# Longitudinal effects of ambient AI scribe use on documentation burden and financial productivity: A quasi-experimental study

[1]RJ Waken, PhD, [2]Sunny S. Lou, MD, PhD, [1]Mackenzie Hofford, MD, [2]Elise Eiden, MS, [1]Carrie Burk, MS,[1]Seunghwan Kim, PhD, [1]Jared Esker, MS, Linying Zhang, PhD, [3]Thomas M. Maddox, MD, SM, [2]Joanna Abraham, PhD, [1]Albert M. Lai, PhD, [4]Sam Bhayani, MD, [5]Deborah O'Dell, [3]Jennifer Schmidt, MD [5]Kayla Paynter, [5]Michele Thomas, MD, [5]Matthew Gerling, MD, [1,3]Philip R.O. Payne, PhD [1,2]Thomas Kannampallil, PhD

[1]Institute for Informatics, Data Science and Biostatistics (I2DB), WashU Medicine
[2]Department of Anesthesiology, WashU Medicine
[3]Department of Medicine, WashU Medicine
[4]Department of Surgery, WashU Medicine
[5]BJC Health System

*Corresponding Author*:
Thomas Kannampallil, PhD
Washington University School of Medicine (WashU Medicine)
660 South Euclid Ave, Campus Box 8054
St Louis, Missouri 63110
thomas.k@wustl.edu
Phone: 314-273-7801

Number of Tables: 2
Number of Figures: 2
Word count: 2995

**Key Points**

**Question:** How do the patterns of use of an ambient Artificial Intelligence (AI) scribe evolve over time?

**Findings**: In this longitudinal, quasi-experimental study on clinician use of an ambient AI scribe, documentation time, note writing patterns and financial productivity evolved over a 150-day period. Compared to the pre-period, note writing time savings increased from 7% (day zero) to 15% (day 150); changes in all other considered outcomes including time outside of scheduled hours, pajama time, note length, note closure <24h, billed work Relative Value Units evolved over the 150-day period.

**Meaning**: Clinician use of ambient AI scribes showed persistent changes in patterns of use over a 150-day period, highlighting a gradual adaptation process and the need for longitudinal assessment.

**Abstract**

**Importance:** Clinician adoption and adaptation of new tools evolve over time. Prior studies of ambient Artificial intelligence (AI) scribes have primarily relied on single time-point measurements (e.g., pre-post), potentially obfuscating their true impact on outcomes.

**Objective**: To investigate longitudinal effects of an AI scribe tool on patient encounter-level outcomes.

**Design**: Case series across 48 weeks (24 pre, 24 post) per clinician.

**Setting**: Primary care clinical encounters occurring between 01/05/24 and 10/31/25.

**Participants**: Primary care clinicians (attending physicians and advanced practice providers).

**Exposure**: Ambient AI scribe introduction to clinical workflow, indexed to study day zero, per clinician.

**Main outcomes and measures**: Encounter-level measurements of documentation time (note writing time, time outside of scheduled hours (TOSH), pajama time), note writing patterns (note length, note closure <24h) and clinician's billed work Relative Value Units (wRVU) with a focus on changes from pre-period outcomes at Day 0 and 150.

**Results**: 220 primary care clinicians (Mean age=43.7, 70.9% females; 56.4% physicians) from 36 clinics, conducting 314,845 patient encounters were included. All outcomes evolved from day zero to day 150 and are compared back to pre-period levels. There was evidence of an immediate 7% decrease on average in note writing time at day zero (Incidence Rate Ratio, IRR 0.93, 95%CI [0.89, 0.96]), intensifying to a 15% decrease by day 150 (IRR 0.85, 95%CI [0.83, 0.87]). There was no evidence of a change in pajama time or TOSH at day zero; however, at day 150, there was evidence of a 18% decrease in pajama time (0.82, 95%CI [0.73, 0.91]) and a 13% decrease in TOSH (0.87, 95%CI [0.77, 0.99]). At day zero, there was evidence of a 5% increase (1.05, 95%CI [1.00, 1.10]) in note length and 31% increase in note closures (1.31, 95%CI [1.13, 1.53]), with both slowly attenuating to pre-period levels by day 150. Although there was no evidence of changes in wRVU at day zero, there was a 2% increase total wRVU at day 150 (1.02, 95%CI [1.01, 1.03]).

**Conclusions and relevance**: Longitudinal changes were gradual, but persistent, underscoring the gradual adaptation of AI scribes, as clinicians situated these tools within their workflows.

## Introduction

The burden of electronic health record (EHR) use has increased exponentially over the past decade.[1,2] A primary contributor to this burden is the increased documentation requirements, owing to policy and payment requirements, increasing clinician workload, burnout, and work-life balance.[3-5] Research has highlighted that every hour of patient-facing time was associated with ~2 hours of additional documentation, potentially detracting clinicians from direct patient interactions, affecting both patient care, and clinician satisfaction.[6]

Artificial Intelligence (AI)-based scribes (henceforth, "AI scribes"), have the potential to reduce clinician time spent on documentation, by generating the draft of a clinical note from the patient-clinician conversation.[7] These tools have transformed clinical note generation and organization into a language-optimized, model-driven report, with the clinician playing the role of an editor.[8]

Although AI scribes have been rapidly adopted and several studies have been conducted, confirmatory evidence regarding their benefits are limited.[7,9] Among the reported studies, there are contrasting findings related to savings in documentation time associated with the use of AI scribes. For example, while some studies have reported savings in note writing time after the implementation of an AI scribe (e.g.,[9-12]), others have reported no or marginal time savings (e.g.,[13,14]). Additionally, almost all studies using self-reported survey-based measures have reported on increased clinician satisfaction, reduced burnout, reduced task load, and perceived savings in note writing time.[11,15-19]

These contrasting findings can be attributed to several methodological and measurement factors adopted in previous studies. First, measurement mechanisms for time-based reporting has relied on vendor-provided metrics, aggregated at a clinician-level, often at bi-weekly or at monthly intervals (e.g., using Epic's Signal platform[9,13]). Such an aggregation fails to capture patient encounter-level variability (e.g., documentation time associated with a specific encounter type) and precludes the ability to control for important encounter-level covariates such as visit type (e.g., face-to-face, telemedicine) or physician productivity (e.g., work relative value units, wRVU). This is the case even for recent RCTs[13,20] that have found limited productivity gains, as highlighted in an accompanying editorial that claimed that AI scribes are not "productivity tools yet."[21] In their editorial, Kim et al also further emphasized that the marginal effect sizes on the productivity gains may be potentially related to the measurement challenges associated with the use of vendor-developed platforms.[21]

Second, studies have almost exclusively relied on pre-post study designs, with measurements at two time-points (e.g., 1-month before/after implementation; see exceptions[13,20]). However, as with any new technology, AI scribe adoption and adaptation evolves over time; as such, single time-point measurements are often unable to capture user adaptations that evolve over time.[22-24]

The primary objective of this study is to conduct a longitudinal, quasi-experimental study using patient-encounter-level measures to ascertain the impact of an AI scribe on outcomes related to documentation time (note writing time, time outside of scheduled hours, pajama time), note writing patterns (note length, note closure) and physician financial productivity (wRVUs) at approximately monthly time points over a 24-week period (e.g., 0-, 30-, 120- and 150-days).

## Methods

### *Study setting, design, and participants*

A longitudinal quasi-experimental study was carried out across outpatient clinics at a large midwestern healthcare system. Primary care clinicians (attending physicians and/or advanced practice providers (APPs)) were provisioned access to an ambient AI scribe (Abridge Inc, New York) integrated within the Epic EHR (Verona, WI) at different time points between June 19, 2024, and July 30, 2025.

Primary care provider (PCP) encounters of clinicians provisioned AI scribe access, occurring between January 5, 2024, and October 31, 2025, were included. Clinicians included attending physicians (either MD or DO) or advanced practice providers (APP; NPs or PAs). Patient encounters were excluded if they contained indicators for non-primary care clinicians, non-PCP encounters, telemedicine encounters, and weekend encounters (see CONSORT Figure 1 for exclusions).

We used an interrupted time series design to investigate the changes in provider-patient encounter level outcomes after AI scribe tool introduction over a 48-week period (24 weeks pre-implementation, 24 weeks post-implementation) centered on first date of a clinician's AI scribe access. In this design all clinicians were considered treated as of the day they were first provisioned access ("day zero"), under an intent-to-treat (ITT) framework, while accounting for typical day-to-day variability in clinician work patterns.

The study protocol was approved by the Institutional Review Board of the Washington University School of Medicine (IRB#202511025).

### *Data*

Data were retrieved from Epic EHR's User Action Log Lite tables in the Clarity database. These tables record pre-specified clinician activities (e.g., note writing, review, orders), time spent on these activities, and along with the associated patient encounters details. For each included patient encounter, we retrieved encounter ID, encounter type (face-to-face, telemedicine, phone), encounter clinician, scheduled appointment hours, whether an AI scribe was used or not (i.e., in the post-period), time spent on note writing activities both during and after the scheduled encounter, note characteristics including length of the note and note closure time, and wRVUs associated with the encounter. In addition, we also captured to the sex and age of both the clinician and patient associated with an encounter. All data were captured at the patient-encounter level.

### *Outcomes and covariates*

There were six outcomes of interest (see Appendix Table 1).

Note writing time was measured as an estimate of the active actions performed on the note including keystrokes, mouse clicks, and mouse movements (e.g., scrolling) with pre-determined timeouts when no activity is detected. Other outcomes included: time outside of scheduled hours (i.e., note writing time outside of scheduled clinic hours, but during a normal workday, 730AM to 5PM); "pajama time" (i.e., time spent on note writing from 530PM to 7AM); note length was number of characters associated with the finalized note; notes signed within the first 24h after the scheduled encounter; and total work RVUs associated with an encounter as reported in the final billing data.

### *Statistical analysis*

To assess the effect of AI scribe tool introduction on each of the outcomes of interest, we modeled typical serial variability in the pre- and post-period using a latent, time series process, and modeled non-linear deviations from that using a generalized additive model specification in the post-period.[25] The latent time series component used the pre-period to tune parameters describing typical serial variability, and in the post-period, accounting for typical serial variability and provided an underlying nowcasted counterfactual, representing the potential untreated outcome; the non-linear intervention effect captured the deviation from the existing serial process, which when combined yields the effect associated with the intervention of interest under the potential outcomes framework.

We fitted a Bayesian non-linear interrupted time series model with a latent autoregressive term accounting for day-to-day practice serial variability in clinician work patterns in a generalized linear modeling framework. Details of the model including model specification, likelihood, prior parameterization, model diagnostic assessment, model predictions, and spline knot specification, are provided in the Supplementary material (Section B).

Two outputs were generated from the interrupted time series model: (a) the average time-varying change from pre-AI scribe period as characterized by the non-linear intervention effect specification characterized as incidence rate ratios in the count response processes and rate ratios in the wRVU process (referred to as IRR hereafter), and (b) the difference in predicted outcomes between the modeled process and counterfactual draws aggregated across the 7 day periods starting at day 0, 30, 60, 90, 120, and 150; this incorporated patient encounter-level prediction error, allowing for the prediction of the effect of AI scribe with a new clinician at day 0 (referred to as “predicted changes” at day 0, day 30, etc. in the results).

Given the focus on descriptive statistical inference over time in the treatment period, we focus on reporting uncertainty via symmetric 95% credible intervals (CIs) based on Markov Chain Monte Carlo draws from posterior simulations across four chains rather than hypothesis testing using p-values.

In sensitivity analyses, we also added secular components that adjust for clinician age group at baseline (<36, 36-55, >55) and sex, and for clinician role (physician vs. APP), and we conducted stratified analyses by clinician role (Supplementary Material Section E).

All models were fit using R version 4.4.1 and Stan version 2.36.0, and all preliminary fits to assess our thin plate spline modeling choices were fit using the mgcv package.[26-28]

## Results

### *General characteristics*

220 clinicians (56% physicians) who worked in 36 clinics were included (Mean age=43.7, 70.9% females). These clinicians were involved in 314,845 patient encounters during the 48-week time-period centered around their first date of AI scribe access (see Figure 1 for CONSORT). Table 1 provides encounter-level pre- and post-AI scribe period summary statistics.

Based on unadjusted analysis, in the post AI scribe period (66.7% of total encounters) the average note writing time [4.39 mins (SD=3.78) vs. 3.83 mins (SD=3.45)], average TOSH [1.23 mins (SD=5.54) vs. 1.19 mins (SD=5.79)] and pajama time [1.71 mins (SD=6.31) vs. 1.62 mins (SD=6.23)] were marginally lower, compared to the pre-period. However, in the post-period,

average note length [8900 characters (SD=8050) vs. 9140 characters (SD=7480)], percentage of notes closed within the first 24 hours [84.2% vs. 84.3%], and wRVU [2.12 (SD=0.88) vs. 2.16 (SD=0.85)] were higher.

Time series for each of the outcomes over the 48-week period are shown in (Supplementary Material (Section C) and were aggregated at the day-level, along with a time zero marker separating the pre- and post-AI scribe periods.

***Interrupted time series of AI scribe effects***

Figure 2 illustrates the post-AI scribe initiation IRRs for each outcome over time with 95% CIs; these represent the average change after AI scribe introduction per encounter while accounting for time varying trends. Table 2 provides a summary of the IRR and predicted changes in outcomes in the clinician week starting at day 0 and day 150 in the post period for a new clinician introduced to the AI scribe along with 95% credible intervals.

At day zero, compared to the pre-period, there was evidence of an immediate 7% decrease on average in note writing time (IRR 0.93, 95%CI [0.89, 0.96]), which continued decrease to 15% below expected time varying trends prior to the AI scribe introduction (IRR 0.85, 95%CI [0.83, 0.87]) by day 150. This translated to approximately 34.5 minutes of note writing time savings per week at day 0 (34.52, 95%CI [10.64, 59.28]) and approximately 68 minutes of note writing time savings per week at day 150 (68.08, 95%CI [50.08, 87.6]).

Compared to the pre-period, there was no evidence of a change in pajama time at day zero (0.96, 95%CI [0.81, 1.14]); however, at day 150 we found evidence of a 18% decrease in pajama time (0.82, 95%CI [0.73, 0.91]). This translated to approximately 34.4 minutes of pajama time savings per week at day 150 (34.45, 95%CI [6.01, 65.21]).

There was no evidence of a change in TOSH at day zero (0.89, 95%CI [0.73, 1.05]); however, at day 150 we found evidence of a 13% decrease in TOSH (0.87, 95%CI [0.77, 0.99]), compared to the pre-period. This translated to approximately 15.6 minutes of TOSH savings per week at day 150 (15.62, 95%CI [-6.49, 39.11]).

There was evidence of increase in note length by 5% (1.05, 95%CI [1.00, 1.10]) at day zero; this translated to a predicted 80 character increase per week at day zero (0.08, 95%CI [-4.16, 4.51]) (per 1000 characters). However, this change slowly attenuated across the post-period, with no evidence of note length differences compared to the pre-period by day 150 (1.02, 95%CI [0.99, 1.05]). Similarly, at day zero, note closures within 24h were 31% higher (1.31, 95%CI [1.13, 1.53]). This was equivalent to approximately 2.8 additional note closures per week at day zero (-2.86, 95%CI [-12.14, 5.0]). By day 150, there was no evidence of differences in note closures within 24h compared to the pre-period (0.93, 95%CI [0.84, 1.02]).

Finally, there was evidence of a 2% increase total wRVU at day 150 (1.02, 95%CI [1.01, 1.03]). This was equivalent to an additional 4.4 wRVUs per week (4.4, 95%CI [-44.52, 52.71]). Additional IRRs and predicted changes at day 30, 60, 90, and 120 are provided in Supplementary material (Section D).

Additional details of the sensitivity analysis by age, sex and clinician type (physician vs. APP) are provided in Supplementary material (Section E).

## Discussion

Using a longitudinal, quasi-experimental study, we found that the use of an AI scribe decreased note writing time by 15% (>60 minutes per week of predicted time savings) and pajama time by 18% (>30 minutes per week) by day 150. The gain on the clinician productivity was marginal (2% increase per week) by day 150.

During the AI scribe period, there were changes from day zero to day 150 for all considered outcomes, highlighting the nuances regarding the adoption and use of novel AI tools. However, changes in outcomes were gradual, but persistent – sometimes changing directions (e.g., note closure <24h and note length) – underscoring the gradual process of adaptation of AI tools in clinical settings, as clinicians situate the tools within their workflows. Besides the two published RCTs,[13,20] to the best of our knowledge, this is one of the first studies offering a longitudinal perspective on AI scribe usage.

In terms of the outcomes, several interesting patterns emerged. At day zero, there was a predicted time savings of approximately half-hour per week, and by day 150, the savings were over an hour per week. Afshar et al, in their recent RCT, reported that the savings in time was approximately 22 minutes per clinician-day, but the method of measurement or the source of the data was unclear.[20] Our metric of note writing time was based on a measure related to active note writing actions (i.e., with recorded keystrokes and mouse movement activity), and is likely a considerable underestimate of time savings.

Although the post-period RVUs were minimal (2% or a predicted 4.4 additional RVUs per week) by day 150, there were considerable differences between physicians and APPs (Supplement, Section E). For example, there were no gains in wRVUs for APPs across the post-period. However, there was a marginal increase in RVUs over the post-period for physicians, with a 3% increase by day 150 (Supplementary material, Section E). This translates to a predicted increase of approximately 6 RVUs per physician or a $194 increase per physician increase in reimbursement for the week starting day 150, using the 2025 Medicare Physician Fee schedule. Our physician financial productivity metrics were higher that the findings from a recent study that reported ~1.8 more RVUs per week among physicians using an AI scribe.[29] It is expected that ambient scribe technology may lead to increased billing and risk-adjustment intensity, potentially prompting a re-calibration from the payer side.[30,31]

The longitudinal measurements also helped uncover nuanced findings highlighting the interplay between adaptation of AI scribe use, aligned with the potential benefits of the AI scribe. For example, a major advantage of using an AI scribe is that a note is generated based on the active patient-clinician conversation; as a result, a clinician does not require mnemonics or personal notes to recall the details when completing the note. This reduces the cognitive load associated with the editing of the note, as the ambient transcript reminds the clinician of the encounter details. We found that the early post-AI scribe period (first 30 days) was associated with ~31% higher note closures within the first 24 hours. After day 90, there was no evidence for note closure within 24h compared to the pre-period (Figure 2). This highlights the potential adaptation clinicians made with the AI scribe, recognizing that they did not have to recall—a cognitively challenging task—the details of an encounter when editing a note, instead relying on the completed transcript for recall, which helps in postponing the completion of the note.

Similar adaptations in clinician behaviors were also observed in note length and pajama time. As has been reported elsewhere, there have been observed increases in note length[7,9] and no

changes in work outside of work (or pajama time)[13,32] after AI scribe implementation. This was also observed in our study immediately after AI scribe implementation (day zero); however, as can be seen in Figure 2, over time, there was a declining trend in the note length, with no evidence of differences in note length by day 150, compared to the pre-period. In contrast, for pajama time there was no evidence of a difference in pajama time till approximately day 90, after which, there was a downward trend, with evidence of time savings by day 150.

From a methodological standpoint, there were several innovative aspects. First, as opposed to almost all prior work that has relied on aggregated, clinician-level metrics provided by vendor platforms (e.g., Epic's Signal), we utilized a patient encounter-level framework. This allowed for the incorporation of important variables (e.g., wRVU, encounter type) for developing homogenous samples. Second, as is evident across the findings, clinicians evolve in their AI scribe use over time. Using single-point measurements (e.g., 1-month before/after) would not capture the true impact of tool adaptations. The current approach allows for characterizing a time-varying effects reflecting AI scribe adaptation over time. This was also implemented in a robust, interrupted time series framework that accounts for serial variability that estimates the treatment effect conservatively through a parsimony enforcing prior. Finally, our encounter-level approach allowed in including all clinicians (physicians, APPs) with AI scribe access to the AI scribe over the study period.

This study has several limitations. The study was conducted in primary care settings from 36 separate clinics that were part of the same health system. Note length was based on the number of characters only; we did not consider the use of smart text or smart phrases (e.g., Epic EHR's text shortcuts) which influences the number of characters in a note. It is possible that clinicians used more of these in the pre-period given the templated structure of outpatient notes. We used an ITT analysis framework, although there were differences in AI scribe usage in the post-period at the clinician level; the primary assumption was that clinician choice of the AI scribe during an encounter based on some rational choices. An in-depth study on how scribe tool availability and use in specific types of encounters and clinician and patient characteristics may highlight additional benefits of these tools. We used an interrupted time series approach to robustly estimate the effect associated with AI scribe tool introduction; the underlying assumption that typical serial variability will follow a consistent pattern across the study period appears reasonable, but a physical randomization approach on observations at the encounter level with longitudinal follow-up would be a superior design.

**Funding**

The funding for this study was provided in part by the Center for Health AI (a joint Washington University School of Medicine and BJC Health System endeavor), the Professional Fulfillment Program (WashU Division of General Internal Medicine) and the Barnes-Jewish Hospital Medical Staff Association.

**Conflict of Interests**

Dr. Kannampallil serves as an editor for an Elsevier journal, outside of this work. The other authors report no conflicts of interest. Dr. Kannampallil is an unpaid advisor on the research steering committee for Abridge. Dr. Maddox is a paid advisor on FDA's Digital Health Advisory committee, an advisory board member for Elion and a board director for the J.F. Maddox Foundation.

## References


1. Cohen GR, Boi J, Johnson C, Brown L, Patel V. Measuring time clinicians spend using EHRs in the inpatient setting: a national, mixed-methods study. *Journal of the American Medical Informatics Association*. 2021;28(8):1676-1682.
2. Overhage JM, McCallie Jr D. Physician time spent using the electronic health record during outpatient encounters: a descriptive study. *Annals of internal medicine*. 2020;172(3):169-174.
3. Apathy NC, Rotenstein L, Bates DW, Holmgren AJ. Documentation dynamics: note composition, burden, and physician efficiency. *Health Services Research*. 2023;58(3):674-685.
4. Gaffney A, Woolhandler S, Cai C, et al. Medical documentation burden among US office-based physicians in 2019: a national study. *JAMA internal medicine*. 2022;182(5):564-566.
5. Moy AJ, Schwartz JM, Chen R, et al. Measurement of clinical documentation burden among physicians and nurses using electronic health records: a scoping review. *Journal of the American Medical Informatics Association*. 2021;28(5):998-1008.
6. Sinsky C, Colligan L, Li L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Annals of internal medicine*. 2016;165(11):753-760.
7. Kanaparthy NS, Villuendas-Rey Y, Bakare T, et al. Real-World Evidence Synthesis of Digital Scribes Using Ambient Listening and Generative Artificial Intelligence for Clinician Documentation Workflows: Rapid Review. *JMIR AI*. 2025;4:e76743.
8. Altschuler S, Huntington I, Antoniak M, Klein LF. Clinician as editor: notes in the era of AI scribes. *The Lancet*. 2024;404(10468):2154-2155.
9. Stults CD, Deng S, Martinez MC, et al. Evaluation of an ambient artificial intelligence documentation platform for clinicians. *JAMA Network Open*. 2025;8(5):e258614-e258614.
10. Tierney AA, Gayre G, Hoberman B, et al. Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. *NEJM Catalyst Innovations in Care Delivery*. 2024;5(3):CAT. 23.0404.
11. Guo Y, Wang J, Hu D, et al. Evaluating ambient artificial intelligence documentation: effects on work efficiency, documentation burden, and patient-centered care. *Journal of the American Medical Informatics Association*. 2025:ocaf180.
12. Ma SP, Liang AS, Shah SJ, et al. Ambient artificial intelligence scribes: utilization and impact on documentation time. *Journal of the American Medical Informatics Association*. 2025;32(2):381-385.
13. Lukac PJ, Turner W, Vangala S, et al. A Randomized-Clinical Trial of Two Ambient Artificial Intelligence Scribes: Measuring Documentation Efficiency and Physician Burnout. *NEJM AI*. 2025;2(12)
14. Liu T-L, Hetherington TC, Dharod A, et al. Does AI-powered clinical documentation enhance clinician efficiency? A longitudinal study. *Nejm Ai*. 2024;1(12):AIoa2400659.

15. Galloway JL, Munroe D, Vohra-Khullar PD, et al. Impact of an artificial intelligence-based solution on clinicians' clinical documentation experience: initial findings using ambient listening technology. *Journal of General Internal Medicine*. 2024;39(13):2625-2627.

16. Pelletier JH, Watson K, Michel J, McGregor R, Rush SZ. Effect of a generative artificial intelligence digital scribe on pediatric provider documentation time, cognitive burden, and burnout. *JAMIA Open*. 2025;8(4):ooaf068.

17. Albrecht M, Shanks D, Shah T, et al. Enhancing Clinical Documentation Workflow with Ambient Artificial Intelligence: Clinician Perspectives on Work Burden, Burnout, and Job Satisfaction. *medRxiv*. 2024:2024.08. 12.24311883.

18. Shah SJ, Devon-Sand A, Ma SP, et al. Ambient artificial intelligence scribes: physician burnout and perspectives on usability and documentation burden. *Journal of the American Medical Informatics Association*. 2025;32(2):375-380.

19. Shin HS, Williams H, Braykov N, Jahan A, Meller J, Orenstein EW. The Influence of Artificial Intelligence Scribes on Clinician Experience and Efficiency among Pediatric Subspecialists: A Rapid, Randomized Quality Improvement Trial. *Applied Clinical Informatics*. 2025;16(04):1041-1052.

20. Afshar M, Ryan Baumann M, Resnik F, et al. A pragmatic randomized controlled trial of ambient artificial intelligence to improve health practitioner well-being. *NEJM AI*. 2025;2(12):AIoa2500945.

21. Kim E, Liu VX, Singh K. AI Scribes Are Not Productivity Tools (Yet). NEJM AI: Massachusetts Medical Society; 2025. p. AIe2501051.

22. Hammer J. Cognitive Adaptation Through AI Collaboration: A Longitudinal Case Study. *Available at SSRN 5525162*. 2025;

23. Berg M. Patient care information systems and health care work: a sociotechnical approach. *International journal of medical informatics*. 1999;55(2):87-101.

24. Berg M. Implementing information systems in health care organizations: myths and challenges. *International journal of medical informatics*. 2001;64(2-3):143-156.

25. Cho S-J. Modelling change processes in multivariate interrupted time series data using a multivariate dynamic additive model: An application to heart rate and blood pressure self-monitoring in heart failure with drug changes. *Journal of the Royal Statistical Society Series C: Applied Statistics*. 2024;73(1):123-142.

26. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2011;73(1):3-36.

27. Team RC. R: a language and environment for statistical computing.[cited 2012 Sep 20]. *Available from: www r-project org*. 2012;

28. *Stan Reference Manual*. Version 2.36.0. 2024. https://mc-stan.org

29. Holmgren AJ, Fenton CL, Thombley R, et al. Ambient Artificial Intelligence Scribes and Physician Financial Productivity. *JAMA Network Open*. 2026;9(1):e2553233-e2553233.

30. Dai T, Kvedar JC, Polsky D. Policy brief: ambient AI scribes and the coding arms race. *npj Digital Medicine*. 2025;8(1):780.

31. Nong P, Neprash HT. Unintended consequences of using ambient artificial intelligence scribes for billing. American Medical Association; 2026:e255771-e255771.

32. Hassan H, Zipursky AR, Rabbani N, et al. Special topic on burnout: clinical implementation of artificial intelligence scribes in healthcare: a systematic review. *Applied Clinical Informatics*. 2025;

## Tables and Figures

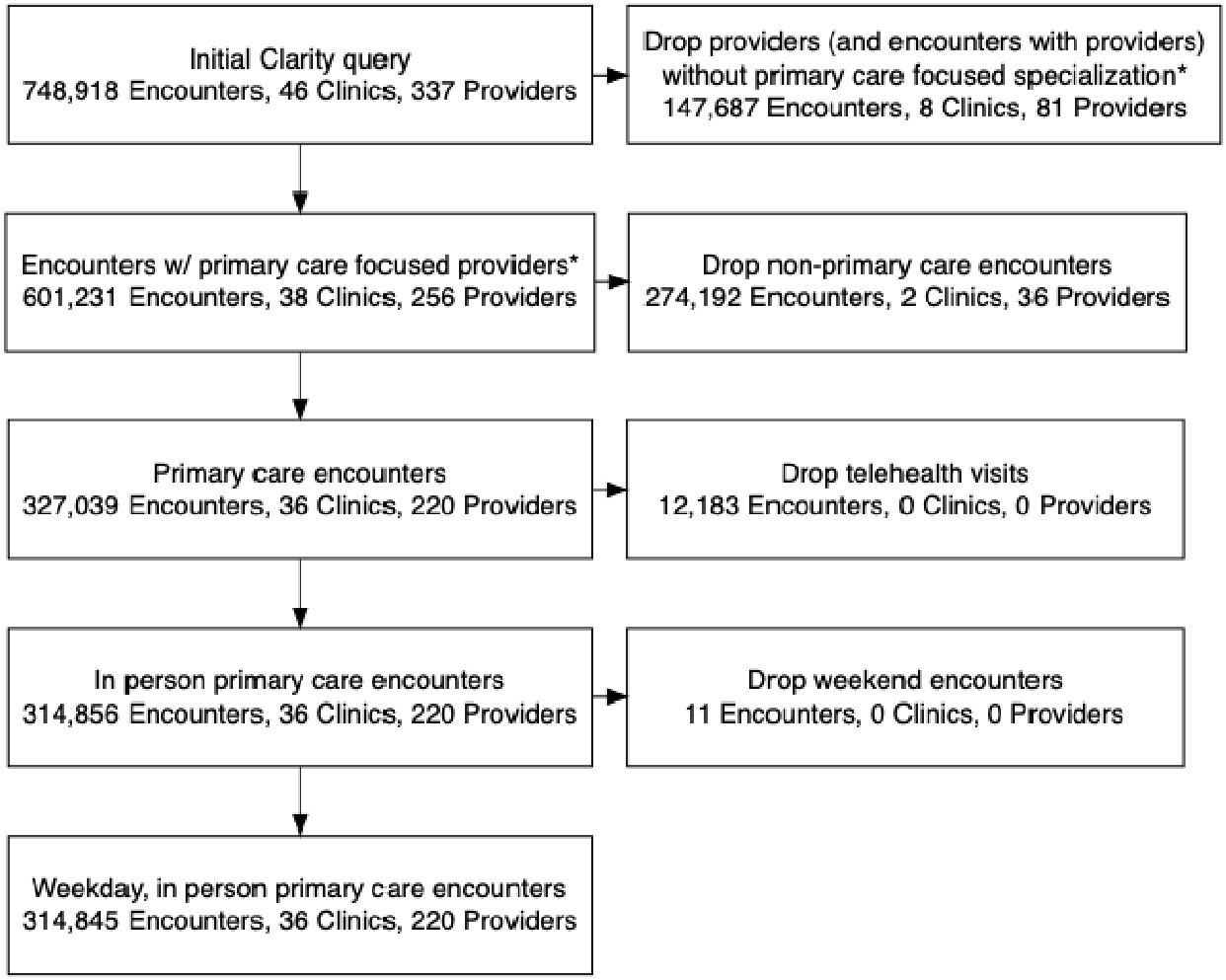
**Table 1**. Summary of clinicians, encounters and outcomes

| | **Pre (N=138252)** | **Post (N=176593)** |
|---|---|---|
| **Clinician specialty** | **Clinicians, Encounters** | **Clinicians, Encounters** |
| Family Medicine | 109, 73359 | 112, 96017 |
| Internal Medicine | 72, 57576 | 75, 72428 |
| Obstetrics and Gynecology | 22, 298 | 21, 328 |
| Pediatrics | 6, 7019 | 6, 7820 |
| **Clinician credentials** | **Clinicians, Encounters** | **Clinicians, Encounters** |
| APP | 86, 31618 | 90, 42459 |
| MD/DO | 123, 106634 | 124, 134134 |
| **Abridge Scribe Usage (%)** | NA | 66.7% |
| **Outcomes** | **Mean (SD)** | **Mean (SD)** |
| Note writing time (Mins) | 4.39 (3.78) | 3.83 (3.45) |
| Pajama time (Mins) | 1.71 (6.31) | 1.62 (6.23) |
| Time Outside of Scheduled Hours (Mins) | 1.23 (5.54) | 1.19 (5.79) |
| Note length (Characters) | 8900 (8050) | 9140 (7480) |
| Note closed <24h | 116423 (84.2%) | 148911 (84.3%) |
| Total work Relative Value Units | 2.12 (0.881) | 2.16 (0.856) |

Table 2. Incidence Rate Ratios (IRR) and predicted changes at day 0 and day 150. *Indicates significant IRRs

| **Outcome measure** | **IRR at Day 0** | **Predicted Change at Day 0** | | **IRR at Day 150** | **Predicted Change at Day 150** |
|---|---|---|---|---|---|
| Note writing time (mins) | 0.93, (0.88, 0.98)* | 34.52, (10.64, 59.28) | | 0.85, (0.82, 0.87)* | 68.08, (50.08, 87.6) |
| Pajama time (mins) | 0.96, (0.81, 1.14) | 5.04, (-24.71, 40.33) | | 0.82, (0.73, 0.91)* | 34.45, (6.01, 65.21) |
| Time Outside of Scheduled Hours (mins) | 0.89, (0.73, 1.05) | 12.65, (-15.11, 43.47) | | 0.87, (0.77, 0.99)* | 15.62, (-6.49, 39.11) |
| Note length (1000 Characters) | 1.05, (1.0, 1.1)* | 0.08, (-4.16, 4.51) | | 1.02, (0.99, 1.05) | -0.05, (-4.36, 4.19) |
| Note closed <24h | 1.31, (1.13, 1.53)* | 2.86, (-5.00 12.14) | | 0.93, (1.02, 0.84) | -1.43, (-10.73, 8.57) |
| Total work Relative Value Units (wRVU) | 1.02, (1.0, 1.04)* | 3.24, (-46.98, 52.72) | | 1.02, (1.01, 1.03)* | 4.4, (-44.52, 52.71) |

**Figure 1.** Consort diagram (within inclusion and exclusion criteria)



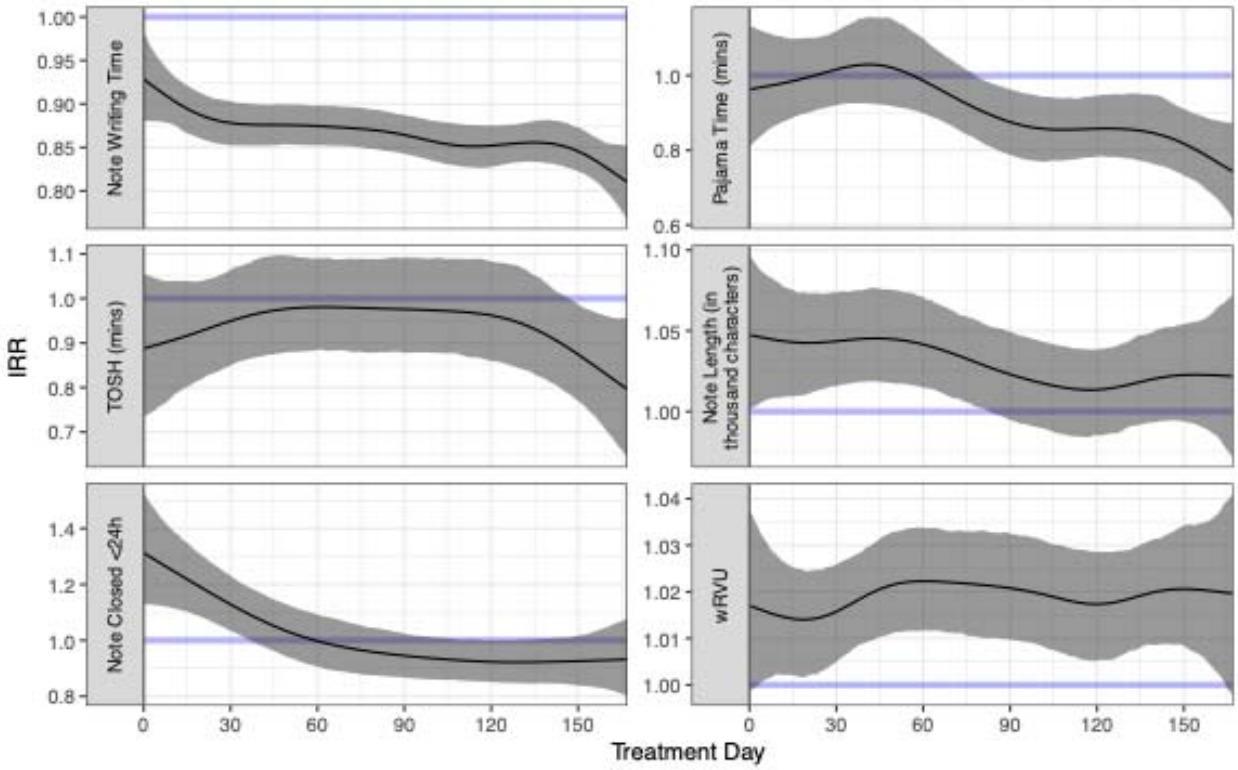



**Figure 2.** IRRs over in the post AI scribe period from day zero to day 168 (i.e., 24 weeks, post-period) for each of the considered outcomes. Note: in the manuscript we report IRRs at specific time intervals of 0, 30, 60, 90, 120 and 150 days corresponding to approximately monthly intervals