Sample Size Re-Estimation Report for Crown Milling Accuracy

Márton Kiss, MD

Table of contents

[Executive Summary 1](#_Toc190888572)

[Background 2](#_Toc190888573)

[Scenario 1: Three-Look Group Sequential Design 2](#_Toc190888574)

[Scenario 2: Two-Look Group Sequential Design 3](#_Toc190888575)

[Remarks 4](#_Toc190888576)

[References 4](#_Toc190888577)

# Executive Summary

This report outlines an unblinded sample size re-estimation strategy for a study comparing the crown accuracy of two milling machines. Initially, a formal sample size calculation was not performed before data collection (2×9 measurements). Now, using data collected to date — and with the safeguard of having been blind to the data until this point — we propose to implement a group sequential design where we would update the sample size after interim analyses and therefore the (missing) initial sample size estimate’s importance is diminished.

A priori, we assumed an effect size (the ratio of the expected difference in RMS values to the standard deviation) of 0.8. With a global two-sided significance level of 0.05 and 80% power, we employ a group sequential design. Two scenarios are considered:

* Scenario 1: A three-look design with interim analyses at approximately 33%, 67%, and 100% of the planned information.
* Scenario 2: A two-look design with the first interim analysis at 34.2% of the total sample size and the final analysis at 100%.

In both scenarios, if a significant difference is observed at an interim look (with an adjusted significance level to account for multiple looks), the study can be terminated early for efficacy. Otherwise, a conditional power analysis will be used to update the sample size based on the interim data.

# Background

When comparing two continuous endpoints (here, the RMS errors from crowns produced by two milling machines), a two-sample t-test is appropriate (the RMS being from a chi-squared distribution with a very high degrees of freedom which may be regarded as normal).

Because no formal a priori sample size calculation was completed, an unblinded sample size re-estimation method is now proposed. Group sequential designs allow for early stopping if efficacy is observed, thereby potentially reducing the number of subjects required while still controlling the overall Type I error rate. This design also relies less on the initial sample size estimate, as the sample size can be adjusted based on the interim data.

The two scenarios below illustrate how the sequential design parameters adjust for multiple looks and detail the planned information rates, cumulative alpha spending, stage-specific significance levels, efficacy boundaries, cumulative power, and estimated sample sizes.

The setup is also known as the Hwang-Shih-DeCani alpha spending framework with a gamma parameter fixed at -1.5.

## Scenario 1: Three-Look Group Sequential Design

This design involves three planned analyses at information rates of approximately 33.3%, 66.7%, and 100% of the total sample size. The table below summarizes the design parameters:

*Sample size calculation for a continuous endpoint*

Sequential analysis with a maximum of 3 looks (group sequential design), two-sided overall significance level 5%, power 80%. The results were calculated for a two-sample t-test, H0: mu(1) - mu(2) = 0, H1: effect = 0.8, standard deviation = 1.

| Stage | 1 | 2 | 3 |
| --- | --- | --- | --- |
| Planned information rate | 33.3% | 66.7% | 100% |
| Cumulative alpha spent | 0.0093 | 0.0247 | 0.0500 |
| Stage levels (two-sided) | 0.0093 | 0.0186 | 0.0366 |
| Efficacy boundary (z-value scale) | 2.600 | 2.354 | 2.090 |
| Lower efficacy boundary (t) | -1.391 | -0.823 | -0.583 |
| Upper efficacy boundary (t) | 1.391 | 0.823 | 0.583 |
| Cumulative power | 0.1750 | 0.5157 | 0.8000 |
| Number of subjects | 18.0 | 36.1 | 54.1 |
| Expected number of subjects under H1 |  |  | 41.7 |
| Exit probability for efficacy (under H0) | 0.0093 | 0.0154 |  |
| Exit probability for efficacy (under H1) | 0.1750 | 0.3407 |  |

Legend:

* (t): treatment effect scale

## Scenario 2: Two-Look Group Sequential Design

This simpler design considers only one interim analysis (at 34.2% of the total information) before the final look at 100% of the data. The following table summarizes the design:

*Sample size calculation for a continuous endpoint*

Sequential analysis with a maximum of 2 looks (group sequential design), two-sided overall significance level 5%, power 80%. The results were calculated for a two-sample t-test, H0: mu(1) - mu(2) = 0, H1: effect = 0.8, standard deviation = 1.

| Stage | 1 | 2 |
| --- | --- | --- |
| Planned information rate | 34.2% | 100% |
| Cumulative alpha spent | 0.0096 | 0.0500 |
| Stage levels (two-sided) | 0.0096 | 0.0437 |
| Efficacy boundary (z-value scale) | 2.589 | 2.017 |
| Lower efficacy boundary (t) | -1.386 | -0.570 |
| Upper efficacy boundary (t) | 1.386 | 0.570 |
| Cumulative power | 0.1772 | 0.8000 |
| Number of subjects | 18.0 | 52.6 |
| Expected number of subjects under H1 |  | 46.5 |
| Exit probability for efficacy (under H0) | 0.0096 |  |
| Exit probability for efficacy (under H1) | 0.1772 |  |

Legend:

* (t): treatment effect scale

# Remarks

### Other 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 26100), using the packages rpact (version 4.1.0; Wassmer G, Pahlke F, 2024), ggplot2 (version 3.5.1; Wickham H, 2016), dplyr (version 1.1.4; Wickham H et al., 2023) and kableExtra (version 1.4.0; Zhu H, 2024).

## 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/>.
* Wassmer G, Pahlke F (2024). rpact: Confirmatory Adaptive Clinical Trial Design and Analysis. R package version 4.1.0, <https://CRAN.R-project.org/package=rpact>.
* 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>.
* Zhu H (2024). kableExtra: Construct Complex Table with ‘kable’ and Pipe Syntax. R package version 1.4.0, <https://CRAN.R-project.org/package=kableExtra>.

### Time of compilation

2025-02-19 20:15:07.375291