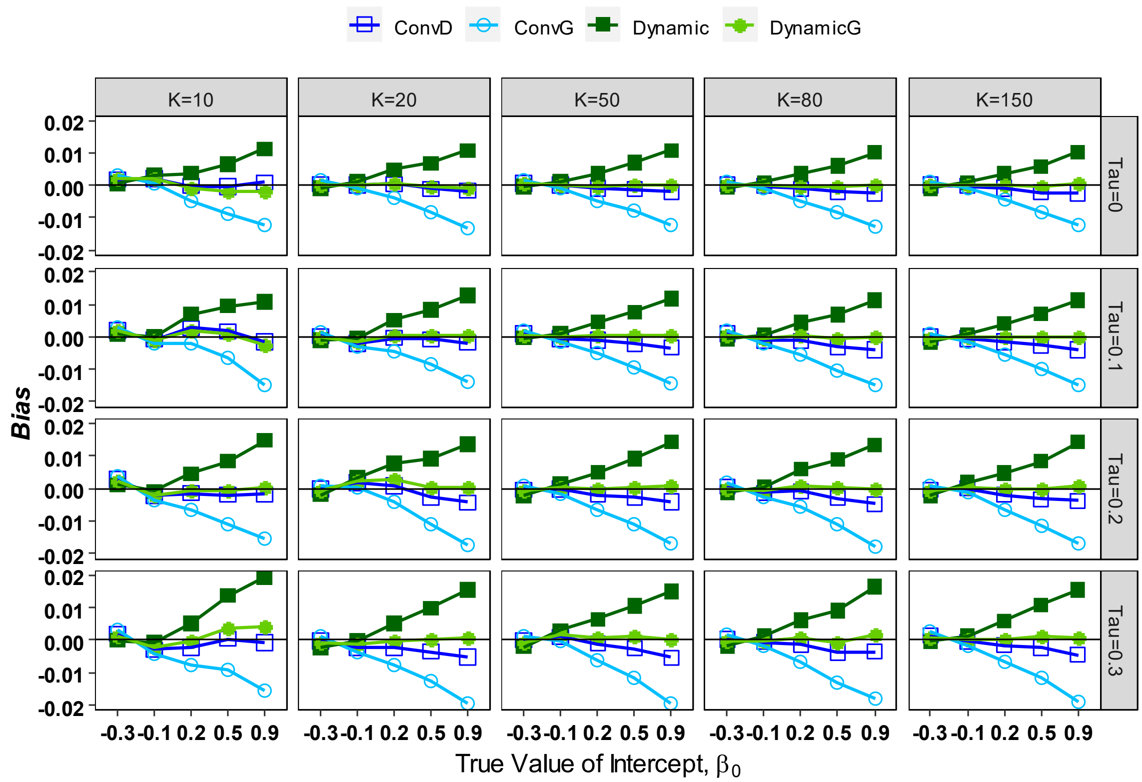




Note. The second conditional mean is the sum of intercept and the regression coefficient of the second predictor $(\beta_2 = .1)$

Figure 19

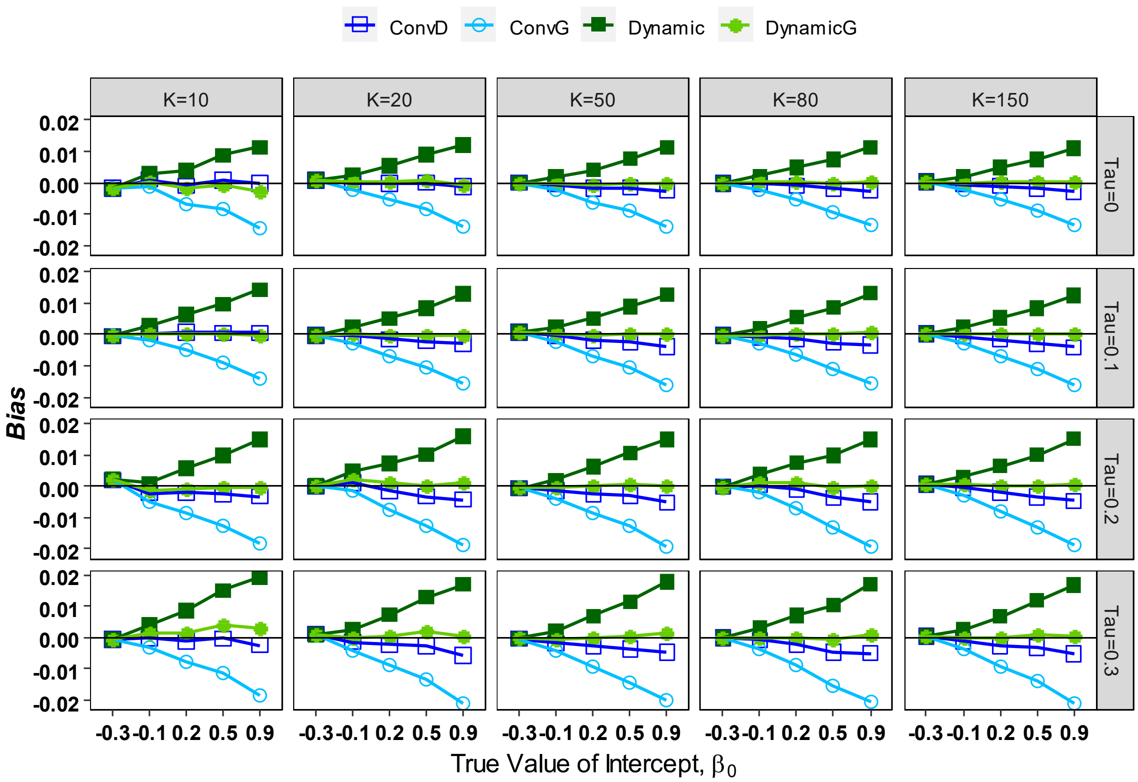
Bias for the Third Conditional Mean for Standardized Mean Difference Effect Size Estimates: Frequentist Analysis



Note. The Third conditional mean is the sum of intercept and the regression coefficient of the first predictor $(\beta_1 = .2)$.

Figure 20

Bias for the Fourth Conditional Mean for Standardized Mean Difference Effect Size Estimates: Frequentist Analysis



Note. The Fourth conditional mean is the sum of intercept and the regression coefficients of both predictors $(\beta_1 + \beta_2 = .3)$.

Figure 21

RMSE for the First Conditional Mean for Standardized Mean Difference Effect Size Estimates: Frequentist Analysis

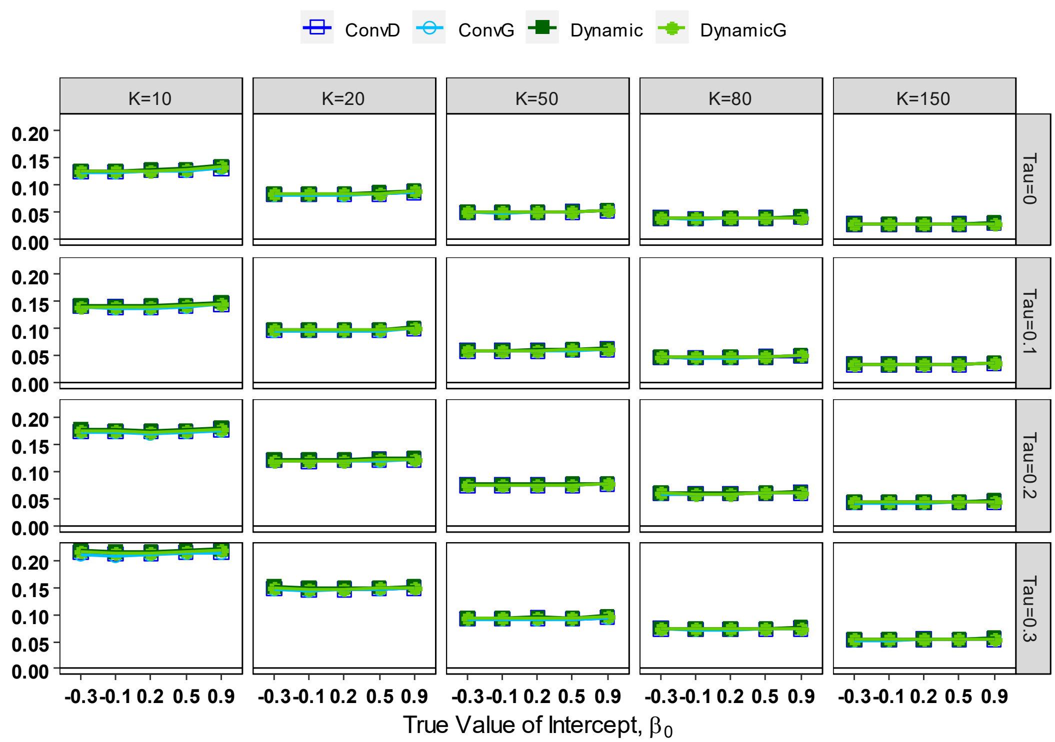
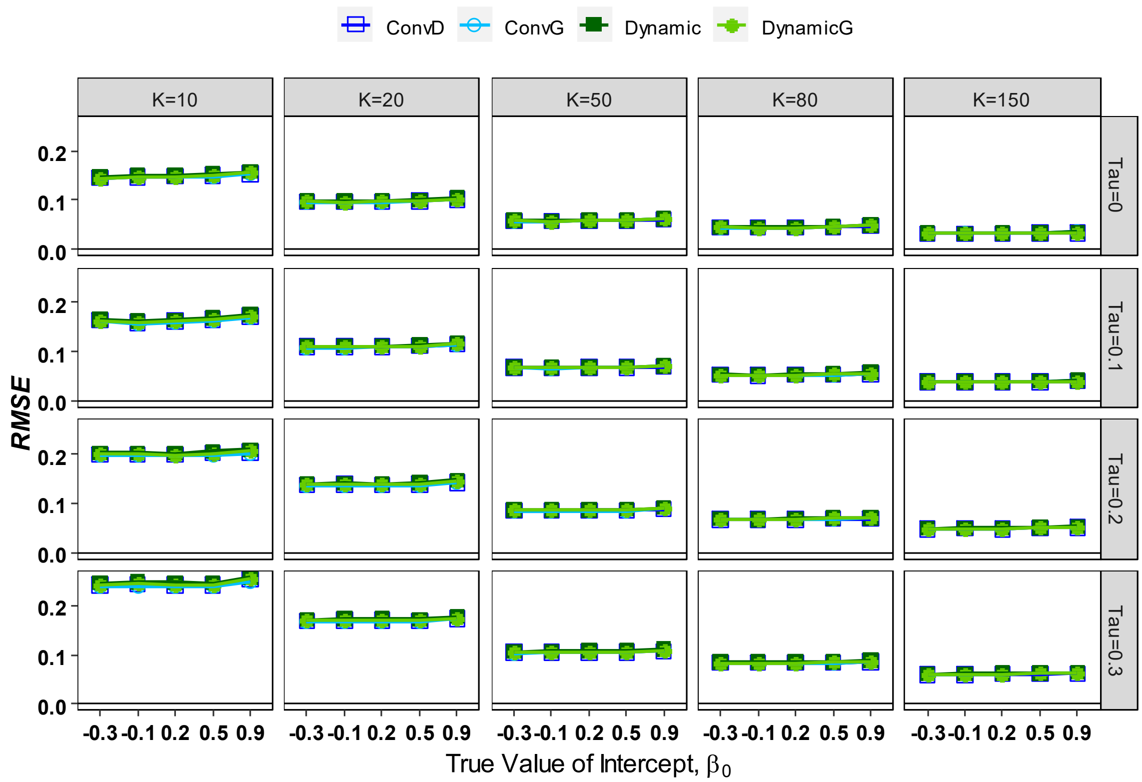


Figure 22

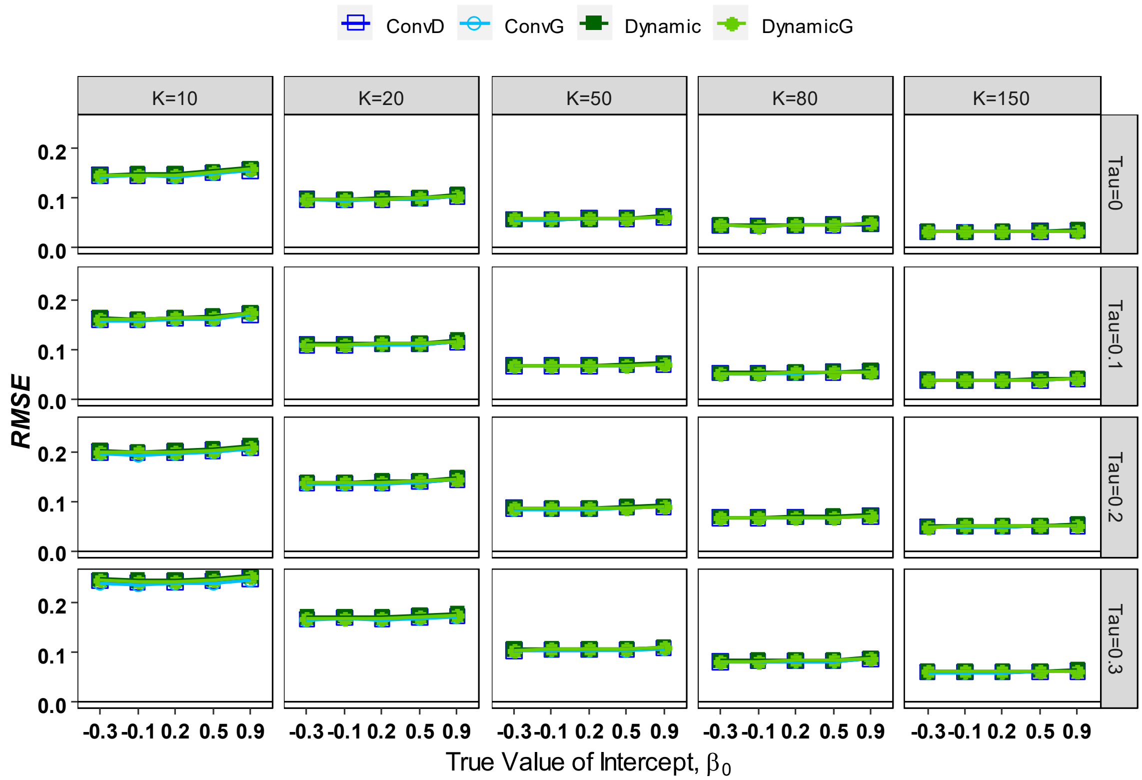
RMSE for the Second Conditional Mean for Standardized Mean Difference Effect Size Estimates: Frequentist Analysis



Note. The second conditional mean is the sum of intercept and the regression coefficient of the second predictor $(\beta_2 = .1)$

Figure 23

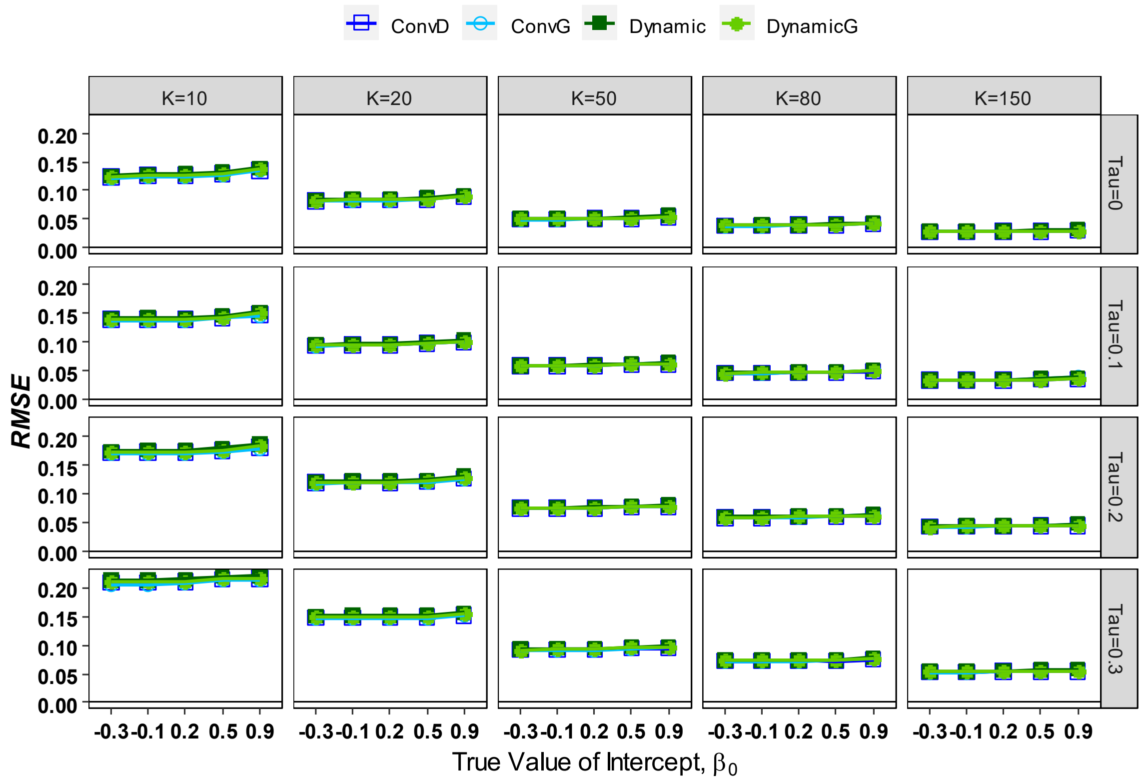
RMSE for the Third Conditional Mean for Standardized Mean Difference Effect Size Estimates: Frequentist Analysis



Note. The Third conditional mean is the sum of intercept and the regression coefficient of the first predictor $(\beta_1 = .2)$.

Figure 24

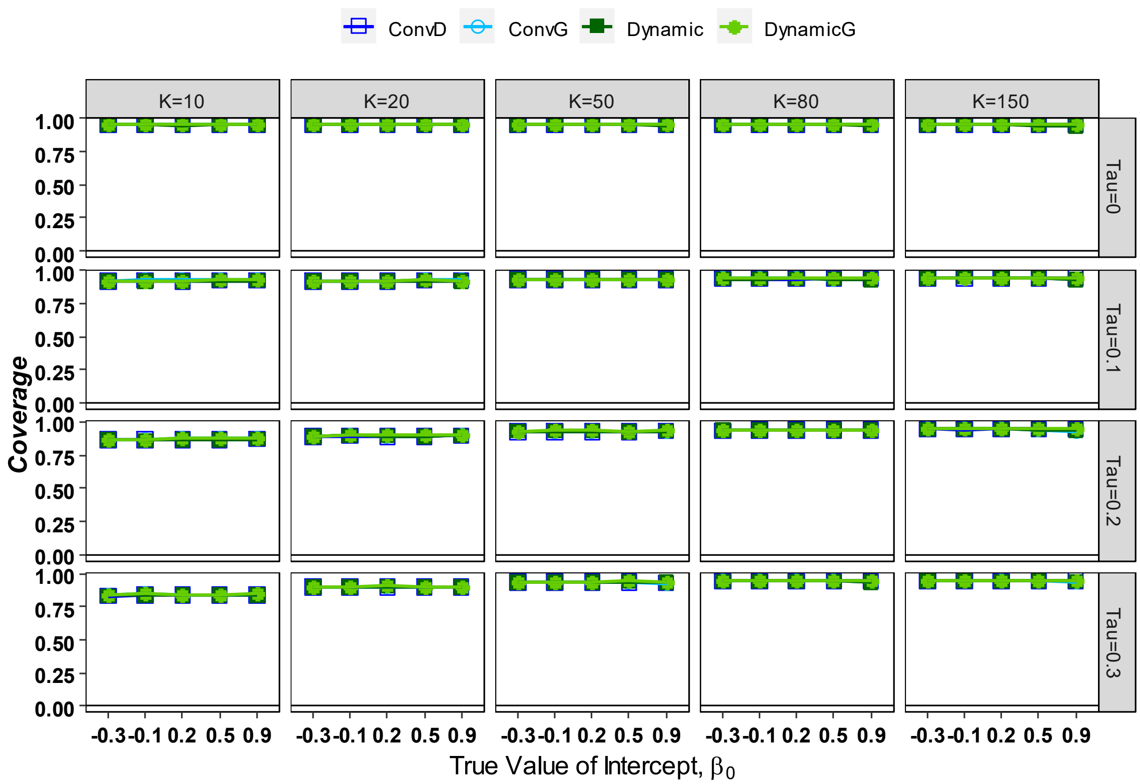
RMSE for the Fourth Conditional Mean for Standardized Mean Difference Effect Size Estimates: Frequentist Analysis



Note. The Fourth conditional mean is the sum of intercept and the regression coefficients of both predictors $(\beta_1 + \beta_2 = .3)$.

Figure 25

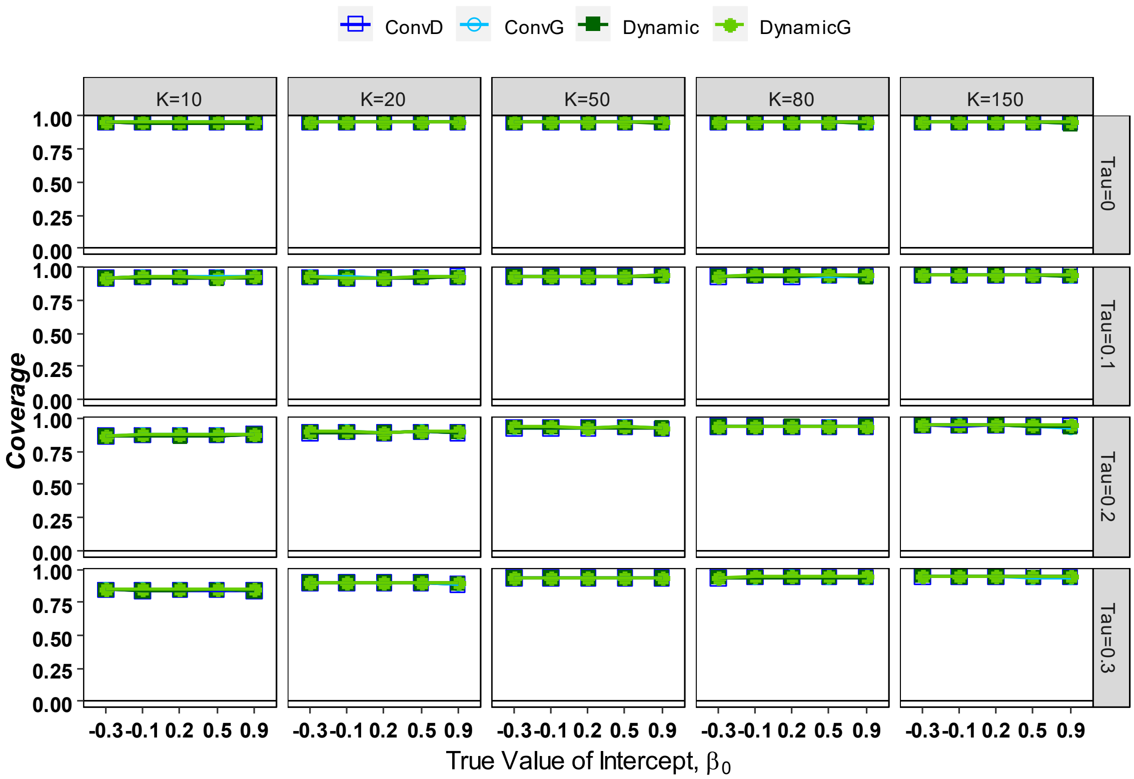
Coverage for the First Conditional Mean for Standardized Mean Difference Effect Size Estimates: Frequentist Analysis



Note. Coverage rate with 95% Nominal probability.

Figure 26

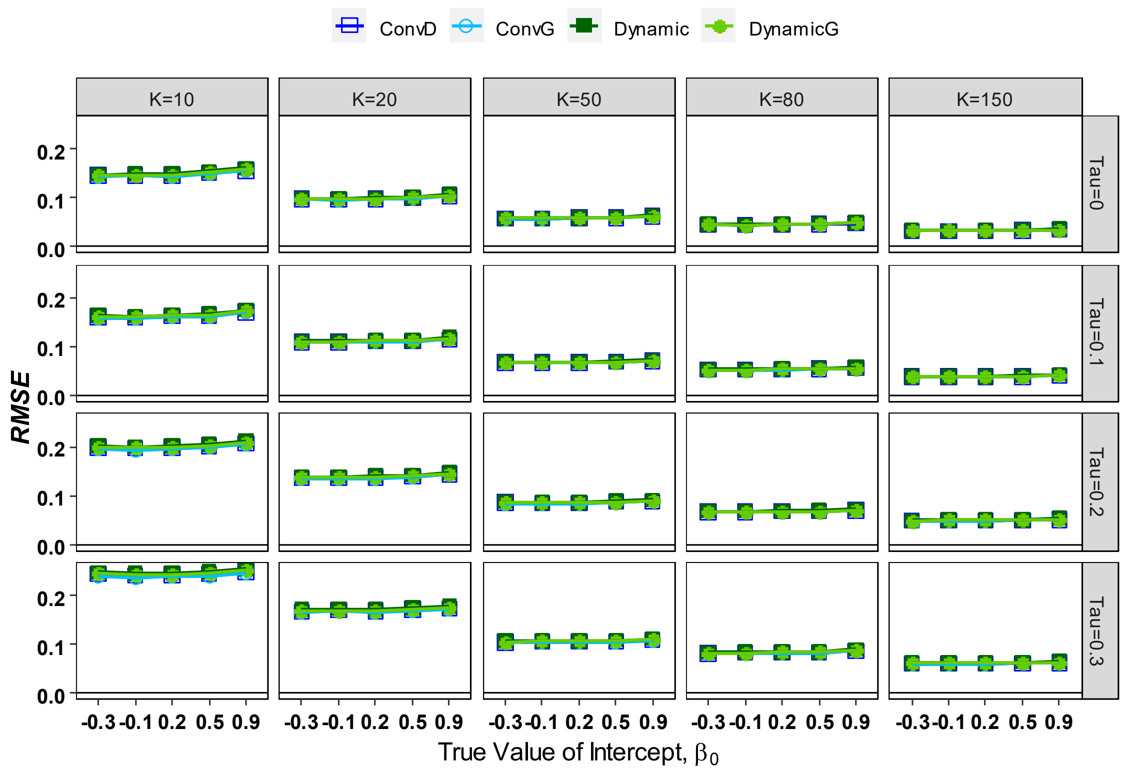
Coverage for the Second Conditional Mean for Standardized Mean Difference Effect Size Estimates: Frequentist Analysis



Note. Coverage rate with 95% Nominal probability. The second conditional mean is the sum of intercept and the regression coefficient of the second predictor $(\beta_2 = .1)$

Figure 27

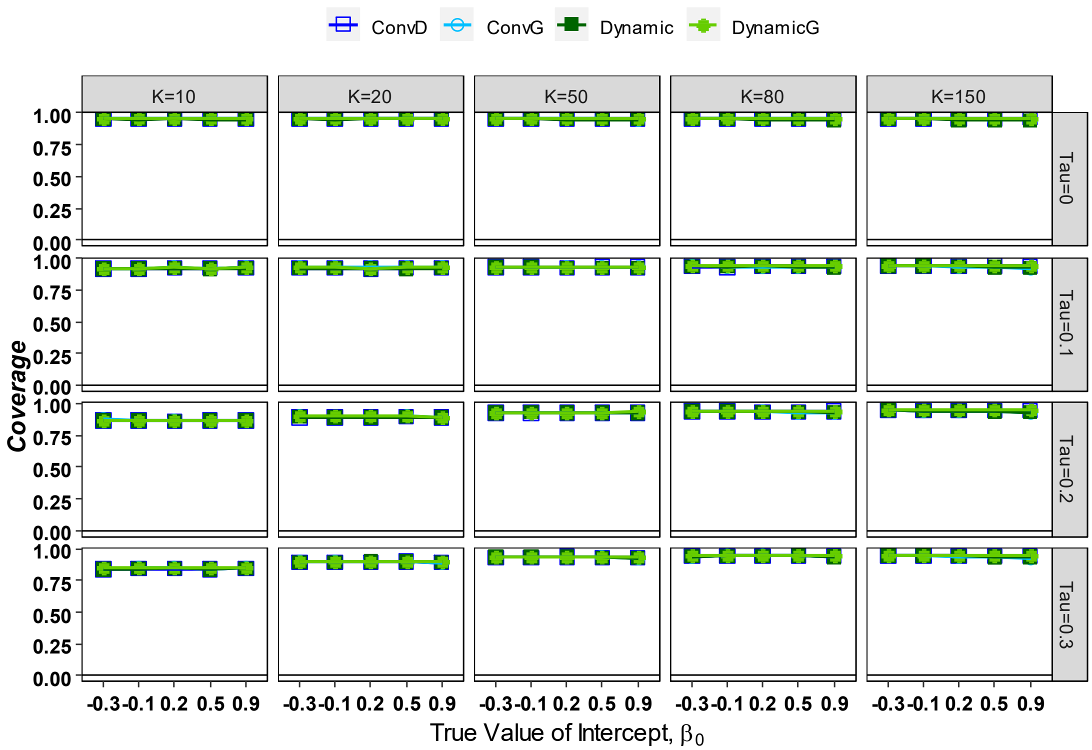
Coverage for the Third Conditional Mean for Standardized Mean Difference Effect Size Estimates: Frequentist Analysis



Note. Coverage rate with 95% Nominal probability. The Third conditional mean is the sum of intercept and the regression coefficient of the first predictor $(\beta_1 = .2)$.

Figure 28

Coverage for the Fourth Conditional Mean for Standardized Mean Difference Effect Size Estimates: Frequentist Analysis



Note. Coverage rate with 95% Nominal probability. The Fourth conditional mean is the sum of intercept and the regression coefficients of both predictors $(\beta_1 + \beta_2 = .3)$.

Figure 29

Bias for Tau for Standardized Mean Difference Effect Size Estimates: Frequentist Analysis

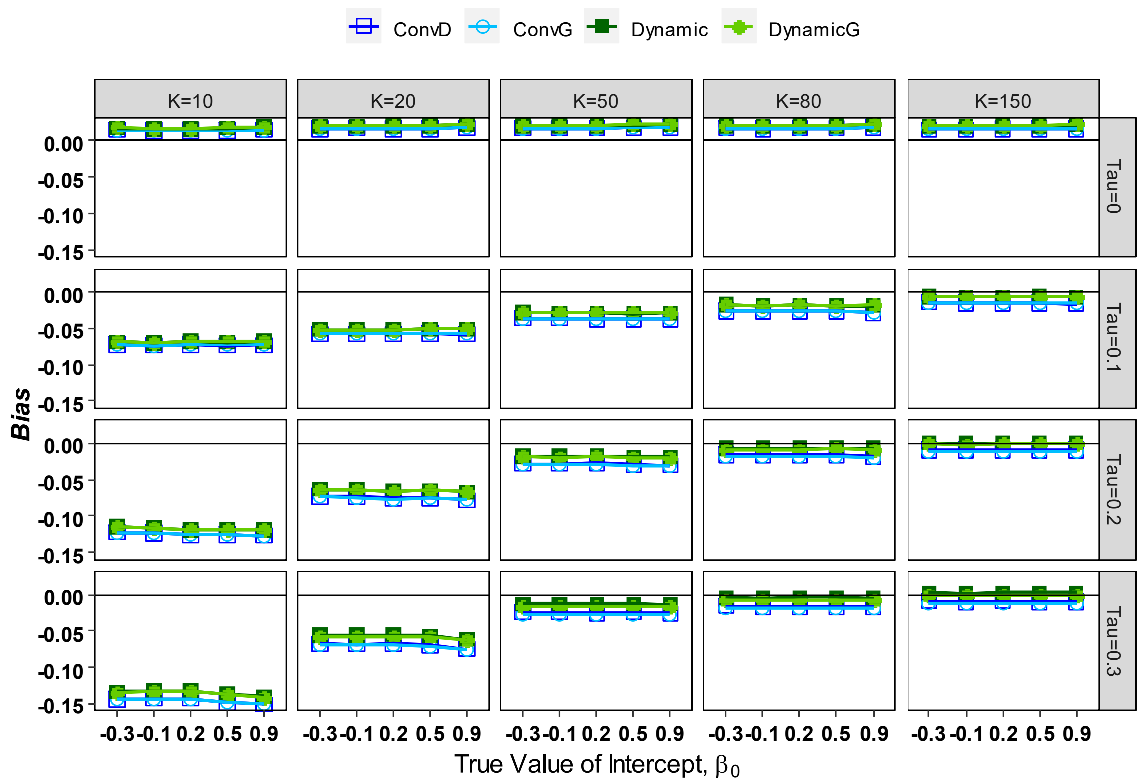


Figure 30

RMSE for Tau for Standardized Mean Difference Effect Size Estimates: Frequentist Analysis

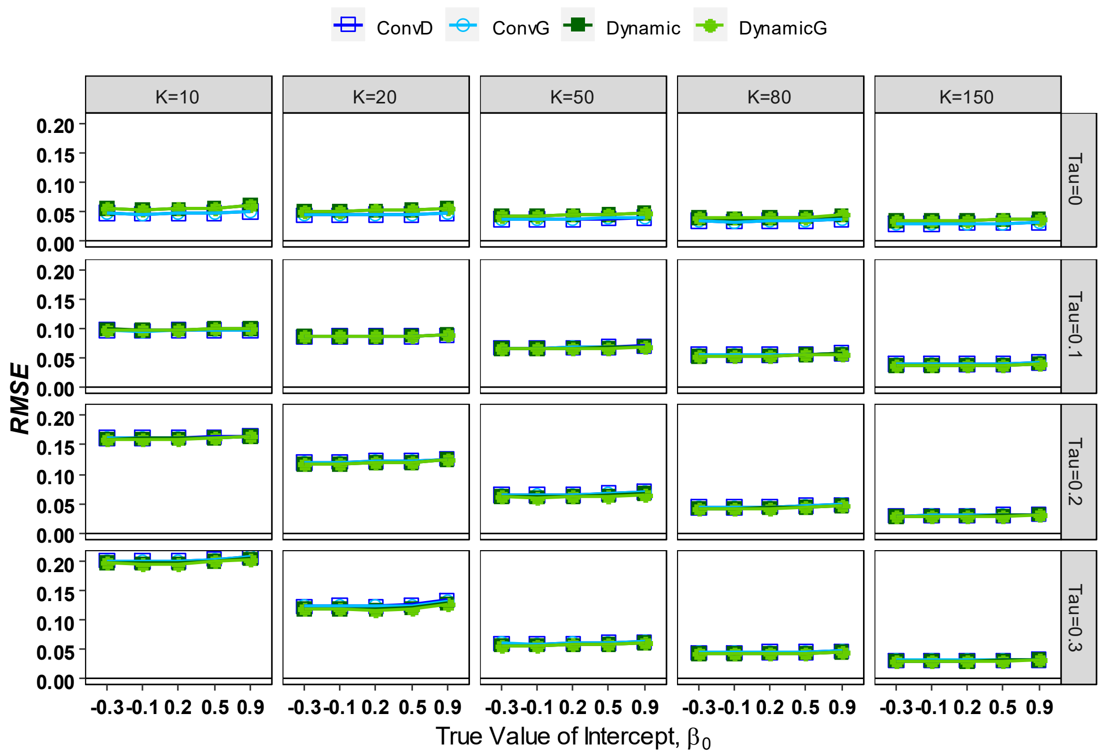


Figure 31

Bias for the First Conditional Mean for Correlation Effect Size Estimates: Bayesian Analysis

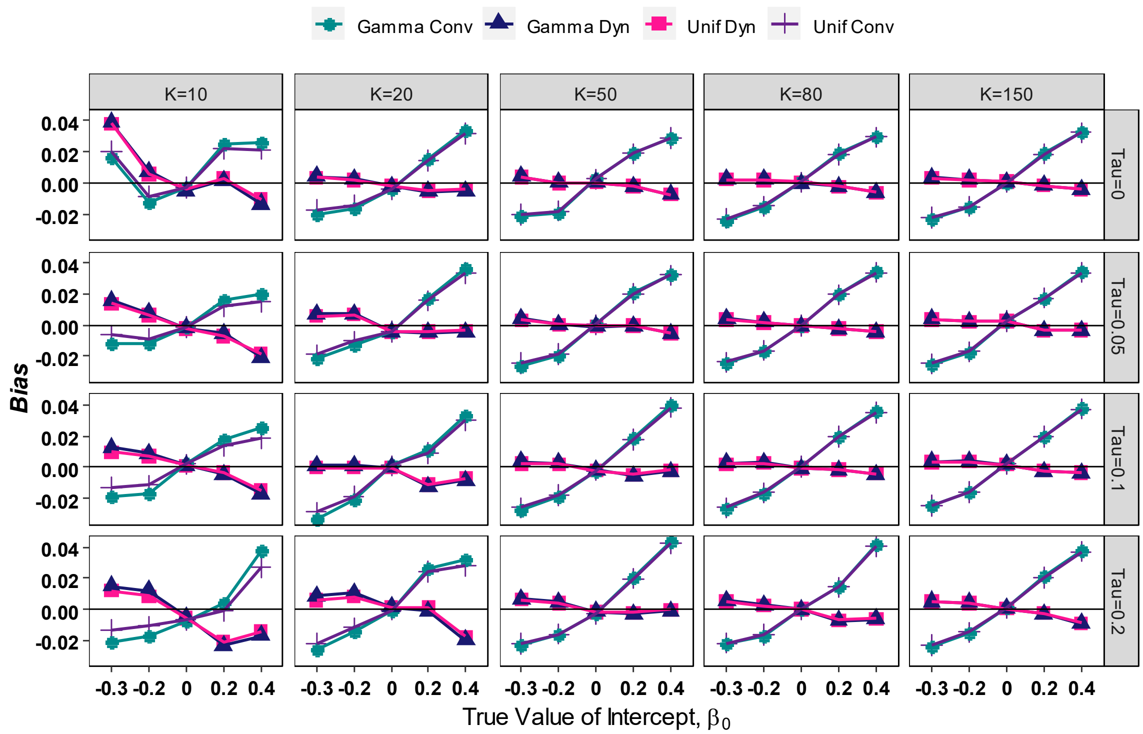
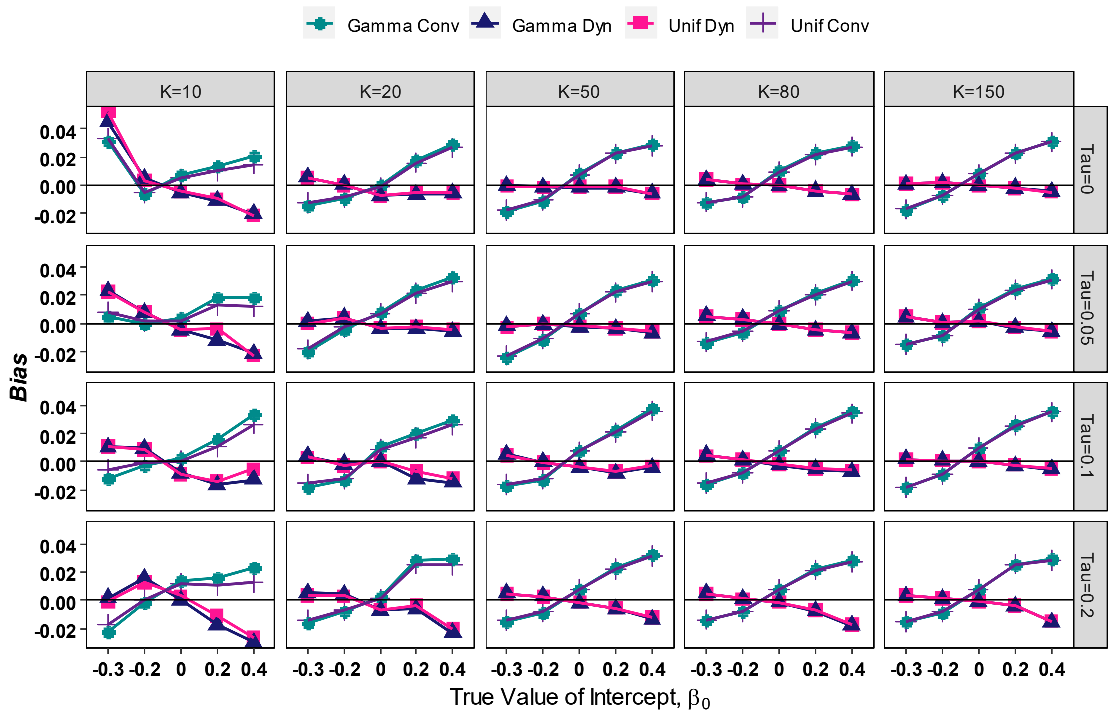


Figure 32

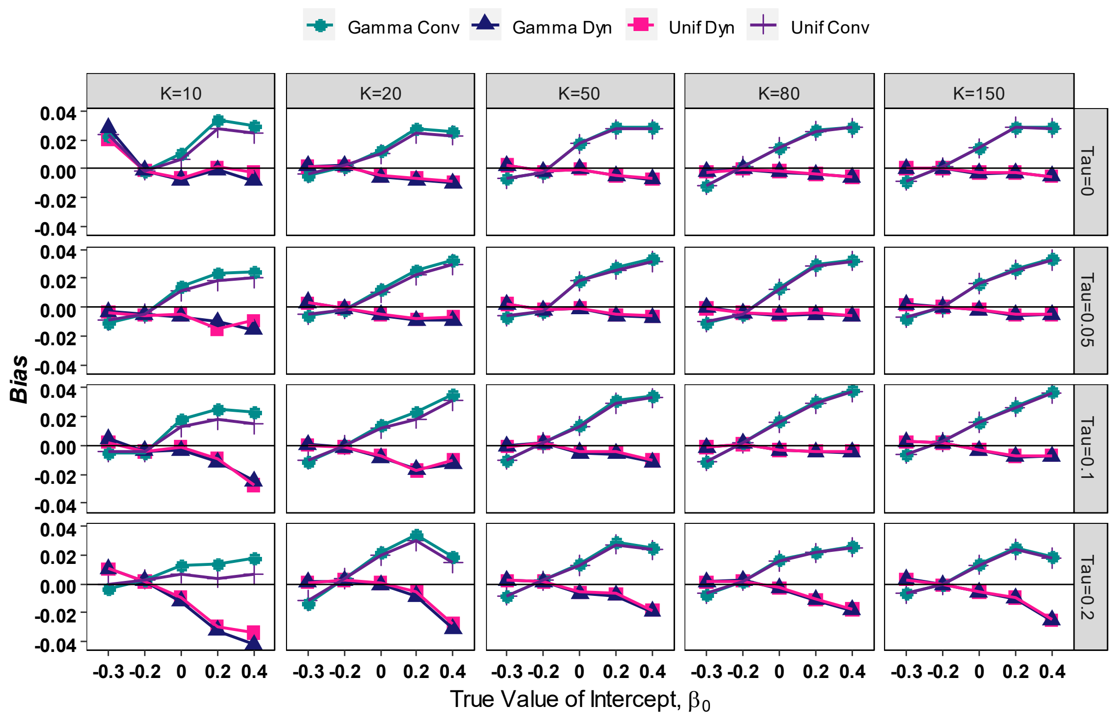
Bias for the Second Conditional Mean for Correlation Effect Size Estimates: Bayesian Analysis



Note. The second conditional mean is the sum of intercept and the regression coefficient of the second predictor $(\beta_2 = .1)$.

Figure 33

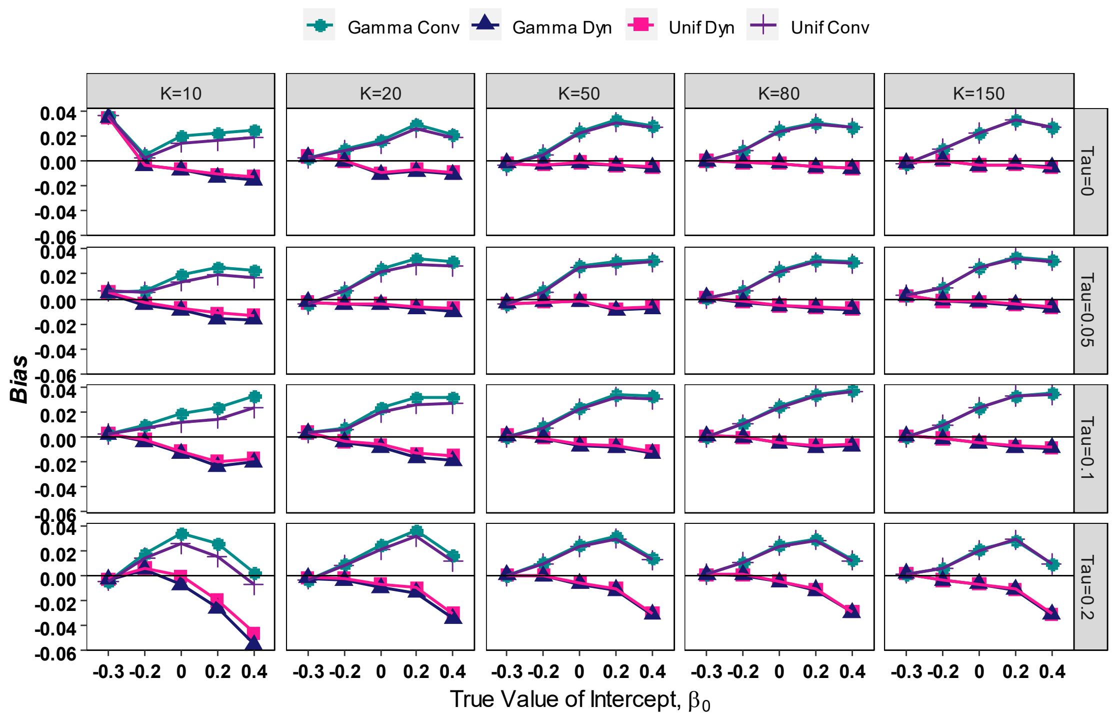
Bias for the Third Conditional Mean for Correlation Effect Size Estimates: Bayesian Analysis



Note. The Fourth conditional mean is the sum of intercept and the regression coefficients of both predictors $(\beta_2 = .2)$.

Figure 34

Bias for the Fourth Conditional Mean for Correlation Effect Size Estimates: Bayesian Analysis



Note. The Fourth conditional mean is the sum of intercept and the regression coefficients of both predictors $(\beta_1 + \beta_2 = .3)$.

Figure 35

RMSE for the First Conditional Mean for Correlation Effect Size Estimates: Bayesian Analysis

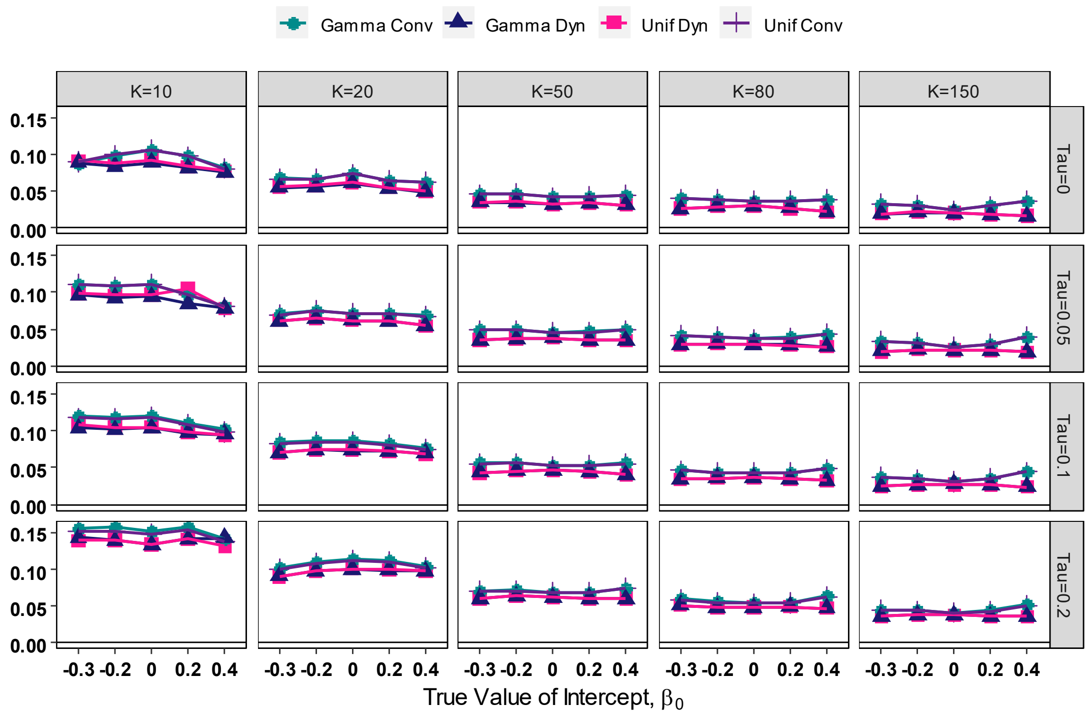
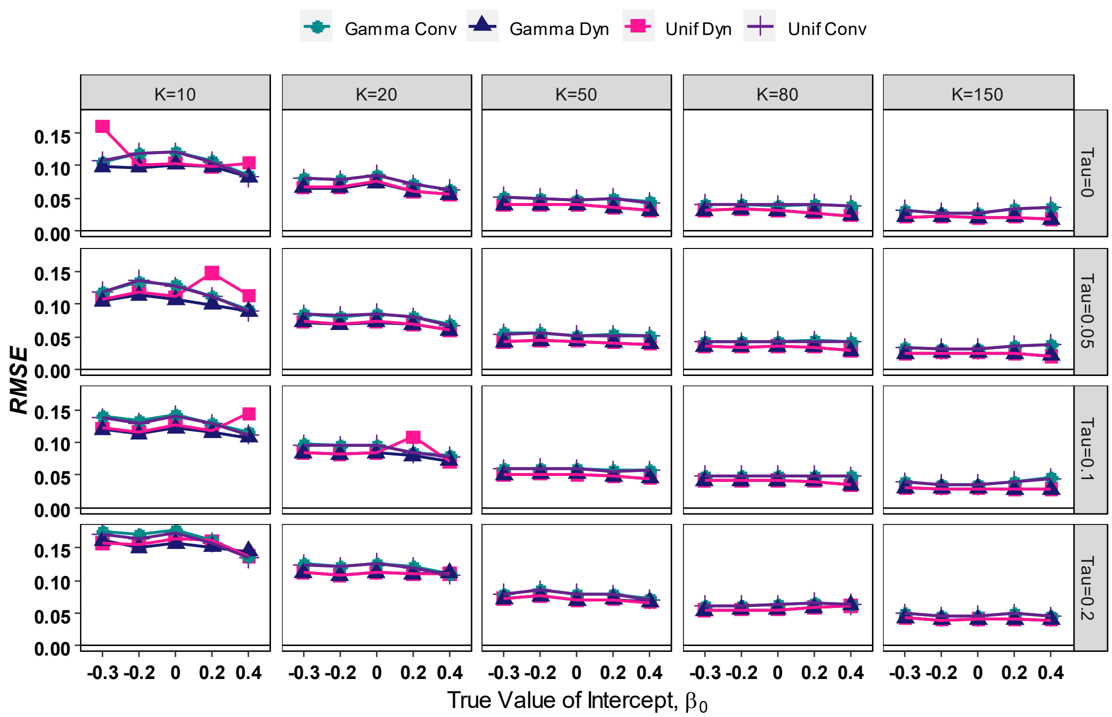


Figure 36

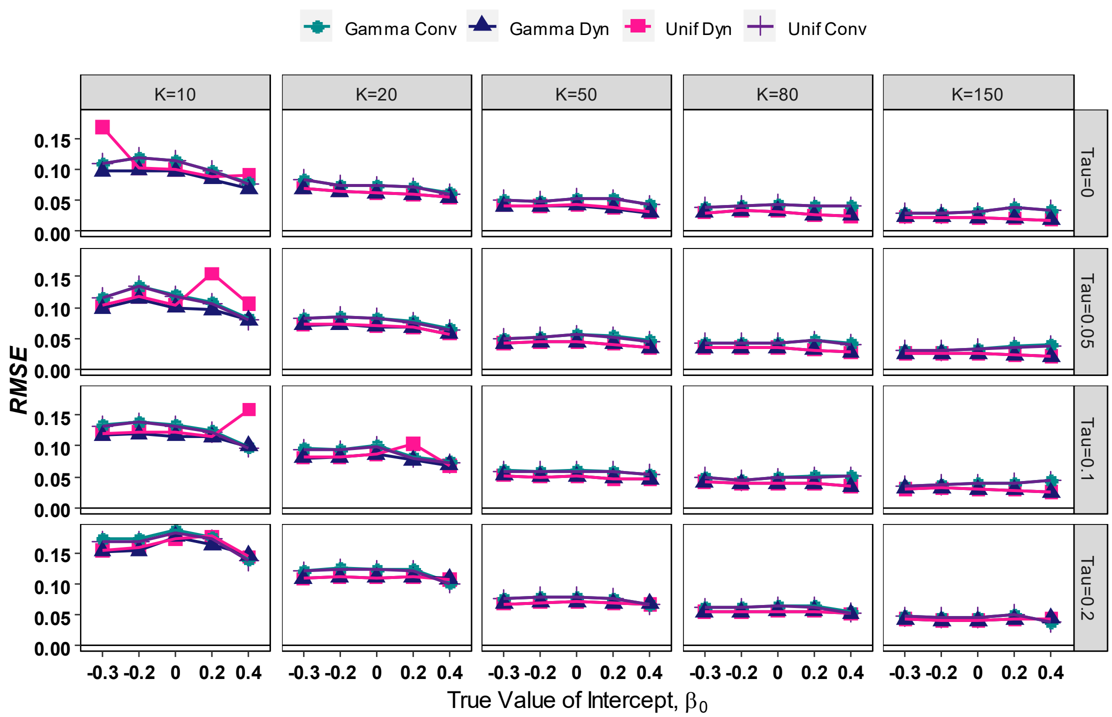
RMSE for the Second Conditional Mean for Correlation Effect Size Estimates: Bayesian Analysis



Note. The second conditional mean is the sum of intercept and the regression coefficient of the second predictor $(\beta_2 = .1)$.

Figure 37

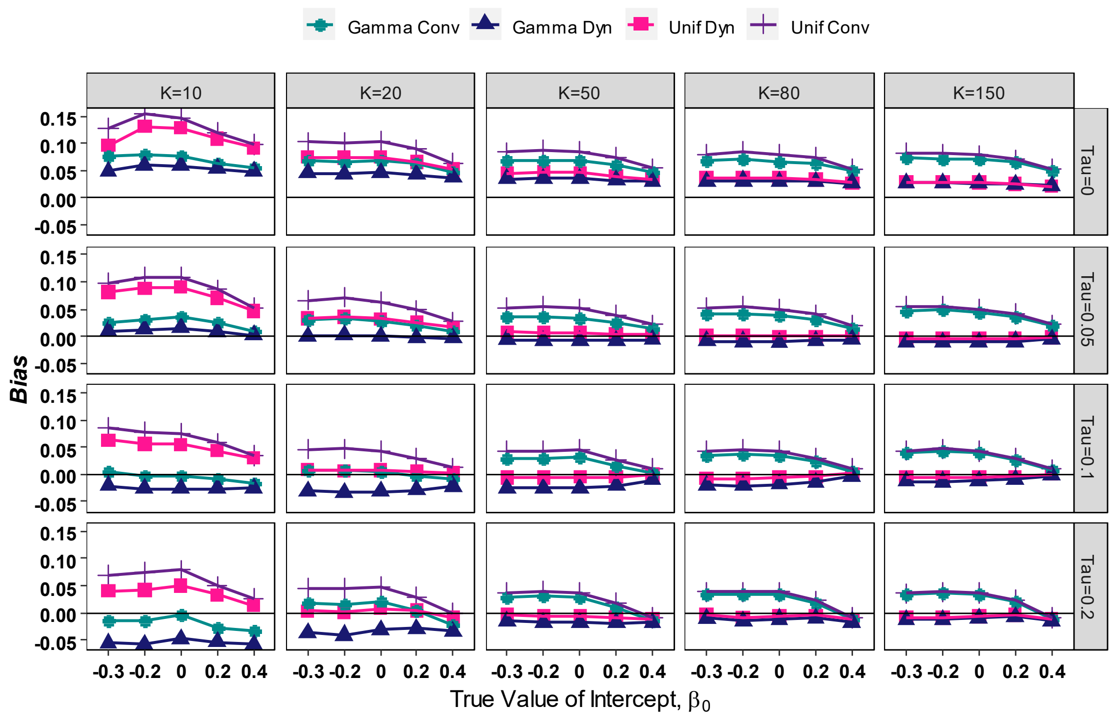
RMSE for the Third Conditional Mean for Correlation Effect Size Estimates: Bayesian Analysis



Note. The Third conditional mean is the sum of intercept and the regression coefficient of the first predictor $(\beta_1 = .2)$.

Figure 38

RMSE for the Fourth Conditional Mean for Correlation Effect Size Estimates: Bayesian Analysis



Note. The Fourth conditional mean is the sum of intercept and the regression coefficients of both predictors $(\beta_1 + \beta_2 = .3)$.

Figure 39

Bias for Tau for Correlation Effect Size Estimates: Bayesian Analysis

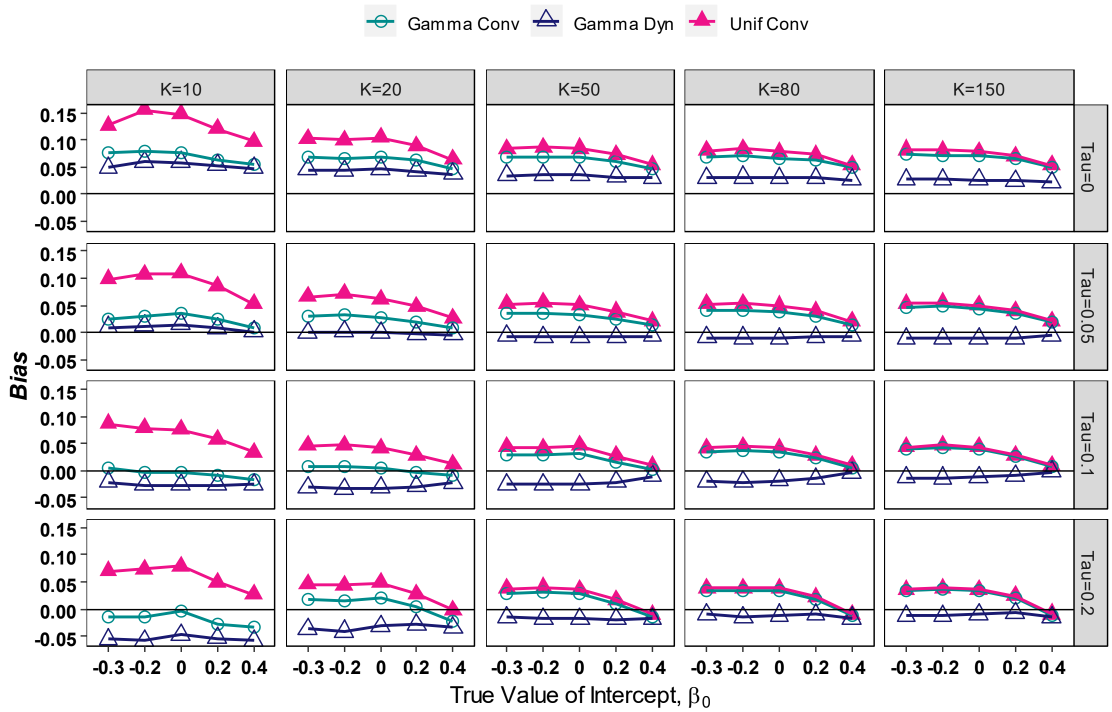


Figure 40

RMSE for Tau for Correlation Effect Size Estimates: Bayesian Analysis

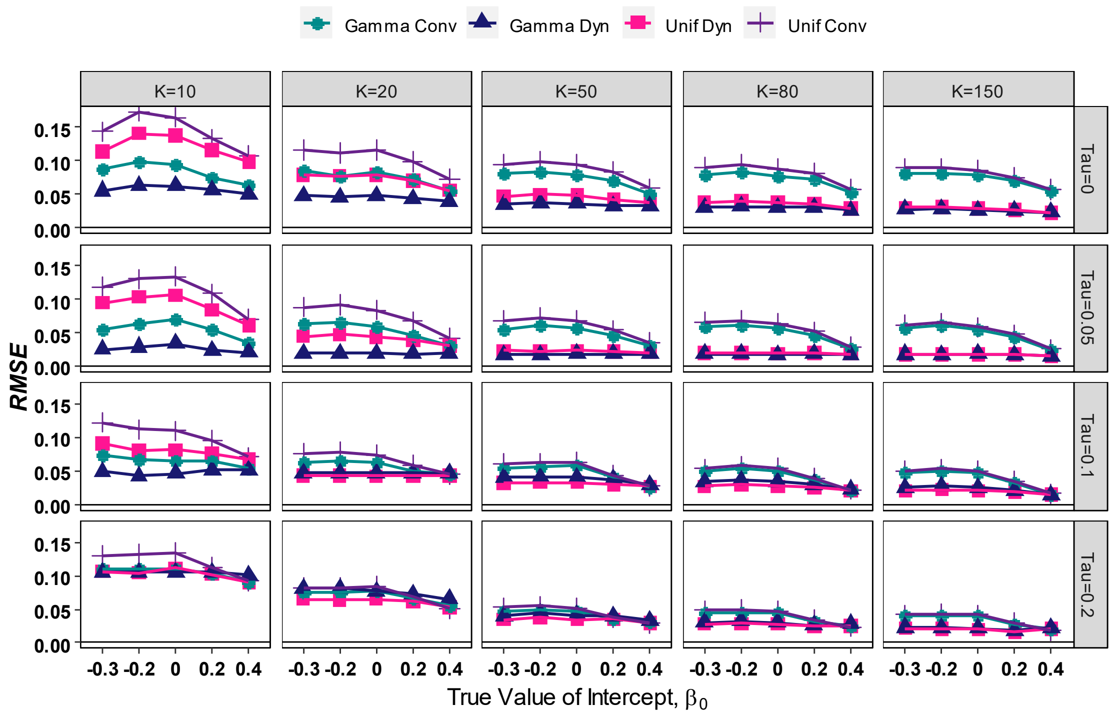


Figure 41

Bias for the First Conditional Mean for Standardized Mean Difference Effect Size Estimates: Bayesian Analysis

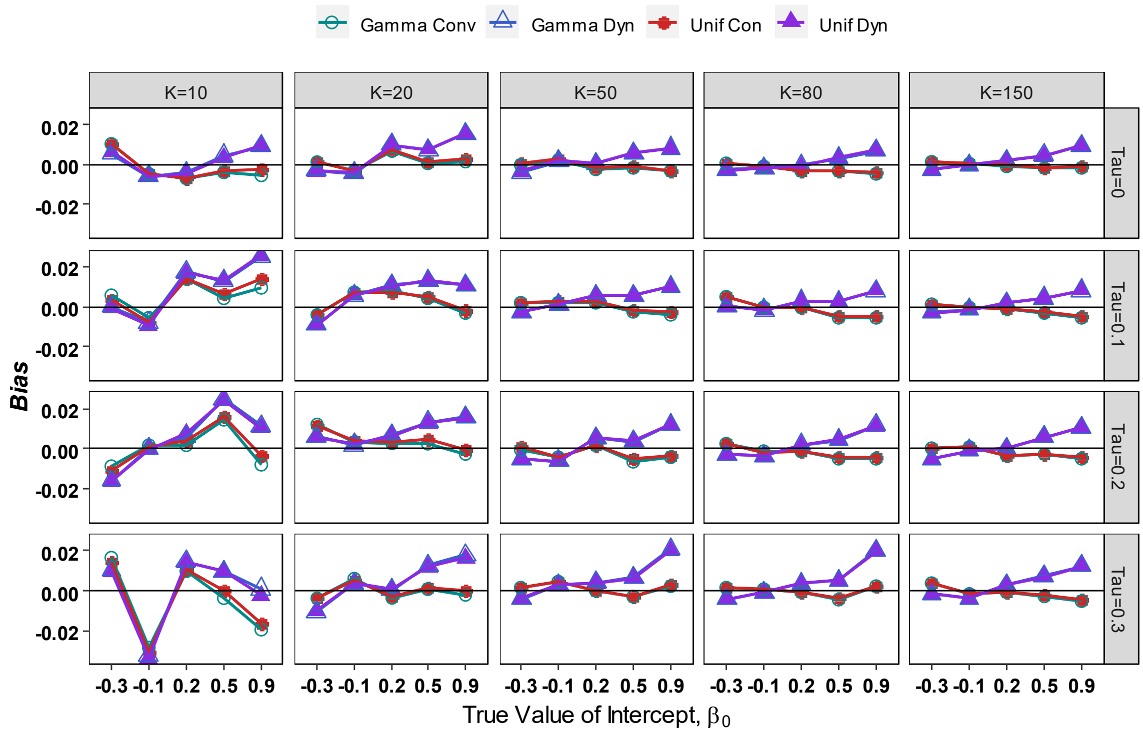
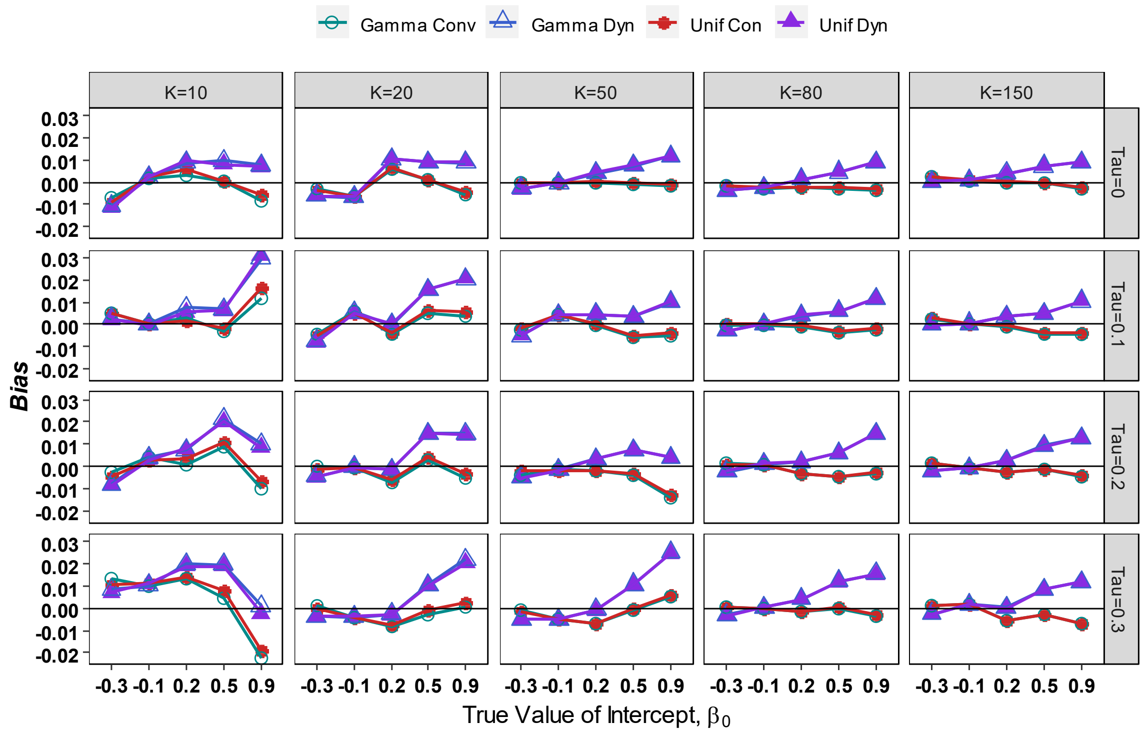


Figure 42

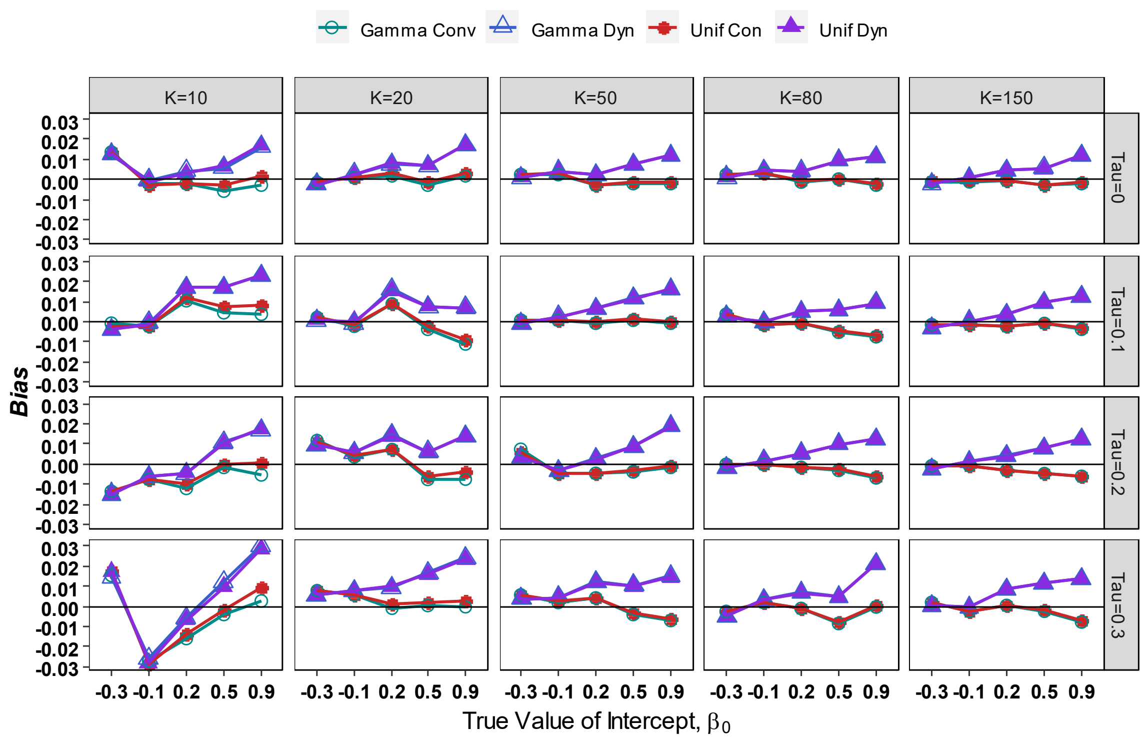
Bias for the Second Conditional Mean for Standardized Mean Difference Effect Size Estimates: Bayesian Analysis



Note. The second conditional mean is the sum of intercept and the regression coefficient of the second predictor $(\beta_2 = .1)$.

Figure 43

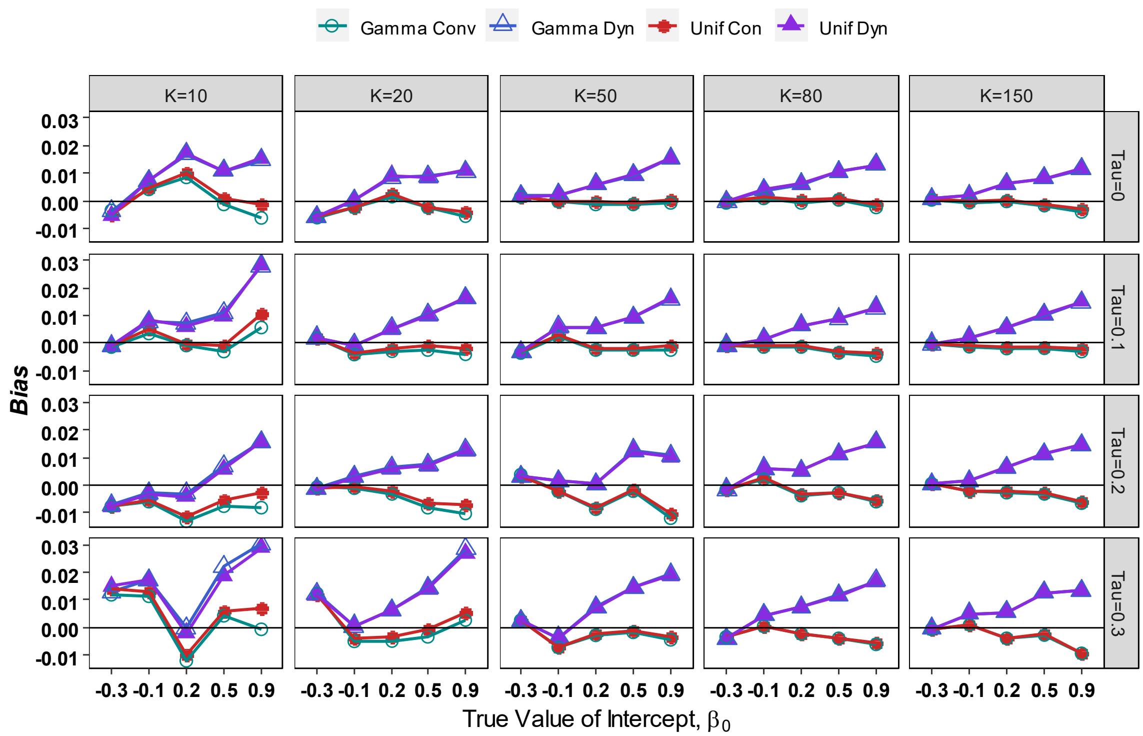
Bias for the Third Conditional Mean for Standardized Mean Difference Effect Size Estimates: Bayesian Analysis



Note. The Third conditional mean is the sum of intercept and the regression coefficient of the first predictor $(\beta_1 = .2)$.

Figure 44

Bias for the Fourth Conditional Mean for Standardized Mean Difference Effect Size Estimates: Bayesian Analysis



Note. The Fourth conditional mean is the sum of intercept and the regression coefficients of both predictors $(\beta_1 + \beta_2 = .3)$.

Figure 45

RMSE for the First Conditional Mean for Standardized Mean Difference Effect Size Estimates: Bayesian Analysis

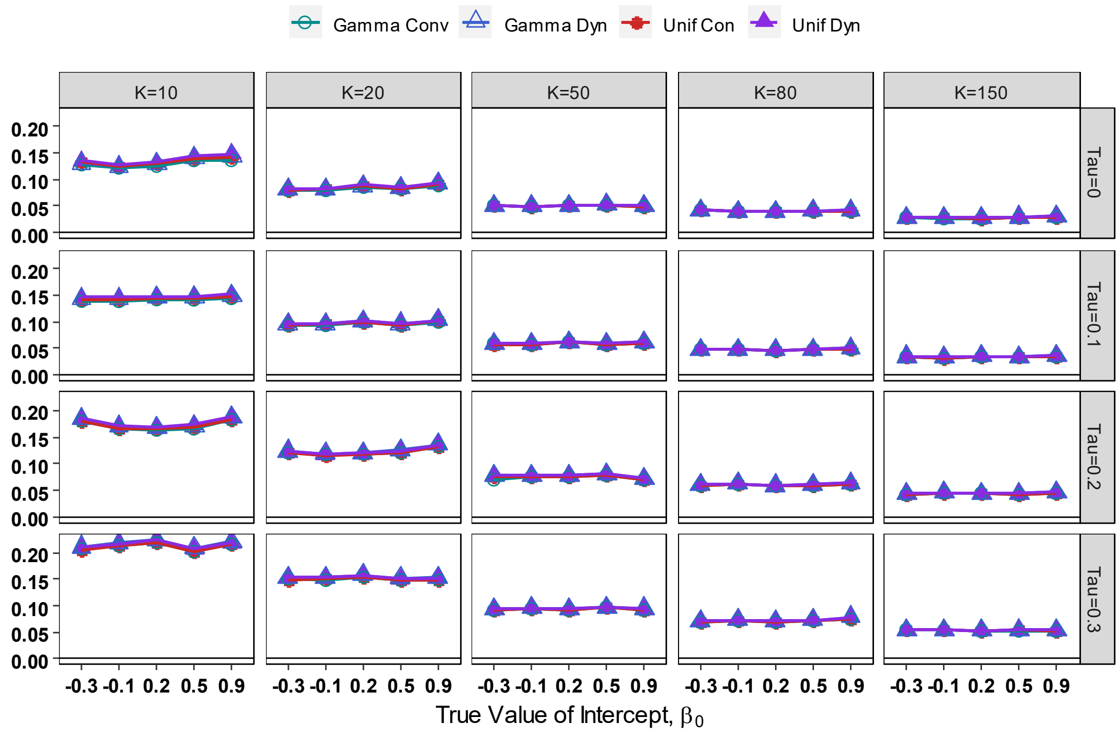
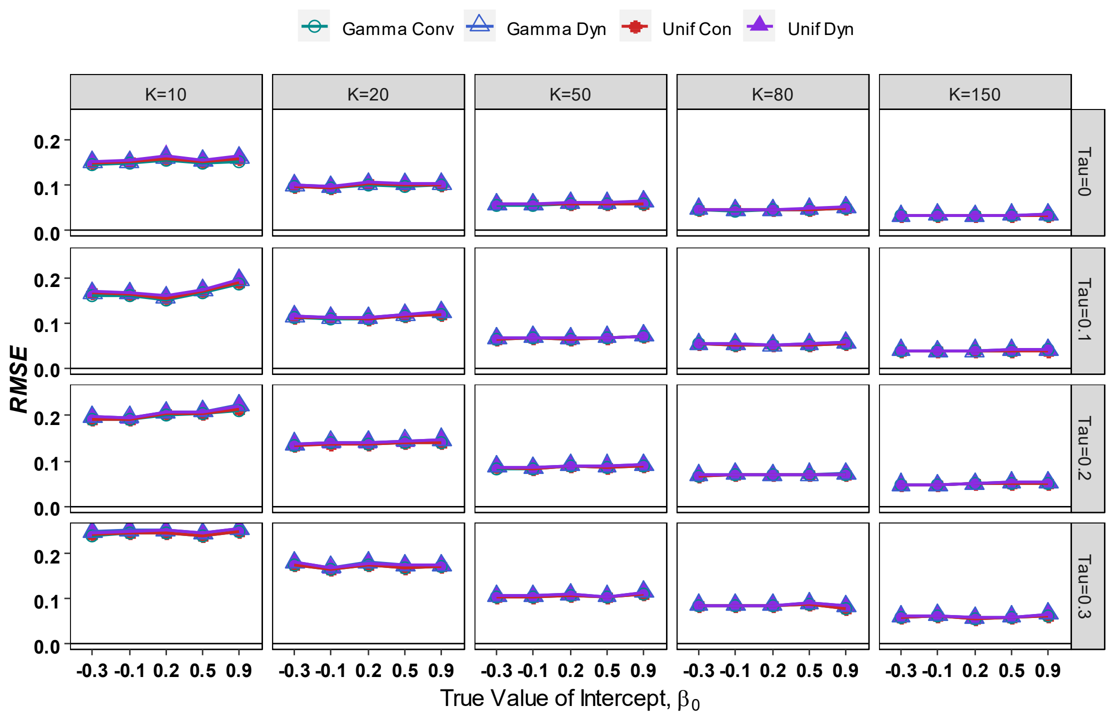


Figure 46

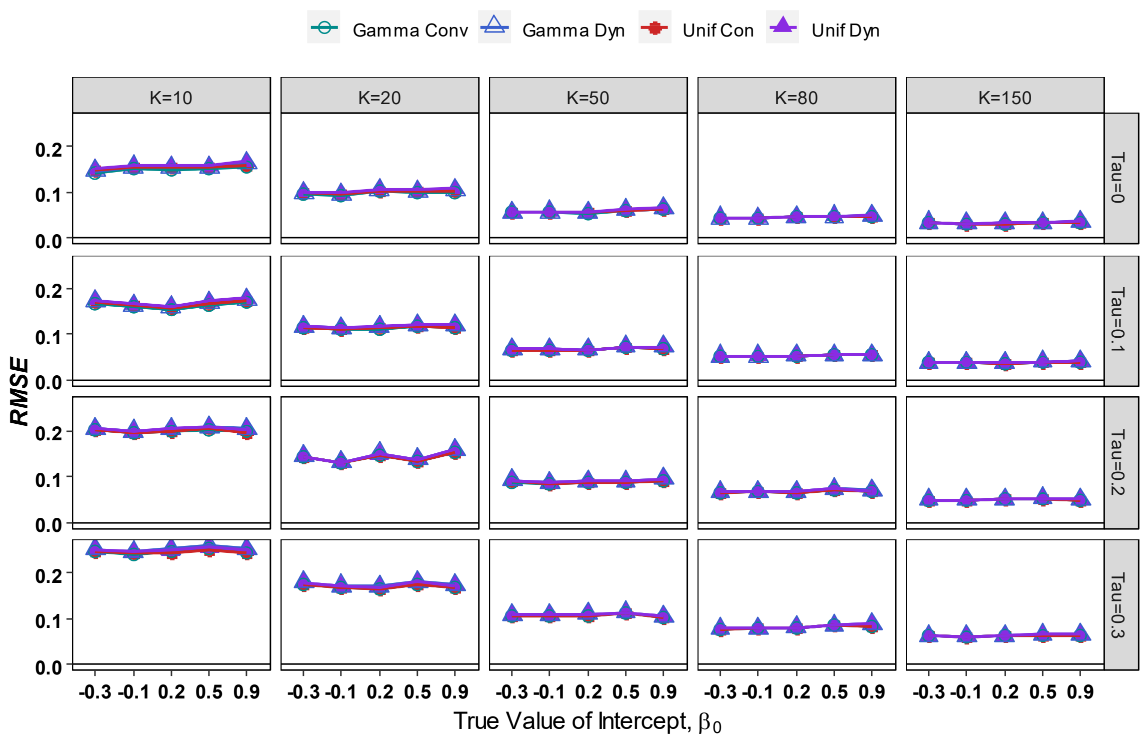
RMSE for the Second Conditional Mean for Standardized Mean Difference Effect Size Estimates: Bayesian Analysis



Note. The second conditional mean is the sum of intercept and the regression coefficient of the second predictor $(\beta_2 = .1)$.

Figure 47

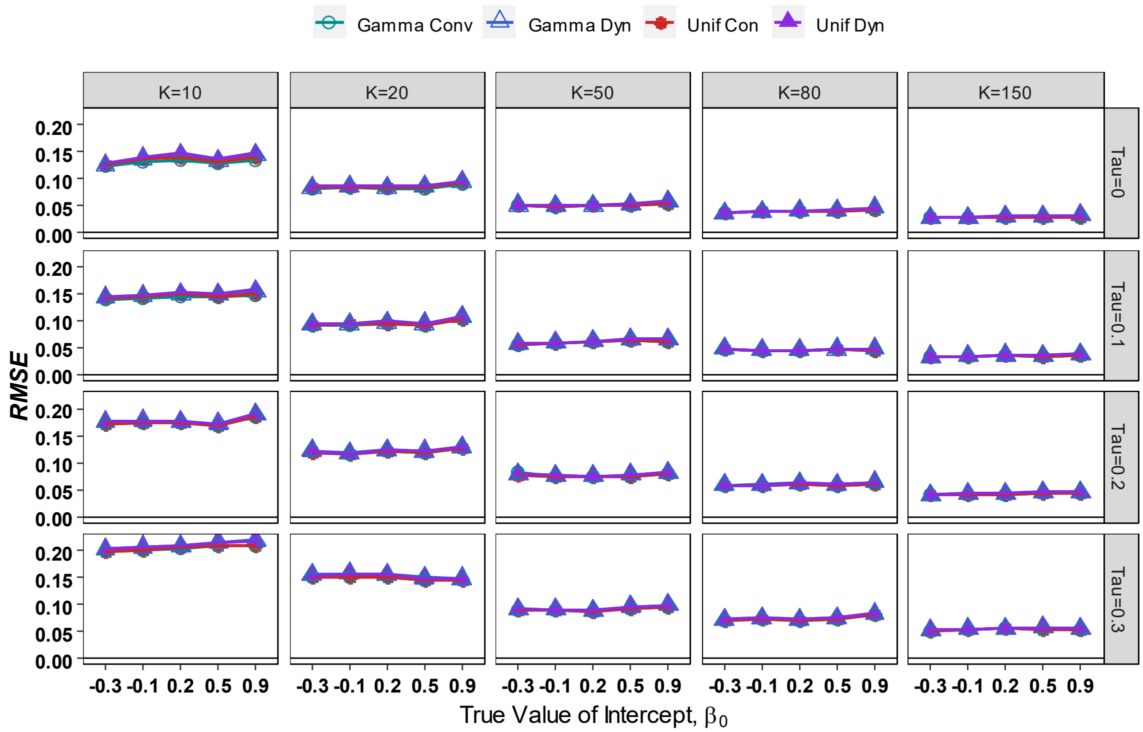
RMSE for the Third Conditional Mean for Standardized Mean Difference Effect Size Estimates: Bayesian Analysis



Note. The Third conditional mean is the sum of intercept and the regression coefficient of the first predictor $(\beta_1 = .2)$.

Figure 48

RMSE for the Fourth Conditional Mean for Standardized Mean Difference Effect Size Estimates: Bayesian Analysis



Note. The Fourth conditional mean is the sum of intercept and the regression coefficients of both predictors $(\beta_1 + \beta_2 = .3)$.

Figure 49

Bias for Tau for Standardized Mean Difference Effect Size Estimates: Bayesian Analysis

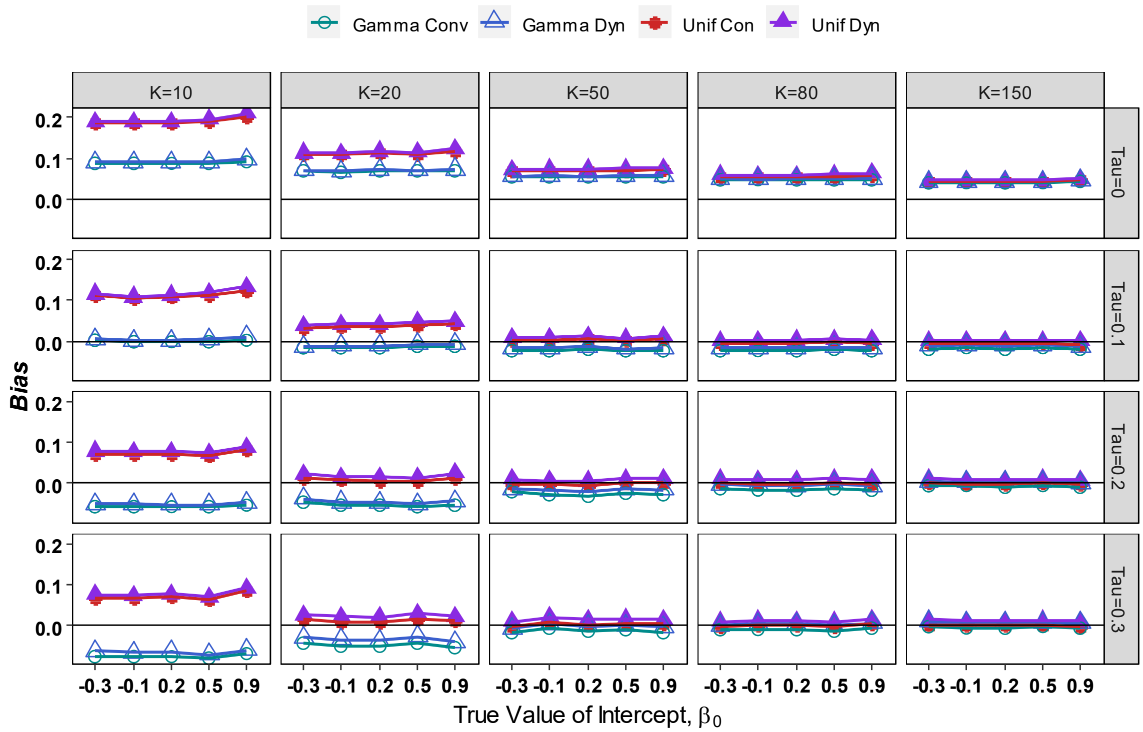


Figure 50

RMSE for Tau for Standardized Mean Difference Effect Size Estimates: Bayesian Analysis

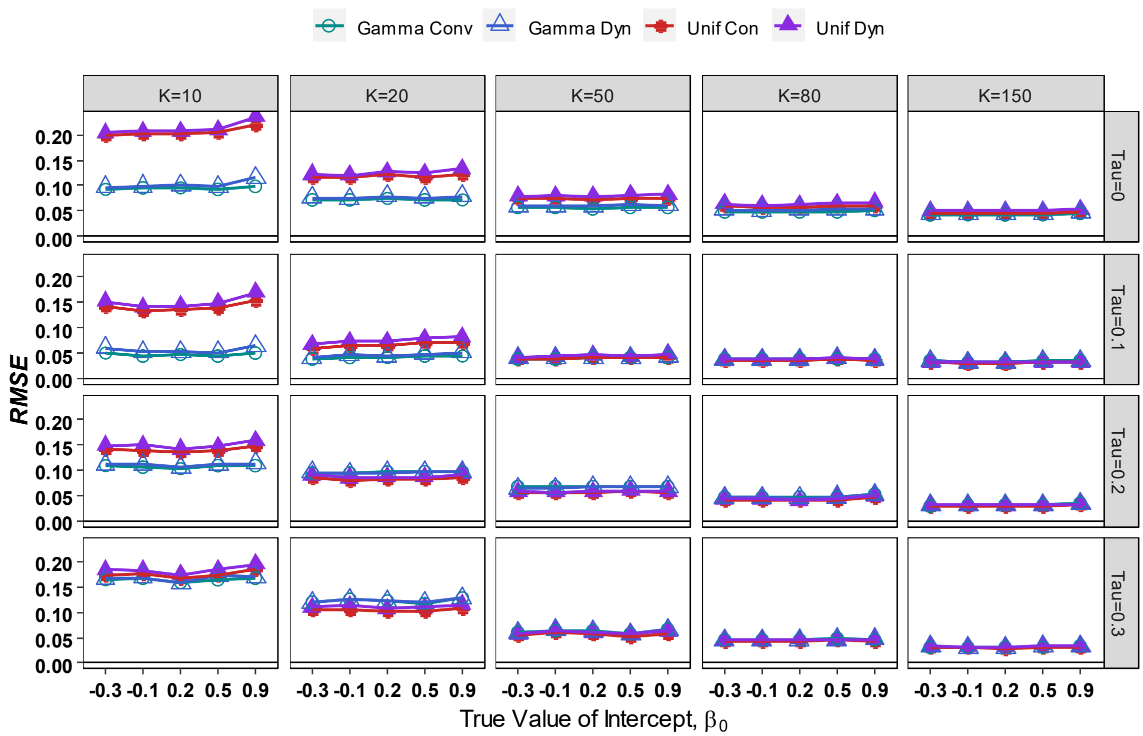
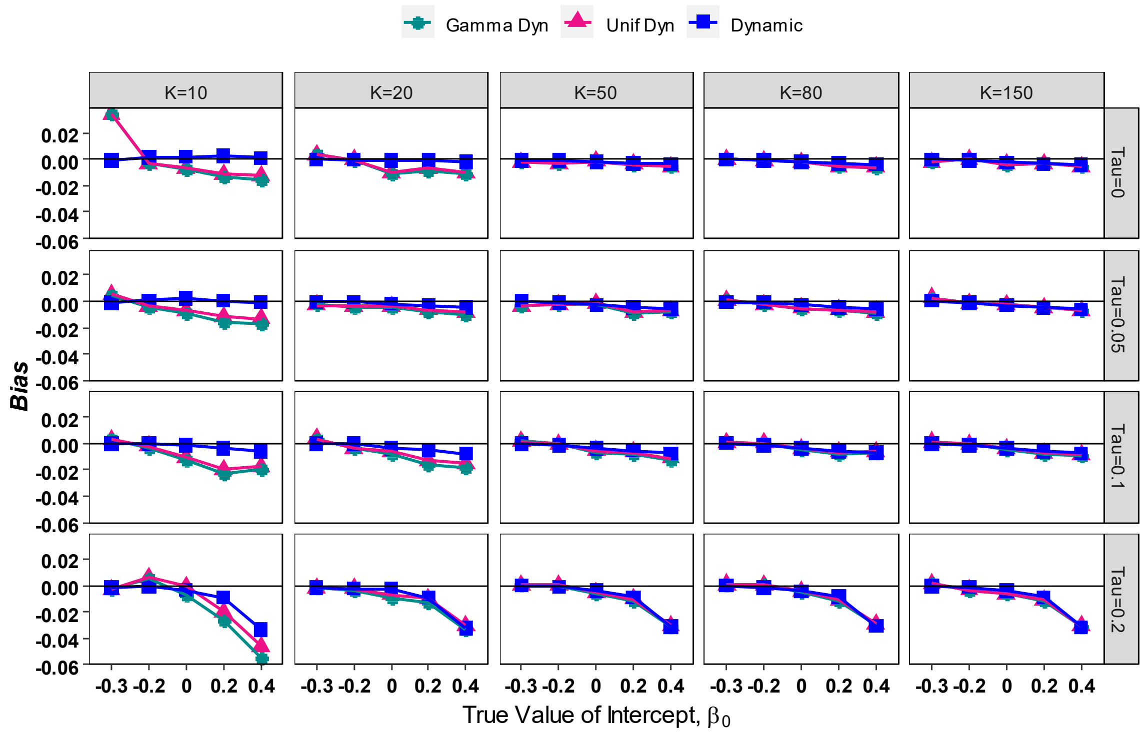


Figure 51

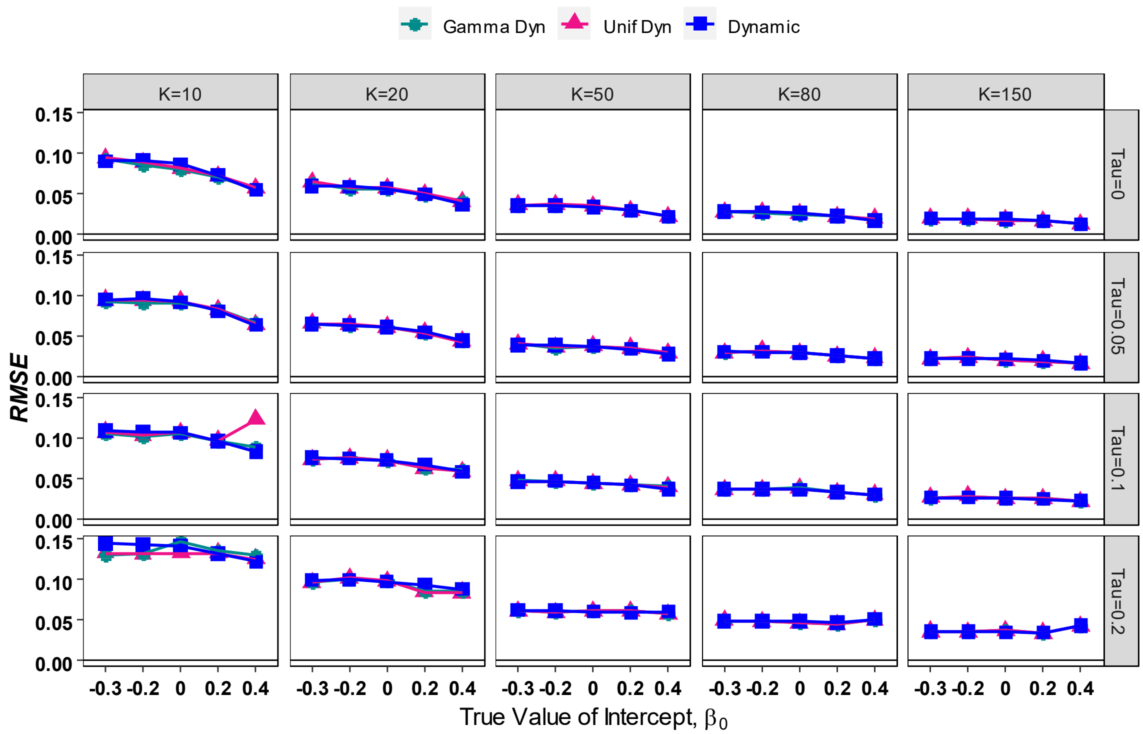
Bias for the Fourth Conditional Mean for Correlation Effect Size: All Dynamic Methods



Note. Plots for the second, third, and fourth conditional mean for all dynamic correlation methods are not included. Results for the other conditional means followed a similar pattern.

Figure 52

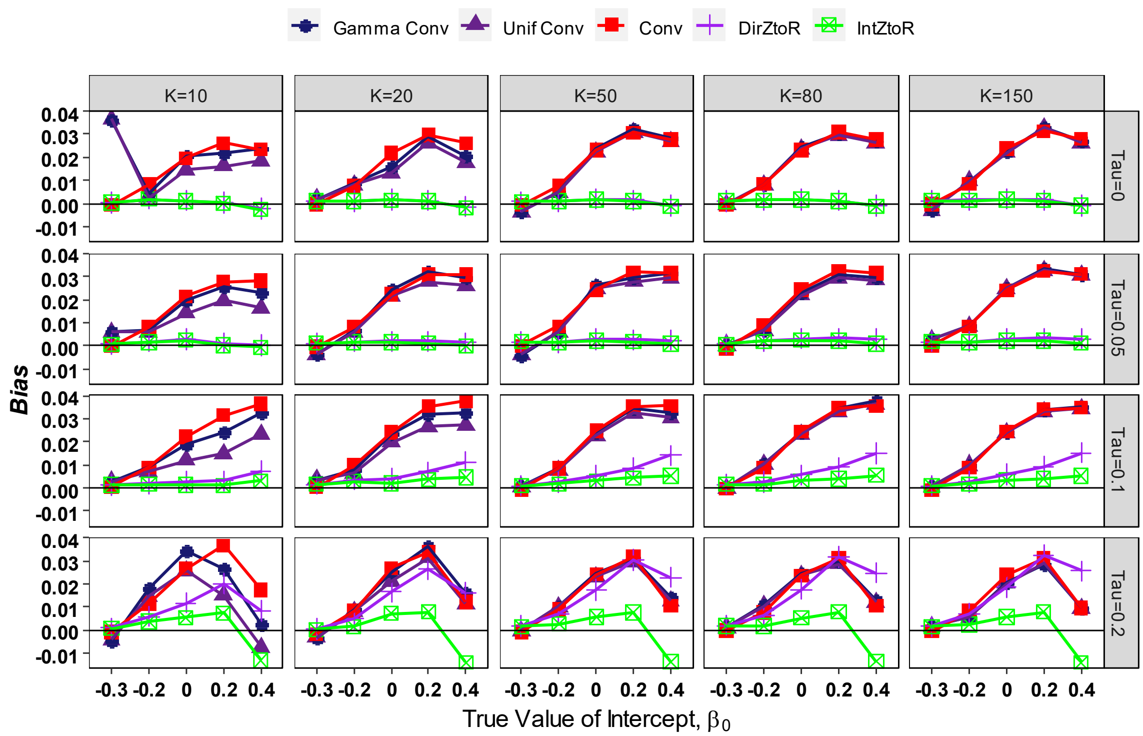
RMSE for the Fourth Conditional Mean for Correlation Effect Size: All Dynamic Methods



Note. Plots for the second, third, and fourth conditional mean for all dynamic correlation methods are not included. Results for the other conditional means followed a similar pattern.

Figure 53

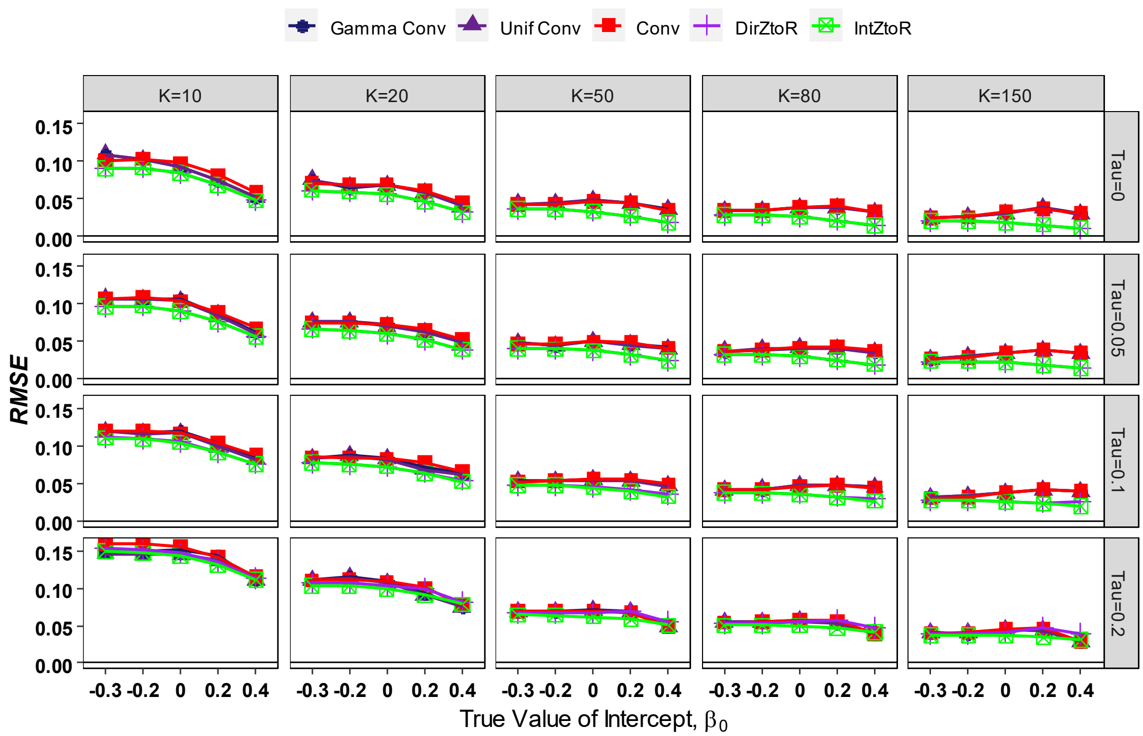
Bias for the Fourth Conditional Mean for Correlation Effect Size: All Conventional Methods



Note. Plots for the second, third, and fourth conditional mean for all conventional correlation methods are not included. Results for the other conditional means followed a similar pattern.

Figure 54

RMSE for the Fourth Conditional Mean for Correlation Effect Size: All Conventional Methods



Note. Plots for the second, third, and fourth conditional mean for all conventional correlation methods are not included. Results for the other conditional means followed a similar pattern.

Figure 55

Bias for Tau for Correlation Effect Size: All Dynamic Methods

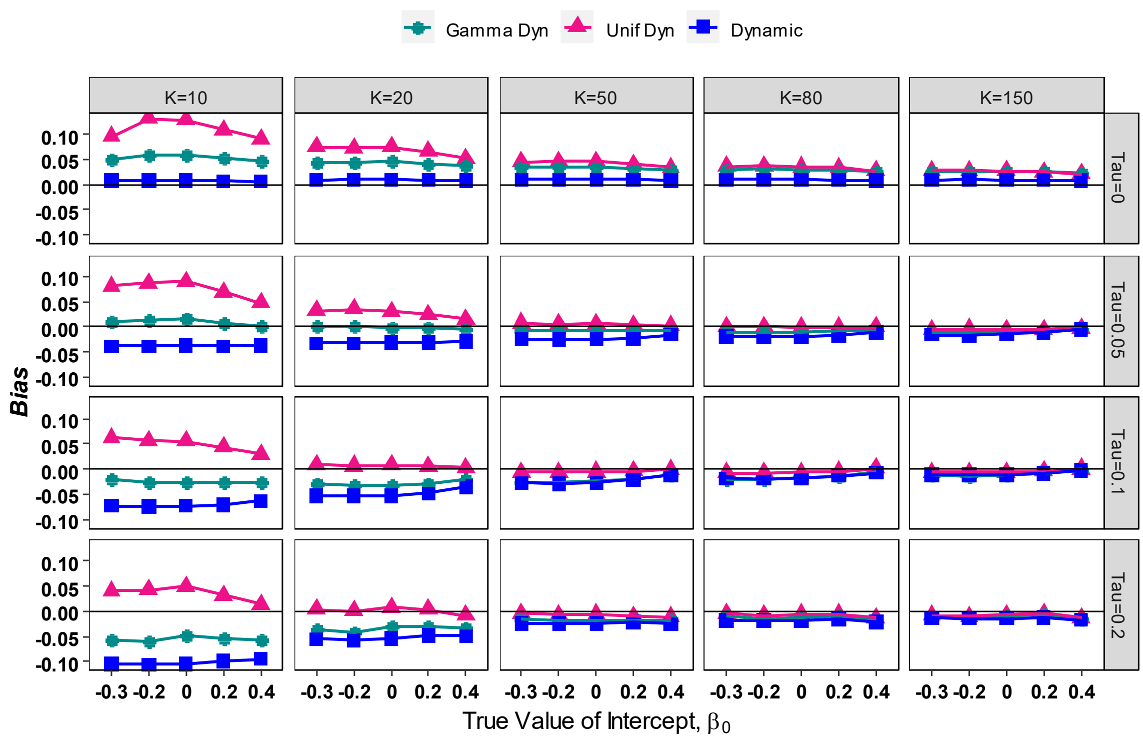


Figure 56

RMSE for the Tau for Correlation Effect Size: All Dynamic Methods

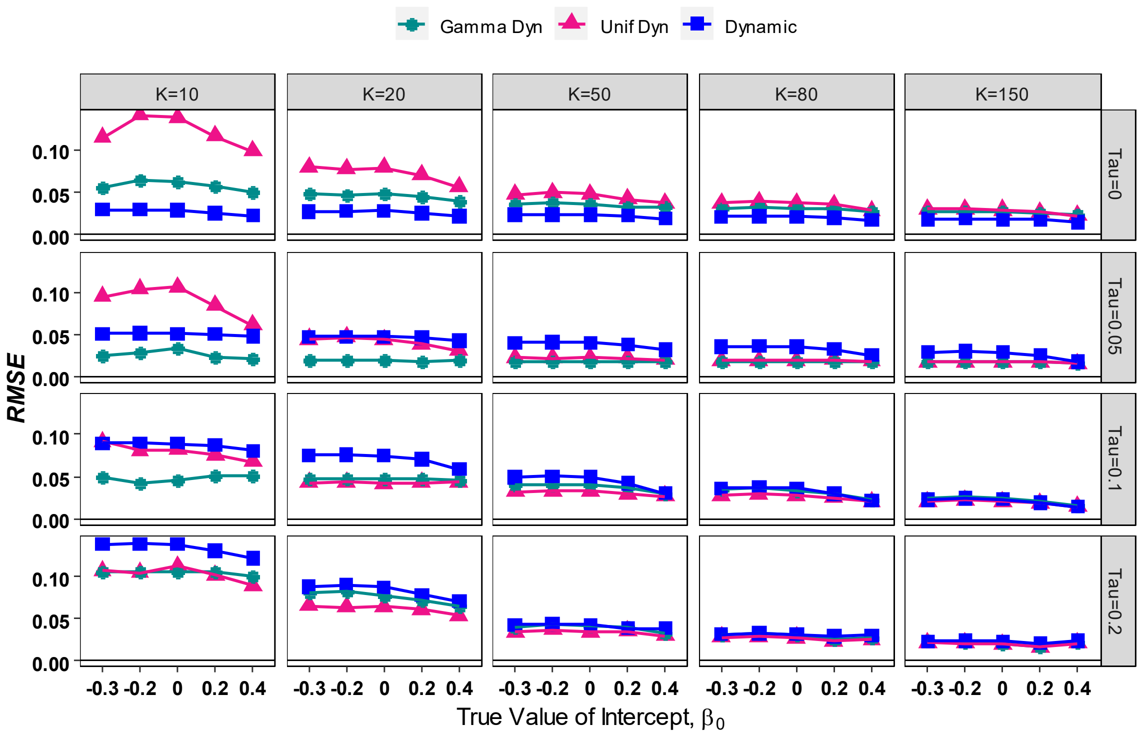
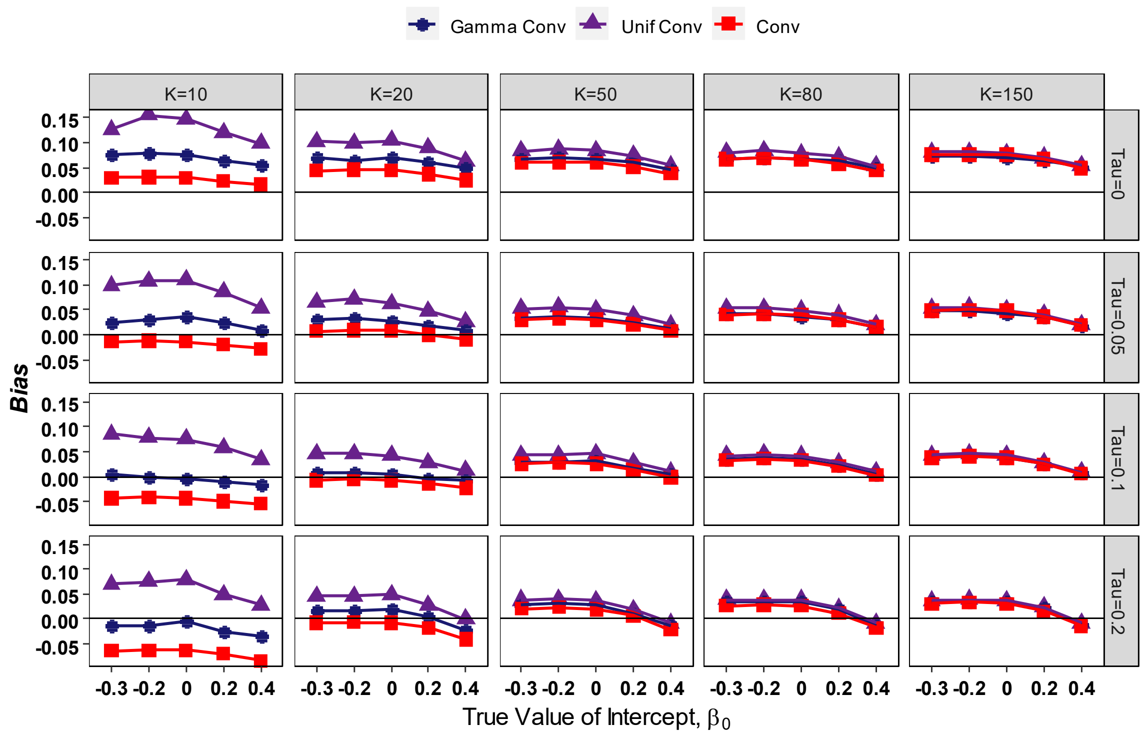


Figure 57

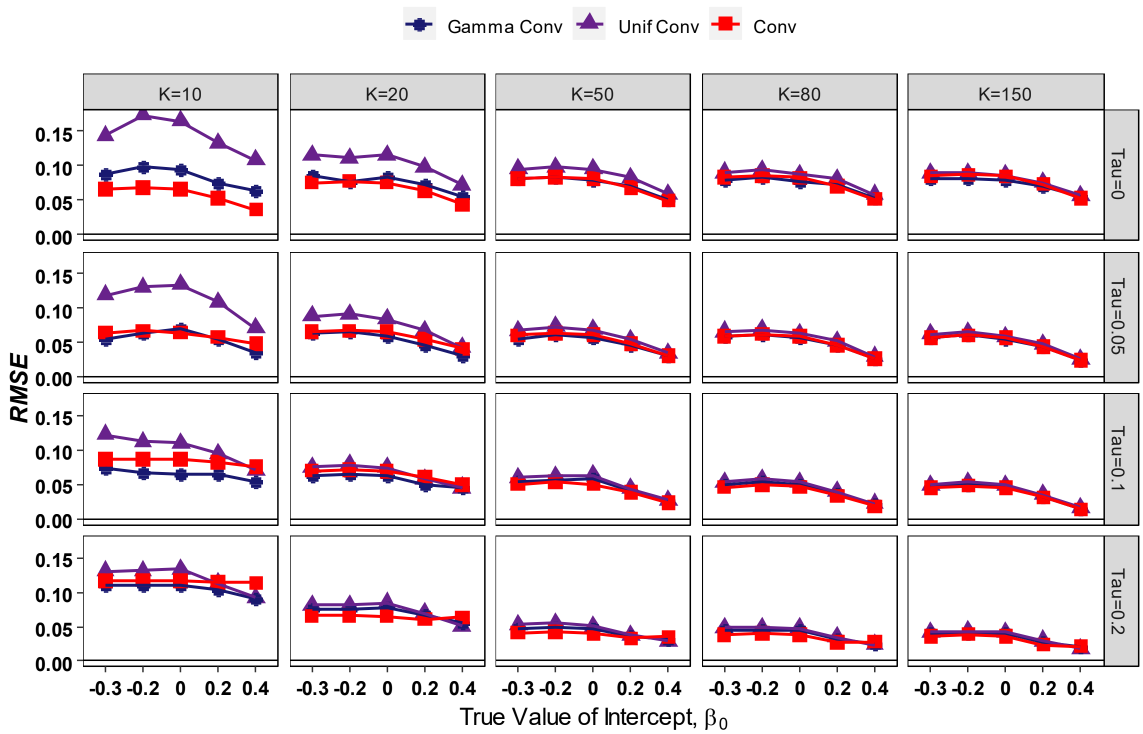
Bias for Tau for Correlation Effect Size: All Conventional Methods



Note. DirZtoR and IntZtoR are excluded from this plot.

Figure 58

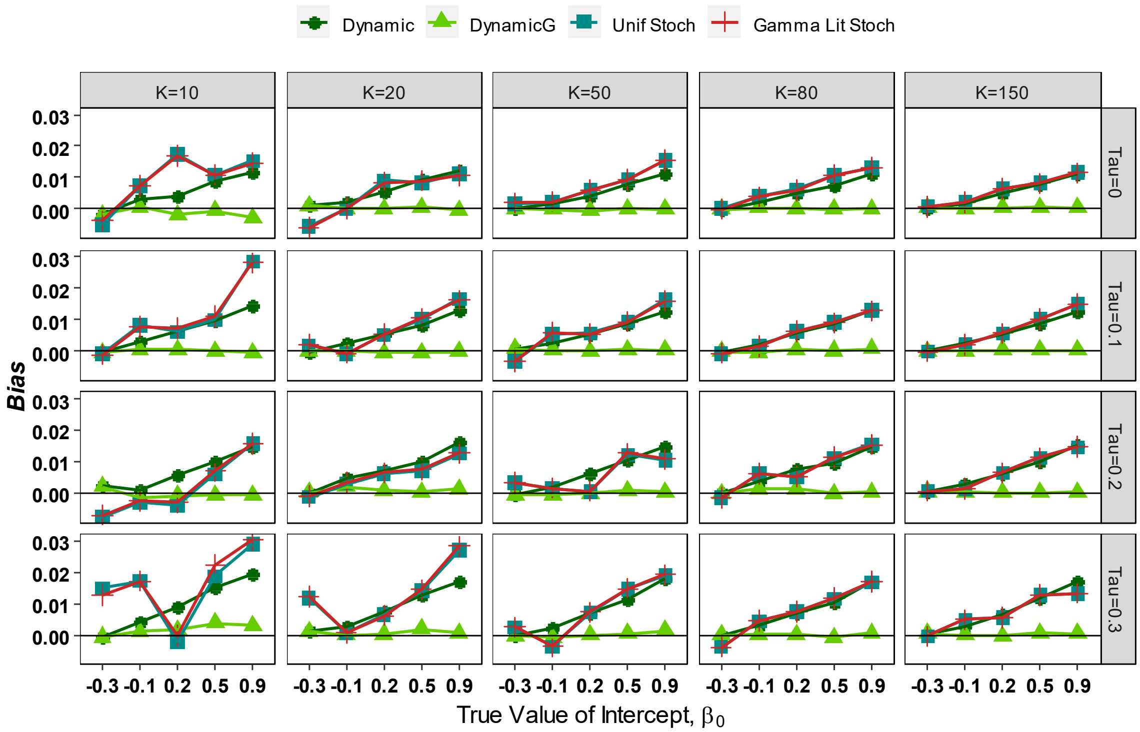
RMSE for the Tau for Correlation Effect Size: All Conventional Methods



Note. DirZtoR and IntZtoR excluded from this plot.

Figure 59

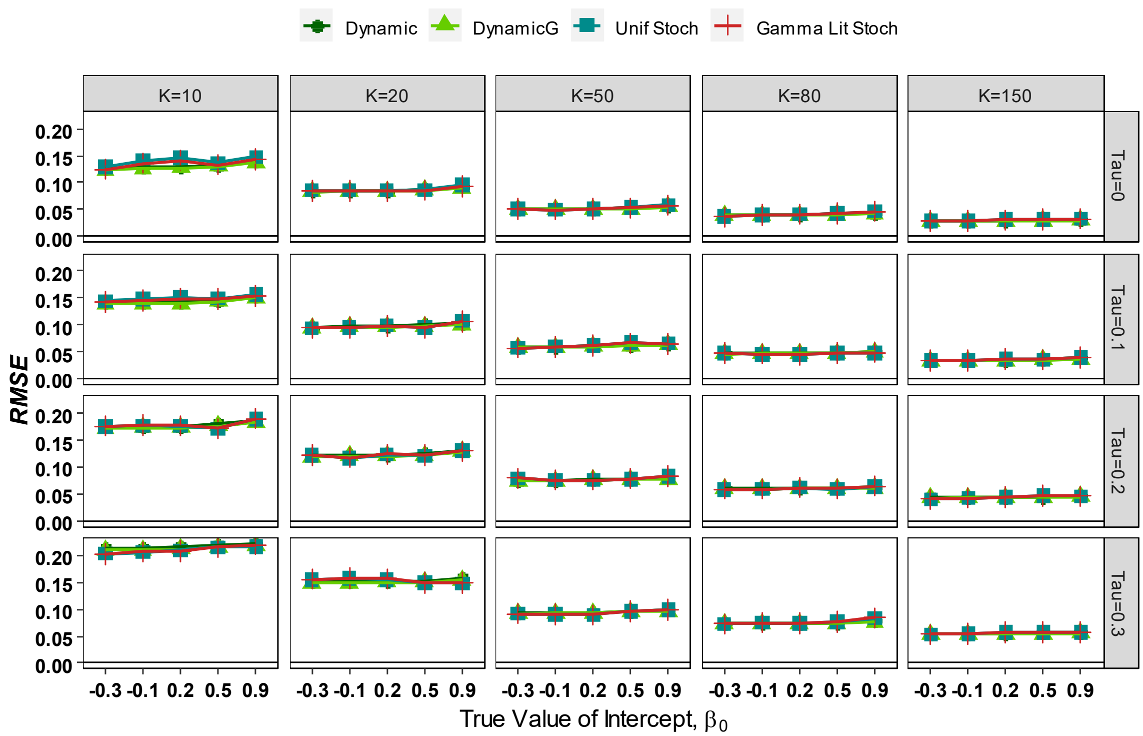
Bias for the Fourth Conditional Mean for Standardized Mean Difference Effect Size: All Dynamic Methods



Note. Plots for the second, third, and fourth conditional mean for all dynamic standardized mean difference methods are not included. Results for the other conditional means followed a similar pattern.

Figure 60

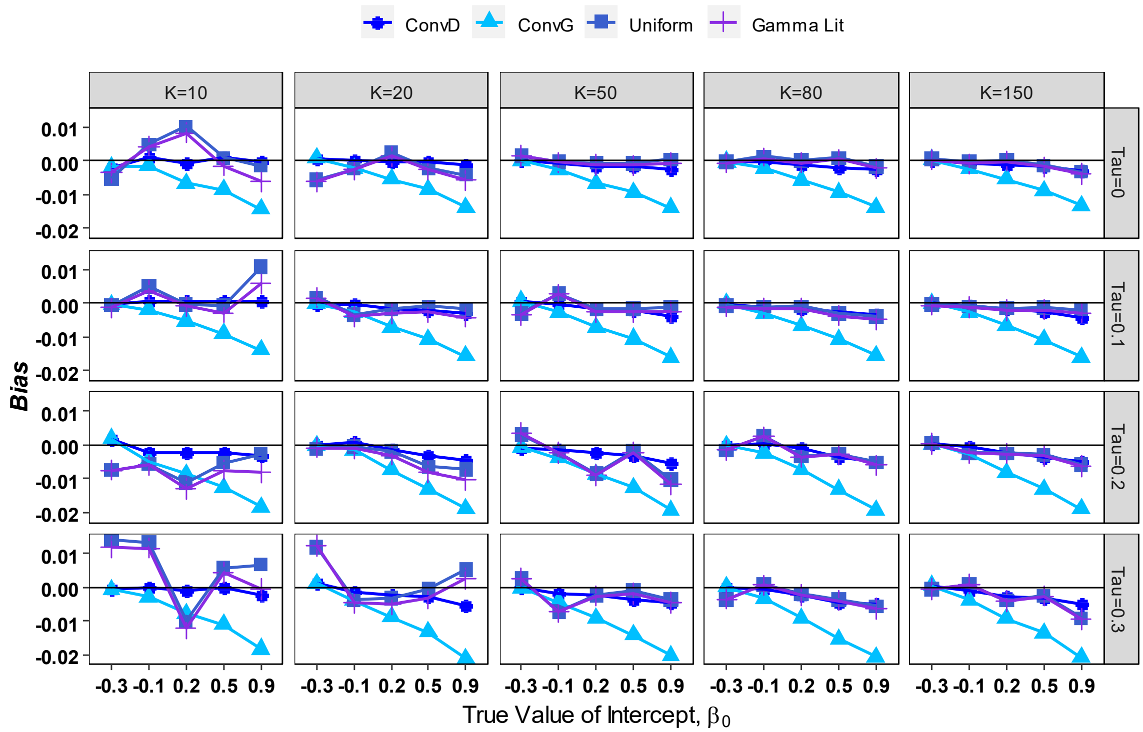
RMSE for the Fourth Conditional Mean for Standardized Mean Difference Effect Size: All Dynamic Methods



Note. Plots for the second, third, and fourth conditional mean for all dynamic standardized mean difference methods are not included. Results for the other conditional means followed a similar pattern.

Figure 61

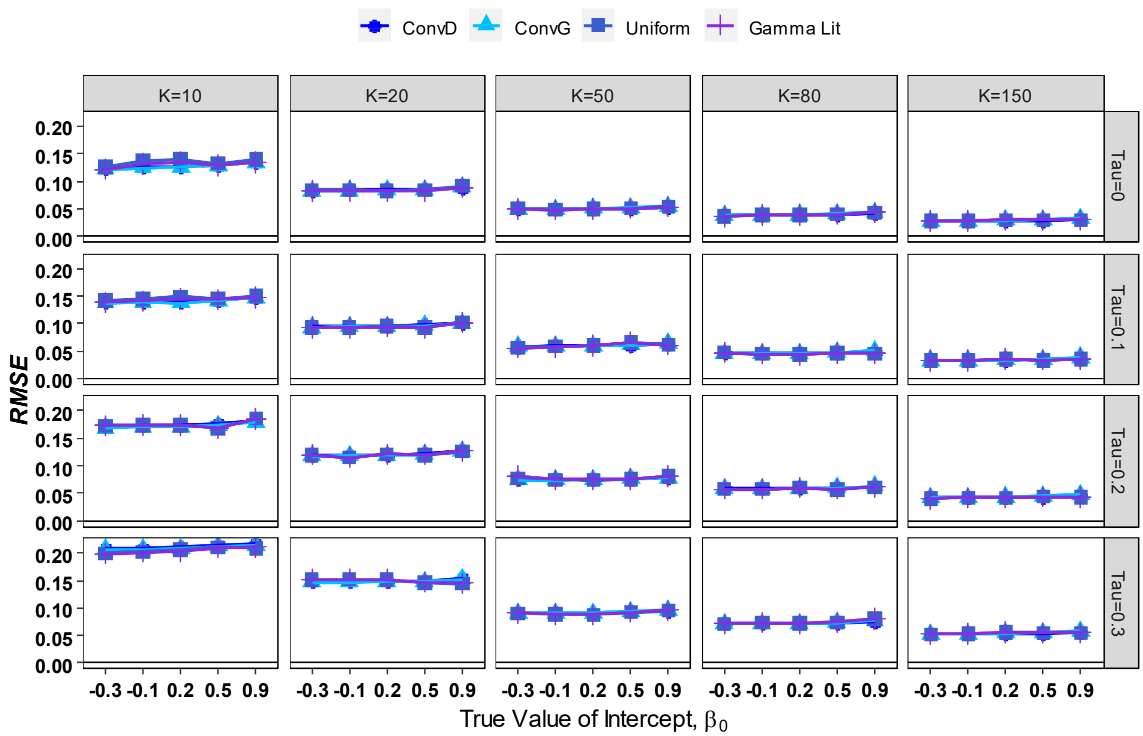
Bias for the Fourth Conditional Mean for Standardized Mean Difference Effect Size: All Conventional Methods



Note. Plots for the second, third, and fourth conditional mean for all conventional standardized mean difference methods are not included. Results for the other conditional means followed a similar pattern.

Figure 62

RMSE for the Fourth Conditional Mean for Standardized Mean Difference Effect Size: All Conventional Methods



Note. Plots for the second, third, and fourth conditional mean for all conventional standardized mean difference methods are not included. Results for the other conditional means followed a similar pattern.

Figure 63

Bias for Tau for Standardized Mean Difference Effect Size: All Dynamic Methods

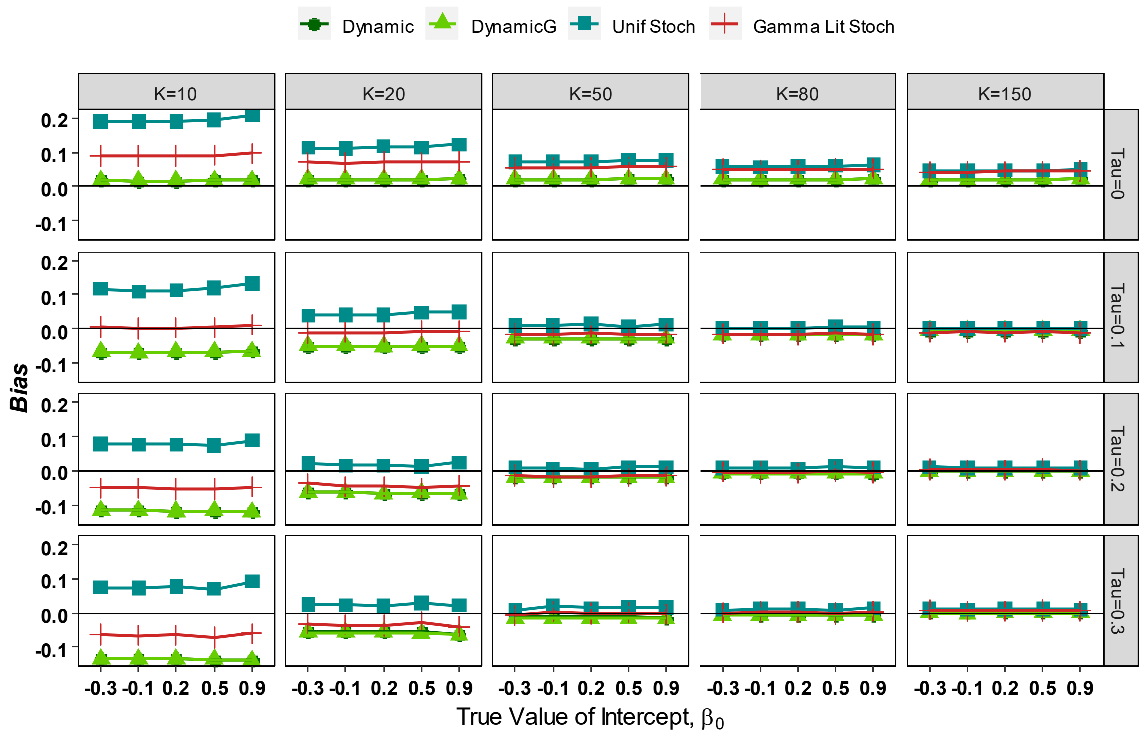


Figure 64

RMSE for the Tau for Standardized Mean Difference Effect Size: All Dynamic Methods

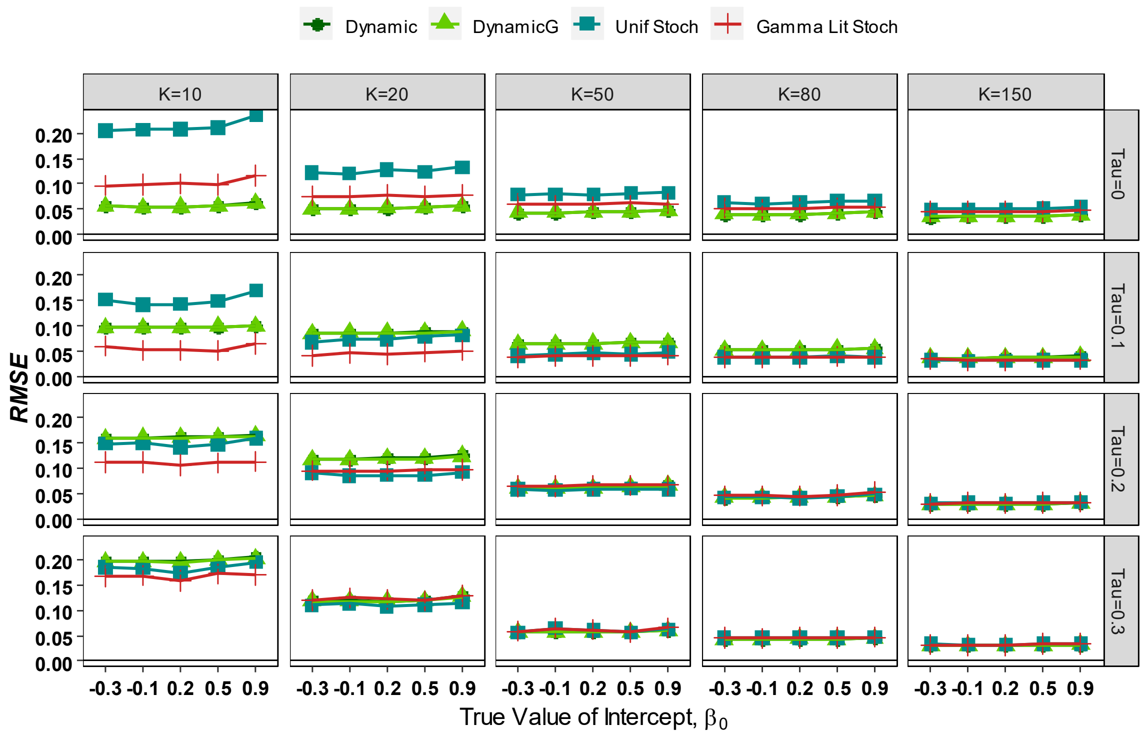


Figure 65

Bias for Tau for Standardized Mean Difference Effect Size: All Conventional Methods

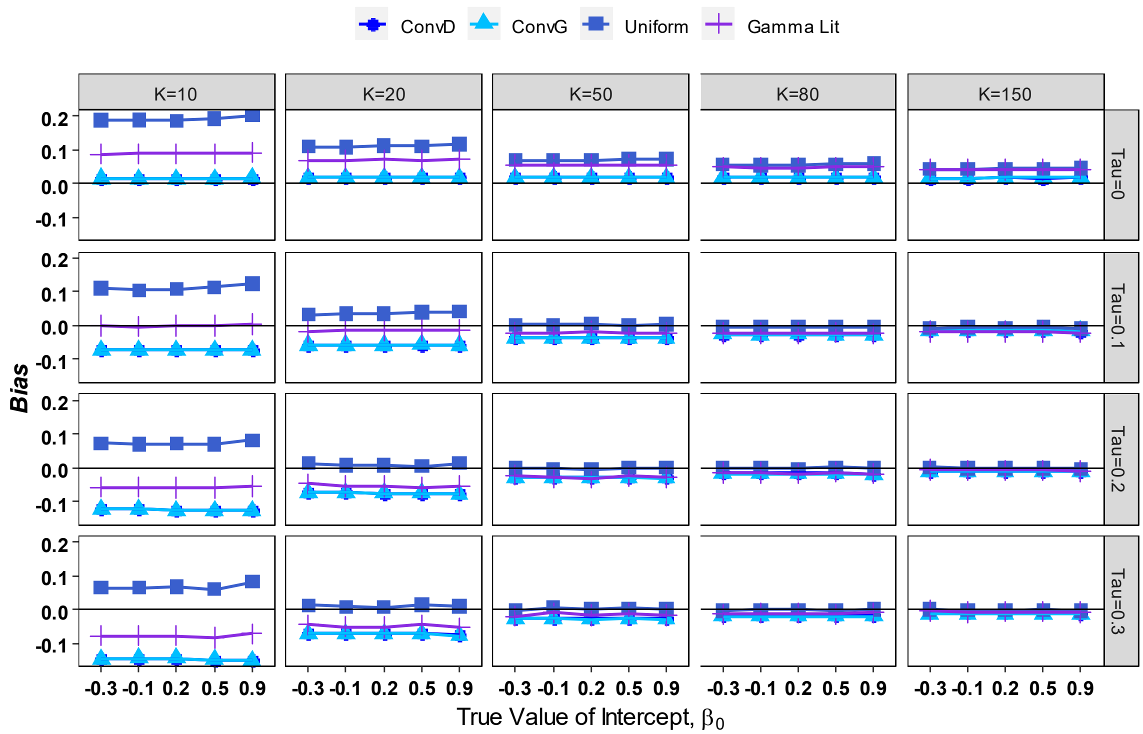
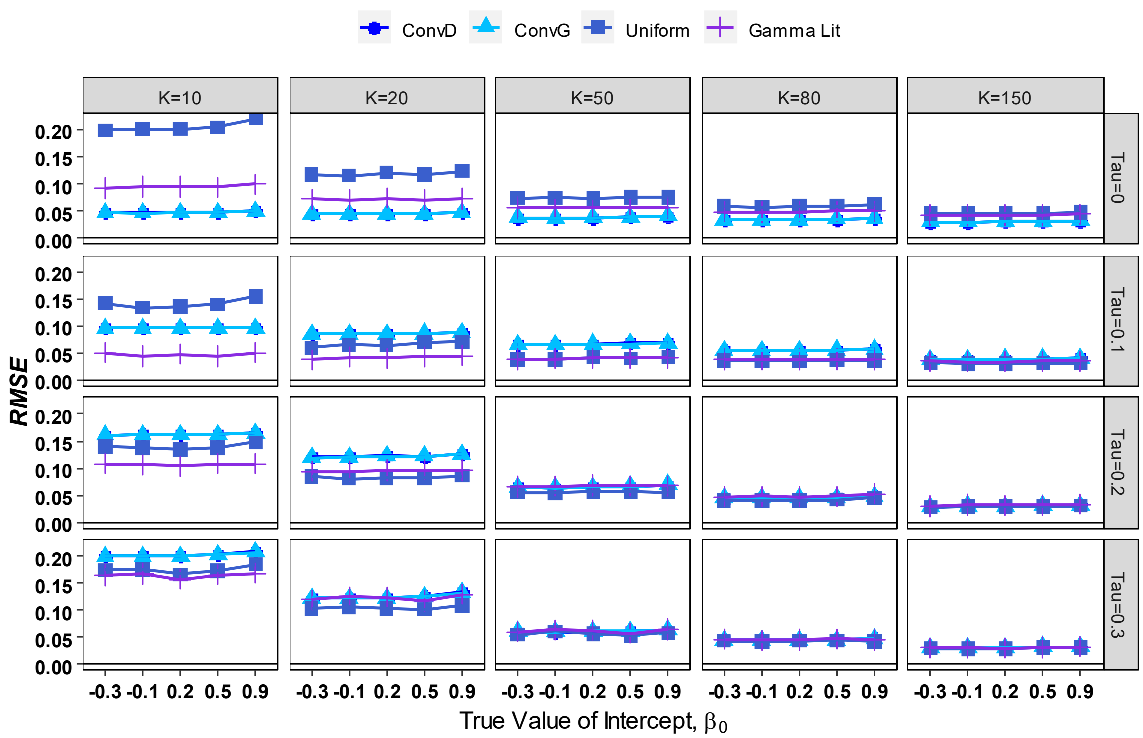


Figure 66

RMSE for the Tau for Standardized Mean Difference Effect Size: All Conventional Methods



Appendix 1
Function for Simulating Sample Sizes

```
function(number){

  retval = rep(0, number)
  proportions = c(0.027202073, 0.060880829, 0.036269430, 0.054404145,
0.059585492, 0.062176166,
                0.071243523, 0.038860104, 0.047927461, 0.034974093, 0.042746114,
0.031088083,
                0.034974093, 0.037564767, 0.041450777, 0.031088083, 0.036269430,
0.032383420,
                0.033678756, 0.034974093, 0.031088083, 0.032383420, 0.032383420,
0.029792746,
                0.015544041, 0.005181347, 0.003886010)
  proportions2<-c(0.02720207, 0.08808290, 0.12435233, 0.17875648,
0.23834197, 0.30051814,
                0.37176166, 0.41062176, 0.45854922,0.49352332, 0.53626943,
0.56735751,
                0.60233161, 0.63989637, 0.68134715, 0.71243523, 0.74870466,
0.78108808,
                0.81476684, 0.84974093, 0.88082901, 0.91321243, 0.94559585,
0.97538860,
                0.99093264, 0.99611399, 1.00000000)


  for(i in 1:number){
   testvar = runif(1)
   testvar2 = 0

   if(testvar <= proportions2[27]) testvar2= round(runif(1,500.5,636.4))
   if(testvar <= proportions2[26]) testvar2= round(runif(1,350.5,500.4))
   if(testvar <= proportions2[25]) testvar2= round(runif(1,250.5,350.4))
   if(testvar <= proportions2[24]) testvar2= round(runif(1,145.5,250.4))
   if(testvar <= proportions2[23]) testvar2= round(runif(1,100.5,145.4))
   if(testvar <= proportions2[22]) testvar2= round(runif(1,78.5,100.4))
   if(testvar <= proportions2[21]) testvar2= round(runif(1,70.5,78.4))
   if(testvar <= proportions2[20]) testvar2= round(runif(1,61.5,70.4))
   if(testvar <= proportions2[19]) testvar2= round(runif(1,55.5,61.4))
   if(testvar <= proportions2[18]) testvar2= round(runif(1,49.5,55.4))
   if(testvar <= proportions2[17]) testvar2= round(runif(1,44.5,49.4))
   if(testvar <= proportions2[16]) testvar2= round(runif(1,40.5,44.4))
   if(testvar <= proportions2[15]) testvar2= round(runif(1,38.5,40.4))
   if(testvar <= proportions2[14]) testvar2= round(runif(1,35.5,38.4))
   if(testvar <= proportions2[13]) testvar2= round(runif(1,32.5,35.4))
```

```
    if(testvar <= proportions2[12]) testvar2= round(runif(1,30.5,32.4))
    if(testvar <= proportions2[11]) testvar2= round(runif(1,28.5,30.4))
    if(testvar <= proportions2[10]) testvar2= round(runif(1,25.5,28.4))
    if(testvar <= proportions2[9]) testvar2= round(runif(1,23.5,25.4))
    if(testvar <= proportions2[8]) testvar2= round(runif(1,21.5,23.4))
    if(testvar <= proportions2[7]) testvar2= round(runif(1,19.5,21.4))
    if(testvar <= proportions2[6]) testvar2= round(runif(1,17.5,19.4))
    if(testvar <= proportions2[5]) testvar2= round(runif(1,15.5,17.4))
    if(testvar <= proportions2[4]) testvar2= round(runif(1,13.5,15.4))
    if(testvar <= proportions2[3]) testvar2= round(runif(1,11.5,13.4))
    if(testvar <= proportions2[2]) testvar2= round(runif(1,9.5,11.4))
    if(testvar <= proportions2[1]) testvar2 = round(runif(1,5.5,9.4))
    retval[i] <-testvar2}


  return(retval)
}
```

Appendix 2

Simulation of Effect Sizes for Correlation

```
n<-samplen(k) #simulating sample sizes of k studies with sampleN function
Desdat<-
matrix(c(rep(rep(1,10),kt),rep(rep(0,4),kt),rep(rep(1,6),kt),rep(rep(0,3),kt),
        rep(rep(1,4),kt),rep(rep(0,3),kt)),ncol=3) #The design matrix
mu<-X%*%Beta #The study effect size parameter.
#Beta is a vector including the intercept (B0), the regression term for the frirst
predictor (B1), and the regression term for the second predictor (B2),

r<-rep(NA,k) ) #placeholder for effect sizes

for(i in 1:k) {
  repeat{
    rho <-rnorm(1,mu[i],sqrt(tau2)) #effectsize parameter for study i
    if (rho<1 && rho>(-1)){
      break}
  }

  cormat <- matrix(c(1,rho,rho,1),2,2) #correlation matrix for parameter

  x <- cbind(rnorm(n[i]),rnorm(n[i]))
  x <- x%*%chol(cormat)  #simulated correlated raw data
  r[i] <- cor(x[,1],x[,2])

}

vr<-(1-r^2)^2 / (n-2) #estimated conditional variance
```

Appendix 3
Simulation of Effect Sizes for Standardized Mean difference

```
n1<-n2<-samplen(k) #simulating sample sizes of k studies with sampleN function
Desdat<-
matrix(c(rep(rep(1,10),kt),rep(rep(0,4),kt),rep(rep(1,6),kt),rep(rep(0,3),kt),
         rep(rep(1,4),kt),rep(rep(0,3),kt)),ncol=3) #The design matrix
mu<-X%*%Beta #The study effect size parameter.
#Beta is a vector including the intercept (B0), the regression term for the frirst
predictor (B1), and the regression term for the second predictor (B2),

d<-rep(NA,k) #placeholder for effect sizes
mui<-rnorm(k,mu,Tau)#Study effect size parameters
for (j in 1:k){

  x1<-rnorm(n[j],0,1)
  x2<-rnorm(n[j],mui[j],1)
  m1<-mean(x1)
  m2<-mean(x2)
  v1<-var(x1)
  v2<-var(x2)
  sp<-sqrt((v1+v2)/2)
  smd=(m2-m1)/sp
  d[j]<-smd
}


vd<-(n1+n2)/(n1*n2) + d^2/(2*(n1+n2)-2) #estimated conditional variance
```

Appendix 4
R Code for Integral z-to-r Transformation

```
#Lower Bound for Z
	LB<-Z-1.96*SEz
#Upper Bound for Z
		UB<-Z+1.96*SEz
#Lower Bound for ζ
	lower <- Z -5*sqrt(t2)
#Upper Bound for ζ
		upper <- Z + 5* sqrt(t2)
	#function for integration
	myf <- function(zest, mu, tau2){ x<-tanh(zest) * dnorm(zest, mean = mu,
sd = sqrt(tau2))
			return(x)
			}
	zint<-integrate(myf, lower = lower, upper = upper, mu = Z, tau2 =
t2)$value
```