Theoretical variance, as a function of population parameters

Delacre Marie[1]

[1] ULB

Theoretical variance, as a function of population parameters

The variance

Cohen's d_s

When variances are equal across populations.

When $\delta_{Cohen} = 0$. When the population effect size is zero, the variance of Cohen's d_s can be simplified as follows:

$$Var_{Cohen's\ d_s} = \frac{N(N-2)}{n_1 n_2 (N-4)}$$

The **variance** of Cohen's d_s is a function of total sample size (N) and the sample size allocation ratio ($\frac{n_1}{n_2}$):

- The larger the total sample size, the lower the variance. The variance tends to zero when the total sample size tends to infinity (see Figure 1);

- The further the sample sizes allocation ratio is from 1, the larger the variance (see Figure 2).

When $\delta_{Cohen} \neq 0$. While the variance of Cohen's d_s still depends on the total sample size and the sample sizes allocation ratio, it also depends on the population effect size (δ_{Cohen}). The larger the population effect size, the larger the variance. Note that the effect of the population effect size decreases when sample sizes increase, as:

$$\lim_{n_1 \to \infty} \left[\frac{df}{df-2} - \left(\frac{\sqrt{\frac{df}{2}} \times \Gamma\left(\frac{df-1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)} \right)^2 \right] = 0$$

$$\lim_{n_2 \to \infty} \left[\frac{df}{df-2} - \left(\frac{\sqrt{\frac{df}{2}} \times \Gamma\left(\frac{df-1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)} \right)^2 \right] = 0$$

$$\lim_{N \to \infty} \left[\frac{df}{df-2} - \left(\frac{\sqrt{\frac{df}{2}} \times \Gamma\left(\frac{df-1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)} \right)^2 \right] = 0$$

All these limits are illustrated in Figure 3.

Glass's d_s

When variances are equal across populations.

When $\delta_{Glass} = 0$. When the population effect size is zero, the variance of Glass's d_s can be simplified as follows:

$$Var_{Glass's\ d_s} = \frac{n_c - 1}{n_c - 3} \left(\frac{1}{n_c} + \frac{1}{n_e} \right)$$

In this configuration, the **variance** of Glass's d_s is a function of the sample sizes of both control (n_c) and experimental (n_e) groups as well as of the sample sizes allocation ratio $\left(\frac{n_1}{n_2}\right)$:

- The larger the sample sizes, the lower the variance (Figure 4);

The sample sizes ratio associated with the lowest variance is not exactly 1 (because of the term $\frac{df}{df-2}$, df depending only on n_c), but is very close to 1 (and the larger the total sample size, the closer to 1 is the sample sizes ratio associated with the lowest variance). The further from this sample size ratio, the larger the variance (see Figure 5).

When $\delta_{Glass} \neq 0$. While the variance of Glass's d_s still depends on the total sample size and the sample sizes allocation ratio, it also depends on the population effect size (δ_{Cohen}). The larger the population effect size, the larger the variance. However, the effect of the population effect size decreases when the control group increases. On the other side, the effect of the population effect size does *not* depend on the size of the experimental group.

$$\lim_{n_c \to \infty} \left[\frac{df}{df-2} - \left(\frac{\sqrt{\frac{df}{2}} \times \Gamma\left(\frac{df-1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)} \right)^2 \right] = 0$$

$$\lim_{n_e \to \infty} \left[\frac{df}{df-2} - \left(\frac{\sqrt{\frac{df}{2}} \times \Gamma\left(\frac{df-1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)} \right)^2 \right] \neq 0$$

These limits are illustrated in Figure 6.

Note: while the sample sizes ratio associated with the lowest variance was very close to 1 with a null population effect size, this is not true anymore when the population effect size is not zero. Indeed, because of the second term in the adddition, when computing the variance, one gives much more weight to the effect size of the control group (see Figure 7), and the larger the population effect size the truer. For example, when δ_{glass}= 4, the lowest variance will occure when n_c is approximately 3 times larger than n_e. When δ_{glass}= 7, the lowest variance will occure when n_c is approximately 5 times larger than n_e, etc.

When variances are unequal across populations, with equal sample sizes.

When $\delta_{Glass} = 0$. When the population effect size is zero, the variance of Glass's d_s can be simplified as follows:

$$Var_{Glass's\ d_s} = \frac{n-1}{n(n-3)} \left(1 + \frac{\sigma_e^2}{\sigma_c^2}\right)$$

Where $n = N/2$ = sample size of each group. The variance is therefore a function of the total sample size and the SD-ratio ($\frac{\sigma_c}{\sigma_e}$):

- The larger the total sample size, the lower the variance (See Figure 8);
- The larger the SD-ratio (i.e. the larger is σ_c in comparison with σ_e), the lower the variance (see Figure 9). However, the effect of the SD-ratio decreases when sample sizes increase, because $\lim_{n(=n_c=n_e) \to \infty} \left[\frac{df}{n(df-2)}\right] = 0$.

***When* $\delta_{Glass} \neq 0$.** While the variance of Glass's d_s still depends on the total sample size and the SD-ratio, it also depends on the population effect size (δ_{Glass}). The larger the population effect size, the larger the variance. However, the effect of the population effect size decreases when the control group increases, as previously explained and illustrated in Figure 6.

When variances are unequal across populations, with unequal sample sizes.

***When* $\delta_{Glass} = 0$.** When the population effect size is zero, the variance of Glass's d_s can be simplified as follows:

$$Var_{Glass's\ d_s} = \frac{n_c - 1}{n_c - 3}\left(\frac{1}{n_c} + \frac{\sigma_e^2}{n_e \sigma_c^2}\right)$$

The variance of Glass's d_s is therefore a function of the total sample size (N), the SD-ratio and the interaction between the sample sizes ratio and the SD-ratio:

- Whatever the SD and sample sizes pairing, increasing n_c and/or n_e will decrease the variance (see Figure 10);
    - The effect of the sample sizes ratio depends on the SD-ratio:
        * We previously mentioned that when $\sigma_c = \sigma_e$, the variance is minimized when sample sizes of both groups are almost identical (see Figure 5), meaning that it is more efficient, in order to reduce variance, to add subjects uniformly in both groups;
        * When $\sigma_e > \sigma_c$, more weight is given to n_e, meaning that it is more efficient, in order to reduce variance, to add subjects in the experiental group (n_e; see bottom plots in Figure 10);
        * When $\sigma_c > \sigma_e$, less weight is given to n_e, meaning that it is more efficient, in order to reduce variance, to add sujects in the control group (n_c; see top plots in Figure 10).

- Finally, there is also a main effect of the SD-ratio: the larger is σ_c in comparison with σ_e, the lower the variance, as we can observe in Figure 11. We can also notice that in Figure 10, the maximum variance is much larger in the two bottom plots (where $\sigma_c < \sigma_e$) than in the two top plots (where $\sigma_c > \sigma_e$).

Note that the effect of the SD-ratio, and the interaction effect between SD-ratio and sample sizes ratio decreases when the sample size of the control group increases (because $\frac{n_c-1}{n_c-3}$ get closer to 1).

When $\boldsymbol{\delta_{Glass} \neq 0}$**.** While the variance of Glass's d_s still depends on the total sample size, the SD-ratio and the interaction between the SD-ratio and the sample sizes ratio, it also depends on the population effect size (δ_{Glass}): the larger the population effect size, the larger the variance. However, the effect of the population effect size decreases when the sample size of the control group increases, as previously explained and illustrated in Figure 6.

Note: when the population effect size was null, when $\sigma_c < \sigma_e$, it was much more efficient to add subjects in the experimental group in order to reduce the variance (because much more weight were given to n_e). When $\delta_{Glass} \neq 0$, it is important to add subjects in both groups in order to reduce the variance (because $\frac{df}{df-2} - \left(\frac{\sqrt{\frac{df}{2}} \times \Gamma\left(\frac{df-1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)}\right)^2$ is only a function of the sample size of the control group). With huge population effect size, it is even always more important to add subjects in the control group (e.g. when $\delta_{Glass} = 30$).

Cohen's d'_s

When variances are equal across populations.

When $\boldsymbol{\delta'_{Cohen} = 0}$**.** When the population effect size is zero, the variance of Cohen's d'_s is computed as following :

$$Var_{Cohen's\ d'_s} = \frac{df}{df-2} \times \frac{N}{n_1 n_2}$$

with

$$df = \frac{4(n_1-1)(n_2-1)}{n_1+n_2-2}$$

.

In this configuration, the degrees of freedom as well as the variance of Cohen's d'_s depends on the total sample size (N) and the sample size allocation ratio $(\frac{n_1}{n_2})$:

- The further the sample size allocation ratio is from 1, the larger the variance (see Figure 12);
- The larger the total sample size, the lower the bias (see Figure 13).

When $\delta_{Glass} \neq 0$***.*** While the variance of Cohen's d'_s still depends on the total sample size and the sample sizes ratio, it also depends on the population effect size (δ'_{Cohen}): the larger the population effect size, the larger the variance. However, the effect of the population effect sizes decreases when the degrees of freedom increase (i.e. when sample sizes increase and/or the sample sizes ratio get closer to 1), as illustrated in Figure 14:

$$\lim_{df \to \infty} \left[\frac{df}{df-2} - \left(\frac{\sqrt{\frac{df}{2}} \times \Gamma\left(\frac{df-1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)} \right)^2 \right] = 0$$

When variances are unequal across populations, with equal sample sizes.

When $\delta'_{Cohen} = 0$***.*** When the population effect size is zero, the variance of Cohen's d'_s can be simplified as follows:

$$Var_{Cohen's\ d'_s} = \frac{df}{df-2} \times \frac{2}{n}$$

Where $n = N/2$=sample size of each group, and $df = \frac{(n-1)(\sigma_1^4+\sigma_2^4+2\sigma_1^2\sigma_2^2)}{\sigma_1^4+\sigma_2^4}$

In this configuration, the degrees of freedom as well as the variance of Cohen's d_s depends on the total sample size (N) and the SD-ratio $\left(\frac{\sigma_1}{\sigma_2}\right)$:

- The further the SD-ratio is from 1, the larger the variance (see Figure 15);
- The larger the total sample size, the lower the variance (see Figure 16).

Note: for a constant SD-ratio, the size of the variance does not matter (see Figure 17).

When $\delta'_{Cohen} \neq 0$. While the variance of Cohen's d'_s still depends on the total sample size and the SD-ratio, it also depends on the population effect size (δ'_{Cohen}): the larger the population effect size, the larger the variance. However, the effect of the population effect size decreases when the degrees of freedom increase (i.e. when the total sample size increases and/or the SD-ratio get closer to 1), as previously illustrated in Figure 14.

When variances are unequal across populations, with unequal sample sizes.

When $\delta'_{Cohen} = 0$. When the population effect size is zero, the variance of Cohen's d'_s can be simplified as follows:

$$Var_{Cohen's\ d'_s} = \frac{df}{df-2} \times \frac{2\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}{\sigma_1^2 + \sigma_2^2}$$

with $df = \frac{(n_1-1)(n_2-1)(\sigma_1^2+\sigma_2^2)^2}{(n_2-1)\sigma_1^4+(n_1-1)\sigma_2^4}$

In this configuration, the degrees of freedom are a function of the total sample size (N) and the interaction between sample sizes ratio $\left(\frac{n_1}{n_2}\right)$ and the SD-ratio $\left(\frac{\sigma_1}{\sigma_2}\right)$:

- The larger the total sample size, the lower the variance (illustration in Figure 18);
- The smallest variance always occurs when there is a positive pairing between variances and sample sizes, because one gives more weight to the smallest variance in the denominator of the df computation and in the numerator in the variance computation. Moreover, the further the SD-ratio is from 1, the further from 1 will also be the sample sizes ratio associated with the smallest variance (see Figure 19). This can be explained by splitting the numerator and the denominator in the DF computation (see the file "Theoretical Bias, as a function of population parameters").

Note: for a constant SD-ratio, the variance does not matter. (See Figure 20).

When $\delta'_{Cohen} \neq 0$***.*** While the variance of Cohen's d'_s still depends on the total sample size and the interaction between the sample sizes ratio and the SD-ratio, it also depends on the population effect size (δ'_{Cohen}): the larger the population effect size, the larger the variance. However, the effect of the population effect size decreases when the degrees of freedom increase (i.e. when the total sample size increases and/or when there is a positive pairing between the sample sizes ratio and the SD-ratio), as previously illustrated in Figure 14.

Shieh's d_s

When variances are equal across populations.

When $\delta_{Shieh} = 0$***.*** When the population effect size is zero, the variance of Shieh's d_s can be simplified as follows:

$$Var_{Shieh's\ d_s} = \frac{df}{(df-2)N}$$

With $df \approx \frac{N^2(n_1-1)(n_2-1)}{n_2^2(n_2-1)+n_1^2(n_1-1)}$. In this configuration, the degrees of freedom as well as the variance of Shieh's d_s depend on the total sample size (N) and the sample size allocation ratio ($\frac{n_1}{n_2}$):

- The further the sample size allocation ratio is from 1, the larger the variance (see Figure 21);

- The larger the total sample size, the lower the variance, whatever the sample sizes ratio is constant (see Figure 22) or not (see Figure 23).

Note: in "Theoretical Bias, as a function of population parameters", we noticed that increasing the sample sizes ratio when adding subjects in only one group could decrease the degrees of freedom. However, due to the "N" in the denominator in the variance computation, even when degrees of freedom decrease due to the fact that one adds subjects only in one group, the variance still decreases (because the denominator in the variance computation increases; see Figure 23).

When $\delta_{Shieh} \neq 0$**.** While the variance of Shieh's d_s still depends on the total sample size and the sample sizes ratio, it also depends on the population effect size (δ_{Shieh}): the larger the population effect size, the larger the variance. However, the effect of the population effect sizes decreases when the degrees of freedom increase (i.e. when sample sizes increase (without increasing the sample sizes ratio) and/or the sample sizes ratio get closer to 1), as previously illustrated in Figure 14.

When variances are unequal across populations, with equal sample sizes.

When $\delta_{Shieh} = 0$**.** When the population effect size is zero, the variance of Shieh's d_s can be simplified as follows:

$$Var_{Shieh's\ d_s} = \frac{df}{(df-2)N}$$

with $df \approx \frac{(\sigma_1^2+\sigma_2^2)^2 \times (n-1)}{\sigma_1^4+\sigma_2^4}$. In this configuration, the degrees of freedom as well as the variance of Shieh's d_s depend on the total sample size (N) and the SD-ratio ($\frac{\sigma_1}{\sigma_2}$).

- The further the SD-ratio is from 1, the larger the variance (see Figure 25);
- The larger the total sample size, the lower the variance (see Figure 26).

Note: for a constant SD-ratio, the size of the variance does not matter (see Figure 27).

When $\delta_{Shieh} \neq 0$**.** While the variance of Shieh's d_s still depends on the total sample size and the SD-ratio, it also depends on the population effect size (δ_{Shieh}): the larger the population effect size, the larger the variance. However, the effect of the population effect sizes decreases when the degrees of freedom increase (i.e. when sample sizes increase and/or the SD-ratio get closer to 1), as previously illustrated in Figure 14.

When variances are unequal across populations, with unequal sample sizes.

When $\delta_{Shieh} = 0$. When the population effect size is zero, the variance of Shieh's d_s can be simplified as follows:

$$Var_{Shieh's\ d_s} = \frac{df}{(df-2)N}$$

With $df \approx \frac{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^2}{\frac{(\sigma_1^2/n_1)^2}{n_1-1}+\frac{(\sigma_2^2/n_2)^2}{n_2-1}}$

In this configuration, the degrees of freedom as well as the variance of Shieh's d_s depend on the sample sizes (n_1 and n_2) and the interaction between the sample sizes ratio ($\frac{n_1}{n_2}$) and the SD-ratio ($\frac{\sigma_1}{\sigma_2}$):

- The larger the total sample size, the lower the variance, whatever the sample sizes ratio is constant (see Figure 28) or not (see Figure 29).

Note: When variances were equal across populations, adding subjects only in the first group had the same impact on the variance than adding subjects only in the second group (see Figure 23). When variances are unequal across groups, this is not true anymore (see Figure 29).

- The smallest variance always occurs when there is a positive pairing between variances and sample size. Moreover, the further the SD-ratio is from 1, the further from 1 will also be the sample sizes ratio associated with the smallest var (see Figure 30).

Moreover, for a constant SD-ratio, the variances don't matter (See Figure 31).

When $\delta_{Shieh} \neq 0$. While the variance of Shieh's d_s still depends on the total sample size and the interaction between the sample sizes ratio and the SD-ratio, it also depends on the population effect size (δ_{Shieh}): the larger the population effect size, the larger the variance. However, the effect of the population effect size decreases when the degrees of freedom increase (i.e. . . .), as previously illustrated in Figure 14.

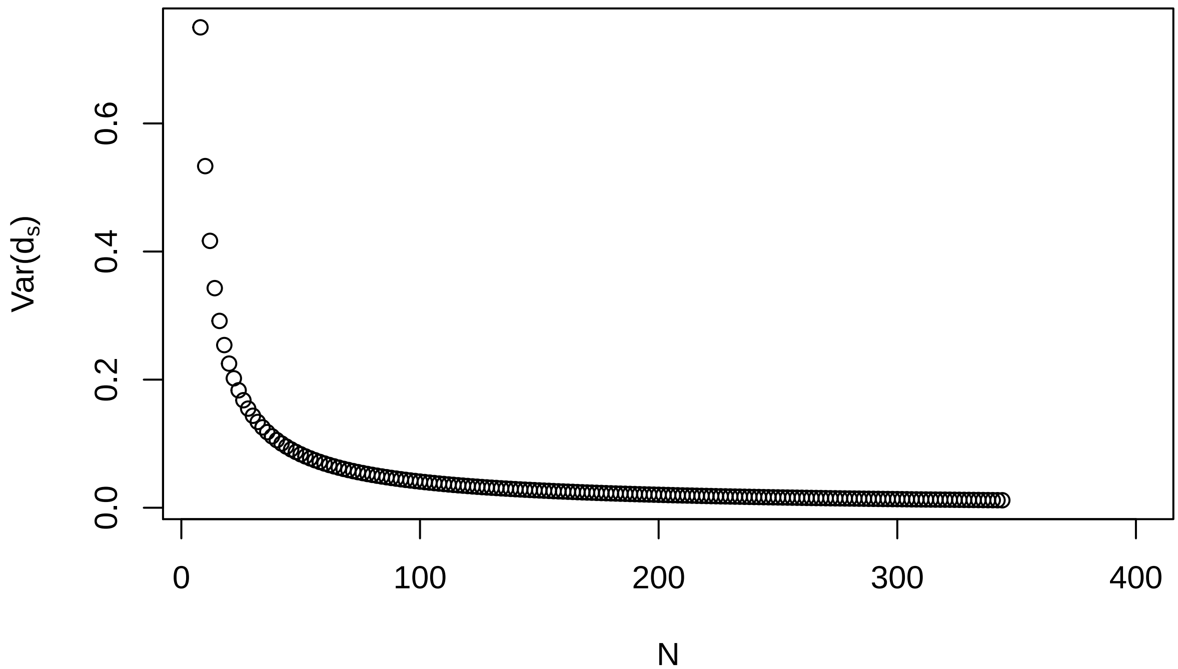

Figure 1. Variance of Cohen's d_s, when variances are equal across groups, as a function of the total sample size (N).

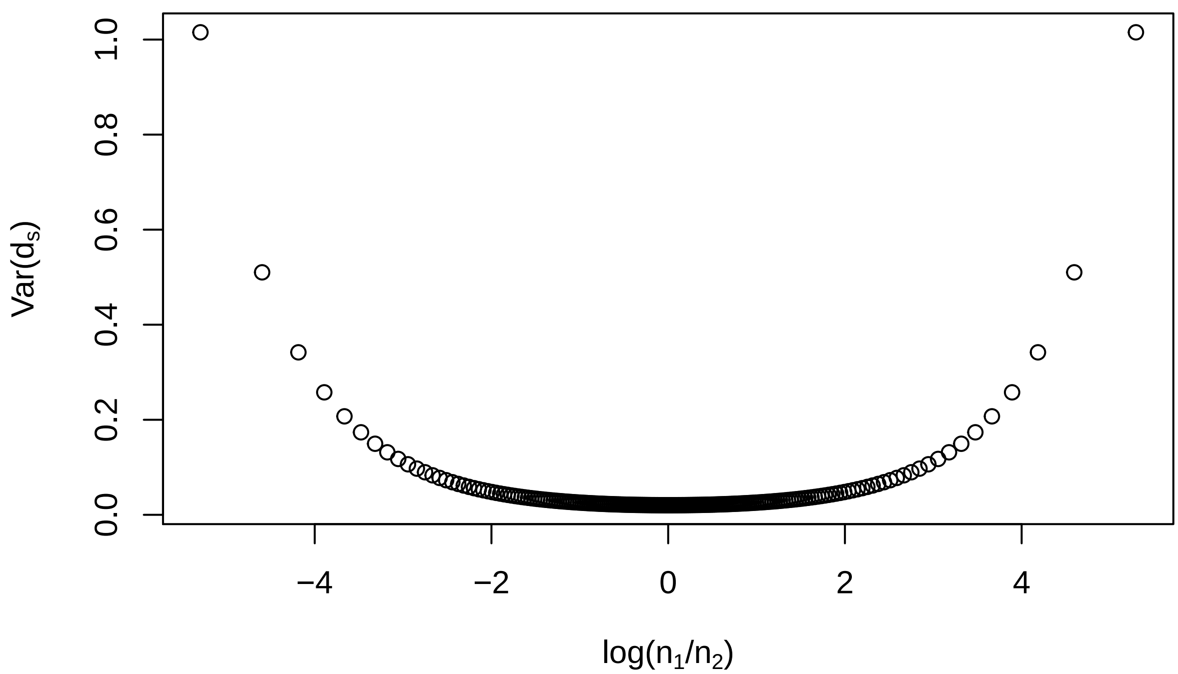

Figure 2. Variance of Cohen's d_s, when variances are equal across groups, as a function of the logarithm of the sample sizes ratio ($log\left(\frac{n_1}{n_2}\right)$).

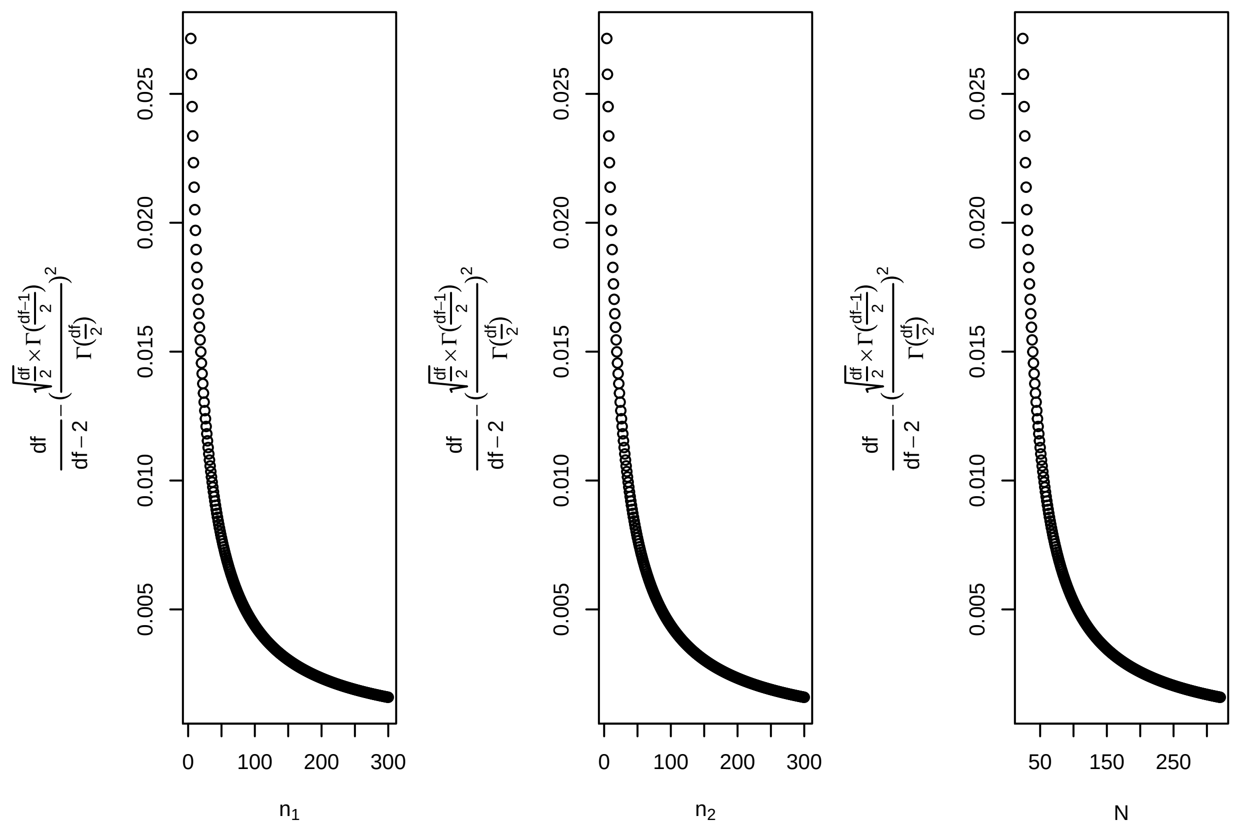

Figure 3. Effect size moderator, when computing the variance of COhen's d_s, as a function of n_1 (left), n_2 (center) and $N = n_1 + n_2$ (right).

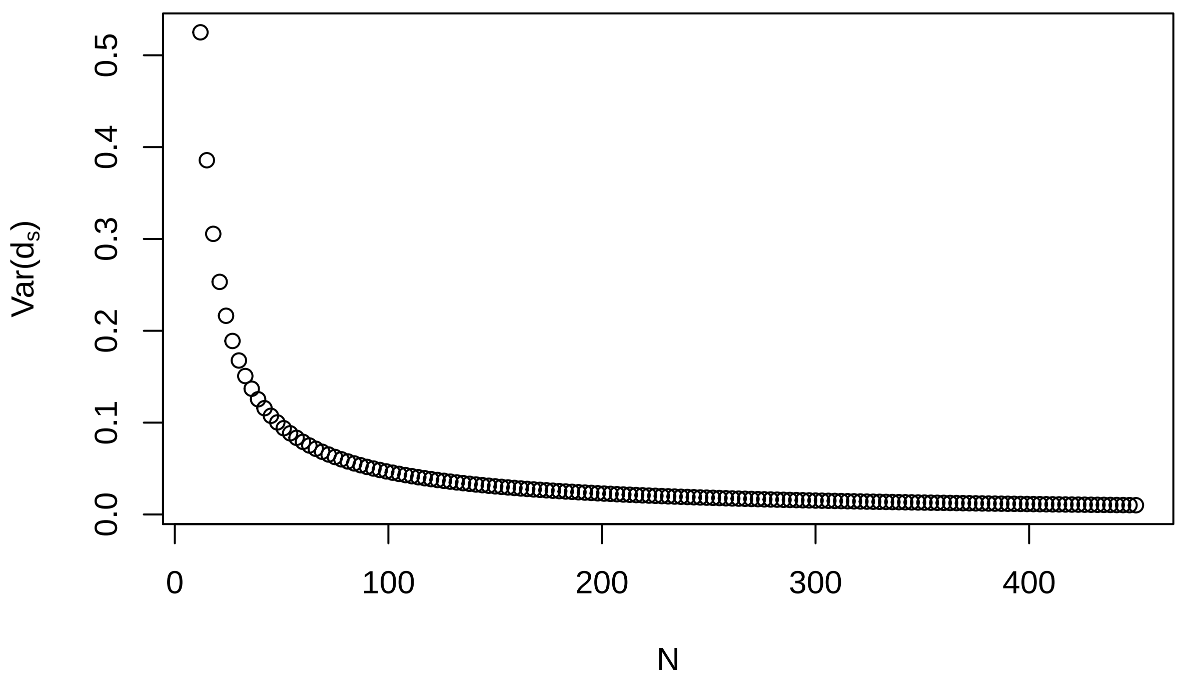

Figure 4. Variance of Glass's d_s, when variances are equal across groups, as a function of the total sample size (N).

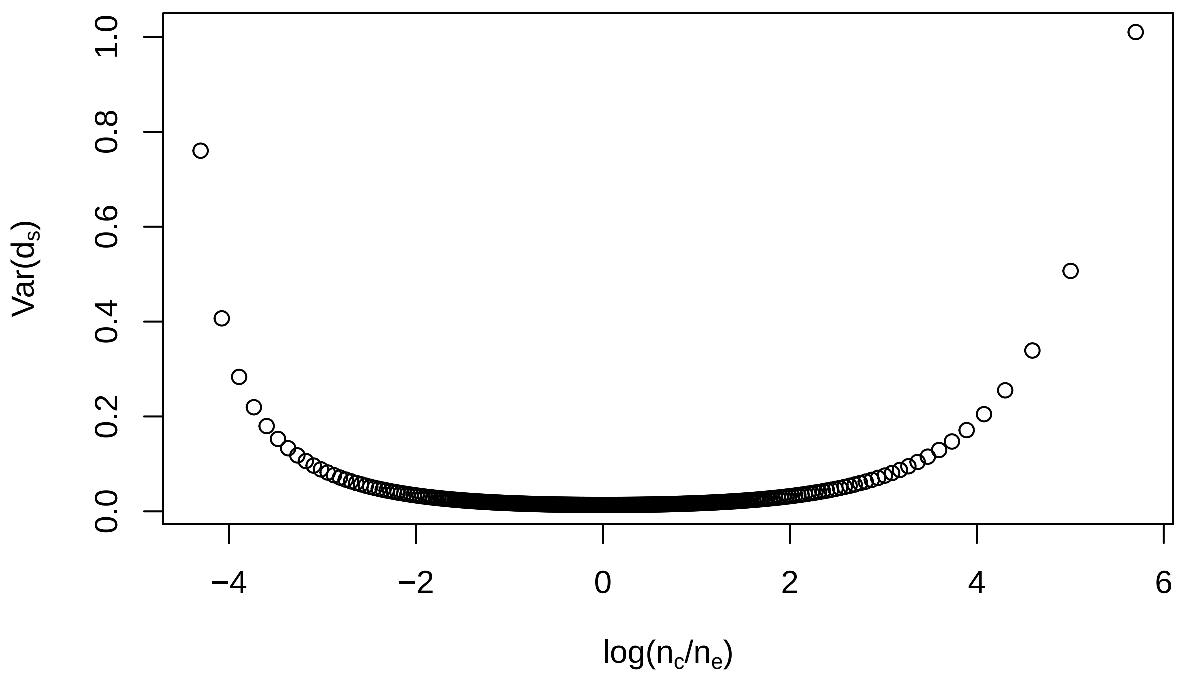

Figure 5. Variance of Glass's d_s, when variances are equal across groups, as a function of the logarithm of the sample sizes ratio ($log\left(\frac{n_c}{n_e}\right)$).

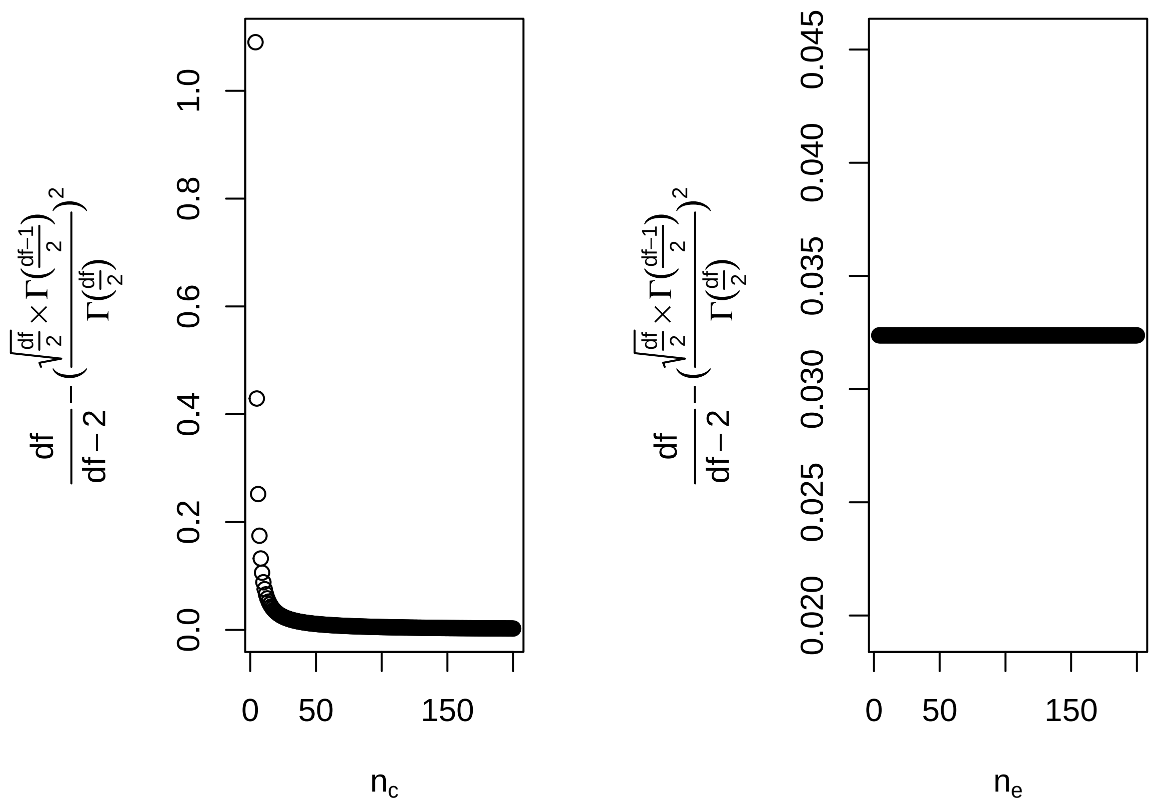

Figure 6. Effect size moderator, when computing the variance of Glass's d_s, as a function of the size of the control group (left) and experimental group (right).

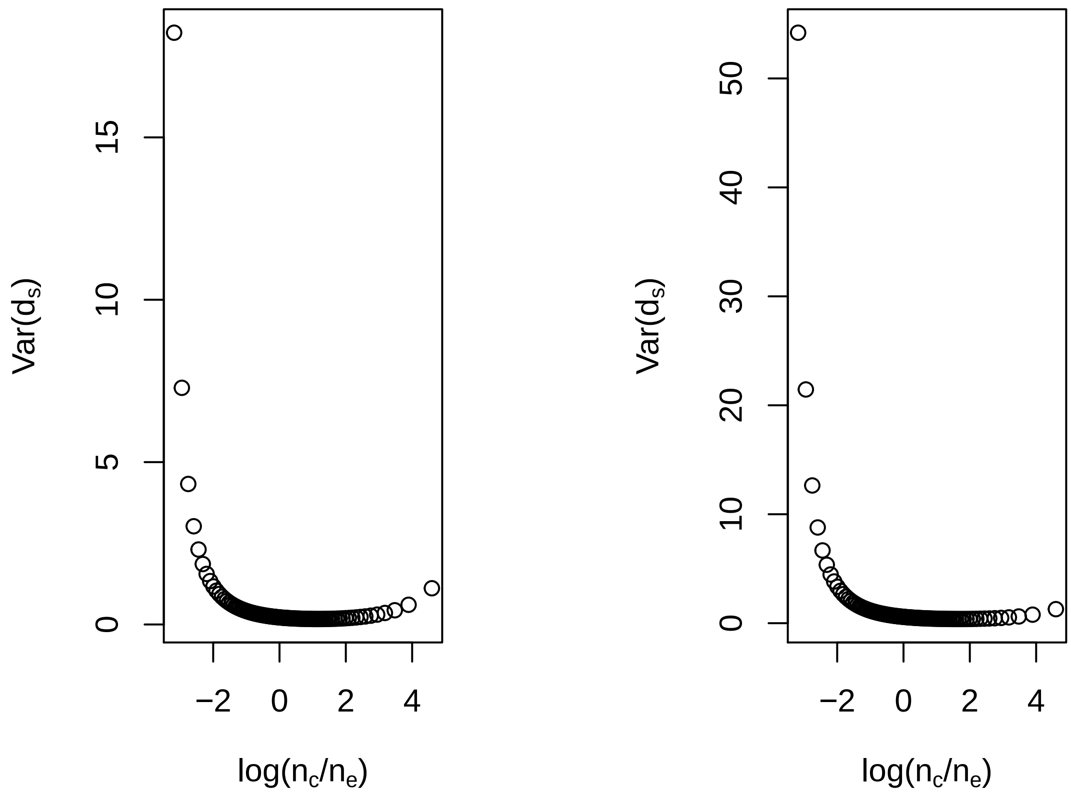

Figure 7. Variance of Glass's d_s, when variances are equal across groups, as a function of the logarithm of the sample sizes ratio ($log\left(\frac{n_c}{n_e}\right)$) when δ_{Glass} equals 4 (left) or 7 (right).

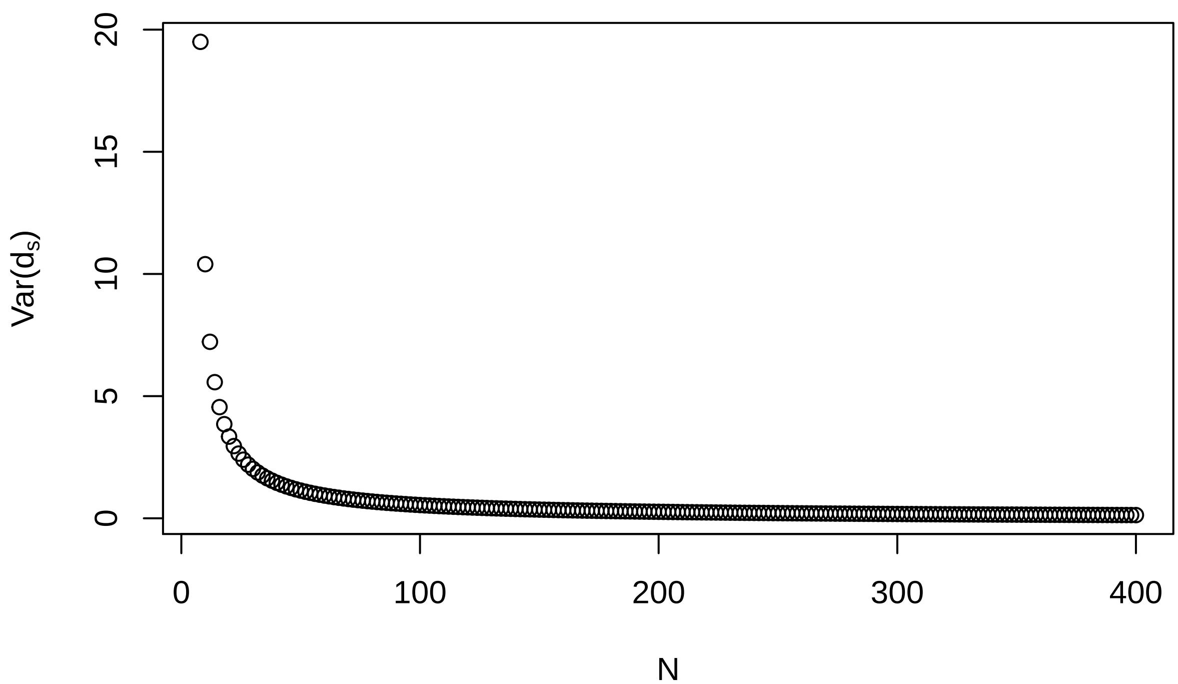

Figure 8. Variance of Glass's d_s, when variances are unequal across groups and sample sizes are equal, as a function of the total sample sizes (N).

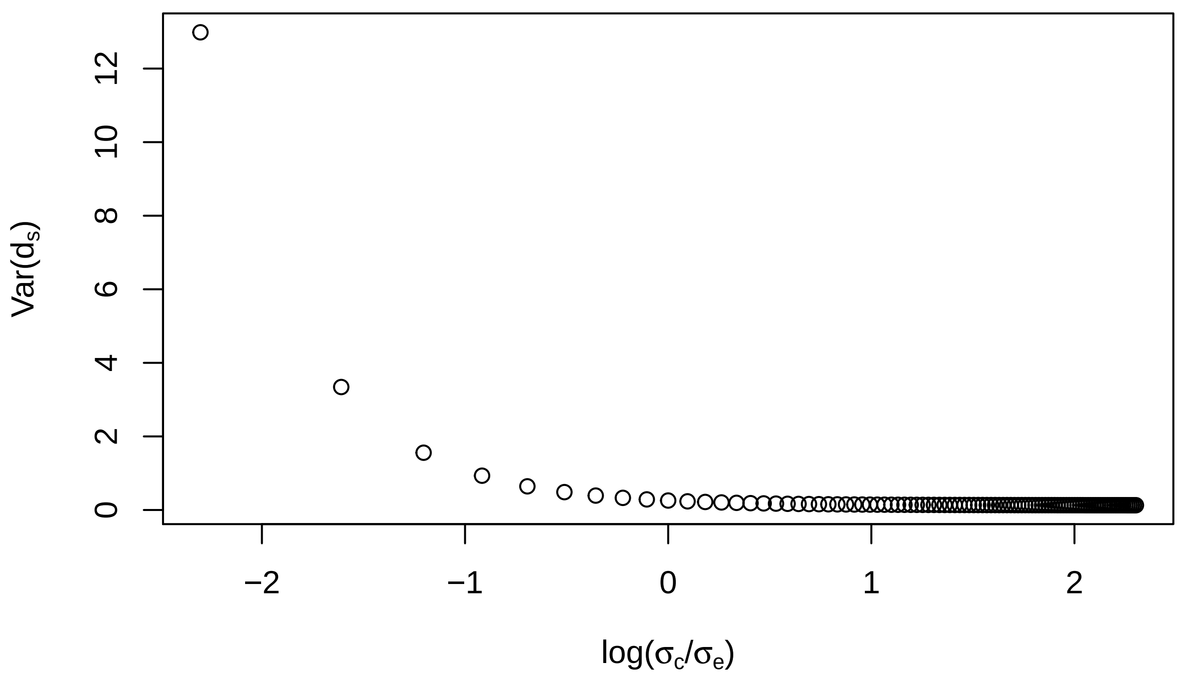

Figure 9. Variance of Glass's d_s, when variances are unequal across groups and sample sizes are equal, as a function of the logarithm of the SD-ratio ($log\left(\frac{\sigma_c}{\sigma_e}\right)$).

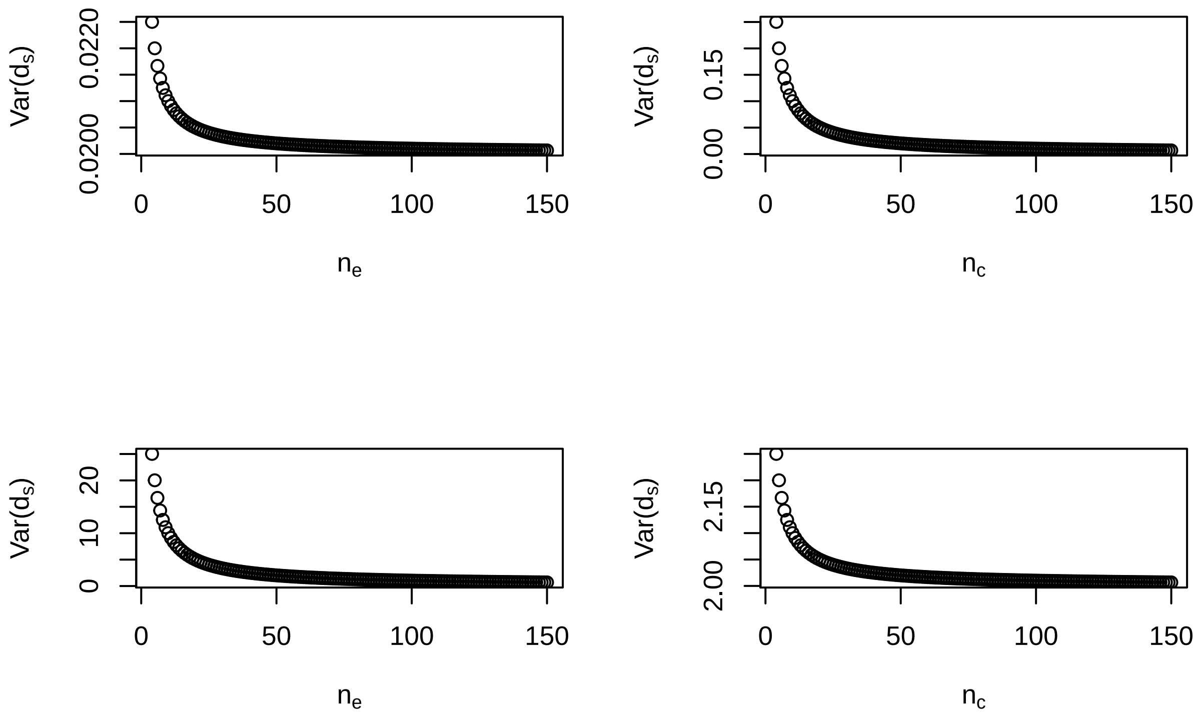

Figure 10. Variance of Glass's d_s, when variances and sample sizes are unequal across groups, as a function of the total sample sizes, when increasing only the control (right) or the experimental (left) group, when $\sigma_c > \sigma_e$ (top plots) or $\sigma_c < \sigma_e$ (bottom plots).

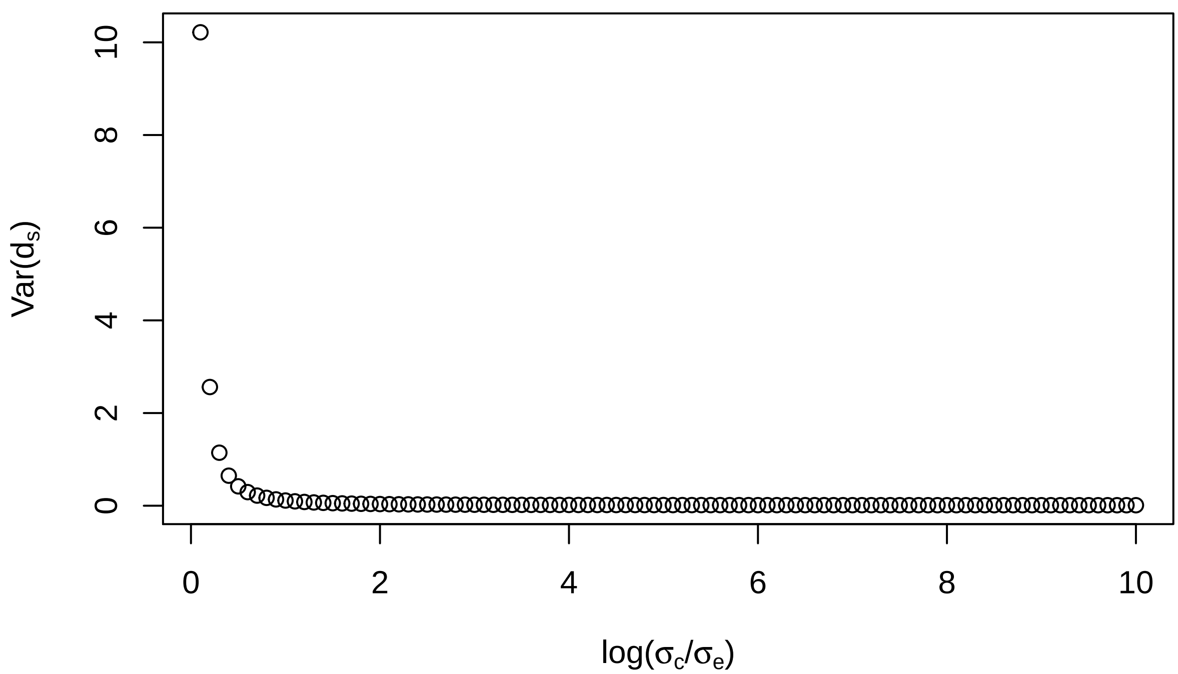

Figure 11. Variance of Glass's d_s, when sample sizes and variances are unequal across groups, as a function of the logarithm of the SD-ratio ($log\left(\frac{\sigma_c}{\sigma_e}\right)$).

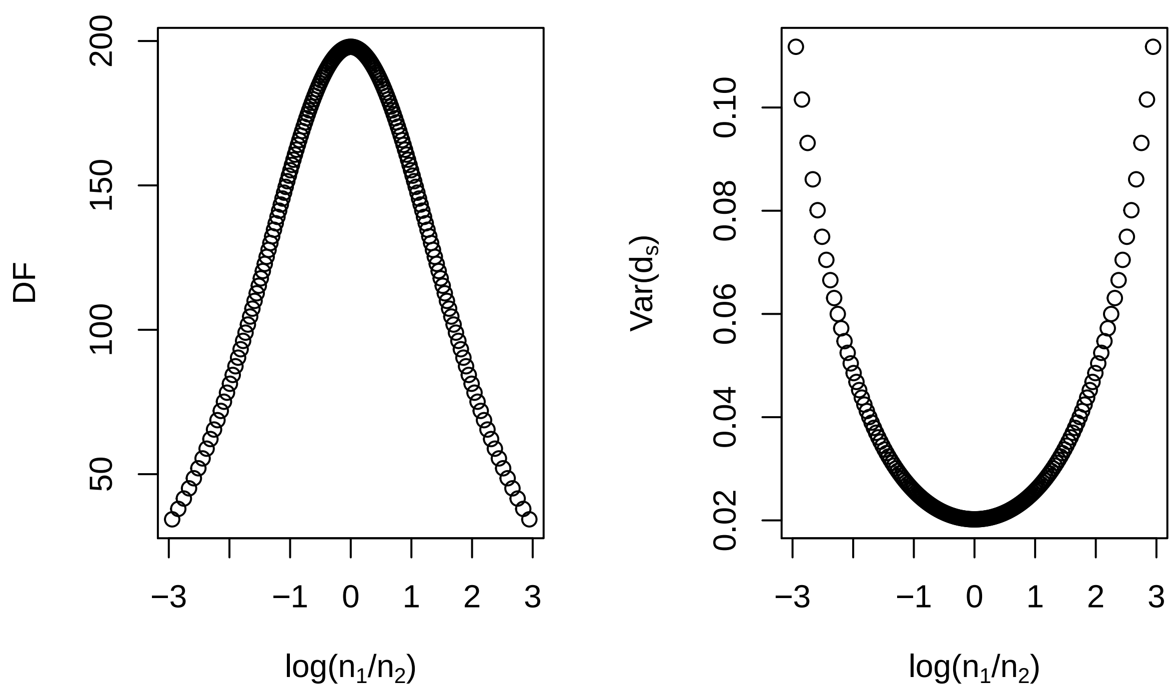

Figure 12. Variance of Cohen's d'_s when variances are equal across groups, as a function of the logarithm of the sample sizes ratio ($log\left(\frac{n_1}{n_2}\right)$).

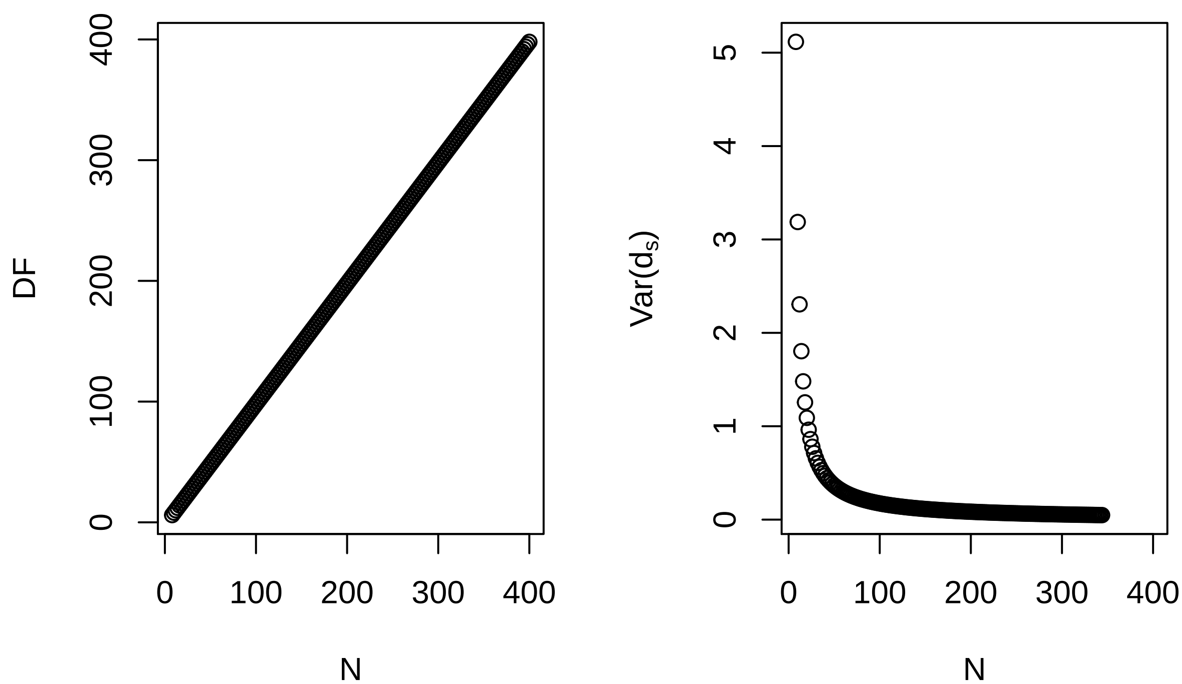

Figure 13. Variance of Cohen's d_s' when variances are equal across groups, as a function of the total sample size (N).

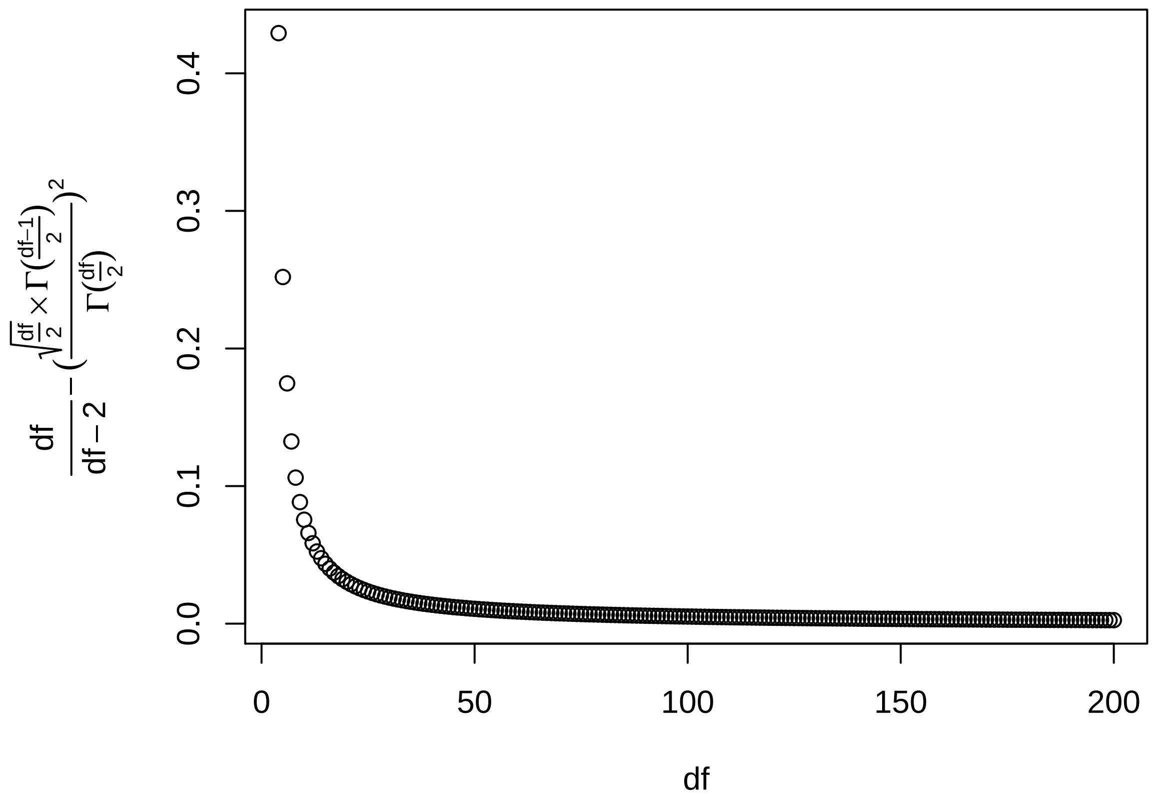

Figure 14. Effect size moderator, as a function of the degrees of freedom.

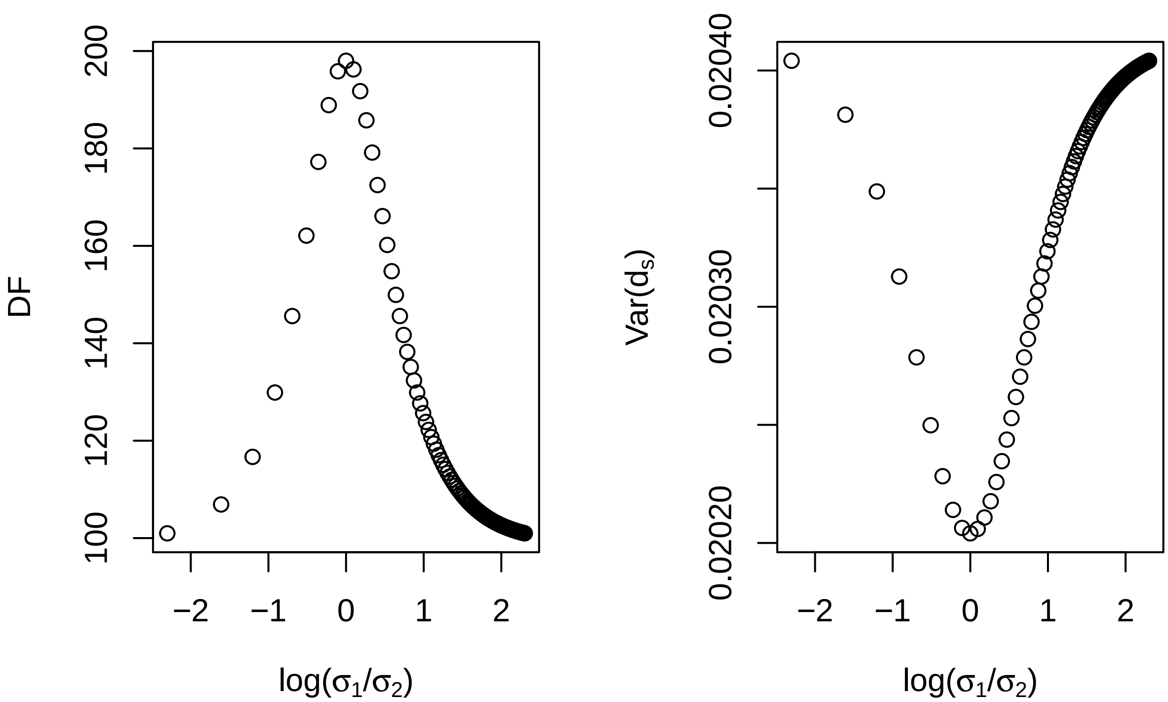

Figure 15. Variance of Cohen's d'_s when variances are unequal across groups and sample sizes are equal, as a function of the logarithm of the *SD*-ratio ($log\left(\frac{\sigma_1}{\sigma_2}\right)$).

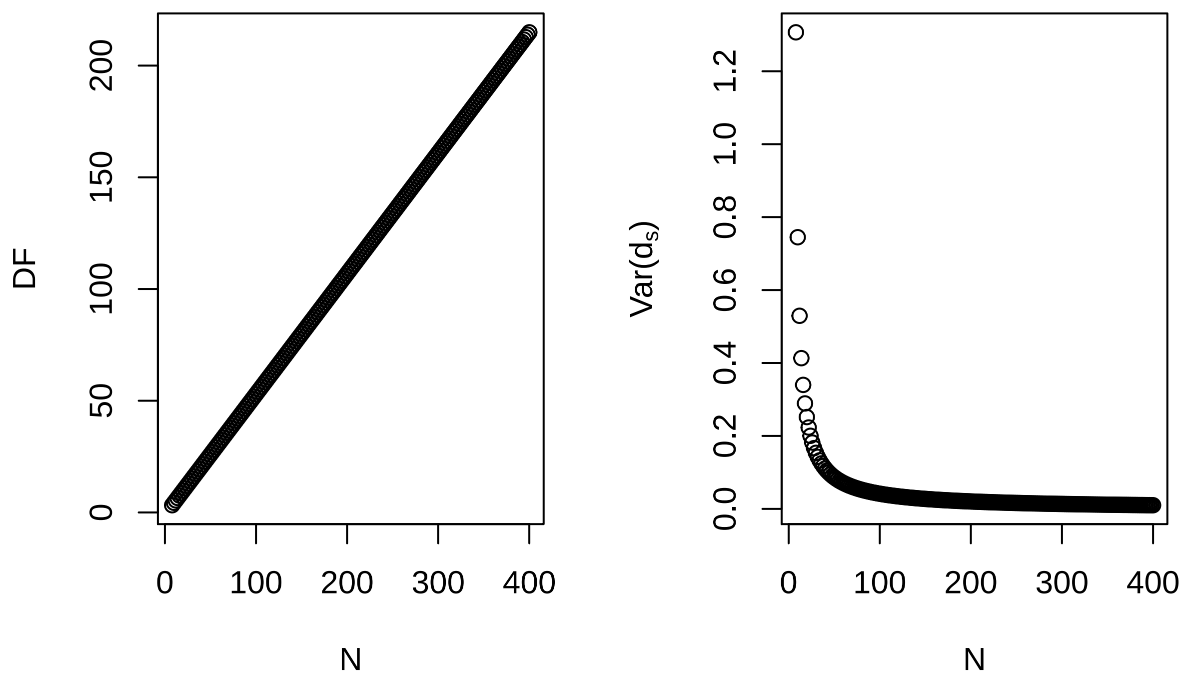

Figure 16. Variance of Cohen's d'_s when variances are unequal across groups and sample sizes are equal, as a function of the total sample size (N).

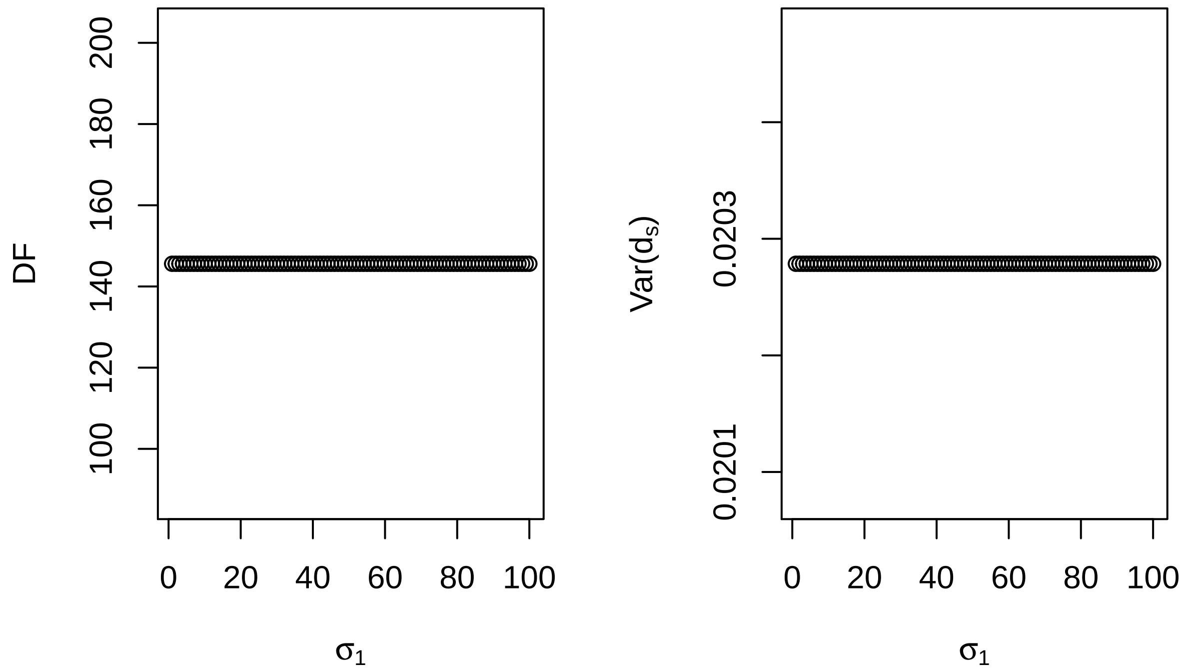

Figure 17. Variance of Cohen's d'_s, when variances are unequal across groups and sample sizes are equal, as a function of σ_1 and σ_2, for a constant *SD*-ratio.

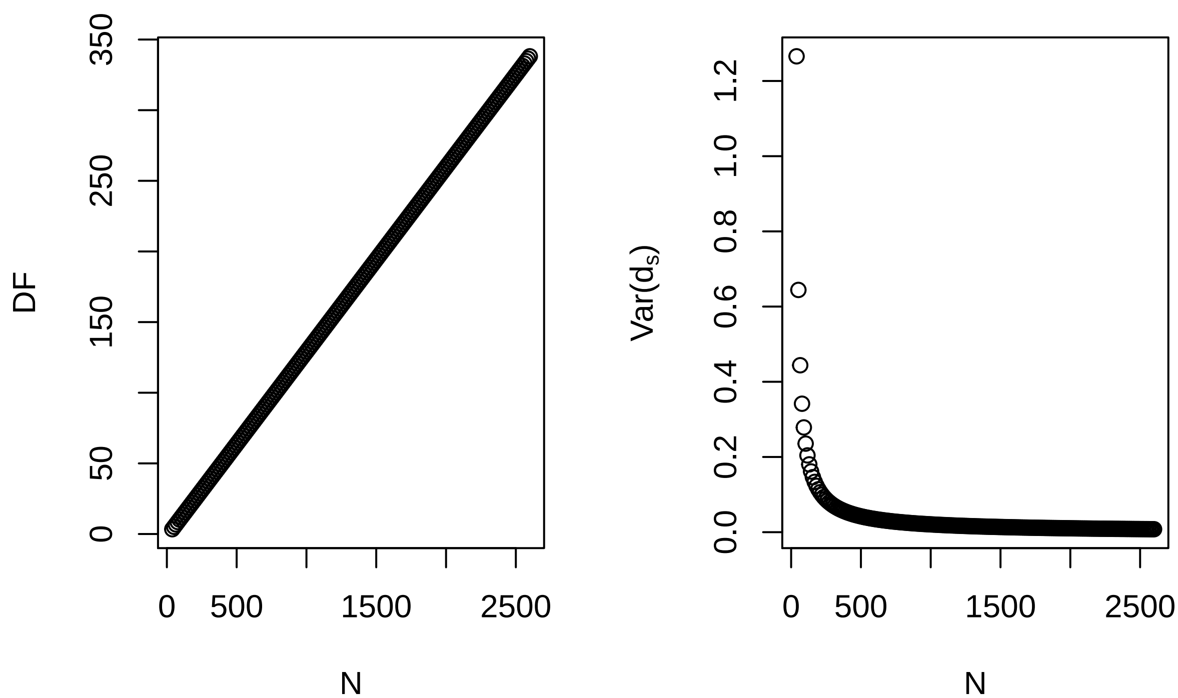

Figure 18. Variance of Cohen's d_s' when variances and sample sizes are unequal across groups, as a function of the total sample size (N).

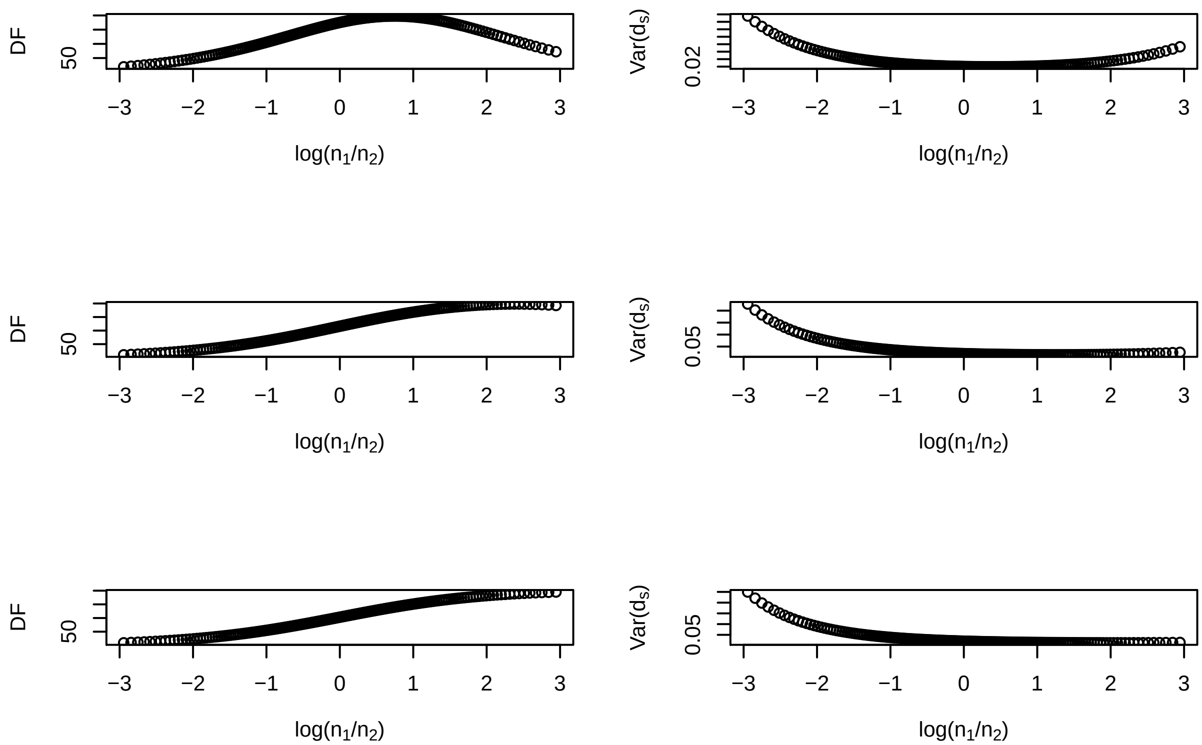

Figure 19. The variance of Cohen's d'_s, when variances and sample sizes are unequal across groups, as a function of the logarithm of the sample sizes ratio ($log\left(\frac{n_1}{n_2}\right)$), when SD-ratio equals 1.46 (first row), 3.39 (second row) or 7 (third row).

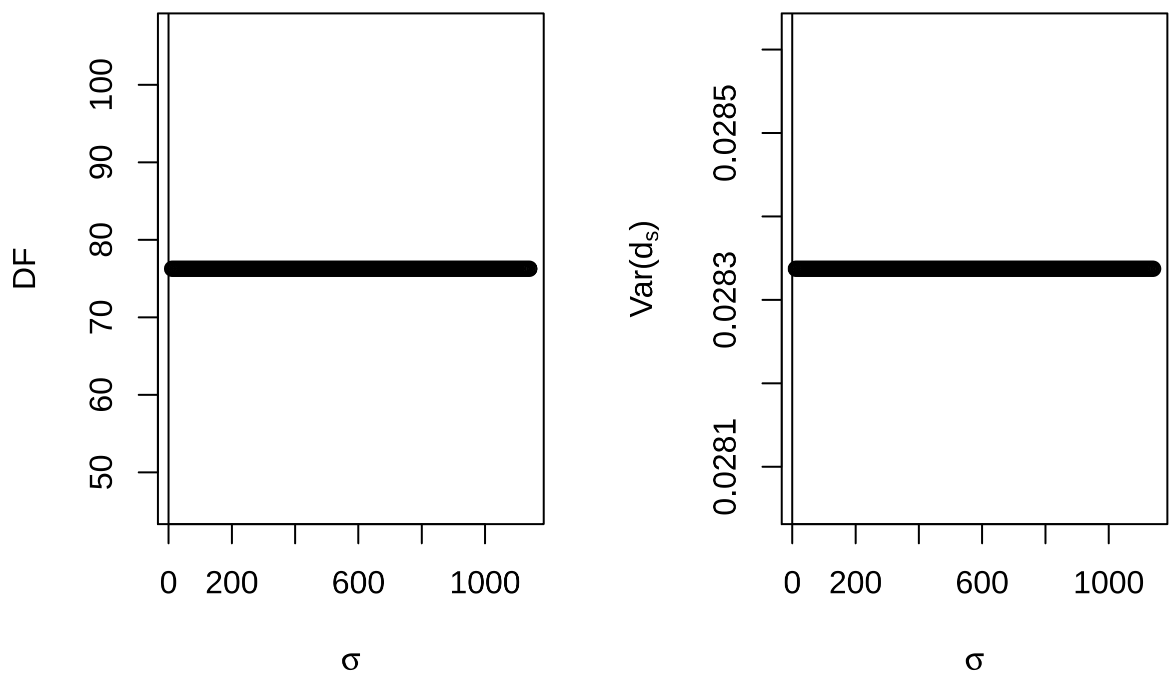

Figure 20. Variance of Cohen's d'_s, when variances and sample sizes are unequal across groups, as a function of σ_1 and σ_2, for a constant *SD*-ratio.

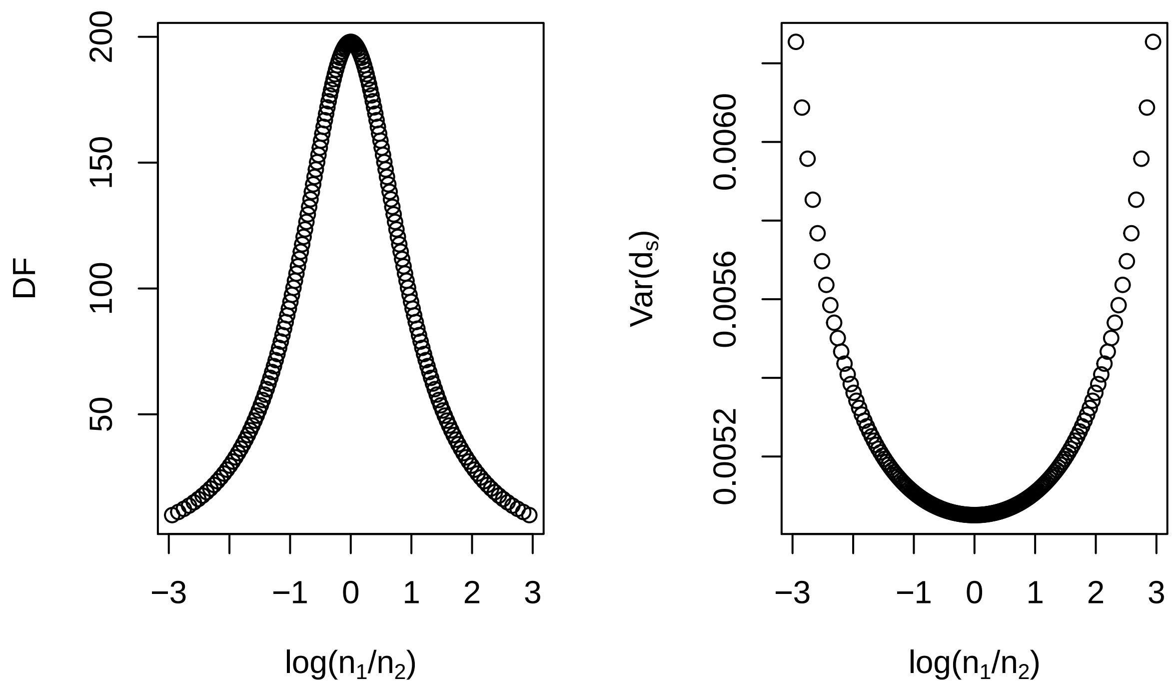

Figure 21. Variance of Shieh's d_s when variances are equal across groups, as a function of the logarithm of the sample sizes ratio ($log\left(\frac{n_1}{n_2}\right)$).

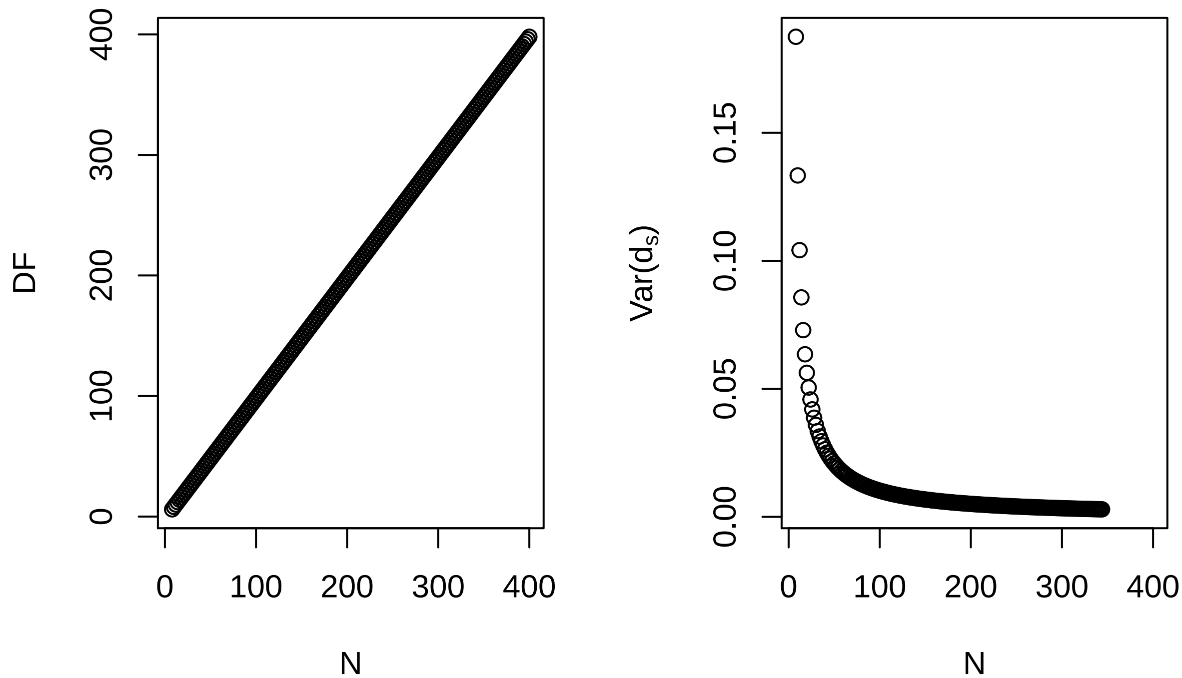

Figure 22. Variance of Shieh's d_s when variances are equal across groups, as a function of the total sample size, for a constant sample sizes ratio ($log\left(\frac{n_1}{n_2}\right)$).

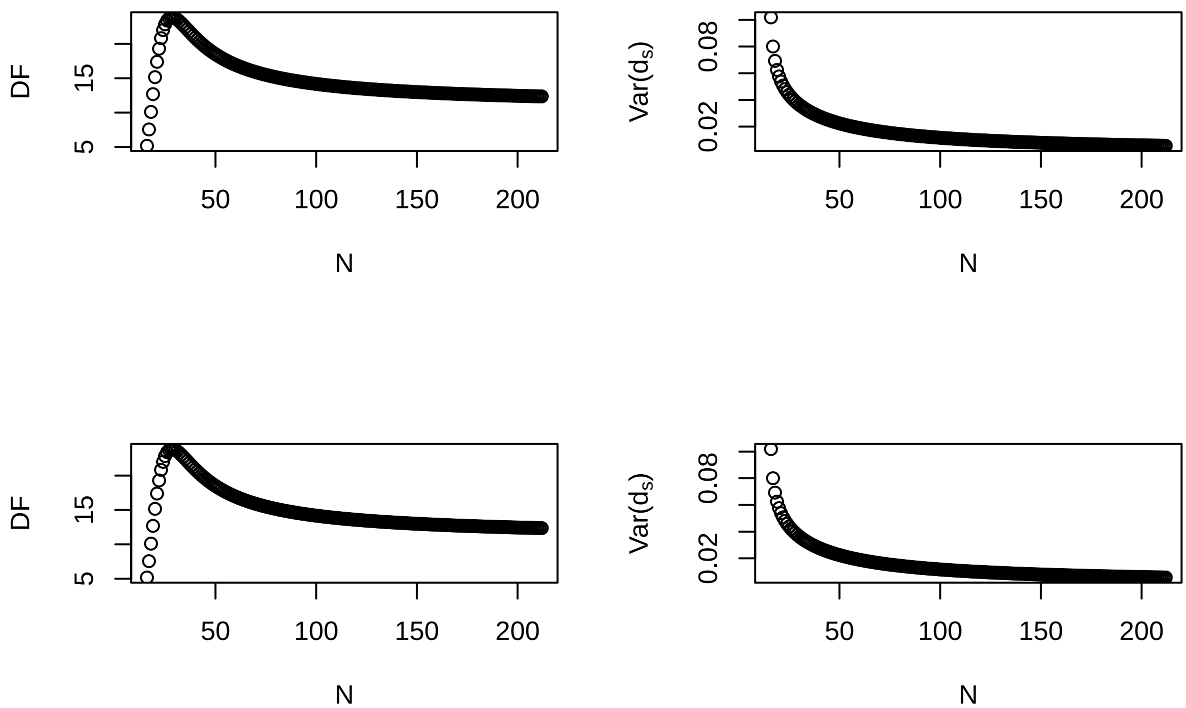

Figure 23. Variance of Shieh's d_s when variances are equal across groups, as a function of the total sample size, when adding subjects only in one group (either in the first group; see top plots; or in the second group; see bottom plots).

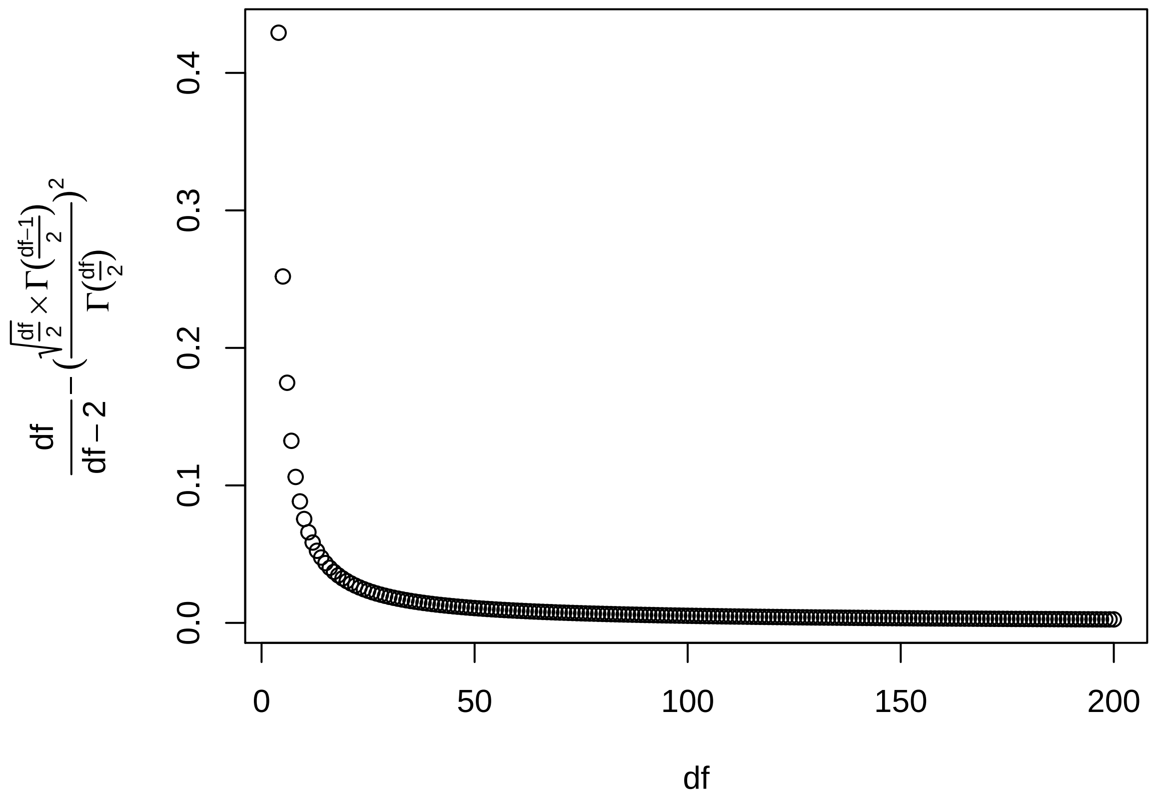

Figure 24. Effect size moderator, as a function of the size of the control (left) or experimental group (right).

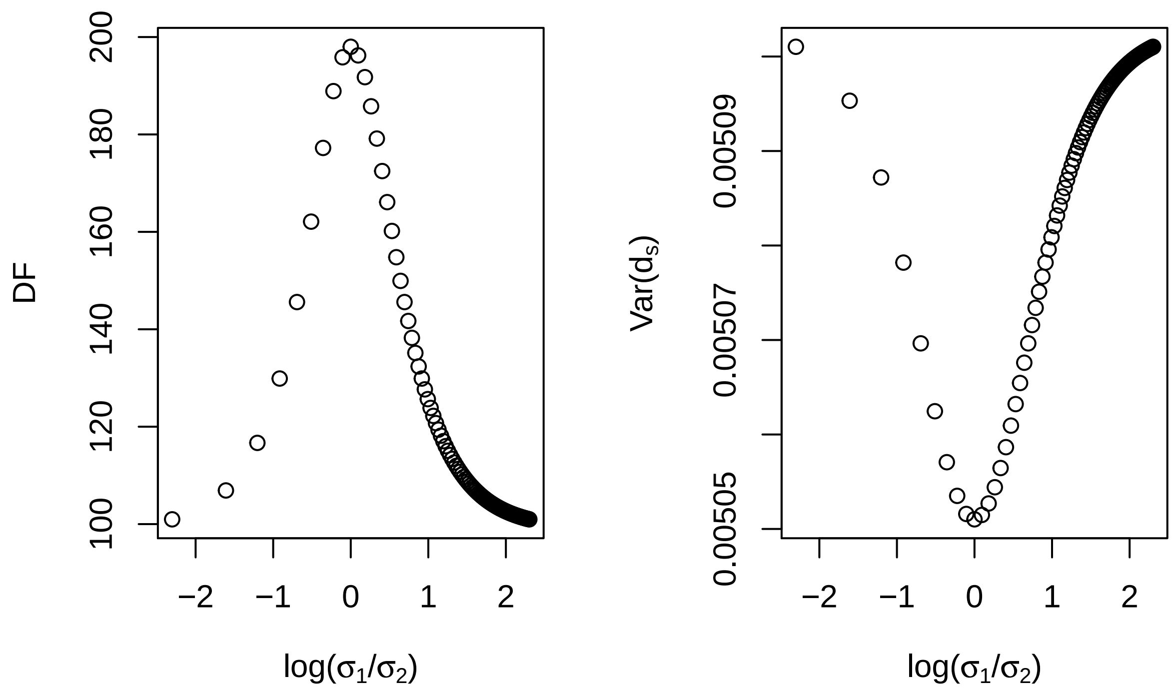

Figure 25. Variance of Shieh's d_s when variances are unequal across groups and sample sizes are equal, as a function of the logarithm of the *SD*-ratio ($log\left(\frac{\sigma_1}{\sigma_2}\right)$).

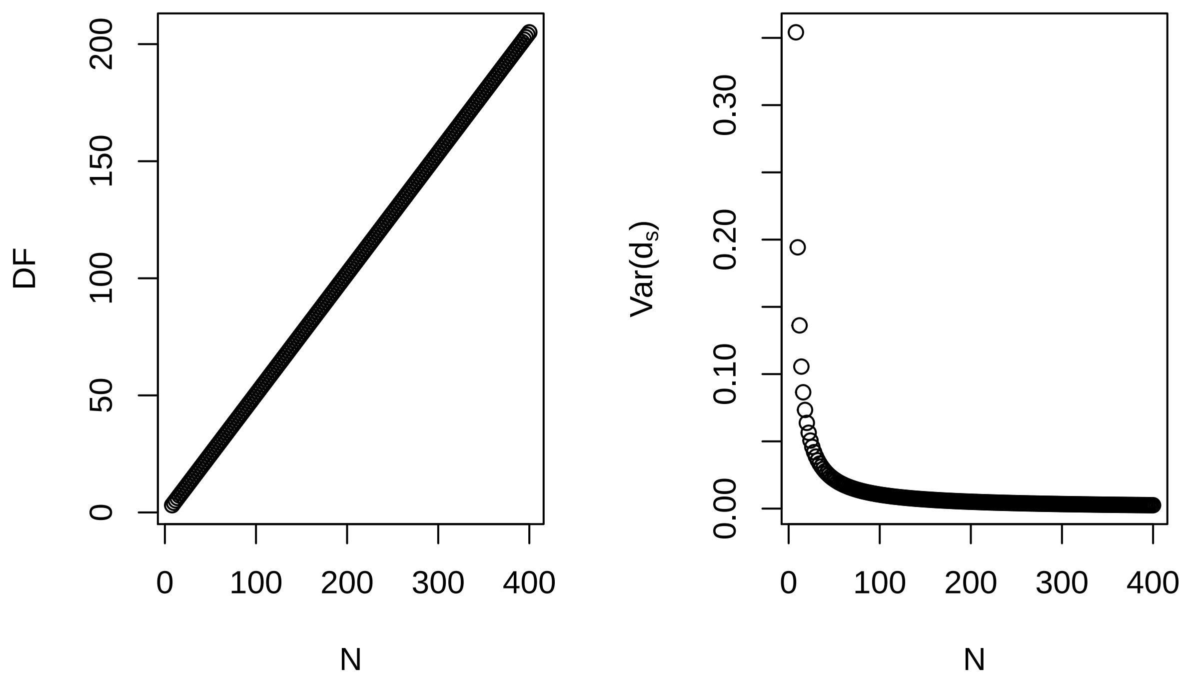

Figure 26. Variance of Cohen's d'_s when variances are unequal across groups and sample sizes are equal, as a function of the total sample size (N).

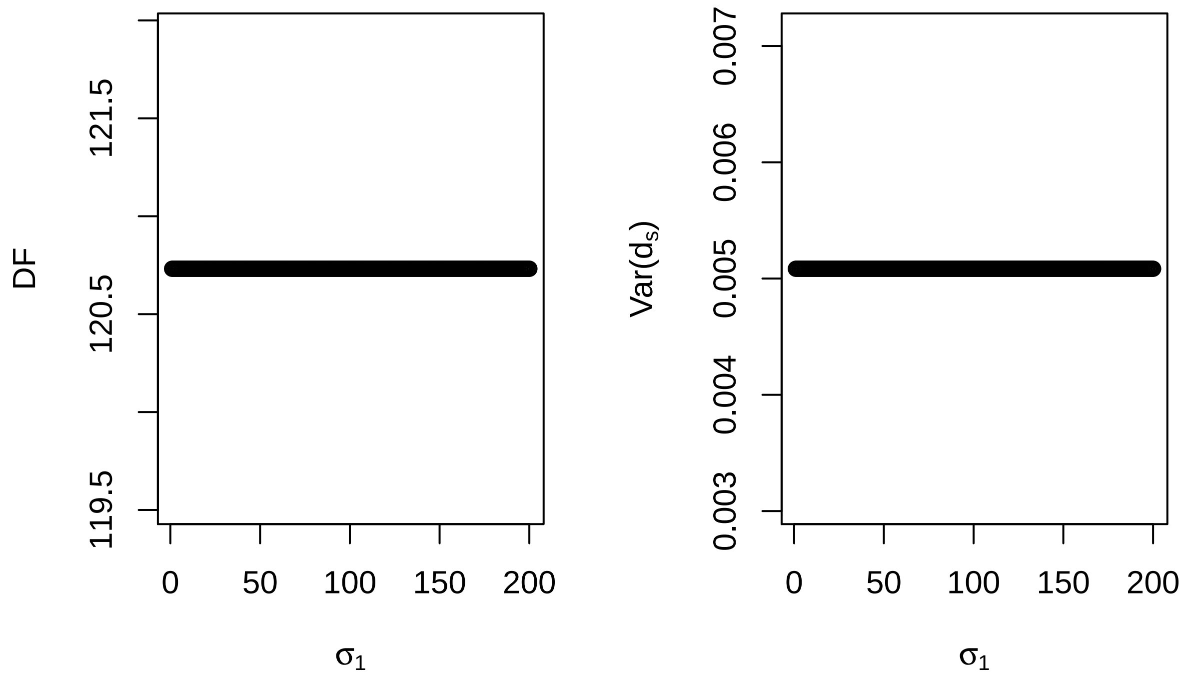

Figure 27. Variance of Shieh's d_s, when variances are unequal across groups and sample sizes are equal, as a function of σ_1 and σ_2, for a constant SD-ratio.

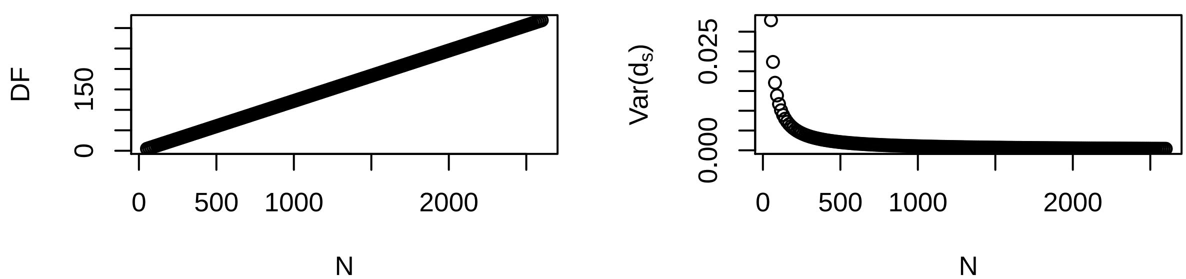

Figure 28. Variance of Shieh's d_s when variances and sample sizes are unequal across groups, as a function of the total sample size, for a constant sample sizes ratio ($log\left(\frac{n_1}{n_2}\right)$).

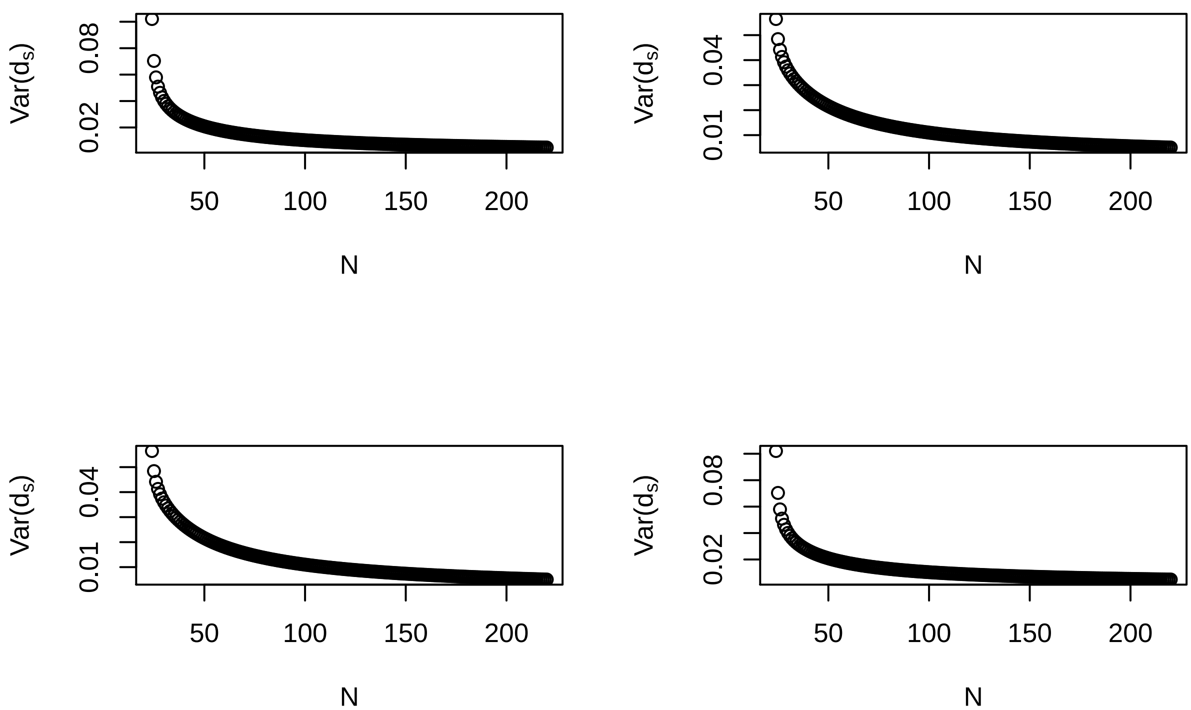

Figure 29. Variance of Shieh's d_s when variances and sample sizes are unequal across groups, as a function of the total sample size, when adding subjects only in one group (either in the first group; see top plots; or in the second group; see bottom plots), and $\sigma_1 > \sigma_2$ (top plots) or $\sigma_1 < \sigma_2$ (bottom plots).

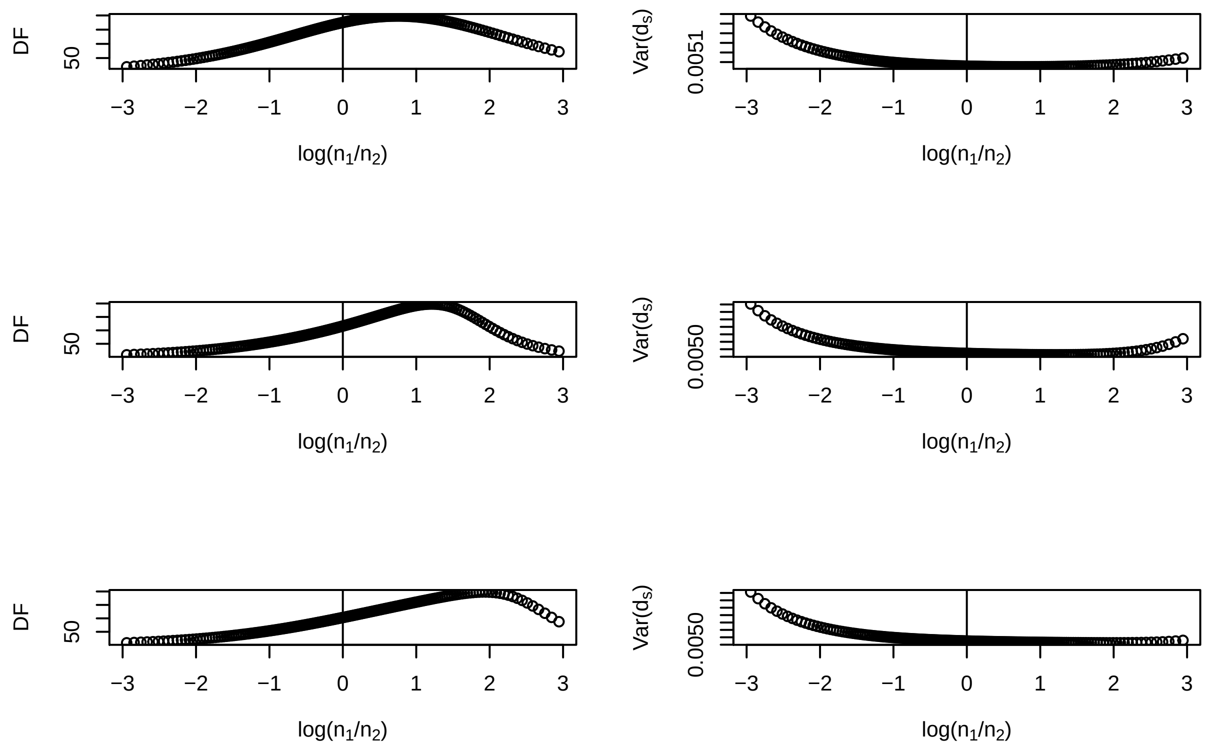

Figure 30. The variance of Shieh's d_s, when variances and sample sizes are unequal across groups, as a function of the logarithm of the sample sizes ratio ($log\left(\frac{n_1}{n_2}\right)$), when SD-ratio equals 1.46 (first row), 3.39 (second row) or 7 (third row).

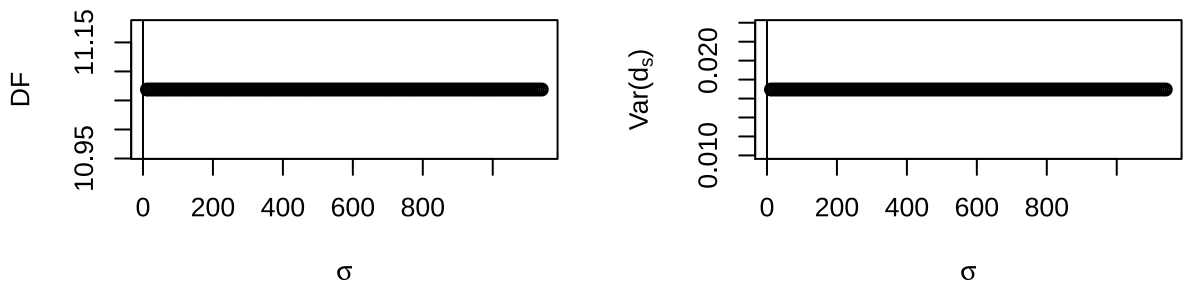

Figure 31. Variance of Shieh's d_s, when variances and sample sizes are unequal across groups, as a function of σ_1 and σ_2, for a constant *SD*-ratio.