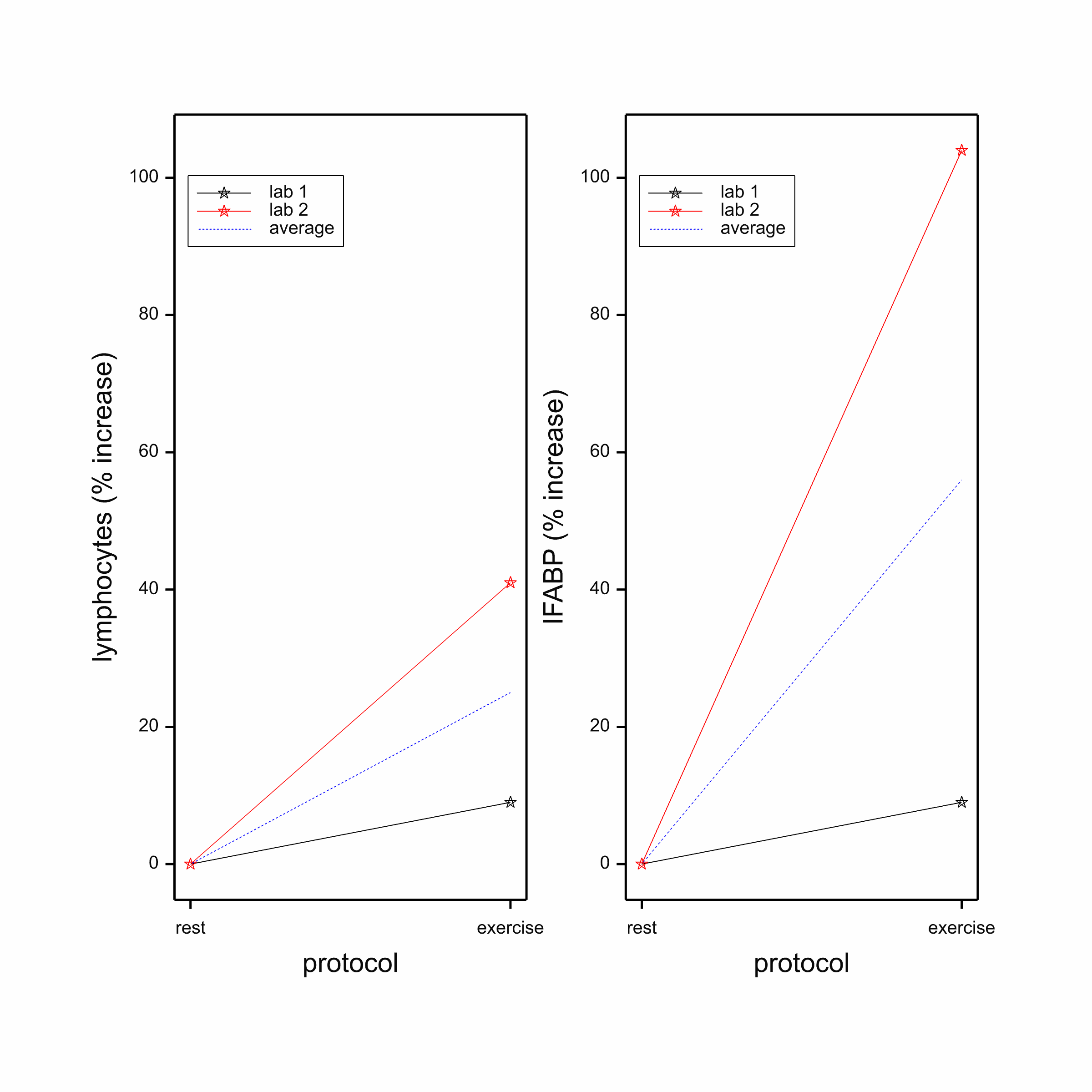
**Statistical design and sample size calculation**

Key parameters of the Tempo! study are lymphocyte concentration, which reflects immunological response to stress, and IFABP concentration, which reflects gut permeability. The coefficients of variation observed for these parameters in the GRINTA! study were 14% (lymphocytes) and 30% (IFABP). Based on these measures of random variation, we calculated minimal effect sizes that can be established for a range of feasible sample sizes by a t test. We assumed a type-1 error of 0.05 and a power of 90%. We considered average effects sizes over the two locations as well as differences in effect sizes between the two locations. An effect size is defined as the increase in lymphocytes or IFABP at some time after start of an exercise protocol versus the change during the same period at rest. Tests for average effect sizes are one sided, because we expect an increase in lymphocytes or IFABP. Tests for differences in effect sizes are two-sided, because it is not known in advance which location has the greater effect size and which one has the smaller effect size. The results are shown in Table yyy.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | lymphocytes | | |  | IFABP | | |
| sample size |  | average | location 1 | location 2 |  | average | location 1 | location 2 |
| 8 |  | 37 | 10 | 64 |  | 91 | -25 | 207 |
| 12 |  | 29 | 10 | 49 |  | 68 | 3 | 133 |
| 16 |  | 25 | 9 | 41 |  | 56 | 9 | 104 |

We exemplify the interpretation of the table by the results for a sample size of 16. Figure qq shows these results graphically. The blue lines in the plots show the average difference between an exercise protocol and the rest protocol that can be picked up using the sample size of 16. For the rest protocol, we expect no difference between IFABP/lymphocyte values taken some time after the start of the protocol and taken at the start. Therefore, the two blue lines start at the point (0,0). The lines further show that a sample size of 16 permits picking up a difference between exercise and rest protocol values of 25% for lymphocytes and 56% for IFABP. That is, the sample size of 16 is sufficiently large to detect that, after some time of exercise, the lymphocytes increase by a factor of (100 + 25)/100 = 1.25. The corresponding factor for IFABP is 1.56.

The red and black lines in the figure show how *different* the effect assessment in the two laboratories must be to be picked up by the statistical tests based on a sample size of 16. The left panel of the figure shows that a difference in effect size for lymphocytes of 32% can be picked up, while the corresponding difference for IFABP is 95%. As the coefficient of variation for the lymphocyte concentration is smaller than for IFABP, the effect sizes that can be established for any given sample size will be smaller for lymphocytes than for IFAPB.



To place the effect sizes in perspective, we show in Table xx the increase in lymphocytes and IFABP half an hour and a full hour after start of the challenge, as observed in GRINTA!

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| protocol | lymphocytes | |  | IFABP | |
|  | 0.5 | 1 |  | 0.5 | 1 |
| 70% EU | 116 | 184 |  | 22 | 128 |
| 70% de | 117 | 153 |  | 36 | 188 |
| 50% | 49 | 86 |  | -7 | 3 |
| 85%/55% | 116 | 179 |  | 22 | 107 |
|  |  |  |  |  |  |

Comparing the effect sizes that can be established with 8-16 volunteers with the sizes from GRINTA!, we see that the observed increase in lymphocytes half an hour after start is likely to be established in a statistically significant way when using a sample size as small as 8. However, this sample size would permit picking up differences between the effect sizes in the laboratories no small than 54%. As the primary objective of Tempo! is to compare different laboratories, the sample size should be sufficiently large to pick up differences if they exist. For this reason, we prefer a sample size of 16.

For IFABP, the increase observed half an hour after start is too small to be established with 8-16 volunteers. However, the sample size of 8 is sufficient to establish the one-hour increases if they are indeed present. The tables further show that inter-laboratory differences in IFABP effects must be very large indeed to be established with sample sizes between 8 and 16. We conclude that lymphocyte concentration is a suitable indicator for laboratory differences, but IFABP is not.

*Statistical design and analysis*

The statistical design of the study consists of replicates of the following Latin Square:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| subject | period | | | |
|  | I | II | III | IV |
| 1 | A | D | B | C |
| 2 | B | A | C | D |
| 3 | C | B | D | A |
| 4 | D | C | A | B |

The letters in the Latin Square denote the four different protocols. It can be checked that every protocol is preceded by every other protocol exactly one time. This permits quantification of carry-over effects.

Based on two replicates of the above Latin Square, and supposing that there are no missing data, an analysis of variance can be carried out, which splits the total variation in components that can be associated with the respective factors. An skeleton analysis of variance table for the difference between a measurement x hour after start and the measurement at the start of a protocol is given below.

|  |  |  |
| --- | --- | --- |
| source of variation | degrees of freedom | |
|  |  |  |
| between subjects |  | 15 |
| laboratory | 1 |  |
| sequence | 3 |  |
| lab x seq | 3 |  |
| residual | 8 |  |
|  |  |  |
| period |  | 3 |
|  |  |  |
| within subjects |  | 45 |
| protocol | 3 |  |
| lab x prot | 3 |  |
| carry-over | 3 |  |
| lab x carry-over | 3 |  |
| residual | 33 |  |
|  |  |  |
| TOTAL |  | 63 |

If there is a moderate loss of data, the analysis can be carried out using a linear mixed model that includes the listed sources of variation.

**Appendix. Detailed effect size calculations for sample sizes of 11-19 volunteers**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | lymphocytes | | |  | IFABP | | |
| sample size |  | average | location 1 | location 2 |  | average | location 1 | location 2 |
| 11 |  | 31 | 10 | 52 |  | 72 | 0 | 144 |
| 12 |  | 29 | 10 | 49 |  | 68 | 3 | 133 |
| 13 |  | 28 | 10 | 47 |  | 65 | 5 | 125 |
| 14 |  | 27 | 10 | 45 |  | 62 | 8 | 117 |
| 15 |  | 26 | 9 | 43 |  | 59 | 8 | 110 |
| 16 |  | 25 | 9 | 41 |  | 56 | 9 | 104 |
| 17 |  | 24 | 9 | 40 |  | 54 | 9 | 99 |
| 18 |  | 23 | 8 | 38 |  | 52 | 10 | 94 |
| 19 |  | 22 | 8 | 37 |  | 51 | 11 | 91 |