- [Sample Size for Individual Responses](#). Disappointingly large.
- [Sport Performance & Science Reports](#). Practitioners' new journal.
  - [Reprint pdf](#) · [Reprint docx](#)

Sample Size for Individual Responses

Will G Hopkins, Institute of Sport Exercise and Active Living, Victoria University, Melbourne, Australia. **Email**.

In the [article on sample-size estimation](#) (Hopkins, 2006), I asserted that sample size for adequate precision for the estimate of the standard deviation representing individual responses in a controlled trial was similar to that for the subject characteristics that potentially explain the individual responses. That assertion was incorrect. In this In-brief item I show that the required sample size in the worst-case scenario of zero mean change and zero individual responses is $6.5n^2$, where n is the sample size for adequate precision of the mean. Since n is usually at least 20, planning for adequate precision of the estimate of individual responses is obviously impractical. Instead, researchers should plan for adequate precision of the subject characteristics and mechanism variables that might explain individual responses, since their sample size in the worst-case scenario is "only" 4n (Hopkins, 2006). The standard deviation for individual responses should still be assessed, because the estimate will be clear for sufficiently large values, and in any case it is important to know how large the individual responses might be, as shown by the upper confidence limit.

The magnitude of individual responses is expressed as a standard deviation, SD_{IR} (e.g., ±2.6% around the treatment's mean effect of 1.8%). The sampling variance (standard error squared) in SD_{IR}^2 is given by statistical first principles as $2V^2/DF$, where $V=SD_{IR}^2$ and DF is the degrees of freedom of the SD_{IR}. V is the difference in the variances of the change scores in the experimental and control groups; hence the sampling variance of SD_{IR}^2 is $2SD_{\Delta E}^4/(n_{IR}-1)$ + $2SD_{\Delta C}^4/(n_{IR}-1)$, where $SD_{\Delta E}$ and $SD_{\Delta E}$ are the standard deviations of change scores in the experimental and control groups, and n_{IR} is the sample size required in each group (assumed equal) to give adequate precision to SD_{IR}. The square root of this expression is the sampling standard error of SD_{IR}^2. In the worst case-scenario, $SD_{IR} = 0$, so $SD_{\Delta E} = SD_{\Delta C} = SD_\Delta$, so the sampling standard error of SD_{IR}^2 is $2SD_\Delta^2/\sqrt{(n_{IR}-1)}$. The sampling standard error of SD_{IR} is not exactly equal to the square root of this expression. In a [simple simulation](#) of a normally distributed variance with mean zero, the expected sampling standard error of the square root of the variance is ~0.80 of the square root of the sampling variance of the variance. Hence the sampling standard error of SD_{IR} is $0.80\sqrt{[2SD_\Delta^2/\sqrt{(n_{IR}-1)}]}$. Since n_{IR} turns out to be very much greater than 1, it follows that the uncertainty in SD_{IR} is inversely proportional to the fourth root of the sample size, whereas the uncertainty in mean effects is inversely proportional only to the square root.

Now, the smallest important value of a standard deviation is half that of a difference or change in a mean (Smith and Hopkins, 2011). Evidence that this rule applies to SD_{IR} is provided by considering how the proportions of positive, trivial, and negative responders change as SD_{IR} increases for a given mean effect of the treatment (Table 1). These proportions were derived with [a spreadsheet](#) that can also be used to investigate how they are impacted by uncertainty in the SD_{IR}. On the reasonable assumption that a difference of 10% in the proportion of responders is substantial, an SD_{IR} of 0.5× the smallest important mean change produces a substantial difference in proportions of responders when the mean change is trivial (0.5× the smallest important change), and an SD_{IR} of 1.0× produces substantial differences in proportions when the mean change is zero or trivial. Larger values of SD_{IR} are needed for substantial changes in proportions when changes in the

mean are substantial. Thus 0.5× the smallest important mean change is an appropriate small-est important value for SD_{IR} in the worst-case scenario of trivial changes in the mean.

Table 1. Proportions of negative, trivial, and positive responders in the population when the mean change and the standard deviation for individual responses (SD_{IR}) are selected fractions and multiples of the smallest important mean change. Proportions in **bold** represent substantial (>10%) differences from the proportion for the same mean change and SD_{IR}=0.

Mean change	SD_{IR}	Proportions of responders (%)		
		Negative	Trivial	Positive
0.0	0.0	0	100	0
0.0	0.5	2	95	2
0.0	1.0	**16**	**68**	**16**
0.5	0.0	0	100	0
0.5	0.5	0	**84**	**16**
0.5	1.0	7	**63**	**31**
1.0	0.0	0	50	50
1.0	0.5	0	50	50
1.0	1.0	2	48	50
1.0	1.5	9	41	50
1.0	2.0	**16**	34	50
2.0	0.0	0	0	100
2.0	0.5	0	2	98
2.0	1.0	0	**16**	**84**
3.0	0.0	0	0	100
3.0	1.5	0	9	91
3.0	2.0	2	**14**	**84**

Proportions were derived with a spreadsheet by assuming individual responses were normally distributed with the given mean change and SD_{IR}.

The standard error for SD_{IR} therefore needs to be 0.5 of the standard error for the change in the mean, when the sample size for the change in the mean (n_Δ) gives adequate precision for zero change in the mean. The standard error for the change in the mean in each group is $SD_\Delta/\sqrt{n_\Delta}$, and the standard error for the difference in the changes is $\sqrt{2}SD_\Delta/\sqrt{n_\Delta}$. So $0.80\sqrt{[2SD_\Delta^2/\sqrt{(n_{IR}-1)}]} = 0.5\sqrt{2}SD_\Delta/\sqrt{n_\Delta}$, from which it follows that $n_{IR} = 1+(0.80/0.5)^4 n_\Delta^2 = 6.5n_\Delta^2$. I have used simulations published in this issue of Sportscience to check that this formula is valid (Hopkins, 2018).

Hopkins WG (2006). Estimating sample size for magnitude-based inferences. Sportscience 10, 63-70

Hopkins WG (2018). SAS programs for analyzing individual responses in controlled trials. Sportscience 22, 1-10

Smith TB, Hopkins WG (2011). Variability and predictability of finals times of elite rowers. Medicine and Science in Sports and Exercise 43, 2155-2160

Reviewer's commentary

This is a very useful contribution to the body of knowledge on treatment heterogeneity. Hopkins has demonstrated that the required sample size for adequate precision of estimation of the SD for individual responses (in the worst-case scenario) is infeasibly large, and no such trial could ever be conducted. For example, consider a conventional parallel-group, before-and-after RCT planned with 90% power at 2-tailed P=0.05 to detect a difference of 3 mmHg in systolic blood pressure with an SD of 10 mmHg, with a correlation between baseline and follow-up measures over the time course of the experiment of r=0.7. Such a study, based on an

ANCOVA analysis model to adjust for chance baseline imbalance, would require 120 participants in each arm. Detecting individual response variance with adequate precision would require up to 93,600 participants per group!

As Hopkins mentions, much smaller and more realistic sample sizes would be needed if the net mean effect (intervention minus control) and the SD for individual responses were substantial. However, he argues persuasively that it is more sensible to design trials with adequate precision to evaluate the effect of putative modifiers of true individual response variance. In this instance the "rule of 4" applies: for any such effect modifier we need 4× the sample size required for the overall net mean effect (480 per arm in the above example). With ever increasing hype surrounding personalized or precision medicine, we need larger trials and appropriate analysis methods to make robust inferences.

Sport Performance & Science Reports

Martin Buchheit, Paris Saint-Germain, 78100 Saint-Germain-en-Laye, Paris, France. **Email**. Reviewer: Will G Hopkins, Institute of Sport Exercise and Active Living, Victoria University, Melbourne, Australia. Sportscience 22, ii, 2018 (sportsci.org/2018/inbrief.htm#spsr. Published Feb 2018. ©2018

The journal *Sport Performance & Science Reports* was launched in November 2017 in response to the frustrations that many applied sport scientists experience with the relevance and dissemination of sport research. As an applied sport scientist working in elite sport, I have found that research is often not aligned toward practitioners' real needs. Furthermore, it is usually written in a difficult academic style and hidden behind a journal subscription. You can find my thoughts on this problem hidden in an invited commentary (Buchheit, 2017). Fortunately I was able to co-publish the commentary in my blog, where you will see that I compared sport scientists with astronauts stuck in orbit, waiting to be rescued. The new journal is a rescue mission.

Articles published in *Sport Performance & Science Reports* are short and straight to the point, with clear practical applications. Busy practitioners can write these articles, improving their relevance. The articles are also published with their accompanying database and statistical spreadsheets for better transparency and learning opportunities for peers. Finally, the editors of the new journal and our colleagues are all frustrated with the flawed traditional reviewing process and the pervasive climate of manuscript rejection. We have therefore opted for post-publication peer review, whereby all articles consistent with the journal's aims and guidelines are published immediately. Authors may then update their articles in response to comments from readers. We hope that our initiative will help bring sport scientists down to earth, where they can lead more rewarding professional lives in the service of sport.

Buchheit, M. (2017). Houston, we still have a problem. International Journal of Sports Physiology and Performance 12, 1111-1114

————