STATISTICAL ANALYSIS PLAN

Developing a Childhood Asthma Risk Passive Digital Marker

Item	Description
Protocol Title	Developing a Childhood Asthma Risk Passive Digital Marker
Protocol Date	August 1, 2022
IRB Number	15873
ClinicalTrials.gov Identifier	NCT05826561
Funding Source	NIH K01HL166436
Study Design	Pilot randomized clinical trial using a Solomon four-group design
Planned Enrollment	100 practicing pediatricians; 25 participants per study arm
Intervention	EHR-integrated machine learning-based Passive Digital Marker (PDM) for childhood asthma risk classification
SAP Version	Version 1.0
SAP Date	August 1, 2022

Table of Contents

1. Administrative Information

This Statistical Analysis Plan (SAP) specifies the planned statistical methods for the study titled "Developing a Childhood Asthma Risk Passive Digital Marker." The document is intended to support NIH grant reporting, IRB documentation, and transparent conduct of the pilot randomized clinical trial. The SAP was finalized before database lock and before investigators conduct the final inferential analyses.

Role/Element	Description
Principal Investigator	Arthur H. Owora, MPH, PhD
Study Sponsor/Funder	NIH K01HL166436
Regulatory Identifier	IRB No. 15873
Protocol Date	August 1, 2022
Study Participants	Practicing pediatricians who provide outpatient pediatric care
Target Population for Clinical Tool	Preschool-aged children with longitudinal EHR data from birth through age 3 years
Planned Enrollment	100 pediatricians, randomized equally to four Solomon design arms, 25 per arm

2. Study Overview

Persistent childhood asthma is difficult to predict during the preschool years because early respiratory symptoms are heterogeneous and may overlap with transient viral wheezing. The study evaluates whether an EHR-integrated machine learning-based Passive Digital Marker (PDM), developed using routinely collected clinical data from birth through age 3 years, is acceptable, usable, feasible to study, and capable of improving pediatrician prognostic accuracy for persistent school-age asthma.

The study uses standardized pediatric patient vignettes derived from authentic longitudinal EHR data. Each participating pediatrician evaluates 10 vignettes. Five vignettes represent children with documented asthma diagnoses between ages 6 and 11 years, and five represent children without documented asthma during that age range. Clinicians assigned to PDM-supported assessment receive the binary PDM risk classification, defined as high or low risk for persistent school-age asthma, when completing prognostic judgments.

3. Study Objectives and Endpoints

3.1 Primary Objectives

The primary objectives are to assess perceived PDM acceptability, perceived PDM usability, and study feasibility among practicing pediatricians. These objectives are aligned with the pilot and implementation-focused purpose of the study, in which the main goal is to determine whether the PDM can be evaluated and accepted in a clinician-facing research workflow.

Primary Endpoint	Operational Definition	Scale/Metric	Primary Summary
Perceived PDM acceptability	Clinician-reported acceptability/behavioral intention to use the PDM if integrated into the EHR	Likert-type behavioral acceptability/behavioral intention scale, 0-5, with higher scores indicating greater acceptability	Mean, SD, median, IQR, 95% CI; proportion with score >=4
Perceived PDM usability	Clinician-reported usability of the PDM decision support display	Modified Healthcare Systems Usability Scale (HSUS), 0-5, with higher scores indicating greater usability	Mean, SD, median, IQR, 95% CI; proportion with score >=4
Study feasibility	Ability to recruit, randomize, and retain clinicians and complete study procedures	Recruitment within planned accrual period; completion of vignette assessments; completion of questionnaires; usable analytic data	Counts, percentages, 95% CIs, and prespecified feasibility thresholds

3.2 Secondary Objective

The secondary objective is to estimate the effect of PDM access on clinician prognostic accuracy for school-age asthma in standardized vignette assessments.

Secondary Endpoint	Operational Definition	Scale/Metric	Primary Summary
Prognostic accuracy	Proportion of correctly classified vignettes for each clinician and across vignette-level observations	Correct classification divided by total vignettes assessed; binary correct/incorrect at vignette level	Mean accuracy by arm; PDM vs standard assessment difference; mixed-effects model accounting for clustering within clinician
Case vignette accuracy	Correct classification among vignettes representing children who later developed asthma	Correct/incorrect for five case vignettes	Mean and 95% CI by condition; Welch-adjusted tests or mixed-effects models as appropriate
Control vignette accuracy	Correct classification among vignettes representing children who did not develop asthma	Correct/incorrect for five control vignettes	Mean and 95% CI by condition; Welch-adjusted tests or mixed-effects models as appropriate

4. Study Design

This is a pilot randomized clinical trial using a Solomon four-group design with pretest and posttest assessments. The Solomon design allows estimation of intervention effects while evaluating whether pretest exposure influences subsequent assessment performance. The four planned study arms are:

1. Standard assessment without PDM support, pretest-posttest design (SAU pre-post), n = 25.
2. Standard assessment without PDM support, posttest-only design (SAU posttest only), n = 25.
3. PDM-supported assessment, pretest-posttest design (PDM pre-post), n = 25.
4. PDM-supported assessment, posttest-only design (PDM posttest only), n = 25.

Pre-post PDM intervention: Clinicians randomized to this arm complete an initial set of asthma prognostic vignette assessments without PDM support, then receive access to the Passive Digital Marker risk classification, and subsequently complete posttest vignette assessments with PDM support. This arm permits evaluation of within-clinician change in prognostic accuracy before versus after exposure to the PDM.

Pre-post SAU: Clinicians randomized to this arm complete an initial set of asthma prognostic vignette assessments without PDM support and then complete posttest vignette assessments, again without PDM support. This arm serves as the standard-assessment comparison condition and permits evaluation of changes in prognostic accuracy over time that may occur because of repeated testing, familiarity with the vignette format, or other non-PDM effects.

Participants evaluate standardized vignettes derived from longitudinal EHR data from birth through age 3 years. The outcome status of each vignette is defined using documented asthma diagnosis between ages 6

and 11 years. Vignette presentation order will be randomized or counterbalanced where feasible to minimize ordering effects. Clinicians are blinded to true vignette case/control status.

5. Analysis Populations

Population	Definition	Primary Use
Randomized population	All clinicians randomized to one of the four study arms	CONSORT-style flow, recruitment, allocation, and retention reporting
Intent-to-treat (ITT) population	All randomized clinicians analyzed according to assigned arm, regardless of degree of exposure or completion, when sufficient endpoint data are available	Primary analysis of feasibility and all randomized-group summaries
Modified ITT/evaluable population	Randomized clinicians who complete at least one endpoint-relevant questionnaire or at least one vignette assessment	Primary analytic population for outcome analyses if no endpoint data are available for a participant
Per-protocol population	Clinicians completing all assigned vignettes and relevant questionnaires without major protocol deviations	Sensitivity analyses
Vignette-level analytic dataset	Each clinician-vignette assessment record, including assigned condition, timing, vignette case status, PDM output when applicable, clinician prediction, and correct/incorrect classification	Mixed-effects analysis of prognostic accuracy

6. Sample Size and Power Considerations

The planned enrollment is 100 pediatricians, with 25 participants randomized to each of the four Solomon design arms. As a pilot study, the trial is not intended to provide definitive hypothesis testing for clinical efficacy. Instead, the sample size is designed to estimate acceptability, usability, feasibility metrics, effect sizes, and variance components to inform a future fully powered pragmatic clinical trial.

Each clinician will evaluate 10 vignettes, resulting in up to 1,000 clinician-vignette assessments if all 100 clinicians complete all vignettes. The design therefore provides precision for clinician-level implementation measures and supports estimation of clustering parameters, including clinician-level intraclass correlation, in vignette-level models. Feasibility estimates will be presented with exact or Wilson 95% confidence intervals, as appropriate.

No formal stopping rule for efficacy is planned because the study is a pilot implementation and decision-support evaluation study. The SAP will avoid overinterpretation of p values and will emphasize estimation, confidence intervals, direction and magnitude of effects, and feasibility criteria.

7. Randomization and Masking

Clinicians will be randomized using a computer-generated allocation sequence to one of four study arms in the Solomon design. Allocation will target equal enrollment of 25 clinicians per arm. Because the intervention consists of providing PDM output to clinicians, participants cannot be blinded to whether they receive decision support. However, clinicians will remain blinded to true vignette outcome status and will not be informed whether an individual vignette represents a child who later developed asthma.

8. Statistical Principles

- **Analysis orientation:** Primary analyses will emphasize estimation and precision because this is a pilot study.
- **Type I error:** Two-sided tests will be used with alpha = 0.05 when inferential tests are reported. P values will be interpreted cautiously and will not be the sole basis for conclusions.
- **Confidence intervals:** 95% confidence intervals will accompany key estimates whenever feasible.
- **Unit of analysis:** Acceptability, usability, and feasibility are clinician-level outcomes. Prognostic accuracy is both a clinician-level proportion and a vignette-level binary outcome.
- **Clustering:** Vignette-level accuracy analyses will account for repeated vignette assessments nested within clinicians using mixed-effects models.

- **Multiplicity:** No formal multiplicity adjustment is planned for this pilot study. Findings across multiple endpoints will be interpreted as exploratory and hypothesis-generating.
- **Covariate adjustment:** Primary analyses will be parsimonious. Exploratory adjusted analyses may include clinician years of practice, specialty, sex/gender where available, and pretest exposure indicators if they are clinically or methodologically relevant.

9. Data Sources and Variable Definitions

9.1 Data Sources

- Clinician enrollment, consent, randomization, and study completion records.
- Clinician demographic and professional background questionnaires.
- Vignette assessment responses, including clinician risk judgment and final classification of whether the child will develop school-age asthma.
- PDM risk classifications shown to clinicians assigned to PDM-supported assessment.
- PDM acceptability/behavioral intention questionnaire responses.
- Modified HSUS usability questionnaire responses.
- Administrative feasibility tracking data, including recruitment timing and completion of assigned procedures.

9.2 Key Derived Variables

Variable	Definition/Coding
Study arm	Four-level categorical variable: SAU pre-post, SAU posttest only, PDM pre-post, PDM posttest only
PDM access	Binary indicator: 1 = PDM-supported assessment; 0 = standard assessment without PDM
Assessment timing	Pretest or posttest assessment, as applicable
Pretest exposure	Binary indicator for assignment to a pretest-posttest arm
Vignette case status	Binary indicator: 1 = child had documented asthma diagnosis between ages 6 and 11 years; 0 = no documented asthma diagnosis
Clinician prediction	Clinician classification or risk judgment regarding school-age asthma development
Correct classification	Binary indicator: 1 = clinician prediction matches true vignette case status; 0 = prediction does not match true status
Clinician-level prognostic accuracy	Number of correctly classified vignettes divided by total completed vignettes for that clinician in a given timing/condition
Acceptability score	Mean or total score from the behavioral acceptability/behavioral intention scale, transformed or reported on 0-5 scale as specified in the final codebook
Usability score	Mean or total modified HSUS score, transformed or reported on 0-5 scale as specified in the final codebook
Feasibility completion	Binary and proportion metrics for recruitment, retention, vignette completion, and questionnaire completion

10. Primary Outcome Analyses

10.1 Perceived PDM Acceptability

Acceptability will be summarized among clinicians exposed to the PDM and, where relevant, among all clinicians who complete the acceptability instrument. The primary summary will include mean, standard deviation, median, interquartile range, minimum, maximum, and 95% confidence interval. The proportion of clinicians with acceptability scores >=4 on the 0-5 scale will also be reported with a 95% confidence interval.

If item-level data are available, internal consistency will be summarized using Cronbach alpha. Item-level distributions will be tabulated to identify ceiling effects, floor effects, or items indicating potential barriers to PDM adoption. No formal hypothesis test is required for the primary acceptability endpoint; the analysis will

focus on whether scores indicate favorable perceived acceptability and whether the observed precision is adequate for planning future work.

10.2 Perceived PDM Usability

Usability will be summarized using the modified HSUS score among clinicians exposed to the PDM and, where applicable, among all clinicians who complete usability items. The primary summary will include mean, standard deviation, median, interquartile range, minimum, maximum, 95% confidence interval, and the proportion with usability score >=4 on the 0-5 scale.

Item-level usability results will be summarized descriptively. Qualitative or open-text usability concerns, if collected, will be coded narratively and summarized as implementation considerations. Internal consistency will be estimated if the number and structure of usability items support reliability assessment.

10.3 Study Feasibility

Feasibility will be evaluated using prespecified operational metrics. The feasibility analysis will report counts, percentages, and 95% confidence intervals for each metric. Feasibility conclusions will be based on whether the study can recruit and randomize the target sample, retain clinicians, and obtain complete vignette and questionnaire data.

Feasibility Metric	Numerator	Denominator	Planned Interpretation
Recruitment feasibility	Number of clinicians enrolled and randomized	Target enrollment of 100 clinicians	Feasible if recruitment reaches the planned sample within the planned accrual period or yields sufficient enrollment to support trial planning
Allocation feasibility	Number randomized into each arm	25 planned per arm	Feasible if allocation is approximately balanced and deviations are explainable
Vignette completion	Number completing all assigned vignette assessments	Number randomized or evaluable	Feasible if completion is high and missingness is limited
Questionnaire completion	Number completing acceptability/usability measures when assigned	Number eligible for questionnaire completion	Feasible if questionnaire completion is high and item missingness is limited
Data usability	Number with analyzable endpoint data	Number randomized	Feasible if endpoint data support planned descriptive and model-based analyses

11. Secondary Outcome Analyses

11.1 Clinician-Level Prognostic Accuracy

Clinician-level prognostic accuracy will be calculated as the number of correctly classified vignettes divided by the total number of vignettes completed by the clinician within each applicable assessment period. Accuracy will be summarized by study arm, PDM access status, and assessment timing using mean, standard deviation, median, interquartile range, range, and 95% confidence intervals.

For simple group comparisons, mean accuracy will be compared between PDM-supported and standard assessment conditions using t tests or Welch-adjusted t tests if variance or group size assumptions are unequal. Nonparametric summaries may be provided as sensitivity analyses if distributions are skewed or bounded at 0 or 1.

11.2 Vignette-Level Mixed-Effects Model

The primary model-based secondary analysis will use the vignette-level correct classification indicator as the outcome. Because each clinician contributes multiple vignette assessments, a mixed-effects model with a random intercept for clinician will be used. The planned model will include fixed effects for PDM access, assessment timing, and the PDM access by assessment timing interaction.

A linear mixed-effects probability model may be used to align with the pilot manuscript analysis and support direct interpretation of coefficients as absolute differences in prognostic accuracy. A logistic mixed-effects model will be conducted as a sensitivity analysis if model convergence and sample size permit.

Planned linear mixed-effects model specification:

The primary model for vignette-level prognostic accuracy will be specified as follows:

$$Correct_{ij} = \beta_0 + \beta_1(PDM_i) + \beta_2(Posttest_{ij}) + \beta_3(PDM_i \times Posttest_{ij}) + u_i + \varepsilon_{ij}$$

where:

$Correct_{ij}$ *is the binary correct/incorrect indicator for vignette j assessed by clinician i.*

PDM_i *indicates whether clinician i was assigned to the Passive Digital Marker condition.*

$Posttest_{ij}$ *indicates whether the vignette assessment occurred during the posttest period.*

$PDM_i \times Posttest_{ij}$ *is the interaction term estimating whether the change in prognostic accuracy from pretest to posttest differs by PDM assignment.*

u_i *is a clinician-specific random intercept that accounts for clustering of vignette assessments within clinicians.*

ε_{ij} *is the residual error term.*

The primary coefficient of interest for the secondary analysis is beta3, the interaction between PDM access and posttest timing. Model estimates will be presented with 95% confidence intervals and p values. The intraclass correlation coefficient will be reported to quantify clustering of vignette assessments within clinician.

11.3 Case and Control Vignette Accuracy

Accuracy will be separately summarized for case vignettes and control vignettes. Case vignette accuracy evaluates identification of children who later developed asthma, whereas control vignette accuracy evaluates identification of children who did not develop asthma. Comparisons between PDM-supported and standard assessment conditions will use Welch-adjusted t tests for clinician-level proportions or mixed-effects models with vignette case status and interaction terms as exploratory analyses.

12. Additional and Exploratory Analyses

- **Pretest sensitization:** The Solomon design will be used to evaluate whether pretest exposure affects posttest prognostic accuracy, acceptability, or usability. Exploratory models may include pretest exposure and pretest exposure by PDM access terms.
- **Clinician characteristics:** Exploratory analyses may examine whether years of practice, specialty, or other clinician characteristics are associated with acceptability, usability, or prognostic accuracy.
- **Vignette difficulty:** If data structure permits, models may include random or fixed effects for vignette to account for differences in vignette difficulty.
- **Agreement with PDM:** Among PDM-exposed clinicians, analyses may describe how often clinician classifications agree with the PDM classification and whether agreement differs by case status.
- **Decision change:** In pretest-posttest arms, analyses may describe the proportion of clinicians whose classification changed between pretest and posttest and whether changes moved toward the true outcome.

13. Missing Data and Protocol Deviations

Missingness will be summarized by endpoint, item, study arm, and timing. The primary approach will be complete-case analysis for each endpoint because the study is a pilot trial and the primary outcomes are direct survey and feasibility measures. The denominator used for each analysis will be clearly reported.

For scale-based endpoints, if a participant is missing a small number of items within a scale, scoring will follow the instrument-specific rule if available. If no prespecified rule exists, a scale score may be calculated when at least 80% of items are completed, using the participant-specific mean of completed items. Sensitivity analyses may compare results using stricter complete-item scoring.

Major protocol deviations will be described before database lock and may include ineligible participant enrollment, failure to receive assigned intervention condition, incomplete vignette exposure, or unblinding to vignette outcome status. Primary analyses will use the ITT or modified ITT approach, with per-protocol analyses used as sensitivity analyses.

14. Safety, Risk, and Monitoring Considerations

The study involves practicing clinicians completing standardized vignettes and questionnaires. It does not assign clinical care to pediatric patients and does not use the PDM to direct real-time patient management during the pilot vignette-based assessment. Therefore, no clinical adverse event analysis is planned. Any unanticipated problems involving risks to participants or others, confidentiality breaches, or significant protocol deviations will be documented and reported according to IRB requirements.

15. Reporting Standards and Tables/Figures

Reporting will align with pilot trial reporting principles and AI clinical intervention guidance, including CONSORT-AI and SPIRIT-AI concepts where applicable. The following tables and figures are planned:

- Table 1. Participant flow, recruitment, randomization, and completion by study arm.
- Table 2. Baseline clinician characteristics by study arm.
- Table 3. Primary outcomes: acceptability, usability, and feasibility summaries.
- Table 4. Secondary outcomes: clinician prognostic accuracy overall and by case/control vignette status.
- Table 5. Mixed-effects model for vignette-level prognostic accuracy.
- Figure 1. CONSORT-style flow diagram of clinician enrollment, allocation, and completion.
- Figure 2. Boxplot or dot plot of prognostic accuracy by study arm and timing.
- Figure 3. Distribution of acceptability and usability scores among PDM-exposed clinicians.

16. Quality Control, Reproducibility, and Software

Analyses will be conducted using validated statistical software, anticipated to be R. Analysis scripts will be version-controlled and archived with the analytic dataset, codebook, SAP, and output tables. Data cleaning will include range checks, verification of randomization assignments, validation of vignette scoring, duplicate record checks, and reconciliation of participant completion status. Any changes to derived variable definitions after SAP finalization will be documented in an amendment or analysis memo.

17. SAP Amendments

Any substantive changes to the planned primary or secondary analyses after SAP finalization will be documented with the date, rationale, affected sections, and whether the change occurred before or after

database lock or unblinding of final outcome analyses. Administrative corrections, formatting edits, and typographical corrections do not require a formal amendment unless they alter analytic meaning.

18. References

1. Centers for Disease Control and Prevention. Most recent asthma data. Updated 2022. Accessed June 22, 2025.
2. Cave AJ, Atkinson LL. Asthma in preschool children: a review of the diagnostic challenges. J Am Board Fam Med. 2014;27(4):538-548. doi:10.3122/jabfm.2014.04.130276
3. Kothalawala DM, Kadalayil L, Weiss VBN, et al. Prediction models for childhood asthma: a systematic review. Pediatr Allergy Immunol. 2020;31(6):616-627. doi:10.1111/pai.13247
4. Owora AH, Tepper RS, Ramsey CD, Becker AB. Decision tree-based rules outperform risk scores for childhood asthma prognosis. Pediatr Allergy Immunol. 2021;32(7):1464-1473. doi:10.1111/pai.13530
5. Owora AH, Jiang B, Shah Y, Gaston B, Boustani M. External validation and update of the pediatric asthma risk score as a passive digital marker for childhood asthma using integrated electronic health records. EClinicalMedicine. 2025;84:103254. doi:10.1016/j.eclinm.2025.103254
6. Ghorayeb A, Darbyshire JL, Wronikowska MW, Watkinson PJ. Design and validation of a new Healthcare Systems Usability Scale for clinical decision support systems: a mixed-methods approach. BMJ Open. 2023;13(1):e065323. doi:10.1136/bmjopen-2022-065323
7. Kaczmarek LD, Kashdan TB, Drazkowski D, et al. Why do people prefer gratitude journaling over gratitude letters? The influence of individual differences in motivation and personality on web-based interventions. Pers Individ Dif. 2015;75:1-6.
8. Liu X, Rivera SC, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020;26(9):1364-1374. doi:10.1038/s41591-020-1034-x
9. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. BMJ. 2020;370:m3210. doi:10.1136/bmj.m3210