AI Adoption and Productivity in Healthcare

Erkmen G. Aslim* Manuel Hoffmann‖ Haonan Yin‖

*University of Vermont & IZA

‖University of California, Irvine

This Abstract Version: *March 26, 2026*

Abstract

Artificial Intelligence (AI) has the potential to reshape work processes and productivity in the health care sector. It presents a unique opportunity to enhance the economic value of professional work by automating time–intensive peripheral tasks, allowing highly skilled practitioners to focus on their core technical expertise. We investigate the adoption, productivity, and work transformation effects of AI scribes – ambient generative AI systems that document clinical conversations – within a large academic health network. Leveraging a staggered rollout and granular data both from electronic health records and the AI scribe platform, we find that adoption of AI scribes is remarkably rapid, reaching near-universal coverage among primary care physicians within fourteen months. Physicians increase their usage post-eligibility and accept the vast majority of AI-generated notes based on audio recordings of physician-patient encounters. As physicians gain experience over twelve months, they transition from passive adopters to efficient users, with shorter recording durations per note as they learn to structure interactions more efficiently for the AI system. By reducing the marginal cost of documentation and alleviating administrative frictions, AI scribes enable a significant reallocation of time toward patient–centered care. Ultimately, this longitudinal analysis demonstrates that when AI tools are designed to address pressing professional burdens, they have the potential to effectively bridge the implementation gap and drive long–term improvements in both the quality and efficiency of care delivery.

JEL-Classification: I10, J0, O33

Keywords: Generative Artificial Intelligence, AI Scribes, Healthcare, Productivity, AI Adoption

Acknowledgments: We are grateful for administrative support from Abridge and the academic health network. This study was reviewed and approved by the Institutional Review Board at the University of Vermont (IRB No. 00000485) and the University of California, Irvine (IRB No. 00000596).

> "We are giving agency back to clinicians, creating clarity for patients, and establishing radical efficiencies to transform healthcare from the conversation up."
>
> — Shiv Rao, MD
>
> CEO and Founder, Abridge

Artificial Intelligence (AI) is predicted to substantially transform the economy, work processes, and productivity (Hoffmann et al., 2024; Brynjolfsson, Li, and Raymond, 2025; Dillon et al., 2025; Bick, Blandin, and Deming, 2025; Hoffmann et al., 2025). In healthcare, one particularly promising application of generative AI is the reduction of documentation burden. Ambient documentation systems (i.e., "AI scribes") aim to alleviate this burden by automatically generating structured clinical notes from audio recordings of patient–physician encounters (Afshar et al., 2025). These systems use speech recognition and large language models to transcribe conversations, extract clinically relevant information, and format it into standardized notes compatible with electronic health records (EHRs). By automating documentation tasks, AI scribes are expected to reduce administrative workload, mitigate physician burnout, allow clinicians to focus more fully on patients during visits, enhance patient–physician interaction, and support more targeted treatment decisions.

If AI scribes meaningfully reduce documentation burden, they should alter physicians' time allocation, workflow, and potentially their well-being. We study the implementation of a generative AI clinical documentation tool, characterize physician adoption and usage intensity, and estimate its effects on documentation efficiency, measured by editing effort, and physician productivity. The tool we examine is Abridge[1], an ambient AI documentation system deployed within a large, multi-hospital academic health network integrated into its electronic health record platform. The network includes six operating hospitals, employs over 15,000 staff, and serves more than one million patients across two states, providing a sizable and geographically diverse clinical setting.

We leverage novel, high-frequency (daily) physician-level proprietary data on AI scribe adop-

[1] https://www.abridge.com

tion, usage intensity, and output quality, linked to detailed productivity metrics at the physician–encounter level. A central advantage of our setting is the ability to construct novel physician-level linkages between AI scribe activity and electronic health record (EHR) data, allowing us to measure objective productivity outcomes such as time spent in the system, note completion timing, appointment closure rates, and documentation turnaround time. This unusually granular integration of AI usage data with operational clinical metrics enables us to observe how generative AI adoption translates into measurable changes in workflow and performance.[2]

Our identification strategy exploits within-provider variation over time in a dynamic difference-in-differences (DID) framework. In the first stage, we estimate the dynamic takeup of AI scribes, where provision of access induces a discontinuous shift in usage over time. Because providers receive access at different times, we employ a heterogeneity-robust DID estimator proposed by Callaway and Sant'Anna (2021), which accommodates staggered treatment timing by comparing early adopters to not-yet-treated providers. We further examine how usage evolves with experience, modeling outcomes as a function of tenure since provision to disentangle the role of learning from the initial adoption response.

We document three sets of results. First, adoption is immediate and widespread: over half of physicians generate their first AI-scribed note within one month of eligibility, rising to near-universal coverage within fourteen months, with approximately 80 to 90 percent of AI-generated text accepted without modification throughout. Second, experience reshapes how physicians interact with the technology. Physicians are 12 to 15 percent more likely to utilize AI-generated notes in their first year, and as they accumulate exposure beyond twelve months, they reduce audio recording durations by nearly five minutes per note — consistent with learning to structure clinical interactions more efficiently for the AI system. Third, experienced physicians shift from passive acceptance to active curation, editing AI-generated text more frequently even as their satisfaction with note quality converges to baseline. Taken together, these efficiency gains in documentation

[2]For a subset of physicians, we additionally observe repeated institutional survey responses, which provide complementary measures of well-being (e.g., self-reported burnout) and time allocation across clinical and administrative tasks. These data allow us to explore mechanisms underlying the observed productivity effects.

create scope for meaningful time reallocation toward patient-centered care.

This paper contributes to the rapidly-evolving literature on the impacts of AI on productivity and the changing nature of professional work. While previous research has focused on AI tools that reduce the cost of core tasks— such as software development (Hoffmann et al., 2024), we document a reverse but complementary effect. In our setting, the AI scribe targets "peripheral" or bureaucratic tasks to enable a focus on core clinical work. This finding serves as a vital complement to the idea that structural frictions within EHR systems act as a significant drag on productivity (Dix, Moran, and Nguyen, 2025). We demonstrate that by improving the inflow of information, AI scribes effectively reduce these administrative frictions, which is as significant to productivity as enhancing core technical outputs.

Second, we provide empirical evidence that there are remarkably few barriers to AI adoption when the tool is specifically designed to alleviate well–documented professional burdens. The rapid pace of adoption and the intensity of usage observed among the physician cohort in our sample serve as a powerful signal of the tool's value through revealed preferences. This suggests that the "implementation gap" often seen with new medical technologies can be bridged when the technology aligns directly with the clinician's desire to return to patient–centered care.

Third, our study offers a unique longitudinal view of the "AI–practitioner" relationship. By utilizing proprietary data on word acceptance and audio duration, we move beyond simple binary adoption metrics. We demonstrate that the value of these documentation tools does not just come from a temporary growth in encounters, but from a sustained increase in documentation volume and a long–term improvement in dictation efficiency as physicians master the tool.

1 Empirical Context

1.1 The AI Scribe Tool

The AI scribe we examine is Abridge, an ambient AI documentation system that automatically generates structured clinical notes from audio recordings of patient–physician encounters.

Founded in 2018 by clinicians, AI scientists, and healthcare experts, Abridge has been deployed across more than 150 healthcare systems in the United States, including major academic medical centers such as UPMC, Yale New Haven Health, Emory Healthcare, Sutter Health, MemorialCare, and the University of Kansas Health System. The platform uses proprietary, fully auditable AI technology purpose-built for healthcare, supporting over 50 clinical specialties, 28 languages, and all care settings.

At its core, Abridge listens to the natural conversation between a physician and a patient during a clinical encounter, transcribes the audio using speech recognition, extracts clinically relevant information, and formats it into a structured clinical note compatible with the electronic health record (EHR). The system is designed to run passively in the background, requiring minimal physician interaction during the visit itself.

— Figure 1 about here —

Figure 1 displays the Abridge web-based note review interface, which consists of three primary panels. The left panel shows the physician's worklist – a chronological schedule of patient encounters for the day, each associated with a recording duration. The center panel is the note editor, where the AI-generated clinical note is displayed in standard medical documentation format with sections for History of Present Illness, Physical Exam, and Assessment & Plan. At the bottom of this panel, physicians can rate the note quality using a five-star system, copy the full note text to clipboard, and mark the encounter as complete. The right panel provides the conversation transcript, allowing physicians to verify the AI-generated content against the original patient–physician dialogue.

The example note shown illustrates a comprehensive clinical encounter for a patient presenting with shortness of breath and chest tightness. The AI system has organized the conversation into a structured HPI narrative, a system-by-system physical exam, and a problem-oriented Assessment & Plan with specific action items – including lymph node biopsy recommendations and medication management. Notably, the transcript panel on the right preserves the natural, conversational tone of the encounter, demonstrating how the system distills unstructured dialogue into formal clinical

documentation. The interface also features a "Copy All" button and a "Mark as Done" workflow indicator, reflecting the tool's integration into the physician's documentation pipeline, where the reviewed note is subsequently transferred to the electronic health record.

2 Methodology

The staggered rollout of AI scribes across the health network generated plausibly exogenous variation in physician eligibility over time, enabling a difference-in-differences (DID) framework to estimate causal effects.[3] We observe the full roster of physicians practicing within the network through administrative EHR data, allowing us to construct a panel at the physician level and follow both adopters and non-adopters over time. This comprehensive coverage enables us to include physician fixed effects and estimate within-physician changes in productivity associated with eligibility and subsequent tool use. To recover static treatment effects, we estimate a pooled two-way fixed effects model:

$$y_{it} = \alpha + \beta \cdot \text{Post}_t + \gamma_i + \delta_t + \varepsilon_{it} \tag{1}$$

where y_{it} is the outcome for physician i in period t, $\text{Post}_t = \mathbf{1}[t \geq t^*]$ is an indicator equal to one after the pooled eligibility provisioning date t^*, γ_i denotes physician fixed effects, δ_t denotes month fixed effects, and ε_{it} is an idiosyncratic error term clustered at the physician level. The coefficient β captures the average within-physician change in the outcome following eligibility. To trace out dynamic effects, we estimate the following model:

$$y_{it} = \alpha + \sum_{\ell \neq -1} \beta_\ell \cdot \mathbf{1}[t - t_i^* = \ell] + \gamma_i + \delta_t + \varepsilon_{it} \tag{2}$$

where ℓ indexes event time relative to the eligibility date t_i^*, and β_ℓ captures the average treatment effect ℓ periods before or after eligibility. The period $\ell = -1$ serves as the reference. Coefficients for $\ell < 0$ provide pre-trend tests of the parallel trends assumption; coefficients for $\ell \geq 0$ trace the

[3] Importantly, physicians were not informed of the precise timing of eligibility in advance, reducing the likelihood of anticipatory behavioral responses prior to treatment.

dynamic productivity response to eligibility. To address the heterogeneous treatment effect bias inherent in standard TWFE estimators, we follow Callaway and Sant'Anna (2021) and estimate heterogeneity-robust aggregated ATTs.

To better understand learning dynamics, we estimate a separate model that replaces event time with experience - defined as time elapsed since the physician first adopted the tool:

$$y_{it} = \alpha + \sum_{b \neq 0} \theta_b \cdot \mathbf{1}[\text{Experience}_{it} = b] + \gamma_i + \delta_t + \varepsilon_{it} \tag{3}$$

where $\text{Experience}_{it} = b$ indicates that physician i is in experience bin b at time t, and θ_b captures the average outcome in bin b relative to the reference bin $b = 0$. Physician fixed effects γ_i absorb time-invariant heterogeneity across clinicians, and time fixed effects δ_t control for seasonality. The coefficients θ_b trace the within-physician productivity trajectory as experience with the tool accumulates, isolating learning from selection into adoption.

3 Data

We have obtained the roster of physicians of a large medical provider in the United States. It represents the universe of phycisians of this medical provider which provides near-universal coverage of patients in one state of the United States of America. We then merge the roster of physicians with data from the AI Scribe tool Abridge. This allows us to obtain granular information on AI Adoption, the notes transcribed via the AI Scribe as well as the audio duration of a physician-patient encounter in a session and whether physicians ultimately send the note to the electronic health record (EHR) system and how much of the AI generated content they accepted. [4]

— Table 1 about here —

Table 1 reports summary statistics for the estimation sample, which comprises $10,004$ physician-patient encounters across $1,868$ physicians. AI adoption is substantial, with 73% of physician-

[4] We have further received access to the complete EHR records. The full-scale analysis of the records will be done prior to the NBER SI in the updated working paper version.

month observations reflecting active use of the AI scribe. Conditional on the sample, physicians produce on average 32 notes per month (SD = 52.8), and log on average 573 minutes of audio per physician per month (SD = 981.0). Among AI scribe-specific engagement metrics, 29% of sessions result in the physician retaining the AI-generated note, and among those who accept AI-generated content, physicians incorporate on average 83% of the AI-generated text without modification, suggesting a high degree of trust in the tool among active users. The average note rating is 3.41 out of 5, indicating generally positive physician assessments of output quality.[5]

4 Results

4.1 Main Results

Our primary analysis focuses on all physicians in the health care network.

— Figure 2 about here —

Figure 2 presents event-time evidence on adoption and usage following eligibility within the time frame from May 2024 to Nov 2025. The figure is based on a balanced panel of 1,868 physicians and 35,492 patient–physician encounters observed before and after eligibility based on the raw data .[6] Adoption is immediate and widespread: 56 percent of physicians create their first AI-generated note within the first month of eligibility, and adoption rises steadily to near universal use within fourteen months.

Usage intensity increases in parallel. The number of AI-generated notes per physician rises quickly from approximately 20 in the first month to nearly 80 notes per month after one year. Audio recording time exhibits a similar pattern, increasing sharply upon eligibility and exceeding 1,700 minutes per physician per month by the end of the first year. Taken together, these patterns imply

[5]Tables A3, A1, A2 provide descriptive statistics for specialty (primary care or other), location (headquarter, satellite), and gender (male, female).

[6]Table A4 shows the main effects based on model 1. Figure A1 shows the staggered rollout figures based on equation 2.

a remarkably stable documentation structure of roughly 20 minutes of audio per note, suggesting systematic rather than sporadic use.

Importantly, physicians accept approximately 80 to 90 percent of AI-generated text without modification throughout the post-eligibility period. This high and stable acceptance rate provides a first-order indication of substantial documentation efficiency gains and suggests that the tool is meaningfully integrated into routine clinical workflow rather than used experimentally or intermittently.[7] Such reductions in documentation effort create scope for time reallocation, potentially increasing physician productivity.

The rapid adoption of AI scribes raises a natural question: does experience with the tool further reshape clinical workflow? To examine dynamic effects, we estimate models with physician and year-month fixed effects that exploit variation in exposure duration following eligibility. We define physician experience relative to the first month post-eligibility and group physicians into 2–5 months, 6–12 months, and more than 12 months of AI exposure.

— Table 2 about here —

Table 2 provides a comprehensive overview of how physician interaction with AI scribes evolves across six distinct dimensions of practice. By analyzing these metrics from left to right, we observe a transition from initial adoption surges to long term professional optimization.

The analysis begins with usage and technical engagement (columns 1–2). Upon becoming eligible, physicians demonstrate an immediate and sustained increase in documentation volume, with those in the 2-12 month cohorts generating approximately 30 additional notes compared to the baseline (column 1). Interestingly, column (2) reveals a significant initial increase in average audio duration of 2.44 minutes during the first five months. This likely represents an initial calibration phase where physicians provide more exhaustive input as they adjust to the system. However, as clinicians reach the one–year mark, they significantly reduce their average audio recording duration by 4.73 minutes per note. This suggests that "power users" develop more concise communication

[7]These patterns are confirmed in staggered difference-in-differences models with physician and year-month fixed effects, where the static average treatment effect on the treated (ATT) is positive, large in magnitude, and precisely estimated.

patterns, likely learning how to structure their patient interactions to be more efficiently processed by the AI system.

Following this technical adjustment, we observe the integration and adoption of AI content. Column (3) indicates that during the first year, physicians are 11.8% to 14.5% more likely to "keep" or utilize an AI–generated note relative to the baseline. While this specific propensity settles after one year, the sustained significance in column (1) suggests the tool remains a permanent fixture in the clinical workflow, even as the method of engagement shifts from more passive to more active engagement with AI scribes.

Finally, columns (4) through (6) illustrate the evaluation and curation phase, where the most complex behavioral adjustments occur. The AI acceptance rate (column 4) - the percentage of raw text accepted without editing - declines significantly as experience grows, reaching a reduction of 0.654 for the most senior users. However, this does not indicate a failure of the technology. When viewed alongside column (5), we see that while the propensity to provide a binary "note rating" decreases over time, the "average note rating" in column (6) shows that the initial significant dissatisfaction of -0.641 in the first few months disappears as users cross the one–year mark.

There are multiple plausible explanations for these observed dynamics. An efficiency story suggests that the tool effectively reduces the marginal cost of documentation, allowing for higher patient volume. A learning story is supported by the audio duration and acceptance data, indicating that physicians undergo a cognitive adjustment period before mastering the tool as a collaborative partner. Finally, a compositional change may be at play, where the types of encounters or the profile of physicians utilizing the tool shift as it moves from early adopters to the broader clinical population. Taken together, this progression suggests that the AI scribe journey begins with a focus on quantity and ends with sophisticated curation, where experienced physicians utilize efficiency gains to actively fine–tune notes to their professional standards.

4.2 Heterogeneity

4.2.1 Specialty: Primary Care vs Non Primary Care

We differentiate specialties by primary care physicians and other specialties. We define primary care, as physicians which are in family medicine and general internal medicine, i.e. clinical settings in which documentation demands are substantial and ambient AI tools are particularly well suited to routine outpatient encounters. Next, we show the same statistics as in Figure 2 with the exception of total audio recording for brevity. The total audio recording follows the usage trend for the number of notes quite closely.

— Figure 3 about here —

Figure 3 displays AI adoption, usage as measured in the number of notes, as well as the percent of AI generated text that is accepted in the notes for primary care physicians vs non primary care physicians.[8] Panels A and B show that both groups exhibit zero adoption prior to provisioning, followed by a steep rise that plateaus close to full adoption by months 15 post-provisioning. The adoption trajectories are strikingly similar across the two subgroups, suggesting that specialty type does not meaningfully differentiate whether physicians adopt the tool once it becomes available.

The most pronounced difference between the two groups emerges in usage intensity. Panels C and D reveal that primary care physicians ramp up to approximately 80–100 mean monthly notes by months 10–20, whereas non-primary care physicians reach only around 20–25 notes per month—roughly one quarter of the primary care volume. This pattern is consistent with our characterization of primary care settings as environments with high documentation burdens and routine outpatient encounter structures that are particularly amenable to ambient AI scribing.

Despite these differences in usage intensity, Panels E and F show that both groups accept AI-generated content at similarly high rates. Acceptance rates jump from zero pre-provisioning to approximately 80–85% immediately after provisioning and remain broadly stable thereafter.

[8]Figure A2 shows the staggered rollout figures based on equation 2.

This suggests that conditional on engaging with the tool, physicians across specialties find the AI-generated output largely acceptable, and that the lower usage intensity among non-primary care physicians reflects differences in encounter volume and workflow fit rather than dissatisfaction with output quality.[9]

4.2.2 Location: Headquarter vs Satellites

In a next step, we are interested whether adoption and usage patterns of AI Scribe differ for the headquarter vs satellite hospitals.

— Figure 4 about here —

Figure 4 displays AI adoption, usage as measured in the number of notes, as well as the percent of AI-generated text that is accepted in the notes for headquarter (UVM $= 1$) vs. other locations (UVM $= 0$).[10] Panels A and B show that both groups exhibit zero adoption prior to provisioning, followed by a steep rise that plateaus close to full adoption by approximately 15 months post-provisioning. The adoption trajectories are remarkably similar across the two location types, suggesting that institutional setting does not meaningfully differentiate whether physicians adopt the tool once it becomes available.

Differences between the two groups are more pronounced in usage intensity. Panels C and D reveal that headquarter physicians ramp up to approximately 40–60 mean monthly notes by months 10–20, whereas physicians at other locations reach somewhat higher volumes of approximately 60–100 notes per month, though confidence intervals are wide in both cases. This pattern may reflect differences in patient volume, encounter complexity, or workflow structure across institutional settings.

Despite these differences in usage intensity, Panels E and F show that both groups accept AI-generated content at similarly high rates. Acceptance rates jump from zero pre-provisioning to approximately 80–85% immediately after provisioning and remain broadly stable thereafter.

[9] Table A5 shows further heterogeneous impacts across primary care vs non primary care specialties by experience.
[10] Figure A3 shows the staggered rollout figures based on equation 2.

This suggests that conditional on engaging with the tool, physicians across both headquarter and satellite locations find the AI-generated output largely acceptable, and that any differences in usage intensity reflect workflow and volume factors rather than dissatisfaction with output quality.[11]

4.2.3 Gender: Male vs Female Physicians

Finally, due to substantial variation in adoption of technology across gender as well as differences in the workplace across gender, we are interested whether we find differences across male and female physicians in AI Scribe adoption and usage.

— Figure 5 about here —

Figure 5 displays AI adoption, usage as measured in the number of notes, as well as the percent of AI-generated text that is accepted in the notes for female vs. male physicians.[12] Panels A and B show that both groups exhibit zero adoption prior to provisioning, followed by a steep rise that plateaus close to full adoption by approximately 15 months post-provisioning. The adoption trajectories are strikingly similar across genders, suggesting that gender does not meaningfully differentiate whether physicians adopt the tool once it becomes available.

Turning to usage intensity, Panels C and D reveal a broadly similar pattern across genders. Both female and male physicians ramp up to approximately 40–60 mean monthly notes by months 5–10, continuing to grow to roughly 60–100 notes by months 15–20, with overlapping confidence intervals throughout. Unlike the specialty comparison, where primary care physicians showed markedly higher usage, the gender comparison does not reveal a pronounced divergence in documentation volume, suggesting that encounter structure and specialty mix may be more important determinants of usage intensity than gender per se.

Panels E and F show that both female and male physicians accept AI-generated content at similarly high rates, jumping from zero pre-provisioning to approximately 80–85% immediately

[11] Table A6 shows further heterogeneous impacts across headquarter vs satellites by experience.
[12] Figure A4 shows the staggered rollout figures based on equation 2.

after provisioning and remaining broadly stable thereafter. The near-identical acceptance trajectories across genders further suggest that conditional on engaging with the tool, physicians find the AI-generated output equally acceptable regardless of gender, and that any future divergence in outcomes is more likely to stem from differences in workflow integration than in attitudes toward the tool's output quality.[13]

5 Conclusion

This paper provides novel empirical evidence on the adoption, usage, and productivity effects of AI scribes in a large academic health network. Leveraging granular physician-level data from both the AI scribe platform and electronic health records, combined with a staggered rollout design and heterogeneity-robust difference-in-differences estimators, we document three main findings.

First, adoption is immediate and near-universal. Over half of physicians generate their first AI-scribed note within one month of eligibility, rising to near-universal coverage within fourteen months across all subgroups — regardless of specialty, location, or gender. The strikingly uniform adoption trajectories suggest that when AI tools are designed to directly address well-documented professional burdens, the implementation gap that typically characterizes new medical technologies can be effectively bridged.

Second, experience reshapes how physicians interact with the technology. In the short run, physicians increase documentation volume and audio recording duration, consistent with an initial calibration phase. Over time, however, experienced physicians reduce recording durations by nearly five minutes per note, indicating that they learn to structure clinical interactions more efficiently for the AI system. This transition from passive adoption to active curation — evidenced by increased editing of AI-generated text alongside stable or improving note quality ratings — suggests a deepening and professionalization of the human–AI collaboration over time.

Third, the productivity gains from AI scribes are not uniformly distributed. Primary care

[13]Table A7 shows further heterogeneous impacts across male vs female physicians by experience.

physicians, who face the heaviest documentation burdens, exhibit substantially higher usage intensity than non-primary care physicians, reaching roughly four times the monthly note volume. By contrast, gender and location produce minimal divergence in either adoption or usage trajectories, suggesting that encounter structure and specialty mix are the primary moderators of how much value physicians extract from ambient AI tools. Across all subgroups, AI acceptance rates converge rapidly to approximately 80–85%, indicating that output quality is broadly acceptable regardless of physician characteristics.

Taken together, these findings paint a picture of AI scribes as a technology that successfully reduces the marginal cost of documentation and reallocates physician time toward patient-centered care. By automating peripheral administrative tasks rather than core clinical work, AI scribes represent a complementary rather than substitutive intervention — one that amplifies the value of physician expertise rather than displacing it. Our results suggest that the productivity gains from AI in healthcare may be largest precisely where administrative frictions are most severe, pointing to a broader principle: AI tools that target well-identified professional pain points are likely to achieve both rapid adoption and sustained impact.

References

Afshar, Majid, Mary Ryan Baumann, Felice Resnik, Josie Hintzke, Anne Gravel Sullivan, Graham Wills, Kayla Lemmon, Jason Dambach, Leigh Ann Mrotek, Mariah Quinn et al. 2025. "A pragmatic randomized controlled trial of ambient artificial intelligence to improve health practitioner well-being." *NEJM AI* 2 (12):AIoa2500945.

Bick, Alexander, Adam Blandin, and David Deming. 2025. "The impact of generative ai on work productivity." Tech. rep., Federal Reserve Bank of St. Louis.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at Work." *The Quarterly Journal of Economics* 140 (2):889–942.

Callaway, Brantly and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of Econometrics* 225 (2):200–230.

Dillon, Eleanor W, Sonia Jaffe, Nicole Immorlica, and Christopher T Stanton. 2025. "Shifting work patterns with generative ai." Tech. rep., National Bureau of Economic Research.

Dix, Rebekah, Kelsey Moran, and Thi Mai Anh Nguyen. 2025. "Costs of Technological Frictions: Evidence from EHR (Non-) Interoperability." Tech. rep., Working Paper.

Hoffmann, Manuel, Sam Boysel, Frank Nagle, Sida Peng, and Kevin Xu. 2024. "Generative AI and the Nature of Work." *Revise & Resubmit at Management Science* .

Hoffmann, Manuel, Sam Boysel, Frank Nagle, and Kevin Xu. 2025. "The Generative AI - Gender Puzzle." *Working Paper* .

Figure 1: AI Scribe: Abridge

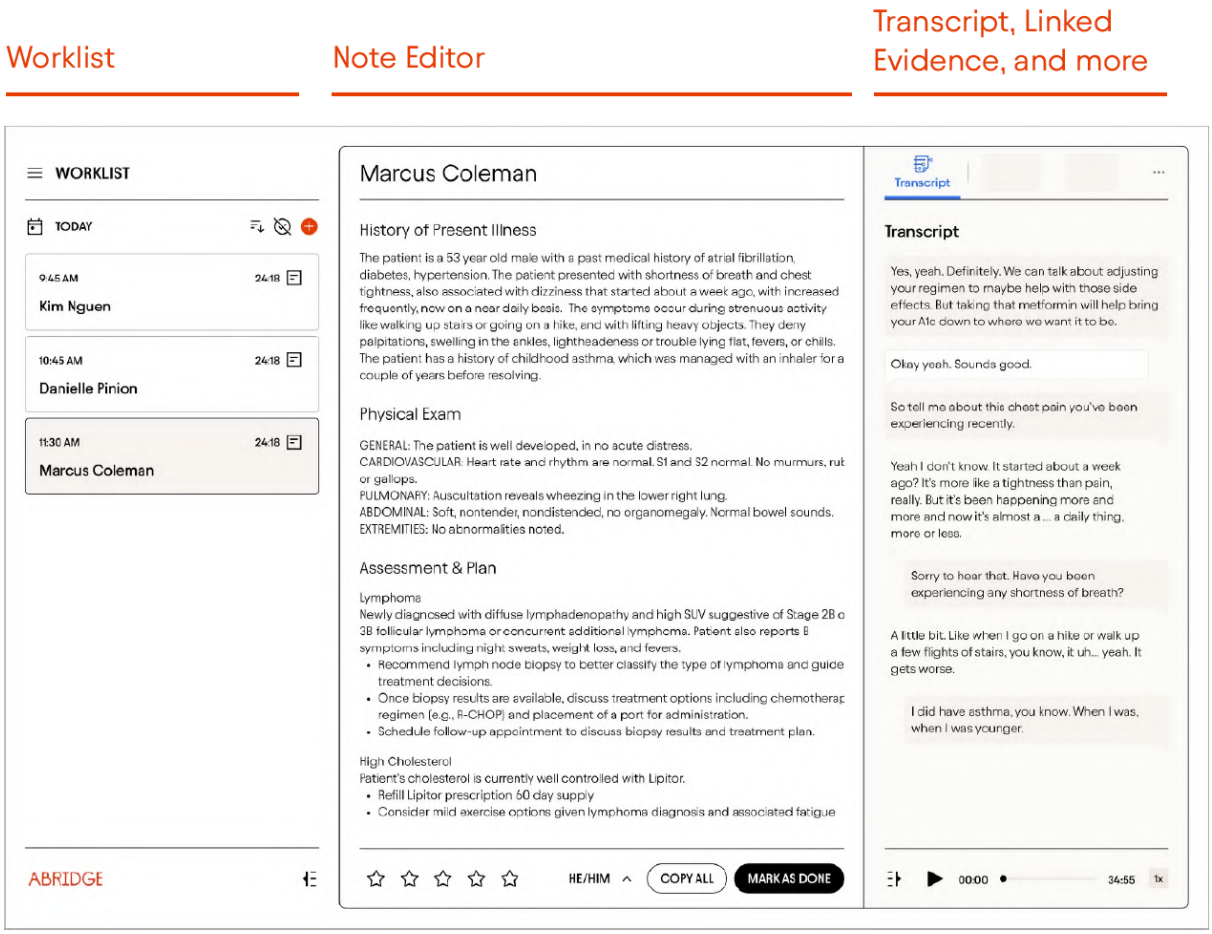
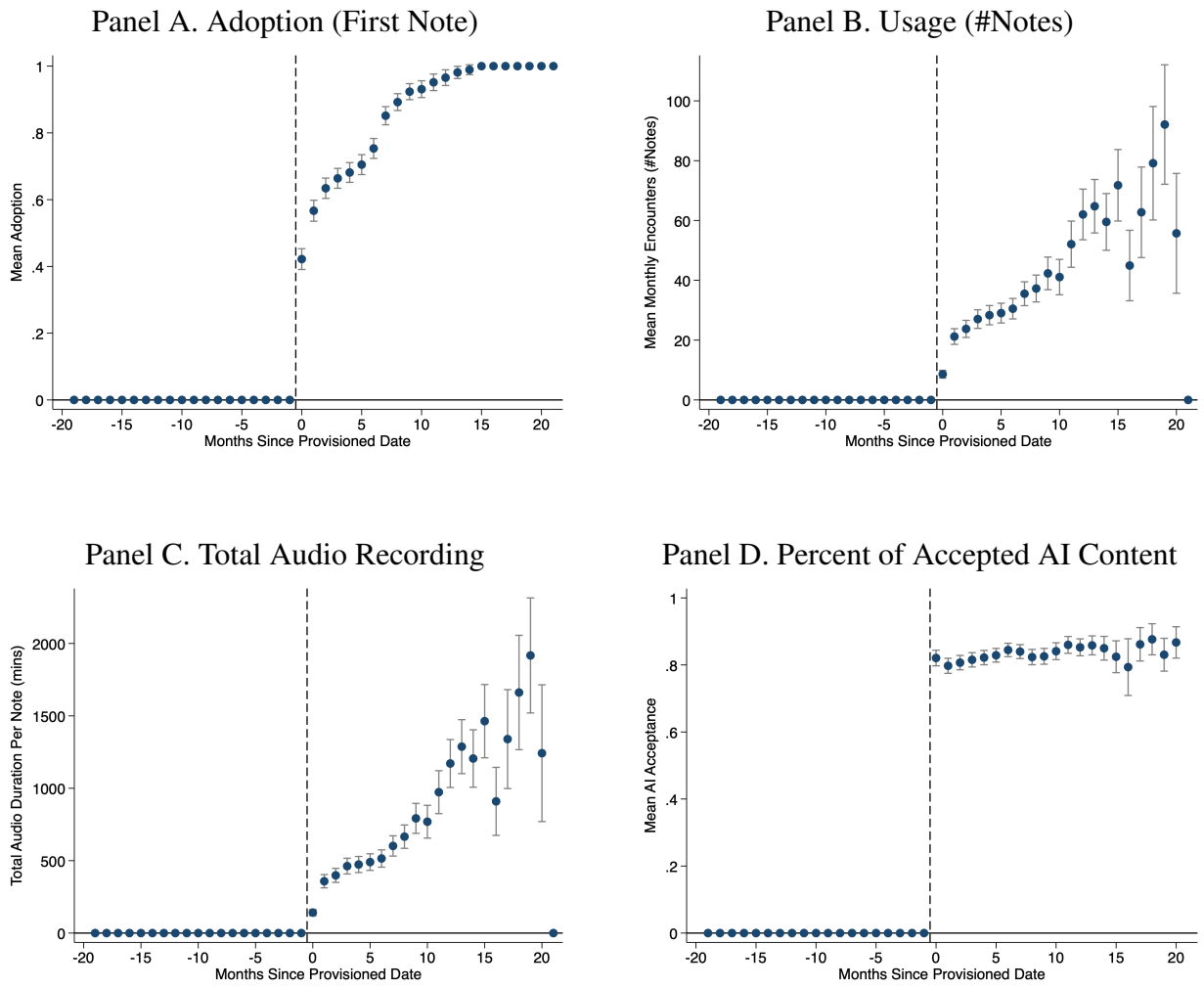


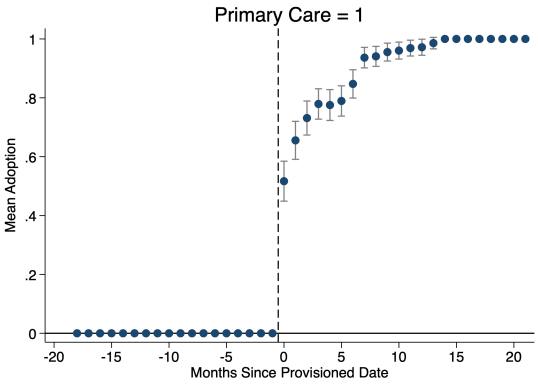
Figure 2: Staggered Rollout: AI Scribes Adoption and Usage



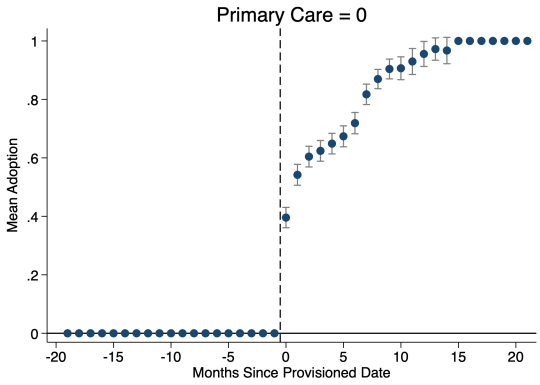
Notes: Panel A illustrates initial adoption of AI scribes, defined as the first AI-generated note following physician eligibility. Panel B reports the number of clinical notes generated using the AI scribe. Panel C displays total audio recording duration per physician. Panel D presents the percentage of AI-generated text accepted without modification, measured as the share of AI-produced content incorporated directly into the final clinical note. This metric serves as a proxy for documentation efficiency. The figures are based on a time period of May 2024 to Nov 2025. 95 percent confidence intervals are shown, with standard errors clustered at the physician level.

Figure 3: Dynamics for Primary Care vs. Non Primary Care

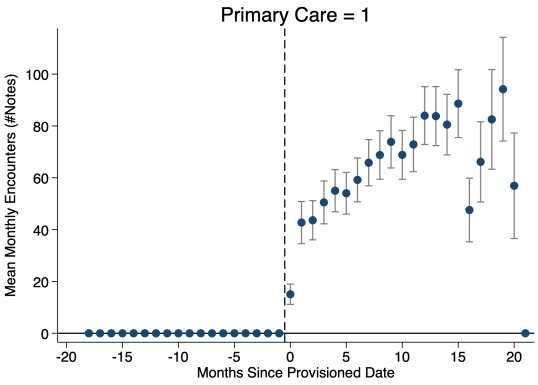
Panel A. Adoption – Primary Care



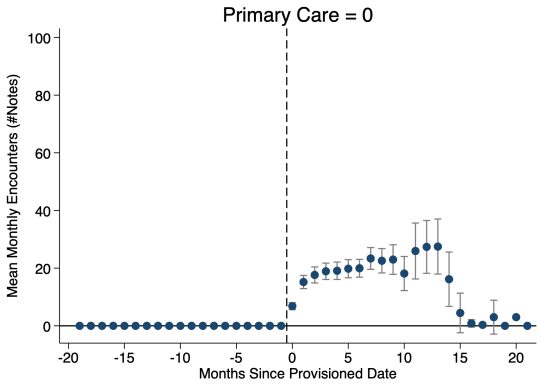
Panel B. Adoption – Non Primary Care



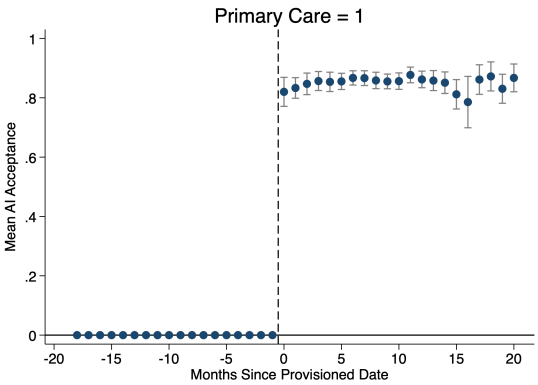
Panel C. Usage (#Notes) – Primary Care



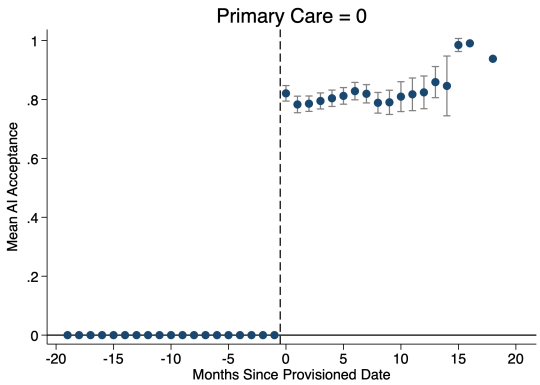
Panel D. Usage (#Notes) – Non Primary Care



Panel E. AI Acceptance – Primary Care



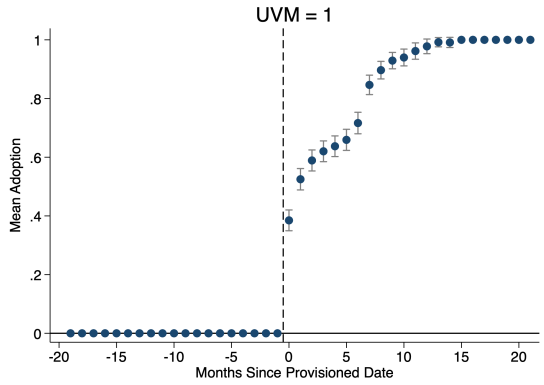
Panel F. AI Acceptance – Non Primary Care



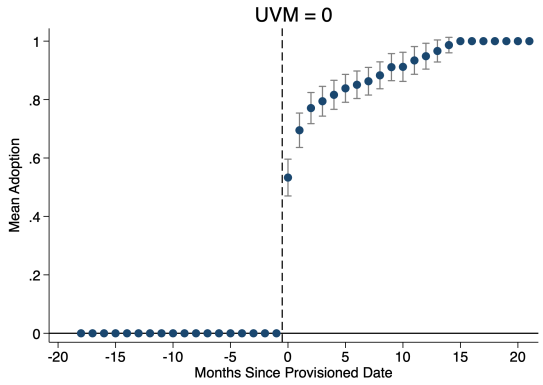
Notes: Panel A and B illustrate AI scribe adoption and usage for primary care and non-primary care physicians, respectively. Panels C and D report mean monthly physician-patient encounters. Panels E and F present the percentage of AI-generated content accepted without modification. The figures are based on a time period of May 2024 to Nov 2025. 95 percent confidence intervals are shown, with standard errors clustered at the physician level.

Figure 4: Dynamics for Headquarter vs. Satellites

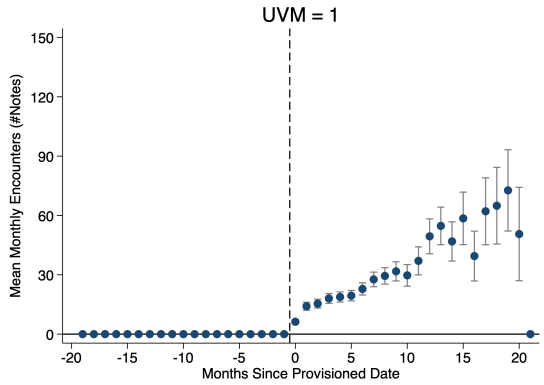
Panel A. Adoption – Headquarter



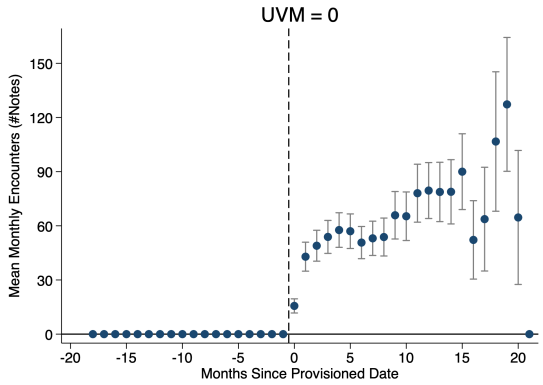
Panel B. Adoption – Satellites



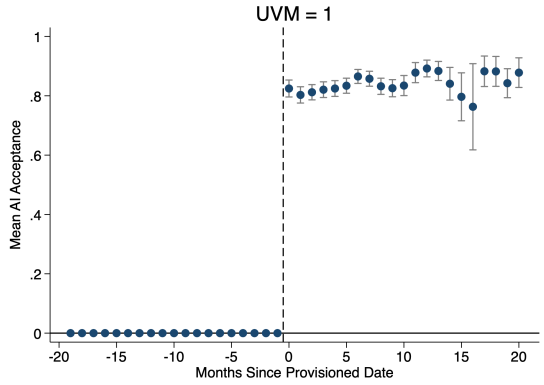
Panel C. Usage (#Notes) – Headquarter



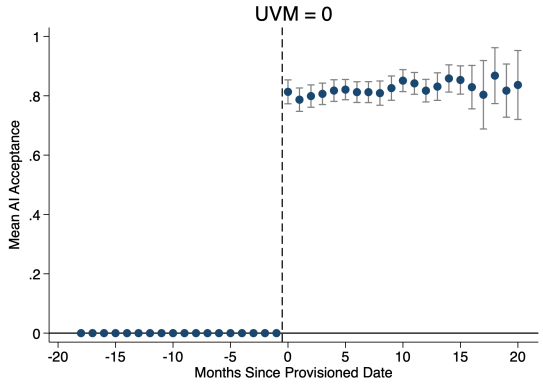
Panel D. Usage (#Notes) – Satellites



Panel E. AI Acceptance – Headquarter



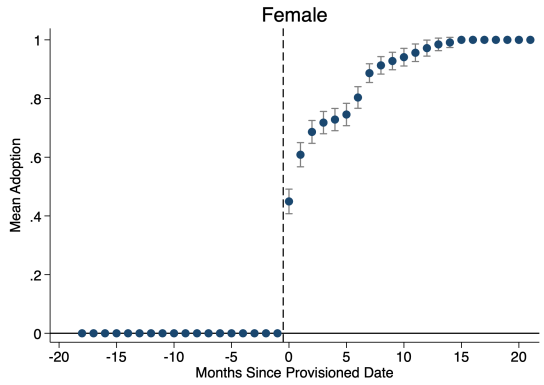
Panel F. AI Acceptance – Satellites



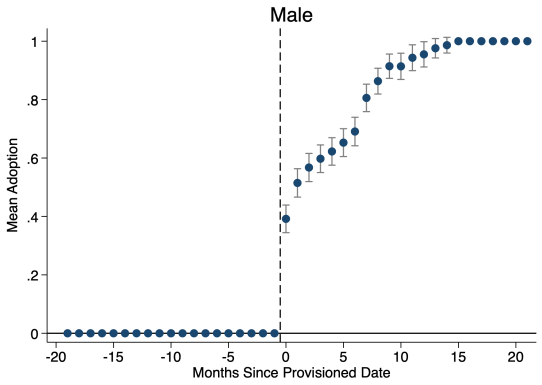
Notes: Panels A and B illustrate AI scribe adoption for headquarter and other locations, respectively. Panels C and D report mean monthly physician-patient encounters. Panels E and F present the percentage of AI-generated content accepted without modification. The figures are based on a time period of May 2024 to Nov 2025. 95 percent confidence intervals are shown, with standard errors clustered at the physician level.

Figure 5: Dynamics for Female vs. Male Physicians

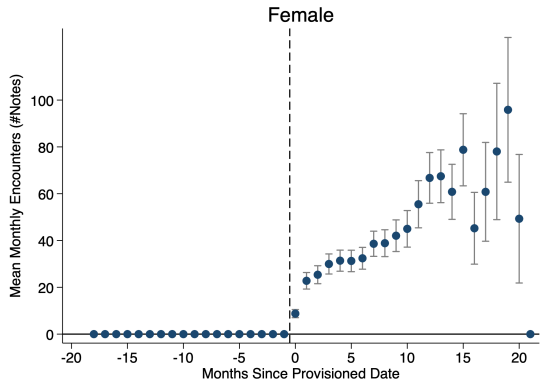
Panel A. Adoption – Female



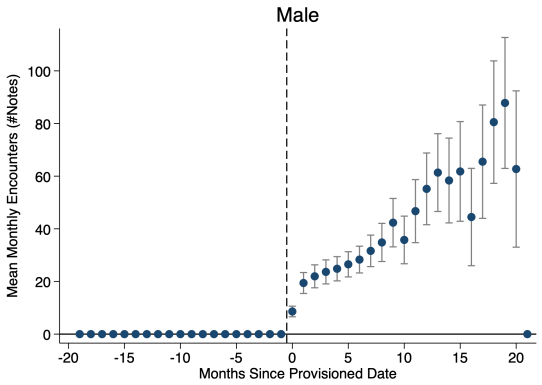
Panel B. Adoption – Male



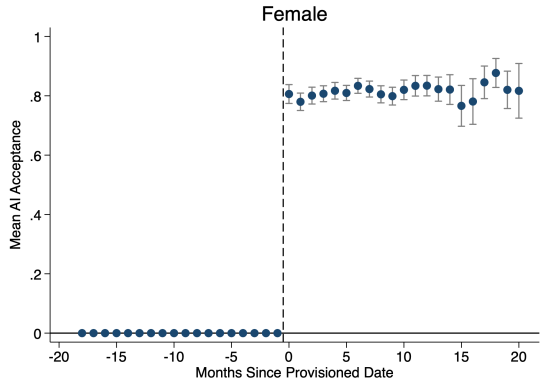
Panel C. Usage (#Notes) – Female



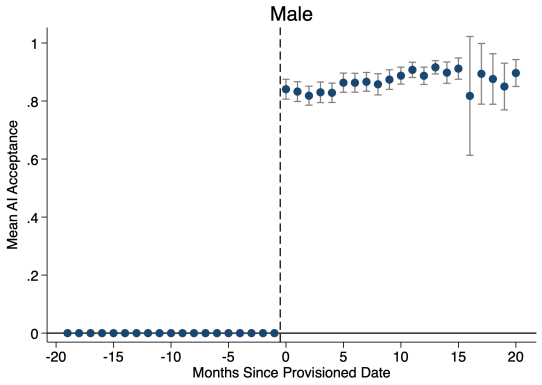
Panel D. Usage (#Notes) – Male



Panel E. AI Acceptance – Female



Panel F. AI Acceptance – Male



Notes: Panels A and B illustrate AI scribe adoption for female and male physicians, respectively. Panels C and D report mean monthly physician-patient encounters. Panels E and F present the percentage of AI-generated content accepted without modification. The figures are based on a time period of May 2024 to Nov 2025. 95 percent confidence intervals are shown, with standard errors clustered at the physician level.

Table 1: Descriptive Statistics

	Mean	SD	Min	Max
AI Adoption	0.73	0.44	0.00	1.00
Notes (Monthly)	31.95	52.83	0.00	420.00
Audio Duration (Min/Physician)	572.55	980.97	0.00	6614.12
Keep AI Notes	0.29	0.45	0.00	1.00
Accepted AI Content	0.83	0.16	0.00	1.00
Avg Note Rating	3.41	1.23	1.00	5.00
Physician-Patient Encounters	10,004			
Physician Observations	1,868			

Notes: Descriptive statistics are shown at the physician-month level. AI Adoption is an indicator equal to one if the physician used the AI scribe in a given month. Notes (Monthly) is the count of AI-scribed notes per physician per month. Audio Duration measures total recorded audio in minutes per physician per month. Keep AI Notes is an indicator for retaining the AI-generated note in the EHR system. Accepted AI Content is the share of AI-generated content accepted by the physician, conditional on use. Avg Note Rating is the share of notes rated positively by the physician.

Table 2: Heterogeneous Effect of Experience on Usage

	(1) Usage (#Notes)	(2) Avg Audio Duration	(3) Keep AI Notes	(4) AI Acceptance Rate	(5) Note Rating	(6) Avg Percent Effort Reduction
Exp. 2–5 Months	0.174***	2.185***	0.060***	-0.074***	-0.056***	-0.007
	(0.012)	(0.308)	(0.013)	(0.011)	(0.011)	(0.009)
Exp. 6–12 Months	0.215***	1.957***	0.063***	-0.161***	-0.104***	0.013
	(0.014)	(0.346)	(0.015)	(0.016)	(0.012)	(0.013)
Exp. over 12 Months	-0.249***	-4.352***	-0.149***	-0.741***	-0.154***	0.012
	(0.015)	(0.309)	(0.013)	(0.012)	(0.012)	(0.019)
Observations	35492	35492	35492	28404	35492	2807
R-squared	0.774	0.605	0.503	0.902	0.312	0.735

Notes: This table contains usage intensity, audio efficiency, and note quality metrics. All specifications include physician and year–month fixed effects. The table is based on a time period of May 2024 to Nov 2025. Standard errors clustered at the physician level are reported in parentheses.

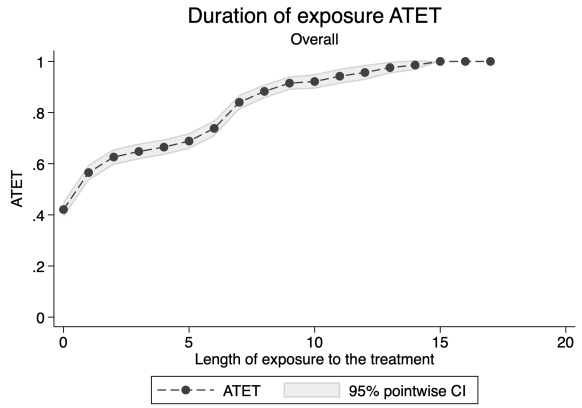
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A Appendix

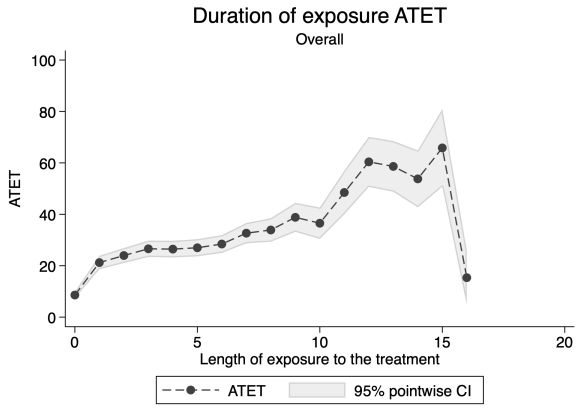
A.1 Figures

Figure A1: Staggered Rollout: AI Scribes Adoption and Usage

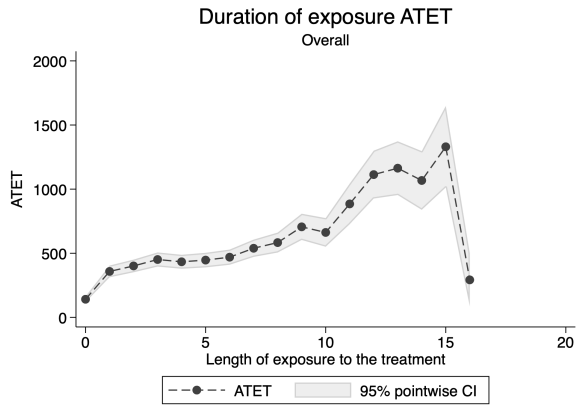
Panel A. Adoption (First Note)



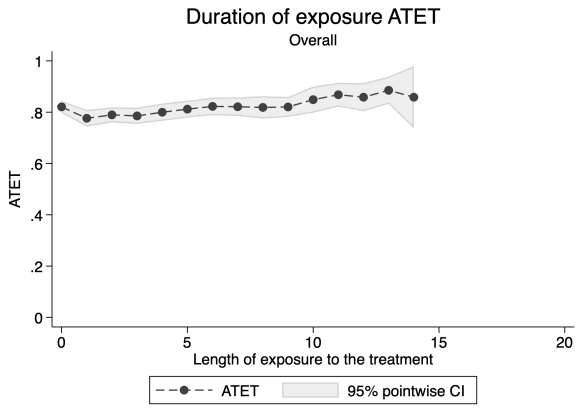
Panel B. Usage (#Notes)



Panel C. Total Audio Recording



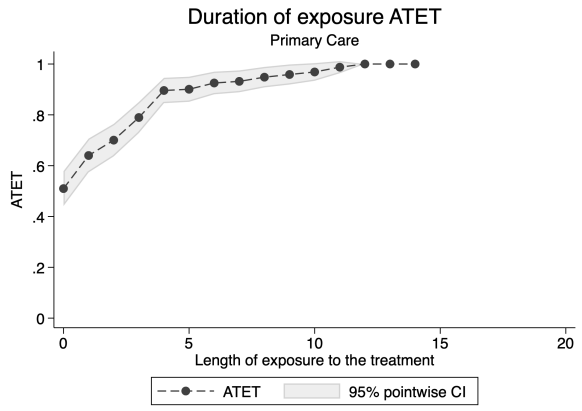
Panel D. Percent of Accepted AI Content



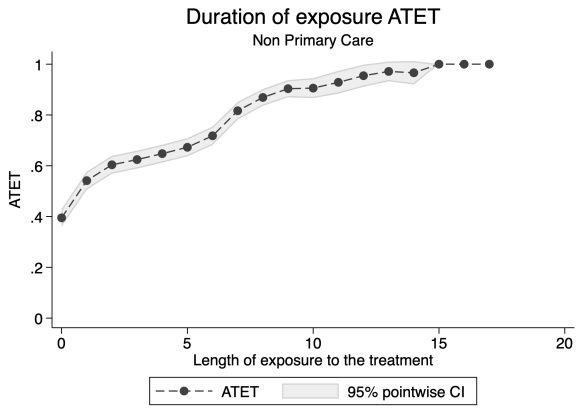
Notes: Panel A illustrates initial adoption of AI scribes, defined as the first AI-generated note following physician eligibility. Panel B reports the number of clinical notes generated using the AI scribe. Panel C displays total audio recording duration per physician. Panel D presents the percentage of AI-generated text accepted without modification, measured as the share of AI-produced content incorporated directly into the final clinical note. This metric serves as a proxy for documentation efficiency. The figures are based on a time period of May 2024 to Nov 2025. 95 percent confidence intervals are shown, with standard errors clustered at the physician level.

Figure A2: Dynamics for Primary Care vs. Non Primary Care

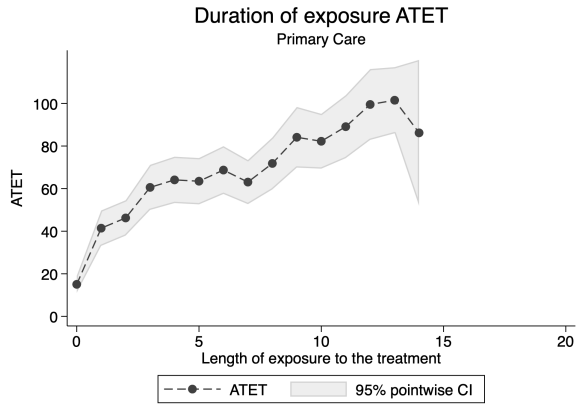
Panel A. Adoption – Primary Care



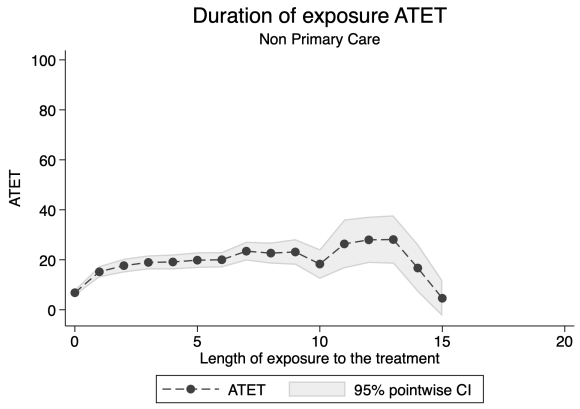
Panel B. Adoption – Non Primary Care



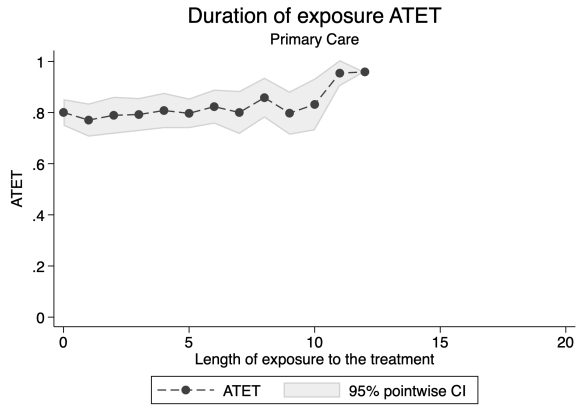
Panel C. Usage (#Notes) – Primary Care



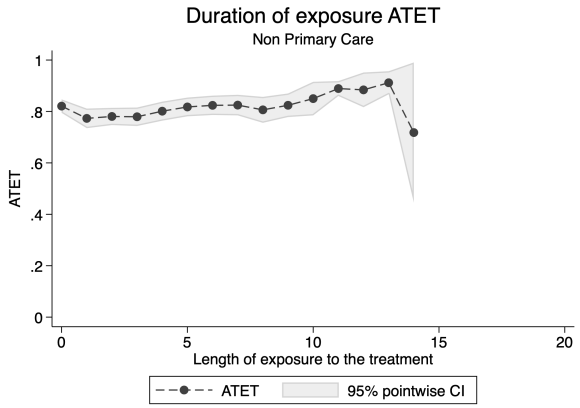
Panel D. Usage (#Notes) – Non Primary Care



Panel E. AI Acceptance – Primary Care



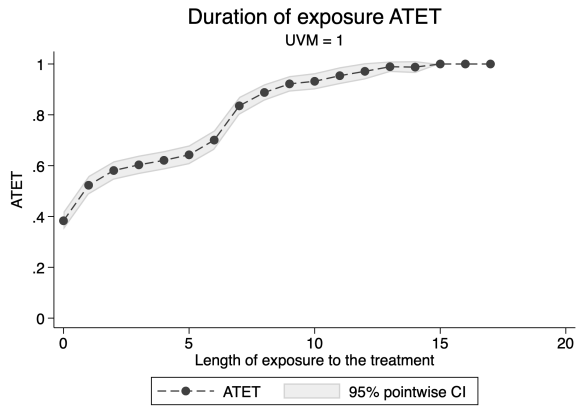
Panel F. AI Acceptance – Non Primary Care



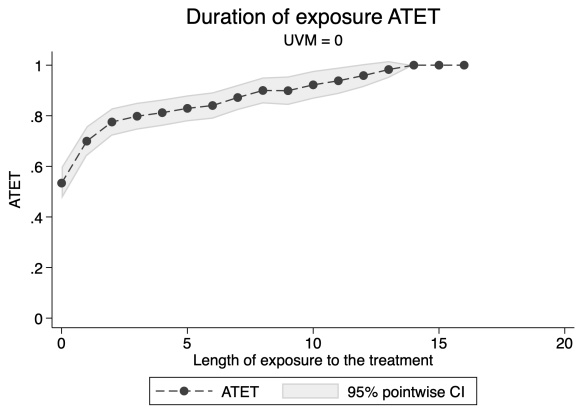
Notes: Panels A and B illustrate AI scribe adoption for primary care and non-primary care physicians, respectively. Panels C and D report mean monthly physician-patient encounters. Panels E and F present the percentage of AI-generated content accepted without modification. The figures are based on a time period of May 2024 to Nov 2025. 95 percent confidence intervals are shown, with standard errors clustered at the physician level.

Figure A3: Dynamics for Headquarter vs. Satellites

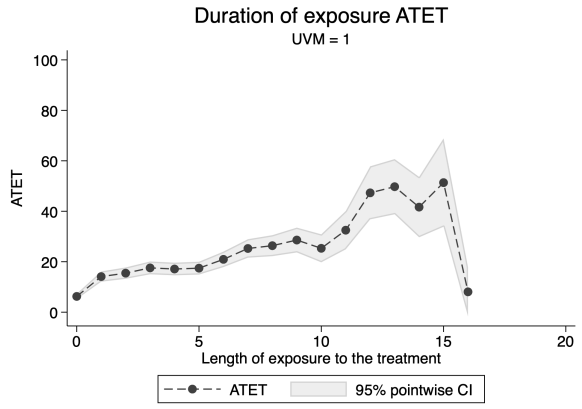
Panel A. Adoption – Headquarter



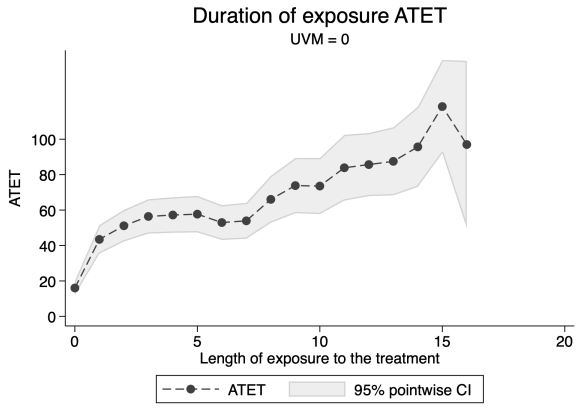
Panel B. Adoption – Satellites



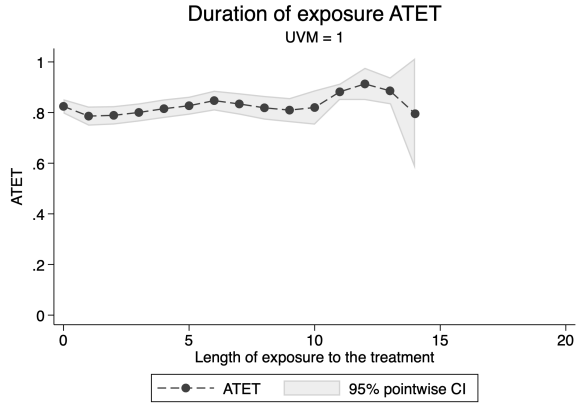
Panel C. Usage (#Notes) – Headquarter



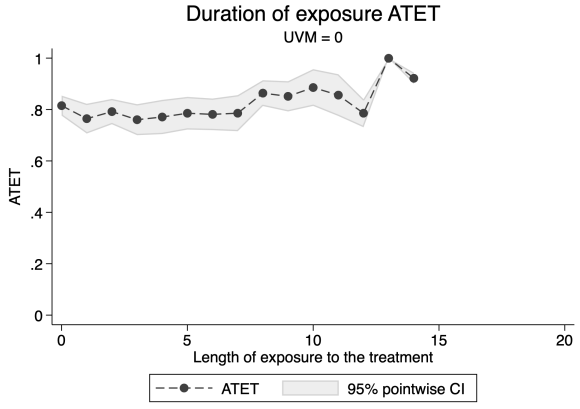
Panel D. Usage (#Notes) – Satellites



Panel E. AI Acceptance – Headquarter



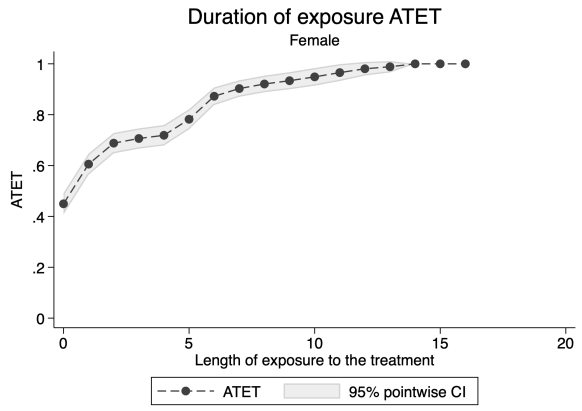
Panel F. AI Acceptance – Satellites



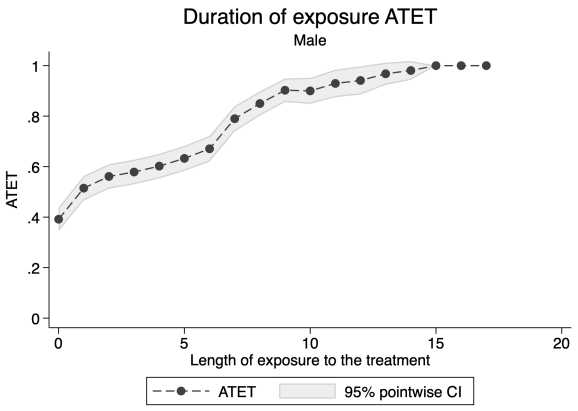
Notes: Panels A and B illustrate AI scribe adoption for headquarter and satellite locations, respectively. Panels C and D report mean monthly physician-patient encounters. Panels E and F present the percentage of AI-generated content accepted without modification. The figures are based on a time period of May 2024 to Nov 2025. 95 percent confidence intervals are shown, with standard errors clustered at the physician level.

Figure A4: Dynamics for Female vs. Male Physicians

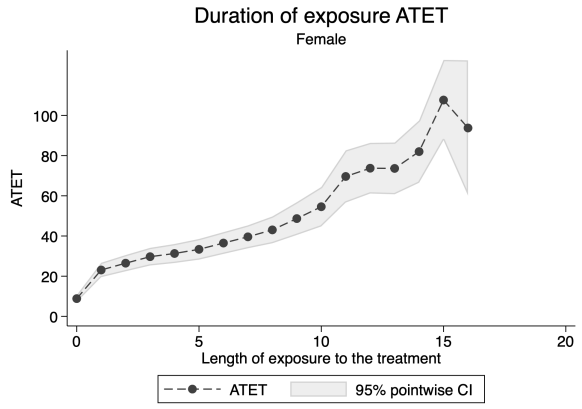
Panel A. Adoption – Female



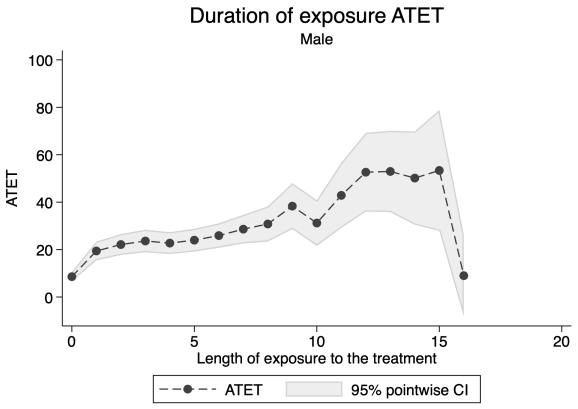
Panel B. Adoption – Male



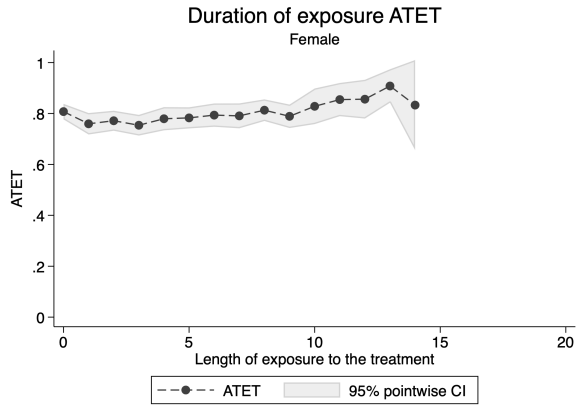
Panel C. Usage (#Notes) – Female



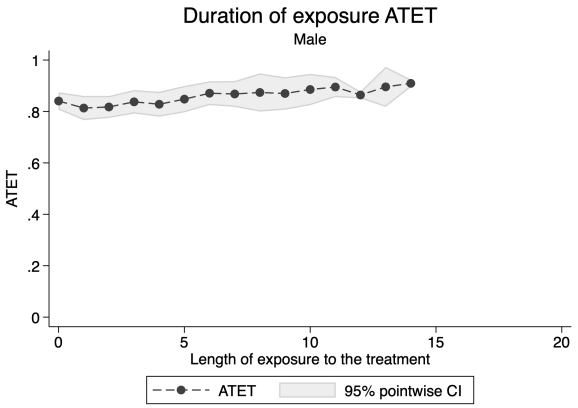
Panel D. Usage (#Notes) – Male



Panel E. AI Acceptance – Female



Panel F. AI Acceptance – Male



Notes: Panels A and B illustrate AI scribe adoption for female and male physicians, respectively. Panels C and D report mean monthly physician-patient encounters. Panels E and F present the percentage of AI-generated content accepted without modification. The figures are based on a time period of May 2024 to Nov 2025. 95 percent confidence intervals are shown, with standard errors clustered at the physician level.

A.2 Tables

Table A1: Descriptive Statistics: Specialty

	Primary Care				Non-Primary Care			
	Mean	SD	Min	Max	Mean	SD	Min	Max
AI Adoption	0.86	0.35	0.00	1.00	0.67	0.47	0.00	1.00
Notes (Monthly)	59.83	65.45	0.00	344.00	18.08	38.33	0.00	420.00
Audio Duration (Min/Physician)	1182.04	1317.14	0.00	6614.12	269.40	549.14	0.00	4596.59
Keep AI Notes	0.40	0.49	0.00	1.00	0.24	0.43	0.00	1.00
Accepted AI Content	0.85	0.14	0.00	1.00	0.81	0.18	0.01	1.00
Avg Note Rating	3.41	1.27	1.00	5.00	3.40	1.17	1.00	5.00
Observations	3323				6681			

Table A2: Descriptive Statistics: Location

	UVM				Non-UVM			
	Mean	SD	Min	Max	Mean	SD	Min	Max
AI Adoption	0.70	0.46	0.00	1.00	0.83	0.38	0.00	1.00
Notes (Monthly)	22.62	38.79	0.00	247.00	55.26	72.42	0.00	420.00
Audio Duration (Min/Physician)	440.56	817.95	0.00	6514.49	902.40	1242.30	0.00	6614.12
Keep AI Notes	0.25	0.43	0.00	1.00	0.40	0.49	0.00	1.00
Accepted AI Content	0.83	0.16	0.00	1.00	0.82	0.17	0.02	1.00
Avg Note Rating	3.35	1.16	1.00	5.00	3.51	1.31	1.00	5.00
Observations	7145				2859			

Table A3: Descriptive Statistics: Gender

	Female				Male			
	Mean	SD	Min	Max	Mean	SD	Min	Max
AI Adoption	0.77	0.42	0.00	1.00	0.68	0.46	0.00	1.00
Notes (Monthly)	34.40	54.07	0.00	420.00	29.01	51.01	0.00	410.00
Audio Duration (Min/Physician)	633.27	1004.71	0.00	6514.49	497.25	946.20	0.00	6614.12
Keep AI Notes	0.32	0.47	0.00	1.00	0.26	0.44	0.00	1.00
Accepted AI Content	0.81	0.17	0.01	1.00	0.86	0.15	0.00	1.00
Avg Note Rating	3.39	1.23	1.00	5.00	3.44	1.23	1.00	5.00
Observations	5736				4155			

Table A4: Main Effects

	(1) AI Adoption	(2) Avg Audio Duration	(3) Keep AI Notes	(4) AI Acceptance Rate	(5) Note Rating	(6) Avg Note Rating
Post	0.588***	8.078***	0.245***	0.807***	0.091***	0.000
	(0.014)	(0.296)	(0.011)	(0.009)	(0.006)	(.)
Observations	35492	35492	35492	28404	35492	2807
R-squared	0.813	0.616	0.503	0.974	0.294	0.733

Standard errors in parentheses

Individual and time fixed effects included.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Experience Effects - Primary Care vs Non Primary Care

	(1) Usage (#Notes)	(2) Avg Audio Duration	(3) Keep AI Notes	(4) AI Acceptance Rate	(5) Note Rating	(6) Avg Percent Effort Reduction
Panel A: Primary Care						
Exp. 2–5 Months	0.134***	2.019***	0.123***	-0.113***	-0.051*	0.005
	(0.025)	(0.697)	(0.032)	(0.025)	(0.027)	(0.017)
Exp. 6–12 Months	0.095***	1.435*	0.151***	-0.273***	-0.116***	0.009
	(0.029)	(0.822)	(0.032)	(0.029)	(0.028)	(0.021)
Exp. over 12 Months	-0.190***	-4.193***	-0.050*	-0.649***	-0.192***	0.007
	(0.031)	(0.731)	(0.028)	(0.026)	(0.027)	(0.026)
Observations	6365	6365	6365	4363	6365	1296
R-squared	0.829	0.706	0.555	0.904	0.360	0.703
Panel B: Non-Primary Care						
Exp. 2–5 Months	0.199***	2.326***	0.029**	-0.046***	-0.056***	-0.008
	(0.014)	(0.347)	(0.013)	(0.011)	(0.013)	(0.012)
Exp. 6–12 Months	0.291***	2.168***	0.015	-0.054***	-0.104***	0.012
	(0.016)	(0.385)	(0.016)	(0.015)	(0.013)	(0.018)
Exp. over 12 Months	-0.293***	-4.702***	-0.184***	-0.788***	-0.134***	0.004
	(0.018)	(0.342)	(0.015)	(0.013)	(0.012)	(0.034)
Observations	29127	29127	29127	24041	29127	1511
R-squared	0.744	0.516	0.457	0.924	0.249	0.751

Standard errors in parentheses

Individual and time fixed effects included.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Experience Effects - Headquarter vs Satellites

	(1) Usage (#Notes)	(2) Avg Audio Duration	(3) Keep AI Notes	(4) AI Acceptance Rate	(5) Note Rating	(6) Avg Percent Effort Reduction
Panel A: Headquarter						
Exp. 2–5 Months	0.156***	2.330***	0.055***	-0.075***	-0.047***	0.002
	(0.013)	(0.380)	(0.015)	(0.014)	(0.012)	(0.013)
Exp. 6–12 Months	0.192***	3.048***	0.079***	-0.204***	-0.081***	0.026
	(0.017)	(0.466)	(0.018)	(0.020)	(0.013)	(0.017)
Exp. over 12 Months	-0.189***	-3.747***	-0.143***	-0.745***	-0.156***	0.027
	(0.017)	(0.370)	(0.014)	(0.015)	(0.014)	(0.027)
Observations	15238	15238	15238	9867	15238	1691
R-squared	0.730	0.568	0.466	0.921	0.298	0.684
Panel B: Satellites						
Exp. 2–5 Months	0.191***	2.042***	0.107***	-0.099***	-0.060**	-0.019
	(0.026)	(0.520)	(0.032)	(0.019)	(0.027)	(0.013)
Exp. 6–12 Months	0.194***	0.669	0.071**	-0.182***	-0.128***	-0.008
	(0.029)	(0.594)	(0.034)	(0.027)	(0.026)	(0.019)
Exp. over 12 Months	-0.322***	-4.750***	-0.187***	-0.719***	-0.183***	-0.020
	(0.033)	(0.570)	(0.032)	(0.021)	(0.026)	(0.024)
Observations	20254	20254	20254	18537	20254	1116
R-squared	0.824	0.679	0.563	0.888	0.344	0.806

Standard errors in parentheses

Individual and time fixed effects included.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Experience Effects - Female vs Male Physicians

	(1) Usage (#Notes)	(2) Avg Audio Duration	(3) Keep AI Notes	(4) AI Acceptance Rate	(5) Note Rating	(6) Avg Percent Effort Reduction
Panel A: Female						
Exp. 2–5 Months	0.190***	3.046***	0.090***	-0.081***	-0.057***	-0.012
	(0.017)	(0.432)	(0.018)	(0.015)	(0.017)	(0.013)
Exp. 6–12 Months	0.217***	2.364***	0.088***	-0.166***	-0.116***	0.001
	(0.019)	(0.464)	(0.020)	(0.020)	(0.017)	(0.016)
Exp. over 12 Months	-0.260***	-4.624***	-0.143***	-0.725***	-0.172***	-0.002
	(0.021)	(0.425)	(0.018)	(0.016)	(0.016)	(0.023)
Observations	19171	19171	19171	15262	19171	1757
R-squared	0.788	0.615	0.516	0.899	0.332	0.719
Panel B: Male						
Exp. 2–5 Months	0.148***	1.037**	0.021	-0.060***	-0.060***	-0.002
	(0.018)	(0.436)	(0.019)	(0.017)	(0.016)	(0.013)
Exp. 6–12 Months	0.205***	1.297**	0.031	-0.150***	-0.093***	0.032
	(0.022)	(0.527)	(0.023)	(0.028)	(0.017)	(0.020)
Exp. over 12 Months	-0.241***	-4.059***	-0.159***	-0.764***	-0.136***	0.043
	(0.023)	(0.458)	(0.021)	(0.018)	(0.017)	(0.032)
Observations	15903	15903	15903	12829	15903	1045
R-squared	0.758	0.592	0.485	0.908	0.273	0.768

Standard errors in parentheses

Individual and time fixed effects included.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$