Evaluating the use of speech recognition for electronic health record documentation

Tobias Hodgson
BSc Comp Sci, MMgmt, MBA

Centre for Health Informatics
Australian Institute of Health Innovation
Faculty of Medicine and Health Sciences
Macquarie University

This work is submitted for the degree of Doctor of Philosophy

2019

Table of contents

Abstract

The near-universal adoption of electronic health records (EHRs) has made electronic clinical documentation an essential reporting method for most practicing clinicians. Identifying the fastest and safest way for clinicians to enter data into electronic medical records (EMRs) is a matter of urgent importance.

Speech recognition (SR) is an increasingly popular input modality for EHR based clinical documentation, yet there is little research evaluating its use. Previous empirical research has focused on SR as a replacement for dictation and transcription, rather than on its effect on safety, efficiency and usability as an input modality for EHRs.

In order to address this knowledge gap, an experiment was performed and then replicated (E1 and E2) with Emergency Department (ED) doctors to compare the impact of input modality used i.e. keyboard and mouse (KBM) or SR, on clinical documentation. In a controlled environment, doctors were asked to complete tasks of varying complexity within a commercial EHR (E1 8, E2 4). Safety (number and type of errors during documentation), efficiency (the time taken to complete tasks), and usability (doctors' ease of use perception) of the EHR configured with each input modality were measured.

The experiments found no safety or efficiency gains when ED doctors utilised SR for EHR based clinical documentation. For complex tasks, SR assisted documentation took significantly longer to complete when compared to KBM (E1: 18.40%, P=0.009, CI: 9.61-47.73; E2: 16.94%, P=0.01, CI: 11.86-48.26). Overall tasks completed with SR resulted in more non-typographical errors when compared to KBM (E1: KBM 32, SR 138; P<0.01, CI: -1.87--1.16; E2: KBM 26, SR 137, P<0.01, CI: -2.01--1.17).

Potential patient harm errors were significantly greater with SR for simple tasks. SR had a negative impact on the usability score (SUS score: KBM 67 vs. SR 61, P=0.045, CI: 0.14-12.00). Overall, SR was perceived to require more training and support than KBM.

Declaration

I hereby certify that the content of this thesis has not previously been submitted for a higher degree to any other university or institution.

I declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, except where due acknowledgement is made within the thesis. Any contribution made to the research by others, with whom I have worked at Macquarie University or elsewhere, is explicitly acknowledged. I also declare that the intellectual content of this thesis is the product of my own work.

The research presented in chapters 3 through to 5 was approved by the Sydney Local Health District Human Research Ethics Committee – Concord Repatriation General Hospital (CRGH) on 5 January 2015 (LNR/14/CRGH/272). Macquarie University acknowledged the study as externally approved on the 22 January 2015.

Site specific approvals were granted for CRGH (LNRSSA/14/CRGH/273), Liverpool Hospital (LNRSSA/15/LPOOL/68), Royal North Shore Hospital (RNSH) (LNRSSA/15/HAWKE/43) and Manly Hospital (LNRSSA/15/HAWKE/48). (See section 7.1)

Tobias Hodgson

Co-author agreement

The substantive part of this thesis comprises four publications, each of which was prepared in collaboration with at least one co-author. The doctoral candidate is the first author on all publications and made the largest contribution in each paper. Further details of co-author contributions are described in the introductory sections preceding each paper. Each co-author has given permission to include these publications in this thesis as indicated by the signatures below.

I as a co-author agree to the inclusion in this doctoral thesis of publications to which I contributed.

Co-author name	Signature	Date
Professor Enrico Coiera		
Associate Professor Farah Magrabi		

Acknowledgements

This thesis would not have been possible without the invaluable support of my principal supervisor Professor Enrico Coiera and associate supervisor Associate Professor Farah Magrabi, both of whom provided advice, guidance and inspiration throughout my candidacy. My supervisors assisted greatly with the direction of my research and co-authored the publications listed within this thesis. I have gained insight from observing how Professor Coiera approached issues from different perspectives and how he frequently found innovative solutions.

I acknowledge a number of fellow PhD students who added greatly to the experience during this educational journey, in particular Diana Arachi and David Lyell who also performed 2nd inter-rater duties during data analysis.

My thanks also go to colleagues at the Australian Institute of Health Innovation, especially those in the Centre for Health Informatics whose ground-breaking research kept me enthused throughout my time at the centre.

I appreciate the assistance provided by staff and management teams at participating hospitals who helped to facilitate the study and I also acknowledge the study participants for volunteering and fitting the experiment(s) into their busy schedules.

Finally, I would like to thank my family and friends for their patience and support throughout this academic endeavour.

List of abbreviations

The following abbreviations are used throughout this document:

AIHI	Australian Institute of Health Innovation
AI	Artificial Intelligence
AT	Activity Theory
BCI	Brain Computer Interface
CHI	Centre for Health Informatics
CI	Confidence Interval
CPOE	Computerised Physician Order Entry
CRGH	Concord Repatriation General Hospital
CSUQ	Computer System Usability Questionnaire
DT	Dictation and Transcription
ED	Emergency Department
EHR	Electronic Health Record
EMR	Electronic Medical Record
HCI	Human Computer Interaction
HREC	Human Research Ethics Committee Sydney Local Health District
IoT	Internet of Things
ISO	International Organization for Standardization
IT	Information Technology
KBM	Keyboard and Mouse
LHD	Local Health District
MQ	Macquarie University
NHS	National Health Service
NLP	Natural Language Processing
NSLHD	Northern Sydney Local Health District
ONCHIT	Office of the National Coordinator of Health Information Technology
PACS	Picture Archiving and Communication System
PPH	Potential Patient Harm
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RIS	Radiology Information System
RNSH	Royal North Shore Hospital
SLHD	Sydney Local Health District
SR	Speech Recognition
SSA	Site Specific Approval
SUS	System Usability Scale
SWSLHD	South Western Sydney Local Health District
TAT	Turnaround Time
UCD	User-Centred Design
UMUX	Usability Metric for User Experience

List of outputs

The following is a list of the outputs created by the work undertaken as part of this thesis.

Peer-reviewed journal articles

Tobias Hodgson, Enrico Coiera; Risks and benefits of speech recognition for clinical documentation: a systematic review, *Journal of the American Medical Informatics Association*, Volume 23, Issue e1, 1 April 2016, Pages e169–e179, https://doi.org/10.1093/jamia/ocv152

Tobias Hodgson, Farah Magrabi, Enrico Coiera; Efficiency and safety of speech recognition for documentation in the electronic health record, *Journal of the American Medical Informatics Association*, Volume 24, Issue 6, 1 November 2017, Pages 1127–1133, https://doi.org/10.1093/jamia/ocx073

Tobias Hodgson, Farah Magrabi, Enrico Coiera; Evaluating the Efficiency and Safety of Speech Recognition within a Commercial Electronic Health Record System: A Replication Study, *Applied Clinical Informatics Journal*, Volume 9, Issue 2, May 2018, Pages 326-325, https://doi.org/10.1055/s-0038-1649509

Tobias Hodgson, Farah Magrabi, Enrico Coiera; Evaluating the usability of speech recognition to create clinical documentation using a commercial electronic health record, *International Journal of Medical Informatics*, Volume 113, May 2018, Pages 38-42, https://doi.org/10.1016/j.ijmedinf.2018.02.011

1. Introduction

1.1 Background

The use of electronic health records[1] (EHRs) has become virtually ubiquitous worldwide. There has been an overwhelming uptake within certain healthcare sectors such as primary care clinicians where the majority utilise electronic clinical documentation (United States 87%, Netherlands 99%, New Zealand and Norway 97%, United Kingdom 96%, Australia at 95%).[1, 2] Hospital adoption of EHRs currently lags behind primary care with lower rates of implementation (United States 81%, United Kingdom 50%, Australia 75%).[3-6]

However, the adoption of EHRs within hospitals is rapidly closing this gap with numerous national initiatives continuing to stimulate uptake. The most notable instances include the United States' 'Meaningful Use' program (Medicare and Medicaid Electronic Health Record Incentive Programs) and the NHS England's 'Paperless 2020' target. The Australian Government's move to the 'My Health Record' opt out model has also increased EHR uptake.[7-9] These programs (and countless others) are driving the growth of the global EHR market at a rapid rate with some market analysts estimating a 5.7% compound annual growth rate until 2025.[10] It is likely that it is only a matter of time before virtually all health records become electronic.

With the current move towards the elimination of paper-based documentation in digital healthcare settings, it is crucial to provide clinicians with the tools and skills to deliver care with the use of EHR systems as quickly and safely as possible. It is essential that all clinicians with access to an electronic patient record have confidence that the displayed patient information is current and accurate, otherwise the many expected benefits of an EHR will go unrealised and detrimental effects on patient outcomes may occur.

[1]Also, electronic medical records (EMR)

1.1.1 EHR benefits

EHR systems have the potential for numerous benefits across clinical, organisational and societal outcomes.[11, 12]

For care delivery, EHRs provide near-instant access to clinical notes and test results along with the ability to easily share information electronically. EHR systems have been shown to provide an overall increase in documentation quality and completeness over paper records.[13, 14] EHRs have now made it possible for records to be shared, viewed and acted upon by any number of other healthcare professionals in any location, with access and permission to view the patient record.[15]

EHRs are foundational to realising the benefits of clinical decision support systems for both diagnostic and management decisions. These systems assist with aspects including information management via knowledgebases, attention focusing by highlighting potential issues, and patient-specific recommendations with diagnostic tools.[16]

Patient outcomes may be improved with less repetition of tests and reduced treatment times. Greater patient education and increased involvement in their treatment may also lead to enhanced results.[17, 18]

Organisational benefits include potential economic benefits through improved billing and cost reductions.[18] With improved visibility and the ability to scrutinise a wider variety of data, performance gains may be found more easily via EHR reporting and statistical analysis.

Broader societal outcomes are improved with the enhanced ability to conduct research utilising electronic data, which should result in a more rapid research cycle.[12, 18] There are many benefits of secondary use of EHRs data for research, both clinical and basic, including simple access to historical data for retrospective studies and ease of identification of potential participants for prospective studies. Large scale public health related research on epidemics and the spread of disease are also increasingly possible due to the digitalisation of health records.[19]

1.1.2 EHR risks

EHR documentation also brings with it risks to clinical, organisational and societal outcomes by introducing new opportunities for errors into clinical workflow.

Risks with electronic records may arise when the record is not available, accessible or usable. Availability concerns may occur due to system uptime, access, or capacity issues, along with lack of physical access to a computer terminal for the clinicians themselves due to space restrictions. Accessibility issues may come from security or interoperability concerns, while issues related to how usable a record is could be due to quality of data or technical design.[20] In some cases, using EHRs can increase the data entry burden on clinicians, leading to lower levels of physician satisfaction, increased rates of burnout, changes in the patient-physician interaction and the introduction of new classes of error and alert fatigue.[21-24]

Risk to patients may occur in the form of treatment delays, inaccurate record information, patient privacy and emotional distress associated with greater personal access and control.[12]

For organisations there are also additional overheads in terms of equipment, training and support that are required when implementing an EHR system. These often result in a temporary loss in productivity and reduced staff satisfaction that ultimately may lead to increased staff turnover leading to additional resourcing issues.

EHR systems are often integrated within a greater national digital health record or personal health record system.[25] Whether a distributed model with data remaining with the source of origin but shared throughout the network, or a consolidated single repository-based model, the increased interoperability of individual healthcare systems has highlighted the need for the accuracy and validity of clinical documentation entries. This poses a real risk that incorrect information may be propagated across systems leading to decisions being made based on flawed data with potential patient harm consequences.

While various types of EHR risks have been identified, research within this thesis focuses on the specific clinical risks arising from the impact of, increased errors, reduced EHR documentation efficiency and ease of use.[26]

The prevalence of EHR related safety concerns and the ever-increasing amounts of data being entered into medical record systems make the pursuit for the safest and most efficient method of performing electronic clinical documentation a matter of urgent importance.[27-29]

1.1.3 Role of interaction modality on risks and benefits

The physical method of entering information into the EHR is an often-overlooked aspect that bears a direct influence on the efficiency and quality of the clinical documentation produced. While the option of paper-based notes is phasing out, there are an ever-increasing number of input options available for digital clinical reporting. These include: keyboard & mouse (KBM), mobile devices, touch screens, digital pens, speech recognition (SR), head-mounted displays, glasses, virtual reality and augmented reality. New methods are constantly entering the marketplace, including some seemingly antiquated methods showing great promise, such as the use of human scribes to assist clinicians with real time electronic documentation.[30, 31]

SR is one of the more popular documentation options available to clinicians (often referred to as voice recognition or automatic speech/voice recognition). SR systems have been commercially available for medical reporting for over two decades and have been adopted in specific clinical settings such as radiology reporting, yet, SR is not uniformly used across all clinical domains.[32-34] The use of SR for tasks beyond straight dictation of text such as direct EHR reporting, filling diagnostic templates, system navigation and advanced controls are all relatively untested. However, SR is often implemented specifically to assist with these very tasks, despite the lack of evidence of benefit and potential for unintended consequences on care delivery.[20]

1.2 Thesis aim

The aim of this research was to evaluate the use of SR for clinical documentation within an EHR, identifying whether the introduction of SR was associated with any measurable benefit or risk.

An experimental study was designed and subsequently replicated to provide empirical data on the 1) Efficiency, 2) Safety, and 3) Usability of EHR based clinical documentation with both SR and the more traditional input modality of KBM. These experiments allowed for direct comparison of documentation undertaken with the two modalities, while completing a standard set of documentation tasks in a controlled environment. The studies were designed to closely model the real-world tasks and systems used by Emergency Department (ED) doctors on a daily basis. Commercial EHR and SR systems were utilised and configured to imitate the recently implemented state health based ED SR solution available to participants.

Tasks undertaken included: the operation and navigation of system menus, fields, pick lists and patient charts, item and task selections, as well as data entry.

1) Efficiency

The length of time required for clinicians to complete electronic documentation is a crucial aspect of EHR use that is often compared with the speed of using traditional paper-based systems. A systematic review of 23 studies found that introduction of EHR based documentation using KBM decreased documentation times for nursing (-24.5% bedside, -23.5% central terminals) while at the same time it increased documentation times for physicians (+17.5% at the bedside, +238.4% centralised computerised provider order entry).[35] However, the impact of using SR on EHR documentation times is unknown.

The aim of the efficiency component of the thesis was thus to provide empirical evidence for whether the introduction of SR provided any tangible advantage for clinical documentation task completion times when compared to the more traditional KBM input modality.

2) Safety

Safety risks introduced with the use of EHRs cover the areas of system design, implementation and use.[36-38] Types of issues observed are well documented and can be classified as either human, machine or a combination of both.[26] Human factors include staffing/training, cognitive load and failure to perform duty, machine based factors include transfer and general technical, or human/machine due to input and output problems. Each bring with them the potential to cause patient harm both at an individual patient level and on a large scale. These all often play a major role in patient harm events and the safe use of EHR systems is crucial to their success.[26, 39]

Given that patient safety should be a primary concern of any clinical system, this thesis also explored the safety of EHR documentation when performed with two common digital input modalities (KBM and SR) by comparing the number, type and severity of error observed during clinical documentation.

The aim of the safety aspect of the thesis was to provide empirical evidence of whether the use of different input modality (KBM or SR) had any significant effect on the number and type of error that occurred when performing clinical documentation within an EHR system.

3) Usability

Usability is a critical yet often-undervalued aspect in the success or failure of an EHR implementation. To be successful, EHR systems must not only be technically robust, but should also be easily usable by clinicians.[40]

Usability can be measured in numerous ways including ergonomic attributes, user mental effort required, and ease of use or user acceptability.[41] The most common tools and instruments used for measuring perceived usability include the Computer System Usability Questionnaire (CSUQ), System Usability Scale (SUS) and the Usability Metric for User Experience (UMUX), all of which were found to be comparable.[42]

There are various guidelines for EHR system usability evaluation used to assess and compare EHR systems in numerous previous studies, the majority of which advocate the use of the SUS.[43-47] This research utilised the SUS to examine the usability of the EHR system while being operated with two different input modalities (KBM & SR).[43-47] Clinicians scored the usability of the EHR system with each input modality (KBM and SR) after exposure to each configuration while performing a defined set of documentation tasks.

The usability aspect of the thesis aimed to identify any links between system usability and the method of data entry, while also investigating whether any significant change in clinician opinion on an EHR system usability occurred with the introduction of an alternative input modality. It specifically sought to determine if the introduction of SR improved or reduced system usability, with a secondary goal of assessing whether a difference in system usability was associated with a significant performance variation.

1.3 Thesis structure

This thesis is built around a core of four individual but related peer-reviewed papers. Each paper addresses a specific aspect of EHR based clinical documentation. The thesis structure and current position is summarised in the model below. (See Figure 1)

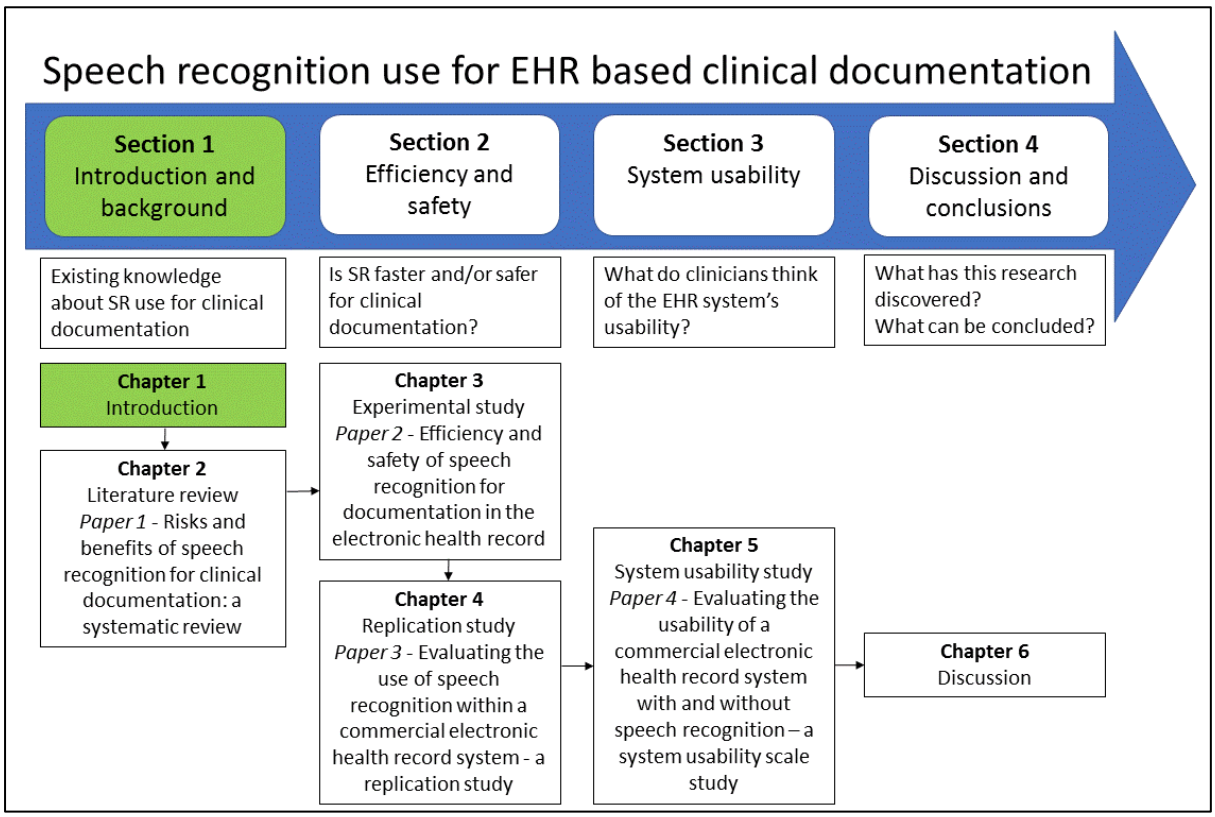


Figure 1: Thesis structure and position – Chapter 1

Section 1 – Introduction and background

The opening section of this thesis includes a chapter based on the first of four publications, a literature review of existing research into the use of SR for electronic clinical documentation. This systematic review summarises existing knowledge of SR use for EHR based clinical documentation.

Section 2 – Efficiency and safety

The second section reports original empirical research in two chapters with results of experimental studies on SR efficiency and safety. Chapter 3 is based on a paper reporting the results of the initial experimental study, and chapter 4 contains a paper reporting on the follow-up replication study.

Section 3 – System usability

The penultimate section investigates the subjective usability of the EHR systems with each input modality (SR and KBM) as reported by clinicians. Chapter 5 reports the results of a system usability study of the EHR and SR systems employed.

Section 4 – Discussion and conclusion

Chapter 6 concludes the thesis by providing an overall analysis of the findings of the research and summarises some wider implications that can be drawn from this work.

1.4 Chapter references

1. Schoen C, Osborn R, Doty MM, Squires D, Peugh J, Applebaum S. A survey of primary care physicians in eleven countries, 2009: perspectives on care, costs, and experiences. Health Affairs. 2009;28(6):w1171-w83.

2. Connors WR. Leading by example: MHCA and SATVA work together to produce a white paper on successful EHR implementations. Behavioral healthcare. 2006;26(5):49-51.

3. Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital "advanced use" divide. Journal of the American Medical Informatics Association. 2017;24(6):1142-8.

4. Henry J, Pylypchuk Y, Searcy T, Patel V. Adoption of electronic health record systems among US non-federal acute care hospitals: 2008-2015. ONC Data Brief. 2016;35:1-9.

5. Deloitte M. Digital health in the UK: an industry study for the office of life sciences. Deloitte Creative Studio. 2015.

6. Australian Government ADHA. Who is using digital health? Online: Australian Digital Health Agency; 2018 [cited 2018 02/07/2018]; Available from: https://www.myhealthrecord.gov.au/about/who-is-using-digital-health.

7. NHS England. Next steps on the NHS five year forward view. London: NHS England2017.

8. US Government CfMMS. Electronic Health Records (EHR) Incentive Programs. Baltimore, MD 21244: U.S. Centers for Medicare & Medicaid Services; 2018 [cited 2018 2018/04/16]; Available from: https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentiveprograms/.

9. Australian Government DoH. My Health Record: National Opt-out. Canberra ACT 2601, Australia2018 [cited 2018 2018/04/16]; Available from: http://www.health.gov.au/internet/main/publishing.nsf/Content/my-health-record-national-opt-out.

10. Transparency Market Research. Electronic Health Records Market - Global Industry Analysis, Size, Share, Growth, Trends, and Forecast 2017 - 2025. Pune, India: 2017.

11. Hoover R. Benefits of using an electronic health record. Nursing2018. 2016;46(7):21-2.

12. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. Risk management and healthcare policy. 2011;4:47.

13. Jang J, Yu SH, Kim C-B, Moon Y, Kim S. The effects of an electronic medical record on the completeness of documentation in the anesthesia record. International journal of medical informatics. 2013;82(8):702-7.

14. Tang PC, LaRosa MP, Gorden SM. Use of computer-based records, completeness of documentation, and appropriateness of documented clinical decisions. Journal of the American Medical Informatics Association. 1999;6(3):245-51.

15. Gunter TD, Terry NP. The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions. Journal of medical Internet research. 2005;7(1).

16. Musen MA, Middleton B, Greenes RA. Clinical decision-support systems. Biomedical informatics: Springer; 2014. p. 643-74.

17. Raposo VL. Electronic health records: Is it a risk worth taking in healthcare delivery? GMS health technology assessment. 2015;11.

18. Menachemi N, Brooks RG. Reviewing the benefits and costs of electronic health records and associated patient safety technologies. Journal of medical systems. 2006;30(3):159-68.

19. Sandhu E, Weinstein S, McKethan A, Jain SH. Secondary uses of electronic health record data: benefits and barriers. Joint Commission journal on quality and patient safety. 2012;38(1):34-40.

20. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. Journal of the American Medical Informatics Association. 2004;11(2):104-12.

21. Shanafelt TD, Dyrbye LN, Sinsky C, Hasan O, Satele D, Sloan J, et al., editors. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. Mayo Clinic Proceedings; 2016: Elsevier.

22. Bates DW, Cohen M, Leape LL, Overhage JM, Shabot MM, Sheridan T. Reducing the frequency of errors in medicine using information technology. Journal of the American Medical Informatics Association. 2001;8(4):299-308.

23. Tai-Seale M, Olson CW, Li J, Chan AS, Morikawa C, Durbin M, et al. Electronic health record logs indicate that physicians split time evenly between seeing patients and desktop medicine. Health Affairs. 2017;36(4):655-62.

24. Khairat S, Burke G, Archambault H, Schwartz T, Larson J, Ratwani RM. Perceived Burden of EHRs on Physicians at Different Stages of Their Career. Applied clinical informatics. 2018;9(02):336-47.

25. Roehrs A, da Costa CA, da Rosa Righi R, de Oliveira KSF. Personal health records: a systematic literature review. Journal of medical Internet research. 2017;19(1).

26. Kim MO, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. Journal of the American Medical Informatics Association. 2017;24(2):246-50.

27. Menon S, Singh H, Giardina TD, Rayburn WL, Davis BP, Russo EM, et al. Safety huddles to proactively identify and address electronic health record safety. Journal of the American Medical Informatics Association. 2017;24(2):261-7.

28. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health information science and systems. 2014;2(1):3.

29. Feldman B, Martin EM, Skotnes T. Big data in healthcare hype and hope. October 2012 Dr Bonnie. 2012;360.

30. Gidwani R, Nguyen C, Kofoed A, Carragee C, Rydel T, Nelligan I, et al. Impact of scribes on physician satisfaction, patient satisfaction, and charting efficiency: a randomized controlled trial. The Annals of Family Medicine. 2017;15(5):427-33.

31. Coiera E, Kocaballi B, Halamaka J, Laranjo L. The digital scribe. npj Digital Medicine. 2018;1(1):58.

32. Johnson M, Lapkin S, Long V, Sanchez P, Suominen H, Basilakis J, et al. A systematic review of speech recognition technology in health care. BMC Medical Informatics and Decision Making. 2014;14(1):94.

33. Herman SJ. Speech recognition and the creation of radiology reports. Applied Radiology. 2004;33(5):23-8.

34. Lawrence D. Can you hear me now? Voice recognition for the EMR has made big strides, and many say meaningful use requirements will accelerate adoption. Healthcare informatics : the business magazine for information and communication systems. 2009;26(12).

35. Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health records on time efficiency of physicians and nurses: a systematic review. Journal of the American Medical Informatics Association. 2005;12(5):505-16.

36. Meeks DW, Smith MW, Taylor L, Sittig DF, Scott JM, Singh H. An analysis of electronic health record-related patient safety concerns. Journal of the American Medical Informatics Association. 2014;21(6):1053-9.

37. Coiera E, Magrabi F. Guide to health informatics. CRC press; 2015. p. 195-220.

38. Sittig DF, Singh H. Defining health information technology–related errors: New developments since To Err Is Human. Archives of internal medicine. 2011;171(14):1281-4.

39. Magrabi F, Ong M-S, Runciman W, Coiera E. An analysis of computer-related patient safety incidents to inform the development of a classification. Journal of the American Medical Informatics Association. 2010;17(6):663-70.

40. Armijo D, McDonnell C, Werner K. Electronic health record usability. Interface Design Considerations James Bell Associates The Altarum Institute. 2009.

41. Bevan NK, Kirakowski J, editors. J. And Maissel, J.(1991). What is usability. Proceedings of the 4th International Conference on HCI, Stuttgart; 1991.

42. Lewis JR. Measuring perceived usability: The CSUQ, SUS, and UMUX. International Journal of Human–Computer Interaction. 2018:1-9.

43. Arellano P, Bochinski, J., Boone, E., Grayson, R., Phua, J., & Uehara, R. Selecting an EMR for your practice: Evaluating usability. Healthcare Information and Management Systems Society (HIMSS): Chicago, IL. 2010.

44. Scarlat A. Electronic health record : a systems analysis of the medications domain. Boca Raton: CRC Press; 2012.

45. Drummond Group. 2015 EHR Certification. Website: Drummond Group; 2015 [cited 2017 2017/09/27]; Available from: https://www.drummondgroup.com/products/certified-products/2015-edition-ehr-products.

46. Drummond Group. 2014 EHR Certification. Website: Drummond Group; 2014 [cited 2017 2017/09/27]; Available from: https://www.drummondgroup.com/products/certified-products/2014-edition-ehr-products.

47. Yen P-Y, Bakken S. Review of health information technology usability study methodologies. Journal of the American Medical Informatics Association. 2011;19(3):413-22.

2. Literature review: Risks and benefits of speech recognition for clinical documentation

2.1 Chapter background

At present little is known about the way that SR is utilised within clinical settings. While SR has been adopted successfully in some areas such as radiology reporting, it has not been uniformly used across all clinical domains.[1, 2] How and where SR is used, along with any potential benefits and risks to documenting with this input modality, are relatively unknown.

In order to address this knowledge gap, a systematic review of existing studies was undertaken to identify where and how SR is currently used for documentation within clinical settings. (See Figure 2)

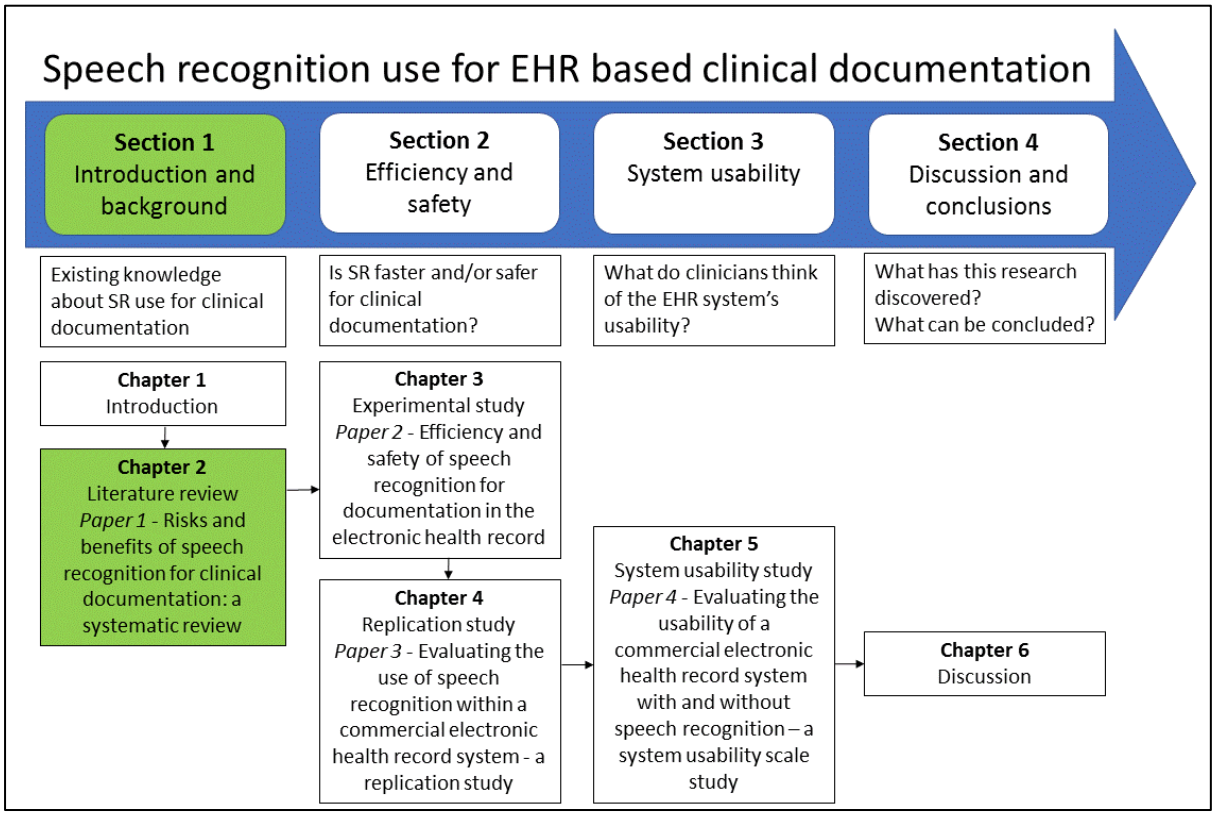


Figure 2: Thesis structure and position – Chapter 2

2.1.1 Aim

The aim of this review was to evaluate existing research studies that addressed SR use within clinical documentation, summarise previous findings and describe the benefits and risks associated with the use of SR system for clinical documentation tasks. A secondary aim was to explore whether SR performance in clinical documentation tasks has improved over time as this technology class has matured.

2.1.2 Method

Elements of both the Cochrane and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodologies were followed while undertaking the systematic review. Guidelines and protocols detailed within the Cochrane handbook and the PRISMA statement and checklist were followed to ensure systematic review requirements were met.[3, 4]

Existing studies analysing speech (or voice) recognition and healthcare related documentation were identified, screened and assessed for relevance by two reviewers, and articles that met requirements after screening and assessment process were then included within the review.

Perhaps unsurprisingly, due to the prevalence of SR in the field of radiology, fifteen of the twenty-three articles within the review were radiology documentation-based studies.

Four major vendors provided the SR technologies reported within the studies: IBM, Philips, Nuance, and Agfa. The sale of Philips Speech Recognition Systems to Nuance in 2008 led to the Nuance solution becoming the most frequently reported within the review studies, and it is arguably the most common SR technology currently used worldwide. (See Figure 3)

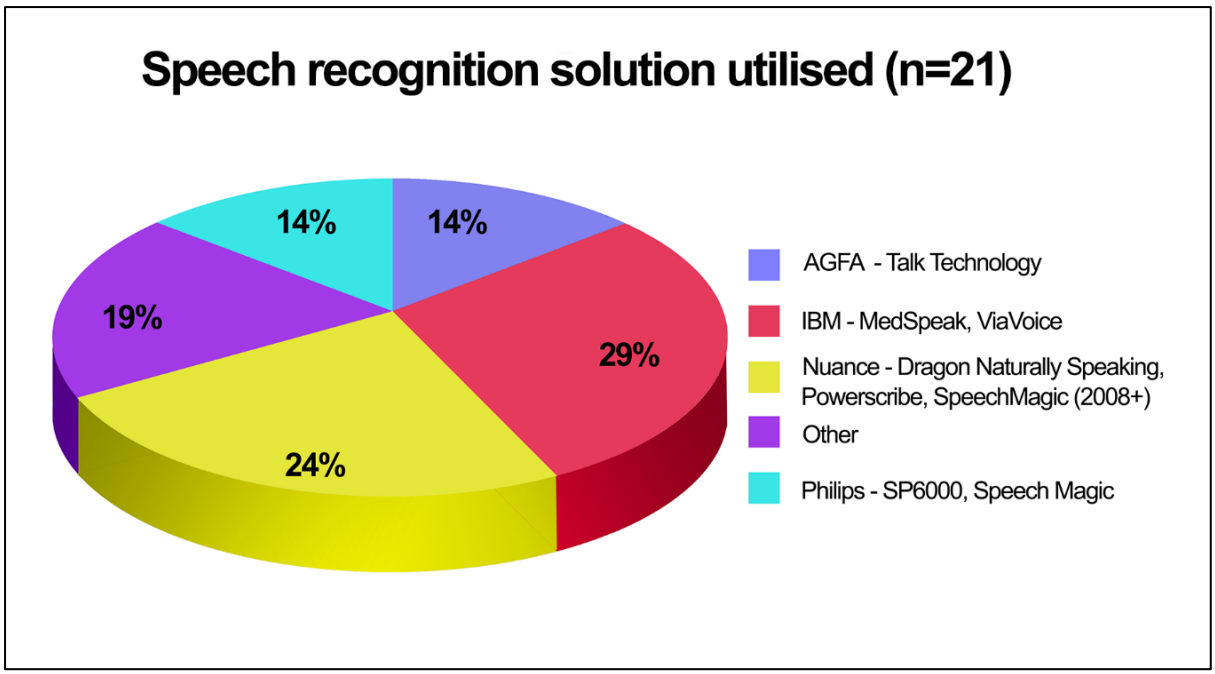


Figure 3: Speech recognition system breakdown – SR solution reported by review studies (n=21).

2.1.3 Results

The limited studies available for review (n=23) showed significant heterogeneity, with considerable differences in both data quality and trial design. The efficiency of documentation with SR was assessed in many of the studies, however SR efficiency was predominately compared to dictation and transcription (DT) services where voice recordings were captured on tape/disc and physically sent off for human transcription. Document editing time increased while using SR when compared to DT in four of six studies (+1876.47% to –16.50%). Dictation time similarly increased in three of five studies (+91.60% to –25.00%). Turnaround time (TAT), the time taken for the entire process from report creation to completion and submission, consistently improved using SR when compared to DT (16.41% to 82.34%). Across all studies the improvement in TAT was 0.90% per year.

The safety of SR was assessed in terms of accuracy rates and errors introduced. SR accuracy was reported in ten studies (88.90% to 96.00%) and appeared to improve 0.03% per year as the technology matured. Text output errors were the primary focus for error assessment within the studies. The mean number of documentation errors per report increased when using SR (0.05 to 6.66) when compared to DT (0.02 to 0.40).

2.1.4 Discussion

Existing studies reported conflicting results both for and against the use of SR for clinical documentation. Dictation time efficiency gains with the use of SR were seen in two studies while three studies showed decreased efficiencies. Faster editing times were reported for SR in two studies, while they were found to be slower in four studies. The number of words per report with the use of SR was found to decrease in three studies and increased in one. Within all studies that measured errors, SR was shown to increase the number of errors per report. Four studies reported some form of financial benefit for SR whilst three reported financial costs.

The modest increase in accuracy of SR over time observed may indicate that this aspect has now improved, which aligns with SR software vendors' current claims of accuracy rates of up to 99%.[5]

The literature review revealed a significant gap in the existing research on SR assisted EHR clinical documentation, with only a few studies that compared SR to DT for clinical report creation, and none that explored the use of SR for general input and navigation of the EHR. Only one study directly compared SR to the use of KBM, which is the most common input modality for EHR documentation.

2.1.5 Conclusions

The literature review highlights that, although SR is a potentially valuable tool for clinical documentation, any benefits must be weighed against potential time penalties, the introduction of additional errors, and unclear cost-benefit.

There is currently no substantial evidence that SR offers any significant benefit or cost for clinical documentation. Many questions such as the most suitable setting, task and users for SR documentation remain unanswered, along with numerous unexplored opportunities for SR use within the medical domain.

In particular, it is important to address the knowledge gap identified within the literature review of a lack of existing studies that address SR use for EHR clinical report creation, general input and navigation when compared to other input modalities.

The pertinent issues raised by the review are:

1) Existing evidence for SR clinical use is solely dictation based.
2) SR is used for EHR based clinical documentation in ways beyond straight dictation, including the operation and navigation of system menus, fields, pick list and patient charts, item and task selections, and data entry.
3) There is no evidence for improved efficiency, safety or usability with the use of SR for performing the tasks listed in 2).

2.2 Article contributorship statement

Tobias Hodgson (TH) and Enrico Coiera (EC) conceived the study and its design. TH conducted the research, the primary analysis, and the initial drafting of the paper. EC contributed to the analysis and drafting of the paper and both TH and EC approved the final manuscript. TH is the corresponding author.

2.3 Paper 1 - Risks and benefits of speech recognition for clinical documentation: a systematic review

Journal of the American Medical Informatics Association, Volume 23, Issue e1, 1 April 2016, Pages e169–e179, https://doi.org/10.1093/jamia/ocv152

Pages 20-30 of this thesis have been removed as they contain published material. Please refer to the citation above for details of the article contained in these pages.

2.4 Chapter references

1. Herman SJ. Speech recognition and the creation of radiology reports. Applied Radiology. 2004;33(5):23-8.

2. Lawrence D. Can you hear me now? Voice recognition for the EMR has made big strides, and many say meaningful use requirements will accelerate adoption. Healthcare informatics : the business magazine for information and communication systems. 2009;26(12).

3. Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions: John Wiley & Sons; 2011.

4. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS medicine. 2009;6(7):e1000097.

5. Nuance Communications. Dragon NaturallySpeaking 13 Premium Data Sheet - Nuance Communications. 2014:2.

2.5 Paper 1 references

1. Johnson M, Lapkin S, Long V, Sanchez P, Suominen H, Basilakis J, et al. A systematic review of speech recognition technology in health care. BMC Medical Informatics and Decision Making. 2014;14(1):94.

2. Herman SJ. Speech recognition and the creation of radiology reports. Applied Radiology. 2004;33(5):23-8.

3. Lawrence D. Can you hear me now? Voice recognition for the EMR has made big strides, and many say meaningful use requirements will accelerate adoption. Healthcare informatics: the business magazine for information and communication systems. 2009;26(12).

4. Neustein A, Markowitz JA. Mobile Speech and Advanced Natural Language Solutions: Springer; 2013.

5. Bliss MF. Speech recognition for the health professions: (using Dragon NaturallySpeaking): Prentice Hall; 2005.

6. Madisetti V. Video, Speech, and Audio Signal Processing and Associated Standards: CRC Press; 2009.

7. Paulett JM, Langlotz CP. Improving language models for radiology speech recognition. Journal of Biomedical Informatics. 2009 2//;42(1):53-8.

8. Gales M, Young S. The application of hidden Markov models in speech recognition. Foundations and Trends in Signal Processing. 2008;1(3):195-304.

9. Eddy SR. What is a hidden Markov model? Nat Biotech. 2004 10//print;22(10):1315-6.

10. Indurkhya N, Damerau FJ. Handbook of natural language processing: CRC Press; 2012.

11. Lafferty J, McCallum A, Pereira FC. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

12. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine. 2009;151(4):264-9.

13. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin. 1968;70(4):213.

14. Hundt WS, O.;Scharnberg, B.;Hold, M.;Kohz, P.;Lienemann, A.;Bonél, H.;Reiser, M. Speech processing in radiology. European Radiology. 1999;9(7):1451-6.

15. Rana DSH, G.;Shepstone, L.;Pilling, J.;Cockburn, J.;Crawford, M. Voice recognition for radiology reporting: Is it good enough? Clinical Radiology. 2005;60(11):1205-12.

16. Pezzullo JAT, G. A.;Rogg, J. M.;Davis, L. M.;Brody, J. M.;Mayo-Smith, W. W. Voice recognition dictation: Radiologist as transcriptionist. Journal of Digital Imaging. 2008;21(4):384-9.

17. Mohr DNT, D. W.;Pond, G. R.;Kamath, J. S.;De Vos, C. B.;Carpenter, P. C. Speech recognition as a transcription aid: A randomized comparison with standard transcription. Journal of the American Medical Informatics Association. 2003;10(1):85-93.

18. Issenman RMJ, I. H. Use of voice recognition software in an outpatient pediatric specialty practice. Pediatrics. 2004;114(3):e290-e3.

19. Vorbeck FB-S, A.;Kettenbach, J.;Huebsch, P. Report generation using digital speech recognition in radiology. European Radiology. 2000;10(12):1976-82.

20. Al-Aynati MMC, K. A. Comparison of voice-automated transcription and human transcription in generating pathology reports. Archives of Pathology and Laboratory Medicine. 2003;127(6):721-5.

21. Bhan SNC, C. L.;Norman, G. R.;Ali, S. H. Effect of voice recognition on radiologist reporting time. Canadian Association of Radiologists Journal. 2008;59(4):203-9.

22. Rosenthal DIC, F. S.;Dupuy, D. E.;Kattapuram, S. V.;Palmer, W. E.;Yap, R. M.;Levine, L. A. Computers in radiology: Computer-based speech recognition as a replacement for medical transcription. American Journal of Roentgenology. 1998;170(1):23-5.

23. Chapman WWA, D.;Fiszman, M.;Haug, P. J. Contribution of a speech recognition system to a computerized pneumonia guideline in the emergency department. Proceedings / AMIA Annual Symposium AMIA Symposium. 2000:131-5.

24. Ramaswamy MRC, G.;Esch, O.;Fanning, D. D.;VanSonnenberg, E. Continuous speech recognition in MR imaging reporting: Advantages, disadvantages, and impact. American Journal of Roentgenology. 2000;174(3):617-22.

25. Ilgner JD, P.;Westhofen, M. Free-text data entry by speech recognition software and its impact on clinical routine. Ear, Nose and Throat Journal. 2006;85(8):523-7.

26. Koivikko MP, Kauppinen T, Ahovuo J. Improvement of report workflow and productivity using speech recognition—a follow-up study. Journal of digital imaging. 2008;21(4):378-82.

27. Krishnaraj A, Lee JK, Laws SA, Crawford TJ. Voice recognition software: effect on radiology report turnaround time at an academic medical center. American Journal of Roentgenology. 2010;195(1):194-7.

28. Kanal KM, Hangiandreou NJ, Sykes AM, Eklund HE, Araoz PA, Leon JA, et al. Initial evaluation of a continuous speech recognition program for radiology. Journal of Digital Imaging. 2001 2001/03/01;14(1):30-7.

29. Smith NTB, R. A.;Pettus, D. C.;Jones, B. R.;Quinn, M. L.;Sarnat, A. Recognition accuracy with a voice-recognition system designed for anesthesia record keeping. Journal of Clinical Monitoring. 1990;6(4):299-306.

30. Zemmel NJ, Park SM, Schweitzer J, O'Keefe JS, Laughon MM, Edlich RF. Status of voicetype dictation for windows for the emergency physician. The Journal of emergency medicine. 1996;14(4):511-5.

31. McGurk SB, K.;Macfarlane, T. V.;Duncan, K. A. The effect of voice recognition software on comparative error rates in radiology reports. British Journal of Radiology. 2008;81(970):767-70.

32. Basma SL, B.;Jacks, L. M.;Rizk, M.;Scaranelo, A. M. Error rates in breast imaging reports: Comparison of automatic speech recognition and dictation transcription. American Journal of Roentgenology. 2011;197(4):923-7.

33. David GC, Chand D, Sankaranarayanan B. Error rates in physician dictation: quality assurance and medical record production. International journal of health care quality assurance. 2014;27(2):99-110.

34. Quint LEQ, D. J.;Myles, J. D. Frequency and Spectrum of Errors in Final Radiology Reports Generated With Automatic Speech Recognition Technology. Journal of the American College of Radiology. 2008;5(12):1196-9.

35. Chang CAS, R.;Jolley, D. Non-clinical errors using voice recognition dictation software for radiology reports: A retrospective audit. Journal of Digital Imaging. 2011;24(4):724-8.

36. Belton K, Dick R. Voice-recognition technology: key to the computer-based patient record. Journal (American Medical Record Association). 1991 Jul;62(7):27-32, 6, 8.

37. Clark S. Implementation of voice recognition technology at Provenant Health Partners. Journal of AHIMA/American Health Information Management Association. 1994;65(2):34, 6, 8-, 6, 8.

38. Nuance Communications. Dragon NaturallySpeaking 13 Premium Data Sheet - Nuance Communications. 2014:2.

39. Coiera E. Guide to health informatics third edition: CRC Press; 2015.

40. Coiera E. Technology, cognition and error. BMJ Qual Saf 2015 (in press). 2015.

41. Coiera E, Aarts J, Kulikowski C. The dangerous decade. Journal of the American Medical Informatics Association. 2012;19(1):2-5.

42. Chleborad KD, T.;Zvara, K.;Hippmann, R.;Ivancakova, R.;Zvarova, J.;Smidl, L.;Trmal, J.;Psutka, J. Evaluation of voice-based data entry to an electronic health record system for dentistry. Biocybernetics and Biomedical Engineering. 2013;33(4):204-10.

3. Experimental study: Efficiency and safety of speech recognition for documentation in the EHR

3.1 Chapter background

The previous chapter established that current research evidence for the use of SR for EHR based clinical documentation is sparse. There is some evidence for SR as a replacement for DT or as a typing substitute, but it is currently not possible to clearly articulate the tasks and clinical settings in which SR use is of benefit, and where it should perhaps be avoided.

This research gap prompted an experimental study designed to assess SR use in terms of safety and efficiency while completing EHR documentation tasks. (See Figure 4)

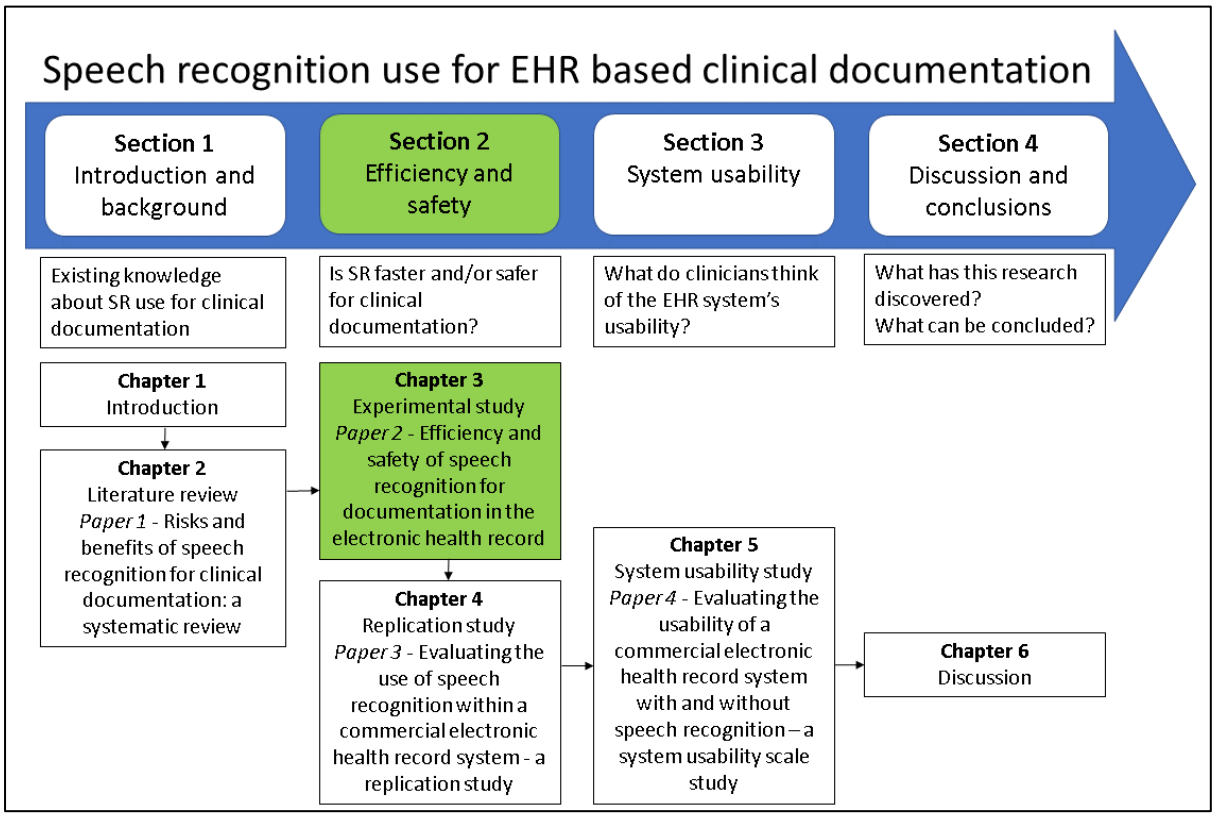


Figure 4: Thesis structure and position – Chapter 3

3.1.1 Hypotheses

This research sets out to test two primary hypotheses and one secondary hypothesis:

Primary hypothesis

1 Speech recognition controlled electronic health record documentation, navigation and interaction is more efficient than keyboard and mouse, i.e. speech recognition assisted documentation is faster.

2 Speech recognition controlled electronic health record documentation, navigation and interaction is safer than with keyboard and mouse, i.e. fewer errors will be committed while documenting with speech recognition.

Secondary hypothesis

Speech recognition controlled electronic health record documentation, navigation and interaction is less impacted by interruptions than keyboard and mouse, i.e. interruptions will result in less additional time and fewer errors to speech recognition documentation than keyboard and mouse.

3.1.2 Method

A within-subject experimental study was undertaken with 35 ED doctors, with each assigned standardised clinical documentation tasks to undertake using a commercial EHR system. Participants navigated the EHR and documented patient information for simulated patients using both KBM and SR as the input modality.

Participant recruitment

The recruitment of clinicians was predominately undertaken via the ED directors at each of the individual hospitals involved in the study. Overall, the hospital managers involved were enthusiastic and supportive of the study and proved invaluable in the recruitment, scheduling and logistics of the experiment trials. Scheduling and attendance of trial sessions rather than recruitment was the major

challenge faced. This was problematic as clinicians worked various shifts, often rotating within departments or hospitals across the entire local health district (LHD). Even with sessions scheduled and clinicians on shift, priority was often (correctly) given to the demands of a busy ED. As a result, numerous trial sessions were run throughout the entire 24-hour roster to obtain the required number of participants.

Task breakdown

The tasks assessed within the trials were representative of those commonly undertaken during EHR documentation as performed by participants on a daily basis. Clinician observations and discussions with senior ED clinicians led to the design of a set of tasks. Each task was performed a total of four times by each subject, in a randomly assigned order by either KBM or SR and with or without interruption.

The following tasks were completed by participants during the experiment trials:

- *Patient assignment* – participants were assigned a specified, yet to be treated, patient waiting within the EHR system.
- *Patient assessment* – participants were provided a basic ED assessment of the patient and were asked to enter data for: present complaint, chief complaint, history of present illness, histories - past medical history and family and social history.
- *Vital signs* – participants looked up the latest existing vital sign entries for the patient and documented these results within their notes.
- *Diagnosis* – participants added a diagnosis for the patient into the EHR.
- *Orders* – participants placed an order for a specified test to be undertaken on the patient.
- *Discharge* – participants began the discharge process and created a discharge summary for the patient.

In addition, participants dealt with any navigational, authentication, confirmation or interruptions that occurred while completing these tasks, e.g. entering credentials during sign-off.

A non-production IT environment with EHR and SR systems was created for participant sessions configured similarity to the state health based system that subjects used routinely. Monitoring and data capture systems along with a video and audio capture device were used to record each individual participant session for detailed analysis post trial.

3.1.3 Results

EHR documentation times were found to be slower overall when undertaken with SR when compared to KBM (18.11% slower, KBM 140.09s, SR 165.46s, P=0.001, CI: 9.87-33.91). This inefficiency by SR was observed for both simple and complex tasks (simple 16.95%, KBM 112.38s, SR 131.44s, P=0.050, CI: -0.07-30.50 and complex 18.40%, KBM 170.48s, SR 201.84s, P=0.09, CI: 9.61-47.73).

Increased errors were observed overall when documenting with SR compared to KBM (KBM 32, SR 138, P<0.001, CI: -1.87--1.16). This increase in errors for SR documentation was observed for both simple (KBM 9, SR 75, P<0.001, CI: -2.42--1.35) and complex tasks (KBM 23, SR 63, P<0.001, CI: -1.63--0.66).

Errors across minor, moderate, and major potential patient harm (PPH) types were found to be significantly higher with SR for both simple and complex tasks: major PPH simple tasks (KBM 2, SR 29, P<0.001, CI: -1.09--0.46) and complex tasks (KBM 11, SR 21, P=0.005, CI: -0.80--0.14), moderate PPH simple tasks (KBM 0, SR 13, P=0.008, CI: -0.71--0.11), minor PPH simple tasks (KBM 7, SR 33, P=0.002, CI: -1.11--0.34) and complex tasks (KBM 9, SR 34, P<0.001, CI: -2.40--1.43). The exception was moderate PPH errors during complex tasks, where there was no significant difference observed (KBM 3, SR 8, P=0.083, CI: -0.17--0.03).

There was no association found between the number of errors observed and participant clinical role, skill or period of experience with the EHR or SR. The introduction of interruptions made no statistically significant difference to the number of errors for complex tasks and simple tasks using SR.

3.1.4 Discussion

The study's results show statistically significant differences in the outcome of using KBM compared with SR, in that SR created documentation took longer and was associated with a higher number of errors. These results do not appear to be influenced by the degree of EHR experience, SR training, or seniority. The overall level of errors observed with EHR use (KBM & SR) is potentially of concern. However, what is an acceptable level of errors for EHR documentation is beyond the scope of this research.

The poor results for SR within the study may have been due to a number of factors including cognitive limitations and system integration:

Cognitive limitations

It is possible that the poor results observed by SR during the experiment were, as Shneiderman proposes, that the SR utilises the same cognitive resources of short-term and working memory as problem solving and recall. Conversely, the hand-eye coordination required for KBM uses different and additional resources leaving problem solving and recall resources unaffected.[1] This fundamental difference in cognitive processing between the two modalities is a potential explanation for poor SR performance, in particular while undertaking complex tasks.

System integration

Far more integration/system errors occurred while tasks were being undertaken with SR. These included some high potential patient harm errors such as misrecognition of words (e.g. 'follow' instead of 'FLU', 'fractured' instead of 'ruptured'). Similar SR technology related failures have been identified as a contributing factor in several actual patient harm cases, confining the need for close scrutiny of documentation created with the assistance of SR.[2]

It is possible that some of the integration errors observed within this study may have been unique to the specific implementation used in this study and might be avoided through a redesign or reconfiguration of inter-system communication methods.

3.1.5 Conclusion

The search for the safest and most efficient method of clinical documentation remains a work in progress. The use of SR to drive interactive clinical documentation in the EHR requires careful evaluation since current generation implementations may require significant development before they are adequately safe and effective.

Given the ubiquitous nature of electronic records in healthcare, and the substantial cost of documentation in terms of clinician time and patient safety, the foundational act of interacting with an electronic record requires far closer attention, and substantially higher research priority, than it currently receives.

3.2 Article contributorship statement

TH, EC, and Farah Magrabi (FM) conceived the study and its design. TH conducted the research, the primary analysis, and the initial drafting of the paper. EC and FM contributed to the analysis and drafting of the paper, and TH, EC, and FM approved the final manuscript. TH is the corresponding author.

3.3 Paper 2 - Efficiency and safety of speech recognition for documentation in the electronic health record

Journal of the American Medical Informatics Association, Volume 24, Issue 6, 1 November 2017, Pages 1127–1133, https://doi.org/10.1093/jamia/ocx073

Journal of the American Medical Informatics Association, 24(6), 2017, 1127–1133
doi: 10.1093/jamia/ocx073
Advance Access Publication Date: 27 July 2017
Research and Applications

Research and Applications

Efficiency and safety of speech recognition for documentation in the electronic health record

Tobias Hodgson, Farah Magrabi, and Enrico Coiera

Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, North Ryde, Australia

Corresponding Author: Tobias Hodgson, Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, L6 75 Talavera Rd, North Ryde, NSW 2109, Australia. E-mail: tobias.hodgson@hdr.mq.edu.au. Phone: +61 2 9850 2496

Received 2 February 2017; Revised 12 June 2017; Accepted 17 June 2017

ABSTRACT

Objective: To compare the efficiency and safety of using speech recognition (SR) assisted clinical documentation within an electronic health record (EHR) system with use of keyboard and mouse (KBM).
Methods: Thirty-five emergency department clinicians undertook randomly allocated clinical documentation tasks using KBM or SR on a commercial EHR system. Tasks were simple or complex, and with or without interruption. Outcome measures included task completion times and observed errors. Errors were classed by their potential for patient harm. Error causes were classified as due to IT system/system integration, user interaction, comprehension, or as typographical. User-related errors could be by either omission or commission.
Results: Mean task completion times were 18.11% slower overall when using SR compared to KBM ($P = .001$), 16.95% slower for simple tasks ($P = .050$), and 18.40% slower for complex tasks ($P = .009$). Increased errors were observed with use of SR (KBM 32, SR 138) for both simple (KBM 9, SR 75; $P < 0.001$) and complex (KBM 23, SR 63; $P < 0.001$) tasks. Interruptions did not significantly affect task completion times or error rates for either modality.
Discussion: For clinical documentation, SR was slower and increased the risk of documentation errors, including errors with the potential to cause clinical harm compared to KBM. Some of the observed increase in errors may be due to suboptimal SR to EHR integration and workflow.
Conclusion: Use of SR to drive interactive clinical documentation in the EHR requires careful evaluation. Current generation implementations may require significant development before they are safe and effective. Improving system integration and workflow, as well as SR accuracy and user-focused error correction strategies, may improve SR performance.

Key words: patient safety, electronic health record, speech recognition, documentation, medical errors

BACKGROUND AND SIGNIFICANCE

Electronic health records (EHRs) are rapidly becoming mandatory for clinical documentation worldwide, and can result in improved documentation completeness and accuracy.[1] The use of EHRs is, however, often associated with increased documentation time for clinicians compared to use of paper records.[2] Speech recognition (SR) systems are seen as an alternative input modality that may be faster and more acceptable to clinicians than the use of keyboard and mouse (KBM) alone.

In the clinical dictation setting, such as radiology results reporting, SR can decrease overall document turnaround time compared to transcription services.[3] SR is also considered a viable option for clinicians who cannot touch-type or are untrained in EHR use.[4,5]

SR is an increasingly common input modality across a range of consumer devices, from smartphones to refrigerators,[6] and has

become the standard method for documentation in specialty areas such as radiology results reporting. Given the significant time clinicians spend on the ubiquitous task of clinical documentation, it is surprising that the benefits of SR for EHR documentation and navigation have been relatively unexplored. Our recent review of the literature identified only a few studies that compared SR to dictation and transcription for clinical report creation. None explored the use of SR for general input and navigation of the EHR, and very few directly compared SR to the use of KBM, the most common input modality for EHR documentation.[3]

OBJECTIVE

The objective of this study was to compare the impact of using SR on the efficiency and safety of EHR documentation compared to KBM alone, using a common commercial EHR in a controlled experimental setting. The 2 input modalities were compared for efficiency (measured by time to complete tasks) and safety (measured by occurrence of documentation errors). The potential influence of task complexity and interruptions on documentation performance was also examined.

MATERIALS AND METHODS

A within-subject experimental study was undertaken with emergency department (ED) physicians, with each assigned 8 standardized clinical documentation tasks using a commercial EHR. Participants navigated the EHR and documented patient information for simulated patients. The order of task completion was allocated randomly, with half of the tasks assigned to KBM and half to SR.

The 8 documentation tasks were representative of those commonly undertaken within an EHR by ED physicians and included patient assignment, patient assessment, diagnosis, orders, and patient discharge. Tasks were chosen in consultation with senior ED clinicians, who did not further participate in the trials. All simulated patients had active records available in the experimental version of the standard ED EHR (see Supplementary Appendix A).

To allow for variation in task complexity, 4 of the 8 tasks were designed to be simple and 4 complex. Complexity was measured by the number of subtasks, with the simple tasks having 2 subtasks and complex tasks having 4. Given that interruptions are commonplace in clinical settings and can contribute to documentation error, 4 of the 8 tasks included a randomly assigned interruption condition.[7,8] Interrupts were generated by a popup with a multiple-choice question taken from an Australasian College for Emergency Medicine fellowship exam practice set, and occurred at the same predefined stage of task completion. A similar interrupt was generated for simple and complex tasks in the KBM and SR conditions (see Figure 1).

The clinical software used for the experiment was the Cerner Millennium suite with the FirstNet ED component (v2012.01.30) and Nuance Dragon Medical 360 Network Edition UK (version 2.0, 12.51.200.072) SR software. Both were configured to replicate the operation of the EHR that participants used daily. All user actions, down to individual keystrokes, were automatically logged with recording software. Session EHR screens and audio were also separately recorded with a High-Definition Multimedia Interface capture device (see Supplementary Appendices B and C).

Thirty-five participants volunteered from 3 urban teaching hospitals in Sydney, Australia, from an eligible population of approximately

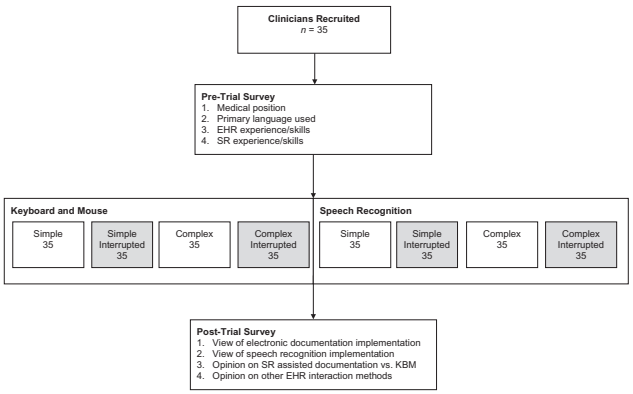


Figure 1. Experiment conceptual design.

100 ED clinicians. To be eligible, subjects must have previously completed training in the EHR system, including specific SR training (EHR 4 h, SR 2 h). Clinicians were excluded if they had a pronounced speech impediment or physical disability that might affect system use.

It was estimated that a sample size of 27 clinicians would be sufficient to test for differences in time efficiency and error rates when using a t-test with a significance level of 0.05 and power of 0.95. Calculations were performed using G*Power (v3.1).[9]

Participants self-assessed their prior experience with clinical documentation using the EHR and the 2 input methods through a pre-study questionnaire. A brief exit questionnaire solicited opinions on the technologies used and their implementation within their workplace. The study was approved by the university and participating hospitals' ethics committees. The trials took place over 2 months, commencing March 2015.

OUTCOME MEASURES

The efficiency of each input modality was measured by the time taken to complete a documentation task, including separate measurements for all subtasks. Task completion times excluded time spent during any interruptions. The time required to launch the dictation dialog box was also deducted, as this time was unrelated to either user behavior or SR time but was associated with the way in which the SR system was locally integrated with the EHR.

The safety of documentation performance was assessed by the number of errors observed. Each observed error was assigned a label in 3 categories by 2 reviewers (TH and DL, see Supplementary Appendix D for the trial errors observed and their assigned labels).

1. Potential for Patient Harm (PPH): This was an assessment of risk that an error had a major, moderate, or minor impact on patient outcomes, based on the scale within the US Food and Drug Administration 2005 guidance document.[10]
2. Error Type: The nature of the error was separated into 3 classes: (A) integration/system: associated with technology (including software, software integration, and hardware); (B) user: associated with user action; and (C) comprehension: related to comprehension (eg, user adds words to or omits words from the prescribed task). Errors could be assigned to > 1 class within this label set.
3. User Error Type: Where user errors occurred, they were assigned 1 of 2 additional labels: (A) omission: errors that occurred when

a participant failed to complete an assigned task, and (B) commission: errors that occurred when participants incorrectly executed an assigned task.

The labels for Error Type were not mutually exclusive, and some errors could have multiple labels assigned. Minor typographical errors, such as missing full stops or incorrect capitalization, were treated as a discrete category, as they had no potential for harm and could not be easily assigned a Type category.

Definitions for these error classes are contained in Supplementary Appendix E. The process for allocating observed errors to class and the inter-rater agreement are presented in Supplementary Appendix D. Inter-rater agreement was calculated using Cohen's kappa.[11] Initial observed agreement was robust across all three categories: (1) 81.48% agreement, $\kappa = 0.694$, $P < .001$, 95% confidence interval [CI], 0.476–0.912. (2) 88.89% agreement, $\kappa = 0.800$, $P < .001$, 95% CI, 0.590–1.010. (3) 85.19% agreement, $\kappa = 0.778$, $P < .001$, 95% CI, 0.584–0.971. All discrepancies between the reviewers were resolved through discussion.

Statistical comparisons were made for efficiency and safety outcome variables on equivalent tasks using both KBM and SR, ie, simple or complex tasks. Aggregate data across all task types were reported, but heterogeneity in task type precluded statistical testing. Since the study data do not follow normal distribution, only nonparametric statistical tests, including Wilcoxon signed rank and Mann-Whitney tests, were undertaken using IBM SPSS Statistics v24.0.0.0. For those statistical tests that ranked paired observations, comparisons were possible only where values for both input modalities were available. In cases where a task had no value for 1 input modality (such as a missed or incomplete task), the pair was excluded.

RESULTS

There were 19 female and 16 male participants, who identified themselves predominantly as senior clinicians (24/35). Most rated their SR skills lower than their EHR skill level, with a mean of 3.9/5 for KBM and 2.7/5 for SR (on a scale from 1 = very poor to 5 = excellent). Many participants had <1 year of SR experience (19/35) while few had <1 year EHR experience (7/35).

Efficiency

Across all 4 experimental conditions (simple task KBM, simple task SR, complex task KBM, and complex task SR), there was no association found between task completion time and clinical role ($P = .100$–$.588$), skill level ($P = .073$–$.452$), or period of experience with the EHR or SR ($P = .303$–$.717$). Interruptions had no effect on task completion times for either input modality across both task types (see Supplementary Appendix F).

Overall, complex tasks took 52.82% longer to complete than simple tasks (mean: complex 185.63 s, simple 121.47 s; $P < .001$; 95% CI, 50.55–75.34). Clinical documentation took significantly longer to complete using SR compared to KBM, with mean time to complete all tasks 18.11% longer when using SR (KBM 140.09, SR 165.46; $P = .001$; 95% CI, 9.87–33.91). This significant difference in mean task completion time held for both simple (KBM 112.38, SR 131.44; $P = 0.050$; 95% CI, −0.07, 30.50) and complex (KBM 170.48, SR 201.84; $P = .009$; 95% CI, 9.61–47.73) tasks (see Table 1 and Figure 2).

Safety

A total of 170 errors was observed, with a significant difference in the number of errors (excluding typographical errors) via input modality (KBM 32, SR 138) (see Figure 3). This increase in errors observed via SR held for both simple (KBM 9, SR 75; $P < .001$; 95% CI, 1.50–2.50) and complex (KBM 23, SR 63; $P < .001$; 95% CI, 0.50–1.50) tasks. There were many typographical errors observed with both input modalities (KBM 213, SR 252), and no significant difference in typographical error frequency by input modality for either simple (KBM 142, SR 133; $P = .345$; 95% CI, −1.00, 0.50) or complex (KBM 71, SR 119; $P = .600$; 95% CI, −0.50, 0.50) tasks was found.

There was no association found between number of errors observed and participant clinical role, skill, or period of experience with the EHR or SR. The introduction of interruptions made no statistically significant difference in the number of errors for complex tasks and simple tasks using SR. However, there was a significant difference in errors observed for simple tasks using KBM (KBM 9, KBMI 7; $P = .018$; 95% CI, 0.00–1.00).

Potential for patient harm

There were significant increases in the occurrence of all classes of PPH errors when using SR across both task types: major PPH simple task (KBM 2, SR 29; $P < .001$; 95% CI, 0.50–1.00), and complex task (KBM 11, SR 21; $P = .005$; 95% CI, 0.00–0.50), moderate PPH simple task (KBM 0, SR 13; $P = .008$; 95% CI, 0.00–0.50), minor PPH simple task (KBM 7, SR 33; $P = .002$; 95% CI, 0.50–1.00), and complex task (KBM 9, SR 34; $P < .001$; 95% CI, 1.50–2.50). The exception was moderate PPH errors during complex tasks, where there was no significant difference (KBM 3, SR 8; $P = .083$; 95% CI, 0.00–0.00) (see Table 2).

Error type

There were significant increases in the occurrence of integration/system, user, and comprehension errors while using SR across both task types: integration/system errors simple task (KBM 0, SR 56; $P < .001$; 95% CI, 1.00–2.00) and complex task (KBM 2, SR 36; $P < .001$; 95% CI, 1.00–2.50), user errors simple task (KBM 5, SR 18; $P = .007$; 95% CI, 0.00–0.50) and complex task (KBM 17, SR 26; $P = .002$; 95% CI, 0.00–0.50), and comprehension errors complex task (KBM 4, SR 17; $P < .001$; 95% CI, 1.00–1.50), with the exception of comprehension errors during simple tasks, where no significant difference was found (KBM 4 SR 5; $P = .317$; 95% CI, 0.00–0.00) (See Table 2).

User error type

There were significant increases in the occurrence of omission errors via SR across both task types: simple task (KBM 2, SR 14; $P = .001$; 95% CI, 0.00–0.50) and complex task (KBM 6, SR 6; $P = .005$; 95% CI, 0.00–0.50). However, there were no significant differences in the occurrence of commission errors via either input modality: simple task (KBM 7, SR 5; $P = .719$; 95% CI, 0.00–0.00) and complex task (KBM 15, SR 21; $P = .201$; 95% CI, 0.00–0.50) (see Table 2).

DISCUSSION

This is the first study, to our knowledge, that compares the impact of input modality on the safety and efficiency of clinical documentation using an EHR in a controlled experimental environment. The study's results show statistically significant differences in outcomes when using KBM compared to SR, with SR created records taking longer and associated with a higher error rate. While these results do not appear to be influenced by the degree of EHR or SR experience,

Table 1. Efficiency summary

Task Type	N	Mean task completion time (s)	Max task completion time (s)	Min task completion time (s)	Difference in completion time SR vs KBM (%)	Z	Wilcoxon P-value	95% CI
KBM – simple and complex	65	140.09	257.63	49.38				129.56, 151.45
SR – simple and complex	60	165.46	327.45	71.67				150.56, 180.37
Total KBM vs total SR	58				118.11	−3.302	0.001	9.87, 33.91
Simple task KBM	34	112.38	214.99	49.38				103.15, 122.42
Simple task SR	31	131.44	285.91	71.67				117.01, 146.94
Simple task KBM vs simple task SR	31				116.95	−1.960	0.050	−0.07, 30.50
Complex task KBM	31	170.48	257.63	110.31				157.26, 183.80
Complex task SR	29	201.84	327.45	138.87				185.71, 219.26
Complex task KBM vs complex task SR	27				118.40	−2.643	0.009	9.61¸47.73
Tasks with interruptions								
Simple task KBM with interrupt	34	112.22	247.26	60.40				101.22, 124.77
Simple task KBM vs simple task KBM with interrupt	34				−99.85	−0.710	0.478	−13.19, 7.69
Simple task SR with interrupt	32	134.41	250.20	67.69				119.78, 149.08
Simple task SR vs simple task SR with interrupt	29				102.26	−0.465	0.642	−19.75, 28.12
Complex task KBM with interrupt	31	186.65	355.90	102.94				168.91, 207.55
Complex task KBM vs complex task KBM with interrupt	29				109.49	−0.465	0.642	−3.39, 33.60
Complex task SR with interrupt	31	216.45	345.70	133.13				196.82, 238.76
Complex task SR vs complex Task SR with interrupt	26				107.24	−1.664	0.096	−1.96, 49.12

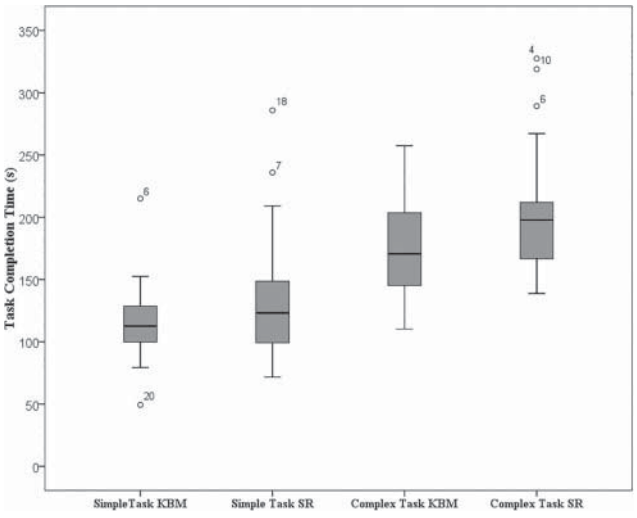


Figure 2. Boxplot of task completion time for simple and complex tasks via input modality.

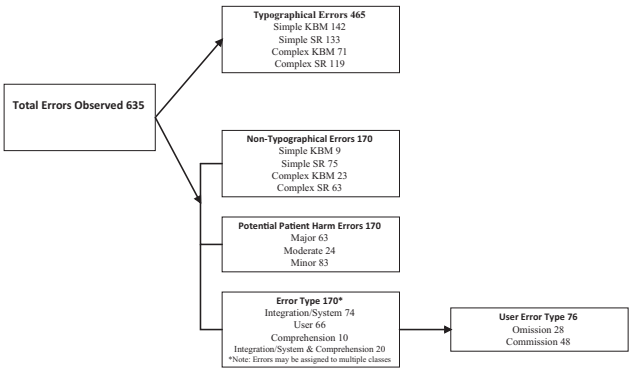


Figure 3. Error framework: overview of the breakdown of errors by class.

skill level, or seniority, it is possible that results might be different with changes in task type, participants, setting, or technology. For the tasks undertaken in this study, creating EHR clinical documentation with the assistance of SR was significantly slower (18.11%) than using KBM alone, and this difference held for both simple (16.95%) and complex (18.40%) documentation tasks.

These results are different from those of a recent observational study that found no difference in charting time when a voice driven EHR at one site was compared to a KBM driven EHR at another.[12] However, the study sample size was too small to determine if this result was statistically significant, and the EHRs at these sites were different and may themselves have accounted for the time difference. Further, patient acuity was different at the 2 sites, resulting in different admission rates and case mix. Consequently, it is hard to draw

Table 2. Error summary

	Non-interrupts					Interrupts					
						KBM vs KBMI			SR vs SRI		
	KBM	SR	Z	Wilcoxon *P*-value	95% CI	Z	*P*-value	95% CI	Z	Wilcoxon *P*-value	95% CI
Total errors observed	**245**	**390**									
Non-typographical	***32***	***138***									
Simple	9	75	−4.370	0.000	1.50, 2.50	−2.373	0.018	0.00, 1.00	−1.115	0.265	−0.50, 1.50
Complex	23	63	−4.085	0.000	0.50, 1.50	−0.294	0.769	0.00, 0.50	−1.086	0.278	−0.50, 1.00
Potential patient harm	***32***	***138***									
Major	**13**	**50**									
Simple	2	29	−3.616	0.000	0.50, 1.00	−2.496	0.013	0.00, 0.50	−1.099	0.272	−0.50, 0.50
Complex	11	21	−2.799	0.005	0.00, 0.50	−1.425	0.154	0.00, 0.50	−0.317	0.751	−0.50, 0.50
Moderate	**3**	**21**									
Simple	0	13	−2.636	0.008	0.00, 0.50	−1.732	0.083	0.00,0.00	−0.474	0.635	0.00, 0.00
Complex	3	8	−1.732	0.083	0.00, 0.00	−1.732	0.083	0.00, 0.00	−2.648	0.008	0.00, 0.50
Minor	**16**	***67***									
Simple	7	33	−3.101	0.002	0.50, 1.00	−0.686	0.493	0.00, 0.00	−0.198	0.843	−0.50, 0.50
Complex	9	34	−4.637	0.000	1.50, 2.50	−1.809	0.070	0.00, 0.50	−2.532	0.011	−1.50, 0.50
Error type	***32***	***158***									
Integration/system	**2**	**92**									
Simple	0	56	−4.598	0.000	1.00, 2.00	−1.414	0.157	0.00, 0.00	−1.792	0.073	−1.00, 0.00
Complex	2	36	−4.460	0.000	1.00, 2.50	−0.577	0.564	0.00, 0.00	−1.008	0.314	−1.00, 0.50
User errors	**22**	**44**									
Simple	5	18	−2.681	0.007	0.00, 0.50	−2.194	0.028	0.00, 0.50	−1.397	0.162	0.00, 0.50
Complex	17	26	−3.133	0.002	0.00, 0.50	−2.675	0.007	0.00, 0.50	−0.210	0.834	−0.50, 0.50
Comprehension	**8**	**22**									
Simple	4	5	−1.000	0.317	0.00, 0.00	−0.707	0.480	0.00, 0.00	−2.207	0.027	0.00, 0.50
Complex	4	17	−4.361	0.000	1.00, 1.50	0.000	1.000	0.00, 0.00	−3.006	0.003	−1.00, 0.00
User error type	***30***	***46***									
Omission	**8**	**20**									
Simple	2	14	−3.207	0.001	0.50, 0.00	−1.265	0.206	0.00, 0.00	−2.138	0.033	−0.50, 0.00
Complex	6	6	−2.840	0.005	0.00, 0.50	−0.632	0.527	0.00, 0.00	−2.500	0.012	−0.50, 0.00
Commission	**22**	**26**									
Simple	7	5	−0.359	0.719	0.00, 0.00	−1.186	0.236	0.00, 0.00	−1.588	0.112	0.00, 0.50
Complex	15	21	−1.279	0.201	0.00, 0.50	−0.188	0.851	0.00, 0.50	−0.471	0.637	−0.50, 0.00
Typographical	***213***	***252***									
Simple	142	133	−0.944	0.345	−1.00, 0.50	−4.702	0.000	−2.50, 1.50	−1.038	0.299	−1.00, 0.50
Complex	71	119	−0.525	0.600	−0.50, 0.50	−4.511	0.000	−3.00, 1.50	0.303	0.303	−0.50, 1.00

Bold values indicate general totals and italic bold corresponds to category totals.

any conclusions from this study, and comparison with the controlled experiment reported here is not possible.

Interestingly, task interruptions did not significantly increase task completion times in our study. This is consistent with another controlled experiment, where interruptions did not affect electronic prescribing tasks.[13] A previous observational study on the impact of interruptions on ED physicians' task time showed a reduction in task time following interruptions, possibly because physicians hurry to complete tasks after being interrupted when they are under time pressure.[14] Our results suggest that documentation tasks might not be easily time compressed, and that interruption penalties might therefore be incurred with other tasks associated with patient care.

In this study, more errors occurred when tasks were completed using SR than KBM alone, and this held for both simple and complex tasks.

The occurrence of serious errors with major PPH was significantly greater using SR (simple task: KBM 2, SR 29; complex task: KBM 11, SR 21) (Table 3). Any increase in these errors that could result in death or serious injury to patients requires further examination.

Far more integration/system errors (such as network transmission delays causing SR commands to either not complete or go to the incorrect location in a chart or the wrong chart) occurred while tasks were being undertaken with SR. These included some highly potential patient harm errors, such as misrecognition of words (eg, "follow" instead of "FLU," "fractured" instead of "ruptured"). It is possible that some of these integration errors could be avoided through a redesign or reconfiguration of inter-system communication methods.

Documentation error rates were high independent of input modality, and if replicated in real-world situations would raise questions about the acceptable baseline level of errors for an EHR system. Recent real-world studies have indeed shown that clinical documentation accuracy can be poor, with one study finding accurate documentation rates of 54.4% via paper and 58.4% with

Table 3. Major potential patient harm errors

Error observed	Incorrect patient	Incorrect blood glucose level	Incorrect test/order collection date entered	Data lost during text transfer (no EHR record created)	Data entered in incorrect EHR field	Section of EHR missed	Incorrect significant word entered (diagnostic)	Significant word entered (diagnostic)	Misrecognition of word by SR	Element of EHR down, eg, vitals	Total errors
Simple task KBM	0	0	0	0	1	0	1	0	0	0	2
Simple task SR	0	0	0	2	1	0	0	0	26	0	29
Complex task KBM	1	3	3	0	2	1	0	0	0	1	11
Complex task SR	0	12	6	0	0	1	0	1	1	0	21

EHRs.[15] Our results thus support calls to improve the quality of clinical documentation independent of modality of input.

There continue to be rapid changes in SR technologies; a recent study reports SR systems that now exceed human performance for conversational speech.[16] One recent study examined the potential for EHR transcription using SR together with natural language processing and found potential time and usability benefits for electronic documentation.[17] These results are in accord with our recent systematic review that identified benefits (and risks) of using SR for clinical documentation.[3]

SR as a use model is typically associated with dictation tasks where users are simply creating large volumes of text. Dictation time benefits may be realized in part because the user interaction and screen navigation tasks associated with electronic documentation are not present, unlike the real-world use cases within this study.

SR systems have carved out a role in specific areas of clinical documentation, such as report dictation in radiology,[3] but there has been little attention paid to the potential risks and benefits of using SR as a primary mode of data entry and system navigation in the EHR. The results of this study, which was designed to replicate real world tasks and used a working clinical system with real clinical users, give some pause for thought. SR appears to reduce efficiency by taking more time than KBM, and also appears to increase the rates of error, both minor and serious.

Some of the errors and inefficiencies appear to be inherent in current-generation SR engines and their fit to the task of clinical documentation within EHRs. EHRs are designed with KBM input in mind, and simply expecting SR to directly replace KBM in such a design may be inappropriate. Designing EHRs that are purpose-built for SR may yield different workflows and design constructs, and consequently different results. It is possible that at least some of these weaknesses with SR can be mitigated with system redesign.

The use of SR to drive interactive clinical documentation in the EHR thus requires careful evaluation, and it cannot be assumed that it is a safe and effective replacement for KBM in this particular setting. Current generation implementations may require significant changes before they are considered safe and effective. Improving system integration and workflow, as well as further developing SR accuracy and enhancing the capacity of user interface design to allow clinicians to detect and correct errors, may improve SR performance.

LIMITATIONS

Several factors could hinder the results of this study being generalized to other clinical settings or information systems. The study used a routine and standardized version of a widely used commercial clinical record system for EDs, integrated with a common commercial clinical SR system. However, other EHR and SR systems might differ in their individual performance, and different approaches to integrating the two could also lead to varying results.

SR performance may be affected by extrinsic factors, such as microphone quality, background noise level, and user accent. Equally, the tasks created for this study were intended to be representative of typical clinical documentation work in an ED, but different tasks in other settings could yield different outcomes. For example, dictation of investigation reports at high volume by expert clinicians might yield better time performance and recognition rates, although our previous review did not identify this.[3]

The lack of any impact of interruptions should be generalized only cautiously, given that the impact of interruptions is a complex

phenomenon, and many different variables must interact for interruptions to have a negative effect on performance.

While the study was not specifically designed to test for the effects of clinical role, skill, or period of experience on outcome variables and was underpowered to test for these effects individually, it was possible to demonstrate a significant absence of correlation when all SR and KBM tasks were pooled. While there were no statistically significant differences in performance observed based on experience, skill, or seniority, the fact that participants volunteered for this study may have introduced a recruitment bias. All participants had completed training in the use of both the EHR and SR, and it is quite possible that results would be different, and potentially worse, if users had no such training.

CONCLUSION

The search for the safest and most efficient method of clinical documentation remains a work in progress. Emerging technologies such as mobile devices, touch pads, pen control, head-mounted displays, virtual reality, and wearable technology all have different affordances and are likely to work better in some contexts and provide varying results over a wide range of tasks.[18] Given the ubiquitous nature of electronic records in healthcare, and the substantial cost of documentation in terms of clinician time and patient safety, the foundational act of interacting with an electronic record requires far closer attention, and substantially higher research priority, than it currently receives.

ACKNOWLEDGMENTS

David Lyell performed the second reviewer functions for error classification. Thanks are offered to the staff and management teams at participating hospitals, in particular Dr Richard Paoloni and Dr Marty Sterrett, who both assisted greatly with study design and trial session coordination.

FUNDING

This work was supported by the National Health and Medical Research Council Centre for Research Excellence in eHealth (APP1032664).

COMPETING INTERESTS

The authors declare that they have no competing interests.

CONTRIBUTORS

TH, EC, and FM conceived the study and its design. TH conducted the research, the primary analysis, and the initial drafting of the paper. EC and FM contributed to the analysis and drafting of the paper, and TH, EC, and FM approved the final manuscript. TH is the corresponding author.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

REFERENCES

1. Coiera E. *Guide to Health Informatics*, 3rd ed. London: CRC Press; 2015.
2. Chiang MF, Read-Brown S, Tu DC, *et al*. Evaluation of electronic health record implementation in ophthalmology at an academic medical center (an American Ophthalmological Society thesis). *Trans Am Ophthalmol Soc*. 2013;111:70
3. Hodgson T, Coiera E. Risks and benefits of speech recognition for clinical documentation: a systematic review. *J Am Med Inform Assoc*. 2016;23(e1):e169–e79.
4. Shea S, Hripcsak G. Accelerating the use of electronic health records in physician practices. *New Engl J Med*. 2010;362(3):192–95.
5. Communications N. *Speech Recognition Accelerating the Adoption of Electronic Health Records* (white paper). Nuance Communications; 2012.
6. Wolpin S. Samsung's new refrigerator is also a giant tablet. 2016. http://www.techlicious.com/blog/samsung-refrigerator-family-hub-touchscreen//. Accessed November 19, 2016.
7. Coiera EW, Jayasuriya RA, Hardy J, Bannan A, Thorpe MEC. Communication loads on clinical staff in the emergency department. *Med J Australia*. 2002;176(9):415–18.
8. Coiera E. The science of interruption. *BMJ Qual Safety*. 2012;21(5):357–60.
9. Faul F, Erdfelder E, Lang A-G, Buchner A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39(2):175–91.
10. Food U, Administration D. Guidance for the content of premarket submissions for software contained in medical devices. *Center Devices Radiol Health*. 2005.
11. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70(4):213.
12. dela Cruz JE, Shabosky JC, Albrecht M, *et al*. Typed versus voice recognition for data entry in electronic health records: emergency physician time use and interruptions. *Western J Emerg Med*. 2014;15(4):541.
13. Magrabi F, Li SY, Day RO, Coiera E. Errors and electronic prescribing: a controlled laboratory study to examine task complexity and interruption effects. *J Am Med Inform Assoc*. 2010;17(5):575–83.
14. Westbrook JI, Coiera EW, Dunsmuir WTM, *et al*. The impact of interruptions on clinical task completion. *Qual Safety Health Care*. 2010;19(4):284–89.
15. Yadav S, Kazanji N, Narayan K, *et al*. Comparison of accuracy of physical examination findings in initial progress notes between paper charts and a newly implemented electronic health record. *J Am Med Inform Assoc*. 2017;24(1):140–44.
16. Xiong W, Droppo J, Huang X, *et al*. *Achieving Human Parity in Conversational Speech Recognition*. 2016;arXiv preprint arXiv:1610.05256.
17. Kaufman DR, Sheehan B, Stetson P, *et al*. Natural language processing–enabled and conventional data capture methods for input to electronic health records: a comparative usability study. *JMIR Med Inform*. 2016;4(4):e35.
18. PricewaterhouseCoopers. *The Wearable Future. Consumer Intelligence Series*. 2014. Accessed October 2014.

3.4 Chapter references

1. Shneiderman B. The limits of speech recognition. Communications of the ACM. 2000;43(9):63-5.

2. Topaz M, Schaffer A, Lai KH, Korach ZT, Einbinder J, Zhou L. Medical Malpractice Trends: Errors in Automated Speech Recognition. Journal of medical systems. 2018;42(8):153.

3.5 Paper 2 references

1. Coiera E. Guide to Health Informatics (3rd Edition). 3rd ed. London: CRC Press; 2015.

2. Chiang MF, Read-Brown S, Tu DC, Choi D, Sanders DS, Hwang TS, et al. Evaluation of electronic health record implementation in ophthalmology at an academic medical center (an American Ophthalmological Society thesis). Transactions of the American Ophthalmological Society. 2013;111:70.

3. Hodgson T, Coiera E. Risks and benefits of speech recognition for clinical documentation: a systematic review. Journal of the American Medical Informatics Association. 2016;23(e1):e169-e79.

4. Shea S, Hripcsak G. Accelerating the Use of Electronic Health Records in Physician Practices. New England Journal of Medicine. 2010;362(3):192-5.

5. Communications N. Speech Recognition Accelerating the Adoption of Electronic Health Records (White Paper). 2012.

6. Wolpin S. Samsung's New Refridgerator IS Also a Giant Tablet. Techlicious.com2016 [cited 2016 19/11/2016]; Article]. Available from: http://www.techlicious.com/blog/samsung-refrigerator-family-hub-touchscreen/.

7. Coiera EW, Jayasuriya RA, Hardy J, Bannan A, Thorpe MEC. Communication loads on clinical staff in the emergency department. Medical Journal of Australia. 2002;176(9):415-8.

8. Coiera E. The science of interruption. BMJ quality & safety. 2012;21(5):357-60.

9. Faul F, Erdfelder E, Lang A-G, Buchner A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods. 2007;39(2):175-91.

10. Food U, Administration D. Guidance for the content of premarket submissions for software contained in medical devices. Center for Devices and Radiological Health. 2005.

11. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin. 1968;70(4):213.

12. dela Cruz JE, Shabosky JC, Albrecht M, Clark TR, Milbrandt JC, Markwell SJ, et al. Typed versus voice recognition for data entry in electronic health records: emergency physician time use and interruptions. Western Journal of Emergency Medicine. 2014;15(4):541.

13. Magrabi F, Li SY, Day RO, Coiera E. Errors and electronic prescribing: a controlled laboratory study to examine task complexity and interruption effects. Journal of the American Medical Informatics Association. 2010;17(5):575-83.

14. Westbrook JI, Coiera EW, Dunsmuir WTM, Brown BM, Kelk N, Paoloni R, et al. The impact of interruptions on clinical task completion. Quality and Safety in Health Care. 2010;19(4):284-9.

15. Yadav S, Kazanji N, Narayan K, Paudel S, Falatko J, Shoichet S, et al. Comparison of accuracy of physical examination findings in initial progress notes between paper charts and a newly implemented electronic health record. Journal of the American Medical Informatics Association. 2017;24(1):140-4.

16. Xiong W, Droppo J, Huang X, Seide F, Seltzer M, Stolcke A, et al. Achieving Human Parity in Conversational Speech Recognition. arXiv preprint arXiv:161005256. 2016.

17. Kaufman DR, Sheehan B, Stetson P, Bhatt AR, Field AI, Patel C, et al. Natural Language Processing–Enabled and Conventional Data Capture Methods for Input to Electronic Health Records: A Comparative Usability Study. JMIR Med Inform. 2016;4(4):e35.

18. PricewaterhouseCoopers. The Wearable Future. Consumer Intelligence Series [Internet]. 2014 October 2014.

4. Replication study: Evaluating the use of speech recognition within a commercial EHR system

4.1 Chapter background

The previous chapter described an experimental study that showed no significant benefits for SR as an input modality either in terms of efficiency or safety. These results were unexpected and have major implications for the use of SR for clinical documentation.

One of the limitations of the study was that some of the errors and time delays that occurred while documenting with SR may have been attributable to EHR and SR system integration factors specific to the experimental setup. These limitations include issues such as elements of a command not completing, and it is possible that SR may have performed significantly better had these integration factors been addressed. Therefore, a review of the technical set-up used during the experiment was conducted to identify any factors that might have biased the study results by hampering the performance of SR.

This analysis extended from system integration through to user workflows and interaction. Once these issues were addressed, a replication of the original experimental study was undertaken to test the performance under the revised technical conditions. (See Figure 5)

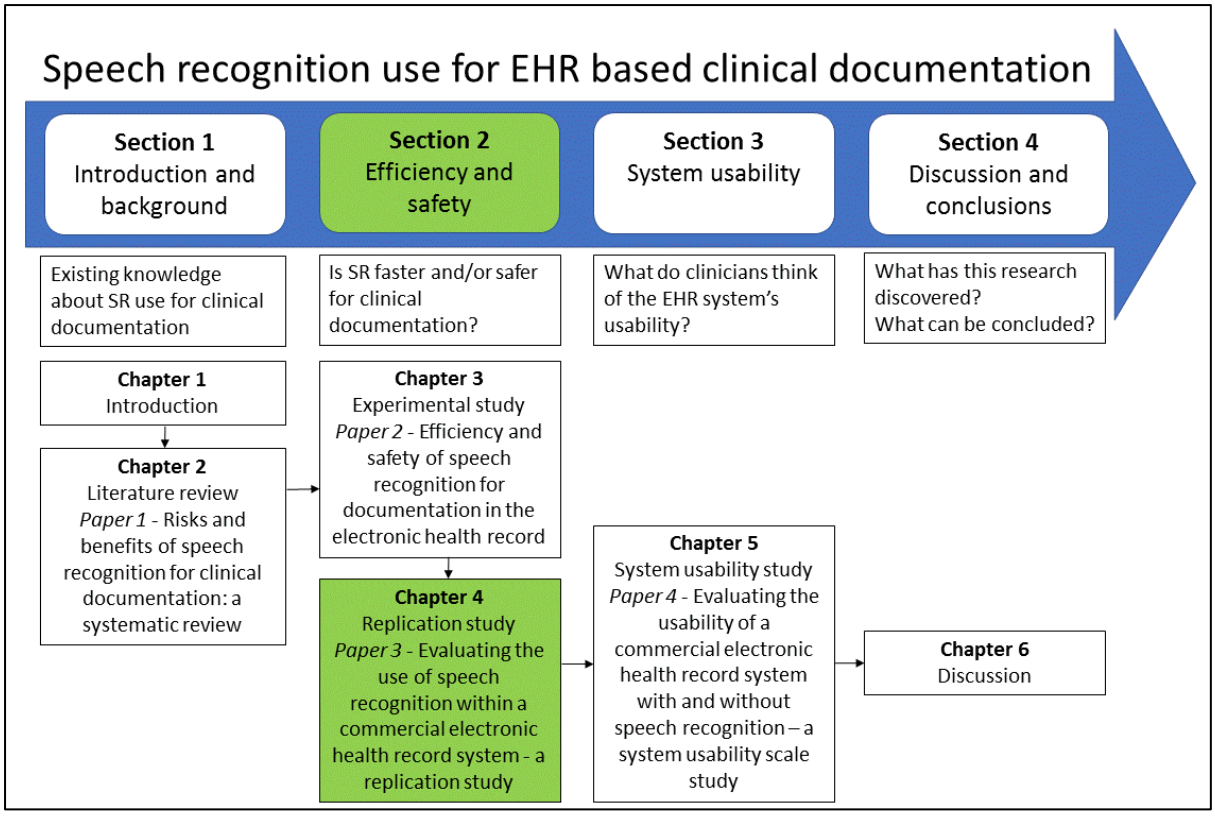


Figure 5: Thesis structure and position – Chapter 4

4.1.1 Hypothesis

This replication study tested the same primary hypothesis as the initial study and introduced a secondary hypothesis which compares the replication study results to those of the initial study:

Primary hypothesis

1 Speech recognition controlled electronic health record documentation, navigation and interaction is more efficient than keyboard and mouse, i.e. speech recognition assisted documentation is faster.

2 Speech recognition controlled electronic health record documentation, navigation and interaction is safer than with keyboard and mouse, i.e. fewer errors will be committed while documenting with speech recognition.

Secondary hypothesis

System optimisation will lead to improved efficiencies and increased safety during electronic health record documentation with speech recognition, i.e. clinical documentation will be faster

and fewer errors will be committed while documenting with speech recognition after system optimisation.

4.1.2 Method

The tasks and participants across both the initial experiment and replication study were directly comparable. The same types of tasks were undertaken, with the same number of participants (35) involved in both the initial experiment (E1) and the replication study (E2). A similar ratio of male to female participants was achieved (male E1 16, E2 18; female E1 19, E2 17). Thirteen participants were involved in both experiments (6 male, 7 female).

The methods used for this experiment (Experiment 2) were the predominately same as those within Experiment 1 (See chapter 3) with the exception that the number of tasks was reduced from eight to four by eliminating those cases in which an external interruption occurred. Updated versions of software for both the EHR and SR systems were also used where available.

Replication studies

Replications studies are typically undertaken to either validate prior findings or to examine the generality of initial findings.[1] There are numerous forms of replication studies available including exact, partial, quasi partial and quasi conceptual, each having their own utility depending on the purpose of the replication.[1, 2] (See Table 1)

Table 1: Forms of replication studies (from Coiera et al 2018[2])

Replication study type	Example study	Utility of replication study design
Exact (or close) replication	A laboratory study of the usability of a specific computerised physician order entry (CPOE) system is repeated in a different laboratory using the exact same protocol and system	High fidelity replications test the validity of an earlier study
Partial replication	A clinical trial of a CPOE system is repeated using the same system in a similar clinical environment, using an identical implementation strategy, and enrolling comparable groups of patients and clinicians	Modest level fidelity replications test the validity of an earlier study when it is not possible to undertake high fidelity studies

Replication study type	Example study	Utility of replication study design
Conceptual replication	Following a trial of a CPOE system in a clinical setting that shows mortality effects, the general hypothesis that all CPOE systems increase mortality rates is tested by using a different CPOE system, with a different implementation strategy, clinical setting and research subjects	Conceptual studies test the generalizability of past results, by sharing common hypotheses but using different clinical settings or methods
Quasi replication (partial)	To test the impact of implementation strategies on mortality rates after a particular CPOE is trialled, the same CPOE system is now tested in a comparable setting, but use a different implementation strategy	Quasi-replications seek to extend earlier experiments by including novel elements or hypotheses to build on the prior work, not just replicate it
Quasi replication (conceptual)	With evidence that CPOE use is associated with mortality changes, researchers test if this is generalizable to other system classes. They test the hypothesis that many clinical systems can affect mortality rates with an experiment using electronic health records and measuring mortality effects	The lowest fidelity form of replication, these studies help test the generality of prior results, but do not allow strong conclusions when their results conflict with earlier studies

This partial replication study used the same EHR and SR systems as the earlier experimental study, along with implementation and integration modifications to address identified limitations that may have impaired a fair comparison between input modalities. This study allowed for the validation of the initial experiment and the determination of the effect (if any) of experiment limitations that could be addressed by system reconfiguration.

A review of the technical set-up used for Experiment 1 was conducted to identify any factors that might have biased the study results by hampering the performance of SR. The analysis extended from low-level network integration through to user workflows and interaction.

To assist identification of such potential factors, an analysis of all the identified issues and errors in Experiment 1 was conducted to determine if a system issue might have contributed to the error. The human computer interaction framework of activity theory (AT) was used to assist with this process.

4.1.3 Activity theory

AT is a descriptive, explanatory and generative theory, providing concepts, explanations and design application respectively.[3] AT has been utilised while designing and assessing HCI systems for over two decades.[4]

The AT framework and checklist were utilised to assist with the identification of issues within the initial experiment.[5] Similar issues were grouped into one of seven categories: command reliability, system stability, patient safety, workflow usability, quality of data, typographical issues, and recognition and documentation issues. Discussions with clinicians, IT and general management were undertaken to determine which issues were of critical importance, those that were acceptable to leave, and those that would not be addressed due to constraints or limitations. Resolutions were then identified including: command changes, domain change, integration modifications, software option enabled, and revised software versions. (See 4.3 Paper 3 - Table 1)

The experiment protocol was then amended to include these revisions prior to running the replication study. Modifications were designed to improve command reliability both in execution and completeness, in system stability for a more robust solution, workflow refinements, data quality enhancements, reduction of typographical issues, along with improved recognition and documentation. (See 4.3 Paper 3 - Table 2)

4.1.4 Results

Complex tasks completion times were significantly slower to complete while using SR when compared to KBM (E2 16.94%; KBM 191.89s, SR 224.39s; P=0.001), replicating results from the initial study (E1 18.40%; KBM 170.48s, SR 201.84s; P=0.009).

Comparing the performance of subjects who participated in both experiments showed no significant difference in mean task completion times: simple tasks via KBM (E1 120.4s, E2 118.1s, P=0.308, CI: -20.39-11.90), simple tasks via SR (E1 129.3s, E2 129.5s, P=0.286, CI: -45.27-26.68), complex tasks via

KBM (E1 179.4s, E2 159.7s, P=0.059, CI: -37.30–2.66), and complex tasks via SR (E1 205.1s, E2 208.4s, P=0.814, CI: -31.64-61.58).

Errors (non-typographical) observed were significantly higher with SR compared to KBM for both simple (E2 KBM 3, SR 84, P<0.001, CI: -2.99--1.64) and complex tasks (E2 KBM 23, SR 53, P=0.001, CI: -1.32--0.39) again replicating earlier results (E1 simple KBM 9, SR 75, P<0.001, CI: -2.42--1.35 and complex KBM 23, SR 63, P<0.001, CI: -1.63--0.66). Repeat participants showed no significant difference in the number of errors observed between experiments 1 and 2.

Potential patient harm (PPH) errors were significantly greater with SR for both experiments across all simple tasks: minor (E1 KBM 7, SR 33, P=0.002, CI: -1.14--0.37 and E2 KBM 1, SR 42, P<0.001, CI: -1.89--0.69), moderate (E1 KBM 0, SR 13, P=0.008, CI: -0.74--0.11 and E2 KBM 1, SR 7, P=0.034, CI: -0.34--0.06) and major (E1 KBM 2, SR 29, P<0.001, CI: -1.11--0.46 and E2 KBM 1, SR 35, P<0.001, CI: -1.20--0.80). However, for complex tasks a significant difference was only observed for minor PPH (E1 KBM 9, SR 34, P<0.001, CI: -1.06--0.37 and E2 KBM 10, SR 44, P<0.001, CI: -1.40--0.60).

Typographical errors were reduced significantly in the replication study compared to the initial experiment (E1 465, E2 150, P<0.001, CI: 2.00-3.00).

4.1.5 Discussion

A series of modifications were made to the SR and EHR integration in order to minimize any potential bias in the original experimental set up towards KBM and to optimize the performance of SR. Several improvements were seen in the performance of SR, but these were insufficient to fundamentally change the overall results.

While there were no statistical differences in overall error rates (non-typographical) despite these technical improvements, the number of error types observed was reduced, with eleven types observed in Experiment 1 being eliminated.

As the variety of EHR documentation methods and technologies available to clinicians continues to grow there may be additional opportunities for SR to excel. Decision support systems also have the potential to mitigate some of the errors and problems observed within this study.[6-8]

4.1.6 Future study replication

Further replication of the experiments repeated in this thesis would be welcomed in order to both validate existing results and increase the generalisability of any conclusions drawn. This replication may take the form of conceptual replication with a different clinical setting/methods/participants, or quasi-replication, with a different implementation strategy or conceptually similar, such as testing alternative input modalities used within EHRs.

Some of the elements of the experiment configuration that may be changed in future replications include:

1) *Experiment settings* – Alternative integration configurations between the systems involved (EHR and SR).
2) *Systems* - These studies utilised standardised versions of a widely used commercial clinical record system for EDs integrated with a common commercial clinical SR system. Ideally, additional studies would be undertaken utilising systems from other vendors.
3) *Documentation tasks* – These studies focused on typical ED documentation tasks, while there are numerous others from within different medical realms yet to be assessed.
4) *Participants* – Users who have received alternative training or are from other medical fields may produce significantly different results than the participants of these studies.

4.1.7 Conclusion

This replication study added further evidence in support of the results of the initial experiment, with both primary hypotheses again disproved. SR controlled EHR documentation, navigation and

interaction was again found to be both slower and generated more errors than KBM based documentation.

System optimisation brought several improvements in the performance of SR, but these were insufficient to fundamentally change the overall results.

Replication studies are to be encouraged in health informatics to ensure that unusual or highly impactful single studies are not acted upon without careful effort to ensure that their findings are valid and indeed generalisable to other experiments and working settings. This would address possible over-dependence on the results of a single study and avoid basing additional work on unconfirmed foundations.

4.2 Article contributorship statement

TH, EC, and FM conceived the study and its design. TH conducted the research, the primary analysis, and the initial drafting of the paper. EC and FM contributed to the analysis and drafting of the paper, and TH, EC, and FM approved the final manuscript. TH is the corresponding author.

4.3 Paper 3 - Evaluating the use of speech recognition within a commercial electronic health record system - a replication study

Clinical Informatics Journal, Volume 9, Issue 2, May 2018, Pages 326-325, https://doi.org/10.1055/s-0038-1649509

Pages 59-68 of this thesis have been removed as they contain published material. Please refer to the citation above for details of the articles contained in these pages.

4.4 Chapter references

1. Nakagawa S, Parker TH. Replicating research in ecology and evolution: feasibility, incentives, and the cost-benefit conundrum. BMC biology. 2015;13(1):88.

2. Coiera E, Ammenwerth E, Georgiou A, Magrabi F. Does health informatics have a replication crisis? Journal of the American Medical Informatics Association. 2018.

3. Kaptelinin V, Nardi BA. Acting with technology: Activity theory and interaction design: Mit Press; 2006.

4. Valecha R, Kashyap M, Rajeev S, Rao R, Upadhyaya S. An Activity Theory Approach to Specification of Access Control Policies in Transitive Health Workflows. 2014.

5. Kaptelinin V, Nardi BA, Macaulay C. Methods & tools: The activity checklist: a tool for representing the "space" of context. interactions. 1999;6(4):27-39.

6. Castaneda C, Nalley K, Mannion C, Bhattacharyya P, Blake P, Pecora A, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. Journal of clinical bioinformatics. 2015;5(1):4.

7. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? Journal of biomedical informatics. 2009;42(5):760-72.

8. Jaspers MWM, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. Journal of the American Medical Informatics Association. 2011;18(3):327-34.

4.5 Paper 3 references

1. Hodgson T, Coiera E. Risks and benefits of speech recognition for clinical documentation: a systematic review. Journal of the American Medical Informatics Association. 2016;23(e1):e169-e79.

2. Hodgson T, Farah M, Coiera E. Efficiency and safety of speech recognition for documentation in the electronic health record. Journal of the American Medical Informatics Association. 2017.

3. Estimating the reproducibility of psychological science. Science. 2015;349(6251).

4. Freedman LP, Cockburn IM, Simcoe TS. The Economics of Reproducibility in Preclinical Research. PLOS Biology. 2015;13(6):e1002165.

5. Nakagawa S, Parker TH. Replicating research in ecology and evolution: feasibility, incentives, and the cost-benefit conundrum. BMC biology. 2015;13(1):88.

6. Coiera EA, E; Georgiou, A; Magrabi, F. Does health informatics have a replication crisis? Journal of the American Medical Informatics Association. 2018.

7. Engestrom Y. Activity theory as a framework for analyzing and redesigning work. Ergonomics. 2000;43(7):960-74.

8. Kaptelinin V, Nardi BA, Macaulay C. Methods & tools: The activity checklist: a tool for representing the “space” of context. interactions. 1999;6(4):27-39.

9. Food U, Administration D. Guidance for the content of premarket submissions for software contained in medical devices. Center for Devices and Radiological Health. 2005.

10. Castaneda C, Nalley K, Mannion C, Bhattacharyya P, Blake P, Pecora A, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. Journal of clinical bioinformatics. 2015;5(1):4.

11. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? Journal of biomedical informatics. 2009;42(5):760-72.

12. Jaspers MWM, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. Journal of the American Medical Informatics Association. 2011;18(3):327-34.

13. Gross MT. The Need for Replication Studies—Is It Really a Done Deal? : JOSPT, Inc. JOSPT, 1033 North Fairfax Street, Suite 304, Alexandria, VA 22134-1540; 1997.

5. System usability study: Evaluating the usability of speech recognition to create clinical documentation using a commercial EHR

5.1 Chapter background

In the previous two chapters the efficiency and safety of using SR as input modality for clinical documentation within an EHR was examined. This chapter explores the usability of the EHR system with the two input modalities (KBM or SR). (See Figure 6)

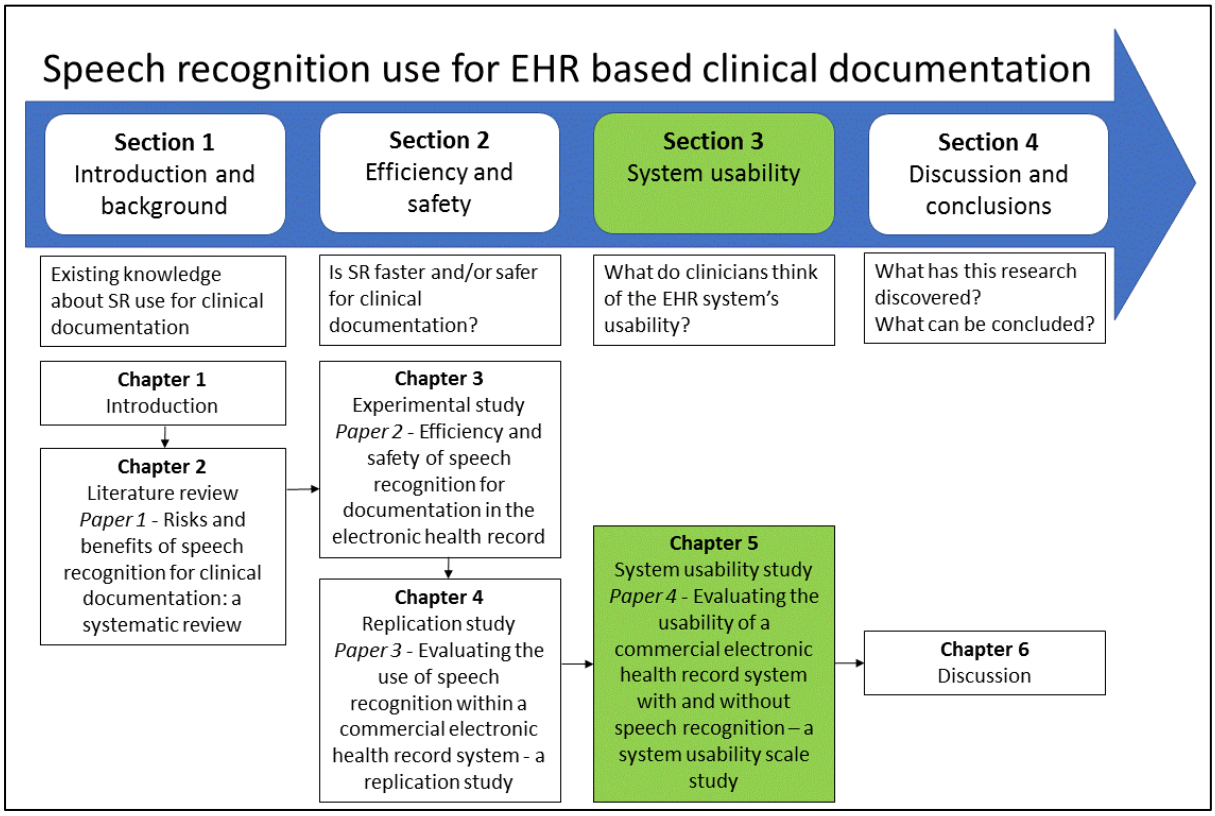


Figure 6: Thesis structure and position – Chapter 5

The success of an EHR system is not gauged purely by how technically robust it is, how usable clinicians find the system is also crucial to EHR success. Usability is so essential that the US government has established stipulations to ensure that usability is considered throughout the EHR system

development process.[1] As many top EHR vendors are US based these conditions are relevant to EHRs globally.[2]

The usability of a system can make or break its real world implementation. Should clinicians find a system difficult to use they will often attempt to find alternatives, workarounds or other methods of task completion, where possible avoiding systems that they find difficult to use altogether.[3]

Usability as defined by the International Organization for Standardization (ISO 9241-11:2018) is the 'extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use'.[4] The earlier work in this thesis based on the experimental and replication study has gone some way in addressing the first two of these goals: effectiveness and efficiency. A comparative assessment of EHR usability with either KBM or SR has not yet been undertaken.

The lack of previous research into the effects of utilising SR as an input modality on the usability of an EHR system led to the design and undertaking of this study. This analysis provides a gauge of clinician opinion of the usability of an EHR with and without the assistance of SR. This subjective element of the thesis attempts to complement the earlier objective analysis of efficiency and safety.

5.1.1 Hypotheses

The following hypotheses were tested by the usability study:

1 Speech recognition controlled electronic health record documentation, navigation and interaction is more usable than keyboard and mouse, i.e. clinicians find an electronic health record system with speech recognition more usable than the same electronic health record system with keyboard and mouse as the input modality.

2 Improved system usability leads to improved system performance, i.e. more usable systems are faster to use and produce fewer errors.

5.1.2 Methods

There are various guidelines for EHR system usability evaluation, the majority of which advocate the System Usability Scale (SUS), which has been used to assess and compare numerous EHR systems in previous studies.[5-9] SUS is a freely available set of 10 standard questions that 'seeks the subjective opinion' of participants.[10] Subsequent factor analysis of the SUS scale identified that the scale can be broken down into the two factors of usability and learnability, providing another element of information available from the data.[11]

SUS was selected to evaluate the subjective opinion of the EHR system with each modality, and was the tool employed for the usability study component of the thesis.[10]

Participants of the study were the same 35 ED doctors who participated in the replication reported in chapter 4 of this document. The SUS questionnaire was undertaken immediately at the conclusion of the replication study tasks. Responses were based solely on their experiences with the systems used during the replication study session, disregarding any prior encounters they may have had with either technology.

5.1.3 Results

A significant difference in SUS score results was observed between the two system configurations (EHR with KBM 67, EHR with SR 61; P=0.045), illustrating that the choice of input modality may significantly affect the usability of a system.

Factor analysis showed that the SUS learnability component proved to be the major difference between the two system configurations. Significant differences were found between the learnability related scores by input modality (EHR with KBM 72, EHR with SR 55; P<0.001). The EHR system with SR was viewed as requiring far more training and support to use than the same EHR system with KBM.

The performance data obtained during the replication study provided an opportunity to test whether SUS scores were associated with any objective difference in system performance. A significant

association was found between usability and task efficiency (P=0.002). However, safety (as measured by error rates) was not associated with SUS scores (P=0.905).

5.1.4 Discussion

The results of this study did not support the primary hypothesis that the SR would increase clinicians' perceived usability of an EHR system. Rather, the introduction of SR as an input modality reduced the overall perceived usability of the EHR system for study participants.

The secondary hypothesis was supported for the performance element of efficiency: the configuration with higher usability was found to be associated with significantly faster documentation. However, no association was found between usability and safety. One possible explanation for this is that clinicians were generally not aware of errors occurring during documentation (otherwise they would have addressed them) and therefore these errors did not affect their experience with the system.

SUS questions focusing on learnability provided the major proportion of overall SUS score discrepancies between the two configurations. This could be due to participant unfamiliarity, an overly complicated SR system or poor integration between SR and EHR systems. The additional demands of learnability with SR is a major cause of not only lower usability scores but may also result in a lack of success in real world implementations, as SR requires additional and continued efforts in terms of training, support and integration.

An informal analysis of the usability issues faced by participants during these trials was undertaken for both EHR system configurations (EHR with KBM and EHR with SR). There were numerous usability issues observed with the EHR system that affected both input modalities including: screen layout and design, long lists, excessive pop-ups, screen traversing and elements of commands not completing. There were also issues that occurred solely with one of the modalities, including navigational issues through to system crashes. (See Appendix 3)

These issues highlight the need for clinician consultation throughout EHR design to ensure that systems meet the needs and desires of those utilising the system. As elements of EHR systems often develop at different rates and with different goals in mind, the number, method and variety of separately developed components often cause usability issues. Additional menu, list and icons are introduced with increasing high priority that lead to a patchwork feel and nature of the underlining EHR system. Constant review and optimisation of the overarching system design is required to ensure that usability remains high as additional components, input modality etc. are introduced.

Perhaps a fundamental difference in cognitive loading between the two modalities is a potential explanation for poor SR performance, given the different types of cognitive resources utilised when documenting with SR compared to KBM.[12]

The SR configuration analysed within this study did not appear well suited to the navigational and functional elements required when documenting in current EHRs. SR fared poorly when compared to the point and click navigation possible with KBM based input. This is likely to continue until EHR systems are designed with SR as a core element from the outset, rather than a component bolted on as an afterthought.

5.1.5 Limitations

Perhaps the greatest difficulty when assessing the usability of any system is the potential for participants' biases. Participants of this study were asked to score the SUS scale based solely on their perception of the particular EHR and SR system being tested, yet they were likely to have also brought existing biases from previous exposure to similar systems and technologies. With the near mandatory use of EMRs, most clinicians had previous exposure to various EHR systems throughout their training and careers, be they positive or negative experiences.

5.1.6 Conclusion

The usability of EHR systems overall, independent of input modality, is an area that requires continued attention and development. While EHR systems have been utilised for decades, their user interfaces are still rudimentary when compared to those of other industries. Additional analysis into the causes of the poor SUS results found within this study, together with research into potential areas of improvements would be a worthwhile exercise to enhance the usability of EHR systems overall.

The addition of an SR component to an EHR system may cause a significant reduction in system usability. Until SR is a core part of the overall system design, it is likely to remain a cause of additional overheads, particularly in terms of support and training. These overheads, along with any other benefits and drawbacks, should be thoroughly assessed prior to introducing SR as an input modality.

The introduction of an alternative input modality into a system is a decision that should not be taken lightly; the method of interaction can bear significant differences in both usability and system performance.

5.2 Article contributorship statement

TH, EC and FM conceived the study and its design. TH conducted the research, the primary analysis, and the initial drafting of the paper. EC and FM contributed to the analysis and drafting of the paper, and TH, EC, and FM approved the final manuscript. TH is the corresponding author.

5.3 Paper 4 - Evaluating the usability of speech recognition to create clinical documentation using a commercial electronic health record

International Journal of Medical Informatics, Volume 113, May 2018, Pages 38-42,

https://doi.org/10.1016/j.ijmedinf.2018.02.011

Pages 77-81 of this thesis have been removed as they contain published material. Please refer to the citation above for details of the article contained in these pages.

5.4 Chapter references

1. US Department of Health and Human Services. Health information technology: Standards, implementation specifications, and certification criteria for electronic health record technology, 2014 edition; Revisions to the permanent certification program for health information technology. Federal Register. 2012;77(171):1-130.

2. Lohse S. IDC Perspective: 2017 HealthTech Rankings — Global Benchmarks for the Vendors of Healthcare IT. Framingham, MA: International Data Corporation; 2017.

3. Spencer D. What is usability? [Paper]: Step Two Designs; 2004; Available from: http://www.steptwo.com.au/papers/kmc_whatisusability/.

4. International Organization for Standardization. Ergonomics of human-system interaction. Part 11: Usability: Definitions and concepts: International Organization for Standardization; 2018.

5. Arellano P, Bochinski, J., Boone, E., Grayson, R., Phua, J., & Uehara, R. Selecting an EMR for your practice: Evaluating usability. Healthcare Information and Management Systems Society (HIMSS): Chicago, IL. 2010.

6. Scarlat A. Electronic health record : a systems analysis of the medications domain. Boca Raton: CRC Press; 2012.

7. Drummond Group. 2015 EHR Certification. Website: Drummond Group; 2015 [cited 2017 2017/09/27]; Available from: https://www.drummondgroup.com/products/certified-products/2015-edition-ehr-products.

8. Drummond Group. 2014 EHR Certification. Website: Drummond Group; 2014 [cited 2017 2017/09/27]; Available from: https://www.drummondgroup.com/products/certified-products/2014-edition-ehr-products.

9. Yen P-Y, Bakken S. Review of health information technology usability study methodologies. Journal of the American Medical Informatics Association. 2011;19(3):413-22.

10. Brooke J. SUS-A quick and dirty usability scale. Usability evaluation in industry. 1996;189(194):4-7.

11. Lewis JR, Sauro J, editors. The factor structure of the system usability scale. International conference on human centered design; 2009: Springer.

12. Shneiderman B. The limits of speech recognition. Communications of the ACM. 2000;43(9):63-5.

5.5 Paper 4 references

1. Armijo D, McDonnell C, Werner K. Electronic health record usability. Interface Design Considerations James Bell Associates The Altarum Institute. 2009.

2. Spenceer D. What is usability? : Step Two; 2004.

3. Health UDo, Services H. Health information technology: Standards, implementation specifications, and certification criteria for electronic health record technology, 2014 edition; Revisions to the permanent certification program for health information technology. Federal Register. 2012;77(171):1-130.

4. Nuance Communications Inc. Speech Recognition: Accelerating the Adoption of Electronic Medical Records. Burlington, MA Nuance Communications; 2008.

5. Winscribe. Winscribe Quick Speech Recognition Enhances EMR Usability and Provides Speech-Enabled Documentation for Physicians Takapuna, Auckland: Winscribe; 2015.

6. M*Modal. Got speech recognition? Franklin, TN: M*Modal; 2014.

7. Bevan NK, Kirakowski J, editors. J. And Maissel, J.(1991). What is usability. Proceedings of the 4th International Conference on HCI, Stuttgart; 1991.

8. Arellano P, Bochinski, J., Boone, E., Grayson, R., Phua, J., & Uehara, R. Selecting an EMR for your practice: Evaluating usability. Healthcare Information and Management Systems Society (HIMSS): Chicago, IL. 2010.

9. Scarlat A. Electronic health record : a systems analysis of the medications domain. Boca Raton: CRC Press; 2012.

10. Drummond Group. 2015 EHR Certification. Website: Drummond Group; 2015 [cited 2017 2017/09/27]; Available from: https://www.drummondgroup.com/products/certified-products/2015-edition-ehr-products.

11. Drummond Group. 2014 EHR Certification. Website: Drummond Group; 2014 [cited 2017 2017/09/27]; Available from: https://www.drummondgroup.com/products/certified-products/2014-edition-ehr-products.

12. Yen P-Y, Bakken S. Review of health information technology usability study methodologies. Journal of the American Medical Informatics Association. 2011;19(3):413-22.

13. Hodgson T, Magrabi F, Coiera E. Efficiency and safety of speech recognition for documentation in the electronic health record. Journal of the American Medical Informatics Association. 2017;24(6):1127-33.

14. Dawson L, Johnson M, Suominen H, Basilakis J, Sanchez P, Estival D, et al. A usability framework for speech recognition technologies in clinical handover: A pre-implementation study. Journal of Medical Systems. 2014 May 15;38(6):56.

15. Vogel M, Kaisers W, Wassmuth R, Mayatepek E. Analysis of documentation speed using web-based medical speech recognition technology: randomized controlled trial. Journal of Medical Internet Research. 2015;17(11).

16. Brooke J. SUS-A quick and dirty usability scale. Usability evaluation in industry. 1996;189(194):4-7.

17. Sauro J. Measuring usability with the system usability scale (SUS). Denver, Colorado: MeasuringU; 2011.

18. Likert R. A method of constructing an attitude scale. Scaling: A sourcebook for behavioral scientists. 1974:233-43.

19. Lewis JR, Sauro J, editors. The factor structure of the system usability scale. International conference on human centered design; 2009: Springer.

20. Needle S. Speech Recognition Software Can Increase Ease-of-Use for Doctors. CTS Medical Blog: CTSGuides; 2012.

21. Middleton B, Bloomrosen M, Dente MA, Hashmat B, Koppel R, Overhage JM, et al. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. Journal of the American Medical Informatics Association. 2013;20(e1):e2-e8.

22. Nielson J. User Satisfaction vs. Performance Metrics. Nielsen Norman Group; 2012.

6. Discussion

6.1 Chapter background

The research reported in this thesis found that EHR based clinical documentation with the use of SR as an input modality was inferior to the use of KBM in terms of efficiency (time to complete documentation), safety (number of errors observed) and perceived usability (SUS scores).

This chapter examines the conclusions that can be drawn from the research and provides an overall summation of the thesis. (See Figure 7)

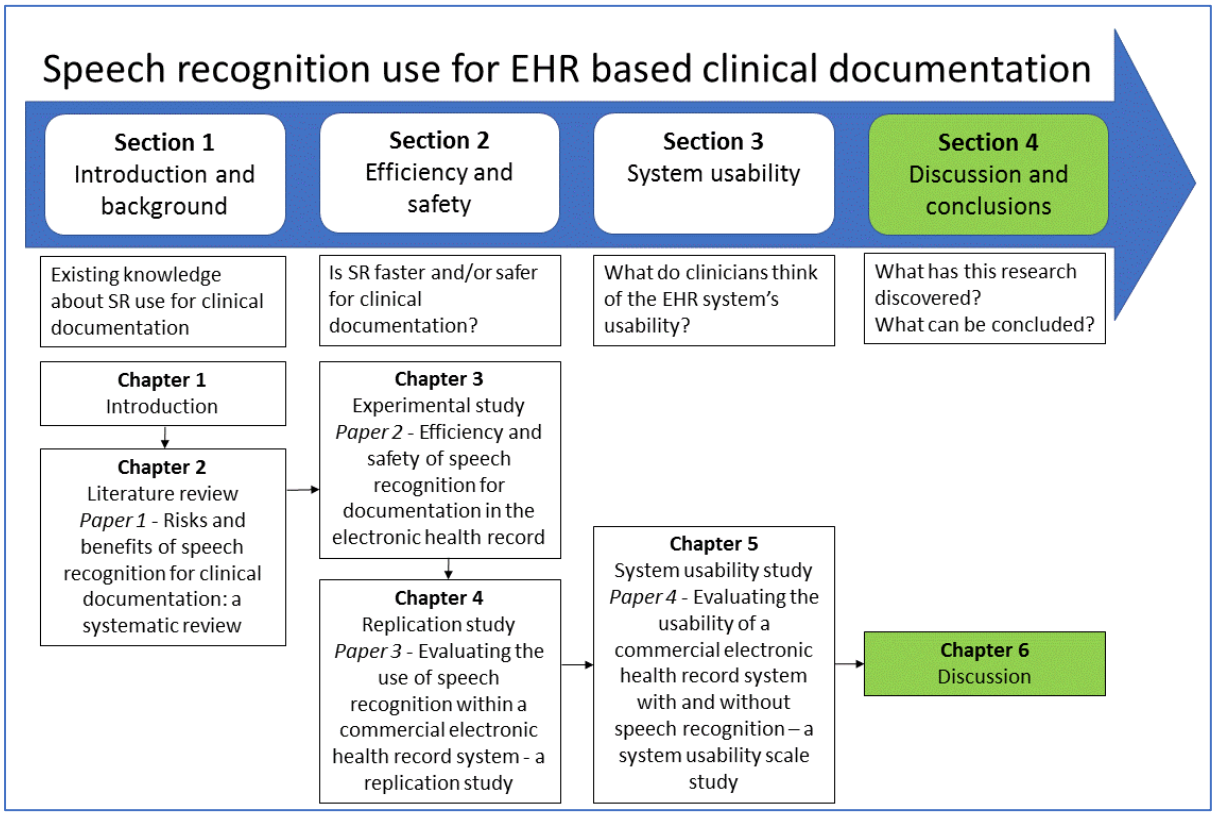


Figure 7: Thesis structure and position – Chapter 6

6.2 Results summary

Efficiency

Across both experiments, complex tasks took significantly longer to complete than simple tasks (mean complex E1 185.63s, simple 121.47s, P<0.001, CI: 50.55-75.34; E2 185.6s, simple 121.5s, P<0.001, CI: -86.20-53.65). This was as expected, with complex tasks having twice the number of subtasks (4) than simple tasks (2).

Clinical documentation took significantly longer to complete using SR when compared to KBM while completing complex tasks for both experiments (E1 KBM 170.48, SR 201.84, P=0.009, CI: 9.61-47.73; E2 KBM 191.9s, SR 224.4s, P=0.009, CI: 11.86-48.26). (See Table 2) However, mean simple task completion times were only significantly slower via SR during the first experiment and there was no significant difference found during the second experiment (E1 KBM 112.38, SR 131.44, P=0.050, CI: -0.07-30.50; E2 KBM 126.4s, SR 126.8s, P=0.701, CI: 6.73-13.20). (See Table 2) Participants who repeated experiment participation demonstrated no significant difference in task completion times between Experiments 1 and 2.

There were no differences in mean task completion times observed for any of the four individual task types between experiments: simple tasks KBM (E1 112.4s, E2 126.4s, P=0.060, CI: -26.04-0.44), simple tasks using SR (E1 131.4s, E2 126.8s, P=0.646, CI: -17.30-10.91), complex tasks KBM (E1 170.5s, E2 191.9s, P=0.199, CI: -39.03-7.39), and complex tasks SR (E1 201.8s, E2 224.4s, P=0.230, CI: -42.93-9.96). (See Table 2)

SR was thus found to be a significantly less efficient method of clinical documentation than KBM, with the exception of simple tasks, where SR completion times were found to be on a par with KBM. There were no tasks undertaken where SR assisted documentation performed significantly faster than KBM documentation.

Table 2: Efficiency summary table

Task completion times Experiment 1						
	Simple tasks KBM time (s)	**Simple tasks SR time (s)**	**Complex tasks KBM time (s)**	**Complex tasks SR time (s)**	**Combined simple & complex KBM (s)**	**Combined simple & complex SR (s)**
Mean	112.38	131.44	170.48	201.84	140.09	165.46
Max	214.99	285.91	257.63	327.45	257.63	327.45
Min	49.38	71.67	110.31	138.87	49.38	71.67
Task completion times Experiment 2						
	Simple tasks KBM time (s)	**Simple tasks SR time (s)**	**Complex tasks KBM time (s)**	**Complex tasks SR time (s)**	**Combined simple & complex KBM (s)**	**Combined simple & complex SR (s)**
Mean	126.39	126.78	191.89	224.39	159.61	176.29
Max	214.54	175.40	349.84	400.01	349.84	400.01
Min	74.14	67.69	104.38	124.42	74.14	67.69

Task completion times Experiment 1 vs Experiment 2				
Efficiency tasks	**Mean task completion time comparison %**	**N**	**Mann-Whitney P-value**	**95% CI**
Simple task KBM E2 vs Simple task KBM E1	112.46%	34	0.060	-26.04-0.44
Simple task SR E2 vs Simple task SR E1	96.46%	31	0.646	-17.30-10.91
Complex task KBM E2 vs Complex task KBM E1	112.56%	31	0.199	-39.03-7.39
Complex task SR E2 vs Complex task SR E1	111.17%	29	0.230	-42.93-9.96

Safety

There was a significant difference in the number of errors (non-typographical) that were observed via input modality for both experiments (E1 170, KBM 32 SR 138, P<0.001, CI: -1.865--1.163; E2 163, KBM 26 SR 137, P<0.001, CI: -2.005--1.167) with the number of SR created documentation errors far exceeding those created with KBM. (See Table 3)

There was no difference in the total number of non-typographical errors between the two rounds of experiments (E1 170, E2 163, P=0.660, CI: -0.279, 0.379). (See Table 3)

Potential patient harm (PPH) errors were significantly greater with SR for both experiments across all simple tasks: minor (E1: KBM 7, SR 33; P=0.002, CI: 0.50-1.00; E2: KBM 1, SR 42, P<0.001, CI: -1.89--0.69), moderate (E1: KBM 0, SR 13, P=0.008, CI: -0.74--0.11; E2: KBM 1, SR 7, P=0.034, CI: -0.34--0.06) and major (E1: KBM 2, SR 29, P<0.001, CI: -1.11--0.46: E2: KBM 1, SR 35, P<0.001, CI: -1.20--0.80).

However, for complex tasks a significant difference was only observed for minor PPH (E1: KBM 9, SR 34, P<0.001, CI: 1.50-2.50; E2: KBM 10, SR 44, P<0.001, CI: -1.40--0.60). (See Table 3)

Significant differences in the number of errors observed between the two experiments were found for some error types: use errors with complex tasks with SR (E1 26, E2 8, P=0.009, CI: 0.00-1.00), comprehension errors with complex tasks with SR (E1 17, E2 0, P<0.001), and omission errors with complex tasks with SR (E1 6, E2 0, P=0.010). (See Table 3)

There was no overall difference in the number of errors observed between experiments for repeat participants (E1 69, E2 57, P=0.311, CI: 0.00–1.00).

There was a significant difference in the occurrence of typographical errors by input modalities observed during the first experiment (E1 KBM 213, SR 252, P=0.037, CI: -1.00-0.00) but not the second experiment (E2 KBM 86, SR 64, P=0.483 CI: -0.00--0.00). The volume of typographical errors was significantly different between the first and second experiment (E1 465, E2 150, P<0.001, CI: 2.00–3.00) and this occurred across all task types: simple tasks with KBM (E1 142, E2 57, P<0.001, CI: 2.00–4.00), simple tasks with SR (E1 133, E2 40, P<0.001, CI: 2.00–4.00), complex tasks with KBM (E1 71, E2 29, P<0.001, CI: 1.00–2.00), and complex tasks with KBM (E1 119, E2 24, P<0.001, CI: 2.00–3.00). (See Table 3)

To summarise the safety results: across both experiments there were a significantly higher number of errors (non-typographical) observed while documenting with SR when compared to KBM.

Typographical errors were, however, reduced significantly in the replication study compared to the initial experiment (E1 465, E2 150, P<0.001, CI: 2.00-3.00). (See Table 3)

Table 3: Error summary table

Experiment 1			***Experiment 2***			***Experiment 1 vs. Experiment 2 (Mann-Witney)***	
						P-values	
	KBM	**SR**		**KBM**	**SR**	**KBM**	**SR**
Total errors observed	**245**	**390**	**Total errors observed**	**112**	**201**		0.6595
Non-typographical	***32***	***138***	***Non-typographical***	***26***	***137***	***Non-typographical E1 vs. E2***	
simple	9	75	simple	3	84	simple	0.814	0.897
complex	23	63	complex	23	53	complex	0.921	0.226
Potential patient harm	***32***	***138***	***Potential patient harm***	***26***	***137***	***Potential patient harm E1 vs. E2***	
Major	**13**	**50**	**Major**	**9**	**43**	**Major**		
simple	2	29	simple	1	35	simple	0.842	0.229
complex	11	21	complex	8	8	complex	0.428	0.310
Moderate	**3**	**21**	**Moderate**	**6**	**8**	**Moderate**		
simple	0	13	simple	1	7	simple	N/A	0.664
complex	3	8	complex	5	1	complex	0.828	0.152
Minor	**16**	**67**	**Minor**	**11**	**86**	**Minor**		
simple	7	33	simple	1	42	simple	0.677	0.916
complex	9	34	complex	10	44	complex	1.000	0.245
Error type	***32***	***158***	***Error type***	***26***	***139***	***Error type E1 vs. E2***		
Integration/system	**2**	**92**	**Integration/system**	**1**	**98**	**Integration/system**		
simple	0	56	simple	0	53	simple	N/A	0.530
complex	2	36	complex	1	45	complex	0.842	0.224
Use errors	**22**	**44**	**Use errors**	**24**	**36**	**Use errors**		
simple	5	18	simple	3	28	simple	0.828	0.162
complex	17	26	complex	21	8	complex	0.538	**0.009**
Comprehension	**8**	**22**	**Comprehension**	**1**	**5**	**Comprehension**		
simple	4	5	simple	0	5	simple	0.039 (C-S)	1.000
complex	4	17	complex	1	0	complex	0.538	**0.000 (C-S)**
Use error type	***30***	***46***	***Use error type***	***25***	***41***	***Use error type E1 vs E2***		
Omission	**8**	**20**	**Omission**	**3**	**21**	**Omission**		
simple	2	14	simple	1	21	simple	1.000	0.190
complex	6	6	complex	2	0	complex	0.414	**0.010 (C-S)**
Commission	**22**	**26**	**Commission**	**22**	**20**	**Commission**		
simple	7	5	simple	2	12	simple	0.668	0.224
complex	15	21	complex	20	8	complex	0.344	0.023
Typographical	***213***	***252***	***Typographical***	***86***	***64***	***Typographical E1 vs. E2***		
simple	142	133	simple	57	40	simple	**0.000**	**0.000**
complex	71	119	complex	29	24	complex	**0.000**	**0.000**

Usability

The usability of the EHR system in Experiment 2 had significantly different mean SUS scores for the two input modalities KBM 67 and SR 61 (P=0.045, CI: 0.14-12.00). Factor analysis showed no difference in the SUS sub element of usability between the EHR with KBM or SR (KBM 65, SR 62, P=0.255, CI: -2.59-9.47). There was however, a significant difference in the mean SUS scores associated with learnability (KBM 72, SR 55, P<0.001, CI: 9.76-23.45).

This suggests that the element of learnability was a large factor in the different scores received by KBM and SR. SR was seen as comparatively more difficult to learn than KBM. A negative correlation between SUS score difference and efficiency difference was observed. For either modality, faster documentation times were associated with higher usability scores.

However, there was no association between the SUS usability score and number of errors occurring. In all instances SR documentation produced a similar (higher) number of errors.

Potential influencing factors

The poor performance of SR in the experiments may have been due to numerous factors including:

I. **Suboptimal system(s) implementation** – one or both systems (EHR or SR) tested may not be best suited for the documentation tasks performed during the study.

   It is possible that the systems used within this study were not the most appropriate for the tasks assessed. However, while other commercially available EHR or SR systems may indeed have fared better, the systems utilised for the studies were based on those implemented by the state public health system and were routinely used by study participants for similar (if not the same) tasks. The EHR and SR systems tested are also amongst the most widely deployed globally. Replication of the experimental study with alternative products (both EHR and SR) by independent researchers would be welcomed.

II. **Inadequate training and support** – participants may not have been sufficiently trained in one or both systems or not have enough relevant experience.

Certainly, improved training and support for both systems is recommended, in particular with SR, as learnability (training and support) was the factor identified as the most telling within the SUS questionnaire and the aspect where SR faired significantly worse than KBM. Best efforts were made to mitigate any training and support variance by only enrolling participants who had received at least the state deemed minimum training (EHR 4h, SR 2h). However, KBM controlled EHR use typically far outweighs any time spent utilising SR.

III. **Poor configuration/integration** – the system configuration or integration between the EHR and SR system may have been inadequate.

Potential integration issues were identified while analysing the first experiment and wherever possible were mitigated prior to the repetition study being undertaken. The configuration and integration utilised were representative of the optimal configuration possible with the technical limitations offered to a state-wide implementation. There are perhaps yet more preferable configurations available. However, these generally relate to stand alone installations without any bandwidth/infrastructure limitations that are not representative of real world use.

IV. **Immaturity of the technology** – the technology, or combination of technologies, was still in development.

Both EHR and SR systems are constantly evolving and improving, and they have been commercially available for decades. Given the widespread real-world use, while clinical documentation using SR is in its infancy, it is now sufficiently mature for critical evaluation.

V. **Tasks/type of documentation** – the type of EHR documentation tasks assessed may have disadvantaged SR compared to KBM.

There is a chance that the tasks selected were biased against SR. However, this study replicated and assessed as closely as possible a real-world application of the systems following a recent state-wide SR implementation.

There may be alternative tasks (other medical specialties), participants (other clinicians) or environments (quieter, less interruptions/distractions) where benefits and advantages of SR are better utilised.

VI. **Cognitive resources** – SR reduces the available cognitive resources available for performing tasks.

There is a case that the use of SR consumes the available short-term and working memory available for performing tasks, which would lead to reduced performance when compared to input modalities such as KBM that do not consume these same resources.[1] It may be that there are limitations to the extent that speech can be optimally utilised with current EHR designs, perhaps focusing on simple tasks, leaving complex tasks to be undertaken via visual or physical modalities.

EHR limitations

The results of the studies undertaken as part of this thesis in part reflect the limitations of the EHR. Baseline error rates and overall usability issues that affect not only SR but also KBM show that EHR systems still have many limitations.

EHR systems as an overarching patient care infrastructure attempt to solve conflicting demands. Numerous competing stakeholders bring a variety of different requirements. These range from clinicians who desire ease of use and efficiency to administrators with demands for greater documentation for billing, funding, reporting and auditing purposes.

There is a thus a trade-off between safety, usability and economic concerns with EHR system design. While patient safety should always be the primary concern, clinician, budgetary, technology and technical support constraints all must also be considered. For example, increased safety could be

achieved with additional access restrictions, mandatory fields or further confirmation screens but these may be too intrusive to clinician users. As noted by Grudin's work on the disparity between who does the work and who gets the benefits, there is a limit to the additional work that will be undertaken by clinicians before they will refuse to use the system or find a workaround.[2] There is a bureaucratic push for the separation of clinical entries into EHRs for billing/auditing purposes. This is in direct opposition to the once common practice of a simple free text entry that was typical in paper-based documentation.

Modern EHR systems thus attempt to be all things to a variety of stakeholders. However, it may not be possible to ensure that the best patient outcomes are achieved while at the same time meeting the requirements and demands of all parties involved.

Hidden benefits and harms of SR

There are numerous benefits and harms that could arise from the use of SR that were not assessed within this study. There is scope for future work to potentially explore some of these including:

- *Quality of documentation* – there may be a variation in the quality, content or length of documentation when completed via different input modalities.
- *Economic evaluation* – cost-benefit analysis should be undertaken on an individual basis for each potential installation. As the costs of systems, installation, training, maintenance and support differ, the economic aspects of implementing SR need specific evaluation.
- *Workflow modifications* – there are likely to be downstream process consequences and workflow modifications incurred with the introduction of SR that need consideration. e.g. the time and place that documentation occurs may be modified.
- *Clinician satisfaction* – clinicians may take pride in using leading edge technologies such as SR. Alternatively, clinicians may resent the pressure to use an input modality forced upon them by hospital management. Lower levels of physician satisfaction and increased rates of burnout are also possible outcomes.[3]

Emerging technologies

There are many exciting emerging technologies that have the potential to complement and leverage SR as an input modality. Artificial Intelligence (AI) is one such area of rapid growth across many industries, with a strong focus on Healthcare.[4] Medical chatbots and AI systems may revolutionise the way that patients (and practitioners) research their health condition, offering a range of services from simple reminders to take medicine, to performing a diagnosis and suggesting treatment.[5]

Natural Language Processing (NLP) as a tool for EHR data capturing is an area where SR may prove beneficial.[6] Research undertaken as part of this thesis has shown that SR may be less than satisfactory when used to enter data into systems which report within individual fields. NLP can supplement SR and allow for speech oriented documentation that removes the need for compartmentalised documentation.

In the distant future, SR may be superseded by technologies now in their infancy such as a Brain-Computer Interfaces (BCIs).[7]

Research real world outcomes

Beyond contributions to research literature, the work reported here has produced several translational outcomes.

The experiments conducted during this research were based on the systems in production within a state based public healthcare system. Various integration and workflow modifications were identified and then implemented as part of the replication study optimisation. These system improvements were communicated to the healthcare organisations involved and they have now successfully implemented many of the modifications from this research within their production EHR/SR environment.

The major change implemented was the method of integration between EHR and SR systems, with multiple health districts now implementing the same technique used for the replication study. This

involved the introduction of an improved method of communication and integration between the locally run SR, and remote session accessed EHR systems. A component (vSync) allows the local SR client application to communicate efficiently within the remote Citrix server published EHR application.

Anecdotal feedback from the clinicians has reported that this has resulted in an increase in system reliability, functionality and robustness, with a reduction of errors, while also allowing for a simplification of workflow and SR based documentation practices.

Initially, these changes were made at one local health district (LHD) and due to the success of that implementation, other LHDs have followed suit.

6.3 Original contributions

This research provides the first rigorous study of the efficiency and safety of speech recognition assisted clinical documentation in an EHR system.

Chapter 4 contains one of only a few clinical replication studies in the healthcare informatics field, which remain uncommon but should be considered essential where study results are unexpected or have significant implication. Replication studies are crucial to avoid over-dependence on the results of a single study and for validation of previously reported results.

6.4 Conclusions

With the rapid emergence of new technologies that may aid clinician documentation such as wearable devices, head mounted displays, virtual and augmented reality to name but a few, SR has the potential to work in conjunction with and complement many digital health technologies.

Continued development of SR audio recording devices and clinical documentation processes means that there should be constant re-evaluation of SR's place in clinical documentation.

Given the ubiquitous nature of electronic records in healthcare, and the substantial cost of documentation in terms of clinician time and patient safety, the foundational act of interacting with

an electronic record requires far closer attention, and substantially higher research priority, than it currently receives.

Research primary learnings

The primary learnings that can be taken from the research undertaken as part of this thesis are set out in Table 4 below.

Table 4: Research primary learnings

Section 1 Introduction	
Findings	**Learnings**
Ch1 Introduction EHR use is virtually ubiquitous.	EHRs bring numerous benefits & risks. EHR input modality may affect electronic documentation.
Ch2 Literature review There is a significant gap in research on SR assisted clinical documentation.	SR is used for EHR based clinical documentation in many uses beyond straight dictation, including the operation and navigation of system menus, fields, pick list and patient charts, item and task selections, and data entry. However, there is no evidence for improved efficiency, safety or usability of SR for performing documentation tasks beyond pure dictation.
Section 2 Efficiency and safety	
Findings	**Learnings**
Ch3 Experimental study SR assisted EHR documentation was found to perform poorly in respect to safety and efficiency when compared to KBM.	SR assisted documentation was found to be significantly slower than KBM documentation. SR documentation generated significantly more errors than KBM based documentation. Improving system integration and workflow may improve SR performance.
Ch4 Replication study The replication study results largely mirrored those of the initial study; SR based EHR documentation performed poorly when compared to KBM.	SR documentation was again found to be both slower and generated more errors than KBM based documentation. System optimisation provided some improvements to experimental results.
Section 3 System usability	
Findings	**Learnings**
Ch5 Usability study The introduction of SR was found to reduce the perceived usability of the EHR system.	The addition of an SR component to an EHR system caused a significant reduction in system usability, primarily due to the sub-element of learnability where SR was seen to require additional training and support compared to KBM.

Section 4 Discussion and conclusion	
Findings	**Learnings**
Ch6 Discussion This research provided little evidence that the introduction of SR was of benefit. System optimisation proved valuable. Further research into the introduction of an input modality on EHR documentation is recommended.	Study results had SR faring poorly compared to KBM for the ED based task undertaken. Real world applications of system optimisation modifications were implemented. The impact of new technologies should be thoroughly explored and analysed prior to introduction to live clinical systems.

The use of speech recognition to drive interactive clinical documentation in the EHR requires careful consideration, and it is not safe to assume that it is currently entirely safe or effective. With clinicians from numerous medical fields accessing the same electronic medical record systems for documentation, it is unwise to assume that what works for department X will work for department Y; even minor differences in requirements, processes or procedures may cause significantly different results.

The search for the safest and most efficient method of clinical documentation remains a work in progress. Emerging technologies such as mobile devices, touch pads, pen control, head-mounted displays, virtual reality, and wearable technology all have different affordances. The place for each of these technologies within the healthcare industry needs individual assessment.

The assumption that the introduction of a new technology will improve clinical outcomes is often incorrect, as exposed through this research into the evaluation of SR for EHR based clinical documentation. Far greater care and assessment needs to be undertaken before the introduction of new components or technologies to an EHR system to ensure that any negative outcomes are mitigated, and benefits realised.

6.5 Chapter references

1. Shneiderman B. The limits of speech recognition. Communications of the ACM. 2000;43(9):63-5.

2. Grudin J. Groupware and social dynamics: Eight challenges for developers. Communications of the ACM. 1994;37(1):92-105.

3. Shanafelt TD, Dyrbye LN, Sinsky C, Hasan O, Satele D, Sloan J, et al., editors. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. Mayo Clinic Proceedings; 2016: Elsevier.

4. Purdy M, Daugherty P. Why artificial intelligence is the future of growth, Accenture. 2016.

5. Topol EJ. The patient will see you now: the future of medicine is in your hands: Tantor Media; 2015.

6. Kaufman DR, Sheehan B, Stetson P, Bhatt AR, Field AI, Patel C, et al. Natural Language Processing–Enabled and Conventional Data Capture Methods for Input to Electronic Health Records: A Comparative Usability Study. JMIR Med Inform. 2016;4(4):e35.

7. Nuyujukian P, Kao JC, Ryu SI, Shenoy KV. A Nonhuman Primate Brain–Computer Typing Interface. Proceedings of the IEEE. 2017;105(1):66-72.

Appendices

Appendix 1 - Ethics approval letters

1.1 Lead HREC & SSA CRGH

Concord Repatriation and General Hospital (CRGH) part of the Sydney Local Health District (SLHD) was the Lead Human Research Ethics Committee (HREC).

Site Specific Approvals (SSA) were also obtained by Northern Sydney Local Health District (NSLHD) for Manly and Royal North Shore Hospitals (RNSH) and South Western Sydney Local Health District (SWSLHD) for Liverpool Hospital.

Appendix 1.1 of this thesis has been removed as it may contain sensitive/confidential content

1.2 Site specific authorisations

Site specific authorisation was provided by both non-lead Human Research Ethics Committee (HREC)

Local Health Districts (LHDs).

Appendix 1.2 of this thesis has been removed as it may contain sensitive/confidential content

Appendix 2 - Participation information and consent form

A separate site-specific version of the master form was produced for each of the sites involved: Manly Hospital, Concord Hospital, Liverpool Hospital and the Royal North Shore Hospital.

INFORMATION FOR PARTICIPANTS

STUDY TITLE: The impact of speech recognition as a navigation and interaction method for EHR clinical documentation in terms of safety, efficiency and cognitive load, when compared to keyboard & mouse interaction.

INVESTIGATORS: Tobias Hodgson

PROJECT SPONSOR: N/A

INTRODUCTION:

You are invited to take part in a research study into the effectiveness of speech recognition for clinical documentation. This Participant Information Sheet will tell you about what is involved in the study and help you decide whether or not you wish to take part. Please read this information carefully. If there is anything you do not understand or if you feel you need more information about anything, please ask. Before you make a decision, please feel free to talk things over with a relative, a friend or your own doctor.

WHY HAVE I BEEN ASKED TO TAKE PART?

You have been invited to take part in this research because you are one of the ED Clinicians that have been trained and have the opportunity to use speech recognition during your clinical documentation into the EHR system

WHAT IS THE PURPOSE OF THIS RESEARCH?

- The study will provide information on the safety, efficiency and ease of use of using speech recognition in comparison to keyboard & mouse interaction.
- By having real clinicians perform task into an EMR using both methods we will be able to directly compare the two methods of interaction.
- The results may provide information as to the areas where: speech recognition is best used, should be improved or avoided altogether. This may lead to addition research as to best improve EHR interaction options.

DO I HAVE TO TAKE PART IN THE STUDY?

Taking part in any research is entirely voluntary. If you do decide to take part you can withdraw at any time without having to give a reason. Please be assured that, whatever your decision, it will not affect your relationship with medical staff.

WHAT WILL HAPPEN TO ME IF I TAKE PART?

- The consent form will be signed prior to any assessments being performed
- A timeslot will be organised for participation
- Participants will undertake eight randomised tasks four to be undertaken via typical keyboard & mouse and four utilising speech recognition.
- Task will be performed under varied clinical conditions (potential interrupts)
- A brief entry and exit questionnaire will also be undertaken
- Participation should take no longer than one session of 30-40 minutes
- Research will be monitored with data logging and video capturing systems.
- There are no costs involved for any participants

WHAT DO I HAVE TO DO?

- The only prerequisites to participate in the study other than being familiar with the EHR system (Cerner's FirstNet) and the speech recognition system (Nuance's Dragon Medical).
- Participants will complete clinical documentation tasks following provided test scripts.

WHAT WILL HAPPEN TO MY TEST DATA?

- Data will be stripped of any identifying information and used only for research purposes
- Data created or accessed during the study will be stored on encrypted drives (both HDD and USB) and only accessible with the appropriate key and backups created on a daily basis.
- Microsoft BitLocker Drive Encryption will be the encryption software utilised through the study.

WHAT ARE THE POSSIBLE BENEFITS OF TAKING PART?

While we intend that this research study to further clinical documentation and the method of EHR interaction, this may occur in the future, it may not be of direct benefit to you.

WHAT ARE THE RISKS OF TAKING PART?

There are no tangible risks in participating in this study

Appendix 2, page 3 of 4 of this thesis has been removed as it may contain sensitive/confidential content

The impact of speech recognition as a navigation and interaction method for EHR clinical documentation in terms of safety, efficiency and cognitive load, when compared to keyboard & mouse interaction.

PARTICIPANT CONSENT FORM

I, ……..…….……[name]

of……[address]

have read and understood the Information for Participants for the above named research study and have discussed the study with

…… ……...

- I have been made aware of the procedures involved in the study, including any known or expected inconvenience, risk and of their implications as far as they are currently known by the researchers.
- I freely choose to participate in this study and understand that I can withdraw at any time.
- I also understand that the research study is strictly confidential.
- I hereby agree to participate in this research study.

Name (Please Print): ..... .

Signature:..... **Date:** .....

Name of Person who conducted informed consent discussion (Please Print): Tobias Hodgson

……

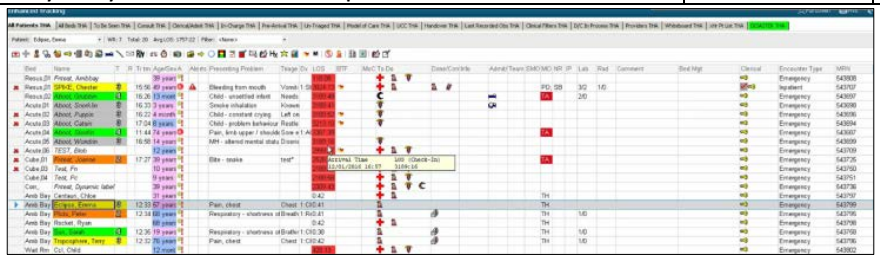
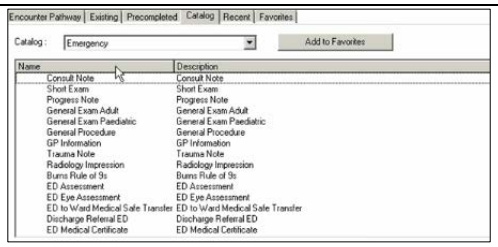
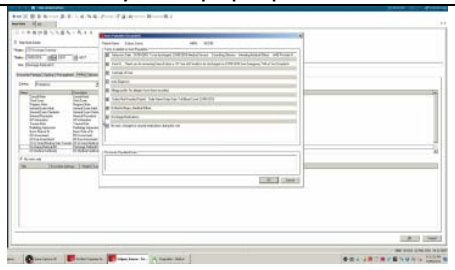
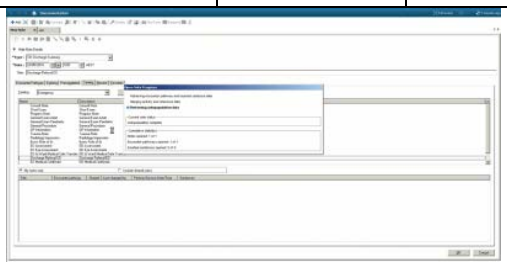
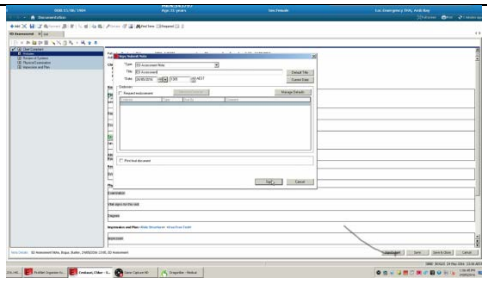
Signature of Person who conducted informed consent discussion:

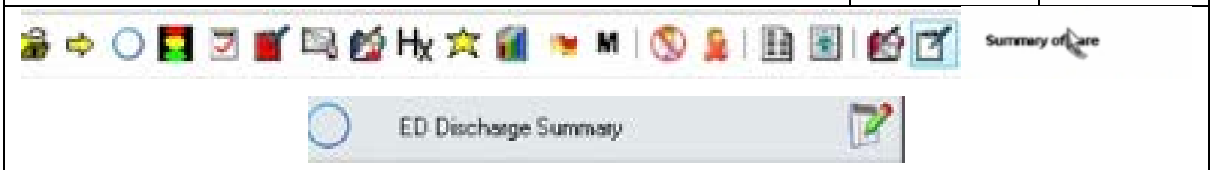
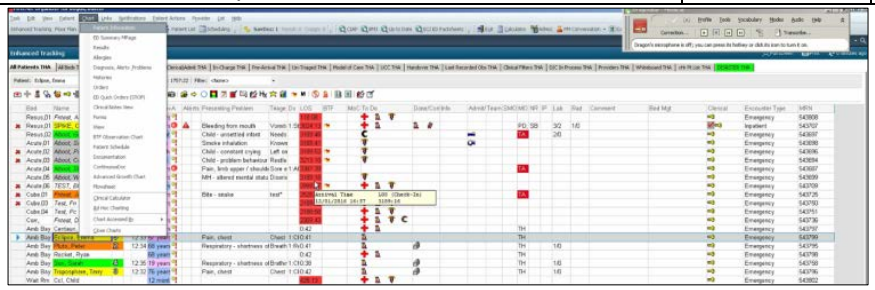
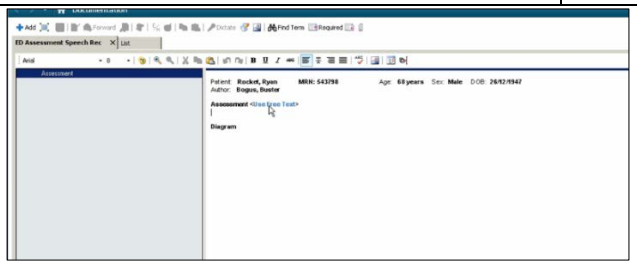
Signature:..... **Date:** .....

Appendix 3 - Observed usability issues table

Issues observed with the usability of the EHR system that affected either input modality are shown below with relevant screenshots.

Table 5: Observed usability issues

<table>
<tr><th>Issue</th><th>EHR & KBM</th><th>EHR & SR</th></tr>
<tr><td>Cluttered screens
Numerous overly busy screens with multiple icons and colours etc.</td><td>✓</td><td>✓</td></tr>
<tr><td colspan="3">  </td></tr>
<tr><td>Long lists
Various long drop lists and selection boxes.</td><td>✓</td><td>✓</td></tr>
<tr><td colspan="3"> 
</td></tr>
<tr><td>Excessive pop-up boxes
Numerous pop-up boxes of different sizes, shapes and positions. No consistency with pop-up location or button placement.</td><td>✓</td><td>✓</td></tr>
<tr><td colspan="3"> 
 </td></tr>
<tr><td>Screen traversing
Many tasks involve traversing back and forth across several screens/pop-ups to complete button selection.</td><td>✓</td><td>✓</td></tr>
<tr><td colspan="3">  </td></tr>
</table>

Issue	EHR & KBM	EHR & SR
Navigation options Multiple and inconsistent methods of navigation available to user. Via top bar, side bar, icons, process pages, text links, right click etc.	✓	
		
Input modality switching Constant switching between tasks via microphone and mouse for some elements of navigation/documentation.		✓
Automated navigation commands Auto navigation commands are visible and interactive. Often leading to confusion and manual interruption.		✓
		
Cursor placement Where and how to select the location for text entry is not always obvious.		✓
		
System crashes EHR system crashes were observed.		✓
Command went nowhere or to wrong place/chart Numerous occurrences of navigational commands not completing.		✓
All elements of command did not complete Several occurrences of elements of a command not completing.	✓	✓

Appendix 4 – Supplementary materials paper 1

Supplementary materials available via online journal publication of the paper.

A) Table A1: SWOT analysis for speech recognition use in clinical documentation

Strengths	**Weaknesses**
• **Reduced TAT** – 8 studies found SR provided a reduction in TAT during report creation. • **Increase in reported data** – 3 studies found an increase in words captured per report. While an increase in data captured is often positive, this could also potentially be a weakness.	• **Increased editing time** – 4 studies found SR provided an increase in document editing time. • **Increased number of errors per report** – 6 studies found SR use increase in the number of errors per report.
Opportunities	**Threats**
• **Financial savings** – 4 studies found SR provided a reduction (or potential of reduction) in costs. • **Technology development** – as SR technology matures accuracy and integration improves while overheads reduce.	• **System overheads** – installation, integration, support and training overheads when implementing a SR system are high. • **Dissatisfaction of clinicians** – SR has been found to reduce staff satisfaction for some clinicians when implemented.

B) Table B1: Data extraction template

Admin/Extraction Information

Reference:	
Title:	
Author(s):	
Date:	
Publication Type: (full report, chapter, abstract, letter etc.)	
Publication:	
Author Contact Details:	
Notes:	

1st Reviewer Name:	
2nd Reviewer Name:	
Date of Extraction/Appraisal:	
Study Summary:	
Study Language	
Article Language	

Methodology

Study Type: (RCT, Quasi-RCT, Longitudinal, Retrospective, Observational, other)	
Start Date:	
End Date:	
Duration:	
Ethical Approval: (Y/N/Unclear)	
Randomisation Details:	
Blinding Details:	

Participants

Population Description:	
No. Subjects:	
Determinants: (English as a 2nd language ESL, Qualification, Experience etc.)	
Notes:	

Criteria

Inclusion Criteria: Was the study original? Does it add to the literature? Was the design sensible?	
Exclusion Criteria: Not relevant to SR Article not English etc.	
Inclusion of Study: (Y/N)	
Reason for Study Exclusion:	
Notes:	

Study Aims

Aim of Study:	
Design:	
Units of Observation:	
Setting and Context: (e.g. clinical reports, Radiologists)	
Method of Recruitment:	
Missing Data:	
Notes:	

Interventions

Intervention indicator, factor or exposure	
Intervention Question e.g. Does SR speed	
Economic Variables e.g. Additional costs due to interventions	

Speech Recognition Details

Format (Template, Free text, Navigation etc.)	
SR Technology	
SW Technology	
HW Technology	

Key Findings

Key Findings:	
Key Conclusions:	
Overall SR Position (For/Against/Neutral)	
Other studies References:	
Notes:	

Study Limitations

Limitation of Study:	
Study Funding Source:	
Any Conflict of Interest:	
Notes:	

C) Table C1: JBI MAStARI – Critical Approval Checklist

<u>RCT and Pseudo</u>
Was the assignment to groups truly random?
Were participants blinded to allocation?
Was allocation to treatment groups concealed from the allocator?
Were the outcomes of people who withdrew described and included in the analysis?
Were those assessing outcomes blind to allocation?
Were the control and treatment groups comparable at entry?
Were groups treated identically other than for the named interventions?
Were outcomes measured in the same way for all groups?
Were outcomes measured in a reliable way?
Was appropriate statistical analysis used?
<u>Comparable Cohort / Case Control Studies</u>
Is the sample representative of the population as a whole?
Are participants of similar ilk?
Were confounding factors identified and strategies to deal with them stated?
Were outcomes assessed using objective criteria?
Was follow up carried out over a sufficient time period?
Were the outcomes of people who withdrew described and included in the analysis?
Were outcomes measured in a reliable way?
Was appropriate statistical analysis used?
<u>Descriptive / Case Series Studies</u>
Was the study based on a random or pseudo-random sample?
Were the criteria for inclusion in the sample clearly identified?
Were confounding factors identified and strategies to deal with them stated?
Were outcomes assessed using objective criteria?
If comparisons were being made, was there sufficient description of the groups?
Was follow up carried out over a sufficient time period?
Were the outcomes of people who withdrew described and included in the analysis?
Were outcomes measured in a reliable way?
Was appropriate statistical analysis used?

Appendix 5 – Supplementary materials paper 2

Supplementary materials available via online journal publication of the paper.

A) Participant tasks

Mary Mercury

Mary Mercury is a 61 year old female who presented to the emergency department with a headache/migraine today. She is yet to be seen by a doctor.

Using the keyboard & mouse, complete the following tasks:

A. Assign yourself as Mary Mercury's provider – use the "Assign provider" icon

B. Perform an ED assessment on Mary Mercury – use the "Documentation" icon

Enter only the following data within the note:
Chief Complaint - Present Complaint: Migraine headache
Chief Complaint - History of Present Illness: Frontal headache that developed into nausea.
Histories - Past Medical History: 5 year history of migraines and headaches.
Histories - Family & Social History: Lives with husband and son.

Veronica Venus

Veronica Venus is a 55 year old female who presented to the emergency department with sun stroke today. She is yet to be seen by a doctor.

Using the speech recognition wherever possible, complete the following tasks:

A. Assign yourself as Veronica Venus's provider – use/say the "Assign provider Speech Rec" command.

B. Perform an ED assessment on Veronica Venus– use/say the "ED Assessment Speech Rec" and "ED Assessment Template" commands

Enter only the following data within the note:
Chief Complaint - Present Complaint: Nausea and sunburn.
Chief Complaint - History of Present Illness: Nausea and vomiting after sun exposure.
Histories - Past Medical History: High cholesterol, under control via diet.
Histories - Family & Social History: Smokes 20 cigarettes a day.

Eric Earth

Eric Earth is a 33 year old male who presented to the emergency department with back pain today. He is yet to be seen by a doctor.

Using the keyboard & mouse, complete the following tasks:

C. **Assign yourself as Eric Earth's provider** – use the "Assign provider" icon

D. **Perform an ED assessment on Eric Earth**– use the "Documentation" icon

Enter only the following data within the note:
Chief Complaint - Present Complaint: Back pain.
Chief Complaint - History of Present Illness: Exacerbation of long-term back pain.
Histories - Past Medical History: Long history of degenerative joint disease.
Histories - Family & Social History: Lives with elderly mother and sister.

Matthew Mars

Matthew Mars is a 65 year old male who presented to the emergency department with leg pain today. He is yet to be seen by a doctor.

Using the speech recognition wherever possible, complete the following tasks:

A. **Assign yourself as Matthew Mars's provider** – use/say the "Assign provider Speech Rec" command.

B. **Perform an ED assessment on Matthew Mars** – use/say the "ED Assessment Speech Rec" and "ED Assessment Template" commands

Enter only the following data within the note:
Chief Complaint - Present Complaint: Leg pain.
Chief Complaint - History of Present Illness: Intermittent stabbing pain in right calf.
Histories - Past Medical History: Removal of ruptured spleen 5 years ago.
Histories - Family & Social History: Lives with elderly mother and sister.

James Jupiter

James Jupiter is a 77 year old male who presented to the emergency department with a fever today. He is already assigned as your patient and has been triaged as category 3.

Using the keyboard & mouse, complete the following tasks:

A. **View James Jupiter's vital signs** – use the "ED Summary MPage" icon

B. **Add a Diagnosis for James Jupiter** – use the "Diagnosis, Alerts & Problems" icon

Enter only the following data within the note:
Diagnosis: Fever.
Diagnosis Comments: BGL is "X.X"
(X.X = the value in James's vital signs)

C. **Add an Order for James Jupiter** – use the "Add Order" icon

Enter only the following data within the note:
Order: Full Blood Count (FBC).
Current Clinical History (Mandatory): Fever.

D. **Create a Discharge Note for James Jupiter** – use the "Depart Process" icon & then the ED Discharge Summary Icon

Enter only the following data within the note:
Visit Information - Summary of Care: The patient has been experiencing a fever.
Health Status – Add Diagnosis: *Include the "Active" Diagnosis in the note.*

Sally Saturn

Sally Saturn is a 40 year old female who presented to the emergency department with breathing issues today. She is already assigned as your patient and has been triaged as category 5.

Using the speech recognition wherever possible, complete the following tasks:

A. **View Sally Saturn's vital signs** – use (say) the "ED Summary MPage Speech Rec" command

B. **Add a Diagnosis for Sally Saturn** – use (say) the "Diagnosis Speech Rec" command

Enter only the following data within the note:
Diagnosis: Closed fracture of wrist.
Diagnosis Comments: BGL is "X.X"
(X.X = the value in Sally's vital signs)

C. **Add an Order for Sally Saturn** – use (say) the "Order Speech Rec" command

Enter only the following data within the note:
Order: Full Blood Count (FBC).
Current Clinical History (Mandatory): Wrist pain.

D. **Create a Discharge Note for Sally Saturn** – use (say) the "Discharge Referral Speech Rec" command.

Enter only the following data within the note:
Visit Information - Summary of Care: Patient appeared with acute wrist pain.
Health Status – Add Diagnosis: *Include the "Active" Diagnosis in the note.*

Uri Uranus

Uri Uranus is a 22 year old male who presented to the emergency department with Loss of consciousness today. He is already assigned as your patient and has been triaged as category 3.

Using the keyboard & mouse, complete the following tasks:

A. **View Uri Uranus's vital signs** – use the "ED Summary MPage" icon

B. **Add a Diagnosis for Uri Uranus**– use the "Diagnosis, Alerts & Problems" icon

Enter only the following data within the note:
Diagnosis: Loss of consciousness.
Diagnosis Comments: BGL is "X.X"
(X.X = the value in Uri's vital signs)

C. **Add an Order for Uri Uranus** – use the "Add Order" icon

Enter only the following data within the note:
Order: Full Blood Count (FBC).
Current Clinical History (Mandatory): LOC.

D. **Create a Discharge Note for Uri Uranus** – use the "Depart Process" icon & then the ED Discharge Summary Icon

Enter only the following data within the note:
Visit Information - Summary of Care: An episode of loss of consciousness.
Health Status – Add Diagnosis: *Include the "Active" Diagnosis in the note.*

Nancy Neptune

Nancy Neptune is a 58 year old female who presented to the emergency department with flu-like symptoms today. She is already assigned as your patient and has been triaged as category 5.

Using the speech recognition wherever possible, complete the following tasks:

A. **View Nancy Neptune's vital signs** – use (say) the "ED Summary MPage Speech Rec" command

B. Add a Diagnosis for Nancy Neptune – use (say) the “Diagnosis Speech Rec” command

Enter only the following data within the note:
Diagnosis: Flu-like symptoms.
Diagnosis Comments: BGL is "X.X”
(X.X = the value in Nancy’s vital signs)

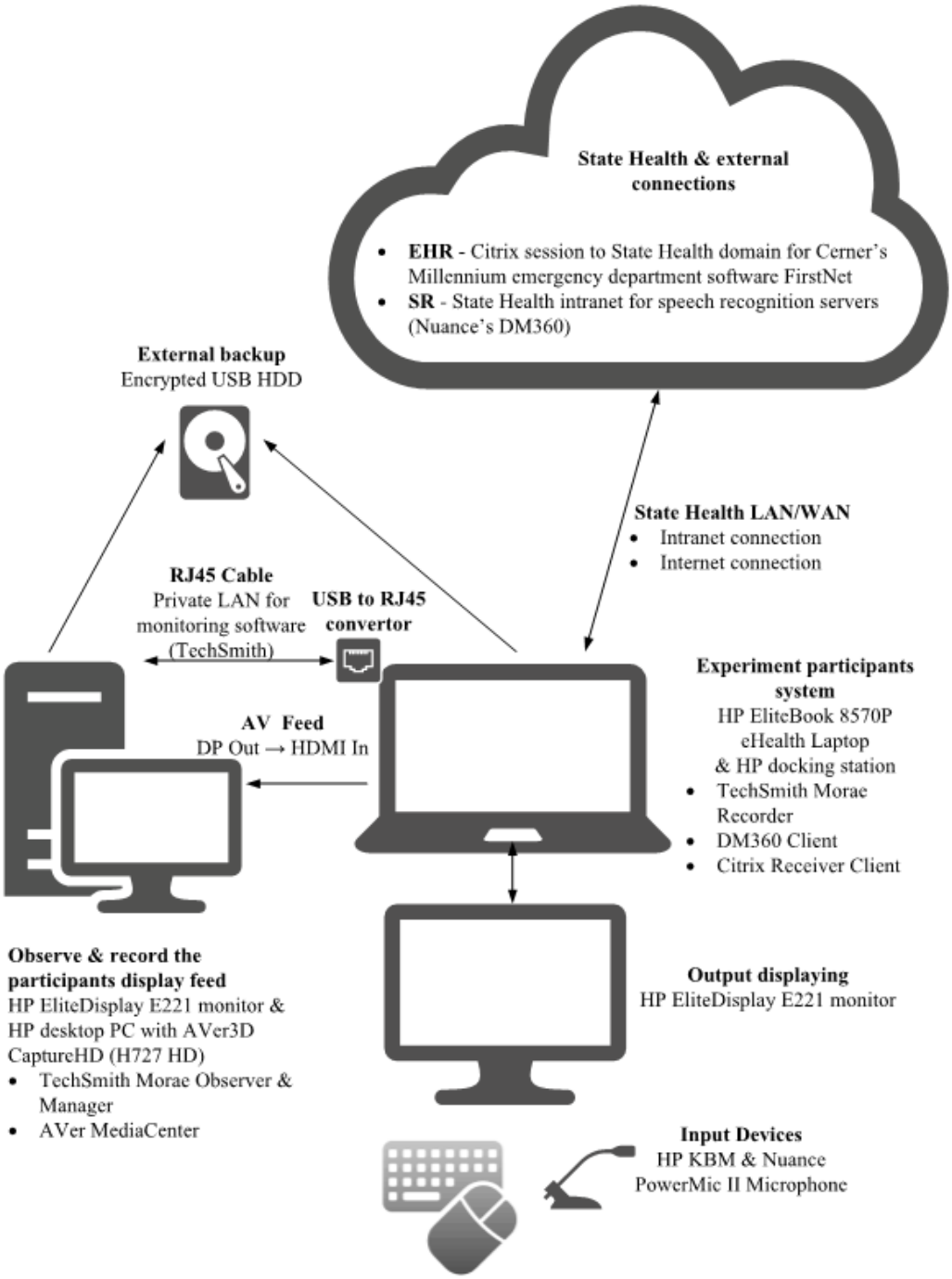
C. Add an Order for Nancy Neptune – use (say) the "Order Speech Rec" command

Enter only the following data within the note:
Order: Full Blood Count (FBC).
Current Clinical History (Mandatory): FLU.

D. Create a Discharge Note for Nancy Neptune – use (say) the “Discharge Referral Speech Rec” command.

Enter only the following data within the note:
Visit Information - Summary of Care: Patient appeared with flu-like symptoms.
Health Status – Add Diagnosis: *Include the “Active” Diagnosis in the note.*

B) Experiment technical design



Materials

Participants were provided physical task sheets (including variables to input) and reference guides. Interrupts, pre- & post-trial questionnaires were electronic and displayed on screen.

Software systems

The system(s) used within the study were established after an investigation of the commonalities and necessities of currently available systems within the Australian marketplace. The chosen software packages were: representative of "Generic" EHR system, and speech recognition systems, covered all core elements for standard clinical documentation and were available for both research access and commercial use within Australia.

A test domain (including testing versions of EHR software) was used, with fictional pre-defined clinician and patient data created that was replicated for each study participant. Clinicians were logged into the system(s) with credentials specifically generated for the trial, no real/live patient data was accessed during the trial.
Permission to utilize these test domains was sought at both the local health district (LHD) level and through State Health.

EHR software

The EHR software used within the trial was FirstNet the ED information management system component of Cerner's Millennium Health Network Architecture suite of products (version 2012.01.30).

Speech recognition hardware & software

The speech recognition software was Nuance's Dragon Medical 360 Network Edition UK (version 2.0, 12.51.200.072).

Survey creation software

The survey component of TechSmith's Morae (v3.3.3) was used to provide pre & post-test surveys along with task interruptions.

Monitoring software

Data logging/screen capturing software

Specific data logging software allowed real time monitoring of task steps, completion time, method comparison and error capturing. TechSmith's Morae (v3.3.3) usability testing software was used throughout the trial.

Display feed capturing

An AverMedia AVer3D CaptureHD card was also used to recording in real-time each trial session. The HDMI display signal (participants monitor feed) was saved as videos. This feed was combined with a modified system tray clock showing seconds as well as hours/minutes (7+ taskbar Tweaker version 4.5.10) which allowed for a secondary method of session playback and calculation of time task durations.

Dragon Medical 360 Network Edition data

Trends or very simple usage statistics were available within the speech recognition software management console.

Hardware

A State Health laptop was configured to access the test domain (HP EliteBook 8570p). A microphone (Nuance PowerMic II), USB keyboard & mouse (Microsoft Wired Desktop 600), and display monitor (HP EliteDisplay E221) were utilized by participants throughout the study.

D) Error class categorization tables

	Potential Patient Harm Errors			*Integration/System, User or Comprehension (Error Type)*			*Omission or Commission Errors (User Error Type)*	
Observed Error	**Minor**	**Moderate**	**Major**	**Integration /System**	**User**	**Comprehension**	**Omission**	**Commission**
Incorrect patient			X		X			X
Incorrect patient - user corrected		X			X			X
No BGL entered		X			X		X	
Incorrect BGL entered			X		X			X
Incorrect test order collection date entered			X		X			X
Data lost during text transfer (no EHR record created)			X		X			X
Data entered in wrong EHR field			X		X			X
Section of EHR missed			X		X		X	
Clinician closed EHR	X				X			X
User added trivial word (e.g. "and")	X					X		X
Omitted trivial word (e.g. "is")	X					X	X	
Incorrect trivial word entered	X				X			X
Incorrect trivial word entered - user corrected	X				X			X
Incorrect significant word entered (diagnostic)			X		X			X
Omitted significant word (diagnostic)			X		X		X	
Template brackets not removed (Accept defaults missed)	X				X		X	
Incorrect method of EHR menu navigation used	X				X			X
Word mangled (letters repeated or cut off)	X			X		X		
Word mangled - user corrected (letters repeated or cut off)	X			X		X		
System added trivial word (e.g. "and")	X			X		X		
Misrecognition of word by SR			X	X				
Misrecognition of word by SR - user corrected		X		X				
All elements of command did not complete	X			X				
Command went nowhere or to wrong place/chart	X			X				
EHR slow - System lag	X			X				
EHR crashed	X			X				
Element of EHR down e.g. vitals			X	X				

Typographical Errors
Missing full stop
Capitalization error
Missing comma(s)
Hyphen error
Plural form error (missing/added "s")
Spelling error

E) Error class definition tables

Potential Patient Harm Errors (after FDA Guidelines for Industry)	
Minor	Failures or latent design flaws would not be expected to result in any injury to the patient.
Moderate	Failures or latent design flaws could result in non-serious injury to the patient.
Major	Failures or latent flaws could result in death or serious injury to the patient.

Integration, User or Spelling & Grammatical Errors (Error Type)	
Integration	Errors caused by the SR system, the EHR systems or interactions between the two. Some errors occurred because of network effects external to either user or clinical software and included delays or failures in execution of user requests, or introduction of errors in input such as mangled word errors.
User	Errors that are typically semantic in nature, including: not entering data, entering incorrect data, and errors in navigation of the system whilst attempting a task.
Comprehension	Errors that may affect the comprehension of the documentation such as additional/missing words.

Omission or Commission Errors (User Error Type)	
Omission	A task or element of a task was not completed.
Commission	A task or part of a task was completed incorrectly.

Typographical Errors	
Typographical	Typographical errors such as missing full stops or capitalisation errors

F) Self-reported data analysis

Experience	% Change in Mean Task Completion Time	N Total	Z	P-Value
Simple Task KBM Low Experience vs Simple Task KBM High Experience	106.23%	16	-0.207	0.717
Simple Task SR Low Experience vs Simple Task SR High Experience	117.33%	13	-1.363	0.538
Complex Task KBM Low Experience vs Complex Task KBM High Experience	91.33%	13	-0.594	0.393
Complex Task SR Low Experience vs Complex Task SR High Experience	90.03%	11	-1.600	0.303
Clinical Role Seniority	**% Change in Mean Task Completion Time**	**N Total**	**Z**	**P-Value**
Simple Task KBM Junior vs Simple Task KBM Senior	107.77%	10	-1.172	0.100
Simple Task SR Junior vs Simple Task SR Senior	120.34%	9	-0.533	0.349
Complex Task KBM Junior vs Complex Task KBM Senior	113.41%	8	-1.540	0.267
Complex Task SR Junior vs Complex Task SR Senior	102.63%	9	-1.718	0.588
Skill Level	**% Change in Mean Task Completion Time**	**N Total**	**Z**	**P-Value**
Simple Task KBM Low vs Simple Task KBM High	99.43%	14	-0.282	0.452
Simple Task SR Low vs Simple Task SR High	112.25%	13	-1.013	0.368
Complex Task KBM Low vs Complex Task KBM High	90.00%	12	-1.334	0.307
Complex Task SR Low vs Complex Task SR High	86.74%	11	-1.511	0.073

G) Participant instructions

Hello Dr. XYZ,

Thank you for agreeing to be part of this trial today. You are going to be performing some clinical documentation tasks with the Electronic Health Record (EHR) system FirstNet, as used within this hospital's Emergency Department.

Tasks will be performed via either keyboard & mouse, or with the assistance of Dragon Medical 360 speech recognition software. The study should take less than an hour of your time and the outcomes of the trial will help to improve electronic clinical documentation in this and other hospitals.

We are not assessing your clinical skills only your interactions with the EHR system.

Please complete the PIS & consent forms now if you have not already.
(Provide the forms.)

For the duration of this trial please complete all tasks using the exact steps and methods as described in the guides that will be provided, even if you typically use a different or superior method.

Where data is provided to enter into the system, please enter only the data exactly as shown: including values, spelling, punctuation & capitalisation etc. even if they seem irrelevant. Please ensure that you review any documentation before submitting/signing off.

There are a varied set of tasks that will be performed in several scenarios:
Provider Assignment, ED Assessment, Vitals, Diagnose, Order & Discharge Referral

Questions & surveys may appear randomly during the trial, please stop and answer these before continuing. These questions are not scored, but please answer honestly and openly.

If informed that a similar order already exists, click "order anyway". Also, ignore any existing notes of the same type that may exist for the patient.

We will now run through example tasks to be undertaken using both methods, keyboard & mouse and speech recognition on the test patients Peter Pluto and Sarah Sun.
Please now login to Dragon with your own profile (look this up prior to session)

You will complete all tasks in FirstNet as Dr. Bogus and are already logged into FirstNet as that provider. If needed the username and login are bogus/bogus.

Appendix 6 – Supplementary materials paper 3

Supplementary materials available via online journal publication of the paper.

Supplementary Material

Appendix A Materials and Methods

Detailed Materials and Methods Used throughout the Experiment

A within-subject experimental study was undertaken with emergency department (ED) physicians, with each assigned four standardized clinical documentation tasks using a commercial electronic health record (EHR). Participants navigated the EHR and documented patient information for simulated patients. The order of task completion was allocated randomly, with half of the tasks assigned to keyboard and mouse (KBM) and half to speech recognition (SR).

The four documentation tasks were representative of those commonly undertaken within an EHR by ED physicians and included patient assignment, patient assessment, diagnosis, orders, and patient discharge. Tasks were chosen in consultation with senior ED clinicians, who did not further participate as subjects in the trials. All simulated patients had active records available in the experimental version of the standard ED EHR.

To allow for variation in task complexity, four of the eight tasks were designed to be simple and four complex. Complexity was measured by the number of subtasks, with the simple tasks having two subtasks and complex tasks having four.

The clinical software used for the experiment was the Cerner Millennium suite with the FirstNet ED component (v2015.01.11) and Nuance Dragon Medical 360 Network Edition UK (v2.4.2) speech recognition software. Both were configured to replicate the operation of the EHR that subjects used daily. All user actions were automatically logged, down to individual keystrokes, with recording software. Session EHR screens and audio were also separately recorded with a high-definition multimedia interface capture device.

Thirty-five participants volunteered from three urban teaching hospitals in Sydney, Australia, from an eligible population of approximately 100 ED clinicians. To be eligible, subjects must have previously completed training in the EHR system, including specific SR training (EHR: 4 hours, SR: 2 hours). Clinicians were excluded if they had a pronounced speech impediment or physical disability that might affect system use.

It was estimated that a sample size of 27 clinicians would be sufficient to test for differences in time efficiency and error rates when using a *t*-test with a significance level of 0.05 and power of 0.95. Calculations were performed using G*Power (v3.1).

A System Usability Scale (SUS) questionnaire was completed at the end of the trial to gather participants' options on the EHR and SR systems. The results of this questionnaire are to be examined in a separate article.

The study was approved by the university and participating hospitals' ethics committees. The trials took place over two separate 2-month periods, commencing March 2015 for Experiment 1 and May 2016 for Experiment 2.

Appendix B Participants' Tasks

Tasks Undertaken by Trial Participants

Chloe Centauri

Chloe Centauri is a 31-year-old female who presented to the emergency department (ED) with neck pain today. She is yet to be seen by a doctor.

Using the **keyboard and mouse**, complete the following tasks:

A. Assign yourself as Chloe's provider—use the "Assign provider" icon.

B. Perform an ED assessment on Chloe—Use the "Documentation" icon.

Enter only the following data within the note:

Chief complaint—Present complaint: Neck pain.

Chief complaint—History of present illness: Aggravation of long-term neck issue.

Histories—Past medical history: 7-year history of degenerative joint disease.

Histories—Family and social history: Lives with elderly mother and father.

Ryan Rocket

Ryan Rocket is a 68-year-old male who presented to the ED with arm pain today. He is yet to be seen by a doctor.

Using **speech recognition** wherever possible, complete the following tasks:

A. Assign yourself as Ryan's provider—Use/say the "Assign provider Speech Rec" command.

B. Perform an ED assessment on Ryan.

- Use/say the "ED Assessment Speech Rec" and "ED Assessment Template" commands.

Enter only the following data within the note:

Chief complaint—Present complaint: Arm pain.

Chief complaint—History of present illness: Short-term stabbing pain in right forearm.

Histories—Past medical history: Removal of ruptured appendix 5 years ago.

Histories—Family and social history: Lives with wife and two children.

Terry Troposphere

Terry Troposphere is a 76-year-old male who presented to the ED with chest tightness today. He is already assigned as your patient and has been triaged as category 3.

Using the **keyboard and mouse**, complete the following tasks:

A. View Terry's vital signs and note latest BGL—Use "ED Summary MPage" icon.

B. Add a diagnosis for Terry—Use "Diagnosis, Alerts & Problems" icon.

Enter only the following data within the note:

Diagnosis: Chest tightness.

Diagnosis comments: BGL is "X.X"

(X.X = the BGL value found in Terry's vital signs)

C. Add an order for Terry—Use the "Add Order" icon.

Enter only the following data within the note:

Order: Full blood count (FBC).

Current clinical history (mandatory): Chest tightness.

Clinician collect: No, **Collection date/time:** Today/Now.

D. Create a discharge note for Terry—Use the "Depart Process" icon and then "ED Discharge Summary" icon.

Enter only the following data within the note:

Visit information—Summary of care: The patient appeared with chest tightness.

Health status—Add diagnosis: *Include the "Active" diagnosis in the note.*

Emma Eclipse

Emma Eclipse is a 57-year-old female who presented to the ED with chest discomfort today. She is already assigned as your patient and has been triaged as category 3.

Using **speech recognition** wherever possible, complete the following tasks:

A. View Emma's vital signs and note the latest BGL—Use (say) the "ED Summary MPage Speech Rec" command.

B. Add a diagnosis for Emma—Use (say) the "Diagnosis Speech Rec" command.

Enter only the following data within the note:

Diagnosis: Chest discomfort.

Diagnosis Comments: BGL is "X.X."

(X.X = the BGL value found in Emma's vital signs)

C. Add an order for Emma—Use (say) the "Order Speech Rec" command.

Enter only the following data within the note:

Order: Full blood count (FBC).

Current clinical history (mandatory): Chest discomfort.

Clinician collect: No, **Collection date/time:** Today/Now.

D. Create a discharge note for Emma.

– Use (say) the "Discharge Referral Speech Rec" command.

Enter only the following data within the note:

Visit information—Summary of care: The patient appeared with chest discomfort.

Health status—Add diagnosis: *Include the "Active" diagnosis in the note.*

Appendix C Experiment Specifics

Specific Details of Experiment Systems

Materials

Participants were provided physical task sheets (including variables to input) and reference guides. The system usability survey was paper based, provided at the conclusion of each trail.

Software Systems

The system(s) used within the study were established after an investigation of the commonalities and necessities of currently available systems within the Australian marketplace. The chosen software packages were representative of common EHR systems, and speech recognition systems, covered all core elements for standard clinical documentation and were available for both research access and commercial use within Australia.

A test domain (including testing versions of EHR software) was used, with fictional predefined clinician and patient data created that was replicated for each study participant. Clinicians were logged into the system(s) with credentials specifically generated for the trial; no real/live patient data were accessed during the trial.

Permission to utilize these test domains was sought at both the local health district (LHD) level and through State Health.

EHR software

The EHR software used within the trial was FirstNet, the ED information management system component of Cerner's Millennium Health Network Architecture suite of products (E1 v2012.01.30, E2 v 2015.01.11).

Speech Recognition Hardware and Software

The speech recognition software was Nuance's Dragon Medical 360 Network Edition UK (E1 v2.0 - 12.51.200.072, E2 v2.4.2–12.51.214.037/045 with vSync enabled).

Monitoring Software

Data logging/screen capturing software: Specific data logging software allowed real-time monitoring of task steps, completion time, method comparison, and error capturing. TechSmith's Morae (v3.3.3) usability testing software was used throughout the trial.

Display Feed Capturing

An Elgato Game Capture HD60 was utilized as a secondary method of session recording.

Dragon Medical 360 Network Edition Data

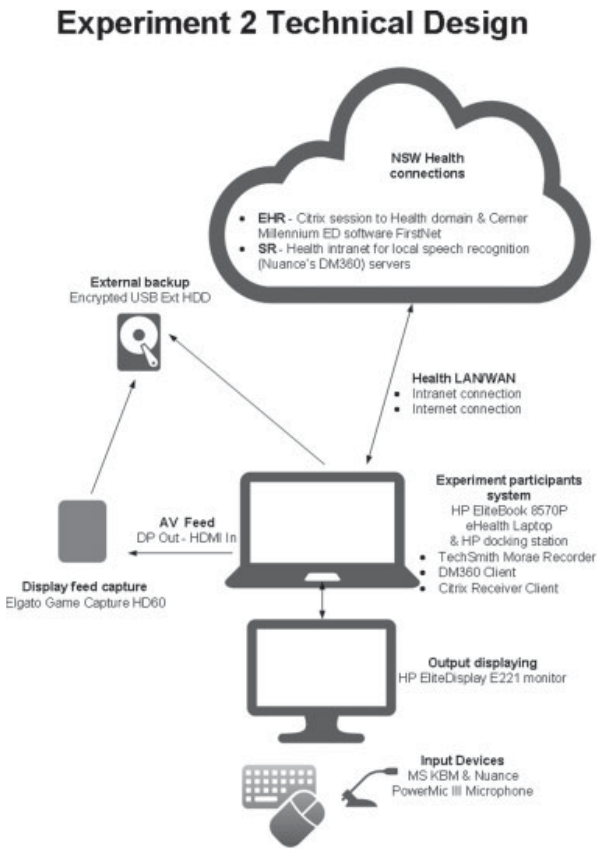
Trends or very simple usage statistics were available within the speech recognition software management console.

Hardware

A State Health laptop was configured to access the test domain (HP EliteBook 8570p). A microphone (E1 Nuance PowerMic II, E2 Nuance PowerMic III), USB keyboard and mouse (Microsoft Wired Desktop 600), and display monitor (HP EliteDisplay E221) were utilized by participants throughout the study.

Appendix D Experiment Technical Design

Diagram of Experiment Technical Configuration



Appendix E: Issues and Errors Observed

Summary of Issues and Errors Observed during Experiment 1

	Observed errors and issues
01	Incorrect patient
02	Incorrect patient—user corrected
03	No BGL entered
04	Incorrect BGL entered
05	Incorrect order collection date/method entered
06	Section of EHR missed
07	Data entered in incorrect EHR field
08	Section of EHR missed
09	Clinician closed EHR
10	User added trivial word (e.g., "and")
11	Omitted trivial word (e.g., "is")
12	Incorrect trivial word entered
13	Incorrect trivial word entered—user corrected
14	Incorrect significant word entered (diagnostic)
15	Omitted significant word (diagnostic)
16	Template brackets not removed (accept defaults missed)
17	Incorrect method of EHR menu navigation used
18	Word mangled (letters repeated or cut off)
19	Word mangled—user corrected (letters repeated or cut off)
20	Additional unnecessary word(s) (e.g., "and")
21	Misrecognition of word by SR
22	Misrecognition of word by SR—user corrected
23	All elements of command did not complete
24	Navigational command went nowhere or to wrong place/chart
25	EHR slow—system lag
26	EHR crashed
27	Element of EHR down, e.g., vitals
	Typographical errors
28	Missing full stop
29	Capitalization error
30	Missing comma(s)
31	Hyphen error
32	Plural form error (missing/added "s")
33	Spelling error

Appendix F Error Classification Tables

Errors Observed and Their Assigned Labels

	Potential patient harm errors			Integration/System, user or comprehension (error type)			Omission or commission errors (user error type)	
Observed error	**Minor**	**Moderate**	**Major**	**Integration/ System**	**User**	**Comprehension**	**Omission**	**Commission**
Incorrect patient—user corrected		X			X			X
No BGL entered		X			X		X	
Incorrect BGL entered			X		X			X
Data entered in incorrect EHR field			X		X			X
Section of EHR missed			X		X		X	
User-added trivial word (e.g., "and")	X					X		X
Omitted trivial word (e.g., "is")	X					X	X	
Incorrect significant word entered (diagnostic)			X		X			X
Omitted significant word (diagnostic)			X		X		X	
Template brackets not removed (accept defaults missed)	X				X		X	
Incorrect method of EHR menu navigation used	X				X			X
Additional unnecessary word (e.g., "and")	X			X		X		
Misrecognition of word by SR			X	X				
Misrecognition of word by SR—user corrected		X		X				
All elements of command did not complete	X			X				
Command went nowhere or to wrong place/chart	X			X				
EHR crashed	X			X				
Element of EHR down, e.g., vitals			X	X				
Typographical errors								
Missing full stop								
Capitalization error								
Hyphen error								
Plural form error (missing/added "s")								
Spelling error								
Space error								

Appendix G Repeat Participants Error Summary Table

Error Summary Table for Repeat Participants

Errors								
Experiment 1			Experiment 2			M-W Experiment 1 vs. Experiment 2		
	KBM	SR		KBM	SR		KBM	SR
Total errors observed	103	142	Total errors observed	62	84		*p*-Values	
Non-typographical	18	51	Non-typographical	8	49	Non-typographical		
Simple	9	23	Simple	2	31	Simple	0.457	0.682
Complex	9	28	Complex	6	18	Complex	0.473	0.106
Potential patient harm	*18*	*51*	*Potential patient harm*	*8*	*49*	*Potential patient harm*		
Major	6	22	Major	4	20	Major		
Simple	2	12	Simple	1	15	Simple	0.317	0.589
Complex	4	10	Complex	3	5	Complex	0.564	0.096
Moderate	2	6	Moderate	2	3	Moderate		
Simple	0	1	Simple	1	3	Simple	0.317	0.317
Complex	2	5	Complex	1	0	Complex	0.317	0.025
Minor	10	23	Minor	2	26	Minor		
Simple	7	10	Simple	0	13	Simple	0.109	0.608
Complex	3	13	Complex	2	13	Complex	0.655	1.000
Error mechanism	*18*	*57*	*Error mechanism*	*8*	*49*	*Error mechanism*		
Integration/System	1	28	Integration/System	0	33	Integration/System		
Simple	0	16	Simple	0	18	Simple	1.000	0.952
Complex	1	12	Complex	0	15	Complex	0.317	0.477
Use errors	11	22	Use errors	8	13	Use errors		
Simple	5	7	Simple	2	10	Simple	0.257	0.317
Complex	6	15	Complex	6	3	Complex	1.000	**0.006**
Comprehension	6	7	Comprehension	0	3	Comprehension		
Simple	4	0	Simple	0	3	Simple	0.196	0.392
Complex	2	7	Complex	0	0	Complex	0.277	0.406
Error genotype	*17*	*23*	*Error genotype*	*8*	*17*	*Error genotype*		
Omission	5	12	Omission	1	8	Omission		
Simple	2	7	Simple	1	8	Simple	0.317	0.655
Complex	3	5	Complex	0	0	Complex	0.083	0.059
Commission	12	11	Commission	7	9	Commission		
Simple	7	0	Simple	1	6	Simple	0.131	0.034
Complex	5	11	Complex	6	3	Complex	0.783	**0.005**
Typographical	*85*	*91*	*Typographical*	*54*	*35*	*Typographical*		
Simple	56	46	Simple	37	20	Simple	0.143	**0.010**
Complex	29	45	Complex	17	15	Complex	0.022	0.002

Appendix 7 – Supplementary materials paper 4

Supplementary materials available via online journal publication of the paper.

A) EHR task experiment overview

This appendix provides details of the electronic health record (EHR) tasks undertaken immediately prior to the System Usability Scale (SUS) questionnaire. The undertaking of these task formed the basis of participant's opinion of the systems which were measured within the usability study.

These tasks were part of a replication study of the authors earlier experiment assessing the efficiency and safety of speech recognition for EHR documentation.[1] A within subject experimental study where participants performed simulated clinical documentation tasks within an EHR with and without the assistance of speech recognition (SR).

Tasks performed

Tasks undertaken were representative of typical clinical documentation duties performed by emergency department clinicians and included: patient assignment, patient assessment, viewing vital signs, performing diagnosis, creating orders, and patient discharge. A total of four tasks were undertaken by participants, a simple task (2 sub-tasks) and a complex task (4 sub-tasks), both performed twice utilizing both input modalities keyboard and mouse (KBM) and SR.

Tasks Allocation

The order of task completion was randomized with the assistance of Research Randomizer software (Version 4.0). Half of the task were performed with keyboard and mouse as the input modality and half with the assistance of speech recognition.

Participants

Thirty-five emergency department clinicians volunteered from three urban teaching hospitals in Sydney, Australia. Previous training in the EHR system, including specific SR training (EHR 4h, SR 2h) was required for participation.

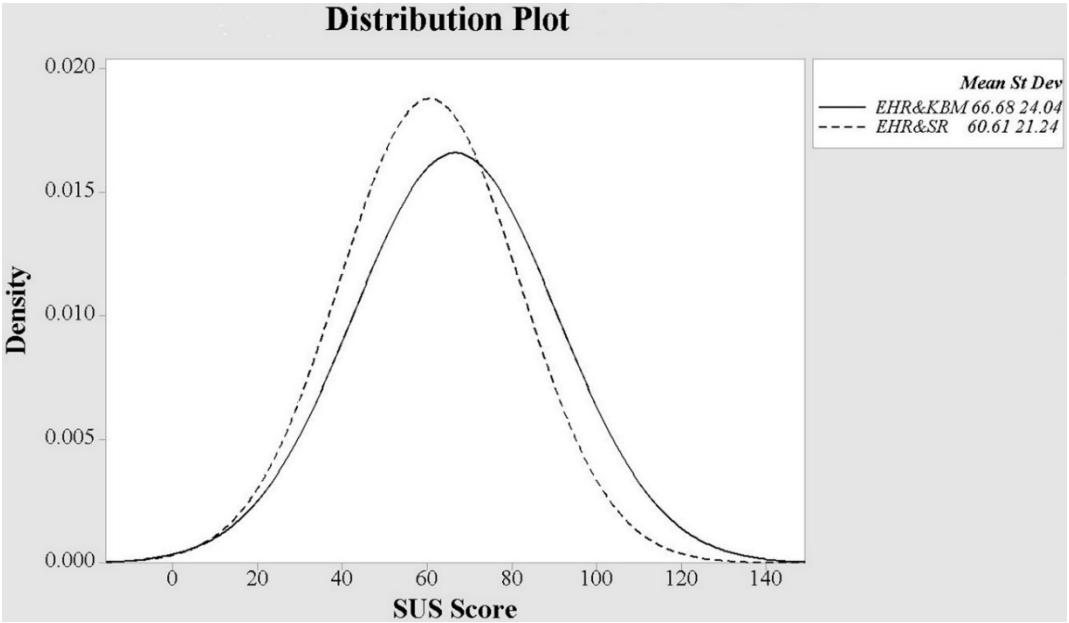
Systems Utilised

The clinical software used for the experiment was the Cerner Millennium suite with the FirstNet ED component (v2015.01.11) and Nuance Dragon Medical 360 Network Edition UK (v2.4.2) speech recognition software.

B) SUS score summary table

SUS Scores EHR, EHR with SR							
System	**Total**	**N**	**Mean**	**Max**	**Min**	**St. Dev**	**P-Value**
EHR with KBM	2333.75	35	66.68	100.00	0.00	24.04	0.115
EHR with SR	2121.25	35	60.61	100.00	0.00	21.24	0.118

C) Distribution plot of EHR and KBM vs EHR and SR



D) System usability scale questionnaire (after Brooke, Digital Equipment Corporation, 1986)

See over page

System Usability Scale

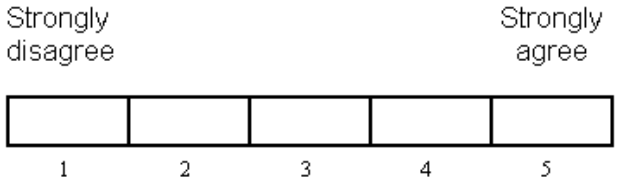
Based solely on the experience with the system you have just gained during the trial (not previous experience/knowledge).

Please rate the following statements on the Electronic Health Record system (EHR/EMR/FirstNet) on a scale of 1 to 5 (1-Strongly Disagree, 5-Strongly Agree).

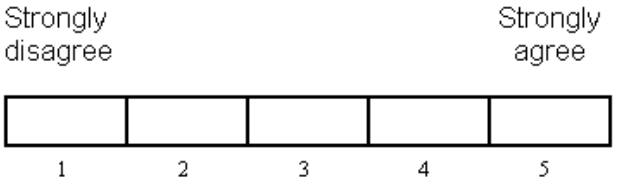
Firstly with keyboard & mouse as the input modality, and then secondly with the assistance of speech recognition (Dragon).

Q1. I think that I would like to use this system frequently

Keyboard & Mouse

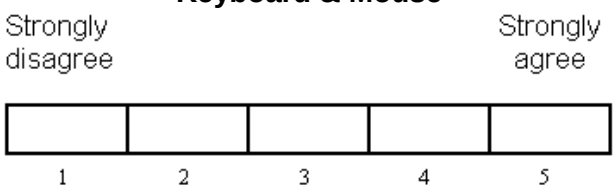


Speech Recognition Assisted

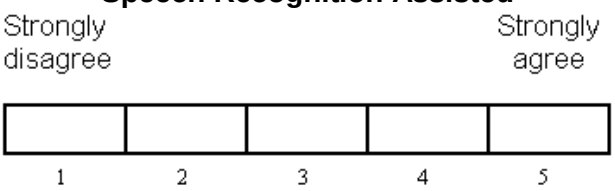


Q2. I found the system unnecessarily complex

Keyboard & Mouse

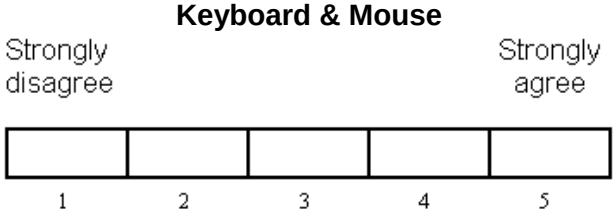


Speech Recognition Assisted

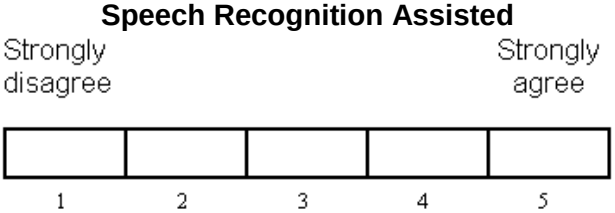


Q3. I thought the system was easy to use

Keyboard & Mouse

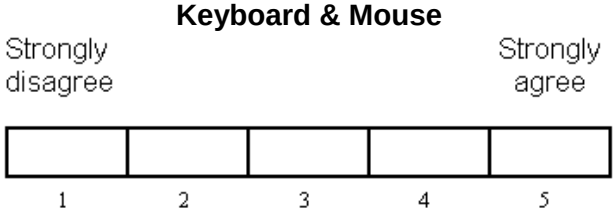


Speech Recognition Assisted

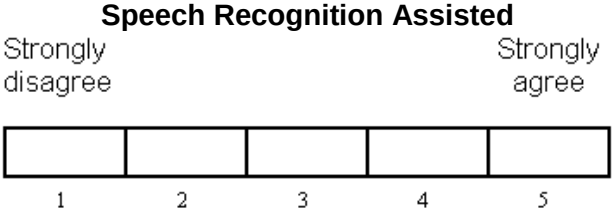


Q4. I think that I would need the support of a technical person to be able to use this system

Keyboard & Mouse



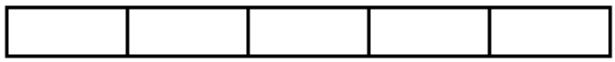
Speech Recognition Assisted



Q5. I found the various functions in this system were well integrated

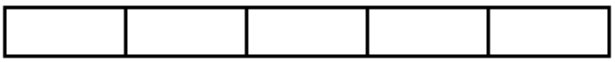
Keyboard & Mouse

Strongly disagree — Strongly agree



Speech Recognition Assisted

Strongly disagree — Strongly agree



Q6. I thought there was too much inconsistency in this system

Keyboard & Mouse

Strongly disagree — Strongly agree

Speech Recognition Assisted

Strongly disagree — Strongly agree

Q7. I would imagine that most people would learn to use this system very quickly

Keyboard & Mouse

Strongly disagree — Strongly agree

Speech Recognition Assisted

Strongly disagree — Strongly agree

Q8. I found the system very cumbersome to use

Keyboard & Mouse

Strongly disagree — Strongly agree

Speech Recognition Assisted

Strongly disagree — Strongly agree

Q9. I felt very confident using the system

Keyboard & Mouse

Strongly disagree — Strongly agree

Speech Recognition Assisted

Strongly disagree — Strongly agree

Q10. I needed to learn a lot of things before I could get going with this system

Keyboard & Mouse

Strongly disagree — Strongly agree

Speech Recognition Assisted

Strongly disagree — Strongly agree