Supplementary materials for *Estimation and inference for step-function selection models in meta-analysis with dependent effects*

Contents

A Score and Hessian of the step-function marginal log likelihood

From Equation (8), the log of the marginal likelihood contribution for effect size estimate i from study j is given by

$$l_{ij}^{M}\left(\boldsymbol{\beta},\gamma,\boldsymbol{\zeta}\right) \propto \log w\left(y_{ij},\sigma_{ij};\boldsymbol{\zeta}\right) - \frac{1}{2}\frac{\left(y_{ij}-\mathbf{x}_{ij}\boldsymbol{\beta}\right)^{2}}{\exp(\gamma)+\sigma_{ij}^{2}} - \frac{1}{2}\log\left(\exp(\gamma)+\sigma_{ij}^{2}\right) - \log A\left(\mathbf{x}_{ij},\sigma_{ij};\boldsymbol{\beta},\gamma,\boldsymbol{\zeta}\right). \tag{A1}$$

The components of the score contribution of study j given in Equations (11), (12), and (13) involve derivatives of $l_{ij}^{M}\left(\boldsymbol{\beta},\gamma,\boldsymbol{\zeta}\right)$ with respect to all model parameters. Let $w_{ij} = w\left(y_{ij},\sigma_{ij};\boldsymbol{\lambda}\right)$, $\mu_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta}$, $\eta_{ij} = \exp(\gamma)+\sigma_{ij}^{2}$, and $A_{ij} = A\left(\mathbf{x}_{ij},\sigma_{ij};\boldsymbol{\beta},\gamma,\boldsymbol{\zeta}\right)$ as defined in Equation (6). The score contribution of study j for the meta-regression coefficients has the general form

$$\mathbf{S}_{\beta j} = \sum_{i=1}^{k_j} a_{ij}\mathbf{x}_{ij}'\left(\frac{\left(y_{ij}-\mu_{ij}\right)}{\eta_{ij}} - \frac{1}{A_{ij}}\frac{\partial A_{ij}}{\partial \mu_{ij}}\right). \tag{A2}$$

The score for the variance parameter has the general form

$$S_{\gamma j} = \sum_{i=1}^{k_j} a_{ij}\tau^{2}\left(\frac{\left(y_{ij}-\mu_{ij}\right)^{2}-\eta_{ij}}{2\eta_{ij}^{2}} - \frac{1}{A_{ij}}\frac{\partial A_{ij}}{\partial \eta_{ij}}\right). \tag{A3}$$

The score for the h^{th} selection parameter has the general form

$$S_{\zeta_h j} = \sum_{i=1}^{k_j} a_{ij}\lambda_{h}\left(\frac{1}{w_{ij}}\frac{\partial w_{ij}}{\partial \lambda_{h}} - \frac{1}{A_{ij}}\frac{\partial A_{ij}}{\partial \lambda_{h}}\right), \tag{A4}$$

for $h = 1, ..., H$. For the step-function selection model, the derivatives of w_{ij} and A_{ij} are given by

$$\frac{\partial w_{ij}}{\partial \lambda_h} = I\left(\alpha_h < p_{ij} \leq \alpha_{h+1}\right) \tag{A5}$$

$$\frac{\partial A_{ij}}{\partial \mu_{ij}} = \frac{1}{\eta_{ij}^{1/2}} \sum_{h=0}^{H} \lambda_h \left[\phi\left(c_{h+1,ij}\right) - \phi\left(c_{hij}\right)\right] \tag{A6}$$

$$\frac{\partial A_{ij}}{\partial \eta_{ij}} = \frac{1}{2\eta_{ij}} \sum_{h=0}^{H} \lambda_h \left[c_{h+1,ij}\phi\left(c_{h+1,ij}\right) - c_{hij}\phi\left(c_{hij}\right)\right] \tag{A7}$$

$$\frac{\partial A_{ij}}{\partial \lambda_h} = B_{hij}, \tag{A8}$$

with c_{hij} and B_{hij} as defined in Equation (7).

The sub-components of the Hessian matrix from study j have the following forms:

$$\mathbf{H}_j^{\beta\beta'} = \frac{\partial}{\partial \boldsymbol{\beta}'}\mathbf{S}_{\beta j} = \sum_{i=1}^{k_j} a_{ij}\mathbf{x}'_{ij}\mathbf{x}_{ij}\left(\frac{1}{A_{ij}^2}\left(\frac{\partial A_{ij}}{\partial \mu_{ij}}\right)^2 - \frac{1}{A_{ij}}\frac{\partial^2 A_{ij}}{\partial \mu_{ij}^2} - \frac{1}{\eta_{ij}}\right) \tag{A9}$$

$$\mathbf{H}_j^{\beta\gamma} = \frac{\partial}{\partial \gamma}\mathbf{S}_{\beta j} = \sum_{i=1}^{k_j} a_{ij}\mathbf{x}'_{ij}\tau^2\left(\frac{1}{A_{ij}^2}\frac{\partial A_{ij}}{\partial \mu_{ij}}\frac{\partial A_{ij}}{\partial \eta_{ij}} - \frac{1}{A_{ij}}\frac{\partial^2 A_{ij}}{\partial \mu_{ij}\partial \eta_{ij}} - \frac{y_{ij}-\mu_{ij}}{\eta_{ij}^2}\right) \tag{A10}$$

$$\mathbf{H}_j^{\beta\zeta_h} = \frac{\partial}{\partial \zeta_h}\mathbf{S}_{\beta j} = \sum_{i=1}^{k_j} a_{ij}\mathbf{x}'_{ij}\lambda_h\left(\frac{1}{A_{ij}^2}\frac{\partial A_{ij}}{\partial \mu_{ij}}\frac{\partial A_{ij}}{\partial \lambda_h} - \frac{1}{A_{ij}}\frac{\partial^2 A_{ij}}{\partial \mu_{ij}\partial \lambda_h}\right) \tag{A11}$$

$$\begin{aligned}\mathbf{H}_j^{\gamma\gamma} = \frac{\partial}{\partial \gamma}S_{\gamma j} = \sum_{i=1}^{k_j} a_{ij}\Bigg[&\tau^2\left(\frac{\left(y_{ij}-\mu_{ij}\right)^2 - \eta_{ij}}{2\eta_{ij}^2} - \frac{1}{A_{ij}}\frac{\partial A_{ij}}{\partial \eta_{ij}}\right)\\ &+ \tau^4\left(\frac{1}{A_{ij}^2}\left(\frac{\partial A_{ij}}{\partial \eta_{ij}}\right)^2 - \frac{1}{A_{ij}}\frac{\partial^2 A_{ij}}{\partial \eta_{ij}^2} - \frac{2(y_{ij}-\mu_{ij})^2 - \eta_{ij}}{2\eta_{ij}^3}\right)\Bigg]\end{aligned} \tag{A12}$$

$$\mathbf{H}_j^{\gamma\zeta_h} = \frac{\partial}{\partial \zeta_h}S_{\gamma j} = \sum_{i=1}^{k_j} a_{ij}\tau^2\lambda_h\left(\frac{1}{A_{ij}^2}\frac{\partial A_{ij}}{\partial \eta_{ij}}\frac{\partial A_{ij}}{\partial \lambda_h} - \frac{1}{A_{ij}}\frac{\partial^2 A_{ij}}{\partial \eta_{ij}\partial \lambda_h}\right) \tag{A13}$$

$$\begin{aligned}\mathbf{H}_j^{\zeta_f\zeta_h} = \frac{\partial}{\partial \zeta_h}S_{\zeta_f j} = \sum_{i=1}^{k_j} a_{ij}\Bigg[&\lambda_f\lambda_h\left(\frac{1}{A_{ij}^2}\frac{\partial A_{ij}}{\partial \lambda_f}\frac{\partial A_{ij}}{\partial \lambda_h} - \frac{1}{w_{ij}^2}\frac{\partial w_{ij}}{\partial \lambda_f}\frac{\partial w_{ij}}{\partial \lambda_h}\right)\\ &+ I(f=h)\lambda_h\left(\frac{1}{w_{ij}}\frac{\partial w_{ij}}{\partial \lambda_h} - \frac{1}{A_{ij}}\frac{\partial A_{ij}}{\partial \lambda_h}\right)\Bigg],\end{aligned} \tag{A14}$$

where the second partial derivatives of A_{ij} are given by

$$\frac{\partial^2 A_{ij}}{\partial \mu_{ij}^2} = \frac{1}{\eta_{ij}} \sum_{h=0}^{H} \lambda_h \left[c_{h+1,ij} \phi(c_{h+1,ij}) - c_{hij} \phi(c_{hij}) \right] \tag{A15}$$

$$\frac{\partial^2 A_{ij}}{\partial \mu_{ij} \partial \eta_{ij}} = \frac{1}{2\eta_{ij}^{3/2}} \sum_{h=0}^{H} \lambda_h \left[(c_{h+1,ij}^2 - 1)\phi(c_{h+1,ij}) - (c_{hij}^2 - 1)\phi(c_{hij}) \right] \tag{A16}$$

$$\frac{\partial^2 A_{ij}}{\partial \mu_{ij} \partial \lambda_h} = \frac{\phi\left(c_{h+1,ij}\right) - \phi\left(c_{hij}\right)}{\eta_{ij}^{1/2}} \tag{A17}$$

$$\frac{\partial^2 A_{ij}}{\partial \eta_{ij}^2} = \frac{1}{4\eta_{ij}^2} \sum_{h=0}^{H} \lambda_h \left[(c_{h+1,ij}^3 - 3c_{h+1,ij})\phi(c_{h+1,ij}) - (c_{hij}^3 - 3c_{hij})\phi(c_{hij}) \right] \tag{A18}$$

$$\frac{\partial^2 A_{ij}}{\partial \eta_{ij} \partial \lambda_h} = \frac{c_{h+1,ij}\phi\left(c_{h+1,ij}\right) - c_{hij}\phi\left(c_{hij}\right)}{2\eta_{hij}}. \tag{A19}$$

B Sandwich estimator for the augmented, re-weighted Gaussian likelihood estimator

The augmented, re-weighted Gaussian likelihood (ARGL) estimators are defined as the solution to the estimating equations given in Equation (21) and (22). Sandwich variance approximations for the ARGL estimators involve the estimating equations and their Jacobian.

Let $\mathbf{J}$ denote the Jacobian matrix of the ARGL estimating equations, given by

$$\mathbf{J}(\boldsymbol{\beta}, \gamma, \boldsymbol{\zeta}) = \sum_{j=1}^{J} \frac{\partial \mathbf{M}_j(\boldsymbol{\beta}, \gamma, \boldsymbol{\zeta})}{\partial(\boldsymbol{\beta}' \gamma \boldsymbol{\zeta}')}. \tag{B1}$$

The exact form of the Jacobian is described below. Let $\tilde{\mathbf{M}}_j = \mathbf{M}_j(\tilde{\boldsymbol{\beta}}, \tilde{\gamma}, \tilde{\boldsymbol{\zeta}})$ and $\tilde{\mathbf{J}} = \mathbf{J}(\tilde{\boldsymbol{\beta}}, \tilde{\gamma}, \tilde{\boldsymbol{\zeta}})$ denote the score vectors and Jacobian matrix evaluated at the solutions to the ARGL estimating equations. A cluster-robust sandwich estimator is then given by:

$$\mathbf{V}^{ARGL} = \tilde{\mathbf{J}}^{-1} \left(\sum_{j=1}^{J} \tilde{\mathbf{M}}_j \tilde{\mathbf{M}}_j' \right) \left(\tilde{\mathbf{J}}^{-1} \right)'. \tag{B2}$$

The Jacobian of (22) is given by

$$\mathbf{J} = \sum_{j=1}^{J} \begin{bmatrix} \mathbf{J}_j^{\beta\beta'} & \mathbf{J}_j^{\beta\gamma'} & \mathbf{J}_j^{\beta\zeta'} \\ \mathbf{J}_j^{\gamma\beta'} & \mathbf{J}_j^{\gamma\gamma'} & \mathbf{J}_j^{\gamma\zeta'} \\ \mathbf{J}_j^{\zeta\beta'} & \mathbf{J}_j^{\zeta\gamma'} & \mathbf{J}_j^{\zeta\zeta'}, \end{bmatrix}$$

where

$$
\begin{aligned}
\mathbf{J}_j^{\beta\beta'} &= -\sum_{i=1}^{k_j} a_{ij} \frac{\mathbf{x}'_{ij}\mathbf{x}_{ij}}{w_{ij}\eta_{ij}} \\
\mathbf{J}_j^{\beta\gamma} &= -\sum_{i=1}^{k_j} a_{ij} \mathbf{x}'_{ij} \tau^2 \left(\frac{y_{ij} - \mu_{ij}}{w_{ij}\eta_{ij}^2} \right) \\
\mathbf{J}_j^{\beta\zeta_h} &= -\sum_{i=1}^{k_j} a_{ij} \mathbf{x}'_{ij} \lambda_h \left(\frac{y_{ij} - \mu_{ij}}{w_{ij}^2\eta_{ij}} \right) \\
\mathbf{J}_j^{\gamma\beta'} &= \left(\mathbf{J}_j^{\beta\gamma} \right)' \\
\mathbf{J}_j^{\gamma\gamma} &= -\sum_{i=1}^{k_j} a_{ij} \left[\frac{\tau^4}{w_{ij}} \left(\frac{2(y_{ij} - \mu_{ij})^2 - \eta_{ij}}{2\eta_{ij}^3} \right) - \frac{\tau^2}{w_{ij}} \left(\frac{(y_{ij} - \mu_{ij})^2 - \eta}{2\eta_{ij}^2} \right) \right] \\
\mathbf{J}_j^{\gamma\zeta_h} &= -\sum_{i=1}^{k_j} a_{ij} \frac{\tau^2 \lambda_h}{w_{ij}^2} \left(\frac{(y_{ij} - \mu_{ij})^2 - \eta_{ij}}{2\eta_{ij}^2} \right) \\
\mathbf{J}_j^{\zeta_h\beta'} &= \left(\mathbf{H}_j^{\beta\zeta_h} \right)' \\
\mathbf{J}_j^{\zeta_h\gamma} &= \mathbf{H}_j^{\gamma\zeta_h} \\
\mathbf{J}_j^{\zeta_f\zeta_h} &= \mathbf{H}_j^{\zeta_f\zeta'_h},
\end{aligned}
$$

with $\mathbf{H}_j^{\beta\zeta_h}$, $\mathbf{H}_j^{\gamma\zeta_h}$, and $\mathbf{H}_j^{\zeta_f\zeta_h}$ as given in Equations (A11), (A13), and (A14).

C Further details about bootstrapping

C.1 Other bootstrap sampling methods

The main manuscript described a two-stage cluster bootstrapping technique. We also considered two other bootstrap sampling schemes.

In the clustered bootstrap scheme, each pseudo-sample is generated by randomly drawing J clusters of observations with replacement from the original sample. In contrast to the two-stage cluster bootstrap, sampled clusters are left intact rather than perturbed. Because clusters are drawn with replacement, some will necessarily be included multiple times and some clusters might not be included in a given pseudo-sample. Following the notation in the main text, let $a_j^{(b)}$ be a first-stage weight for cluster j and $a_{ij}^{(b)}$ be the weight assigned to observation i in cluster j for pseudo-sample b. The clustered bootstrap process is then equivalent to drawing $a_1^{(b)}, ..., a_J^{(b)}$ from a multinomial distribution with J trials and equal probability on each of J categories and then setting $a_{ij}^{(b)} = a_j^{(b)}$ for $i = 1, ..., k_j$ and $j = 1, ..., J$.

The fractional random weight bootstrap follows a very similar process in which each pseudo-sample is generated by assigning a random weight to every cluster. Rather than using a multinomial distribution, the weights follow independent exponential distributions with mean 1, so that the weights for pseudo-sample b are generated as $a_j^{(b)} \sim Exp(1)$, with $a_{ij}^{(b)} = a_j^{(b)}$ for $i = 1, ..., k_j$ and $j = 1, ..., J$. A crucial difference between these bootstrap techniques is that the fractional random weight bootstrap puts strictly positive weight on every cluster of observations in every pseudo-sample, whereas the clustered bootstrap and two-stage bootstrap can assign zero weight to some clusters[1]. If the existence of an estimator hinges on the inclusion of one or a small number of clusters, this will create computational problems for the non-parametric bootstrap but not for the fractional random weight bootstrap.

C.2 Bootstrap confidence interval construction

We describe three methods for constructing a $1-2\alpha$ confidence interval (CI) from a set of B bootstrap replications. Consider a parameter θ that is a scalar component of $\boldsymbol{\beta}$, γ, or $\boldsymbol{\zeta}$. Let $\hat{\theta}$ denote an estimator of θ with sandwich variance estimator V. Let $\hat{\theta}_b^*$ denote the same estimator computed from bootstrap pseudo-sample b, with corresponding sandwich variance estimator V_b^*. Let $\hat{\theta}_{(1)}^*, ..., \hat{\theta}_{(B)}^*$ denote the pseudo-sample estimators sorted in ascending order. An estimator for the standard error of $\hat{\theta}$ can be computed by taking the standard deviation of the $\hat{\theta}_1^*, ..., \hat{\theta}_B^*$.

First, the percentile CI is calculated by taking the α and $1-\alpha$ quantiles of the bootstrap distribution, with end-points

$$\left[\hat{\theta}_{((B+1)\alpha)}^*, \hat{\theta}_{((B+1)(1-\alpha))}^*\right].$$

Second, the so-called "basic" CI pivots the bootstrap distribution around the original estimator $\hat{\theta}$. Its end-points are given by

$$\left[2\hat{\theta} - \hat{\theta}_{((B+1)(1-\alpha)}^*, 2\hat{\theta} - \hat{\theta}_{((B+1)\alpha)}^*\right].$$

Third, a studentized CI uses the bootstrap distribution of t-statistics rather than point estimators. The t statistic for pseudo-sample b is computed as $t_b^* = \left(\hat{\theta}_b^* - \hat{\theta}\right) / \sqrt{V_b^*}$. The studentized CI is computed using the percentiles of the bootstrap distribution of $t_1^*, ..., t_B^*$, taking

$$\left[\hat{\theta} - \sqrt{V} \times t_{((B+1)(1-\alpha)}^*,\ \hat{\theta} - \sqrt{V} \times t_{((B+1)\alpha)}^*\right].$$

Fourth, the bias-corrected-and-accelerated (BCa) CI is similar to the percentile CI in that its end-points are defined by quantiles of the bootstrap distribution. However, instead using the α and $1-\alpha$ quantiles, it uses quantiles that are adjusted to take into account the bias of the estimator and the degree to which its sampling variance is related to the underlying

parameter, as measured using an acceleration coefficient. These adjustments are defined in terms of the empirical influence function, which we approximate using a leave-one-cluster-out jackknife. The jackknife influence value for cluster j is $\hat{\theta} - \hat{\theta}^{+}_{-j}$, where $\hat{\theta}^{+}_{-j}$ denotes the estimator of θ computed while leaving out the observations in cluster j for $j = 1, ..., J$. The acceleration coefficient is then

$$\hat{a} = \frac{\sum_{j=1}^{J} \left(\hat{\theta} - \hat{\theta}^{+}_{-j}\right)^3}{6\left[\sum_{j=1}^{J} \left(\hat{\theta} - \hat{\theta}^{+}_{-j}\right)^2\right]^{3/2}}.$$

The bias coefficient is calculated as the proportion of the bootstrap distribution that falls below the original estimator:

$$\hat{\beta} = \frac{1}{B}\sum_{b=1}^{B} I\left(\hat{\theta}^{*}_{b} < \hat{\theta}\right).$$

With the acceleration and bias coefficients defined, define the adjustment function $f(\alpha)$ as

$$f(\alpha) = \Phi\left(\Phi^{-1}(\hat{\beta}) + \frac{\Phi^{-1}(\hat{\beta}) + \Phi^{-1}(\alpha)}{1 - \hat{a}\left[\Phi^{-1}(\hat{\beta}) + \Phi^{-1}(\alpha)\right]}\right)$$

for $0 < \alpha < 1$. The end-points of the BCa CI are then given by

$$\left[\hat{\theta}^{*}_{((B+1)\times f(\alpha))}, \hat{\theta}^{*}_{((B+1)\times f(1-\alpha))}\right].$$

Notably, the basic and studentized confidence intervals depend on the scale of the parameter θ, and the accuracy of their coverage levels therefore depends on the parameterization. In contrast, the percentile and bias-corrected-and-accelerated confidence intervals are invariant to transformation of θ.

C.3 Extrapolating coverage rates

The simulations reported in the main manuscript involved using an extrapolation technique suggested by Boos and Zhang[2] to estimate the coverage levels of bootstrap confidence intervals for $B = 1999$ bootstrap re-samples, when each replication of the

simulation involved only $B = 399$ re-samples. We used $R = 2000$ replications of the simulation process. For each replication, we computed an estimator $\hat{\theta}$ of parameter θ and a set of bootstrap re-samples $\hat{\theta}_b^*$ for $b = 1, ..., 399$. With these data, we can construct a confidence interval for θ by drawing $d \leq B$ of the bootstrap re-samples at random without replacement, then using these replications in the calculations described in Section C.2. We repeat this process for $d = 49, 99, 199, 299, 399$. Let C^{dr} be an indicator for whether the confidence interval constructed from d re-samples covers θ, for $r = 1, ..., R$. To extrapolate coverage, we fit a linear regression

$$C_{dr} = \alpha + \beta \left(\frac{1}{d} - \frac{1}{1999} \right) + e_{dr}$$

by ordinary least squares, with standard errors clustered by replication r to account for dependence in C_{dr} computed from the same replication of the simulation process. We take the estimate of α as the coverage rate of confidence intervals constructed from $B = 1999$ re-samples.

To validate this extrapolation technique, we ran additional simulations where we generated confidence intervals based on $B = 1999$ re-samples and computed coverage rates directly. Due to high computational demand, we conducted this exercise for a small subset of the conditions examined in the full simulation study, which we selected because they varied in the estimate coverage rates of percentile confidence intervals. Specifically, we looked conditions with $J = 15$ studies, an average effect size of $\mu = 0.2$, typical primary study sample sizes, a heterogeneity ratio of $\omega^2/\tau_B^2 = 0.5$, outcome correlation $\rho = 0.8$, between-study heterogeneity of $\tau_B \in \{0.05, 0.45\}$, and probability of selection $\lambda_1 \in \{0.05, 0.20\}$. We used the two-stage bootstrap re-sampling process because it showed the best performance in the larger simulation study.

Figures C1 and C2 illustrate the extrapolation results for the composite maximum

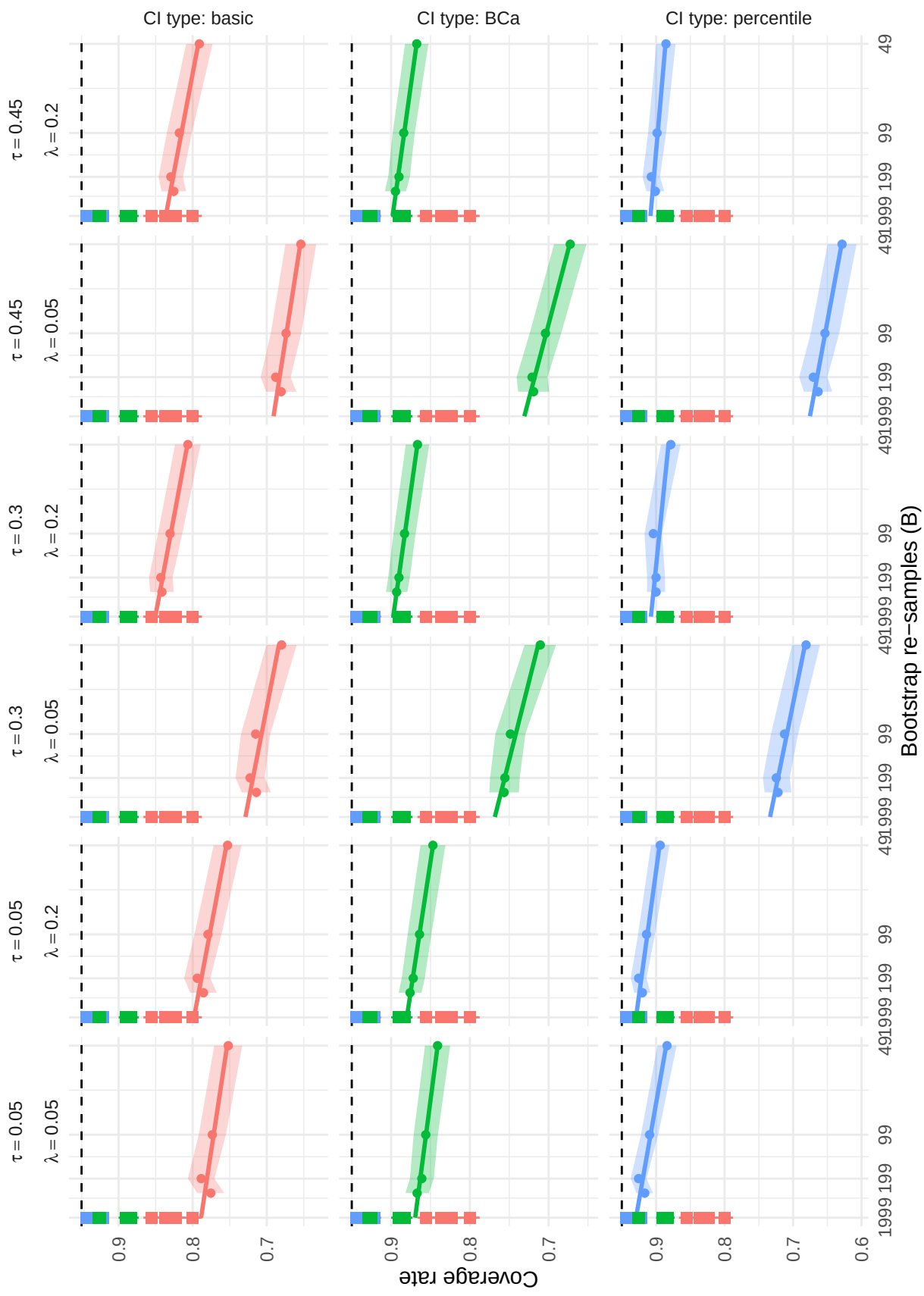

Figure C1
Actual and extrapolated coverage rates of selected bootstrap confidence intervals based on composite maximum likelihood estimation, using two-stage bootstrap resampling. Circular dots represent actual coverage rates used to extrapolate coverage to $B = 1999$ re-samples, with colored bands indicating point-wise Monte Carlo intervals. Squares represent actual coverage rates using $B = 1999$ re-samples, with whiskers indicating point-wise Monte carlo intervals, based on 4000 replications.

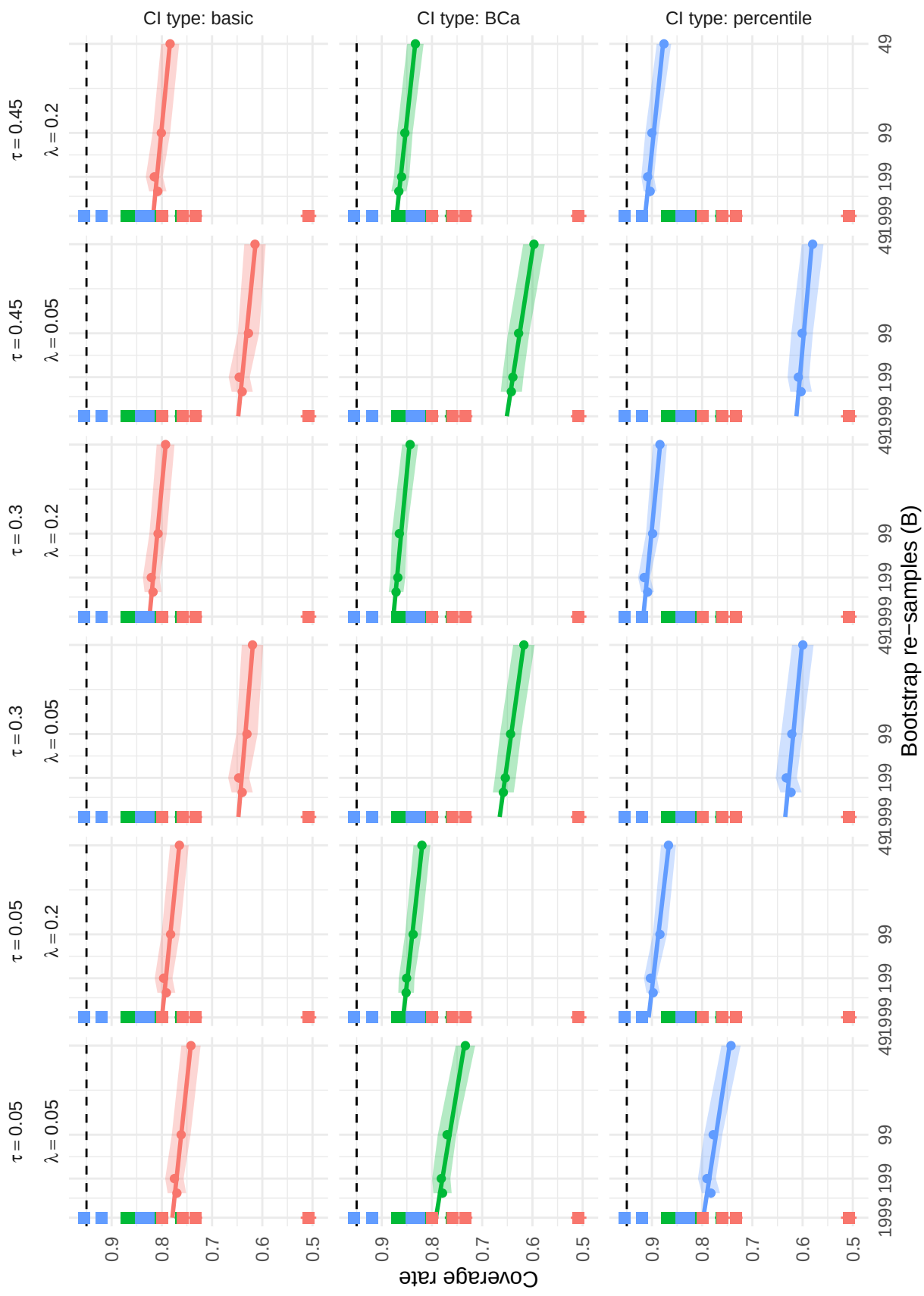

Figure C2
Actual and extrapolated coverage rates of selected bootstrap confidence intervals based on augmented reweighted Gaussian likelihood estimation, using two-stage bootstrap resampling. Circular dots represent actual coverage rates used to extrapolate coverage to $B = 1999$ re-samples, with colored bands indicating point-wise Monte Carlo intervals. Squares represent actual coverage rates using $B = 1999$ re-samples, with whiskers indicating point-wise Monte carlo intervals, based on 4000 replications.

likelihood and augmented, reweighted Gaussian likelihood estimators, respectively. In each figure, the circular points show the estimated coverage rates of a specific type of bootstrap confidence interval computed from $d = 49, 99, 199, 299$, or 399 bootstrap re-samples, with bands representing point-wise Monte Carlo intervals; These data are drawn from the larger simulation study and use $R = 2000$ replications. The horizontal axis of each panel is scaled in proportion to $1/d$, and the straight line in each panel depicts the linear regression fit used for extrapolation. On the left edge of each panel, a square point and vertical whiskers indicate the coverage rates computed from the additional simulations with $B = 1999$ re-samples and $R = 4000$ replications of the simulation process. Comparing the square points to the linear extrapolations provides a sense of the accuracy of the extrapolation technique.

It can be seen that nearly all coverage rates from the additional simulations fall within the Monte Carlo intervals of the extrapolation. Across conditions, estimators, and confidence interval types, discrepancies between extrapolated coverage and directly estimated coverage ranged from -0.1 to 0.3 percentage points, with an average of 0.0 and root mean-squared error of 0.1 percentage points. Thus, the extrapolations computed in the large simulation study appear to accurately predict the coverage rates computed from the additional simulations with $B = 1999$ re-samples.

D Additional simulation results for estimators of average effect size (μ)

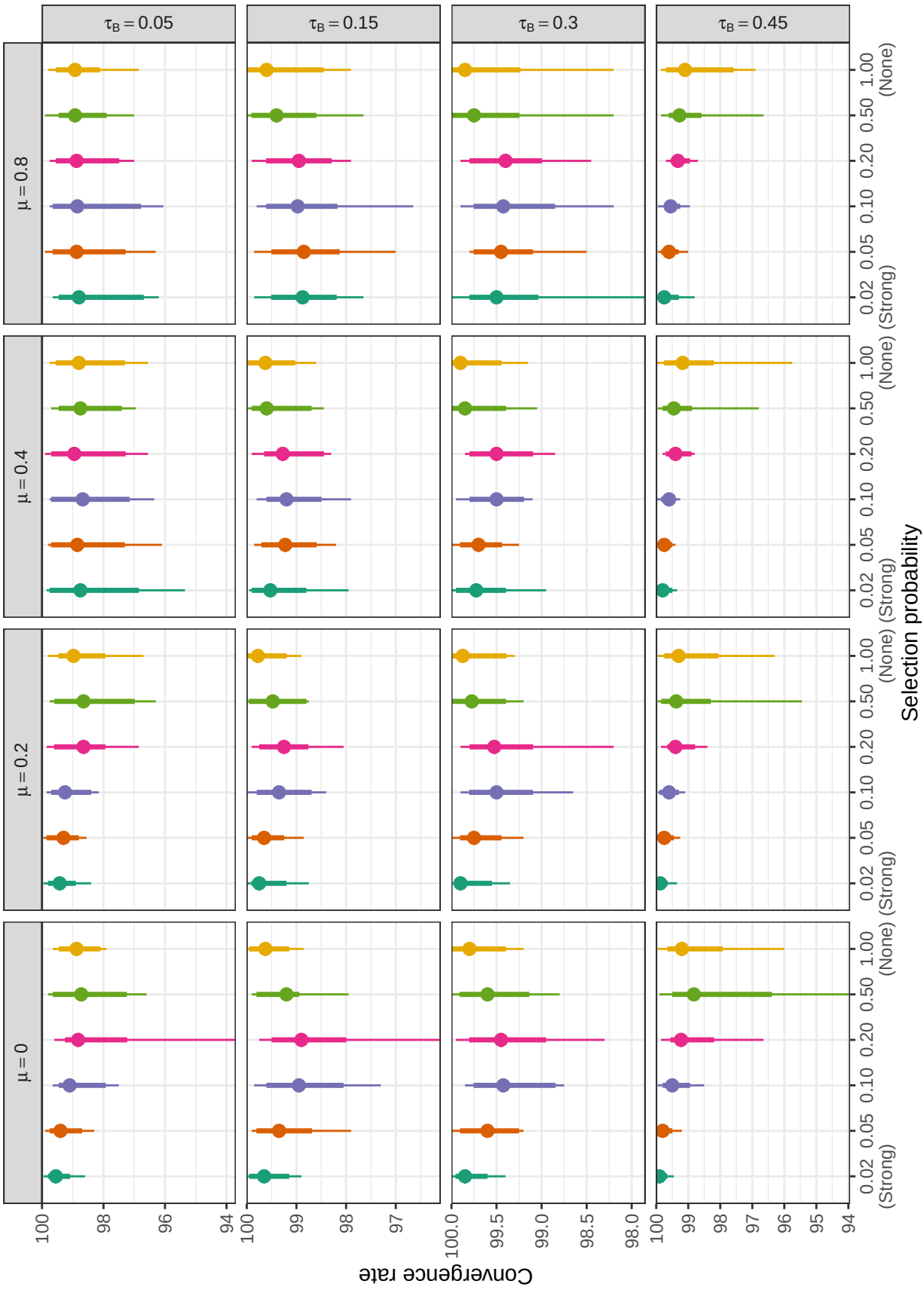

Figure D1

Convergence rates of CML estimators by selection probability, average SMD, and between-study heterogeneity. Points correspond to median convergence rates; thin lines correspond to range of convergence rates; thick lines correspond to inter-decile range.

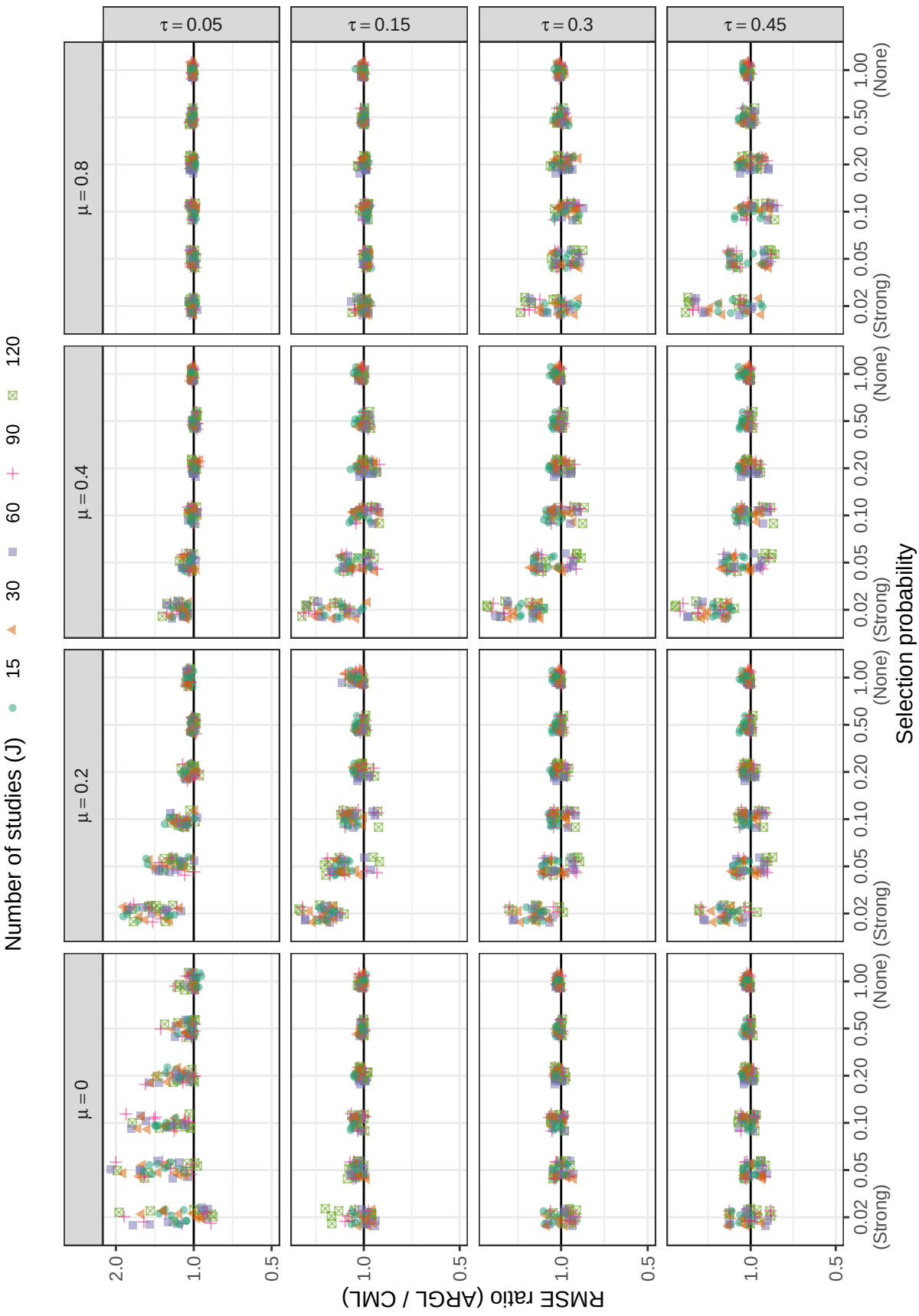

Figure D2
Ratio of root mean-squared error for ARGL estimator to root mean-squared error of CML estimator by selection probability, number of studies, average SMD, and between-study heterogeneity

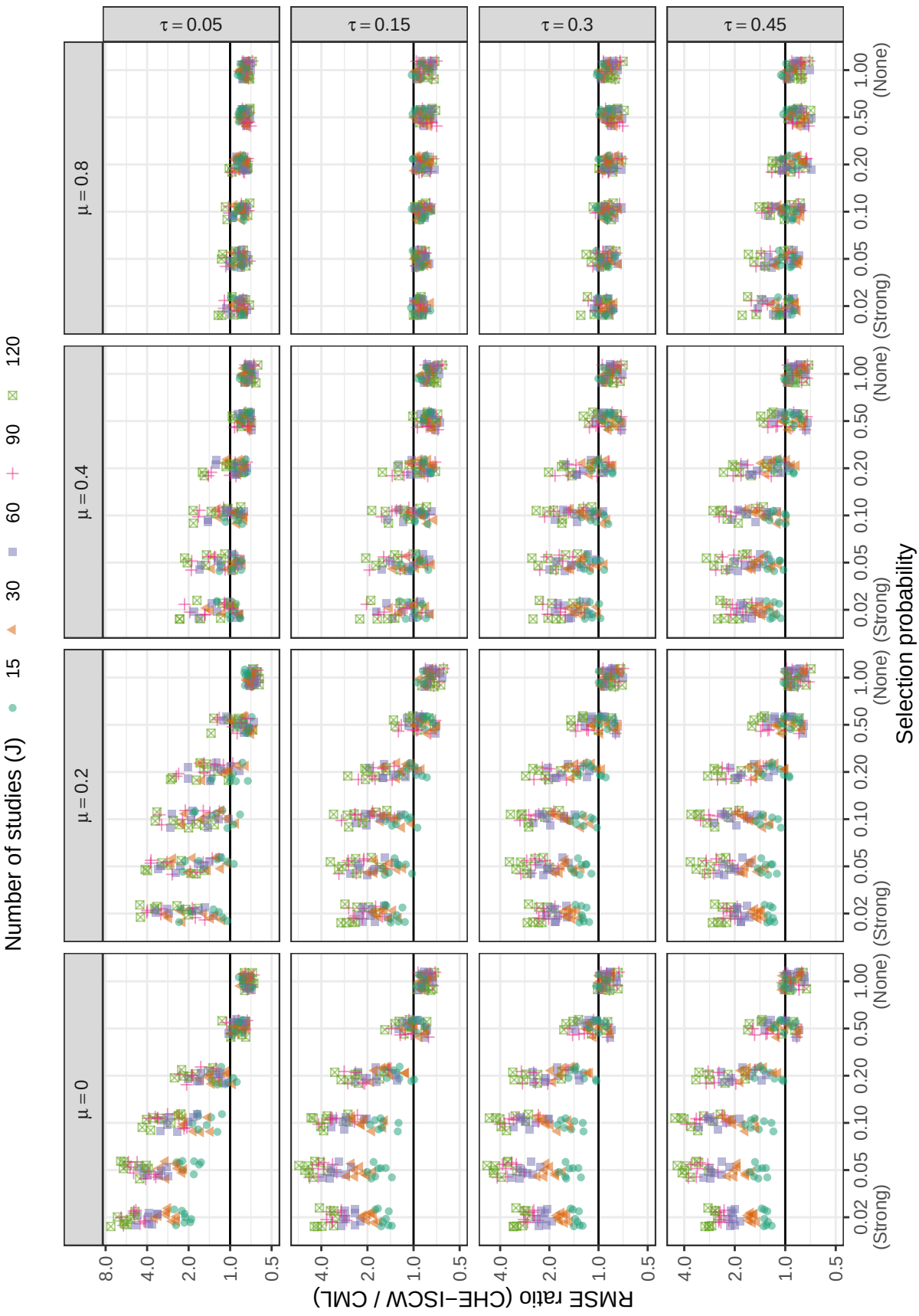

Figure D3
Ratio of root mean-squared error for CHE-ISCW estimator to root mean-squared error of CML estimator by selection probability, number of studies, average SMD, and between-study heterogeneity

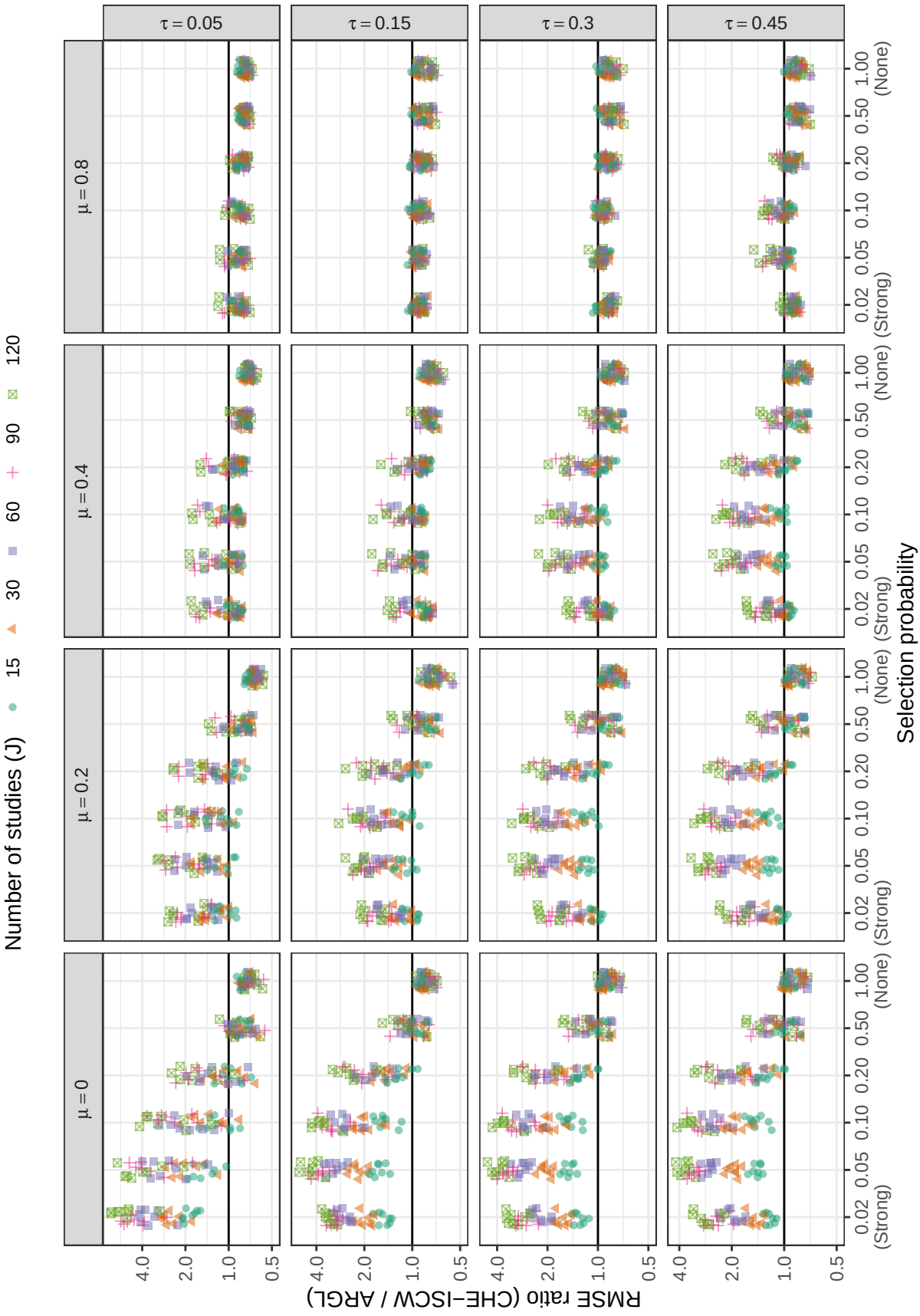

Figure D4
Ratio of root mean-squared error for CHE-ISCW estimator to root mean-squared error of ARGL estimator by selection probability, number of studies, average SMD, and between-study heterogeneity

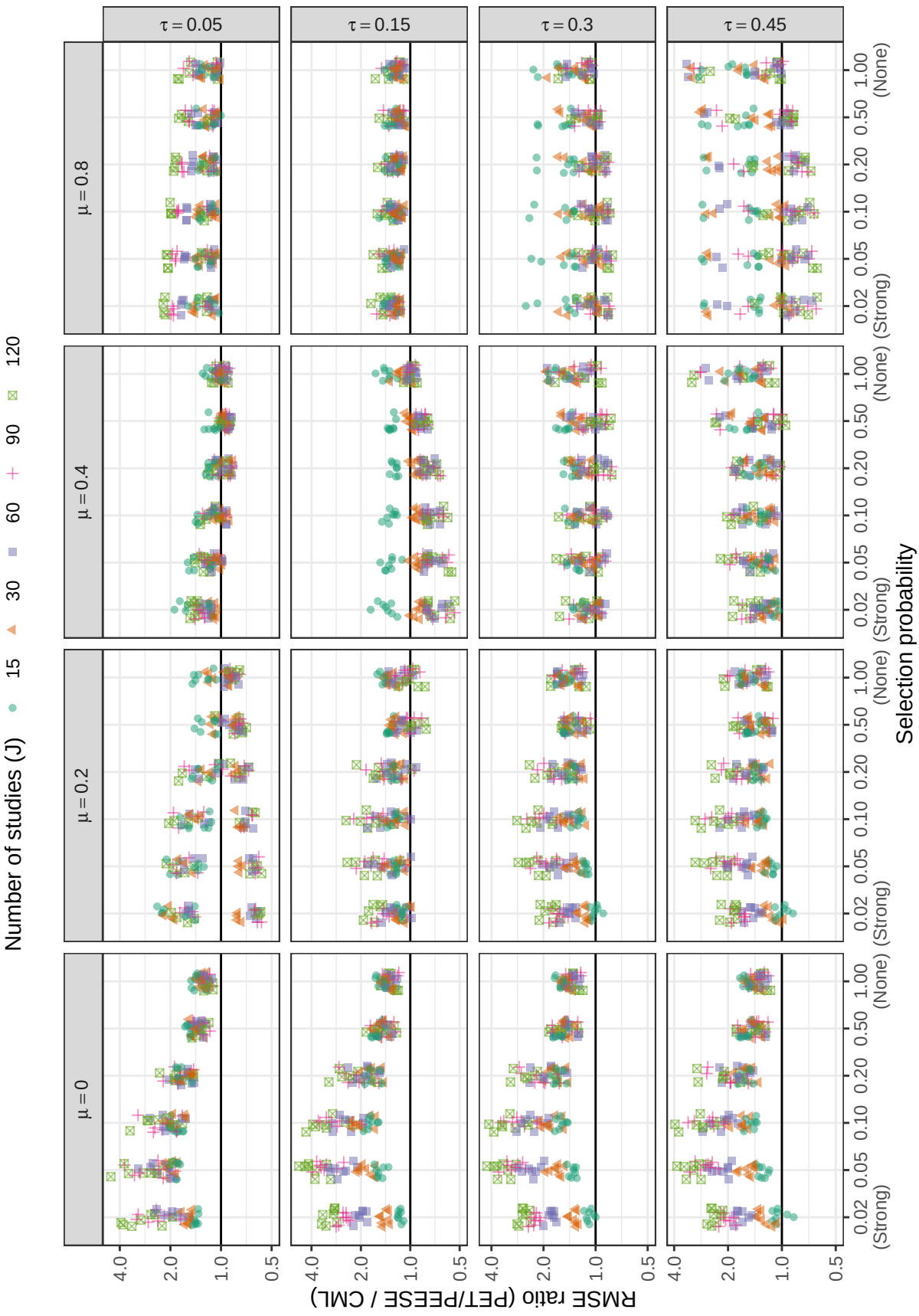

Figure D5

Ratio of root mean-squared error for PET/PEESE estimator to root mean-squared error of CML estimator by selection probability, number of studies, average SMD, and between-study heterogeneity

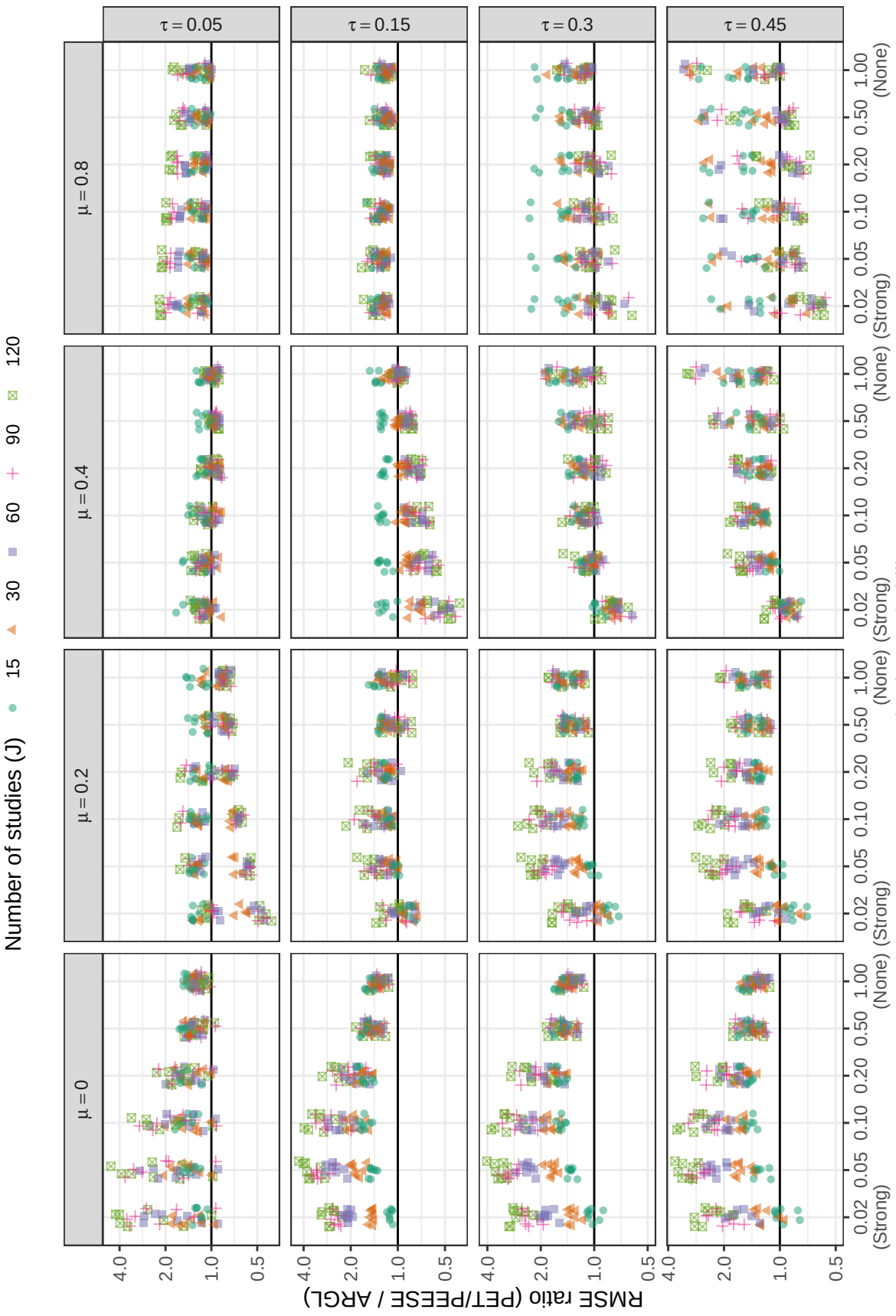

Figure D6
Ratio of root mean-squared error for PET/PEESE estimator to root mean-squared error of ARGL estimator by selection probability, number of studies, average SMD, and between-study heterogeneity

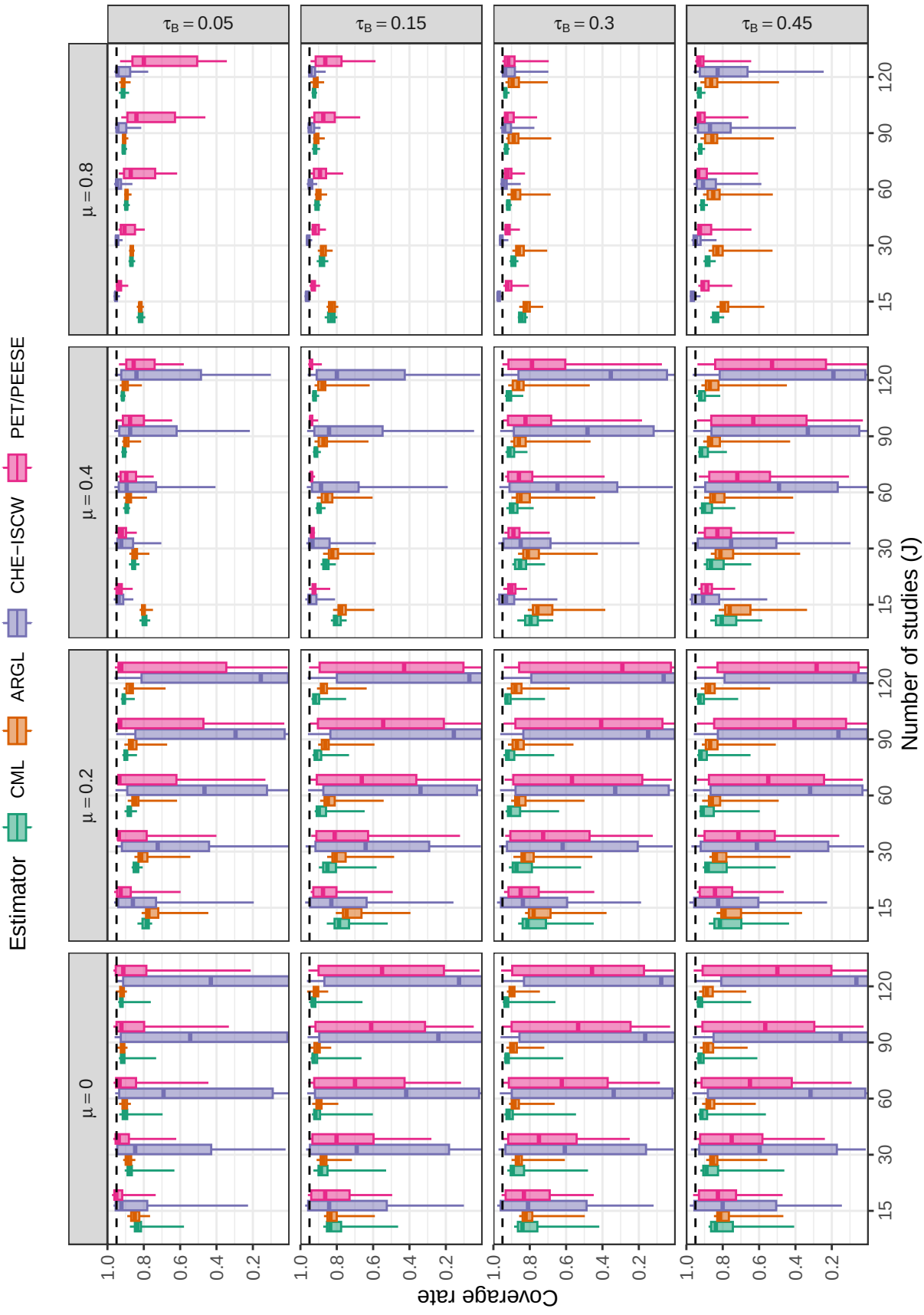

Figure D7

Coverage levels of confidence intervals based for average effect size based on cluster-robust variance approximations, by number of studies, average SMD, and between-study heterogeneity. Dashed lines correspond to the nominal confidence level of 0.95.

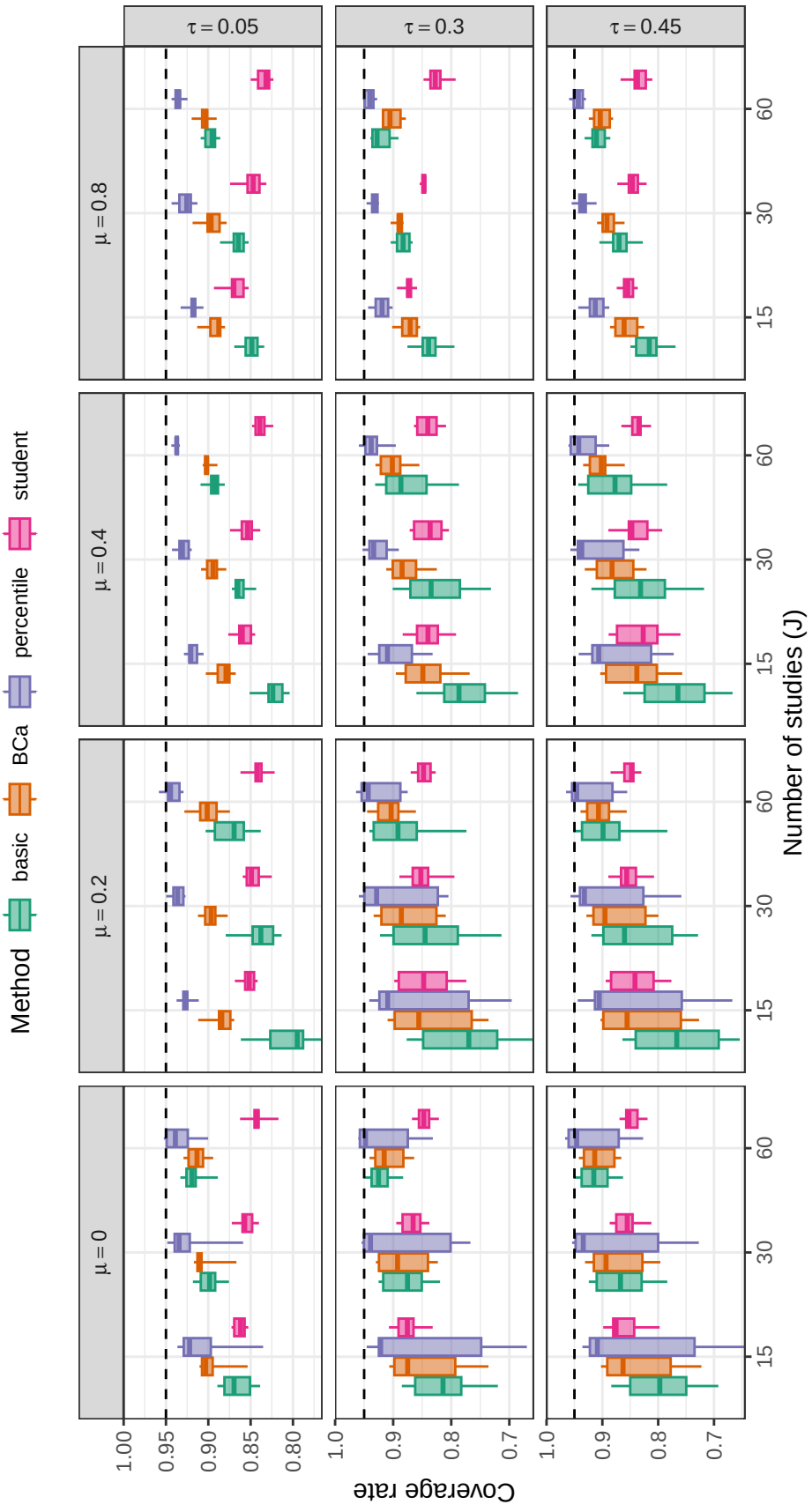

Figure D8
Coverage levels of two-stage bootstrap confidence intervals based on the CML estimator of average effect size by number of studies, average SMD, and between-study heterogeneity. Dashed lines correspond to the nominal confidence level of 0.95.

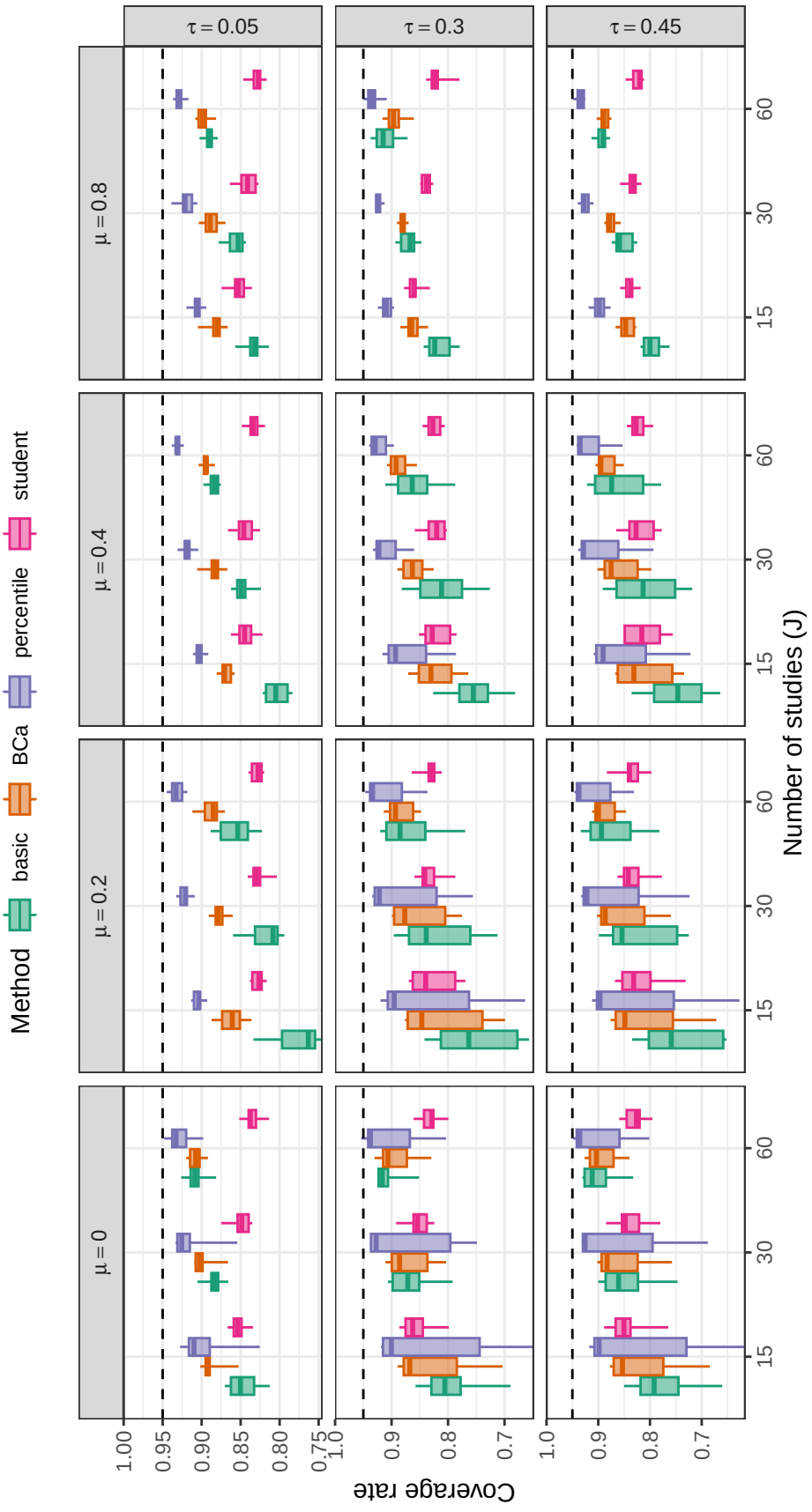

Figure D9
Coverage levels of multinomial bootstrap confidence intervals based on the CML estimator of average effect size by number of studies, average SMD, and between-study heterogeneity. Dashed lines correspond to the nominal confidence level of 0.95.

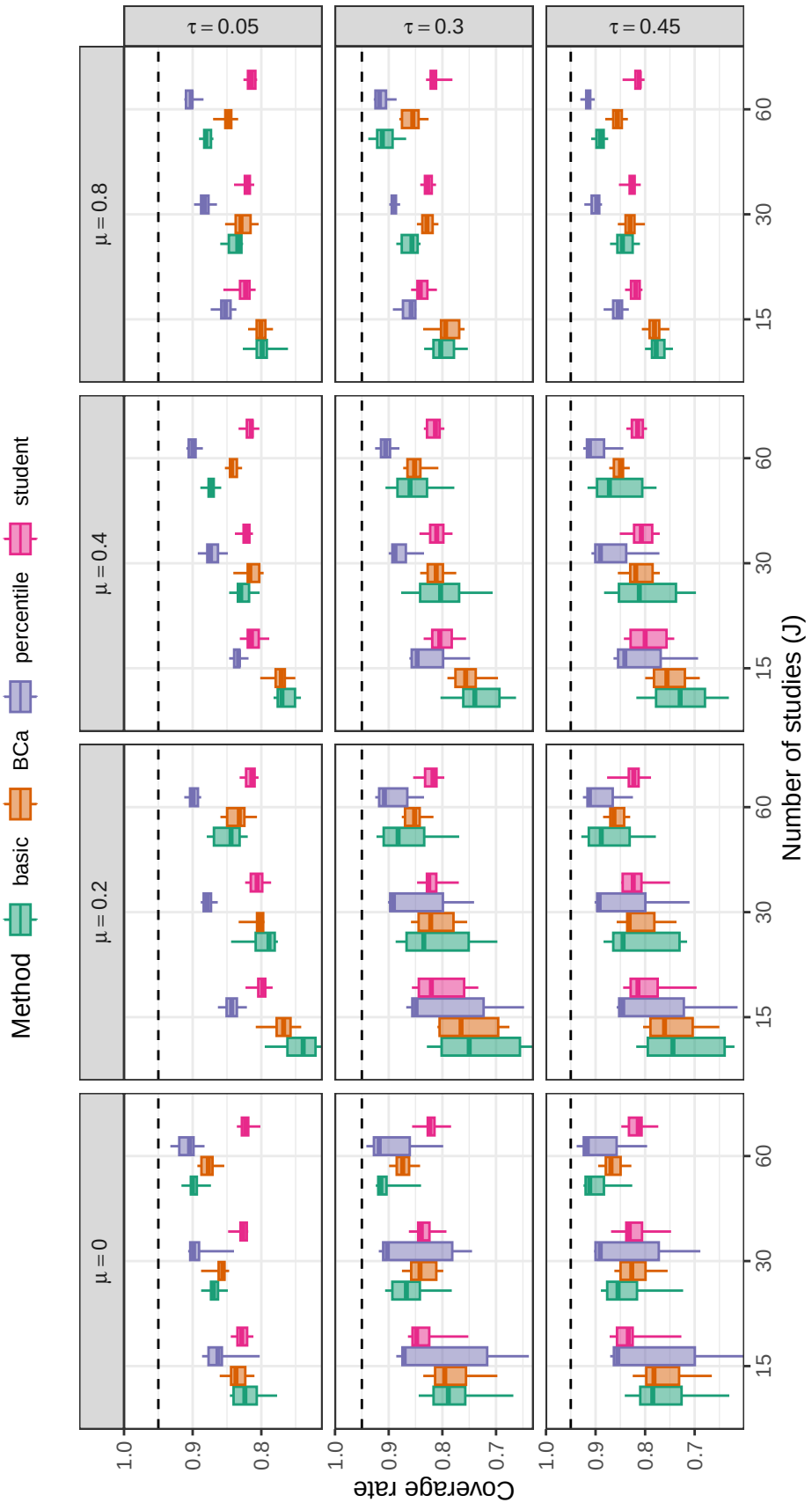

Figure D10
Coverage levels of fractional random weight bootstrap confidence intervals based on the CML estimator of average effect size by number of studies, average SMD, and between-study heterogeneity. Dashed lines correspond to the nominal confidence level of 0.95.

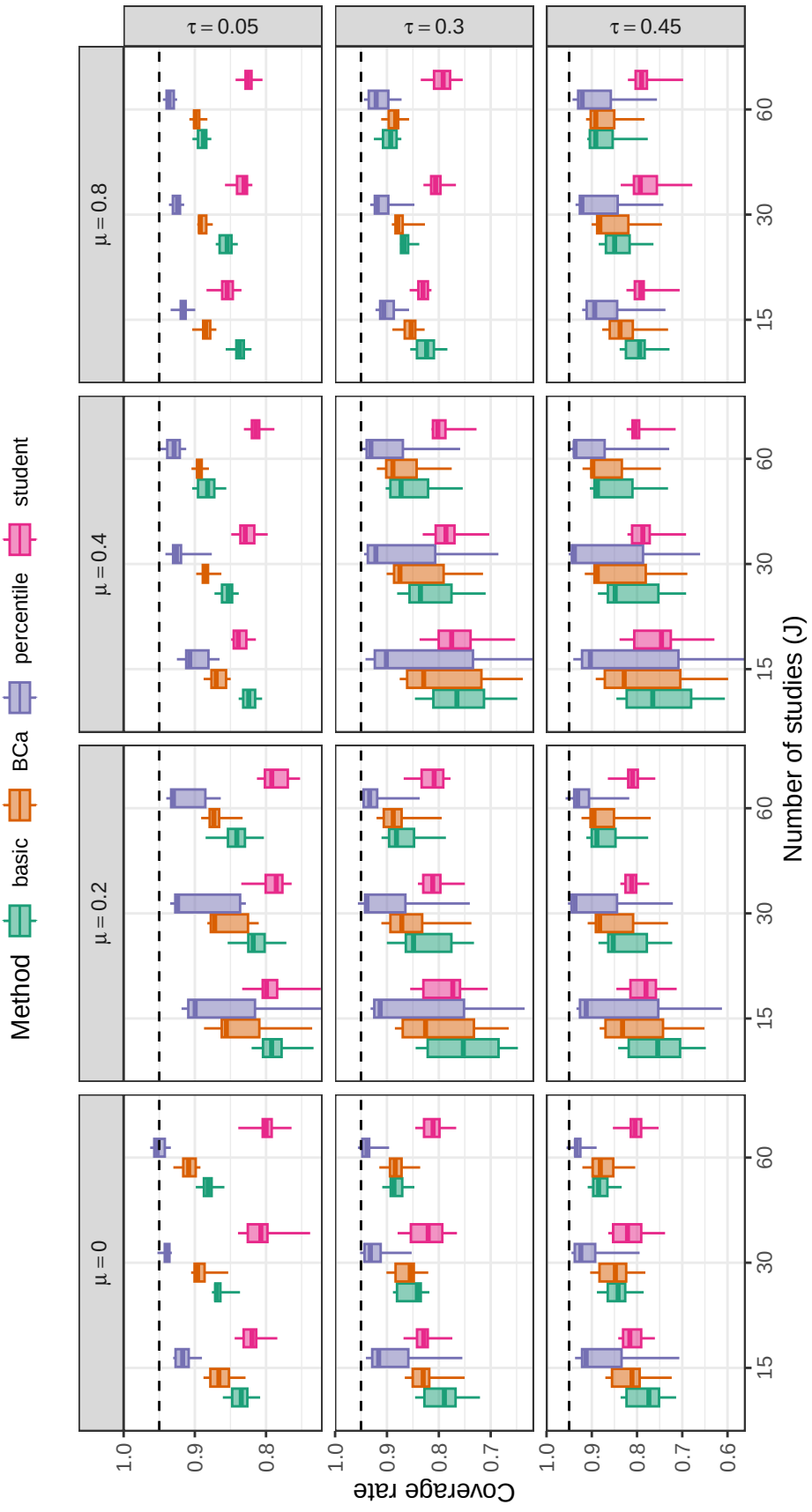

Figure D11
Coverage levels of two-stage bootstrap confidence intervals based on the ARGL estimator of average effect size by number of studies, average SMD, and between-study heterogeneity. Dashed lines correspond to the nominal confidence level of 0.95.

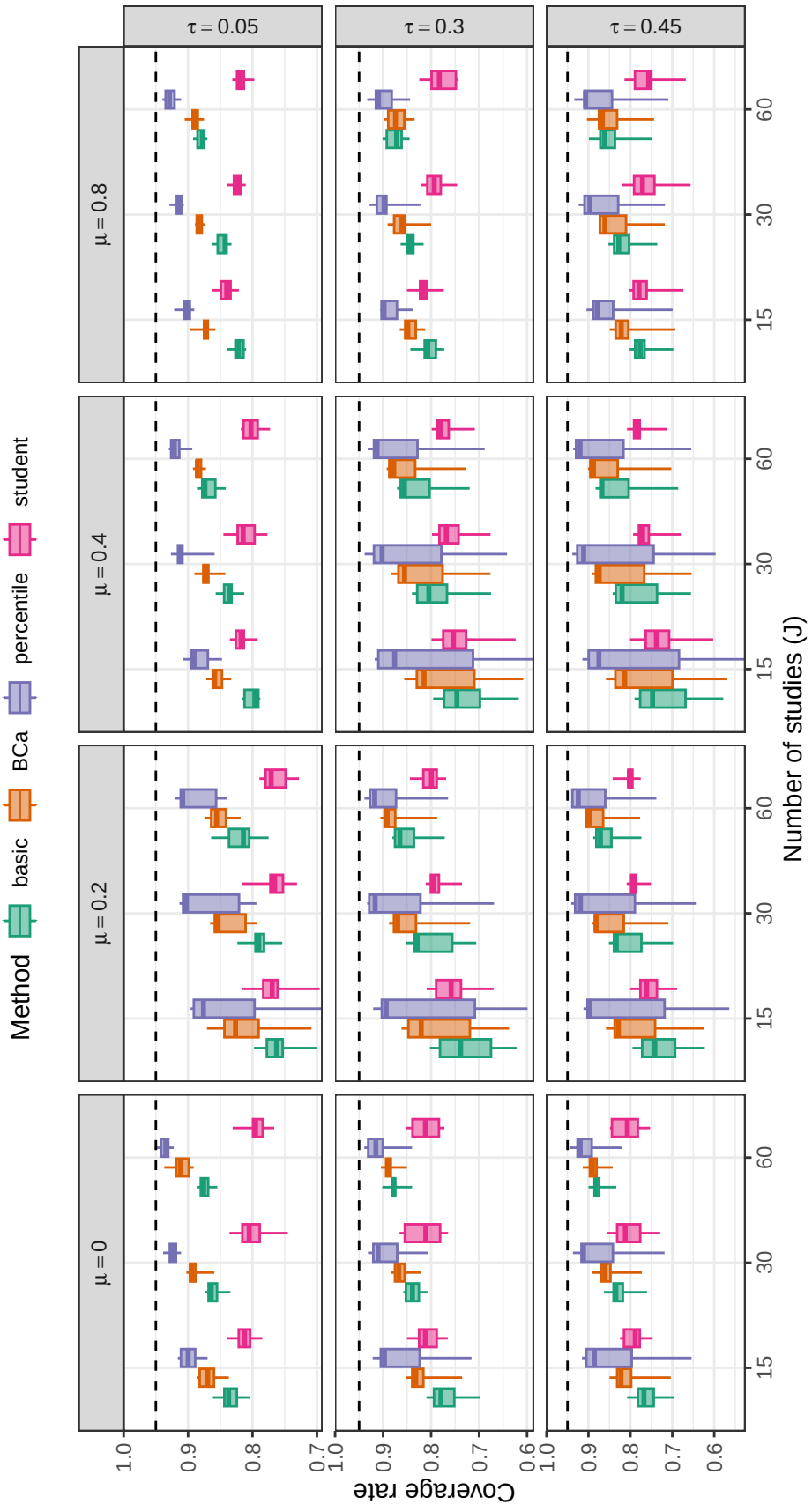

Figure D12
Coverage levels of multinomial bootstrap confidence intervals based on the ARGL estimator of average effect size by number of studies, average SMD, and between-study heterogeneity. Dashed lines correspond to the nominal confidence level of 0.95.

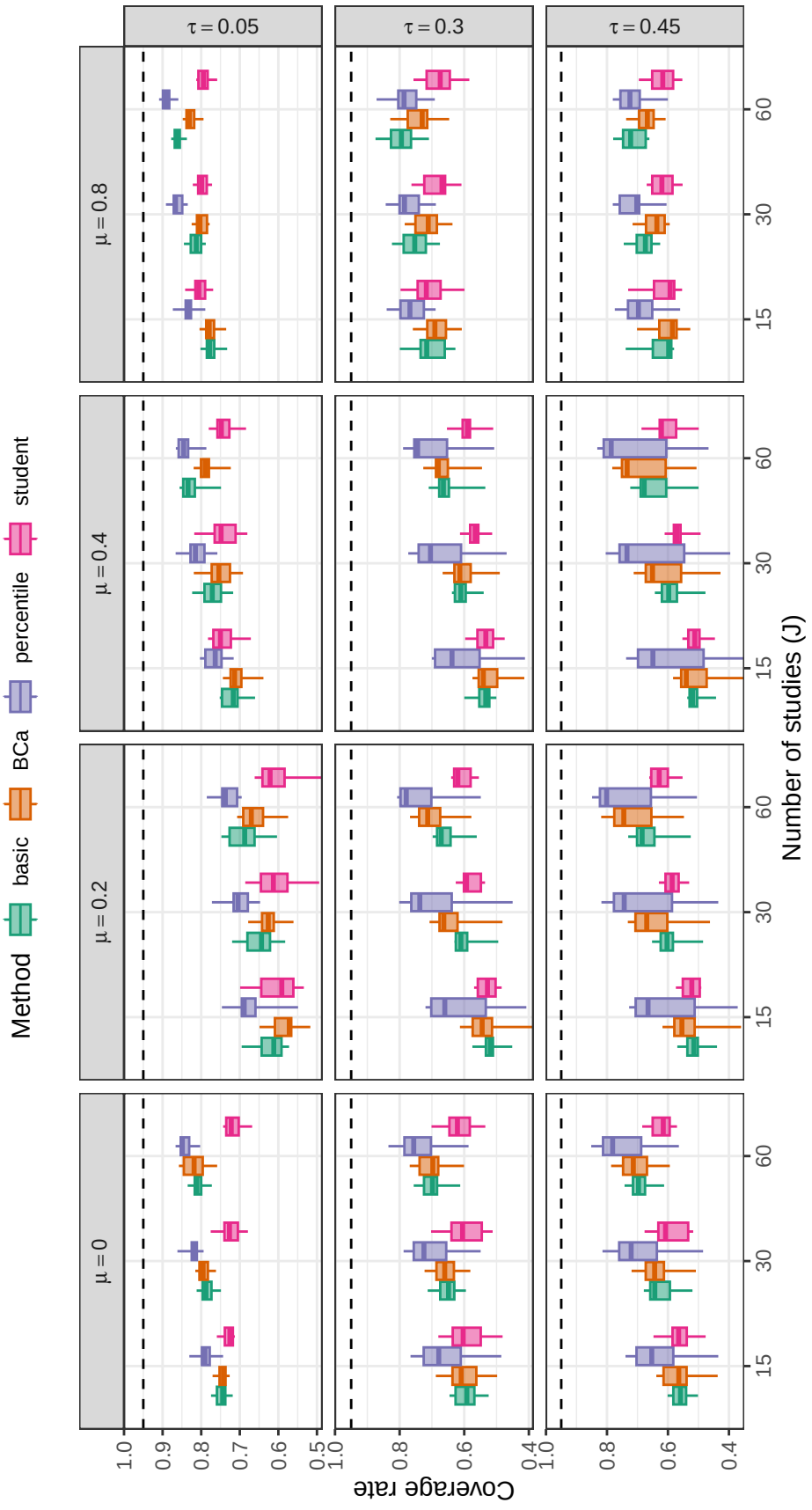

Figure D13
Coverage levels of fractional random weight bootstrap confidence intervals based on the ARGL estimator of average effect size by number of studies, average SMD, and between-study heterogeneity. Dashed lines correspond to the nominal confidence level of 0.95.

E Additional simulation results for estimators of heterogeneity (τ^2)

We briefly consider estimation of the marginal heterogeneity of the effect size distribution, for which the CHE-ISCW, CML, and ARGL methods are all relevant. Figure E1 depicts the bias for the CHE-ISCW, CML, and ARGL estimators of log-heterogeneity $\gamma = \log(\tau^2)$. In most conditions, the estimators are biased in the negative direction. Bias is high for all three estimators in conditions where the between-study heterogeneity is low ($\tau = 0.05$) with bias improving as between-study heterogeneity increases. The CML estimator is less strongly biased than the CHE-ISCW estimator under conditions where there is strong selection. However, the ARGL estimator is badly biased downward, especially under conditions where selection is strong, average SMD is low and between-study heterogeneity is low.

Figure E2 depicts the scaled RMSE for estimators of log-heterogeneity, with Figures E4 and E5 providing greater detail about the relative accuracy of the three methods. Under nearly all conditions, the CML estimator is much more accurate than the ARGL estimator. The RMSE of CML is smaller than that of CHE under some but not all conditions; the former performs better under conditions where selection is strong, average SMD is low, between-study heterogeneity is large, and sample size is large. However, just as with the estimators of average effect size, the CHE estimator is more accurate when selection is mild or absent, leading to a bias-variance trade-off.

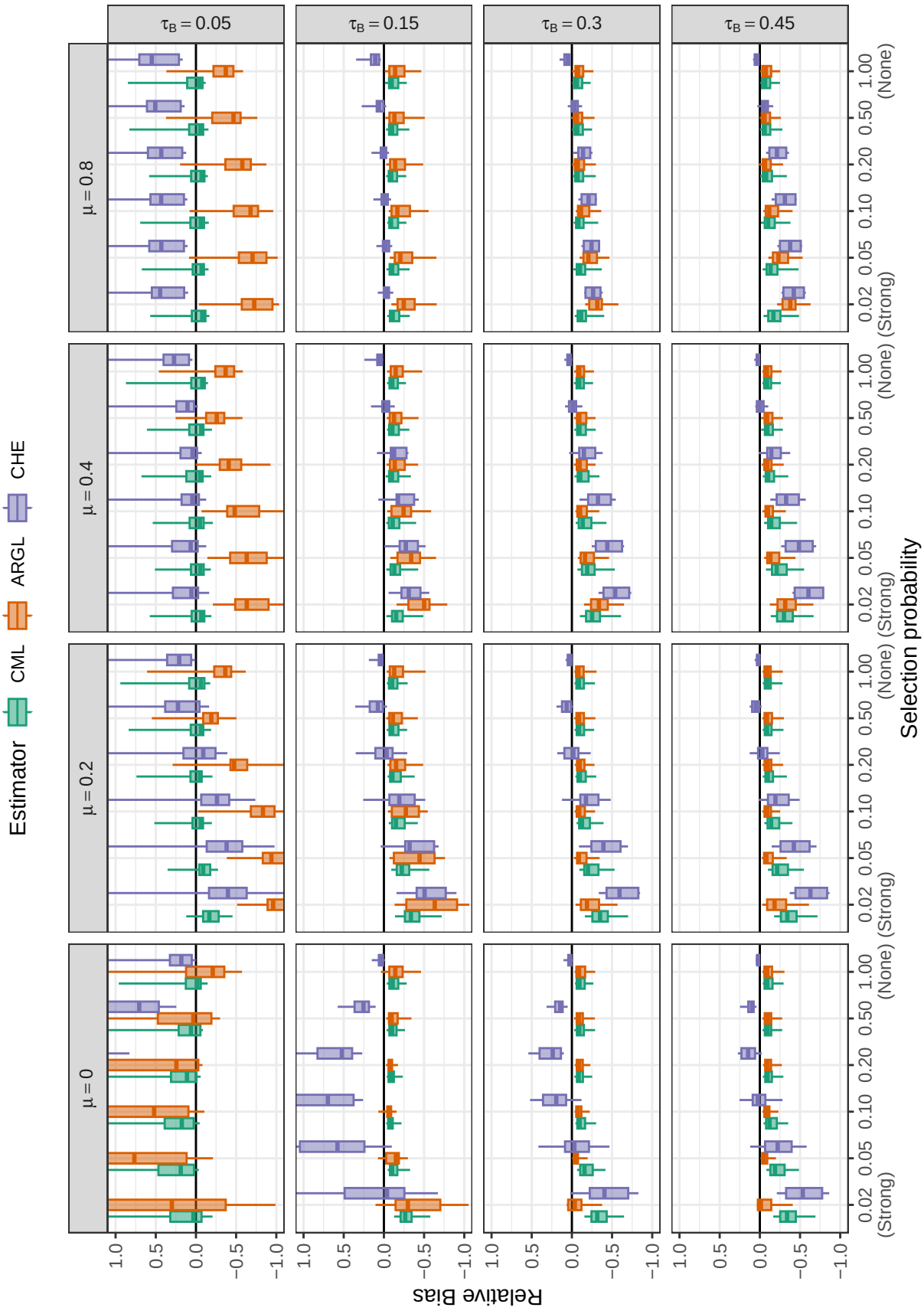

Figure E1
Relative bias of heterogeneity estimators by selection probability, average SMD, and between-study heterogeneity

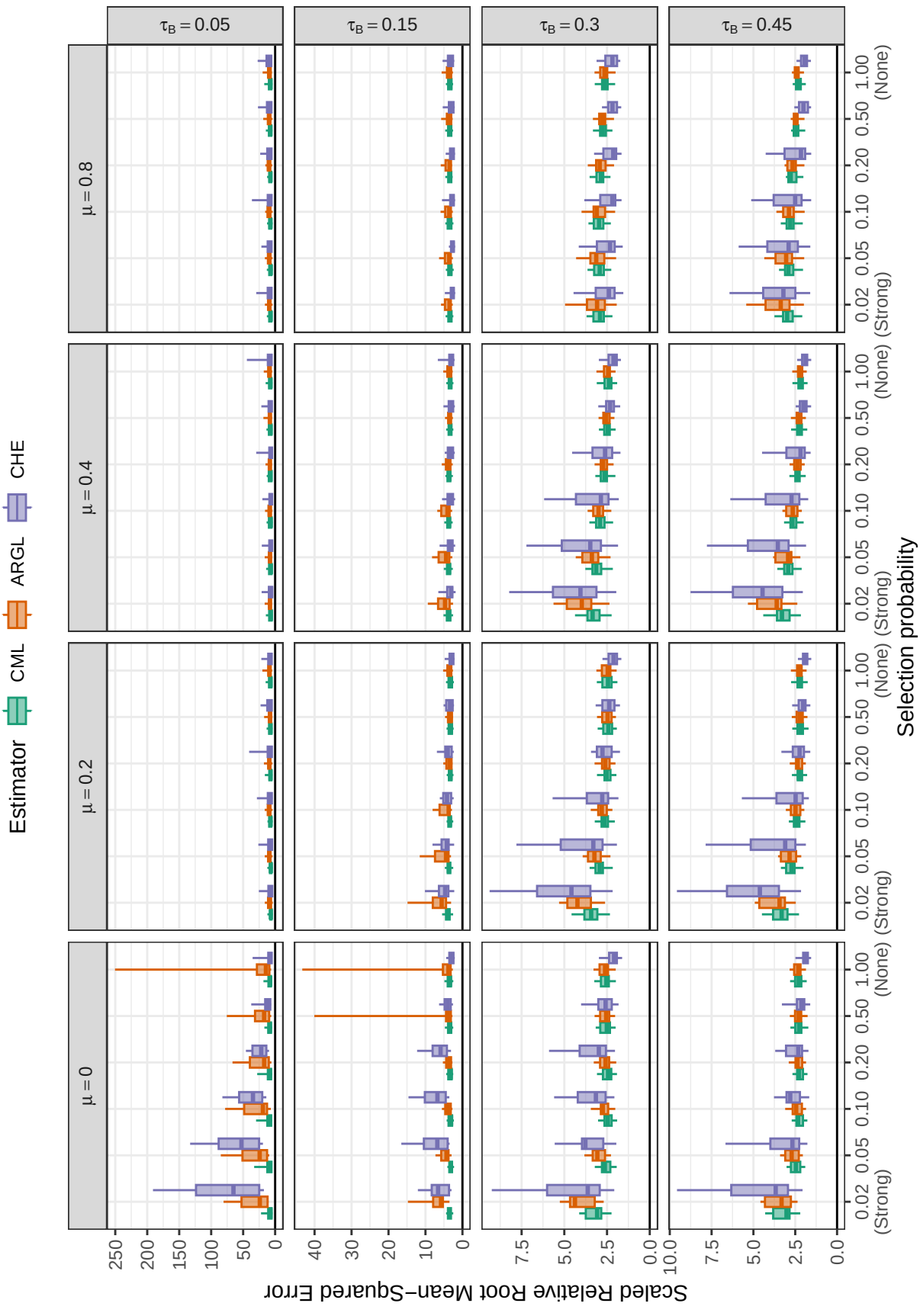

Figure E2
Scaled relative root mean-squared error of heterogeneity estimators by selection probability, average SMD, and between-study heterogeneity

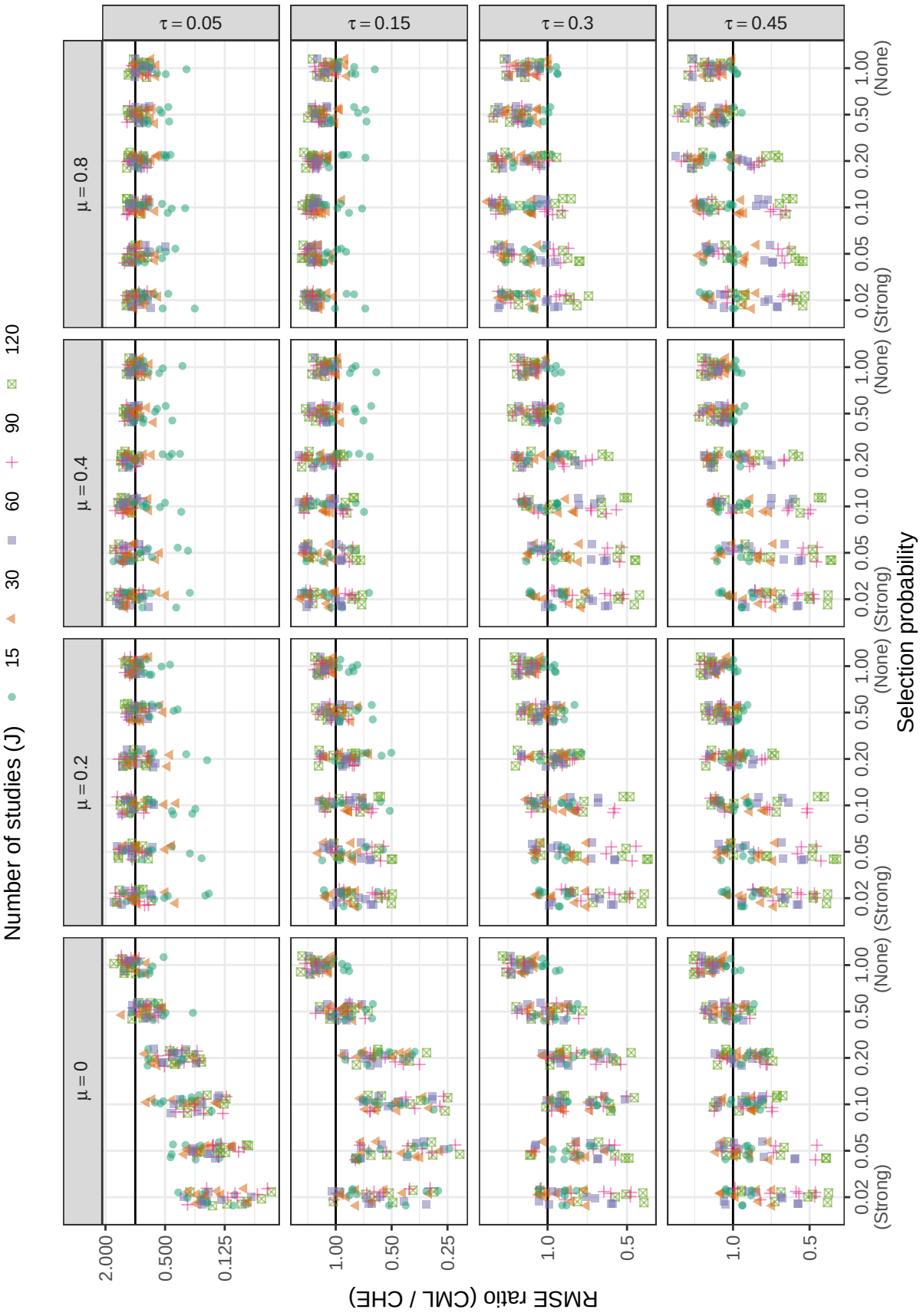

Figure E3
Ratio of root mean-squared error for CML heterogeneity estimator to root mean-squared error of CHE heterogeneity estimator by selection probability, number of studies, average SMD, and between-study heterogeneity

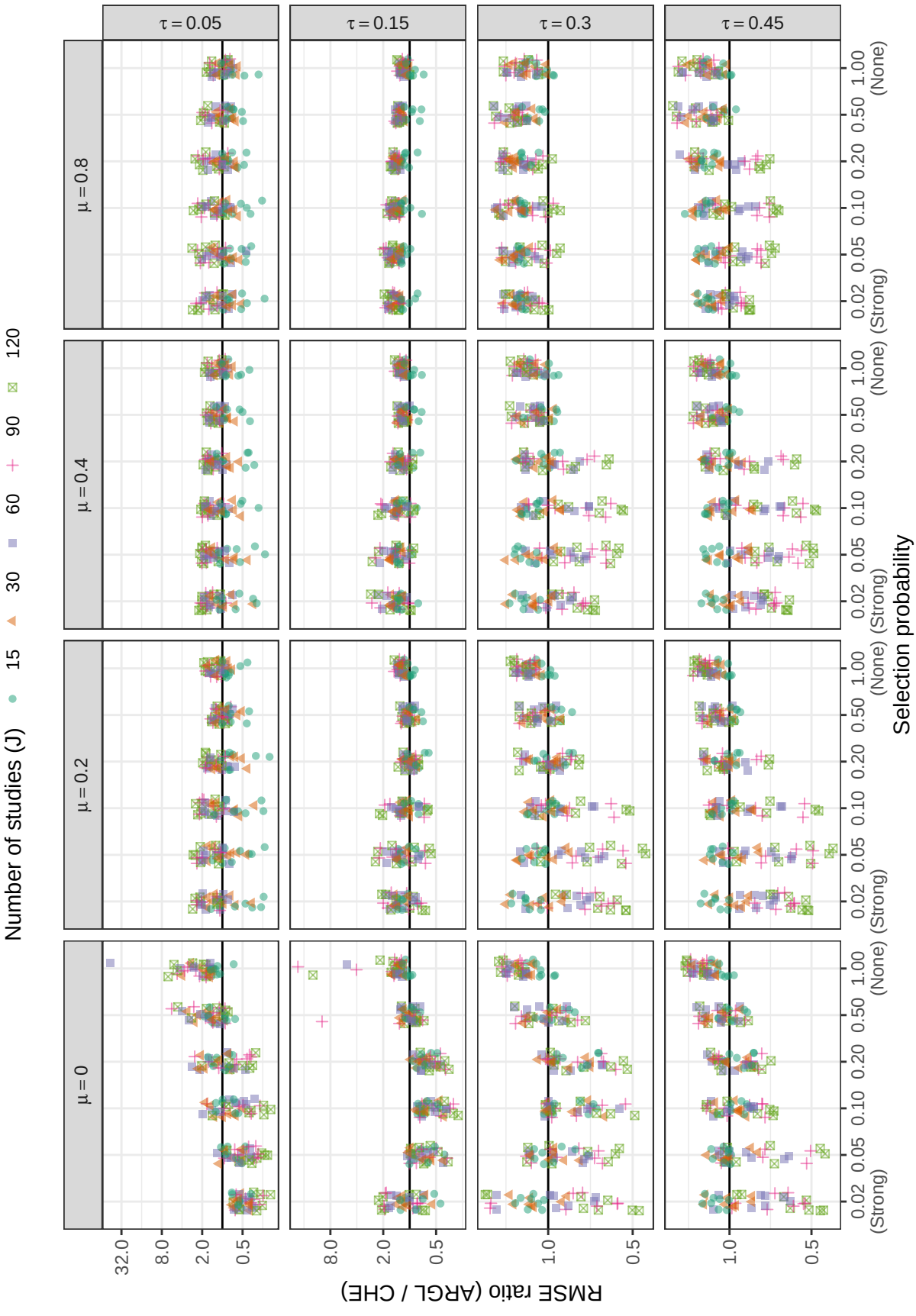

Figure E4
Ratio of root mean-squared error for ARGL heterogeneity estimator to root mean-squared error of CHE heterogeneity estimator by selection probability, number of studies, average SMD, and between-study heterogeneity

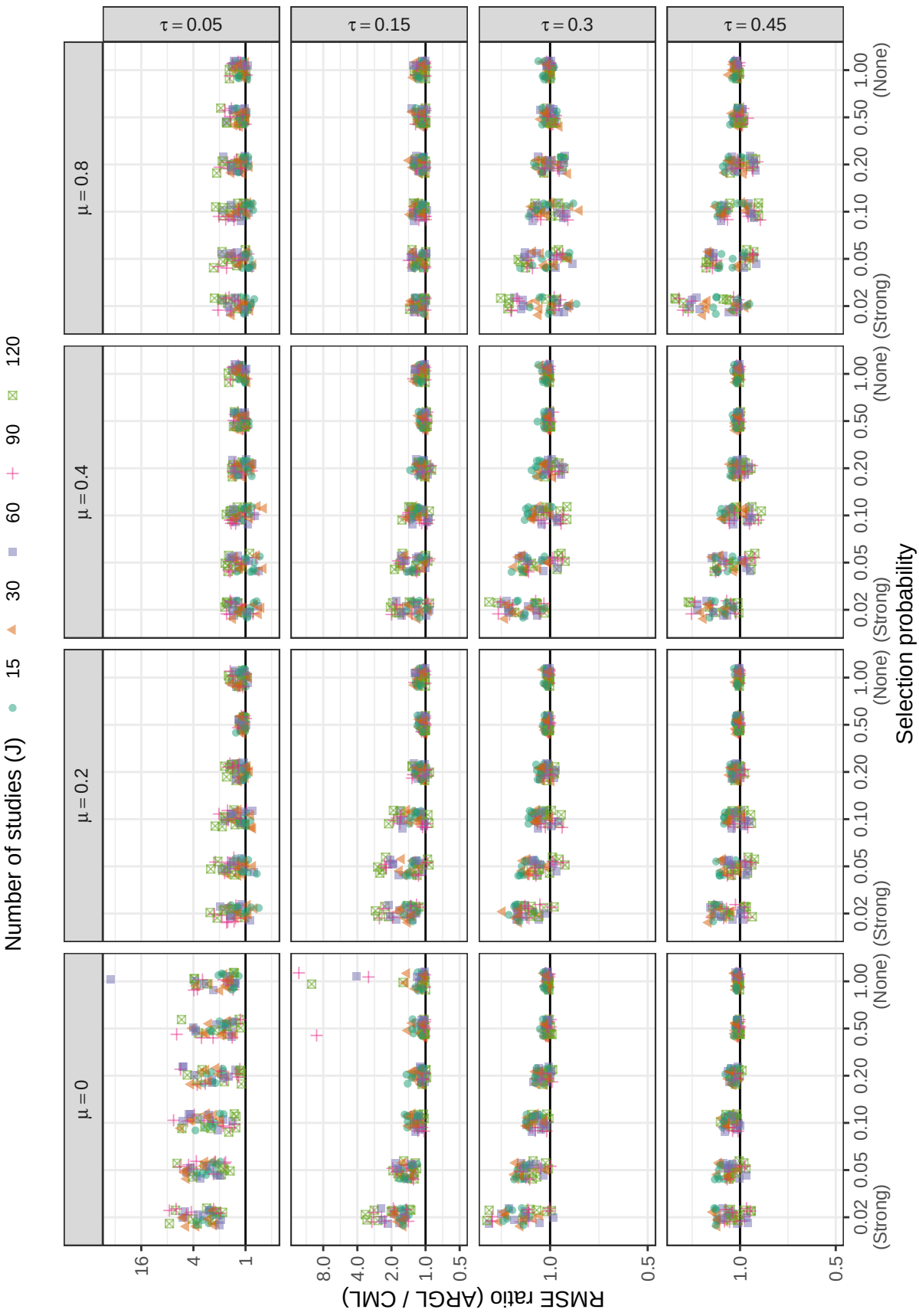

Figure E5
Ratio of root mean-squared error for ARGL heterogeneity estimator to root mean-squared error of CML heterogeneity estimator by selection probability, number of studies, average SMD, and between-study heterogeneity

F Additional simulation results for estimators of selection parameter (ζ_1)

The CML and ARGL methods both estimate the log-selection parameter of the step-function model, a parameter that may be of substantive interest. Figure F1 shows that the bias of the two estimators is similar: Both show bias close to zero under conditions with low average SMD and high between-study heterogeneity but have large negative biases under conditions with strong selection probability and high average SMD, especially with low between-study heterogeneity. Figure F2 depicts the RMSE for the ARGL and CML estimators of the log-selection parameter with Figure F3 providing a more detailed view of the relative accuracy of the ARGL versus the CML estimators. The CML estimator of the selection parameter generally outperforms the ARGL estimator, especially under conditions with strong selection. Figures F4 and F5 depict the confidence interval coverage of the two estimators with various bootstrap confidence interval estimation approaches, using two-stage clustered bootstrap resampling. Broadly, none of the methods provide coverage rates that are close to the nominal 95% level across all conditions examined.

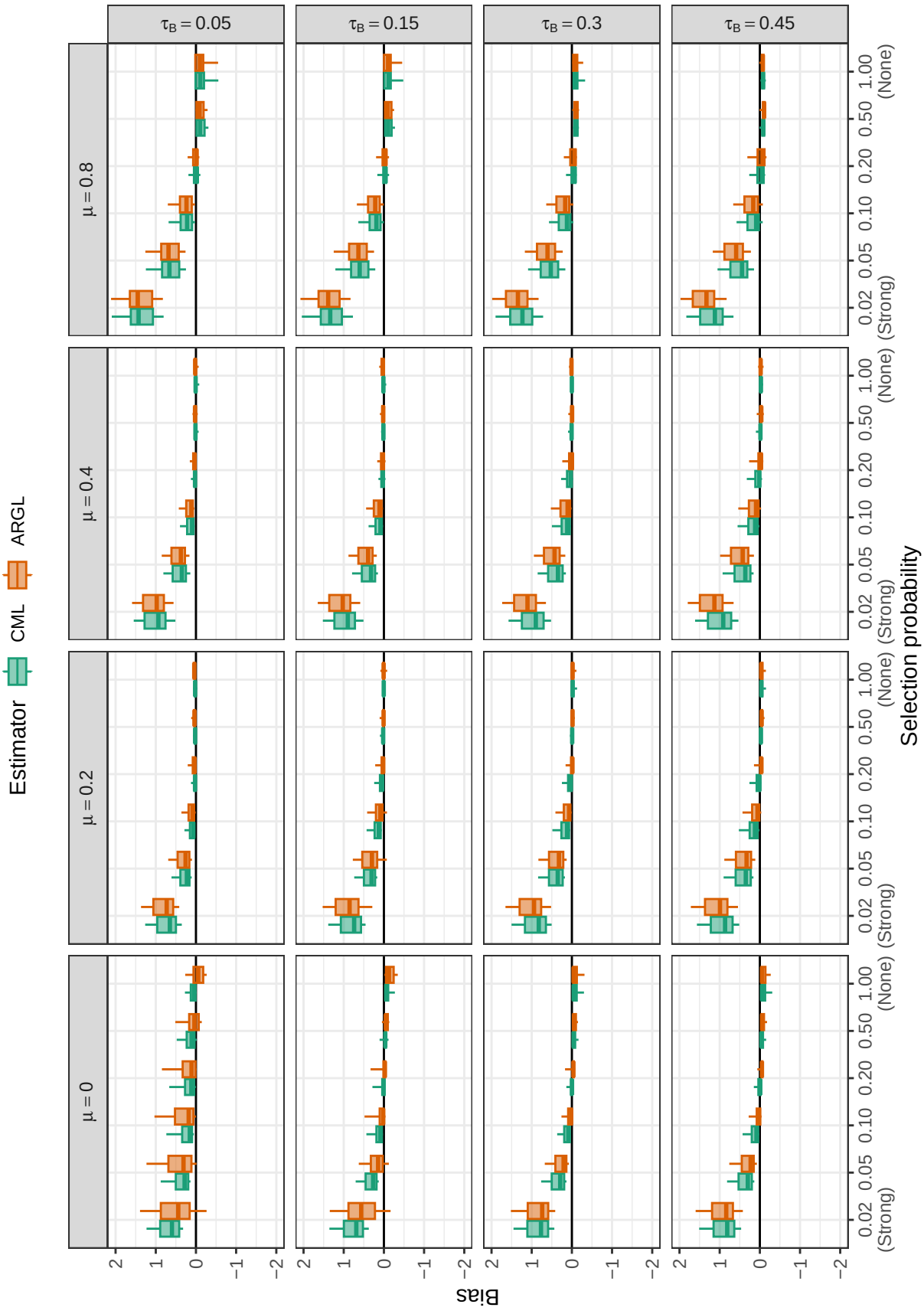

Figure F1
Bias for estimators of log-selection parameter by selection probability, average SMD, and between-study heterogeneity

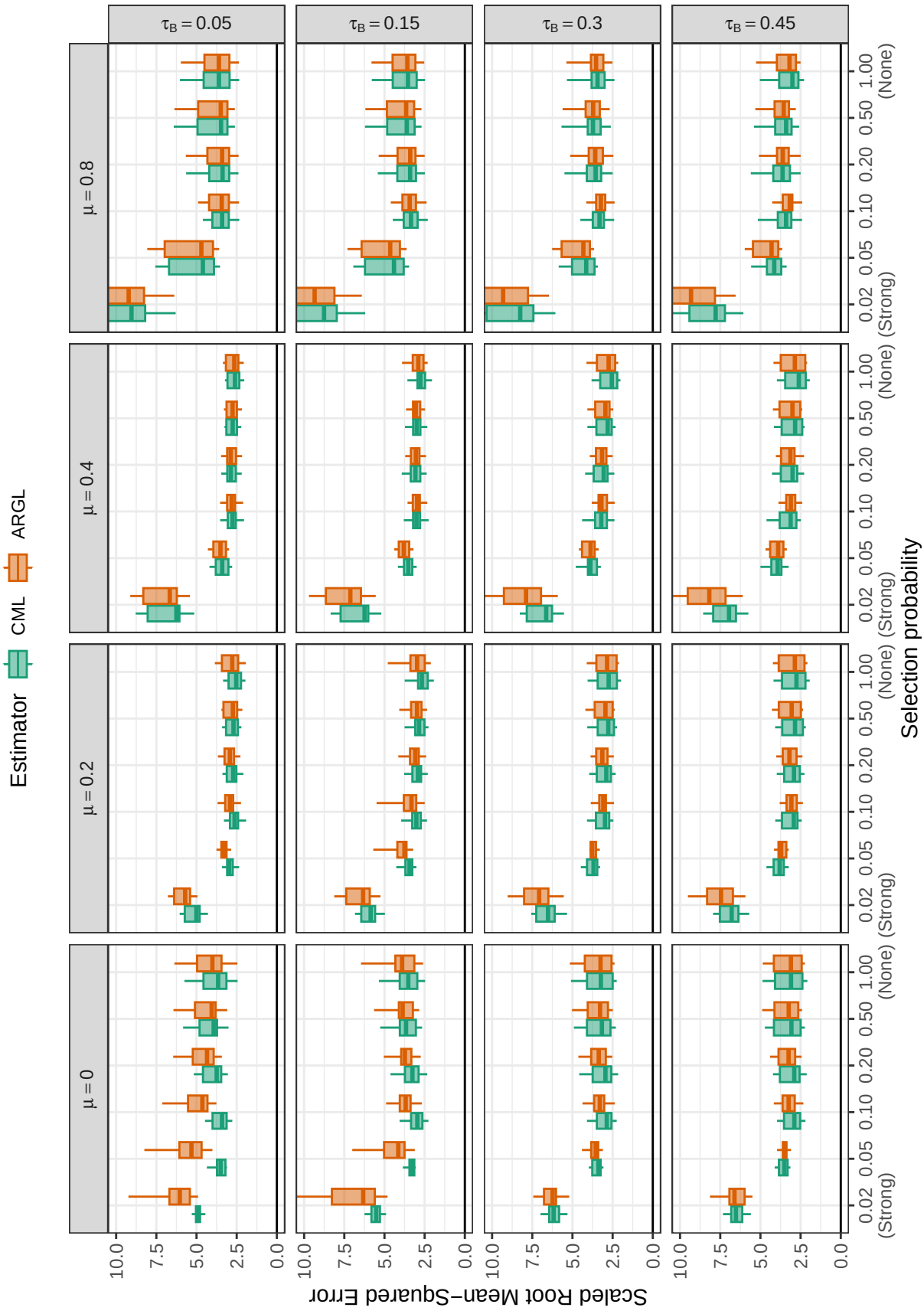

Figure F2
Scaled root mean-squared error for log-selection parameter estimators by selection probability, average SMD, and between-study heterogeneity

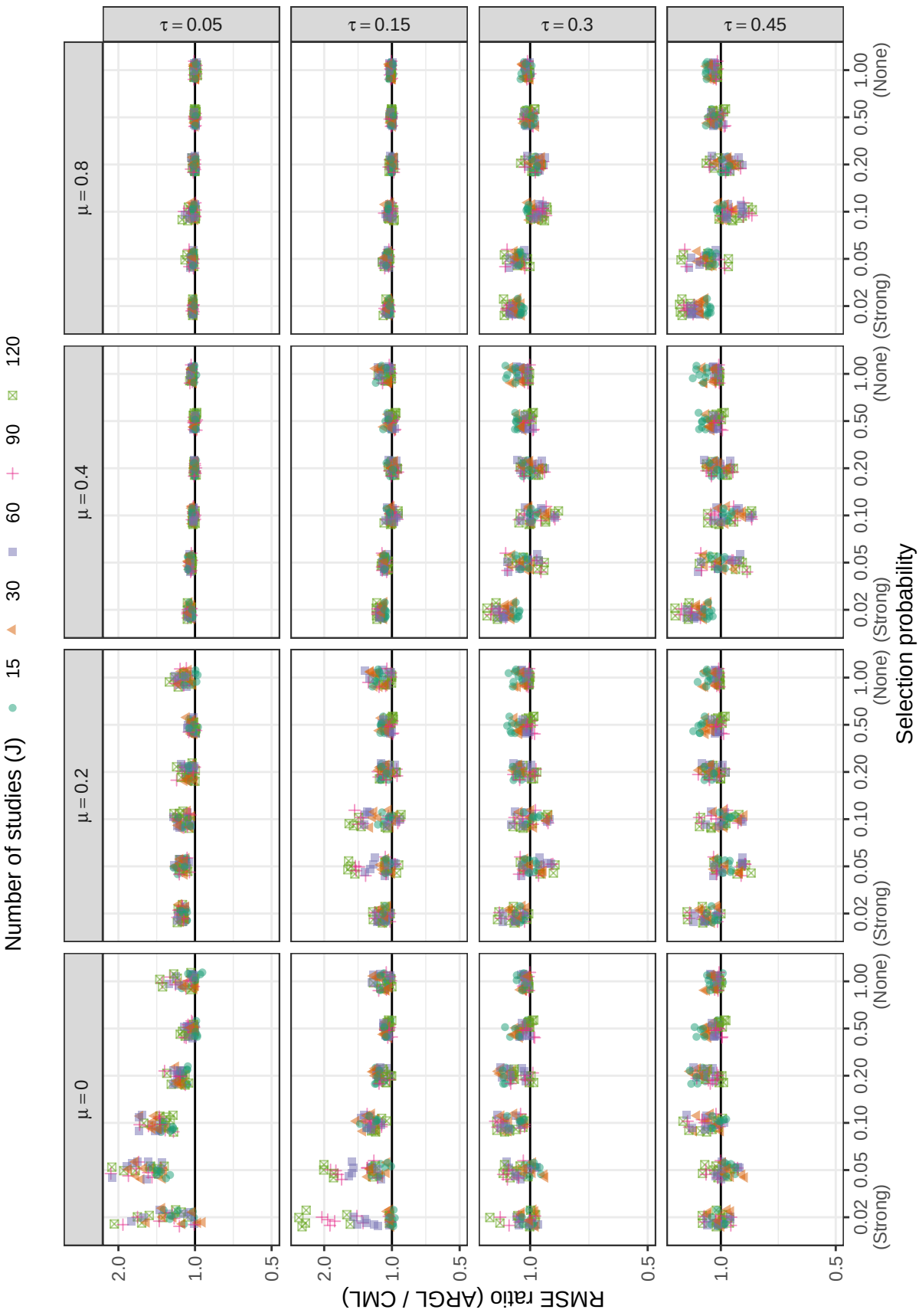

Figure F3
Ratio of root mean-squared error for ARGL log-selection parameter estimator to root mean-squared error of CML log-selection parameter estimator by selection probability, number of studies, average SMD, and between-study heterogeneity

Figure F4
Coverage levels of two-stage bootstrap confidence intervals based on the CML estimator of log-selection parameter by number of studies, average SMD, and between-study heterogeneity. Dashed lines correspond to the nominal confidence level of 0.95.

Figure F5
Coverage levels of two-stage bootstrap confidence intervals based on the ARGL estimator of log-selection parameter by number of studies, average SMD, and between-study heterogeneity. Dashed lines correspond to the nominal confidence level of 0.95.

Additional References

1. Xu L, Gotwalt C, Hong Y, King CB, Meeker WQ. Applications of the fractional-random-weight bootstrap. *The American Statistician*. 2020;74(4):345-358. doi:10.1080/00031305.2020.1731599

2. Boos DD, Zhang J. Monte carlo evaluation of resampling-based hypothesis tests. *Journal of the American Statistical Association*. 2000;95(450):486-492. doi:10.1080/01621459.2000.10474226