**Supplemental Materials**

We varied the population variances ratio

($\sigma\_{1}^2/\sigma\_{2}^2$ = 1/5, 1/2, 1, 2, or 5; for each, the smaller

$\sigma^2$ = 2), the sample sizes (smallest \textit{n} = 20, 50, or 100), and the sample size ratio ($n\_{1}/n\_{2}$ = 1, 2/3, or 1/2).

Additionally, we varied the difference in group means based on Cohen's d

values of 0, .2, .5, and .8 when variances were equal. Importantly, Cohen's d

assumes that the population variances are equal and pools the group variances

just like Student's t test, and it is not well-defined when variances are unequal. Therefore, we used the same differences in group means when variances were unequal. Because we changed the variance ratio by increasing the variance of one group, the mean differences could be considered to represent smaller effects when variances are unequal.

For each condition (i.e., combination of variance ratio, smaller sample size, and sample size ratio), we set the seed of the random number generator to the same value and ran 10,000 simulations. When we report conditions with equal sample sizes and variance ratios of 2 and 5, they are identical to their symmetric conditions with equal sample sizes and variance ratios of 1/2 and 1/5.

**% Because the coverage**

**% probability of a confidence interval is not influenced by the effect size, we**

**% only show the coverage probabilities when the null hypothesis is true.**

**When there was a true cross-over interaction, as shown in Figure 7, the difference in means between groups B and C vs. groups A and D was the same as the simulations of two groups with a medium-sized effect.**

**\subsection{Comparing the Degrees of Freedom of the Two Tests}**

**A different way to decide which test to use might be to compare the**

**df. As the sample variances become unequal, the df**

**of Welch's t test decrease but the df of Student's t test**

**stay the same. Although the df of the two tests typically**

**differ slightly due to sampling error, there might be a point where the difference**

**reliably signals that the population variances**

**are unequal. For example, perhaps when Welch's df**

**are 90\% of Student's, it typically means the variances are unequal.**

**We examined whether this ratio**

**could provide a heuristic for deciding that**

**the variances are unequal and you should use Welch's t test.**

**To our knowledge, this possibility has never**

**been examined. We used the formulas for the**

**df from the two tests to see how the ratio is affected by changes in**

**the group variances and sample sizes.**

**The top Figure 1 shows how the df ratio is affected by changes in the group variances when sample sizes are equal. When the variances are**

**equal, the df ratio is equal to one. When**

**one group has twice the variance of the other, the ratio drops**

**to .90. A useful heuristic might be to assume that the variances are unequal**

**when the ratio falls below 96\%.**

**\begin{figure\*}**

**<<dfratiosDiffvars, echo=FALSE, collapse=TRUE, fig=TRUE, height=10>>=**

**load('dfs.list.RData')**

**df.ratio.Ns <- function(n1, n2, var1, var2) {**

**welch.num <- (var1/n1 + var2/n2)^2**

**welch.denom <- var1^2/(n1^2\*(n1-1)) + var2^2/(n2^2\*(n2-1))**

**classic.df <- n1+n2-2**

**(welch.num/welch.denom)/classic.df**

**}**

**df.ratio.defaults <- list('n1'=50,'n2'=50,'var1'=2,'var2'=2)**

**partial.df <- function(var = 'a', params=df.ratio.defaults){**

**params[[var]]=as.name('x')**

**function(x)do.call(df.ratio.Ns, params)**

**}**

**dfchange.var <- ggplot(data.frame(x=seq(2,10,.1)), aes(x)) +**

**stat\_function(fun=partial.df(var='var2',**

**params=list('n1'=50,'n2'=50,'var1'=2,'var2'=2))) + #NOTE: the sample sizes**

**#don't affect the df ratio**

**labs(y=bquote(frac('df'['Welch'],'df'['Student'])),**

**x=bquote(frac(sigma[1]^2,sigma[2]^2))) +**

**scale\_x\_continuous(breaks=c(2,4,6,8,10), labels=c(2,4,6,8,10)/2)**

**dfchange.Ns <- ggplot(data.frame(x=seq(50,100,.1)), aes(x)) +**

**stat\_function(fun=partial.df(var='n2',**

**params=list('n1'=50,'n2'=50,'var1'=2,'var2'=2))) +**

**labs(y=bquote(frac('df'['Welch'],'df'['Student'])),**

**x=bquote(frac('n'[1],'n'[2]))) +**

**scale\_x\_continuous(breaks=c(50,60,70,80,90,100),**

**labels=c(50,60,70,80,90,100)/50)**

**dfchange.both <- ggplot(data.frame(x=seq(2,10,.1)), aes(x)) +**

**stat\_function(fun=partial.df(var='var2',**

**params=list('n1'=75,'n2'=50,'var1'=2,'var2'=2)), aes(linetype = 'n2')) +**

**stat\_function(fun=partial.df(var='var2',**

**params=list('n1'=50,'n2'=75,'var1'=2,'var2'=2)), aes(linetype = 'n1/2')) +**

**scale\_linetype\_manual(name = 'Sample size ratio',**

**values=c('solid', 'dashed'),**

**labels=c(bquote(frac('n'[1],'n'[2])~'='~frac(3,2)),**

**bquote(frac('n'[1],'n'[2])~'='~frac(2,3)))**

**) +**

**labs(y=bquote(frac('df'['Welch'],'df'['Student'])),**

**x=bquote(frac(sigma[1]^2,sigma[2]^2))) +**

**scale\_x\_continuous(breaks=c(2,4,6,8,10), labels=c(2,4,6,8,10)/2)**

**layout <- matrix(c(1,2,3), nrow=3, byrow=TRUE)**

**multiplot(plotlist=list(dfchange.var, dfchange.Ns, dfchange.both),**

**layout=layout)**

**@**

**\textit{Figure 1.} Degrees of freedom ratio when sample sizes are equal and**

**variances are unequal (top), when variances are equal and sample sizes are**

**unequal (middle), and when both variances and sample sizes are unequal (bottom).**

**\end{figure\*}**

**But look at what happens when the variances are equal but the sample**

**sizes are different (middle of Figure 1). The df ratio decreases as the difference in sample sizes increases**

**The 96\% heuristic would lead you to incorrectly**

**conclude that the variances are unequal when only the sample sizes are unequal.**

**%' \begin{figure}**

**%' <<dfratiosDiffNratios, echo=FALSE, collapse=TRUE, fig=TRUE>>=**

**%'**

**%' ggplot(data.frame(x=seq(50,100,.1)), aes(x)) +**

**%' stat\_function(fun=partial.df(var='n2',**

**%params=list('n1'=50,'n2'=50,'var1'=2,'var2'=2))) +**

**%' labs(y=bquote(frac('df'['Welch'],'df'['Student'])),**

**%x=bquote(frac('n'[1],'n'[2]))) +**

**%' scale\_x\_continuous(breaks=c(50,60,70,80,90,100),**

**%labels=c(50,60,70,80,90,100)/50)**

**%'**

**%' @**

**%'**

**%' \textit{Figure 4.} Degrees of freedom ratio when sample sizes are unequal**

**%and variances are equal.**

**%' \end{figure}**

**The picture becomes even more complicated when both the sample sizes and**

**the variances are unequal (bottom of Figure 1). In this case, the change in the df ratio depends on which group has the**

**larger variance. As the variance of the smaller group**

**increases, there is an immediate drop in the ratio that begins to level off. However, as the variance of the larger group increases,**

**it initially counteracts the**

**unequal sample sizes, and the ratio increases and approaches one before it drops**

**again. Due to the difference in sample sizes, a 96\% heuristic could mislead you**

**into thinking the variances are equal when they are actually 2-3 times**

**different from each other.**

**%' \begin{figure}**

**%' <<dfratiosDiffvarsDiffNratios, echo=FALSE, collapse=TRUE, fig=TRUE>>=**

**%' ggplot(data.frame(x=seq(2,10,.1)), aes(x)) +**

**%' stat\_function(fun=partial.df(var='var2',**

**%params=list('n1'=75,'n2'=50,'var1'=2,'var2'=2)), aes(colour='n2')) +**

**%' stat\_function(fun=partial.df(var='var2',**

**%params=list('n1'=50,'n2'=75,'var1'=2,'var2'=2)), aes(colour='n1/2')) +**

**%' scale\_colour\_manual(values=c('blue', 'red'),**

**%labels=c(bquote(frac('n'[1],'n'[2])~'='~frac(3,2)),**

**%bquote(frac('n'[1],'n'[2])~'= '~frac(2,3)))) +**

**%' labs(y=bquote(frac('df'['Welch'],'df'['Student'])),**

**%x=bquote(frac(sigma[1]^2,sigma[2]^2))) +**

**%' scale\_x\_continuous(breaks=c(2,4,6,8,10), labels=c(2,4,6,8,10)/2)**

**%' @**

**%'**

**%' \textit{Figure 5.} Degrees of freedom ratio when sample sizes are unequal**

**%and variances are unequal.**

**%' \end{figure}**

**In short, there doesn't seem to be a simple heuristic based on the df ratio.**