

## Context

- 1 Evaluation method
- 2 Paragraph for the study protocol
- 3 Details
- 4 Results
- Appendix
- References

# Gabi SmartCare



## Sample Size estimation for the respiratory rate accuracy trial

DNAlytics -  
(mailto:marco.munda@dnalytics.com)

marco.munda@dnalytics.com

21/11/22

## Context

⌚ Sample size calculation to validate respiratory rate (RR) accuracy by comparison of repeated paired readings from the test system and the reference system in a pediatric population: neonates (0 - 28 days); infants (28 days - 2 years); children (2 years - 5 years).

## 1 Evaluation method

Accuracy is stated in terms of the root-mean-square (rms) difference between measured values ( $\hat{x}_i$ ) and reference values ( $x_i$ ), as given by

$$A_{\text{rms}} = \sqrt{\frac{\sum_{i=1}^n (\hat{x}_i - x_i)^2}{n}}$$

$A_{\text{rms}}$  must be below 3 cpm (cycles per minute) to meet the specification.

From a statistical perspective, this specification is achieved ("success") if the **one-sided 95% confidence interval** of  $A_{\text{rms}}$  lies below 3 cpm.

In case of repeated measures per subject, the above formula is modified as follows

$$A_{rms} = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m (\hat{x}_{ij} - x_{ij})^2}{nm}}$$

where  $i = 1, \dots, n$  indexes the subjects, and  $j = 1, \dots, m$  indexes the repeated measures within a subject.

**This document provides information on the sample size requirement for such a study.**

## 2 Paragraph for the study protocol

### 4 paired readings per subject

Assuming that the mean of the absolute differences between measured values and reference values is  $\leq 2$  and that the between-subject and within-subject variances are  $\leq 0.75$ , a minimum sample size of 12 patients (resulting in 48 paired data points) was computed to achieve  $\geq 80\%$  power in demonstrating that  $A_{rms}$  is below 3 cpm, at the 95% confidence level.

### 10 paired readings per subject

Assuming that the mean of the absolute differences between measured values and reference values is  $\leq 2$  and that the between-subject and within-subject variances are  $\leq 0.75$ , a minimum sample size of 11 patients (resulting in 110 paired data points) was computed to achieve  $\geq 80\%$  power in demonstrating that  $A_{rms}$  is below 3 cpm, at the 95% confidence level.

### 20 paired readings per subject

Assuming that the mean of the absolute differences between measured values and reference values is  $\leq 2$  and that the between-subject and within-subject variances are  $\leq 0.75$ , a minimum sample size of 10 patients (resulting in 200 paired data points) was computed to achieve  $\geq 80\%$  power in demonstrating that  $A_{rms}$  is below 3 cpm, at the 95% confidence level.

## 3 Details

### Constructing the confidence interval of $A_{rms}$

A confidence interval of  $A_{rms}$  can be obtained by bootstrap. The resampling scheme must take the hierarchical nature of the data into account (repeated measures). The non-parametric bootstrap simply consists of randomly sampling

subjects with replacement (Davison and Hinkley (1997), Section 3.8). In its parametric version, resample data can be generated according to the classical linear random effects model:

$$d_{ij} = \mu_0 + b_i + \epsilon_{ij} \quad (d_{ij} = \hat{x}_{ij} - x_{ij})$$

with  $b_i \sim N(0, \sigma_b^2)$  and  $\epsilon_{ij} \sim N(0, \sigma_e^2)$ .

 Alternatively, the distributional properties of  $A_{rms}$  are studied in Ndikintum and Rao (2016) based on the above model, under balanced settings (e.g., no missing values). In particular, the expression of the variance of  $A_{rms}^2$  in that paper reveals that it is more beneficial to increase the number of patients rather than the number of replicates per patient. Bayesian estimation is yet another option.

## Sample size calculation

The sample size is determined to reach  $\geq 80\%$  power, i.e. to reach a *probability of success* of  $\geq 80\%$ , using the bootstrap method.

The power of the study has been estimated from 1000 simulations in the following settings:

- The data are generated from the above model in order to reflect the hierarchical nature of the study data (Ndikintum and Rao (2016));
- The number of patients  $n$  varies between 5 and 20;
- The number of paired readings per subject  $m$  is set to 4, 10, and 20;
  - The total number of paired data points equals  $n \times m$ .
- The bias between the two methods ( $\mu_0$ ) varies between 0 and 2;
- The within-subject variability of the differences ( $\sigma_e^2$ ) varies between 0.5 and 1;
- The between-subject variability ( $\sigma_b^2$ ) is set equal to  $\sigma_e^2$ .
  - Worst-case scenario according to Ndikintum and Rao (2016).

The table below reports the value of the population accuracy,  $\sqrt{\sigma_e^2 + \sigma_b^2 + \mu_0^2}$ , for each scenario:

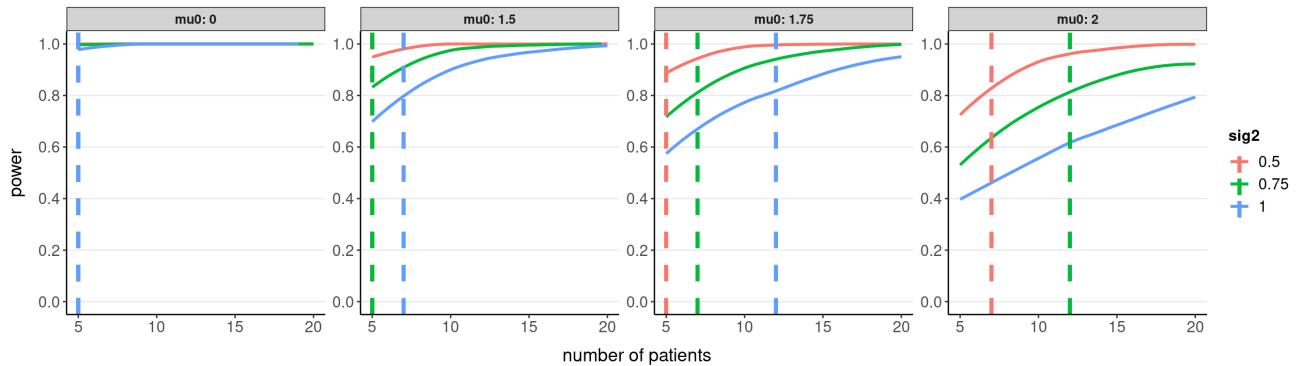
mu0	sig2e	sig2b	Arms
0.00	0.50	0.50	1.00
0.00	0.75	0.75	1.22
0.00	1.00	1.00	1.41
1.50	0.50	0.50	1.80

<b>mu0</b>	<b>sig2e</b>	<b>sig2b</b>	<b>Arms</b>
1.50	0.75	0.75	1.94
1.50	1.00	1.00	2.06
1.75	0.50	0.50	2.02
1.75	0.75	0.75	2.14
1.75	1.00	1.00	2.25
2.00	0.50	0.50	2.24
2.00	0.75	0.75	2.35
2.00	1.00	1.00	2.45

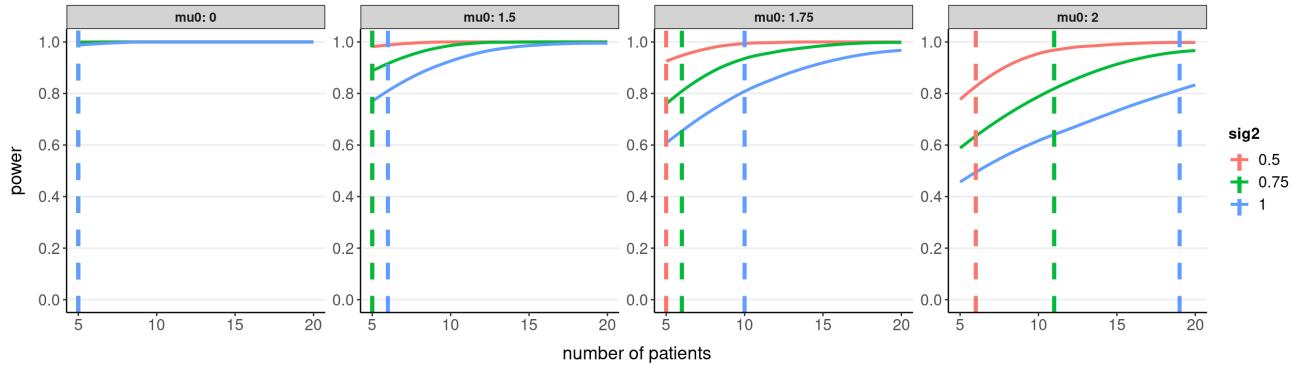
## 4 Results

Power curves (power versus number of patients) are depicted below. The vertical lines indicate the minimum numbers of patients to reach 80% power. The complete results are shown in tabular format in the appendix.

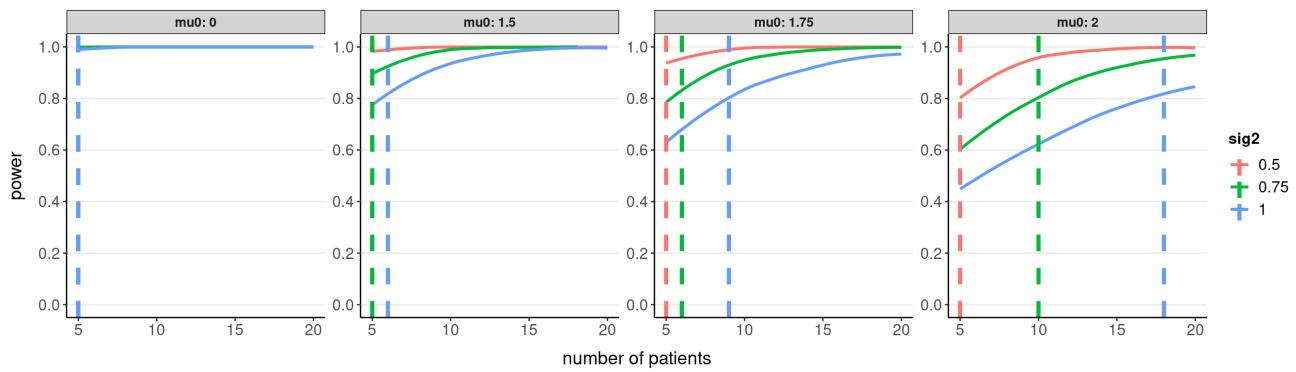
### 4 paired readings per subject



### 10 paired readings per subject



### 20 paired readings per subject



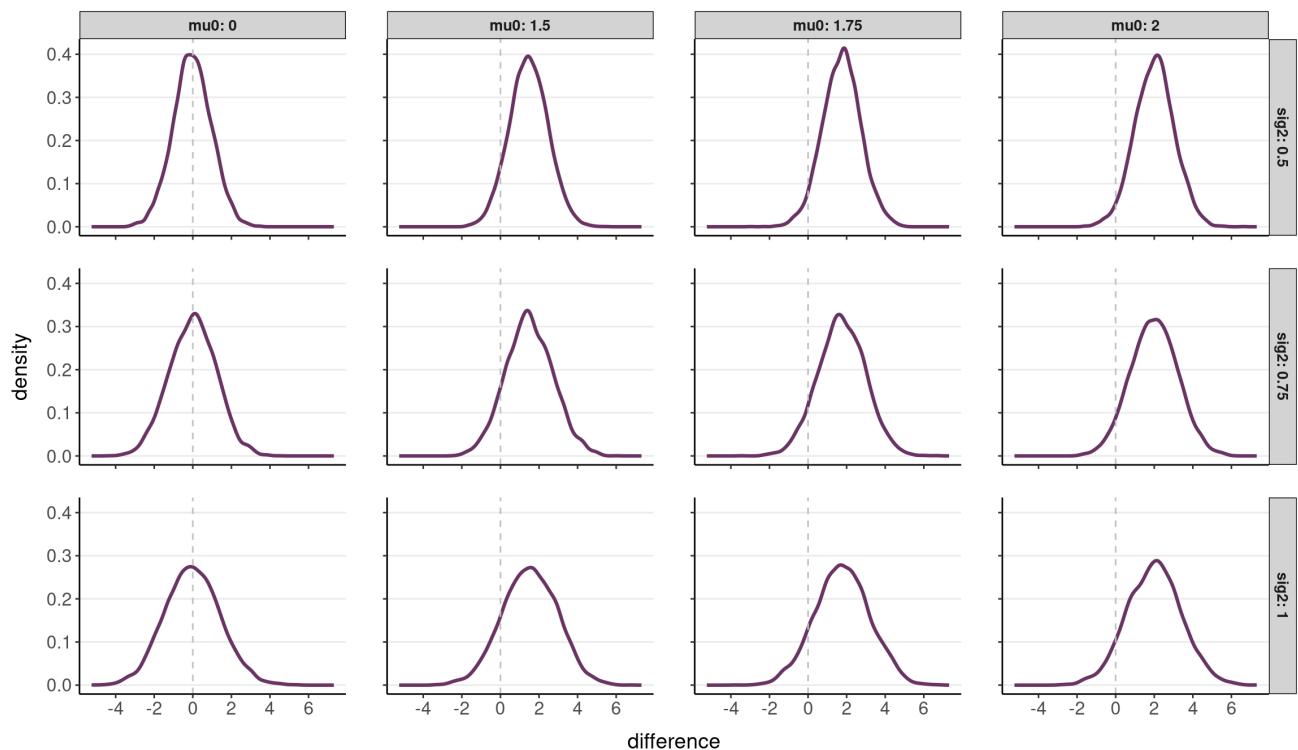
# Appendix

## Simulation results

$\mu_0$	$\sigma^2_e$	$\sigma^2_b$	Arms	$m$	$n$	$p$
0	0.5	0.5	1	4	5	
0	0.5	0.5	1	4	6	
0	0.5	0.5	1	4	7	
0	0.5	0.5	1	4	8	
0	0.5	0.5	1	4	9	
0	0.5	0.5	1	4	10	
0	0.5	0.5	1	4	11	
0	0.5	0.5	1	4	12	
0	0.5	0.5	1	4	13	
0	0.5	0.5	1	4	14	
0	0.5	0.5	1	4	15	
0	0.5	0.5	1	4	16	
0	0.5	0.5	1	4	17	
0	0.5	0.5	1	4	18	
0	0.5	0.5	1	4	19	
0	0.5	0.5	1	4	20	

## Scenario selection

To help selecting an appropriate scenario, the following figure depicts the distribution of the difference between the two paired measures in each setting (data are simulated from the linear random effects model for the purpose of the graph).



## References

Davison, A. C., and D. V. Hinkley. 1997. *Bootstrap Methods and Their Application*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511802843> (<https://doi.org/10.1017/CBO9780511802843>).

Ndikintum, Nfii K., and Marepalli Rao. 2016. “A Special Inference Problem in Repeated Measures Design—Test of Statistical Hypothesis on Accuracy Root Mean Square—Application to Pulse Oximetry Studies.” *Statistics in Biopharmaceutical Research* 8 (1). <https://doi.org/10.1038/s41416-019-0602-7> (<https://doi.org/10.1038/s41416-019-0602-7>).

