

TECHNICAL CASE STUDY | BIOSTATISTICS

Correcting sample size planning in a vaccine non-inferiority study

Independent review of recruitment assumptions, confidence interval methods and analysis readiness

KITE-Ai

A wrong sample size can derail a study.

Review the assumptions before recruitment begins.



Biostatistics & Statistical Programming

Executive assessment

KITE-Ai identified an error in a sample size calculation prepared by another service provider for a paediatric vaccine study. The review revisited planning assumptions and recruitment requirements and recommended Newcombe's Hybrid Score method for the planned comparison of seroconversion proportions.

The sponsor benefit, as reported by KITE-Ai, was avoidance of unnecessary enrolment, time savings and lower overall study costs while retaining the agreed statistical requirements. No participant reduction, monetary saving or timeline improvement is quantified in this report.

Technical distinction

Correcting sample size planning and choosing a confidence interval method are related decisions, but they are not interchangeable. Adopting Newcombe's method alone does not establish that fewer participants are sufficient. The final design needs a separate, reproducible power assessment under its prespecified decision rule.

01 | STUDY QUESTION

Define the endpoint and the non-inferiority claim

Clinical and statistical setting

The case concerns a randomized, active-controlled, multi-dose paediatric vaccine study assessing immunogenicity through seroconversion. The primary comparison follows the vaccination regimen; earlier immunogenicity assessments provide additional descriptive information. Product identities, sponsor, sites, study code and project-specific numerical assumptions are omitted.

Response and contrast

For each treatment group, the binary response must be derived from the protocol-defined baseline serostatus, assay threshold and eligible post-vaccination assessment. Let x_T and x_C denote responders and n_T and n_C denote participants with analysable endpoint data in the designated population.

$p_T = x_T / n_T$; $p_C = x_C / n_C$; $d = p_T - p_C$

Here T is the test vaccine and C is the active comparator. Positive differences favour the test vaccine. The contrast direction must be fixed consistently in the SAP, dataset labels, statistical software ordering and output footnotes.

Hypotheses and decision rule

Let $M > 0$ be the prespecified clinically justified non-inferiority margin on the absolute proportion scale. The hypotheses are:

$H_0: p_T - p_C \leq -M$

$H_1: p_T - p_C > -M$

For a prespecified two-sided confidence interval $[L, U]$ at the matching confidence level, the planned confidence interval rule declares non-inferiority only if $L > -M$. Equality with the boundary does not satisfy this strict rule. Confidence level and one-sided significance level must be reconciled explicitly. [2, 3]

Specification	Required detail
Response	Baseline serostatus, assay result threshold and endpoint assessment.
Timing	Dose-relative visit window and selection of repeat assessments.
Population	Primary population, exclusions and endpoint denominators.
Interpretation	Contrast direction, margin, interval method and strict decision boundary.

The non-inferiority margin requires clinical and statistical justification. It must not be widened merely to achieve a smaller recruitment target. An immunogenicity endpoint also does not, by itself, establish comparative clinical protection. [3]

02 | INTERVAL ESTIMATION

Newcombe Hybrid Score confidence intervals

Why the method matters

For high seroconversion rates, few non-responders may be observed. Interval behaviour can then be sensitive to the method chosen. Newcombe's hybrid-score approach combines Wilson score limits for the two independent proportions. It differs from simply subtracting the two interval endpoints. The exact variant, including continuity correction, must be prespecified. [1, 2]

Construction without continuity correction

For group i in $\{T, C\}$, let p_i be the observed response proportion and n_i its denominator. For a two-sided confidence level $1 - \alpha$, let z be the standard-normal quantile at $1 - \alpha/2$. Wilson limits are obtained from:

$$ci = (p_i + z^2 / (2n_i)) / (1 + z^2 / n_i)$$
$$hi = z / (1 + z^2 / n_i) \times \sqrt{p_i(1 - p_i)/n_i + z^2/(4n_i^2)}$$
$$Li = ci - hi; Ui = ci + hi$$

With $d = p_T - p_C$, combine the component limits as follows:

$$L = d - \sqrt{(p_T - LT)^2 + (UC - p_C)^2}$$
$$U = d + \sqrt{(UT - p_T)^2 + (p_C - LC)^2}$$

Then compare L with $-M$. These equations describe estimation of the observed risk difference; they do not insert the non-inferiority margin into the Wilson component intervals. [2]

Implementation controls

Control	Reason
Independent groups	Use the independent-proportions construction, not a paired-data method.
Ordering	Verify treatment rows and response columns before reading software output.
Confidence setting	Explicitly reconcile ordinary interval ALPHA settings with one-sided NI settings.
Variant	Record continuity correction choice and software/version options.
Rounding	Apply the decision to full-precision limits; round only for presentation.

A Newcombe interval is not a Farrington-Manning test. If a separate hypothesis-test p-value is requested, its method and relationship to the confidence interval decision must be stated. SAS documentation distinguishes these options. [2]

03 | SAMPLE SIZE

Verify power under the intended decision rule

Reconstruct the planning calculation

A complete audit needs the original and revised calculation, assumptions, software settings and an explanation of the discrepancy. Those records are not available for numerical reproduction in this report. Consequently, the exact error mechanism and the size of the claimed recruitment reduction are not independently established here.

Planning input	Audit question
Expected responses	Which assumed test and comparator rates were used, and why?
Margin and power	Are the clinically justified margin, significance level and target power consistent?
Allocation	Are denominators per group or total, and is the allocation ratio preserved?
Evaluability	Is the allowance a proportion of randomized participants, and which losses does it cover?
Method alignment	Does calculated power correspond to the final interval-based decision rule?

Decision-rule power using binomial enumeration

For specified true planning rates q_T and q_C , independent counts follow $X_T \sim \text{Binomial}(n_T, q_T)$ and $X_C \sim \text{Binomial}(n_C, q_C)$. For each possible count pair, construct the prespecified Newcombe interval and evaluate its lower limit.

Power = sum over x_T, x_C of
 $I\{L(x_T, x_C) > -M\} \times \Pr(X_T = x_T) \times \Pr(X_C = x_C)$

Search candidate allocations for the smallest design meeting target power, using stable binomial probability calculations. Verify operating characteristics at the null boundary and across relevant nuisance response rates; nominal coverage does not imply exact finite-sample type I error control. Simulation is an alternative when fully enumerated evaluation is impractical, with a fixed seed and Monte Carlo uncertainty reported.

Inflation and rounding

If N_e is the required evaluable total and r is the anticipated non-evaluable fraction of randomized participants, the starting recruitment calculation is $N_r = \text{ceiling}[N_e / (1 - r)]$. Round further as needed to preserve allocation and operational constraints. Multiplying N_e by $(1 + r)$ is not equivalent.

Avoid double-counting baseline seropositivity, withdrawal and missing endpoints within a single allowance. Method changes are not inherently sample-saving; the effect depends on the assumptions and decision rule. FDA guidance connects NI sample size to margin and power. [3]

04 | POPULATIONS AND DATA

Protect the connection between enrolment and analysis

Population governance

The primary per-protocol analysis and supportive full analysis set assessment need explicit membership rules. Baseline serostatus, exposure, endpoint availability, completion requirements and major deviations must be translated into traceable participant-level flags. Reasons for exclusion should be agreed during blinded review. [4]

Component	Technical requirement
Baseline serostatus	Distinguish negative, positive, missing and unresolved assay status.
Dose and sample links	Derive assessment timing from the correct dose date and actual collection date.
Repeat samples	Apply prespecified selection and repeat-assay rules, preserving traceability.
Major deviations	Use clinically agreed classifications; avoid outcome-driven exclusion rules.
Observed-data FAS	Separate FAS membership from the number contributing observed endpoint data.
Missingness	Summarize endpoint availability and documented reasons by group.

Supportive analyses and interpretation

PP and FAS results should be compared with a reconciliation of contributing participants. Agreement supports interpretation but does not eliminate bias from missing outcomes or exclusions. In non-inferiority studies, neither population should automatically be assumed conservative. [4]

Where an observed-data approach is prespecified, no-imputation reporting still requires transparent missingness summaries. Additional sensitivity analyses should be defined for the actual missing-data risk and estimand; they should not be added selectively after viewing treatment results.

Recruitment monitoring

A blinded evaluability report can distinguish recruited, vaccinated, endpoint-assessed and PP-eligible participants. Age and site summaries may identify operational patterns without revealing treatment allocation. Any adaptation to recruitment must follow the approved protocol and governance process.

A recruitment reduction is justified only when the revised target still supports the planned evaluable population and statistical operating characteristics. Reducing the target without reassessing losses can replace excess recruitment with underpowering.

05 | PROGRAMMING AND VALIDATION

A reproducible SAS and R verification package

Implementation design

KITE-Ai specializes in SAS and R programming with documented validation and quality control for regulatory and non-regulatory studies. For this type of work, the verification package should separate dataset derivation from statistical calculation and decision reporting. The controls below describe the recommended package; completed execution logs were not supplied for this case.

Layer	Deliverable and acceptance check
Input derivation	Approved population and response specifications; traceable xT/nT and xC/nC counts.
SAS calculation	Explicit PROC FREQ risk-difference method and confidence settings; verified table ordering.
Independent R QC	Separate Wilson/Newcombe implementation or verified package with version and settings recorded.
Reconciliation	Compare response counts, risk difference, interval limits and final NI flag at full precision.
Reproducibility	Retain source code, logs, software versions, input provenance and execution records.

Meaningful numerical checks

Boundary responses: all responders, no responders and a single non-responder in either arm. Zero denominators must trigger an explicit unavailable-result state, never a fabricated confidence interval.

Unequal group sizes: verify that each Wilson component uses its own denominator. Swapping treatment labels should negate the estimated difference and reverse the interval signs appropriately.

Margin boundary: verify the strict lower-limit comparison, including values extremely close to -M. Display rounding must not change the computed decision.

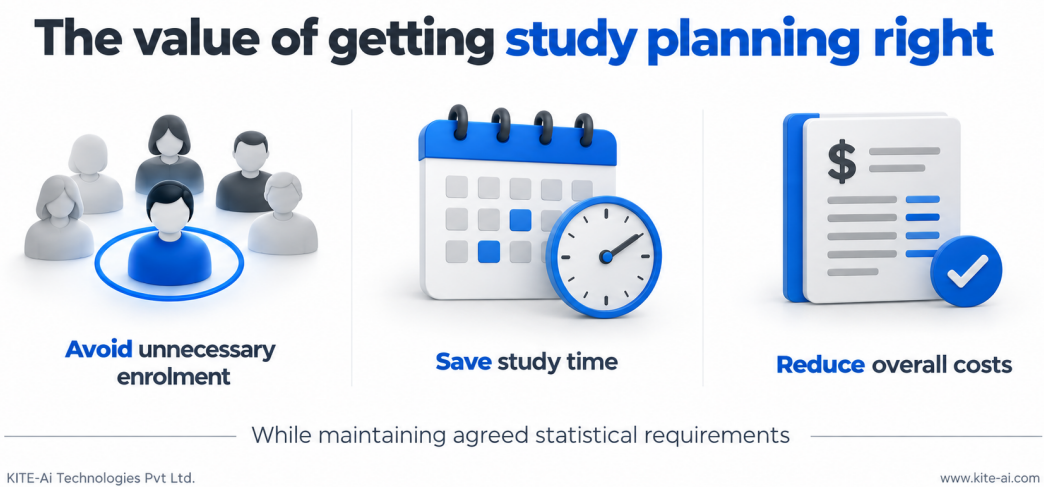
Power evaluation: compare enumerated probabilities with a separately implemented check or simulation; document seeds and uncertainty when simulation is used.

Required primary output

Report the analysis population, responders and denominators, group response percentages, test-minus-comparator difference, prespecified confidence interval, margin and NI decision. Footnotes should identify the method, correction variant, missing-data handling and endpoint timing.

Validation establishes that software implements the agreed specification. It does not independently justify the clinical margin, prove the planning assumptions or demonstrate regulatory acceptance.

Reported benefits and the evidence needed to quantify them



Case outcome

KITE-Ai reports that correcting the planning error helped avoid unnecessary enrolment, save time and reduce sponsor costs while preserving agreed statistical requirements. These are qualitative case outcomes; no before-and-after recruitment totals, dates or cost records are presented.

Benefit	Evidence for a measured claim
Avoided enrolment	Approved original and revised targets, with comparable evaluability assumptions.
Time saved	Recruitment plans and actual dates, distinguishing enrolment duration from fixed follow-up.
Cost reduction	Participant-level variable costs and revised budgets, net of review and amendment costs.
Statistical adequacy	Reproduced planning calculation, power/size evaluation and documented approval.

Cost attribution framework

Net avoided cost = avoided participants x avoidable variable cost per participant + other attributable avoided costs - incremental review and implementation costs. Include only costs that would actually have been incurred; fixed site costs do not necessarily fall when recruitment falls.

Calendar savings also depend on recruitment rate, site capacity and the follow-up schedule. They cannot be inferred directly from a participant reduction. The recommendation is a planning intervention, not evidence of a successful trial result or a demonstrated NI conclusion.

07 | CONCLUSIONS AND REFERENCES

A stronger basis for study commitments

Conclusion

The case illustrates the value of reviewing sample size assumptions, evaluability and analysis specifications together. KITE-Ai identified a planning error and recommended Newcombe's Hybrid Score confidence interval analysis. The reported sponsor benefits were lower unnecessary recruitment, time savings and reduced costs. Quantifying those benefits requires the original and revised calculation records and operational evidence.

Evidence basis

The case narrative follows information supplied by KITE-Ai for this report. No participant data, original vendor calculation, reproduced power output, change-control approval, cost ledger or timeline comparison was available for independent verification. Technical explanations describe the statistical framework and recommended verification work, rather than certifying that every listed control was executed in this engagement.

Methodological references

[1] Newcombe RG (1998). Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine* 17:873-890. [Source](#)

[2] SAS Institute. SAS/STAT User's Guide, PROC FREQ: Risks and Risk Differences; TABLES Statement. Consult the documentation matching the installed software release. [Source](#)

[3] US FDA. Non-Inferiority Clinical Trials to Establish Effectiveness. General NI design principles; not a product-specific vaccine endorsement. [Source](#)

[4] ICH E9. Statistical Principles for Clinical Trials, sections on analysis populations and missing data. [Source](#)

KITE-Ai capabilities

Biostatistics	Statistical programming
Study design and sample size review; SAP and mock shells; randomization; DMC support and TLF review.	SAS and R; SDTM and ADaM datasets; tables, listings and figures; independent QC; CSR and ISS/ISE packages.

Therapeutic areas: Oncology; respiratory (COPD); gastrointestinal; dermatology; hepatology; metabolism and nutrition disorders; reproductive system and breast disorders; infections and infestations.

KITE-Ai Technologies Pvt Ltd.
www.kite-ai.com | info@kite-ai.com | +91 84519 07235
Baner Business Bay, Pashan-Sus Road, Pune Bangalore Highway Pashan Exit, Behind Audi Showroom, Off Mumbai Baner
Anex, Pune, Maharashtra - 411045, India.