

UNIVERSITY OF TECHNOLOGY SYDNEY

Faculty of Engineering and Information Technology

**Empirical Investigation of the Factors Influencing
Data Completeness in Electronic Medical Records**

by

Caihua Liu

A THESIS SUBMITTED
IN FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

Sydney, Australia

2019

Certificate of Original Authorship

I, Caihua Liu declare that this thesis, submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature:

Production Note:
Signature removed prior to publication.

Date:

October 28, 2019

Acknowledgements

This thesis would not be possible without the help and contribution of many others. Most significantly, I am much indebted to my advisor, Professor Didar Zowghi, for her guidance and support throughout my PhD study. I also would like to thank my external supervisor, Dr. Amir Talaei-Khoei who supported me throughout my journey of working on this research program. Professor Didar Zowghi and Dr. Amir Talaei-Khoei always pave the way for me and also their high standard and constructive criticisms have made me a better researcher.

I am also grateful to my collaborators, Professor Veda Catherine Storey, Professor Zhi Jin, and Dr. Jay Daniel. Without their help, my thesis would not be as good as it is now. I would like to thank Associate Professor Tapan Rai and Dr. Jean Lai who contributed valuable suggestions to my survey design. Another appreciation goes to the participants who dedicated their time for my study. I owe a great debt of gratitude to Professor Susan Williams for her insightful advices when I was a visiting student at University of Koblenz-Landau. It has been a pleasure to work with her. I also want to thank the members of Professor Williams' EOTLab, Patrick Nitschke and Clara Nitschke, for their warmest welcome and support when I stayed at Koblenz.

Last but not least, I could not have done this without the encourage and support of my family and my friends. They cheered me up when I was feeling depressed and tried their best to offer me help when I needed it.

Caihua Liu

Sydney, Australia, 2019

Dedications

To my beloved family members, my sister Jianghua Liu, and my parents Kerong Liu and Deling Mo, for their unwavering support and best wishes.

List of Papers

Journal papers:

J-1: **Liu, C.**, Talaei-Khoei, A., Zowghi, D. & Daniel, J. 2017, 'Data completeness in healthcare: A literature survey', *Pacific Asia Journal of the Association for Information Systems*, vol. 9, no. 2, pp.75-100. (Published)(Section 2.1)

J-2: **Liu, C.**, Zowghi, D. & Talaei-Khoei, A. 2019, 'An empirical study of the antecedents of data completeness in electronic medical records', *International Journal of Information Management*. (Published)(Section 6.2)

J-3: **Liu, C.**, Zowghi, D., Talaei-Khoei, A. & Jin Z. 'Empirical study of Data Completeness in Electronic Health Records in China'. (Submitted to the *Pacific Asia Journal of the Association for Information Systems*) (Chapter 7)

J-4: **Liu, C.**, Zowghi, D., & Talaei-Khoei, A. 'Factors influencing data completeness in electronic medical records: A cross-country comparison'.(Will be submitted to a relevant journal) (Chapter 8)

Conference papers:

C-1: **Liu, C.**, Zowghi, D., Talaei-Khoei, A., & Daniel, J. 2018, 'Achieving data completeness in electronic medical records: A conceptual model and hypotheses development', *51st Hawaii International Conference on System Sciences*, University of Hawaii at Manoa, Association for Information Systems IEEE Computer Society Press, Waikoloa Village, Hawaii, pp. 2824-2833. (Published) (Chapter 5)

C-2: **Liu, C.**, Talaei-Khoei, A. & Zowghi, D. 2018, 'Theoretical support for enhancing data quality: Application in electronic medical records', *24th Americas Conference on Information Systems*, Association for Information Systems, New Orleans, LA, pp. 1-10 (Published)(Chapter 4)

Table of Contents

| | |
|--|-------------|
| Certificate of Original Authorship | ii |
| Acknowledgements | iii |
| Dedications..... | iv |
| List of Papers..... | v |
| Table of Contents | vi |
| List of Figures..... | xii |
| List of Tables..... | xiii |
| Abstract..... | xv |
| Chapters..... | 1 |
| Chapter 1 - Introduction | 1 |
| 1.1 Research background and problem statement | 1 |
| 1.1.1 Related studies on factors influencing data completeness in electronic medical records..... | 2 |
| 1.1.2 Related studies on factors influencing data completeness as a dimension of data quality..... | 3 |
| 1.2 Objectives of the research and research questions..... | 4 |
| 1.3 Significances of the research..... | 6 |
| 1.4 Scope of the research..... | 7 |
| 1.4.1 Content scope | 7 |
| 1.4.2 Geographical scope..... | 8 |
| 1.4.3 Subject scope | 8 |
| 1.5 Overview of the research methods..... | 9 |
| 1.5.1 Literature review..... | 10 |
| 1.5.2 Conceptual model development..... | 10 |
| 1.5.3 Empirical evaluation..... | 10 |
| 1.6 Roadmap of the thesis..... | 11 |
| Chapter 2 – Literature Review | 13 |
| 2.1 Landscape analysis of initial set of papers..... | 13 |
| 2.1.1 Introduction..... | 13 |
| 2.1.2 Methods..... | 16 |
| 2.1.3 Results..... | 21 |
| 2.1.4 Discussion | 30 |

| | |
|---|-----------|
| 2.1.5 Conclusion | 40 |
| 2.2 Systematical review of literature related to all aspects of the thesis | 41 |
| 2.2.1 Introduction..... | 41 |
| 2.2.2 Theoretical background | 44 |
| 2.2.3 Research method..... | 45 |
| 2.2.4 Results and discussion | 47 |
| 2.2.5 Conclusion | 52 |
| Chapter 3 – Research Design | 56 |
| 3.1 Qualitative and quantitative approaches..... | 56 |
| 3.2 Research design..... | 57 |
| 3.2.1 Participants sampling | 59 |
| 3.2.2 Instrument development..... | 60 |
| 3.2.3 Data collection | 61 |
| 3.2.4 Data analysis | 65 |
| 3.3 Research assumptions..... | 68 |
| 3.4 Research ethics | 69 |
| 3.4 Conclusion | 70 |
| Chapter 4 – Theoretical Support for the Development of the Conceptual Model..... | 71 |
| 4.1 Introduction..... | 71 |
| 4.2 What is data quality..... | 73 |
| 4.3 Research methods | 74 |
| 4.3.1 Selection of research papers..... | 74 |
| 4.3.2 Conceptualisation of characteristics of related studies | 75 |
| 4.4 Comparison of theoretical support for enhancing data quality | 78 |
| 4.4.1 Breadth dimension | 80 |
| 4.4.2 Depth dimension | 83 |
| 4.4.3 Interaction dimension..... | 84 |
| 4.5 Conclusion | 86 |
| Chapter 5 – Conceptual Model..... | 89 |
| 5.1 Introduction..... | 89 |
| 5.2 Data completeness as a dimension of data quality..... | 90 |
| 5.3 Theoretical support..... | 91 |
| 5.3.1 Review of existing models..... | 91 |
| 5.3.2. Utilising the model of factors influencing data quality management in addressing | |

| | |
|---|------------|
| data completeness in EMR..... | 94 |
| 5.4 Hypotheses development and literature support | 95 |
| 5.4.1. Relationship between clinic director’s support for EMR implementation with regulatory capability for EMR-enabled care processes toward data completeness in EMR..... | 96 |
| 5.4.2. Relationship between clinic director’s support for EMR implementation with EMR alignment to care processes toward data completeness in EMR..... | 97 |
| 5.4.3. Relationship between resources with regulatory capability for EMR-enabled care processes toward data completeness in EMR | 97 |
| 5.4.4. Relationship between resources with EMR alignment to care processes toward data completeness in EMR | 98 |
| 5.4.5. Relationship between resources with EMR integration toward data completeness in EMR..... | 98 |
| 5.4.6. Relationship between regulatory capability for EMR-enabled care processes with clinical staff’s participation toward data completeness in EMR..... | 99 |
| 5.4.7. Relationship between regulatory capability for EMR-enabled care processes with EMR integration toward data completeness in EMR | 100 |
| 5.4.8. Relationship between regulatory capability for EMR-enabled care processes and data completeness in EMR..... | 100 |
| 5.4.9. Relationship between EMR alignment to care processes with clinical staff’s participation toward data completeness in EMR..... | 101 |
| 5.4.10. Relationship between EMR alignment to care processes with EMR integration toward data completeness in EMR..... | 101 |
| 5.4.11. Relationship between clinical staff’s participation and data completeness in EMR | 102 |
| 5.4.12. Relationship between EMR integration and data completeness in EMR..... | 102 |
| 5.4.13. Relationship between clinical staff’s participation and EMR integration toward data completeness in EMR..... | 103 |
| 5.5 Construct operationalisation | 104 |
| 5.6 Conclusion | 104 |
| 5.6.1. Theoretical contributions | 104 |
| 5.6.2. Practical contributions | 105 |
| Chapter 6 – Factors Influencing Data Completeness in Electronic Medical Records in USA..... | 107 |
| 6.1 Data examination | 107 |
| 6.1.1 Evaluation of missing data..... | 107 |
| 6.2.2 Identification of outliers..... | 109 |

| | |
|--|------------|
| 6.2.3 Testing of assumptions underlying multivariate techniques | 110 |
| 6.2 Factors influencing data completeness in electronic medical records in USA..... | 111 |
| 6.2.1 Introduction..... | 112 |
| 6.2.2 Literature review..... | 114 |
| 6.2.3 Hypothesis development..... | 117 |
| 6.2.4 Research method..... | 124 |
| 6.2.5 Results..... | 127 |
| 6.2.6 Discussion..... | 134 |
| 6.2.7 Conclusion | 144 |
| Chapter 7 – Factors Influencing Data Completeness in Electronic Medical Records in China | 146 |
| 7.1. Introduction..... | 146 |
| 7.2 Theoretical background | 149 |
| 7.2.1 Definition of data completeness..... | 149 |
| 7.2.2 Related studies and rationales of using Xiao, Xie and Wan’s model..... | 150 |
| 7.3 Conceptual model | 151 |
| 7.3.1 Clinic Director’s Support for EMR Implementation, Regulatory Capability of EMR-Enabled Care Processes, and EMR Alignment to Care Processes | 152 |
| 7.3.2 Resources, Regulatory Capability of EMR-Enabled Care Processes, and EMR Alignment to Care Processes..... | 153 |
| 7.3.3 Regulatory Capability of EMR-Enabled Care Processes, Clinical Staff’s Participation, EMR Integration, and Data Completeness in EMR | 154 |
| 7.3.4 EMR Alignment to Care Processes, Clinical Staff’s Participation, and EMR Integration..... | 155 |
| 7.3.5 Clinical Staff’s Participation and Data Completeness in EMR | 155 |
| 7.3.6 EMR Integration, Clinical Staff’s Participation, and Data completeness in EMR..... | 156 |
| 7.4. Research method..... | 157 |
| 7.4.1 Measures of constructs and instrument development | 157 |
| 7.4.2 Research sample and data collection | 158 |
| 7.4.3 Sample size | 160 |
| 7.4.4 Control variable | 160 |
| 7.5 Results..... | 162 |
| 7.5.1 Measurement model..... | 162 |
| 7.5.2 Structural model..... | 165 |

| | |
|--|------------|
| 7.6 Discussion..... | 166 |
| 7.6.1 Implications for academic..... | 170 |
| 7.6.2 Implications for practice | 171 |
| 7.6.3 Limitations and future work..... | 173 |
| 7.7 Conclusion..... | 174 |
| Chapter 8 – Factors Influencing Data Completeness in Electronic Medical Records: A Cross-Country Comparison | 175 |
| 8.1 Introduction..... | 175 |
| 8.2 Theoretical background | 178 |
| 8.2.1 Factors influencing data completeness in EMR..... | 178 |
| 8.2.2 Role of culture in data quality..... | 181 |
| 8.3 Research methodology | 184 |
| 8.3.1 Operationalisation of constructs and instrument..... | 184 |
| 8.3.2 Data collection and participants..... | 186 |
| 8.3.3 Research site and sample | 187 |
| 8.3.4 Control variables | 190 |
| 8.4 Data analysis and results..... | 191 |
| 8.4.1 Measurement model..... | 192 |
| 8.4.2 Structural model..... | 194 |
| 8.4.3 Cross-country comparison | 196 |
| 8.5 Discussion..... | 197 |
| 8.5.1 Theoretical implications..... | 200 |
| 8.5.2 Practical implications..... | 201 |
| 8.5.3 Limitations and future study | 204 |
| 8.6 Conclusion..... | 205 |
| Chapter 9 – Conclusion | 207 |
| 9.1 Overview..... | 207 |
| 9.2 Findings and contributions of the research..... | 209 |
| 9.2.1 Findings for the research questions..... | 209 |
| 9.2.2 Theoretical implications..... | 213 |
| 9.2.3 Practical implications..... | 216 |
| 9.3 Limitations of the research | 220 |
| 9.4 Future work..... | 221 |
| References..... | 224 |

| | |
|---|------------|
| Appendices..... | 254 |
| Appendix A: Data extraction form in the literature review | 254 |
| Appendix B: List of the included 54 studies for the SLR..... | 255 |
| Appendix C: Sources of the survey questions..... | 258 |
| Appendix D: Invitation letter | 262 |
| Appendix E: Questionnaire of the factors influencing data completeness in EMR | 264 |
| Appendix F: Normality test | 275 |
| Appendix G: Results of scree test for the USA data samples..... | 278 |
| Appendix H: Results of full collinearity test for inner VIF value for the USA data samples | 279 |
| Appendix I: Results of scree test for the Chinese data samples | 280 |

List of Figures

| | |
|--|--------------|
| Figure 1. The process of identifying the relevant papers in the review of data completeness in healthcare..... | 19 |
| Figure 2. Distribution of study themes about evaluation in the literature review of data completeness in healthcare..... | 27 |
| Figure 3. Lifecycle of data completeness management in healthcare in this review... | 30 |
| Figure 4. Data extraction, synthesis and analysis process in the SLR..... | 47 |
| Figure 5. Study steps of the entire research..... | 58 |
| Figure 6. The model of factors influencing data quality management taken from Xiao, Xie and Wan (2009)..... | 94, 117, 151 |
| Figure 7. The conceptual model of factors influencing data completeness in EMR being examined in this study (Liu et al. 2018)..... | 96, 117, 157 |
| Figure 8. Results of the structural model based in the USA data samples..... | 132 |
| Figure 9. Results of the structural model based in the Chinese data samples... | 166, 195 |
| Figure 10. The conceptual model taken from Liu, Zowghi and Talaei-Khoei (2019).. | 179 |
| Figure 11. Results of the structural model based in the Australian data samples..... | 195 |

List of Tables

| | |
|--|-----|
| Table 1. Alignment of research objective and research questions..... | 5 |
| Table 2. Alignment of research objective and significances of the research..... | 6 |
| Table 3. List of journals selected in the review of data completeness in healthcare.... | 18 |
| Table 4. Distribution on the forms of healthcare records in the review of data completeness in healthcare..... | 21 |
| Table 5. Definition and measure of data completeness in the review of data completeness in healthcare | 23 |
| Table 6. Study themes related to data completeness in healthcare in this review..... | 24 |
| Table 7. Data completeness assessment: Issues and methods in this review..... | 36 |
| Table 8. Determinants of data completeness in healthcare: Difficulties and methods in this review..... | 37 |
| Table 9. Distribution of papers in each database in the SLR..... | 46 |
| Table 10. Categories of the factors that influence data completeness in electronic records in the SLR..... | 48 |
| Table 11. Research activities to address research objectives and research questions... | 57 |
| Table 12. Ending conditions used in the taxonomy development..... | 76 |
| Table 13. Characteristics of related studies on factors influencing data quality..... | 79 |
| Table 14. Characteristics of related studies on factors influencing data quality and the conceptual model proposed in this thesis..... | 92 |
| Table 15. Missing data patterns for the data collected from USA in this research..... | 108 |
| Table 16. Detection results of univariate and multivariate outliers in this research... | 109 |
| Table 17. Results of testing univariate normality in this research..... | 111 |

| | |
|---|-----|
| Table 18. Definition of EMR data completeness taken from Weiskopf et al. (2013) (pp. 831-832)..... | 115 |
| Table 19. Demographics of the USA respondents based on their characteristics..... | 127 |
| Table 20. Results of factor analysis based in the data samples collected from USA... | 129 |
| Table 21. Convergent and discriminant validity of the measurement model based in the data samples collected from USA..... | 131 |
| Table 22. Hypothesis results based in the USA data samples..... | 132 |
| Table 23. Demographics of the Chinese respondents based on their characteristics .. | 160 |
| Table 24. Demographics of the Chinese respondents based on their workplace..... | 161 |
| Table 25. Results of factor analysis based in the data samples collected from China... | 163 |
| Table 26. Convergent and discriminant validity of the measurement model based in the data samples collected from China..... | 165 |
| Table 27. Hypotheses results based in the data samples collected from China..... | 165 |
| Table 28. Theoretical support on the effect of culture on quality management..... | 182 |
| Table 29. Demographic analysis of participants for Australia and China..... | 189 |
| Table 30. Results of factor analysis for the Australian and Chinese data samples..... | 192 |
| Table 31. Results of convergent and discriminant validity for the Australian and Chinese data samples..... | 194 |
| Table 32. Test of differences between individualistic and collectivistic culture..... | 197 |
| Table 33. A summary of dimensions and characteristics of conceptual models for factors influencing data quality taken from Liu, Talaei-Khoei and Zowghi (2018)..... | 212 |

Abstract

Recent advances in electronic medical records (EMR) in healthcare highlight the role of high-quality data, because poor data quality can destabilise clinical decision-making that in turn may impact on the quality of care. Data completeness as an essential dimension of data quality has been the subject of much research in the last three decades, while addressing data completeness in EMR is still a significant challenge in healthcare.

Exploring the factors that influence data completeness could be a starting point to address data quality issues in EMR. Constructing a conceptual model can be an effective exercise for studying complex phenomena such as data completeness in EMR. We do not have a conceptual model of factors influencing data completeness in EMR and specific relationships between these factors remain unclear. This study presents a novel conceptual model of the factors influencing data completeness in EMR and investigates the relationships between them.

This thesis begins with presenting the first literature review of data completeness in healthcare and provides a novel synthesis of the identified challenges that has led to a categorisation of the factors influencing data completeness in electronic records from human, managerial, technical, and external perspectives. The thesis also presents a taxonomy of factors influencing data quality using breadth, depth, and interaction dimensions, and suggests guidelines for developing and comparing the factors influencing data quality.

The empirical evaluation of the proposed conceptual model is carried out by surveying clinical practitioners in Australia, China, and USA. The results of data analysis reveal that the priority of the included factors is different among the three countries. Moreover, cultural differences are identified and highlighted in the relationships between these factors. The findings of this study draw specific attention to the important factors and help clinical practitioners identify which of these areas need to be addressed in order to improve EMR data completeness. In addition, the conceptual model proposed in this thesis can serve as the basis for the development of tools, methods and techniques for addressing data completeness in EMR.