# Protocol: AI Scribe Patient Experience Study

**Official Title:** Evaluation of Patient Experience in Ambient AI Scribe Encounters: A Secondary Analysis of a Randomized Trial

**Funding:** This study was supported by the UCLA Department of Medicine, NIH/NIA R01AG070017-01, K76AG064392-01A1, and K24AG047899, and the NIH/NCATS UCLA CTSI UL1TR001881.

**Conflicts of Interest**: None declared.

**Trial Registration:** Parent trial (ClinicalTrials.gov: NCT06792890), ICTRN (pending)

**Protocol Version Date**: 12/10/2025

**Keywords**

Conditions
- Physician Workflow
- Patient Satisfaction

Keywords
- Artificial Intelligence Scribe
- Randomized Controlled Trial
- Patient Experience
- CG-CAHPS Survey
- Clinical Documentation

## Study Description

**Brief Summary:**

This study's primary aim is to assess the impact of AI scribe tools on patient-reported communication outcomes. Secondary objectives include subgroup analyses to explore heterogeneity in patient experience outcomes based on demographic and clinical encounter characteristics.

**Detailed Description:**

Ambient AI scribe technologies are designed to improve physician workflow by automating documentation tasks during clinical encounters to reduce cognitive burden, and reflect the fastest growing application of generative AI technologies in health care.[1] Our previous study of a two-month randomized controlled trial focused on the effect of ambient scribes on

physician efficiency and burnout.[2] However, there remains a critical evidence gap on the impacts on these technologies on patient-oriented outcomes.

The patient experience component will focus on evaluating the communication composite scores from Clinician and Group Consumer Assessment of Healthcare Providers and Systems (CG-CAHPS) surveys linked to outpatient encounters during the study period. For decades, CG-CAHPS psychometric surveys have been used by the Centers for Medicare and Medicaid Services to assess the patient experience, and they reflect the gold standard tool to measure patient satisfaction.[3] These scores will be adjusted for baseline communication performance, patient demographics, and provider characteristics. Secondary outcomes will include subgroup analyses to identify variations in patient-reported experiences based on demographic and clinical encounter characteristics.

To our knowledge, this protocol reflects the first study to document the real-world effect of generative AI scribes on patient-reported outcome measures.

## Study Objectives

Primary Objective: To evaluate whether ambient AI scribe use is associated with differences in patient-reported communication scores on the CG-CAHPS survey.

Secondary (Exploratory) Objectives:

- To explore heterogeneity of effects across demographic, encounter, and provider subgroups.
- To examine top-box communication outcomes, overall provider rating, and individual communication items.
- To compare three individual groups (Control, Nabla, MicrosoftDAX Copilot) in a descriptive subanalysis.

## Study Design

- Study Type: Retrospective secondary analysis of a completed randomized clinical trial
- Randomization: Provider-level covariate-constrained randomization from the parent trial based on baseline time in notes, burnout, and clinic days per week as per previous study.[2]
- Analysis: Primary analysis will follow provider-level intent-to-treat assignment. All outcomes and analytic approaches were prespecified prior to accessing patient-experience data to minimize selective reporting.
- Study Setting: The study was conducted in ambulatory clinics within the UCLA Health system in the United States.
- Enrollment [Actual]

    - Providers: 238
    - Encounters eligible for analysis: 36235
    - Surveys: 749
- Number of Groups/Cohorts: three cohorts combined into two groups (see below)
- Intervention Time: 2 months (11/4/2024–1/3/2025)
- Target Follow-up Duration: Surveys collected up to 6 months after each eligible encounter.

## Eligibility

**Study Population:** Eligible participants for this study included all adult patients with an encounter with a study provider during the study period. Study providers were select ambulatory care physicians within the UCLA Health system who held at least one half-day of clinic per week. Only English-language encounters are included due to a lack of internal validation for translation capabilities in the vendor technologies. Physicians who typically use a human scribe agreed to forego this assistance while assigned to an intervention or control group. Due to the technical requirements of one vendor, all providers were required to use an iOS portable device. This study was reviewed and approved by the UCLA Institutional Review Board (IRB-24-5425).

**Inclusion Criteria:**

- Patient encounters with ambulatory care physicians within the UCLA Health system who held at least one half-day of clinic per week

**Exclusion Criteria:**

- Encounters with trainee providers (e.g., residents, medical students) and allied healthcare professionals (e.g., RNs, PAs)
- Encounters with attendings who work exclusively with trainees
- Encounters with a human scribe
- Non-English encounters

## Arms and Interventions

| Arm | Description |
|---|---|
| Group A – AI Scribe (Nabla) | Participants used an ambient AI scribe (Nabla) during clinical encounters. |
| Group B – AI Scribe (Microsoft DAX Copilot) | Participants used an ambient AI scribe (Microsoft DAX Copilot) during clinical encounters. |

| | |
|---|---|
| Control Arm – No Scribe | Participants did not use an AI scribe and continued usual clinical documentation practices. |

Randomization and Grouping:
Participants were initially randomized in a 1:1:1 ratio to Group A, Group B, or Control. For primary analyses, Groups A and B are combined into a single “AI Scribe” treatment arm and compared against the Control Arm, as no differential effect by vendor is hypothesized.

Footnote (applies to AI Scribe groups):
AI scribe tools capture physician–patient conversations, generate a transcript, and summarize the transcript into a clinical note automatically added to the EHR. Tools are transcriptional only and do not provide clinical decision support. Physicians inform patients about recording and obtain verbal consent; declinations are tracked.

## Outcome Measures

**Primary Outcome Measure:**

| Outcome Measure | Measure Description | Time Frame |
|---|---|---|
| Mean CG-CAHPS Communication Composite Score (0–100) | CG-CAHPS surveys linked to enrolled physicians and encounters during the study period will be used to calculate the communication composite outcome. A mean CG-CAHPS Communication Composite Score (0–100 scale) derived from the four validated communication items assessing whether the doctor: (1) explained things in an easy-to-understand way, (2) listened carefully, (3) showed respect for what the patient had to say, and (4) spent enough time with the patient.[4] Surveys are eligible if completed within 6 months of an encounter that occurred during the study period. The primary analysis will use adjusted linear regression with robust standard errors clustered at the provider level, adjusting for:<br>- Previous CG-CAHPS scores (provider-level)<br>- Patient reported physician's role at time of encounter (new vs established)<br>- Patient age<br>- Patient sex<br>- Patient race/ethnicity<br>- Patient education level<br>- Patient self-reported overall health score. | Baseline period: 6 months prior to enrollment: 5/4/2024 – 11/3/2024<br>Intervention period: 11/4/2024 – 1/3/2025<br>Follow-up: Surveys completed within 6 months after the encounter |

**Secondary/Exploratory Outcome Measures:**

| Outcome Measure | Measure Description |
|---|---|
| Subgroup analyses of communication composite scores | Subgroup analyses will examine differences in the primary outcome in the following groups: 1) baseline poor communication performers (bottom third), 2) new appointment encounters, 3) encounters with a patient's usual provider (determined by patient survey response), 4) female vs male physicians, 5) generalist vs specialist physicians, 6) high adopting physicians as defined as top 50$^{th}$ percentile of usage of the tool. |
| Mean CG-CAHPS communication composite score, top box | Proportion of CG-CAHPS surveys achieving maximal Communication Composite Scores. Analyses use adjusted regression with provider-clustered robust SEs. |
| Overall provider rating (0–10), continuous | Continuous overall provider rating (0–10) from eligible CG-CAHPS surveys. Modeled using adjusted regression with provider-clustered SEs. |
| Overall provider rating (0–10), top box | Proportion of surveys with top-box overall provider rating, using standard CG-CAHPS definitions. Analyzed using adjusted regression. |
| Single-item communication scores, continuous | Continuous scores for each individual communication domain item. Derived from eligible CG-CAHPS surveys and analyzed using the adjusted regression framework. |
| Three arm comparison of communication composite scores | This will be a subanalysis of the primary outcome but with three arms (control, Nabla, and Microsoft Dax Copilot). |

## Statistical Analysis

The primary outcome will use adjusted linear regression with robust standard errors clustered at the physician level for the full cohort and all subgroup analyses. Covariate selection was based on established predictors of CG-CAHPS communication performance. Secondary outcomes will use binary models for top-box indicators given ceiling effects. A two-sided $p<0.05$ will define significance in the primary outcome, secondary/exploratory outcomes are not confirmatory. Communication domains with partially missing items will be included so long as ≥50% items are answered, as per AHRQ CAHPS Instructions for Analyzing Data from CAHPS Surveys protocol.[5]

## Data Management, Confidentiality, and Sharing

All analyses will use de-identified CG-CAHPS survey data and provider-level metadata. All data remain on secure UCLA Health servers in compliance with UCLA Health policies and HIPAA requirements. CG-CAHPS survey data cannot be publicly shared due to institutional restrictions. Aggregated results may be shared upon reasonable request.

## Harms

This study involves secondary analysis of routine patient-experience surveys and presents no risk to physicians or patients. No adverse events are anticipated or monitored.

## References


1. Tierney, Aaron A., et al. "Ambient artificial intelligence scribes to alleviate the burden of clinical documentation." NEJM Catalyst Innovations in Care Delivery 5.3 (2024): CAT-23.
2. Lukac, Paul J., et al. "Ambient AI Scribes in Clinical Practice: A Randomized Trial." NEJM AI 2.12 (2025): AIoa2501000.
3. Dyer, Naomi, et al. "Psychometric properties of the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) clinician and group adult visit survey." Medical care 50 (2012): S28-S34.
4. Quigley, Denise D., et al. "Associations of the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Clinician and Group Survey Scores with Interventions and Site, Provider, and Patient Factors: A Systematic Review of the Evidence." Journal of Patient Experience 11 (2024): 23743735241283204.
5. Agency for Healthcare Research and Quality. Instructions for Analyzing Data from CAHPS® Surveys: Using the CAHPS Analysis Program. AHRQ, 2015, https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/helpful-resources/analysis/2020-instructions-for-analyzing-data.pdf