

# Simulation Studies in “Three-way Mixed Effect ANOVA to Estimate MRMC Limits of Agreement”

In this R Markdown file, we provide the code for the simulation studies in the manuscript. All the functions and dataset used in this file are in ANOVA.MRMC.LOA R package (<https://github.com/SiWen314/ANOVA.MRMC.LOA>). Two simulation studies were conducted to validate both the simulation and limits of agreement variance estimates.

## Verify the simulation is consistent with the derived theoretical values

By simulating individual WRBM and BRBM differences independently from the model, we can calculate Monte Carlo estimates of the variances of WRBM and BRBM differences and compare them to the theoretical values derived from the model. There were two sets of parameters in this experiment. One set of parameters fixed the case related parameters  $\sigma_C^2 = \sigma_{\tau C}^2 = 1$  and allowed the reader related parameter  $\alpha_R (= \alpha_{\tau R})$  to range from 2 to 20. The other set of parameters fixed the reader related parameters  $\alpha_R = \alpha_{\tau R} = 10$  and allowed the case related parameter  $\sigma_C^2 (= \sigma_{\tau C}^2)$  to range from 0.1 to 2, incrementing by 0.1. In both parameter sets,  $\beta_R = \beta_{\tau R} = 1$ , so that the reader variability is only affected by  $\alpha_R$ . We simulated 100,000 trials with each trial having 4 measurements from 2 readers for a single case under 2 modalities. The Monte Carlo estimates of the variances are the sample variances of the 100,000 independent WRBM and BRBM differences. The relative bias between the Monte Carlo estimates  $\tilde{V}_{WR}^{12}$ ,  $\tilde{V}_{BR}^{12}$  and the derived theoretical values  $V_{WR}^{12}$ ,  $V_{BR}^{12}$  are defined as

$$\text{relative\_bias}(\tilde{V}_{WR}^{12}) = \frac{\tilde{V}_{WR}^{12} - V_{WR}^{12}}{V_{WR}^{12}}$$
$$\text{relative\_bias}(\tilde{V}_{BR}^{12}) = \frac{\tilde{V}_{BR}^{12} - V_{BR}^{12}}{V_{BR}^{12}}$$

```
# First Parameter set
alpha.list = seq(2,20,1)
sigma.list = c(1)

result.validSimulation.alphaR <- validateSimulation(alpha.list, sigma.list)

# Second Parameter set
alpha.list =c(10)
sigma.list =seq(0.1,2,0.1)

result.validSimulation.sigmaC <- validateSimulation(alpha.list, sigma.list)
```

In Figure 1, we show the relative bias between Monte Carlo estimates  $\tilde{V}_{WR}^{12}$ ,  $\tilde{V}_{BR}^{12}$  and the derived theoretical value  $V_{WR}^{12}$ ,  $V_{BR}^{12}$  while varying the reader related parameter  $\alpha_R$  and case related parameter  $\sigma_C^2$ . The dashed horizontal lines in each of the subplots denote the 0 bias between the Monte Carlo estimation and theoretical values. As shown in the figures, the relative bias is distributed tightly around 0 and the absolute value of the relative bias is less than 1% for most of the cases. There is no linear trend of the relative bias with respect to the changes of the parameter settings, which indicates that the bias is independent of the magnitudes of the theoretical value.

```
library(ggplot2)
library(gridExtra)
```

```

# First Parameter set

## Within-Reader
p1 <- ggplot(result.validSimulation.alphaR, aes(x=alpha_R,y=WR_var_relative_bias)) +
  geom_point() +
  geom_hline(yintercept=0, linetype="dashed")+
  ylab(expression(relative_bias~(tilde(V)[WR]^12)))+
  xlab(expression(alpha[R]))+
  ylim(-0.011,0.011)+
  ggtitle(expression("(a) WRBM, " * sigma[C]^2 * " = 1"))

## Between-Reader
p2 <- ggplot(result.validSimulation.alphaR, aes(x=alpha_R,y=BR_var_relative_bias)) +
  geom_point() +
  geom_hline(yintercept=0, linetype="dashed")+
  ylab(expression(relative_bias~(tilde(V)[BR]^12)))+
  xlab(expression(alpha[R]))+
  ylim(-0.011,0.011)+
  ggtitle(expression("(b) BRBM, " * sigma[C]^2 * " = 1"))

# Second Parameter set

## Within-Reader
p3 <- ggplot(result.validSimulation.sigmaC, aes(x=sigma_C,y=WR_var_relative_bias)) +
  geom_point() +
  geom_hline(yintercept=0, linetype="dashed")+
  ylab(expression(relative_bias~(tilde(V)[WR]^12)))+
  xlab(expression(sigma[C]^2))+
  ylim(-0.011,0.011)+
  ggtitle(expression("(c) WRBM, " * alpha[R] * " = 10"))

## Between-Reader
p4 <- ggplot(result.validSimulation.sigmaC, aes(x=sigma_C,y=BR_var_relative_bias)) +
  geom_point() +
  geom_hline(yintercept=0, linetype="dashed")+
  ylab(expression(relative_bias~(tilde(V)[BR]^12)))+
  xlab(expression(sigma[C]^2))+
  ylim(-0.011,0.011)+
  ggtitle(expression("(d) BRBM, " * alpha[R] * " = 10"))

do.call("grid.arrange", c(list(p1,p2,p3,p4), ncol=2))

```

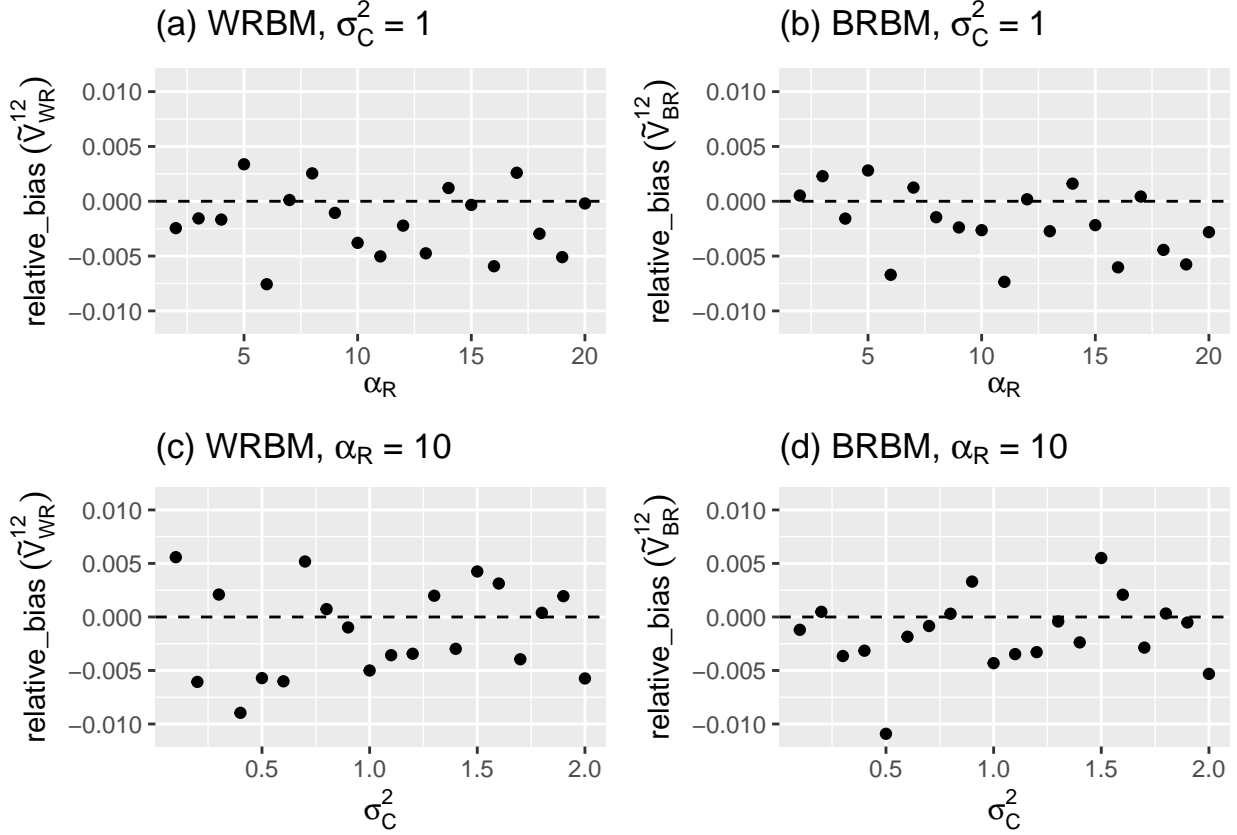


Figure 1 Relative bias between the Monte Carlo estimation ( $T = 100,000$ ) and the derived theoretical value changing over  $\alpha_R$  and  $\sigma_C^2$

### Validate and characterize the MRMC limits of agreement estimates

In this experiment, we compared the variance estimates of WRBM and BRBM differences by ANOVA to the derived theoretical values. We denote the variance estimates of WRBM and BRBM differences for the trial  $t$  by  $\hat{V}_{WR,t}^{12}$  and  $\hat{V}_{BR,t}^{12}$  and we assess the estimates in terms of relative bias and coefficient of variation:

$$relative\_bias(\hat{V}_{WR}^{12}) = \frac{\frac{1}{T} \sum_t \hat{V}_{WR,t}^{12} - V_{WR}^{12}}{V_{WR}^{12}}$$

$$CV(\hat{V}_{WR}^{12}) = \frac{SD(\hat{V}_{WR}^{12})}{V_{WR}^{12}}$$

where  $SD()$  denotes the sample standard deviation across  $T = 1000$  Monte Carlo trials. We tested on the following three sets of parameters: \* Different number of readers  $J = 3, 4, \dots, 10, K = 100, \alpha_R = 10, \sigma_C^2 = 1$  \* Different number of cases  $K = 50, 60, \dots, 150, J = 5, \alpha_R = 10, \sigma_C^2 = 1$  \* Different reader and case variabilities  $(\alpha_R, \sigma_C^2) \in \{3, 4, 6, 11, 21\} \times \{0.1, 0.2, 0.4, \frac{2}{3}, 1\}, J = 5, K = 100$  The range of  $\alpha_R$  and  $\sigma_C^2$  in the third set was selected based on the weights of parameters in theoretical value  $V_{BR}^{12}$ . When  $\beta_R = \beta_{\tau R} = 1, \alpha_R = \alpha_{\tau R}, \sigma_C^2 = \sigma_{\tau C}^2$ ,

$$V_{BR}^{12} = \frac{2\beta_R}{\alpha_R - 1} + 2\sigma_{\tau C}^2 + \frac{2\beta_{\tau R}}{\alpha_{\tau R} - 1} = \frac{4}{\alpha_R - 1} + 2\sigma_C^2$$

To make sure both  $\alpha_R$  and  $\sigma_C^2$  range from low variability to high variability and contribute to  $V_{BR}^{12}$  at the similar level, we selected the parameters so that  $\frac{4}{\alpha_R - 1}$  and  $2\sigma_C^2$  can be at 5 different levels: 0.2, 0.4, 0.8, 4/3, 2. Hence, the theoretical value  $V_{BR}^{12}$  ranged from 0.4 to 4.

```

# First Parameter set
nR.list = seq(3,10,1)
nC.list = c(100)
alpha.list = c(6)
sigma.list = c(0.4)

result.validMRMCVarEstimate.nReader <- validateMRMCVarEstimation(nR.list, nC.list,
                                                                    alpha.list, sigma.list)

# Second Parameter set
nR.list = c(5)
nC.list = seq(50,150,10)
alpha.list = c(6)
sigma.list = c(0.4)

result.validMRMCVarEstimate.nCase <- validateMRMCVarEstimation(nR.list, nC.list,
                                                                alpha.list, sigma.list)

# Third Parameter set
nR.list = c(5)
nC.list = c(100)
alpha.list = c(21,11,6,4,3)
sigma.list = c(0.1,0.2,0.4,2/3,1)

result.validMRMCVarEstimate.parameter <- validateMRMCVarEstimation(nR.list, nC.list,
                                                                    alpha.list, sigma.list)

```

Figure 2 and Figure 3 present the relative bias and CV of the variance estimates of BRBM differences changing over different sets of parameters. The plots for the WRBM differences share the similar pattern, so only the ones for the BRBM difference are shown. In Figure 2, plots (a) and (c) show the results when the number of readers ranges from 3 to 10 and the number of cases is 100, while (b) and (d) shows the results when the number of cases changes from 50 to 150 and the number of readers is 5. In both settings, the relative bias observed is small. This is comparable to the relative bias in the previous study. When we compare the CVs for the two sets of simulation parameters, the CVs of the variance estimates of BRBM in both plots (c) and (d) decreases smoothly as the number of readers or cases increases. This is because as we increase the size of the MRMC study, the variance estimate is more precise for each simulation study. Therefore, the variation across all the simulation studies will decrease.

```

# First Parameter set

## relative bias
p1 <- ggplot(result.validMRMCVarEstimate.nReader,aes(x=nReader, y = BR_var_relative_bias))+
  geom_point() +
  geom_hline(yintercept=0, linetype="dashed")+
  xlab("Number of Readers")+
  ylab(expression(relative_bias~(hat(V)[BR]^12)))+
  ylim(-0.011,0.011)+
  ggtitle(expression("(a) K = 100, " * alpha[R] * " = 6, " * sigma[C]^2 * " = 0.4"))

## CV
p2 <- ggplot(result.validMRMCVarEstimate.nReader,aes(x=nReader, y = BR_CV))+
  geom_point() +
  xlab("Number of Readers")+
  ylab(expression(CV~(hat(V)[BR]^12)))+
  ylim(0,0.2)+

```

```

ggtitle(expression("(c) K = 100, "*alpha[R]*" = 6, "* sigma[C]^2 * " = 0.4"))
# Second Parameter set

## relative bias
p3 <- ggplot(result.validMRMCVarEstimate.nCase,aes(x=nCase, y = BR_var_relative_bias))+
  geom_point() +
  geom_hline(yintercept=0, linetype="dashed")+
  xlab("Number of Cases")+
  ylab(expression(relative_bias~(hat(V)[BR]^12)))+
  ylim(-0.011,0.011)+
  ggtitle(expression("(b) J = 5, "*alpha[R]*" = 6, "* sigma[C]^2 * " = 0.4"))
## CV
p4 <- ggplot(result.validMRMCVarEstimate.nCase,aes(x=nCase, y = BR_CV))+
  geom_point() +
  xlab("Number of Cases")+
  ylab(expression(CV~(hat(V)[BR]^12)))+
  ylim(0,0.2)+
  ggtitle(expression("(d) J = 5, "*alpha[R]*" = 6, "* sigma[C]^2 * " = 0.4"))

do.call("grid.arrange", c(list(p1,p3,p2,p4), ncol=2))

```

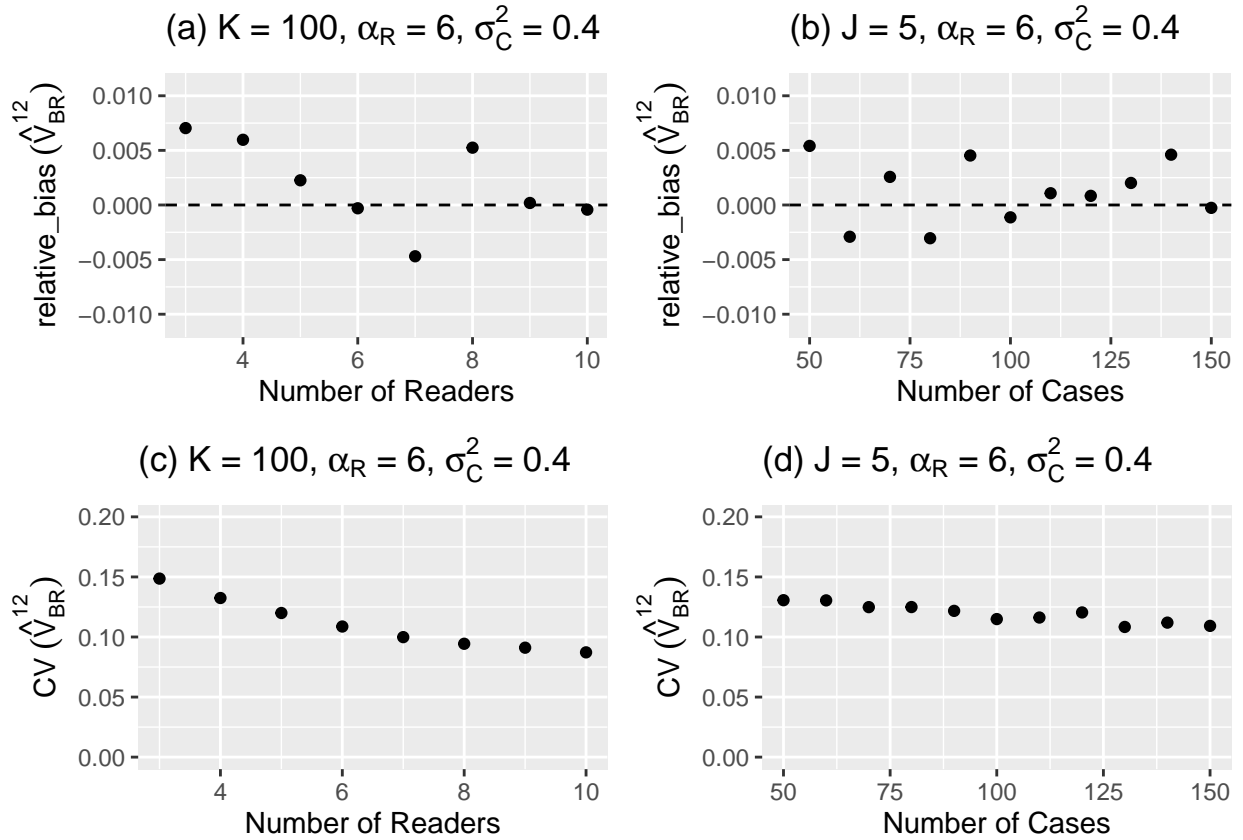


Figure 2 Relative bias and CV of variance estimation for BRBM difference changing over the size of the study

Figure 3 describes how the relative bias and CV of the variance estimate for BRBM differences change over the reader and case related parameters,  $\alpha_R$  and  $\sigma_C^2$ . There are 5 readers and 100 cases in each MRMC study. From Figure 3(a) we can see that most of the points lie within the range of  $\pm 1\%$ . The line with solid dot ( $\alpha_R = 3$ ) has the largest fluctuation comparing to other lines. This is the setting that involves the

largest reader variability, so it is expected that the variance of differences, especially for the between-reader difference, will be less stable than the results for other parameter settings. This can also be confirmed in Figure 3(b); the line for  $\alpha_R = 3$  (line with solid dot) has the highest CV for each  $\sigma_C^2$ . Generally, as  $\alpha_R$  increases, the CV drops, since the reader variability in the study decreases. As we increase the case related parameter  $\sigma_C^2$ , the case variability increases and it will dominate the variability from the readers, since the number of cases is much larger than that for the readers. Therefore, the lines in Figure 3(b) will be closer to each other when the  $\sigma_C^2$  increases from 0.1 to 1.

```
# Third Parameter set

result.validMRMCVarEstimate.parameter$alpha_R <-
  as.factor(result.validMRMCVarEstimate.parameter$alpha_R)
## relative bias
p1 <- ggplot(result.validMRMCVarEstimate.parameter,
             aes(x=sigma_C, y = BR_var_relative_bias, group = alpha_R))+
  geom_point(aes(shape = alpha_R), size = 3) +
  geom_line() +
  geom_hline(yintercept = 0, linetype = 'dashed') +
  xlab(expression(sigma[C]^2))+
  ylab(expression(relative_bias~(hat(V)[BR]^12))) +
  theme(legend.position = 'none')+
  ylim(-0.013,0.013)+
  ggtitle("(a) J=5, K = 100")
## CV
p2 <- ggplot(result.validMRMCVarEstimate.parameter,
             aes(x=sigma_C, y = BR_CV, group = alpha_R))+
  geom_point(aes(shape = alpha_R), size = 3) +
  geom_line() +
  xlab(expression(sigma[C]^2))+
  ylab(expression(CV~(hat(V)[BR]^12))) +
  labs(shape = expression(alpha[R]))+
  ylim(0,0.3)+
  ggtitle("(b) J=5, K = 100")

do.call("grid.arrange", c(list(p1,p2), ncol=2))
```

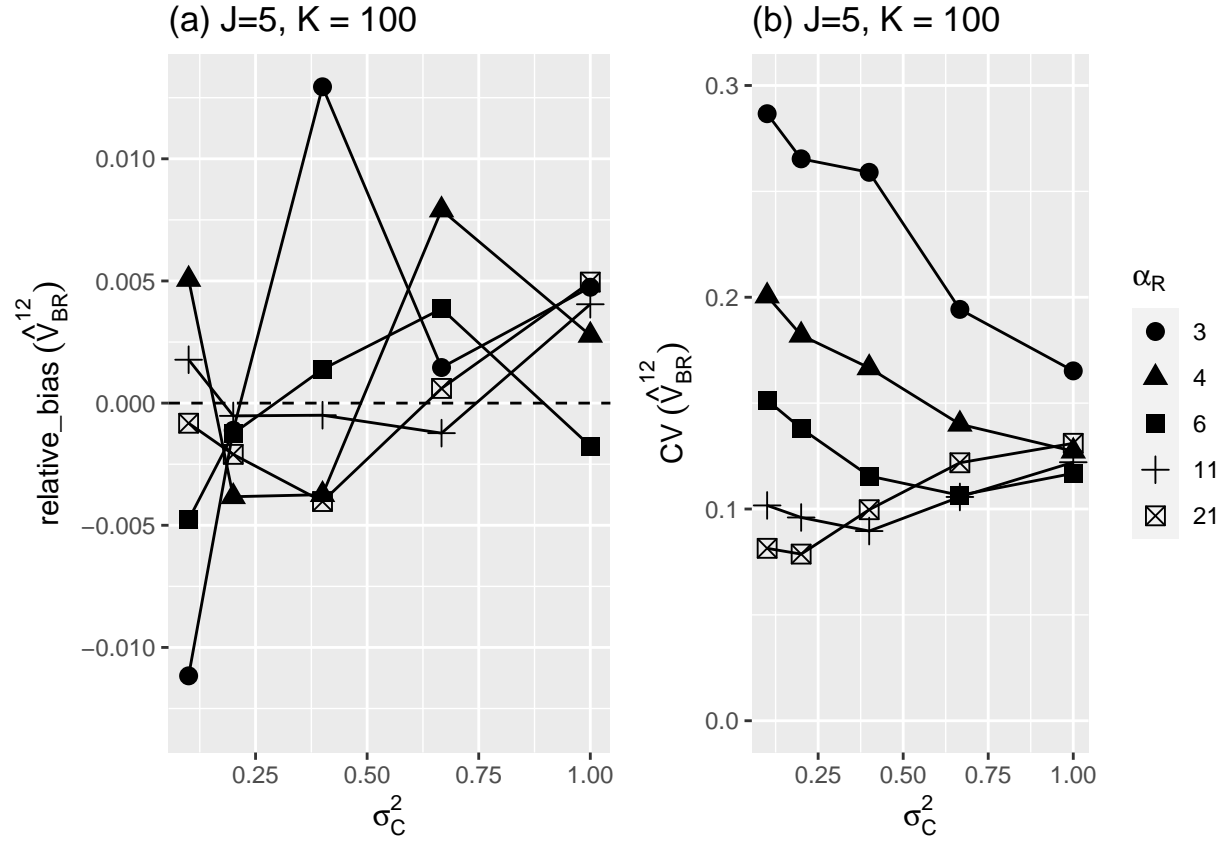


Figure 3 Relative bias and CV of variance estimation for BRBM difference changing over  $\alpha_R$  and  $\sigma_C^2$