ISSN: 2181-3337 | SCIENTISTS.UZ VOLUME 5 ISSUE 7 JULY 2026

AUTOMATIC SPEECH RECOGNITION AND CLINICAL DOCUMENTATION TIME IN TWO HEALTHCARE INSTITUTIONS IN BUKHARA, UZBEKISTAN: A WITHIN-CLINICIAN QUASI-EXPERIMENTAL STUDY

Teshayev Rajabbek
Westminster International University in Tashkent

DOI: 10.67620/sai.57d0012

ABSTRACT

Clinical documentation can consume a substantial proportion of physicians' working time and may limit the operational benefits of healthcare digitalisation. This study examined whether automatic speech recognition (ASR) reduced the time required to create medical documentation in two functioning healthcare institutions in Bukhara, Uzbekistan. A within-clinician quasi-experimental comparison was used: the same clinicians completed documentation by conventional keyboard entry and by ASR-supported voice input. The quantitative analysis covered more than 120 observed patient records. To account for differences in note length, documentation time was standardised as minutes required to produce 100 words. Mean time declined from 8.8 minutes with manual typing to 5.2 minutes with ASR, corresponding to a descriptive reduction of 3.6 minutes per 100 words, or 40.9%. The finding suggests that locally adapted speech input may reduce the labour-time cost of documentation in routine clinical practice. However, the available report did not include dispersion measures, confidence intervals, inferential statistics, transcription accuracy, correction time, or full ethical and technical details. The result should therefore be interpreted as preliminary real-world evidence rather than a definitive causal estimate. Future evaluation should combine time outcomes with note quality, clinically significant error rates, clinician workload, patient consent, privacy safeguards, and implementation costs.

ACCEPTED

02.08.26

Keywords: *automatic speech recognition; clinical documentation; electronic health records; workflow efficiency; digital health; Uzbekistan.*

CC BY

INTRODUCTION

Healthcare digitalisation is frequently presented as a route to improved continuity, accessibility, and organisational efficiency. In practice, however, the introduction of electronic health records (EHRs) can also shift a substantial clerical burden onto clinicians. Time-and-motion research has shown that physicians may spend as much or more time on EHR and desk work as on direct patient interaction, while event-log studies indicate that a considerable proportion of EHR activity continues after scheduled clinical hours (Arndt et al., 2017; Sinsky et al., 2016). Poor usability and intensive EHR interaction have also been associated with professional fatigue and burnout (Khairat et al., 2020; Melnick et al., 2020).

This tension is particularly important in health systems undergoing rapid digital transition. Uzbekistan has prioritised digital health development, but the practical value of digital reform depends on

whether new systems reduce rather than reproduce administrative friction. In facilities with limited specialist capacity, high patient volumes, and mixed paper–digital workflows, documentation time is not merely an individual inconvenience. It is an operational resource that affects scheduling stability, patient flow, clinician attention, and the feasibility of wider digital implementation.

Automatic speech recognition converts spoken language into text and may allow clinicians to enter information more quickly than by keyboard. Earlier clinical speech-recognition studies, however, have reported heterogeneous effects because performance depends on recognition accuracy, speaker adaptation, specialty vocabulary, background noise, correction requirements, and integration with the EHR (Hodgson & Coiera, 2016; Hodgson et al., 2019). More recent ambient documentation systems have renewed interest in voice-enabled workflows by reporting reductions in note-writing time and after-hours EHR work (Duggan et al., 2025; Stults et al., 2025; Tierney et al., 2024). These newer systems are technically different from conventional ASR, but they address the same managerial problem: the time cost of transforming clinical information into a usable medical record.

The linguistic context matters. Clinical ASR systems trained mainly on English-language data cannot be assumed to perform equally well with Uzbek speech, local accents, code-switching, abbreviations, drug names, and specialty-specific terminology. A system may transcribe ordinary language adequately while producing clinically important errors in diagnoses, dosages, anatomical terms, or medication names. Consequently, evaluation should consider both efficiency and documentation quality. Speed alone is insufficient if the generated text requires extensive editing or introduces clinically significant errors.

The present article focuses on a real-world comparison conducted as part of doctoral research in Bukhara. Its primary question is: how much did ASR-supported voice input reduce the time required to produce a standard quantity of medical text compared with manual keyboard entry? A secondary aim is to interpret the operational meaning of the observed difference while clearly distinguishing measured findings from projected managerial benefits.

2. Materials and Methods

2.1 Study design and setting

The study used a within-clinician quasi-experimental design in two functioning healthcare institutions in Bukhara, Uzbekistan. The same clinicians documented clinical information using conventional keyboard entry and ASR-supported voice input. This design reduced between-participant variation because each clinician served as his or her own practical comparator.

Full randomisation was not reported in the supplied manuscript. The design should therefore be regarded as quasi-experimental rather than a randomised trial. The available material also did not specify whether the order of manual and ASR conditions was counterbalanced. Order effects, learning, fatigue, and increasing familiarity with the ASR system may consequently have influenced performance.

2.2 Participants and observations

The quantitative analysis included more than 120 observed patient records. The exact number of participating clinicians, the exact record count, medical specialties, demographic characteristics, recruitment method, and the distribution of observations across the two institutions were not stated in the supplied manuscript. These details must be recovered from the dissertation and reported before journal submission.

Eligibility criteria should specify that participants were clinicians responsible for creating the relevant medical records and able to use both keyboard and voice input. Exclusion criteria should identify incomplete records, interrupted sessions, technical failures, training trials, and observations with missing timing or word-count data. Because these criteria were not included in the source manuscript, no exclusions have been assumed in this revision.

ISSN: 2181-3337 | SCIENTISTS.UZ

2.3 Intervention and comparison

The comparison condition was routine keyboard-based documentation. The intervention condition used ASR-supported voice entry to convert spoken medical narration into text. The supplied document did not identify the software, ASR model, application version, device, microphone, operating system, network requirements, vocabulary adaptation, or whether processing occurred locally or through a cloud service. These technical details are essential for reproducibility and data-governance assessment.

The final paper should also specify whether timing included dictation only or the complete workflow of dictation, recognition, correction, formatting, EHR transfer, and clinician approval. The present analysis can support only the reported measure of text-production time; it cannot establish end-to-end documentation efficiency if correction and verification were excluded.

2.4 Outcome measurement

The primary outcome was documentation time standardised to 100 words of medical text. Standardisation was appropriate because total completion time can be misleading when notes differ in length. For each observation, the preferred calculation is: time per 100 words = total documentation time in minutes divided by total word count, multiplied by 100.

The source manuscript reported mean values of 8.8 minutes per 100 words for manual typing and 5.2 minutes per 100 words for ASR. It did not report standard deviations, medians, interquartile ranges, clinician-level averages, confidence intervals, or the distribution of paired differences. Therefore, this revision presents the comparison descriptively and does not claim statistical significance.

2.5 Statistical analysis

The descriptive percentage reduction was calculated as $[(8.8 - 5.2) / 8.8] \times 100 = 40.9\%$. The phrase “up to 41%” has been replaced with “a descriptive mean reduction of 40.9%,” because the available values are group means rather than a reported maximum.

For a definitive analysis, record-level observations should not be treated as fully independent because multiple records were produced by the same clinicians and clinicians were nested within institutions. The preferred analysis is a linear mixed-effects model with documentation method as a fixed effect and random intercepts for clinician and institution. If distributional assumptions are not met, a log-transformed mixed model or a paired clinician-level nonparametric analysis may be appropriate. Results should include the mean paired difference, 95% confidence interval, effect size, test statistic, exact p value, and sensitivity analyses.

2.6 Ethics, consent, and data governance

The supplied manuscript did not include an ethics committee name, approval number, approval date, clinician-consent procedure, patient-consent procedure or waiver, de-identification process, audio-retention policy, or data-storage location. Because the study involved clinical documentation and potentially identifiable voice or health information, these omissions are critical and must be corrected from the original study records.

Before submission, the authors must state whether actual patient encounters were recorded, whether patient identifiers were captured, whether audio or text was transmitted to an external provider, how long recordings were retained, who had access, and how the final records were verified. Clinician responsibility for reviewing and approving the final note should be explicit.

3. Results

3.1 Documentation time

Across more than 120 observed patient records, the reported mean time required to produce 100 words of medical documentation was 8.8 minutes with manual keyboard entry and 5.2 minutes with ASR-supported voice input. The absolute descriptive difference was 3.6 minutes per 100 words. Relative to manual typing, this corresponds to a 40.9% reduction in mean documentation time.

ISSN: 2181-3337 | SCIENTISTS.UZ

The result indicates that voice input was faster under the observed conditions. It does not, by itself, demonstrate that ASR improved note accuracy, completeness, patient safety, total encounter duration, patient throughput, or clinician well-being. No inferential statistical conclusion is possible from the summary values alone.

Documentation method	Mean min/100 words	Absolute difference	Relative difference	Interpretation
Manual keyboard entry	8.8	Reference	Reference	Observed comparison
ASR-supported voice input	5.2	−3.6 min/100 words	−40.9%	Descriptive reduction

Table 1. Reported documentation-time comparison. Variability and inferential statistics were not available in the supplied manuscript.

3.2 Evidence not reported

The supplied study report did not provide recognition accuracy, word error rate, clinically significant error rate, number of corrections, correction time, note completeness, physician satisfaction, patient outcomes, or implementation cost. These outcomes should not be implied from the time result.

4. Discussion

The observed reduction is operationally important because documentation is repeated across many consultations. A saving of 3.6 minutes per 100 words could accumulate when clinicians produce multiple notes each day. Nevertheless, the study did not measure whether saved documentation time was converted into additional consultations, shorter queues, longer patient interaction, or reduced after-hours work. Such consequences should be presented as plausible hypotheses for future evaluation rather than as demonstrated outcomes.

The direction of the finding is compatible with international research showing that documentation support can reduce time spent on notes. Tierney et al. (2024) reported favourable clinician experiences with ambient documentation, while more recent evaluations have found reductions in EHR note time and documentation burden after implementation of ambient systems (Duggan et al., 2025; Stults et al., 2025). Human-scribe studies also suggest that reducing continuous keyboard entry can improve physician workflow and, in some settings, productivity or patient throughput (Mishra et al., 2018; Walker et al., 2019).

Direct numerical comparison is inappropriate because these studies evaluated different technologies, specialties, healthcare systems, and outcomes. Ambient systems may listen to an entire encounter and generate structured notes, whereas conventional ASR usually transcribes explicit dictation. Human scribes perform additional coordination tasks that speech-recognition software may not reproduce (Tran et al., 2020). The Uzbek study therefore contributes context-specific preliminary evidence rather than proof that all speech-enabled systems yield a 40.9% benefit.

A major unresolved issue is the trade-off between speed and quality. Clinical ASR performance cannot be judged only by general word accuracy because errors in medications, measurements, negation, or diagnoses may be clinically consequential. Clinically aware evaluation frameworks have been proposed to better reflect the medical importance of transcription errors (Shor et al., 2023). Future Uzbek-language evaluation should report general word error rate together with a clinician-rated measure of clinically significant errors and the time required for correction.

Language adaptation is central to implementation. Uzbek medical communication may include Russian terminology, international drug names, abbreviations, and code-switching. A specialised lexicon should therefore be developed and tested by specialty. However, a lexicon alone will not solve workflow problems. The system must also integrate with EHR templates, authentication, data fields, and approval

ISSN: 2181-3337 | SCIENTISTS.UZ

procedures. Otherwise, time saved during dictation may be lost through copying, formatting, or duplicate entry.

Governance is equally important. Speech-enabled documentation can involve sensitive health information and biometric voice data. Institutions should establish patient-notification or consent procedures, data-minimisation rules, secure transmission and storage, retention limits, audit trails, and contractual restrictions on secondary data use. Final responsibility for the clinical record must remain with the clinician. Human review is necessary because even high-performing systems may produce omissions, substitutions, or plausible but incorrect text.

The findings should also be interpreted within the study's methodological limits. The number and characteristics of clinicians were not reported, the sample was drawn from only two institutions, the condition order was not described, and no measure of statistical uncertainty was available. Participants may have differed in typing skill, speaking rate, digital literacy, specialty, and familiarity with the system. A novelty or training effect may also have influenced the observed times. These factors limit causal inference and national generalisability.

5. Practical Implications

A phased pilot strategy is preferable to immediate large-scale deployment. Pilot sites should use predefined success criteria covering total documentation time, correction time, clinically significant errors, note completeness, clinician acceptance, patient consent, technical reliability, and cost per completed note.

Implementation should begin in specialties with relatively structured and repetitive documentation, but performance must be evaluated separately for each specialty. The ASR model should be adapted to Uzbek medical vocabulary and tested under realistic background-noise conditions.

The technology should operate within a clinician-in-the-loop workflow. ASR-generated text should remain a draft until reviewed and approved by the responsible clinician. Institutions should document responsibility for errors and maintain an audit trail of edits and approvals.

Integration with the EHR is necessary. A system that saves time during dictation but requires manual transfer, reformatting, or duplicate entry may not reduce total workload. End-to-end time, rather than dictation time alone, should be the primary implementation outcome.

6. Limitations

This revision is limited by the information contained in the supplied manuscript. It was not possible to verify the exact number of clinicians or records, institutional identities, study dates, participant characteristics, technical configuration, condition order, ethics approval, consent process, or record-level data.

The reported result is based on two mean values without measures of variability. Statistical significance, confidence intervals, and effect size cannot be reconstructed without the original dataset. The study also did not report accuracy, correction burden, note quality, patient outcomes, or cost.

The findings are drawn from two institutions in Bukhara and should not be generalised to all Uzbek healthcare facilities, specialties, languages, or digital infrastructures. Replication with prespecified protocols, larger multisite samples, and complete reporting is required.

7. Conclusion

In two functioning healthcare institutions in Bukhara, ASR-supported voice input was associated with a lower reported mean time for producing medical text than manual keyboard entry. Time per 100 words declined from 8.8 to 5.2 minutes, an absolute descriptive difference of 3.6 minutes and a relative reduction of 40.9%. This finding supports further evaluation of locally adapted ASR as a potential workflow tool in Uzbekistan. It should not yet be interpreted as proof of improved patient throughput, care quality, cost-effectiveness, or national scalability. A submission-ready evaluation requires complete participant and technical reporting, ethics and data-governance documentation, record-level statistical analysis, and simultaneous assessment of speed, correction burden, and clinical accuracy.

ISSN: 2181-3337 | SCIENTISTS.UZ

Declarations

Ethics approval and consent to participate: [INSERT the approving committee, approval number and date, clinician consent procedure, and patient consent or waiver from the original study records.]

Consent for publication: [INSERT applicable statement.]

Data availability: The record-level dataset and analysis materials were not included with the supplied manuscript. The authors should state where de-identified data and code can be accessed, or provide a justified restriction.

Competing interests: The author should declare any financial, institutional, or technical relationship with the ASR software provider.

Funding: [INSERT funding source or state that the study received no specific funding.]

Author contributions: Rajabbek Kaxramonovich Teshayev: conceptualisation, investigation, analysis, and manuscript preparation. Revise this statement if additional contributors qualify for authorship.

REFERENCES

1. Arndt, B. G., Beasley, J. W., Watkinson, M. D., Temte, J. L., Tuan, W.-J., Sinsky, C. A., & Gilchrist, V. J. (2017). Tethered to the EHR: Primary care physician workload assessment using EHR event log data and time-motion observations. Annals of Family Medicine, 15(5), 419–426. https://doi.org/10.1370/afm.2121
2. Duggan, M. J., Gervase, J., Schoenbaum, A., et al. (2025). Clinician experiences with ambient scribe technology to assist with documentation burden and efficiency. JAMA Network Open, 8(2), e2460637. https://doi.org/10.1001/jamanetworkopen.2024.60637
3. Hodgson, T., & Coiera, E. (2016). Risks and benefits of speech recognition for clinical documentation: A systematic review. Journal of the American Medical Informatics Association, 23(e1), e169–e179. https://doi.org/10.1093/jamia/ocv152
4. Hodgson, T., Magrabi, F., & Coiera, E. (2019). Speech recognition for clinical documentation from 1990 to 2018: A systematic review. Journal of the American Medical Informatics Association, 26(4), 324–338. https://doi.org/10.1093/jamia/ocy179
5. Khairat, S., Coleman, C., Ottmar, P., Jayachander, D. I., Bice, T., & Carson, S. S. (2020). Association of electronic health record use with physician fatigue and efficiency. JAMA Network Open, 3(6), e207385. https://doi.org/10.1001/jamanetworkopen.2020.7385
6. Melnick, E. R., Dyrbye, L. N., Sinsky, C. A., Trockel, M., West, C. P., Nedelec, L., Tutty, M. A., & Shanafelt, T. (2020). The association between perceived electronic health record usability and professional burnout among US physicians. Mayo Clinic Proceedings, 95(3), 476–487. https://doi.org/10.1016/j.mayocp.2019.09.024
7. Mishra, P., Kiang, J. C., & Grant, R. W. (2018). Association of medical scribes in primary care with physician workflow and patient experience. JAMA Internal Medicine, 178(11), 1467–1472. https://doi.org/10.1001/jamainternmed.2018.3956
8. Pearlman, K., Wan, W., Shah, S., & Laiteerapong, N. (2025). Use of an AI scribe and electronic health record efficiency. JAMA Network Open, 8(10), e2537000. https://doi.org/10.1001/jamanetworkopen.2025.37000
9. Shor, J., Bi, R. A., Venugopalan, S., Ibara, S., Goldenberg, R., & Rivlin, E. (2023). Clinical BERTScore: An improved measure of automatic speech recognition performance in clinical settings. arXiv. https://doi.org/10.48550/arXiv.2303.05737

ISSN: 2181-3337 | SCIENTISTS.UZ

10. Sinsky, C., Colligan, L., Li, L., Prgomet, M., Reynolds, S., Goeders, L., Westbrook, J., Tutty, M., & Blike, G. (2016). Allocation of physician time in ambulatory practice: A time and motion study in 4 specialties. Annals of Internal Medicine, 165(11), 753–760. https://doi.org/10.7326/M16-0961
11. Stults, C. D., Deng, S., Martinez, M. C., et al. (2025). Evaluation of an ambient artificial intelligence documentation platform for clinicians. JAMA Network Open, 8(5), e258614. https://doi.org/10.1001/jamanetworkopen.2025.8614
12. Teshayev, R. K. (2025). Improving healthcare management tools in Uzbekistan based on foreign experience [Doctoral dissertation, Westminster International University in Tashkent]. [VERIFY the exact registered dissertation title and repository record.]
13. Tierney, A. A., Gayre, G., Hoberman, B., Mattern, B., Ballesca, M., Kipnis, P., Liu, V., & Lee, K. (2024). Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. NEJM Catalyst Innovations in Care Delivery, 5(3). https://doi.org/10.1056/CAT.23.0404
14. Tran, B. D., Chen, Y., Liu, S., & Zheng, K. (2020). How does medical scribes' work inform development of speech-based clinical documentation technologies? A systematic review. Journal of the American Medical Informatics Association, 27(5), 808–817. https://doi.org/10.1093/jamia/ocaa020
15. Walker, K., Ben-Meir, M., Dunlop, W., Rosler, R., West, A., O'Connor, G., Chan, T., Badcock, D., Putland, M., & Hansen, K. (2019). Impact of scribes on emergency medicine doctors' productivity and patient throughput: Multicentre randomised trial. BMJ, 364, l121. https://doi.org/10.1136/bmj.l121

ISSN: 2181-3337 | SCIENTISTS.UZ