Study protocol with statistical analysis plan

Six-Month Change and Follow-up Patterns in AbilityScore During Routine Paediatric Developmental Care in India: Protocol for the PINNACLE-TRAJ Retrospective Multicentre Cohort Study

Koti Reddy Saripalli[1]

[1] Bharath Healthcare Laboratories Private Limited (Pinnacle Blooms Network), Plot No. 50/A, Raghavendra Nagar Colony, Jeedimetla, Suchitra, Hyderabad 500055, Telangana, India.

Correspondence: Koti Reddy Saripalli; kotii@bhclpl.org; ORCID: 0009-0007-7866-0737.

Manuscript BHCL-RET-2026-001 | Version 1.1 | 15 September 2026

Abstract 312 words | Main text 3513 words

Revised manuscript. Not peer reviewed.

Abstract

Background: Repeated developmental assessments collected during routine care can describe change and gaps in reassessment across a service network. Their interpretation depends on comparable measurement over time and inclusion of children who do not return. A clearly defined cohort and transparent analysis of missing follow-up are needed to characterize changes in AbilityScore, a developmental index recorded in paediatric services.

Objective: To estimate mean change in AbilityScore between baseline and a prespecified six-month assessment window in a baseline-defined cohort, and to describe the availability and characteristics of follow-up. Secondary objectives concern longer-term scores, recorded developmental domains and associations with baseline characteristics.

Methods: PINNACLE-TRAJ is a planned retrospective observational study of routine records from eligible Pinnacle Blooms Network centres in India. The proposed database close is 30 June 2026. A documented comparable scoring era and eligible centres will be fixed from deployment and measurement records before outcome inspection. Children aged under 13 years at their first eligible baseline will enter the cohort without any requirement for a subsequent visit. Baselines must permit the full 150–210-day primary window; with the proposed close, the latest permissible baseline is 2 December 2025. The primary endpoint is the comparable score nearest day 180 within that window. Mean within-child change will be estimated with multilevel multiple imputation under a stated missing-at-random assumption and centre-aware uncertainty. Observed-case, weighting and missing-not-at-random analyses will assess sensitivity. Postbaseline care exposure will not define cohort membership or the primary contrast.

Results: Ethics review, study registration and selection of the final scoring era are pending. The eligible cohort size will be established after the required approvals and deployment review. This protocol reports the planned study; results are not yet available for publication.

Conclusions: The study is intended to describe score change and reassessment patterns within an eligible service population. Its uncontrolled design cannot estimate treatment effectiveness, isolate an AI contribution, establish an important clinical difference or provide population developmental norms.

Keywords: child development; routine health data; retrospective cohort; developmental measurement; follow-up; missing data; AbilityScore; India

Introduction

Developmental care involves repeated decisions about a child's capabilities, support needs and participation in everyday activities. Routine assessment records can show how a recorded index changes during care and how consistently children are reassessed. Their value depends on knowing what was measured, whether scoring remained comparable and which children are absent from later assessments. Service records are generated by clinical and operational processes rather than a research visit schedule; an analysis restricted to returning children may therefore describe a selected group.

AbilityScore is described in a deposited methodology report as a clinician-observed composite developmental index ranging from 0 to 1000, interpreted relative to age expectations [1]. A separate protocol outlines concurrent-validity, test–retest and change-interpretation research [2]. Longitudinal analysis additionally requires attention to responsiveness, age adjustment and measurement comparability across software versions. The present study focuses on recorded score change during care and the availability of reassessment.

PINNACLE-TRAJ will examine the index as recorded during routine care. It will distinguish a descriptive change in the score from a treatment effect and from a change that children or families consider important. The first question is deliberately limited to a six-month window with a clear denominator. Longer trajectories and service-exposure associations are secondary. Results will follow the STROBE and RECORD reporting frameworks, including disclosure of record selection, coding rules, linkage and limitations [3,4].

Objectives

The primary objective is to estimate the child-weighted mean change in AbilityScore from the first eligible baseline to the assessment nearest day 180 within days 150–210, in a cohort

defined without reference to future attendance. The estimate will be accompanied by its uncertainty and the proportion with an observed, comparable primary-window score.

Secondary objectives are to describe reassessment at three, twelve and twenty-four months; changes in documented comparable domain indexes; distributions of score change; and variation by prespecified baseline age, recorded diagnostic category, sex, centre, calendar period and recorded equity-programme status. Exploratory analyses will examine temporally defined therapy and home-practice records.

Methods

This is a retrospective, noninterventional, multicentre cohort study of routinely collected records. The research will not assign care, modify treatment or schedule additional assessments. BHCL is the proposed study sponsor and data custodian. Participating centres, data-access responsibilities and independent statistical review arrangements will be documented before research extraction.

Setting, calendar limits and scoring-era selection

The candidate record interval is 1 January 2020–30 June 2026. These dates describe a proposed search boundary, not verified continuous instrument availability. A deployment audit will identify score algorithms, assessment forms, age-norm tables, domain definitions, rater instructions, release dates and centre adoption dates. The primary analysis will use one documented stable scoring system, or versions whose equivalence is supported by an adequate measurement bridge. Similar score distributions before and after a release will not be accepted as proof of equivalence.

The eligible centre-by-calendar intervals will be determined from these records before inspecting study outcome values or changes. An interval must allow every selected baseline the full 210-day primary follow-up opportunity under the comparable system. Where several

suitable intervals exist, all will be included under the same documented rule; intervals will not be selected for favourable trajectories. A scoring-era manifest will identify included versions, centres, dates, exclusions and supporting documents.

The proposed database close is 30 June 2026. Consequently, 2 December 2025 is the latest baseline with a complete 210-day opportunity. A later baseline is administratively ineligible for this horizon, not a missing six-month outcome. If deployment evidence cannot support an adequate comparable interval, the investigators will revise and timestamp the feasibility and design decision before outcome analysis. A revised database close or start date will be an explicit amendment, with recalculated horizon limits.

The deployment audit will also document changes in assessment modality, rater training, centre operation and data capture, including any periods affected by pandemic restrictions or transition to remote assessment. These features may alter either the score or the probability of reassessment. The manifest will distinguish a change in measurement from a disruption in service availability. Prespecified calendar summaries will describe these periods, and sensitivity analyses will exclude a documented incompatible assessment modality when necessary. Calendar adjustment alone will not be treated as evidence that different administration methods measure the same construct.

Assessment windows

Day 0 is the eligible baseline assessment date. All intervals use elapsed calendar days. The target and window limits are operational assessment definitions, not exact biological ages. At each horizon, the eligible risk set is determined from baseline and the database close before examining whether a follow-up occurred.

Participants and baseline definition

Children enter at their first valid AbilityScore assessment within an eligible centre-by-calendar interval, provided their age is at least 0 and less than 13 years and their score,

assessment date, age and scoring version can be established. Baseline is the first eligible observation in this measurement era, not necessarily the child's first-ever therapy contact. Prior recorded care and any earlier assessment will be summarized when available; no arbitrary washout is imposed or claimed. The documented scoring-age range must also cover the child's full primary window, assessed from baseline age and calendar dates; a baseline that would necessarily age beyond that range will be ineligible for this horizon and counted explicitly.

No later assessment, minimum duration of therapy, home-practice activity or score increase is required. Children who discontinue care or transfer remain in the baseline cohort. A stable network child identifier will identify repeat records across centres. Each child contributes one baseline; transfers will not create a new participant or new baseline. The baseline centre defines the primary clustering unit, with cross-centre follow-up recorded.

Test, demonstration and training records will be removed using documented production flags. Duplicate and superseded records will be resolved through record identifiers and completion status; same-day assessments will use the last finalized non-superseded assessment timestamp. Records will not be deleted solely because a score is unusually low, high or declines. Invalid dates, impossible ages or out-of-range scores will be checked against source metadata by the authorized custodian. Unresolved invalid baseline records will be excluded with reasons; invalid follow-up values will be marked unavailable without removing the child.

Cohort flow and record provenance

A flow diagram will distinguish candidate records, deduplicated children, baseline exclusions, the primary cohort, observed primary-window assessments and each category of unavailable outcome. Baseline-era exclusions and later missingness will remain separate.

Counts will reconcile across the flow diagram, tables and analytic datasets; no network-wide service total will be used as a participant denominator.

Measurement and endpoint construction

The recorded AbilityScore total is the primary measurement. Historical totals will not be silently recomputed using a newer algorithm. The measurement appendix accompanying the final scoring-era manifest will describe the scale, component domains, scoring direction, age adjustment, permissible range, administration, rater training and version bridge, including any information withheld for intellectual-property reasons. Restricted details must not prevent a reader from understanding the construct and evaluating comparability.

For each child, select the valid comparable assessment with the smallest absolute distance from day 180 among days 150–210, breaking equal-distance ties in favour of the earlier day. Within a day, use the finalized-record rule above. Define $D_i = Y_{i6} - Y_{i0}$. If no such assessment exists, the endpoint is missing; an out-of-window or incompatible score is not automatically substituted. The same target-distance rule applies to secondary windows.

Score change will be reported in points. Higher scores will be interpreted only according to the documented instrument definition; an age-adjusted increase does not necessarily equal acquisition of the same functional skills at every age. Domain results will be included only for domains with established definitions and comparable historical capture. The planned domains are speech, motor, behavioural, cognitive, school readiness, self-sufficiency and mainstream participation; their exact operational labels and scoring will be confirmed in the manifest.

The proportion with an increase of at least one-half of the primary cohort's baseline standard deviation will be a secondary distributional summary among children with an observed comparable primary-window endpoint. The threshold will be computed once from valid baseline scores and recorded before follow-up outcomes are released. It is not an

AbilityScore minimal clinically important difference, a validated responder definition or evidence that an individual benefited from care. Independent anchors and population-specific interpretation would be needed to support such conclusions [5].

Primary estimand and intercurrent events

The target is the mean six-month window score change in children meeting the baseline definition, with every child weighted equally. Conceptually, it describes the comparable score that would have been recorded under reassessment within the specified window, while care proceeds as it actually occurred. The identifying assumption for unobserved scores is stated in the missing-data analysis; reassessment itself is not assumed to cause or prevent developmental change.

Stopping therapy, recorded home-practice nonparticipation and transfer do not remove children or reset time. An assessment obtained at another eligible network centre remains usable if its measurement version is comparable. No outcome after a documented death will be imputed. Such records will remain in the cohort flow and be reported separately as an undefined rather than ordinary missing score. If a death before day 210 is documented, the numerical score-change estimand will be explicitly reported for children not recorded as having died before the end of the window; it will not be portrayed as an outcome encompassing all baseline children. Routine records may not provide complete vital-status ascertainment, and this limitation will be stated.

Baseline variables and service records

The core baseline variables are score, age in months, sex as recorded, primary recorded diagnostic category, baseline centre, calendar date and recorded equity-programme status. Diagnostic categories will be mapped before outcomes are inspected: autism spectrum disorder, speech/language developmental concern, global developmental delay, attention-deficit/hyperactivity disorder, cerebral palsy, other recorded category, documented absence of

an established category, and unrecorded/unknown category. These are record classifications, not diagnoses made by the study. Multiple diagnoses will be retained as additional flags where available. Missing or unrecorded categories will remain distinct from a documented negative finding.

Family language, prebaseline care, rater identifier, service counts, home-practice activity and documented reasons for discontinuation will be used only if their definitions and capture periods are sufficiently clear. Equity-programme enrolment is a programme indicator, not a validated measure of household income. Documentation of home activity measures recording in the platform and does not establish all activity performed by the family.

For exploratory longitudinal associations, therapy exposure at an assessment on day t will be the recorded number of completed services from day max(0,t-60), inclusive, to day t-30, exclusive, divided by the number of calendar days in that interval and multiplied by 30. Home activity will use the same lagged interval, distinguishing recorded activity, documented zero activity and unknown capture. The lag prevents the predictor from incorporating future sessions or the immediate assessment interval. These variables will not define eligibility or the primary mean-change estimate.

Data preparation, access and quality control

Research extraction will occur only after the required ethics and institutional decisions. An authorized custodian will implement a version-controlled extraction specification with record counts, hashes and a transformation log. Direct identifiers and free-text clinical notes will not be included. Study-specific pseudonyms will replace child and rater identifiers; linkage keys will remain with the custodian, separately protected. Pseudonymization is not treated as proof of anonymity.

Exact elapsed days, baseline age and required calendar variables will be derived accurately inside the protected environment before date minimization. The released analytic file will

contain elapsed days and only the calendar detail required by this SAP. Data will be processed in an approved access-controlled environment with logging, encryption, role-based permissions and a retention schedule agreed with the ethics committee and custodian. No identifiable records will be uploaded to public AI services.

Quality checks will cover duplicate children, impossible chronology, missing versions, domain-total inconsistencies where checkable, abrupt changes in record capture, centre transfers and field completeness. Quality review will not select records on favourable change. A masked preparation stage may examine schemas, counts, missingness and baseline distributions as authorized; follow-up values and change summaries will remain unavailable to the analyst until the plan and code are frozen. Any unavoidable earlier knowledge will be logged.

Sample size and precision

All children satisfying the finalized baseline definition will be considered; the eligible sample size is currently unknown. No network attendance or service count is taken as the study sample. A feasibility report will present the number of children, centres, calendar intervals, baseline distributions and ascertainment opportunities after approvals and before outcome inspection. Precision depends on between-child variation, clustering, missingness and the validity of imputation assumptions. A large administrative database cannot compensate for poor comparability or highly selective reassessment.

This is an estimation study without a minimum effect chosen to demonstrate success. Confidence intervals and the fraction of missing information will be central. Few contributing centres or near-absent follow-up in important strata will restrict inference. If the primary outcome is wholly unavailable, the study will report feasibility and follow-up findings rather than produce a fabricated or entirely extrapolated mean change.

Primary analysis

The primary analysis will use one row per baseline child and the selected six-month endpoint. Missing scores will be multiply imputed under a model that includes centre clustering and prespecified predictors of the score and its availability [6–8]. In each completed dataset, compute the arithmetic mean of D_i across children. Estimate its within-imputation variance by an intercept-only linear model with CR2 bias-reduced cluster-robust variance at the baseline-centre level. Combine point estimates and variances with Rubin's rules and a small-sample degrees-of-freedom adjustment [6,9].

The contrast is marginal mean change, not a regression coefficient for treatment or a covariate-adjusted causal effect. No postbaseline exposure is conditioned on to redefine this contrast. Report the imputed estimate, 95% confidence interval, observed baseline denominator, observed primary-outcome denominator, number and proportion imputed, number of centres and missing-information diagnostics. The unadjusted mean among observed primary-window cases will appear alongside it, clearly identifying the selected population.

The analytic plan assumes that valid endpoints lie from 0 to 1000. Imputation will occur on a logit-transformed scale, with all imputed scores returned to this range and observed scores preserved exactly. Boundary frequencies and observed-versus-imputed distributions will be reported. Modelling details, convergence checks, donor-based sensitivity and failure rules are specified in Supplement S1. Model diagnostics will not justify changing the primary endpoint or deleting inconvenient records.

Missing-data assumptions and sensitivity analyses

Missing at random means that, conditional on the prespecified baseline and auxiliary record information, unobserved window scores have the same conditional distribution as observed scores. This is untestable and may fail if families discontinue for unrecorded improvement,

deterioration, cost, relocation or dissatisfaction. Administrative ineligibility, missed assessment, unavailable data fields and structurally incompatible scoring will be tabulated separately. Missingness patterns and baseline standardized differences will be reported without treating significance tests as proof of comparability.

An alternative analysis will weight observed primary-window changes by their estimated inverse probability of assessment, using a prespecified model and reported positivity and weight diagnostics [10]. This assesses a different modelling route to the same target under its assumptions; it does not validate the missing-at-random premise. The primary estimate will also be repeated with an alternative bounded imputation specification.

Missing-not-at-random sensitivity will shift imputed missing six-month scores both downward and upward by prespecified multiples of the baseline SD, retaining observed scores and respecting the scale boundaries. The full estimate-versus-shift curve will be reported, with any truncation at boundaries and the shifts at which conclusions change. These are assumption scenarios, not estimated corrections [11]. Additional analyses will use a wider 120–240-day window with its own baseline opportunity limit, a narrower 165–195-day window, records with unchanged raters and a documented stable-version subset where relevant. Restricted analyses will state their altered populations.

Secondary and exploratory analyses

Three-month summaries will use the primary cohort; twelve- and twenty-four-month summaries will use nested cohorts with full calendar and measurement opportunity for their respective windows. Observed availability and denominators will accompany every estimate. Comparable domain changes and distribution-based increases will be secondary, with no confirmatory claim based on a favourable subset of domains or horizons.

A secondary mixed-effects model will describe postbaseline scores up to day 760 in the appropriate comparable era. It will include the baseline score as a covariate, but not

simultaneously as a repeated response; the time basis and random-effects structure are prespecified in Supplement S1 [12]. Only times supported by the relevant baseline-defined cohort will be summarized. Children with no later score remain in cohort reporting, although the observed-data likelihood contributes no postbaseline response for them. This model does not replace the primary imputed six-month window estimate.

Baseline age bands will be 0 to under 3, 3 to under 6, 6 to under 9, and 9 to under 13 years. Other planned strata are recorded sex, diagnostic category, equity-programme status, state/union territory and calendar period. Associations will be expressed with estimates and uncertainty; cells with fewer than 50 observed endpoints will not receive model-based subgroup comparisons. This is a stability rule, not proof that 50 observations are adequate. A small-cell disclosure policy may require further suppression.

Lagged care-exposure associations are exploratory and vulnerable to confounding by need and time-varying clinical decisions. Their interpretation will account for the possibility that children receiving more support differ systematically from children receiving less support.

Bias, interpretation and generalisability

Within-child change can reflect maturation, measurement error, regression to the mean, changing assessment practice and selective observation as well as changes during care. Baseline adjustment does not remove all these explanations [13]. The study has no untreated comparison and cannot separate the contribution of individual therapies, home practice, the digital platform or other support.

The eligible centres and families will not constitute a probability sample of Indian children. Generalisability will be assessed through the documented locations, ages, recorded conditions, language availability, measurement era and follow-up patterns. A centre ranking would require additional case-mix, measurement and quality evidence and is not a study

objective. A descriptive increase in an index will not be equated with school admission, independence or a guaranteed outcome.

Ethics, registration and governance

The protocol will be submitted to a competent DHR-registered ethics committee. Approval and any consent waiver are pending. A waiver will be requested for secondary use of minimized routine records, explaining feasibility of recontact, risks, safeguards and effects on participants' rights; the committee may require consent or other conditions. The researcher will not presume a waiver [14]. Multicentre scientific and local implementation review, data-custodian permission and site arrangements will follow the applicable ICMR framework [15]. Research extraction will not begin before the required decisions are documented.

The investigators intend to register the study with CTRI before initiating research data collection if eligible and required or retained as a voluntary commitment. CTRI accepts observational studies, requires ethics documentation and does not accept a retrospective data-collection study after that research collection has begun. A submitted application is not an assigned CTRI registration [16]. ICMJE encourages but does not universally require registration for non-trial observational designs [17]. Applicable ethics, institutional and journal conditions will be recorded; dates and prior study activity will be reported truthfully.

An OSF Secondary Data Preregistration will timestamp the approved analysis plan before research outcome inspection [18]. It will state that the care data already exist and disclose which investigators or company staff have previously accessed relevant records, dashboards or analyses. Earlier analyses, if any, will be identified rather than relabelled as prospectively specified. Repository deposit, registration and peer-reviewed publication will remain separately identified in study communications.

Data processing will be reviewed against Indian requirements applicable at the time of processing, including the Digital Personal Data Protection Act and the phased commencement of its 2025 Rules [19]. Public sharing will include disclosure-reviewed aggregates, analysis code and synthetic examples. A data-access statement will explain restrictions on individual-level records and the resulting limits on independent reproduction of the models.

Study status and dissemination

As of this protocol version, the final eligible scoring era, cohort count, ethics decision and registry identifiers are pending. The deployment review and authorized feasibility checks will determine the final eligible records. A dated study-status statement will accompany registration and submission.

After approvals and the scoring-era decision, the sequence will be: finalize the extraction specification; complete permitted masked feasibility checks; freeze the protocol, SAP and synthetic-data-tested code; timestamp registration and any amendments; release authorized outcome data; execute the analysis; independently check tables and interpretations; and submit the results. Amendments after outcome access will state the date, reason, prior knowledge and affected analyses. No month-by-month completion promise is made.

Results will be reported regardless of the direction or size of change, including inability to estimate the primary target reliably. The protocol and results will be linked, and related analyses of overlapping records will be disclosed. Parent or caregiver input on interpretation and accessible summaries is planned but has not yet occurred. Its actual scope and contributors will be reported. Public-facing summaries will preserve the study population, missingness and noncausal interpretation.

Discussion

The contribution of PINNACLE-TRAJ is a transparent account of a developmental index and its reassessment process within routine Indian services. The baseline-defined denominator makes loss of follow-up visible, while a documented measurement era reduces ambiguity about whether repeated totals are comparable. Separating a prespecified six-month endpoint from exploratory trajectories should make the primary result easier to interpret and reproduce.

The protocol cannot resolve absent measurement validation, recover unrecorded life outcomes or remove unmeasured selection through statistical sophistication alone. The analysis may need to conclude that available records support only limited descriptive claims. Conversely, if measurement and follow-up are sufficiently documented, the findings could identify priorities for independent validation, prospective outcome measurement and future comparative research. Those studies would answer different questions and require their own designs.

Declarations

Ethics approval and consent to participate

Ethics approval and a waiver of individual consent are pending. No approval identifier is claimed. Research extraction will begin only after the applicable ethics, institutional and site permissions are documented. Any requirement for consent or other safeguards will be implemented as directed.

Consent for publication

This protocol contains no identifiable individual participant information, images or case histories. Any future participant-specific material would require an appropriate separate consent and disclosure review.

Registration and protocol availability

CTRI and OSF registration are pending. Study identifier BHCL-RET-2026-001 is an internal identifier, not a registry number. Version 1.1 is dated 15 September 2026. The final scoring-era manifest, SAP, access declaration and subsequent amendments will accompany the registered plan.

Funding and sponsor role

The study is supported by Bharath Healthcare Laboratories Private Limited. No external funding is reported. The sponsor is the data custodian and will provide authorized routine records and organizational resources, subject to ethics and data-governance requirements. Independent statistical review and a commitment to report results regardless of direction are planned.

Competing interests

Koti Reddy Saripalli is Founder and Chairman of Bharath Healthcare Laboratories Private Limited, which operates Pinnacle Blooms Network and develops the system being studied. This leadership and commercial relationship is relevant to the design, conduct and interpretation of the research. Independent statistical review, a frozen analysis plan and transparent reporting of amendments are planned safeguards.

Author contributions

Koti Reddy Saripalli: conceptualisation, methodology, writing—original draft, writing—review and editing, and project administration. Additional contributors will be named according to their actual contributions and authorship eligibility when appointed.

Data and code availability

No study dataset accompanies this protocol. The planned release will include the analysis code, data dictionary, synthetic examples, extraction logic that can safely be disclosed, and

disclosure-reviewed aggregates. Patient-level access will remain subject to ethics, legal and custodian restrictions. Public aggregates will not be described as sufficient to reproduce all individual-level analyses.

Use of generative artificial intelligence

Generative AI was used for drafting assistance, manuscript restructuring, language editing, and support with reference and formatting work. No patient-level data were supplied to these tools. The author retains responsibility for scientific accuracy, source verification and interpretation; tool details will be disclosed according to the receiving journal's policy.

Acknowledgements

None at this protocol stage.

References

1. Saripalli KR, Ramtenki R, Boini SY. AbilityScore: Development and Preliminary Psychometric Characterization of a Standardized Composite Index for Quantifying Child Developmental Readiness [repository report]. Zenodo; 2026 Apr 9. doi:10.5281/zenodo.19482123. https://zenodo.org/records/19482123

2. Saripalli KR, Ramtenki R, Boini SY. Concurrent Validity, Test-Retest Reliability, and Preliminary Minimal Clinically Important Difference of PinnacleAI AbilityScore: A Cross-Sectional Psychometric Validation Study Protocol [repository protocol]. Version 2.1. Zenodo; 2026 Apr 9. doi:10.5281/zenodo.19482476. https://zenodo.org/records/19482476

3. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453–1457. doi:10.1016/S0140-6736(07)61602-X. https://pubmed.ncbi.nlm.nih.gov/18064739/

4. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al.; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015;12(10):e1001885. doi:10.1371/journal.pmed.1001885. https://pubmed.ncbi.nlm.nih.gov/26440803/

5. Wyrwich KW, Norquist JM, Lenderking WR, Acaster S; Industry Advisory Committee of International Society for Quality of Life Research. Methods for interpreting change over time in

patient-reported outcome measures. Qual Life Res. 2013;22(3):475–483. doi:10.1007/s11136-012-0175-x. https://pubmed.ncbi.nlm.nih.gov/22528240/

6. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med. 2011;30(4):377–399. doi:10.1002/sim.4067. https://pubmed.ncbi.nlm.nih.gov/21225900/

7. Enders CK, Mistler SA, Keller BT. Multilevel multiple imputation: a review and evaluation of joint modeling and chained equations imputation. Psychol Methods. 2016;21(2):222–240. doi:10.1037/met0000063. https://pubmed.ncbi.nlm.nih.gov/26690775/

8. Quartagno M, Grund S, Carpenter J. jomo: a flexible package for two-level joint modelling multiple imputation. R J. 2019;11(2):205–228. doi:10.32614/RJ-2019-028. https://journal.r-project.org/articles/RJ-2019-028/

9. Pustejovsky JE, Tipton E. Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. J Bus Econ Stat. 2018;36(4):672–683. doi:10.1080/07350015.2016.1247004. https://www.tandfonline.com/doi/abs/10.1080/07350015.2016.1247004

10. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. Stat Methods Med Res. 2013;22(3):278–295. doi:10.1177/0962280210395740. https://pubmed.ncbi.nlm.nih.gov/21220355/

11. Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: a practical guide. Stat Med. 2020;39(21):2815–2842. doi:10.1002/sim.8569. https://pubmed.ncbi.nlm.nih.gov/32419182/

12. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982;38(4):963–974. doi:10.2307/2529876. https://pubmed.ncbi.nlm.nih.gov/7168798/

13. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. Int J Epidemiol. 2005;34(1):215–220. doi:10.1093/ije/dyh299. https://pubmed.ncbi.nlm.nih.gov/15333621/

14. Indian Council of Medical Research. National Ethical Guidelines for Biomedical and Health Research Involving Human Participants. New Delhi: ICMR; 2017. Available online

15. Indian Council of Medical Research, Bioethics Unit. Frequently asked questions: operational guidelines for single ethics review of multicentre research in India and research ethics committee registration [Internet]. 2026 [cited 2026 Sep 15]. https://ethics.ncdirindia.org/Fre_AQ.aspx

16. Clinical Trials Registry–India. Frequently asked questions [Internet]. Indian Council of Medical Research [cited 2026 Sep 15]. https://www.ctri.nic.in/Clinicaltrials/faq.php

17. International Committee of Medical Journal Editors. Recommendations: clinical trials—registration [Internet; cited 2026 Sep 15]. Available online

18. Center for Open Science. Welcome to registrations and preregistrations: Secondary Data Preregistration template [Internet; cited 2026 Sep 15]. https://help.osf.io/article/330-welcome-to-registrations

19. Ministry of Electronics and Information Technology, Government of India. Digital Personal Data Protection Rules 2025 and enforcement timeline documents [Internet]. 2025 [cited 2026 Sep 15]. Available online

20. Bailey BE, Andridge R, Shoben AB. Multiple imputation by predictive mean matching in cluster-randomized trials. BMC Med Res Methodol. 2020;20(1):72. doi:10.1186/s12874-020-00948-6. https://pubmed.ncbi.nlm.nih.gov/32228491/

Table 1. Prespecified assessment horizons with the proposed 30 June 2026 database close

Horizon	Target day	Inclusive window	Latest baseline for full opportunity
Primary six months	180	150–210	2 December 2025
Secondary three months	90	75–105	17 March 2026
Secondary twelve months	365	330–400	26 May 2025
Secondary twenty-four months	730	690–760	31 May 2024

The primary cohort has the six-month cutoff. The three-month analysis within that cohort does not add later baselines merely because the shorter window permits them. Longer horizons use nested baseline-defined subsets with comparable measurement coverage for their full windows.

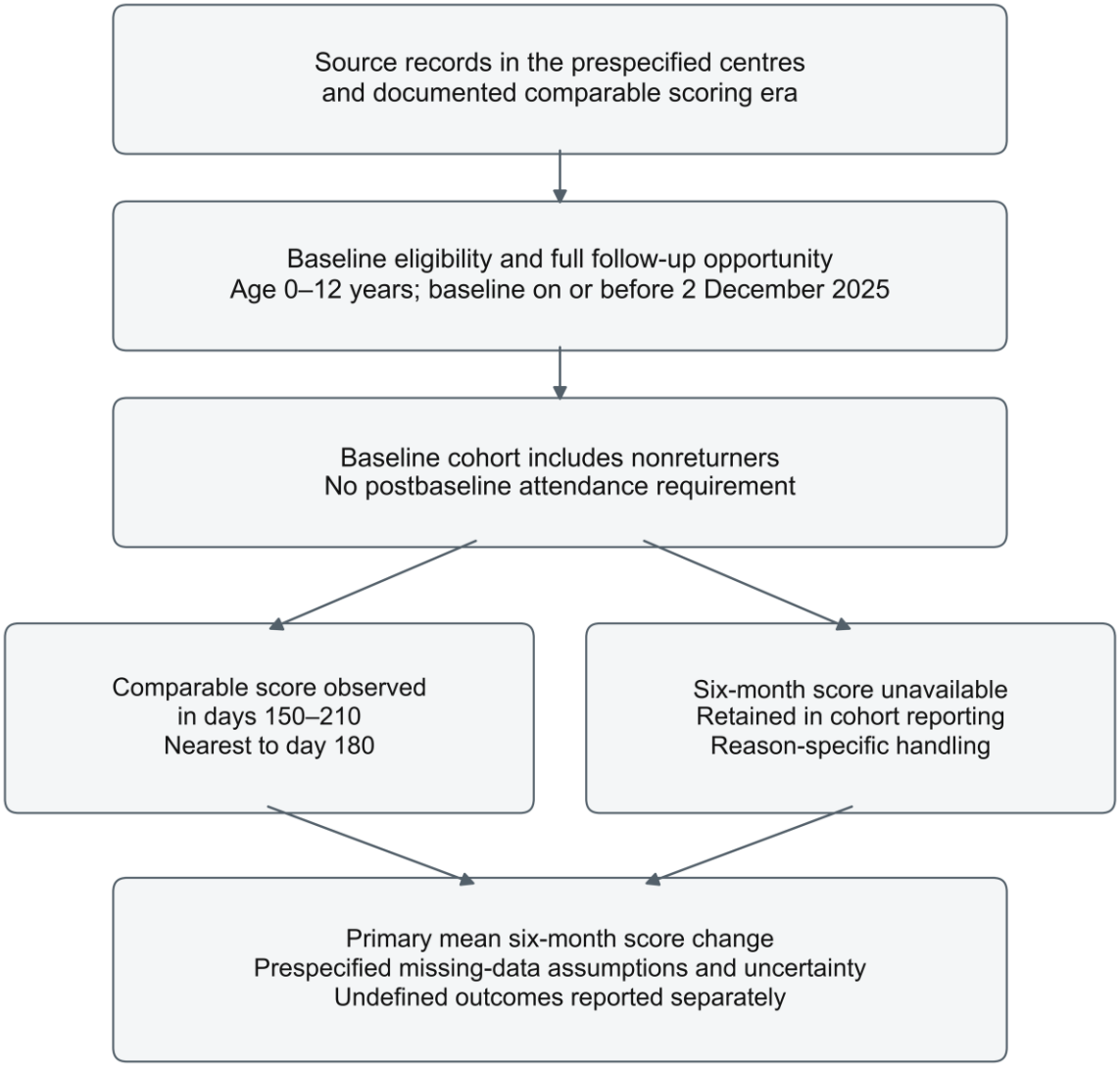

Figure 1. Planned cohort derivation and primary-outcome availability. Protocol schematic only; no study counts are displayed. Candidate routine records pass through deployment-defined centre/version eligibility, deduplication and baseline eligibility. Children remain in the cohort whether a comparable primary-window score is observed or unavailable. The reporting flow will explain unavailable scores, including a separate undefined-outcome category for any recorded death, and reconcile every branch to the baseline denominator.

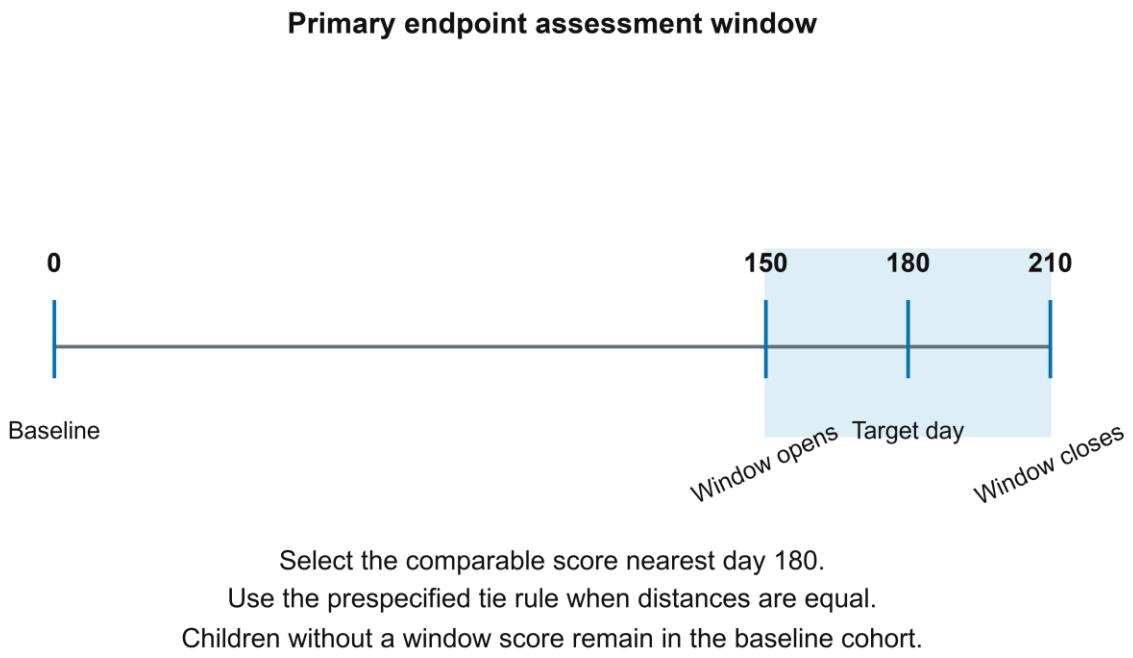

Figure 2. Primary endpoint and complete follow-up opportunity. Protocol illustration only. Baseline is day 0; the primary endpoint is the comparable score nearest day 180 within the inclusive day 150–210 window, with earlier-date ties. A child without a score in that window remains in the baseline cohort. With a proposed database close of 30 June 2026, a baseline must be on or before 2 December 2025 to permit the entire 210-day window. No trajectory or outcome values are shown.

Supplementary material

Companion material to BHCL-RET-2026-001 | 1.1

Supplement S1. Prespecified statistical implementation

Analysis units and datasets. Maintain a baseline file with one row per child, an assessment file with one row per finalized record and a service file with the minimum necessary dates and categories. The primary wide file contains Y0, selected Y3 and Y6, target distances, reason-coded missingness, baseline covariates, baseline centre, transfer flags and defined auxiliary records. Derive the file from a deterministic script; retain excluded-record counts. Do not restrict it to children with follow-up. Long-horizon files are nested by full opportunity and scoring-age coverage, not attendance.

Primary model specification. Let $Z_k = \text{logit}\{\min[0.9995,\max(0.0005,Y_k/1000)]\}$ for k=3,6. Observed scores, including 0 and 1000, remain unchanged in analysis. Jointly model (Z_3, Z_6) using a multivariate normal model with a random intercept vector for baseline centre and a common within-centre residual covariance. Fixed predictors are an intercept; baseline score per 100 points; natural cubic spline of baseline age in months with internal knots at 36, 72 and 108 months and boundary knots at 0 and 156 months; recorded sex; diagnostic category; equity-programme record status; baseline calendar year; and baseline quarter. Categories have an explicit unrecorded level. These are features of available records rather than imputed diagnoses or income.

Auxiliaries. Add log(1+number of completed services in days 0–90) and the corresponding record-completeness indicator only when capture is demonstrably complete or its missingness is identifiable. Include recorded prebaseline-care status as yes/no/unknown. An absent platform home entry will not be coded as no home practice. If a proposed auxiliary field was not collected in the entire eligible era, omit that field globally and document the fact before

unmasking. A predictor with only one observed level or an exactly dependent design column will be removed by a deterministic rank check, not selected by its association with change.

Imputation execution. Use the jomo1rancon implementation of the continuous two-level joint model with random intercepts and common residual covariance [8]. Freeze the exact function, package version, prior settings, predictor matrix, random seed and transformation in the code manifest. Start four chains with 5,000 burn-in iterations each and obtain 25 imputed datasets per chain separated by 1,000 iterations, for 100 imputations. Inspect trace plots, between-chain agreement and effective sample sizes for fixed effects, covariance parameters and the mean missing score. Extend chains if split R-hat exceeds 1.01 or effective sample size is below 400 for these monitored parameters. These are computational checks, not tests of MAR.

Completed-data estimates. Back-transform imputed scores using 1000×inverse-logit(Z), leaving all observed scores exact. Compute Q_m = mean(Y6,m-Y0). Obtain U_m from an intercept-only OLS fit with baseline-centre CR2 variance. Pool using Qbar and T = mean(U_m)+(1+1/M)×var(Q_m). Use the Barnard–Rubin small-sample adjustment with complete-data degrees of freedom set conservatively to the smaller of G-1 and the smallest available CR2 Satterthwaite degrees of freedom across imputations. Archive all ingredients. If fewer than two centres contribute, CR2 is undefined and this planned multicentre inferential analysis will not be presented as executable without a dated amendment. Very small effective degrees of freedom will be flagged as unstable.

Monte Carlo and support checks. Compare Qbar across the four independent imputation batches. Increase the total imputations to 200 and then 400 if the Monte Carlo standard error of Qbar exceeds 10% of its pooled standard error. If convergence or this precision criterion remains unsatisfied, report the problem and seek a documented methodological amendment rather than choose a favourable run. Show observed and imputed score distributions by

baseline-age band and centre, boundary frequencies, number of centres with no observed Y6, and missing-information fractions. Entirely unobserved strata require extrapolation; flag them and provide a sensitivity analysis excluding such strata with its changed target population. Do not claim that diagnostics verify imputed values.

Inverse-probability analysis. Define R6=1 for an observed valid comparable primary-window endpoint. Fit a logistic model for R6 using the fixed baseline predictors above, a centre random intercept, and the defined first-90-day service auxiliary when available. Use stabilized weights mean(R6)/p_i among R6=1 children; normalize them to sum to the observed-case count. Report untruncated weights and estimates, then a prespecified estimate truncating weights at their empirical 1st and 99th percentiles. Report probability distributions, weight range, effective sample size and centre-specific outcome availability. No floor will silently overwrite an estimated probability. A centre with no observed outcome cannot be recovered by weighting; report that support failure. For uncertainty, use 1,000 baseline-centre bootstrap resamples, refitting the probability model each time. Failed fits and the number of successful replicates will be reported; fewer than 900 successful replicates precludes a routine bootstrap interval.

MNAR analysis. For each missing six-month score, add delta×SD0 to its MAR imputation and restrict the result to 0–1000. Use delta values -1.0,-0.75,-0.5,-0.25,0,0.25,0.5,0.75,1.0; SD0 is fixed from the primary baseline cohort. Repeat pooling and report the proportion hitting a boundary for each scenario. This bounded delta procedure is an explicitly chosen sensitivity scenario motivated by controlled imputation, not a fitted model of dropout [11]. Display both adverse and favourable missing-outcome scenarios. Report zero-crossing where relevant without treating it as a clinical-importance threshold.

Alternative imputation. Repeat the primary analysis using the multilevel PMM-draw approach described by Bailey and colleagues, with five donors and 100 imputations,

implemented and checked on synthetic clustered data before unmasking [20]. This tests sensitivity to the transformed-normal specification and preserves donor score bounds. Include the exact adaptation and software/code in the manifest because the cited method was evaluated in cluster trials, whereas this application is a nonrandomized service cohort. If a valid implementation is not available before code freeze, replace this named sensitivity through an explicit pre-analysis amendment rather than silently call ordinary single-level PMM equivalent.

Window sensitivity. For 120–240 days, restrict the original cohort to baseline on or before 2 November 2025 and confirmed comparable scoring coverage through day 240. For 165–195 days, keep the original cohort and use the narrower endpoint. Retain target day 180 and the same tie rule. State the denominator for each; widening a window does not merely increase sample size and may change what six-month change means.

Secondary longitudinal model. Use postbaseline assessments on days 1–760, subject to the relevant comparable measurement coverage. Fit by restricted maximum likelihood: score as response; baseline score per 100, baseline-age spline, recorded baseline covariates, calendar year and a natural cubic spline of elapsed days with internal knots at days 90, 180 and 365 and boundary knots at days 1 and 760 as fixed effects; baseline-centre and child random intercepts plus an uncorrelated child time slope (time in years). Baseline is not a response. If the slope model is singular or fails after scaled predictors and a second optimizer, remove the random slope; if the centre variance is then on the boundary, retain child effects and use centre fixed effects. Log this fixed fallback order. Report residual and boundary diagnostics. No population estimate will be extrapolated to unsupported follow-up horizons.

Secondary uncertainty and multiplicity. Report window means, distributions and availability first. Secondary model summaries are associations conditional on the recorded predictors. Obtain time-specific contrasts with their covariance and average predictions over each stated

baseline-defined horizon cohort; compare predicted follow-up score with that cohort's observed baseline mean. Use centre-aware intervals and document method. Domain and subgroup estimates are exploratory, with 95% intervals described as unadjusted for multiplicity; do not select a few significant findings as confirmed discoveries. Model-based subgroup comparisons require at least 50 observed endpoints per reported level and usable representation across centres. Descriptive results remain subject to disclosure control.

Exposure and rater analyses. Apply the prespecified 30-day lagged exposure window only at assessments after day 60; earlier visits are not used in exposure-association models. Enter therapy count as log(1+services per 30 days) and home-record status as activity/zero/unknown. Add them only to the exploratory model. Rater continuity summaries use observed rater identifiers; missing rater identifiers are not treated as a common rater. A same-rater analysis is a selected-subset sensitivity. Where a stable rater identifier and cross-centre employment records support an additional rater effect, specify it in a pre-unmasking amendment; do not invent it from free text. Domain changes and the half-SD threshold will be reported as observed-case descriptive summaries with their explicit denominators; no domain-specific clinical-response inference is planned. The threshold row will contain the observed count and percentage, not a pooled model effect. For exploratory baseline subgroup modelling, add one baseline characteristic-by-time-spline interaction at a time to the secondary model; do not fit a saturated combination or select interactions by statistical significance.

Reproducibility and exceptions. Publish a package-lock file, software versions, seed, code checksum, extraction schema, scoring-era manifest and synthetic-data examples. Tests must cover duplicate resolution, window boundaries, tie rules, score bounds, transfers, zero follow-up, administrative cutoffs, missing versions and a known synthetic mean. Deviations will be

timestamped with the data knowledge available. Table entries remain empty until approved analysis; no illustrative number is a study result.

Supplement S2. Blank reporting templates

These tables specify reporting fields only. An em dash denotes an unfilled field, not zero participants, zero change or an unavailable published result. Denominators must be shown for each horizon and estimate. Add disaggregated rows only where they follow the prespecified analysis and disclosure rules.

Table S1. Cohort construction and primary-outcome availability — blank template

Stage or reason	Children, n	Centres, n	Definition or denominator
Candidate unique children	—	—	Candidate calendar boundary
Excluded: baseline data invalid/unresolved	—	—	After deterministic validation
Excluded: calendar/version/age coverage	—	—	Prespecified manifest
Primary baseline cohort	—	—	All baseline-eligible children
Observed valid 150–210-day endpoint	—	—	Primary cohort
No assessment in the primary window	—	—	Primary cohort
Only incompatible/invalid window record	—	—	Primary cohort
Recorded death/undefined endpoint	—	—	Primary cohort; report separately
Primary-score analysis denominator	—	—	State any undefined-outcome qualification

Reasons within each exclusion stage must be mutually exclusive or explicitly labelled as overlapping. The final primary-outcome branches must reconcile to the baseline cohort.

Supplement S2B. Baseline characteristics

Table S2. Baseline description and follow-up selection — blank template

Characteristic	All baseline children	Observed six-month score	Missing six-month score	Standardized difference
Children, n	—	—	—	Not applicable

Characteristic	All baseline children	Observed six-month score	Missing six-month score	Standardized difference
Baseline age, mean (SD); median (IQR)	—	—	—	—
Baseline score, mean (SD); median (IQR)	—	—	—	—
Age bands, n/N (%)	—	—	—	—
Recorded sex, n/N (%)	—	—	—	—
Diagnostic categories, n/N (%)	—	—	—	—
Equity record status, n/N (%)	—	—	—	—
Prior care yes/no/unknown, n/N (%)	—	—	—	—
Centre/region/calendar period, n/N (%)	—	—	—	—

Show unrecorded categories. Do not place postbaseline therapy or home exposure in this baseline table. Record undefined outcomes separately.

Supplement S2C. Primary results and sensitivity analyses

Table S3. Six-month mean score change — blank template

Analysis	Eligible N / observed n	Mean change, points	95% CI	Key diagnostic
Primary multiple imputation under MAR	—	—	—	Imputed fraction; information fraction
Observed primary-window cases	—	—	—	Selected population
IPW, untruncated	—	—	—	Effective N; probability support
IPW with truncated weights	—	—	—	Weight range
Alternative PMM-draw imputation	—	—	—	Convergence/donor diagnostics
MNAR delta scenarios	—	—	—	Delta; boundary fraction
Window 120–240 days	—	—	—	Earlier baseline cutoff
Window 165–195 days	—	—	—	Narrower measurement window

Analysis	Eligible N / observed n	Mean change, points	95% CI	Key diagnostic
Same-rater observed subset	—	—	—	Changed population

Expand the MNAR row into the nine prespecified scenarios. Blank cells are not numerical findings. Never label mean change as a therapy effect. MI, multiple imputation; MAR, missing at random; IPW, inverse-probability weighting; MNAR, missing not at random; PMM, predictive mean matching.

Supplement S2D. Secondary horizons and domains

Table S4. Secondary score summaries — blank template

Horizon / domain	Eligible N	Observed n (%)	Observed mean change (SD)	Model estimate (95% CI)
Three-month total	—	—	—	—
Twelve-month total	—	—	—	—
Twenty-four-month total	—	—	—	—
Six-month comparable domain: name	—	—	—	—
Increase of at least 0.5 baseline SD, n/N (%)	—	—	—	Not estimated

Repeat the domain row only for documented comparable domains. For the threshold row, report the observed count and percentage with its denominator, not a mean or pooled model effect. The threshold is not an MCID.

Supplement S2E. Scoring-era and analysis audit

Table S5. Required deployment and analysis manifest — blank template

Manifest item	Value / identifier	Evidence or file	Decision date
Database close	Proposed 30 June 2026	—	—
Eligible baseline start/end	—	Deployment records	—
Centres and operational intervals	—	Centre adoption records	—
Scoring versions and age range	—	Specification/validation bridge	—
Excluded versions and reasons	—	Version audit	—
Ethics decision and permissions	Pending	—	—
Registry identifiers	Pending	—	—
Prior access/analysis declaration	—	Signed investigator/custodian account	—

Manifest item	Value / identifier	Evidence or file	Decision date
Code and SAP freeze	—	Checksum/package lock	—
Authorized outcome-data release	—	Access log	—
Amendments	—	Versioned change log	—

Pending means not documented as completed in this protocol. The template contains no ethics or registry approval claim.