Evaluating Ambient AI Documentation in Ambulatory Practice: A 120-Day Mixed-Methods Pilot Study

Christiane Pimentel-Olatokunbo[1,2] BS, Min Jung ('MJ') Kim[1,3] PhD, R. Dirk Stanley[4] MD MPH, Thomas P. Agresta[1,2,5] MD MBI

[1]Department of Family Medicine University of Connecticut School of Medicine, [2]Center for Quantitative Medicine, UConn Health,
[3]Pat and Jim Calhoun Cardiology Center, UConn Health, [4]Chief Medical Information Officer, UConn Health, [5]Chair of Family Medicine, Trinity Health of New England

Introduction

Background: Physician documentation burden contributes to burnout and reduced patient interaction. Ambient AI tools like ***Nuance DAX CoPilot by Microsoft*** automatically transcribe clinician–patient conversations into structured notes, potentially reducing after-hours charting, enhancing workflow efficiency, and improving patient rapport.

Objective: Evaluate the effect of DAX CoPilot on documentation burden, perceived quality, efficiency, burnout (Mini Z 2.0), and Epic SIGNAL metrics among ambulatory physicians during a 120-day pilot.

Methods

Design: Prospective 120-day pilot with pre- (Day 0), mid- (Day 54), and endpoint (Day 117) surveys.

Participants: 30 volunteer ambulatory physicians across multiple specialties (27 completed both baseline and post surveys; 20 completed all three).

Measures:

- Documentation quality: Accuracy, Completeness, Efficiency, Overall Quality (Likert 1–5)
- Documentation burden (1 = High → 5 = None)
- Job satisfaction & work–life balance
- Burnout (Mini Z 2.0: Supportive Work Environment, Work Pace & EMR Stress, Total Score)
- Epic SIGNAL metrics:
  - Pajama Time (min per scheduled day), Active EMR Time (min per encounter)
  - Analysis: Wilcoxon rank sum tests for paired comparisons; effect sizes (Cohen's r); qualitative thematic analysis of open-ended responses.

Analysis

Qualitative Themes:

- Efficiency & Balance: Less after-hours charting; more time with patients (improved patient rapport).
- Accuracy/Style Gaps: Hallucinations, passive phrasing, missed exam details, repetitive A/P.
- Technical Reliability: Server timeouts, connectivity, authentication friction.
- Customization Needs: Desire for template learning, order pends, and MA-based consent workflows.
- Expansion Interest: 86% support use for learners, inbox triage, and verbal order pends.

Representative Quotes:

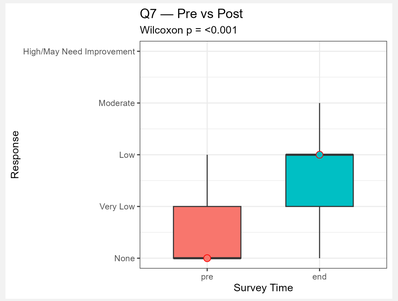
- "I no longer spend hours charting on weekends — DAX saves me ~20 minutes daily."
- "The A/P is not thorough enough; I still rely on my templates."

Results

Three surveys **(PRE-PILOT Ambient AI Documentation Survey, MIDPOINT-PILOT Ambient AI Documentation Survey, and ENDPOINT-PILOT Ambient AI Documentation Survey**) were disseminated to study participants prior to beginning the study (Day 0), at the midpoint of the study (Day 54), and at the end of the study (Day 117), respectively.

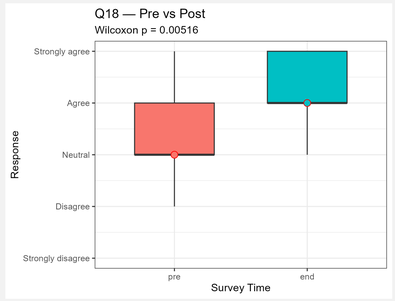
There were 29 documented responses for the **PRE-PILOT**, 22 for **the MIDPOINT-PILOT**, and 29 for the **ENDPOINT-PILOT**) from the 30 volunteer ambulatory providers across multiple specialties.

Current Burden of Clinical Documentation



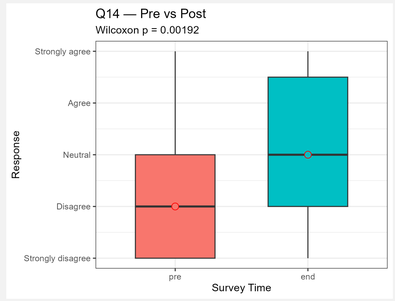
Clinical Documentation Burden improved significantly after DAX Copilot implementation (p<0.0001)

Overall Job Satisfaction



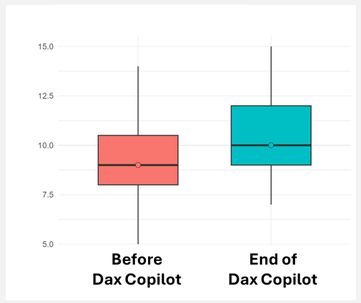
Overall Job Satisfaction improved significantly after DAX Copilot implementation (p = 0.005).

Time Spent on EMR at Home



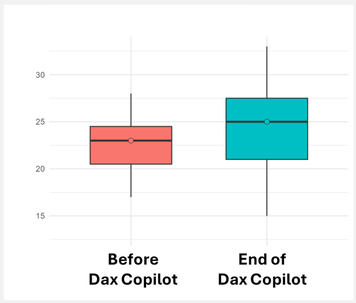
Time spent on EMR at home significantly decreased after DAX Copilot implementation (p = 0.002, large effect size = 0.61)

Subscale 2: Work Pace & EMR Stress



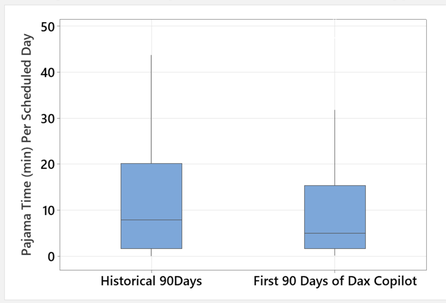
Subscale 2: Work Pace & EMR Stress Improved Significantly after DAX Copilot use (p = 0.005)

Total Mini-Z Score



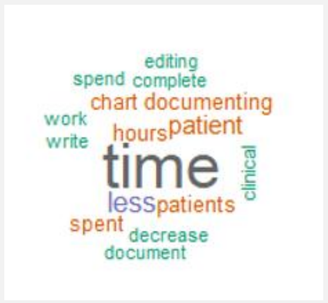
Total Mini-Z Score Significantly Increased after DAX Copilot use (p < 0.0001)

Pajama Time (Min Per Scheduled Day)



Pajama time (minutes spent on EMR after hours per scheduled day) significantly decreased with DAX Copilot use (p = 0.03).

Reported Biggest Advantages of Using AI Dictation in Clinical Practice



Before DAX Copilot Use

During DAX Copilot Use

After DAX Copilot Use

Common themes revolved around efficiency, more patient time, and improved balance (before, during, and after DAX Copilot use).

Discussions/Conclusions

Discussion:

Key Insights: Documentation burden fell sharply within 54 days, sustained at 117 days. 62% cited AI dictation as a major retention factor. Average time savings: ~20 minutes/day; reduced weekend "pajama time."

Impact: DAX CoPilot markedly reduced documentation burden and improved perceived efficiency within two months of deployment.

Challenges: Accuracy variability, integration gaps, and reliability issues must be addressed before wider rollout.

Operational ROI: Strong provider satisfaction and retention benefits justify continued evaluation and staged expansion.

Conclusions:

Ambient AI documentation ***(DAX CoPilot***) substantially improved clinician experience and efficiency in ambulatory settings. Technical refinements and specialty-specific optimization are essential to maximize value and safety.

Recommendations:

- Fix reliability issues (server, login)
- Integrate pended orders & templates
- Specialty checklists for accuracy
- Pilot expansion to learners and inbox triage
- Compare with Epic SIGNAL for objective validation

Next Steps for research:

- Follow clinician cohorts for 12–24 months to assess sustained well-being, job satisfaction, & intent to stay.
- Calculate cost savings from reduced burnout-related turnover, increased patient capacity, and lower documentation time.
- Evaluate patient comfort, perceived privacy, and willingness to consent to recording, and assess patient–provider rapport scores.
- Explore patient expectations, perceived empathy, and communication quality during AI-assisted visits.
- Evaluate impact of AI-augmented documentation on medical education and clinical supervision.

Acknowledgements

[*Self-funded research study*] UConn Health clinicians, Epic & IT teams, Microsoft/Nuance collaboration, and all pilot participants.

References

1. Owens, L. M., Wilda, J. J., Hahn, P. Y., Koehler, T., & Fletcher, J. J. (2024). The association between use of ambient voice technology documentation during primary care patient encounters, documentation burden, and provider burnout. Family practice, 41(2), 86–91. https://doi.org/10.1093/fampra/cmad092
2. Sloss, E. A., Abdul, S., Aboagyewah, M. A., Beebe, A., Kendle, K., Marshall, K., Rosenbloom, S. T., Rossetti, S., Grigg, A., Smith, K. D., & Mishuris, R. G. (2024). Toward Alleviating Clinician Documentation Burden: A Scoping Review of Burden Reduction Efforts. Applied clinical informatics, 15(3), 446–455. https://doi.org/10.1055/s-0044-1787007
3. Misurac, J., Knake, L. A., & Blum, J. M. (2025). The Effect of Ambient Artificial Intelligence Notes on Provider Burnout. Applied clinical informatics, 16(2), 252–258. https://doi.org/10.1055/a-2461-4576