The impact of using AI-powered Voice-to-text Technology for clinical documentation on the quality of care – Systematic Review

Ahmed Alboksmaty
NIHR North West London PSRC

Authors: Ahmed Alboksmaty, Reham Aldakhil, Hutan Ashrafian, Ara Darzi, Ana-Luisa Neves

Background: AI-powered Voice-to-Text Technology (AIVT) has emerged as a promising and evolving tool for reducing the documentation workload during medical consultations. AIVT has the potential to improve the overall medical consultation experience by reducing the need for healthcare providers to focus on screens for documentation, enabling more interaction with patients. This systematic review investigates the influence of AIVT on healthcare quality in primary care and outpatient settings, focusing on seven dimensions of quality: effectiveness, efficiency, safety, patient-centeredness, timeliness, equity, and integration.

Research question: What is the impact of using AIVT for documentation on the quality of care provision during medical consultations in primary care and outpatient settings?

Methods: A comprehensive search was conducted across five databases—Medline, Embase, Global Health, CINHAL, and Scopus—to identify studies published up to September 20, 2024. The inclusion criteria required studies to assess the use of AIVT for medical documentation during consultations in primary care or outpatient settings, compare it to manual or non-AI documentation methods, and report outcomes relevant to healthcare quality. Interrater agreement was measured using Cohen's κ scores. Data related to the seven quality dimensions were extracted, and study quality was evaluated using the Mixed Methods Appraisal Tool. A narrative synthesis approach was employed for analysis.

Results: Of the 1,924 studies identified, nine met the inclusion criteria, comprising data from 524 healthcare professionals, 616 patients, and 1,069 consultations. Most studies (n=7) were conducted in the United States, with two other papers originating from Bangladesh and the Philippines. Improvements were consistently reported in the domains of effectiveness, patient-centeredness, and efficiency, as all studies discussing these dimensions (n=9, 6, and 5, respectively) highlighted benefits such as smoother documentation, reduced administrative burdens, and improved patient-provider interactions. However, findings related to safety were inconclusive; while three of six studies identified risks, including transcription errors, others reported reliable performance. Additionally, four studies emphasized the enhanced timeliness of care enabled by the seamless integration of AIVT with Electronic Health Records. Equity concerns were highlighted in three studies, primarily due to limited participant diversity and reliance on simulation settings rather than real practice.

Conclusion: AIVT shows a potential to enhance documentation efficiency and foster patient-centred care. However, unresolved issues, such as transcription inaccuracies, safety risks, and limited generalisability, reflect the need for further testing in real-world, large-scale contexts. Collaborative efforts among policymakers, researchers, and clinicians are essential to address these challenges and ensure the safe, equitable, and effective adoption of AIVT in healthcare settings.

NIHR | National Institute for Health and Care Research
SafetyNet | NIHR Patient Safety Research Collaboration (PSRC) Network

The impact of using AI-powered voice-to-text technology for clinical documentation on quality of care in primary care and outpatient settings: a systematic review

Aboksmaty A, Aldakhil R, Hayhoe B, Ashrafian H, Darzi A, Neves AL.

Background & Rationale

AI-powered Voice-to-text Technology (AIVT) offers a promising solution to reduce clinicians' burden of documentation during medical consultations, allowing them to devote more time and attention to patient interaction (Figure 1).

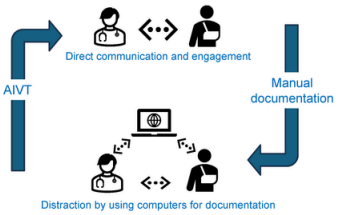

Figure 1: Enhancing patient-centred consultations with AIVT for documentation.

Aims & Objectives

This review assesses the impact of AIVT on care quality in primary care and outpatient settings. Quality is assessed across seven domains (Figure 2), including six from the Institute of Medicine[1] framework and an additional integration domain defined by the WHO and existing literature.

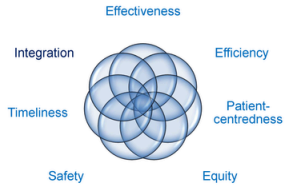

Figure 2: Quality domains considered for the analysis of the systematic review

Methods

A comprehensive search of five databases (Medline, Embase, Global Health, CINHAL, and Scopus) was conducted for studies published up to September 20, 2024 (Figure 3). Eligible studies focused on AIVT for medical documentation in primary care or outpatient settings and reported outcomes related to the quality domains. Cohen's κ scores assessed inter-rater agreement, and methodological quality was evaluated using the Mixed Methods Appraisal Tool.

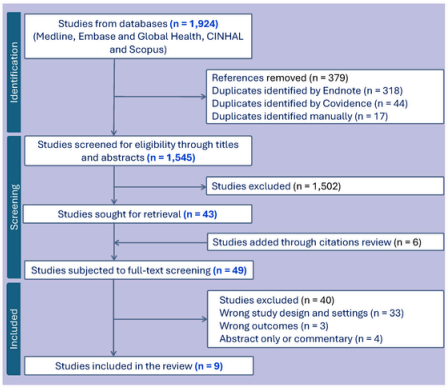

Figure 3: PRISMA flow chart clarifying the screening process and outcomes

Results

The initial search identified 1,924 studies, with nine meeting the inclusion criteria. Cohen's κ scores indicated substantial inter-rater agreement. Most studies (n=7) were from the USA, with one each from Bangladesh and the Philippines (2018–2024). The studies collectively involved 524 healthcare professionals and 616 patients, evaluating 1,069 medical consultations. Studies were conducted in primary care (n=5), outpatient settings (n=2), and both (n=2), with four in simulated environments. Table 1 summarises the key findings of the AIVT impact on the quality domains.

Table 1: Key findings of the AIVT impact on quality domains

Effectiveness	➢ Improved documentation quality, reducing cognitive burden and burnout. ➢ Captured key history points with comparable accuracy to, or even occasionally better than manual transcription. ➢ Inaccuracies noted, with 19.6% being clinically significant[2]. ➢ Performance metrics (e.g., WER, F1 scores) showed mixed results, indicating room for improvement.
Efficiency	➢ Increased documentation speed compared to typing and dictation. ➢ AIVT tools were 2.7× faster for history-taking and over 2× faster for physical examination documentation[3]. ➢ Efficiency improved with system familiarity and more practice ➢ Concerns about additional time spent reviewing and correcting errors.
Safety	➢ Potential improvements in clarity of prescriptions and medical instructions. ➢ Transcription inaccuracies, especially with medication names, posed risks. ➢ Error rates and clinical implications were frequently noted as concerns.
Patient-centredness	➢ Facilitated a more personalised and engaging consultation experience. ➢ Improved patient-provider interactions and communication. ➢ Ethical concerns, requiring informed consent for AIVT use.
Equity	➢ Limited generalisability due to participant homogeneity and simulations. ➢ Exclusion of certain groups (e.g., paediatrics, psychiatry patients, and non-native English speakers). ➢ Factors like patient volume and provider tech familiarity influenced outcomes.
Timeliness	➢ Improved documentation completion rates within 24 hours in AIVT users[4]. ➢ Potential to streamline documentation processes and transfers.
Integration	➢ Most studies reported seamless integration with EHR systems. ➢ Some studies explicitly reported successful integration, while others didn't specify.

Conclusion

Improvement
- Effectiveness
- Efficiency
- Patient-Centredness

! Concerns
- Safety
- Equity

? Limited Evidence
- Timeliness
- Integration

References
(1) Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century Washington (DC): National Academies Press (US); 2001 [Available from: https://www.ncbi.nlm.nih.gov/books/NBK222274/
(2) Goss FR, Blackley SV, Ortega CA, Kowalski LT, Landman AB, Lin CT, et al. A clinician survey of using speech recognition for clinical documentation in the electronic health record. Int J Med Inform. 2019;130:103938.
(3) Wang J, Lavender M, Hoque E, Brophy P, Kautz H. A patient-centered digital scribe for automatic medical documentation. JAMIA Open. 2021;4(1):ooab003.
(4) Haberle T, Cleveland C, Snow GL, Barber C, Stookey N, Thornock C, et al. The impact of nuance DAX ambient listening AI documentation: a cohort study. J Am Med Inform Assoc. 2024;31(4):975-9.

This study/research is funded by the National Institute for Health and Care Research (NIHR) North-West London Patient Safety Research Collaboration (PSRC). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.