Sample Size considerations for the KIPRUN study

Márton Kiss, MD

2024-12-04

Table of contents

[Study Overview and Rationale 2](#_Toc184214251)

[Study Design Summary 2](#_Toc184214252)

[Key Endpoint 3](#_Toc184214253)

[Detailed results 4](#_Toc184214254)

[Remarks 6](#_Toc184214255)

[References 6](#_Toc184214256)

# Study Overview and Rationale

The clinical study aims to refine the current approach to infection monitoring using PCT (Procalcitonin) levels, specifically for patients in the ICU setting. The study would help determine how well a new PCT monitoring method consisting of daily measurements might work compared to the current standard (whose accuracy is not satisfactory), which we refer to as the “gold standard” test (a PCT measurement taken after suspicion of infection has arisen).

The goal would be to improve accuracy for diagnosing infections with the help of PCT and provide better clinical decision-making support, ultimately enhancing patient outcomes.

# Study Design Summary

* **Two Phases:** Since data on the baseline PCT measurements’ relevance is not available yet, the study would be split into two phases. During the first phase, data would be collected, and a specific method would be defined at the end of Phase 1. The method then would be validated in Phase 2.
* **Phase 1: Threshold Setting**
  + **Patient Inclusion:** Patients would be monitored via daily blood draws (eg. every day between 0800-0900). An “event” would be defined as a rise in suspicion of infection significant enough for a physician to order a PCT test. The physician should be blinded against the daily test results but would be free to order PCT tests as required.
  + **Sample size:** At least **60 “events”** are required to observe 7-20 misclassfications (based on the previous study’s data of inappropriate therapy in 23% of cases). The incidence of infections impacts the sample size, eg. if 50% of all ICU patients experience an infection during their stay, 120 patients are needed to be enrolled for an observed 60 events, but if this ratio is 10%, 600 patients needed to be enrolled (in Phase 1).
  + **No Treatment Decision Based on New Test:** Treatment decisions would not be influenced by the new test results in Phase 1 (since the new test would still be under devlopment). Treatment is at the discretion of the physician, consistent with current practices. This avoids the possible ethical burden of treating patients in a manner which is contraindicated by the guidelines as well as making it possible to compare the current best practice with the better alternative (as described later).
  + **Analysis:** Data gathered would be used to develop a specific methodology for the new test using all data observed for a patient (daily monitoring results and the post-suspicion data).
* **Interim analysis**
  + Patient enrollment should be suspended during the evaluation of Phase 2.
  + Activities include:
    - Chart review (responsible: study coordinator).
    - Compilation of an interim report containing recommendations for a test to be used in Phase 2.
    - Recommendation on GO/NO GO for Phase 2 based on the data.
  + Revised sample size for Phase 2, given the characteristics of the developed test.
  + The test results based on the gold standard methodology would not impact the decisions at the interim. Assumptions regarding the performance of the gold standard test are to be taken from the previous study’s data.
* **Phase 2: Validation**
  + **Testing Both Methods:** Both the old (gold standard) and the new PCT tests would be run on all patients.
  + **Treatment Based on New Test:** Treatment decisions would be made based on the new test’s results using the thresholds/methods established in the interim after Phase 1.
  + **Goal:** Confirm if the new test might provide a clearer and more accurate picture of infection status at the time when suspicion for infection is arisen compared to the gold standard. Data regarding the gold standard would be taken from both Phase 1 & Phase 2.
  + **Additional analyses:** Other endpoints, like treatment outcomes should also be investigated, comparing results between Phase 1 & Phase 2, comparing the current best practice with outcomes gained with monitoring PCT.
  + **Expected Improvement:** It is estimated that the new test might be approximately 19% more accurate than the old one. This is based on previous results.
  + **Sample size:** For Phase 2, the evaluation of a **further 60 events** would yield 80% Power to establish a significant gain in accuracy over the gold standard method if the expected gain in accuracy is 19%.

# Key Endpoint

* **Accuracy Comparison:** Measure and compare the **accuracy** of the old and new testing methods in identifying infection risks. Accuracy is the outcome whose investigation would require the lowest sample size. Other characteristics (NPV, PPV, sensitivity, specificity etc.) should be reported.
  + **Old Test Accuracy:** Evaluated using data from both Phases 1 and 2.
  + **New Test Accuracy:** Evaluated based solely on Phase 2 data.
  + **Comparison Aim:** Understand the true value of the new test compared to the current gold standard method.

**Important Notes on Significance and Study Phases:**

* **Sample size requirements:** A total of 120 observed events (for Phases 1 & 2 altogether) would be suitable to develop and validate a method given the assumptions mentioned above. However:
  + This would mean a higher number of enrolled subjects, since not everyone develops an infection, and monitoring has to be started before the occurrance of an infection. If X% of patients suffer an infection in the ICU, then for 120 events, 120/(100%-X%) patients would need to be enrolled.
  + The sample size should be revisited at the interim stage where the characteristics of the method which we would be able to develop are better known (ie. would the new method indeed be 19% more accurate than the gold standard?)
  + Most ethical committees prefer to see the maximum possible subjects to be included in the study. These numbers should be given based on the average length of ICU stay and the incidence of infections (or the chance of one patient developing an infection during their stay).
* **No Significance Level Adjustment Required:**
  + **Threshold Selection Independence:** Data from Phase 1 would only be used to set thresholds (or develop a more elaborate algorithm).
  + **Independent Validation:** Phase 2 would serve as the independent validation phase.
  + **No Data Overlap:** Since there would be no data overlap influencing the optimization and validation phases, adjusting the significance level would not be necessary.
  + **Main analysis:** The accuracy (correctly/incorrectly classified patients) per tests could be compared via a chi-squared test.

# Detailed results

Results were simulated based on the design described above. Assumptions included:

* Statistically significant difference between treatments with a two-sided test at =5%
* The possible gain in accuracy compared to the ABSOLUTE method (19%)
* Ratio of the sample size between Phase 1 / Total sample size (50%)
* Sample size for Phase1 >= 60

The results indicate that observing 60 events in Phase 1 and a further 60 events in phase 2 would yield >80% Power for the study. If the actual gain in accuracy would be lower, the Power would be moderately impacted (ie. at 17%, the actual Power would be >70%).

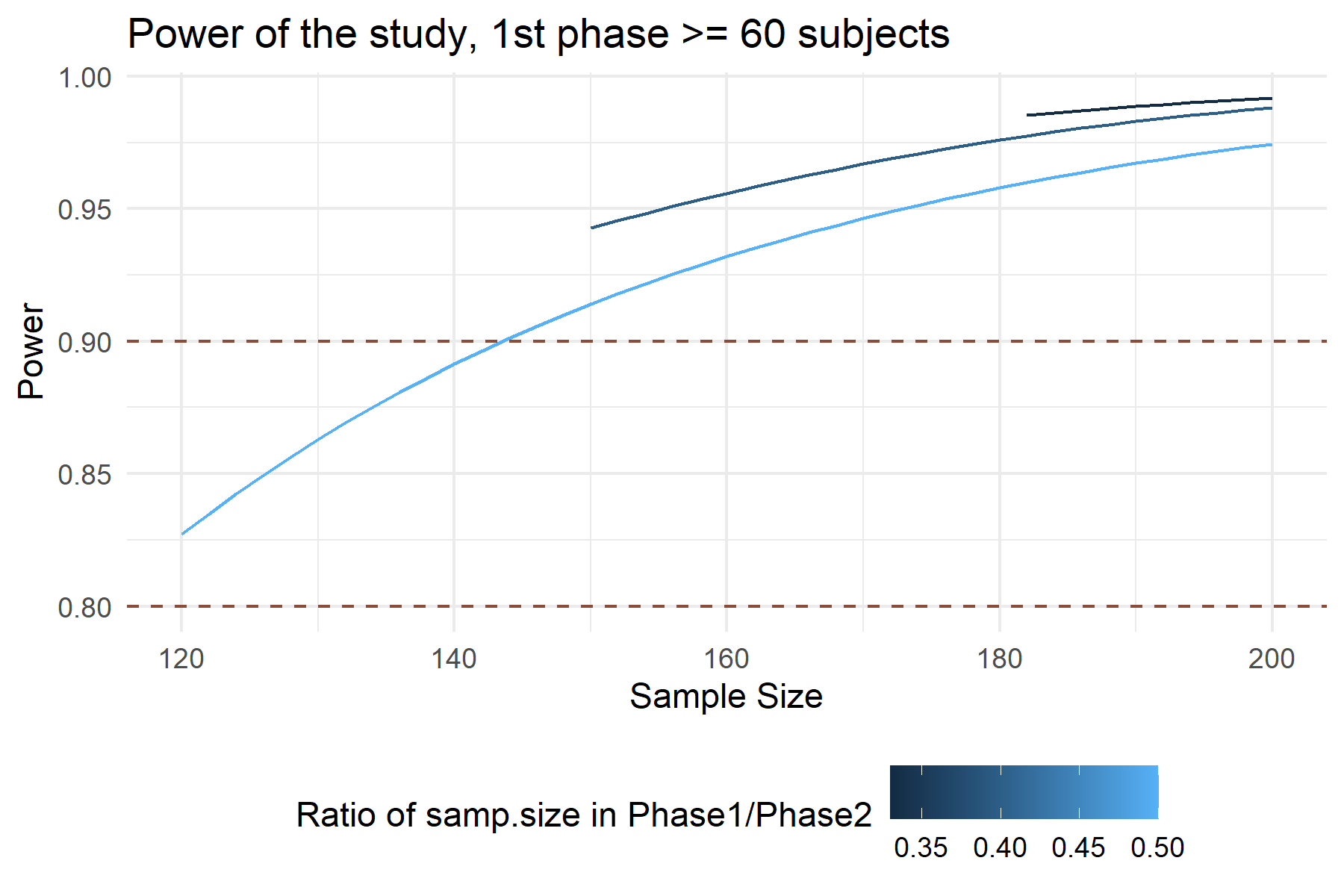


Figure 1

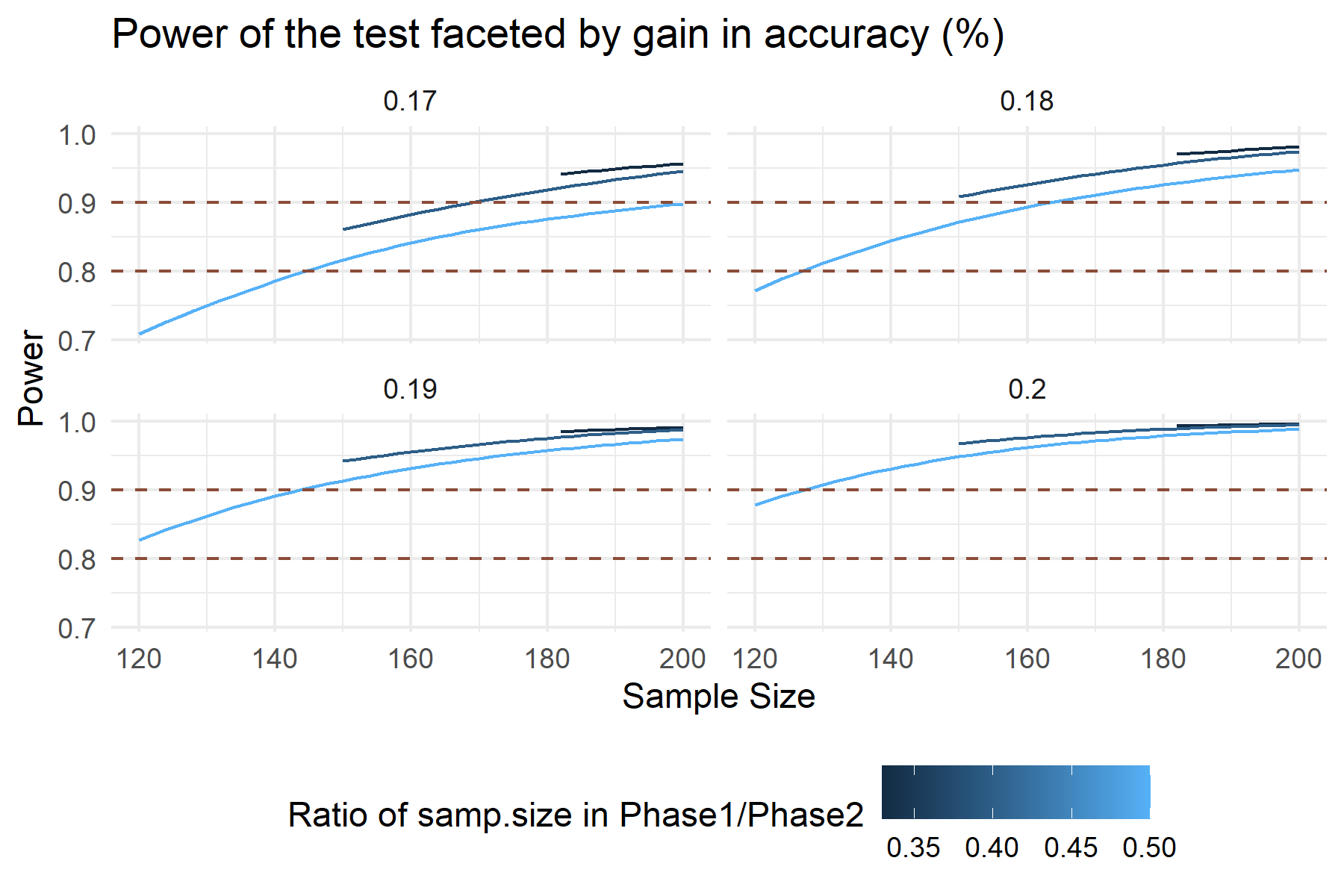


Figure 2

# Remarks

### Information regarding the document’s compilation

Analyses were conducted using the R Statistical language (version 4.4.1; R Core Team, 2024) on Windows 11 x64 (build 22631), using the packages ggplot2 (version 3.5.1; Wickham H, 2016) and dplyr (version 1.1.4; Wickham H et al., 2023).

## References

* R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
* Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
* Wickham H, François R, Henry L, Müller K, Vaughan D (2023). dplyr: A Grammar of Data Manipulation. R package version 1.1.4, <https://CRAN.R-project.org/package=dplyr>.

### Time of compilation

2024-12-04 14:13:20.257424