Pacific Asia Journal of the Association for Information Systems
ISSN 1943-7544
Research Paper doi: 10.17705/1pais.12204 Volume 12, Issue 2 (2020)

Empirical study of Data Completeness in Electronic Health Records in China

Caihua Liu[1,*], Didar Zowghi[2], Amir Talaei-Khoei[3], Zhi Jin[4]

[1]Sun Yat-Sen University, China, liucaih5@mail.sysu.edu.cn
[2]University of Technology Sydney, Australia, Didar.Zowghi@uts.edu.au
[3]University of Nevada Reno, USA; University of Technology Sydney, Australia, atalaeikhoei@unr.edu
[4] Peking University, China, zhijin@pku.edu.cn

Abstract

Background: *As a dimension of data quality in electronic health records (EHR), data completeness plays an important role in improving quality of care. Although many studies of data management focus on constructing the factors that influence data quality for the purpose of quality improvement, the constructs that are developed for interpreting factors influencing data completeness in the EHR context have received limited attention.*

Methods: *Based on related studies, we constructed the factors influencing EHR data completeness in a conceptual model. We then examined the proposed model by surveying clinical practitioners in China.*

Results: *Our results show that the data quality management literature can serve as a starting point to derive a conceptual model of factors influencing data completeness in the EHR context. This study also demonstrates that "resources" should be added as a factor that influences data completeness in EHR.*

Conclusion: *Our resulting conceptual model shows a substantial explanation of data completeness in EHR assessed in this study. Although the proposed relationships between the included factors were previously supported in the literature, our work provides the beginning empirical evidence that some relationships may not be always significantly supported. The possible explanation of these differences has been discussed in the present research. This study thus benefits decision makers and EHR program managers in implementing EHR as well as EHR vendors in the EHR integration by addressing data completeness issues.*

Keywords: Data Completeness, Electronic Health Records, Conceptual Model, China.

Citation: Liu, C., Zowghi, D., Talaei-Khoei, A., & Jin, Z. (2020). Empirical study of Data Completeness in Electronic Health Records in China. *Pacific Asia Journal of the Association for Information Systems, 12*(2), 103-128. https://doi.org/10.17705/1pais.12204

Introduction

Electronic health records (EHR) integrate data from medical systems and repositories, enabling clinical practitioners to access required data on the process and outcomes of patient care (Lin et al., 2019). This aggregated data supports decision making, research, and planning that is increasingly seen as a promising vehicle to improve quality of care (Masrom & Rahimly, 2015). For instance, as noted by Hydari et al. (2018), EHR have led to a 27% drop in events about patient safety in Pennsylvania hospitals over 2005–2012, which improve patient safety and suggest large-scale economic benefits. However, these advances of EHR in healthcare highlight the role of high-quality data, because poor data quality is the principal barrier to effective clinical decision making (Foshay & Kuziemsky, 2014; Kumar et al., 2018).

Data quality is defined in the ISO 25012 Standard as "*the degree to which data satisfy the requirements defined by the product-owner organization*", and can be reflected through its dimensions such as completeness and accuracy (ISO 25000 Portal, 2019). Data completeness that presents the degree to which all needed data for a specific task is available (Wang & Strong, 1996), is viewed as an essential data quality dimension in healthcare for two reasons. First, addressing missing data and interpreting the processed incomplete data require more effort, and incompleteness is considered as the primary challenge in dealing with data availability for reuse in healthcare research and planning (Weiskopf & Weng, 2013). Second, incomplete data generated at the point of care can result in diagnostical errors for a patient. For example, missing data in a clinical diagnostic support system led to poor recommendations on 77% clinical encounters (Berner et al., 2005). Understanding the factors influencing data completeness therefore is of primary importance to address data quality in EHR.

Data completeness in healthcare has only been reviewed partially in prior literature on data quality (Johnson et al., 2015; Mashoufi et al., 2018). A systematic review focusing on data completeness in the area has recently been published (Liu et al., 2017). The existing study themes related to data completeness in the EHR context include: assessing completeness for data practices (Estiri et al., 2019), analyzing determinants contributing to data completeness (Johnson et al., 2017), and developing methods to improve data completeness (Li et al., 2019). The identification of the factors influencing EHR data completeness allows clinical practitioners to discover problem areas, while revealing relationships between them could potentially help preserve complete data or systematically reduce missing data. Hence, establishing interactions between these factors plays an important role in achieving complete EHR data for improving quality of care. However, limited attention paid to empirical investigation of the interactions between the factors.

A large body of literature has investigated factors influencing data quality (e.g. Al-Hiyari et al. (2013), Tee et al. (2007), Xu et al. (2002), and Zellal and Zaouia (2015)). Recent developments in that strand of literature (Xiao et al., 2009; Xu, 2013) propose that data management contributes to data quality. However, the constructs used for explaining the factors influencing data completeness as a dimension of data quality in the context of EHR requires further exploration.

Liu et al. (2020) empirically examined the factors that influence the EHR data completeness, based on the data collected from the medical professionals in Nevada, USA, however, studies that include more clinical practitioners in other countries contribute to better understanding these factors and/or increasing the generalization of research findings. Therefore, in the study a new construct "resources" and new relationships between the constructs are added into the conceptual model of factors influencing data completeness in EHR that extends the prior work (Xiao et al., 2009). Then, we evaluated the extended model through surveying clinical practitioners who were working at healthcare settings [1] across China. Data completeness problems can appear due to various reasons such as missing records for an entity or missing values in a record (Liu et al., 2016). Here we look at data incompleteness as missing record (a row of the table), missing attribute (a column of the table), and missing value (a cell of the table) in EHR. Because data completeness is defined for use and requirements (De Feo & Juran, 2017), four perspectives are proposed by Weiskopf et al. (2013) to define data completeness (documentation, breadth, density, and predictive), which elaborate the most frequently used situations in determining data completeness in the context of EHR. These perspectives are further used to construct and measure EHR data completeness in general in the present work. Data and information are two different concepts, however, data completeness and information completeness are not distinguished here for simplifying the presentation of the work. Accordingly, our study aims to examine the factors that influence EHR data completeness and answer the following question:

Research Question: What are the factors influencing data completeness in electronic health records?

As mentioned, Liu et al. (2020) examined the factors that influence EHR data completeness with the empirical data focussing on medical professionals in Nevada, however, our current study differs from their work in at least four ways. *Firstly*, this study extends the existing work by empirically evaluating the extended conceptual model and testing the proposed hypotheses by surveying clinical practitioners in China. *Secondly*, in this study we developed a Chinese version of the survey questionnaire that can be viewed as a new instrument to measure the factors influencing data completeness in EHR. *Thirdly*, we will show that the significant factors are “resources”, “EHR alignment to care processes”, “regulatory capability for EHR-enabled care processes”, and “EHR integration”, and reveal the extent to which these factors affected EHR data completeness. These findings that differ from the results reported by Liu et al. (2020) are discussed with their possible explanation in the context of the present study. *Lastly*, this study presents further implications in the context of our findings for decision makers at healthcare settings to address EHR data completeness through ensuring sufficient human resources, time, and funding for EHR implementation. We also inform EHR program managers to address EHR data completeness by improving regulations and procedures on data quality management and process management for EHR-enabled care processes, and customizing EHR at healthcare settings and enhancing communication with clinical staff.

The remaining of this paper is organized as follows: we first review related studies in the next section; after that the hypotheses for the factors influencing data completeness in EHR are developed; then the method of operationalization of the constructs, data collection, and analysis are followed; thereafter we provide the results of data analysis and discuss our research findings; and lastly we conclude the present study.

Theoretical Background

Data Completeness

In this section, we give the definition related to data completeness. Researchers indicated that quality is defined for use and requirements (De Feo & Juran, 2017), and this could serve as an analytical framework to differentiate the quality of data and information (Tilly et al., 2017). For example, data is believed to be the fact that result from the observation of physical phenomena, while information is considered to be the output from the data refined via some processes (Fox et al.,1994). Data completeness refers to a data quality dimension, and

[1] Healthcare Settings refer to public, private, and clinical areas of healthcare facilities such as hospitals, clinics, and nursing houses (The NOAH Professionalization Committee, 2018).

therefore the literature of data quality could serve as a basis to construct the factors that influence EHR data completeness. As mentioned in the Introduction, Weiskopf et al. (2013) have defined EHR data completeness from four perspectives. In this study, we also utilized these perspectives to construct and evaluate EHR data completeness, as presented in Appendix A.

Related Studies and Rationales of Using Xiao et al. Model

This section describes related studies on data quality. Since data completeness is a data quality dimension, previous studies of factors that influence data quality could provide some theoretical perspective for constructing a conceptual model of the factors that influence EHR data completeness. For instance, Xu et al. (2002) identified the factors that influence data quality in enterprises and divided them into several groups. Thereafter, a few researchers (e.g. Nord et al. (2005), Xu and Lu (2003), and Xu (2013)) continued the investigation on the factors derived from Xu et al. (2002). Other studies also examined the factors influencing data quality in their proposed conceptual models and gained insights into the relationships between these factors, based on empirical evidence (e.g. Al-Hiyari et al. (2013), Tee et al. (2007), Kokemueller (2011), Xiao et al. (2009), and Wixom and Watson (2001)).

As mentioned in the Introduction, understanding interactions between the factors that influence data quality help better achieve quality-assured data. Therefore, building up relationships between these factors can make prominent contributions to ensuring data quality that is given priority when studying the factors. A few researchers have paid attention to these relationships (Al-Hiyari et al., 2013; Kokemueller, 2011; Tee et al., 2007; Xiao et al., 2009). These four conceptual models therefore serve as a starting point that helps construct factors influencing EHR data completeness in a conceptual model. Since top management commitment and staff members' engagement determine the success of data quality management (Al-Hiyari et al., 2013; Xiao et al., 2009), the participation from both staff members and top management level needs to be taken into account when looking at the factors affecting data quality.

Based on this analysis, the two studies (Al-Hiyari et al., 2013; Xiao et al., 2009) developed more comprehensive models for explaining the factors influencing data quality. However, Al-Hiyari et al. (2013) employed the empirical data collected from students instead of practitioners to evaluate the proposed model, which may have decreased the results' validity. Thus, the model of Xiao et al. (2009) is a better choice of the starting point to construct the factors influencing EHR data completeness in a conceptual model. Their investigated factors are: "top management support (TMS)", "business-IT alignment (BITA)", "capability on the regulation" and "process management (CRPM)", "integration of information systems (IIS)", and "staff participation (SP)", as shown in Figure 1.

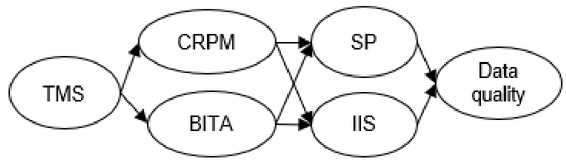


Figure 1 - Xiao et al. (2009)'s Model

Conceptual Model

Based on the model of Xiao et al. (2009), a conceptual model of factors that influence EHR data completeness is derived. Researchers revealed that "resources (including time, human resources, and funding) have positive and significant impacts on the success of organizational implementation for data quality improvement (concerning the extent to which organisations

https://aisel.aisnet.org/pajais/vol12/iss2/4
DOI: 10.17705/1pais.12204

address the issues such as change and/or process management and widespread support to improve data quality) in empirical studies (Kokemueller, 2011; Wixom & Watson, 2001). This success is further positively associated with data quality (Kokemueller, 2011). Essentially, the "organizational implementation success" (OIS) for data quality improvement implies a "capability on the regulation and process management" (CRPM) to improve data quality in organizations. The definition of OIS is similar with the content contained in CRPM from Xiao et al. (2009). Thus, "resources" could also have similar impacts on CRPM to address data completeness and this factor is added into the resulting conceptual model. Furthermore, a body of literature discussed in Section 3.2 also support and help hypothesize the relationships between "resources" and other included factors for addressing data completeness in EHR. Seven included constructs are "resources" (herein, RSC), "clinic director's support for EHR implementation" (herein, DSI, "EHR alignment to care processes" (herein, AGM), "regulatory capability for EHR-enabled care processes" (herein, RGC), "EHR integration" (herein, IGT), "clinical staff's participation" (herein, SPT), and "data completeness in EHR" (herein, DC).

Accordingly, the extended conceptual model includes the participation from both top management level (i.e. clinic director) and staff members (i.e. clinical staff). Meanwhile, the conceptual model considers: (1) the relationships between DC and the factors, and (2) the relationships between these factors. We defined all included constructs being examined in this study, as shown in Table 1.

Table 1 - Definitions of the Constructs that are Consistent with Liu et al. (2020)

Construct	Definition
Clinic director's support for EHR implementation	clinic directors are aware of the importance of EHR implementation toward data completeness and make their contributions to relevant activities (adapted from Xiao et al. (2009))
Resources	human resource, time, and funding needed for conducting EHR implementation (adapted from Wixom and Watson (2001))
Regulatory capability for EHR-enabled care processes	the capability of dealing with regulations formulation and processes management to enable EHR-enabled care processes at healthcare settings (adapted from Xiao et al (2009))
EHR alignment to care processes	EHR fit into workflow during care delivery (adapted from Herzberg et al. (2011) and Xiao et al. (2009))
Clinical staff's participation	clinical staff are aware of the importance of EHR data completeness and make their contributions to addressing EHR data completeness (adapted from Xiao et al. (2009))
EHR integration	EHR are integrated across different data sources, aggregating data for multiple practices (adapted from Aanestad et al. (2017))
Data completeness in EHR"	available EHR data for documentation, breadth, density, and predictive use (used definitions of completeness from Weiskopf et al. (2013))

Clinic Director's Support for EHR Implementation, Regulatory Capability of EHR-Enabled Care Processes, and EHR Alignment to Care Processes

Prior studies indicated that the commitment of top management facilitates the formulation of regulation and enhances the process management in IS implementation (Kokemueller, 2011; Wixom & Watson, 2001). Furthermore, Xiao et al. (2009) reported that attitude and knowledge of decision makers toward data quality could determine how much the managerial support to relevant activities can be carried out as well as regulations and processes supporting these activities can be motivated. As implementation of information systems (IS) involves the process that aligns business with IT, top management level is needed to make decisions on relevant activities to achieve this alignment (Xiao et al., 2009). Similarly, at healthcare settings, regulatory capability of EHR-enabled care processes also relies on clinic directors' support. Clinic directors' understanding and knowledge about EHR could also determine the extent to which communication between clinical staff and IT professionals to implement EHR can be driven (Gopalakrishna-Remani et al., 2018). This alignment facilitates EHR implementation at healthcare settings to achieve complete data in EHR. We proposed:

H1: Higher levels of clinic director's support for EHR implementation are positively associated with regulatory capability for EHR-enabled care processes.

H2: Higher levels of clinic director's support for EHR implementation are positively associated with EHR alignment to care processes.

Resources, Regulatory Capability of EHR-Enabled Care Processes, and EHR Alignment to Care Processes

Since implementation of IS costs time and money, resources cannot be ignored in this implementation (Kokemueller, 2011; Wixom & Watson, 2001). Researchers reported that the amount of time and human resources assigned to address IS implementation could influence the project timeline (Wixom & Watson, 2001). For healthcare settings, without sufficient time, human resources, and funding for EHR implementation, it is unlikely to conduct the activities on institutional and process management for EHR-enabled care processes (Shaw, 2014). As EHR implementation contains a process of (1) alignment between workflow of EMR and care processes as well as (2) communication between clinical staff and IT professionals, users' needs related to EHR integrated into their workflow are required to address by interviewing users about use of the current system (Staff et al., 2016). Hence, a network of users should be organized and supported in the process of interviews, within a set of considerable resources (Kushniruk & Nøhr, 2016). Furthermore, EHR are scattered across multiple systems and data repositories that require large investment on technologies such as integration technologies (Nkanata et al., 2018). EHR integration cannot overlook the role of resources. For healthcare organizations, if their time, IT professionals, and funding for EHR implementation are sufficient, EHR integration could have better chances to be dealt with that contributes to aggregating high-quality data in EHR. We proposed:

H3: Higher levels of resources are positively associated with regulatory capability for EHR-enabled care processes.

H4: Higher levels of resources are positively associated with EHR alignment to care processes.

H5: Higher levels of resources are positively associated with EHR integration.

Regulatory Capability of EHR-Enabled Care Processes, Clinical Staff's Participation, EHR Integration, and Data Completeness in EHR

Researchers (Xiao et al., 2009) indicated that users are needed to follow defined rules and processes for achieving high-quality data when they are using IS. In the context of this study, complete EHR data can be aggregated by navigating staff members with structured rules and procedures of using EHR (Staff et al., 2016). Meanwhile, staff members are stressed in training about the impacts of EHR data completeness that could raise their attention to addressing EHR data completeness in patient care. Since IS integration in an organization can be facilitated by regulation formulation and process management (Xiao et al., 2009), IT professionals are likely to address EHR integration toward EHR data completeness, based on explicit and concrete regulation and process representations of using IS (Kelley et al., 2015). Also, structured regulations and processes designed to deal with using EHR in care processes that drive users

https://aisel.aisnet.org/pajais/vol12/iss2/4
DOI: 10.17705/1pais.12204

to achieve the anticipated outcomes when implementing EHR (e.g. achieving complete data) (McCarthy & Eastman, 2013). We proposed:

H6: Higher levels of regulatory capability for EHR-enabled care processes are positively associated with clinical staff's participation.

H7: Higher levels of regulatory capability for EHR-enabled care processes are positively associated with EHR integration.

H8: Higher levels of regulatory capability for EHR-enabled care processes are positively associated with data completeness in EHR.

EHR Alignment to Care Processes, Clinical Staff's Participation, and EHR Integration

According to Xiao et al. (2009), a good understanding and communication between IT and business play an important role in improving staff members' capability to utilize IS. In the context of this study, healthcare organizations are likely to accept EHR that fit into care processes and clinical staff have better chance to gather and apply high-quality data at the point of care (Herzberg et al., 2011). Meanwhile, communication between clinical staff and IT professionals could provide potential opportunities to (1) deal with mismatch problems between care processes and EHR, and (2) meet clinical staff' requirements for data use during care delivery (Herzberg et al., 2011). When EHR align with care processes, this contributes to well preparing data for different use through integrating required data from multiple data sources (Sherer et al., 2015). We proposed:

H9: Higher levels of alignment of EHR to care processes are positively associated with clinical staff's participation.

H10: Higher levels of alignment of EHR to care processes are positively associated with EHR integration.

Clinical Staff's Participation and Data Completeness in EHR

Kelley et al. (2015) asserted that the lack of knowledge about data input could lead to data incompleteness generated in EHR. If clinical staff do not pay attention to data completeness when they input data, data delays and/or incomplete data could emerge because of human errors (Warsi et al., 2002). Furthermore, if clinical staff are under stress due to the time limitation on recording tasks, more missing items could appear at point of data input; because data is not available at the moment when it requires to be recorded, it could incur incomplete data entered in EHR (Staff et al., 2016; Warsi et al., 2002). We proposed:

H11: Higher levels of clinical staff's participation are positively associated with data completeness in EHR.

EHR Integration, Clinical Staff's Participation, and Data completeness in EHR

Integration technologies assist in determining the extent to which high-quality data is gathered from diverse sources (Xiao et al., 2009). Poor integration of medical systems and repositories could incur data errors, resulting in poor-quality data in EHR (Carvalho et al., 2018). Furthermore, effective data practices from the success of EHR integration help users better understand EHR benefits and further their use of EHR at the point of care (Muthee et al., 2018; Sherer et al., 2015). Thus, staff members are more inclined to take part in the activities for dealing with EHR data completeness. We proposed:

H12: Higher levels of EHR integration are positively associated with data completeness in EHR.

H13: Higher levels of EHR integration are positively associated with clinical staff's participation.

Note that EHR alignment to care processes involves participation from both IT professionals and clinical staff for EHR implementation. Therefore, here we do not hypothesize the influence of "clinical staff's participation" on "EHR alignment to care processes".

Within the conceptual model shown in Figure 2, the solid lines indicate the factors and significant relations among these factors that have been validated in Xiao et al. (2009), while the dotted lines present the proposed new relationships and new factor for the conceptual model.

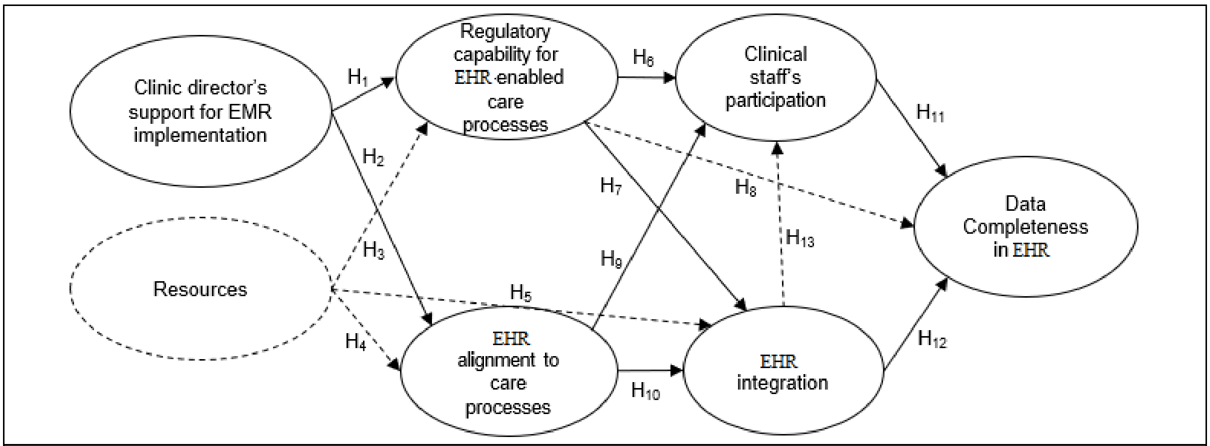


Figure 2 - The Conceptual Model of Factors Influencing Data Completeness in EHR Revised from Liu et al. (2018)

Research Method

Measures of Constructs and Instrument Development

In this study, following the process of instrument development advised by MacKenzie et al. (2011), we developed the measures for all constructs in the conceptual model based on prior studies (see the footnote[2]). As mentioned, Weiskopf et al. 's (2013) four perspectives that have elaborated the most frequently used situations in defining and measuring EHR data completeness are adopted to measure "DC". This allows the present work to better study and understand EHR data completeness. The definitions of these variables for each construct are presented in Appendix A. We designed a survey including two parts: (1) a background questionnaire for participants and their institutions; and (2) a measurement of included constructs using the format of a 5-level Likert item like Xiao et al. (2009)'s study (1 = "Strongly disagree", 2 = "Disagree", 3 = "Neutral", 4 = "Agree", 5 = "Strongly agree"). In this light, we could determine whether or not new hypotheses proposed for the included factors are supported in the present study with a comparison of prior work (Xiao et al., 2009). Our questionnaire items were extracted and revised based on the relevant literature (see the footnote[2]), and then reviewed and modified by experts in the fields of data quality, EHR, survey construction, software development and information systems. Furthermore, the survey instrument was pre-tested by physicians. Our multiple reviews and revisions assist in

[2] The definition and survey sources for each construct are available via the following link: https://www.dropbox.com/s/r9z1izlr3h73vmp/Intrument%20development.docx?dl=0

https://aisel.aisnet.org/pajais/vol12/iss2/4
DOI: 10.17705/1pais.12204

validating instrument development. To decrease marginal errors, we reversely arranged some survey questions as suggested by Wixom and Watson (2001) as well as Xiao et al. (2009). Appendix A presents all survey questions that used to measure the included constructs.

Since the instrument was developed originally in English (all authors are situated at a university in an English-speaking country), it was necessary to translate the questionnaire into Chinese. We followed an adequate translation procedure as summarized by Chen and Boore (2010):

(1) the first author's native language is Chinese, and therefore, the author translated the questionnaire from English to Chinese first and another researcher outside the research group who is bilingual in English and Chinese reviewed the translation;

(2) thereafter we back-translated the questionnaire from Chinese into English;

(3) we repeated steps (1) and (2) in order to achieve accuracy and consistency of the translation.

Research Sample and Data Collection

Online questionnaire can be sent to participants via network that saves expenses for travel and long-distance telephone call. Administration of online questionnaire allows participants to provide complete answer to each question that is not possible in traditional paper-based instruments. Accordingly, we adopted this data collection approach in this study.

Compared with other countries, China has a very large healthcare system (including 32,120 hospitals, 946,490 community health centres, and 19,550 specialized public health institutions (National Health and Family Planning Commission of the People's Republic of China, 2018)), in order to meet the requirements of the growing population on healthcare. This enormous amount of healthcare consumption in China provides massive amount of healthcare data that has a great potential to identify determinants of patient outcomes and improve medical treatment (Li et al., 2017). Furthermore, the Chinese State Council issued its Guiding Opinions on Promoting and Regulating the Development of the Application of Healthcare Big Data, which has attached great importance to gathering and using healthcare data (Liu, 2018). Addressing data quality problems of EHR is thus considered as a priority in Chinese healthcare organizations. We anticipated that we can learn important lessons from clinical practitioners in China about the factors influencing data completeness in EHR.

We implemented the questionnaire in Chinese version using the wjx.cn platform. The final version of the questionnaire was available between March 23, 2018 and June 15, 2018. Our target respondents were practitioners who have established EHR and/or are currently using EHR in their workplaces in China. We believed that the respondents targeted are an appropriate and representative sample to understanding the factors influencing EHR data completeness in the current study.

We firstly sent the link of the questionnaire to the colleagues who have the working/ research experience with Chinese healthcare settings. Then the online questionnaire was distributed by the contact colleagues in their research or working groups via the WeChat app platform. The group members who viewed the invitation letter that includes an introduction of the objectives of our survey and the guide on how to fill in the survey questions and considered that they were the targeted subjects of this study can complete the questionnaire online and forward the link to other colleagues and peers known to them to participate.

Respondent Profile

This study engaged a total number of 150 subjects in China in answering the online questionnaire, and 148 participants completed it. No missing data occurred among these complete responses. Because 3 questionnaires answered from the practitioners whose workplace had not established EHR, finally the 145 responses were remained in data analysis of this study. Table 2 provides our participants' characteristics.

Table 2 - Demographics of Respondents Based on Their Characteristics

Characteristic	Frequency	Percent	Characteristic	Frequency	Percent
Gender			Education Level		
Male	84	57.9	Undergraduate	93	64.1
Female	61	42.1	Postgraduate	68	46.9
Age			Job position		
20 - 39 years	42	29.0	Physician	24	16.6
30 - 39 years	67	46.2	Nurse	14	9.6
>= 40 years	36	24.8	Clinical director/manager	15	10.3
			EHR program manager	41	28.3
Years of experience			Software developer	24	16.6
< 5 years	73	50.3	Others	27	18.6
>= 5 years	72	49.7			
Total	**145**	**100.0**	**Total**	**145**	**100.0**

According to Table 2, 57.9% were male and 42.1% were female; the average age of participants was between 30-39 years; half of participants (49.7%) had been establishing and/or using EHR for more than 5 years. 64.1% participants reported holding a bachelor's degree, and 46.9% indicated gaining a higher qualification with postgraduate study. Regarding job positions of the participants, 16.6% worked as physician, 9.6% as nurse, 10.3% as clinical manager or director, 28.3% as EHR program manager or personal who is familiar with EHR implementation, and 16.6% as software developer. The others are EHR users at healthcare settings such as pharmacists and medical technologists.

Control Variable

In China, provinces, autonomous regions, and municipalities formulate and implement regional health plans and establish medical institutions based on "The National Planning Guideline for the Healthcare Service System (2015–2020)" announced by the centre government (General Office of the State Council of the People's Republic of China, 2015). Thus, healthcare practitioners are regulated under the same healthcare service system. However, the functions of cities could determine the level of medical requirements. It was reported that a higher level of gross domestic product per capita of the city calls for higher requirements for healthcare products and services (Xu, 2018). Regions could have different levels of healthcare development, resulting in dissimilarities in answers for the survey questions. We intended to look at the included factors from clinical practitioners in developed regions in order to learn their experience of addressing data completeness in EHR. Furthermore, as job positions could have an impact on perceived data quality (Tee et al., 2007), they were taken into account for any differences in answers on the factors influencing EHR data completeness in the present study. Hence, we focused on the viewpoints from clinical staff (including physicians, nurses and medical personals who are using EHR during care processes) in developed regions (e.g. Beijing, Shanghai, Jiangsu, Zhejiang, Guangdong, Fujian, and Shandong (National Bureau of Statistics of the People's Republic of China, 2017)) in China. See Table 3.

Table 3 - Demographics of Respondents in China in This Study					
Province	Frequency	Percent	Province	Frequency	Percent
Beijing	*62(29)*	*42.8*	*Jiangsu*	*8*	*5.4*
Fujian	2	*1.4*	Jiangxi	2	1.4
Guangdong	*34(16)*	*23.4*	Liaoning	2	*1.4*
Guangxi	5	3.4	Neimenggu	2	1.4
Hainan	2	1.4	*Shandong*	*3(2)*	*2.1*
Hebei	4	2.8	*Shanghai*	*6(3)*	*4.1*
Henan	1	0.7	Shanxi	3	2.1
Hubei	2	1.4	Xingjiang	1	0.7
Hunan	1	0.7	*Zhejiang*	*5(2)*	*3.4*
Total	**113**	**78**	**Total**	**32**	**22**

Note that the number in the brackets presents the number of clinical staff identified in our sample.

Data Analysis and Results

The unit of data analysis is each questionnaire completed by the respondents. We collected a total of 145 valid responses for data screening and analysis. All the collected questionnaires were de-identified, and the data from the questionnaires was processed by using SPSS for statistical analysis.

In this study we applied Partial Least Squares (PLS) approach to test the conceptual model, since this approach is suitable and useful in the context of formative constructs (like this study) (Sarker et al., 2009; Wixom & Watson, 2001). Firstly, the measurement model was assessed (reliability and validity test); then the structural model was evaluated (hypothesis test), as described below.

Measurement Model

We firstly carried out convergent and discriminant validity to assess the measurement model for all scales. Although our survey instrument was adapted based on multiple instruments and translated into a Chinese version, reviewed and revised by experts and pre-tested by physicians, it is the same as a new instrument and is needed to examine the unidimensionality. According to Gerbing and Anderson (1988), unidimensionality pertains to that all relevant items measure a single underlying trait. Hence, we assessed the unidimensionality for all scales at first. Because unidimensionality cannot be directly measured by PLS, we conducted the scree test based on the dataset by using SPSS, for ascertaining how many latent constructs are contained in the conceptual model (Thakurta et al., 2018). The scree plot shown in Appendix B indicated seven underlying factors that exist in the conceptual model. Thereafter, using smartPLS 3.0, we assessed the factor loading at 300 iterations for each measure in the measurement model. As noted in Hair et al. (2006), the factor loading estimates of measures that are lower than 0.5 should be removed from the measurement model. The factor analysis of this study suggested that all measure variables should be remained in the measurement model. See Table 4.

Table 4 - Results of Factor Analysis

Constructs	Measure Variables	Factor Loading	VIF	Indicator	Mean	Standard Deviation	Factor Loading
DSI	Awareness (DAW)	0.915	4.482	DAW1	4.44	0.639	0.976
				DAW2	4.44	0.669	0.978
	Attitude (DAT)	0.951	6.362	DAT1	4.52	0.671	0.902
				DAT2	4.00	0.816	0.891
	Competency (DCY)	0.940	3.195	DCY1	4.13	0.793	0.959
				DCY2	4.33	0.678	0.945
				DCY3	4.13	0.864	0.944
RSC	Funding (FUN)	0.923	2.963	FUN1	3.77	0.899	1.000
	Human resource (HR)	0.866	2.999	HR1	3.71	0.915	1.000
	Time (TIM)	0.936	2.669	TIM1	3.90	0.934	1.000
RGC	Data quality management (DQM)	0.979	3.284	DQM1	3.90	0.913	0.918
				DQM2	3.92	0.882	0.951
				DQM3	3.75	0.926	0.953
				DQM4	3.75	0.988	0.948
	Process management (PM)	0.928	3.284	PM1	3.69	1.001	0.939
				PM2	3.58	0.997	0.959
				PM3	3.65	0.968	0.970
AGM	Customisation (CUS)	0.848	1.965	CUS1	3.92	0.882	1.000
	Communication (COM)	0.973	1.965	COM1	3.96	0.862	0.931
				COM2	4.04	0.816	0.960
				COM3	3.98	0.874	0.947
SPT	Awareness (SAW)	0.918	1.921	SAW1	4.23	0.731	0.909
				SAW2	4.23	0.731	0.940
				SAW3	4.35	0.683	0.834
	Attitude (SAT)	0.711	1.544	SAT1	4.52	0.610	1.000
	Competency (SCY)	0.800	2.242	SCY1	4.15	0.849	1.000
	Mental status (SMS)	0.745	1.961	SMS1	4.06	0.895	1.000
IGT	Ease of use (EOU)	0.857	1.427	EOU1	3.69	0.919	1.000
	Usefulness (UFN)	0.837	1.528	UFN1	4.10	0.799	0.930
				UFN2	4.12	0.758	0.904
				UFN3	4.25	0.622	0.936
				UFN4	4.21	0.667	0.910
	Compatibility (CMP)	0.591	1.346	CMP1*	2.81	1.049	-0.863
				CMP2	2.96	1.427	0.848
DC	Documentation (DOC)	0.745	1.617	DOC1	3.77	0.921	1.000
	Breadth (BRE)	0.810	1.764	BRE1	3.75	0.837	1.000
	Density (DEN)	0.899	2.976	DEN1	3.79	0.825	1.000
	Predictive (PRE)	0.837	2.414	PRE1	3.87	0.841	1.000

The star superscript means that this measure was reverse coded.

It is worth mentioning here that the conceptual model constructed in the present study contains reflective and formative constructs that should be dealt with discriminatingly (Freeze & Raschke, 2007). Among of the included constructs, reflective constructs include “DSI”, “SPT”, and “DC”. According to Freeze and Raschke (2007), for reflective measures that are caused by the latent construct, Cronbach’s α coefficient can be utilized to assess their internal consistency and reliability. Hence, for reflective constructs we adopted three criteria to examine their convergent validity (Hair et al., 2006): (1) estimates of factor loading should be greater than or equal to 0.5; (2) internal consistency and reliability should be greater than or equal to 0.7; and (3) average variance extracted (AVE) should be greater than or equal to 0.5. The results in Tables 4 and Table 5 indicate that all our reflective constructs were above the suggested level.

In the conceptual model, formative constructs are “RSC”, “AGM”, “RGC”, and “IGT”. For formative measures that are the instances cause the latent construct, it is not appropriate to test internal consistency and reliability for these constructs (Freeze & Raschke, 2007; Hair et al., 2011). Although reflective measures desire multicollinearity, for formative measures the construct can be destabilized by excessive multicollinearity (Jarvis et al., 2003). In order to ascertain there are no significant multicollinearity problems for formative constructs, we tested variance inflation factors (VIF) for these constructs as advised by Hair et al. (2011). The results of VIF were presented within the limited threshold (VIF < 10) (Baabdullah et al. 2019). See Table 4. We also needed to (1) ensure the chosen measures without any conceptual overlap by reviewing relevant literature for the relevance of the measures to their formative construct; and (2) present bivariate correlation (factor loading estimates) between the formative construct and its measures (Cenfetelli & Bassellier, 2009), as shown in Table 4.

As noted in Fornell and Larcker (1981), to address discriminant validity of a measurement model, the AVE square root of each latent variable should be greater than the latent variable's highest squared correlation with any other variable. We also employed this criterion in the current study. Table 5 gives the correlations between all pairs of measures for these constructs computed (Ai et al., 2019), indicating an adequate discriminant validity.

Table 5 - The Measurement Model’s Convergent and Discriminant Validity

Construct	Cronbach's α coefficient	Composite Reliability	AVE	Factor Correlations						
				DSI	RSC	RGC	AGM	SPT	IGT	DC
DSI	0.931	0.956	0.878	1.000						
RSC				0.718	1.000					
RGC				0.657	0.840	1.000				
AGM				0.695	0.793	0.798	1.000			
SPT	0.823	0.883	0.654	0.639	0.639	0.635	0.711	1.000		
IGT				0.608	0.637	0.640	0.716	0.589	1.000	
DC	0.841	0.894	0.680	0.504	0.546	0.664	0.698	0.543	0.762	1.000

Structural Model

The assessment of the structural model was carried out with the results presented in Table 6 after examination of the measurement model. We used bootstrapping with 5, 000 resamples to determine the significance of the paths in the structural model as advised by Hair et al. (2011). The quality of the structural model was assessed based on squared multiple correlations (R^2) as shown in Figure 3.

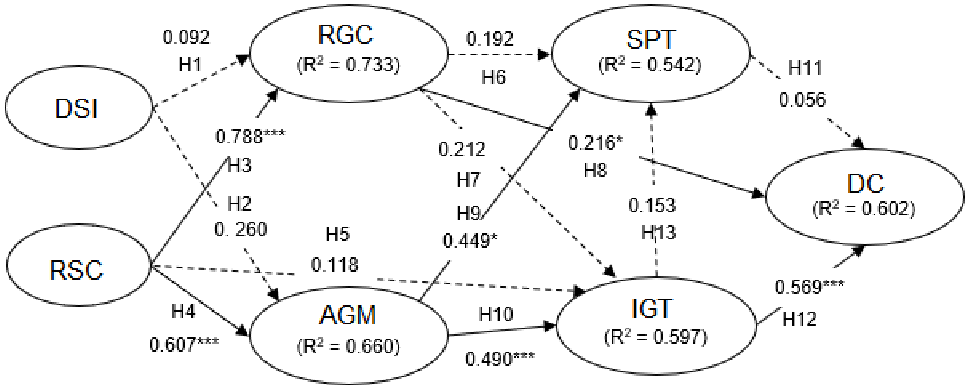
Table 6 - Hypotheses Results in This Study

No.	Hypothesis	Path coefficient (β)	p - value	T-Statistic	Result
H1	DSI -> RGC	0.092	0.455	0.748	Not Support
H2	DSI -> AGM	0.260	0.058	1.894	Not Support
H3	RSC -> RGC	0.788	0.000	8.916	Support

Table 6 - Hypotheses Results in This Study

No.	Hypothesis	Path coefficient (β)	*p* - value	T-Statistic	Result
H4	RSC -> AGM	0.607	0.000	5.134	Support
H5	RSC -> IGT	0.118	0.571	0.566	Not Support
H6	RGC -> DSI	0.192	0.350	0.935	Not Support
H7	RGC -> IGT	0.212	0.282	1.076	Not Support
H8	RGC -> DC	0.216	0.046	1.996	Support
H9	AGM -> SPT	0.449	0.044	2.019	Support
H10	AGM -> IGT	0.490	0.000	3.733	Support
H11	SPT -> DC	0.056	0.732	0.342	Not Support
H12	IGT -> DC	0.569	0.000	4.260	Support
H13	IGT -> SPT	0.153	0.230	1.199	Not Support

Figure 3 indicates that all variances for the endogenous dependent variables were explained, ranging from 0.542 to 0.733 that can be considered as substantial (Thakurta et al., 2018). When “DSI” combined with “RSC” they explained 66.0% of the variance for “AGM” and 73.3% of the variance of “RGC”. “RSC” together with “RGC” and “AGM” explained 59.7% of the variance contained in “IGT”. While the three constructs “AGM”, “RGC”, and “IGT” explained 54.2% of the variance contained in “SPT”. “RGC” along with “IGT” and “SPT” explained 60.2% of the variance for “DC".



“ * ” presents *p*-value < 0.05; “ ** ” presents *p*-value < 0.01; and “ *** ” presents *p*-value < 0.001.

Figure 3 - Structural Model’s Results in This Study

Discussion

This research empirically examined the factors influencing data completeness in the EHR context. We developed the instrument in its Chinese version, and then assessed the measurement model and the structural model based on the perspective of clinical staff from Chinese developed regions. According to perspective of medical staff in Nevada, Liu et al. (2020) showed that all included constructs were the factors influencing data completeness in EHR. However, the results of data analysis of this study reveal that only “resources”, “regulatory capability for EHR-enabled care processes”, “EHR alignment to care processes”, and “EHR integration” were the factors that significantly influence the data completeness in EHR, according to the viewpoints of clinical staff in Chinese developed regions. Additionally, the extent to which these factors influenced data completeness are outlined.

Research Findings

Our empirical findings indicate that both (1) the proposed causal path between "DSI" and "RGC" ($\beta = 0.092$, $p > 0.05$) and (2) the proposed causal path between "DSI" and "AGM", were rejected in this study ($\beta = 0.260$, $p > 0.05$). This may be due to a relatively high level of commitment and support from top management level at healthcare settings in Chinese developed regions. Following the National Planning Guideline on the Healthcare Service System to achieve data sharing and interconnection by implementing EHR has attached much attention from clinic directors. Hence, from the viewpoint of clinical staff, clinic director's support is very basic for EHR implementation that could not serve as a prior factor affecting "RGC" and "AGM". These findings are not similar with (Gopalakrishna- Remani et al., 2018; Kokemueller, 2011; Wixom & Watson, 2001; Xiao et al., 2009).

The results of this study provide strong evidence to support the causal relationships: (1) between "RSC" and "RGC" ($\beta = 0.788$, $p < 0.001$) as well as (2) between "RSC" and "AGM" ($\beta = 0.670$, $p < 0.001$). These results indicate that "RSC" influences "RGC" and "AGM". Resources thus could serve as a factor that influence EHR data completeness, according to the clinical staff's perspective in Chinese developed regions. If healthcare settings have sufficient resources for EHR implementation, (1) regulations formulation and processes management for EHR-enabled care processes and (2) alignment between EHR and care processes could have better chances to be addressed. These findings are consistent with Kokemueller (2011), Shaw (2014), Staff et al. (2016), and Wixom and Watson (2001), in the context of EHR. However, resources failed to contribute to EHR integration with the coefficient value of 0.118 at the significant level greater than 0.05. This finding is not consistent with Nkanata et al. (2018). We noted that the measures of resources adapted from Wixom and Watson (2001) only contained three items (i.e. funding, human resources, and time that are available to support EHR implementation) that concern managerial perspective, while EHR integration seems to rely heavily on technical support. This may open an interesting area for adding measures to assess technical resources and re-examining the role of this construct in addressing EHR data completeness.

Our empirical results do not support: (1) the causal path proposed between "RGC" and "SPT" ($\beta = 0.192$, $p > 0.05$), and (2) the causal path proposed between "RGC" and "IGT" ($\beta = 0.212$, $p > 0.05$). In this study, the respondents were highly educated, and for such respondents using EHR to address data completeness might not be a problem. The integration of EHR (namely, IGT) that concerns more about technical perspective, while regulations formulation and processes management for EHR-enabled care processes (namely, RGC) addresses human and managerial perspective (Liu et al., 2020). Thus, from the clinical staff's viewpoints, the influence of "RGC" on their participation and "IGT" toward DC might not be considered as preferential. However, this study supports the hypothesis for the causal connection between "RGC" and "DC" ($\beta = 0.216$, $p < 0.05$), that is not identified in the study of Xiao et al. (2009). This finding clearly indicates that "RGC" directly affects "DC", which is consistent with McCarthy and Eastman (2013).

Statistical results support that, the proposed relationship between "AGM" and "SPT" ($\beta = 0.449$, $p < 0.05$), and the proposed relationship between "AGM" and "IGT" ($\beta = 0.490$, $p < 0.001$), are supported. These findings point out that "AGM" facilitates "SPT" and meanwhile contributes to "IGT", in the context of this study. The EHR customization and/or a good understanding and communication between clinical staff and IT professionals on EHR technology could help EHR embedded into care processes. In this light, EHR can be better accepted at healthcare settings and aggregate high-quality data from various data sources. These are consistent with the findings reported in the IS literature by multiple researchers ("Herzberg et al., 2011; Sherer et al., 2015; Xiao et al., 2009").

Against expectation, there was no significant relation between "SPT" and "DC" in the present study ($\beta = 0.056$, $p > 0.05$). We found that the endogenous that is explained in the dependent variable "DC" ($R^2 = 0.602$) is relatively weaker than "RGC" and "AGM" as shown in Figure 3.

This implies some hidden factors that have more significant impact on data completeness in EHR not contained in the conceptual model that needs further investigation (e.g. interviews with the clinical staff). This finding is not similar with previous research (Kelley et al., 2015, Staff et al., 2016; Warsi et al., 2002).

As anticipated, the empirical results of the present study report a significant relationship between "IGT" and "DC" (β = 0.569, p < 0.001). Our research demonstrates that higher levels of EHR integration are positively associated with data completeness in EHR. For instance, if EHR systems that are integrated with multiple data sources are of high-quality (e.g. ease of use and/or usefulness that satisfy users' requirements), this will help provide complete data for users. This finding is in an agreement with Carvalho et al. (2018) and Xiao et al. (2009). However, this research rejects the hypothesis of the relationship between "SPT" and "IGT" (β = 0.153, p > 0.05). As shown in Table 4, there is a high level of the participation in the activities of addressing data completeness in EHR, indicating that data quality problems in EHR have received much attention from the clinical staff. While based on their responses, EHR integration was at a moderate level. From the clinical staff's perspective, other underlying factors might play a more important role than the quality of EHR integration in facilitating their participation. This finding is not similar with Muthee et al. (2018) and Sherer et al. (2015).

Implications for Academic

While previous studies have empirically evaluated the factors influencing data quality (e.g. Tee et al. (2007) and Xu (2013)), in this study, we demonstrate that data quality management theoretical constructs can serve as a starting point to understand the factors that influence data completeness in the EHR context. Our study extends the existing work (Liu, Zowghi et al., 2018) by empirically examining the conceptual model and testing the hypotheses for the relationships proposed between the included factors by surveying clinical practitioners in China. The results show that the instrument (in a Chinese version) developed in this study as a reference tool could be used to measure the factors that influence EHR data completeness.

The six constructs of the conceptual model used to interpret the factors influencing EHR data completeness were examined by the empirical data involving clinical staff in Chinese developed regions. Our results show that the "resources" factor that was not contained in the model of Xiao et al. (2009) should be added as one of the factors that has a significant impact on EHR data completeness. Furthermore, Liu et al. (2020) disclosed that all included constructs were the factors influencing data completeness in EHR, according to the viewpoint of medical staff in Nevada, while "clinic director's support for EHR implementation" and "EHR integration" were not significant factors that influence EHR data completeness in our study. We therefore offer some interesting possibilities for exploring the root causes of these differences through more in-depth studies such as interviews.

Additionally, the relationships proposed between the included factors in the conceptual model were tested in this study to link these factors with data completeness. However, some of the hypotheses proposed for these relationships that were previously used in the literature, are found to have no significance in the current study. We have explained possible causes for these differences in Section 6.1. Our results also reveal a new mechanism on addressing EHR data completeness: higher levels of "regulatory capability of EHR-enabled care processes" are positively associated with "data completeness in EHR", which was not significant in the model of Xiao et al. (2009). The extent to which the relations between these factors in this study were different from the those from Liu et al. (2020), resulting in further implications for addressing EHR data completeness in the context of China. Our research thus may provide the foundations to a number of similar studies to theorize the deterministic factors that influence data completeness or other dimensions of data quality (e.g. accuracy and consistency) in multiple applications and explore the underlying relationships between these factors.

https://aisel.aisnet.org/pajais/vol12/iss2/4
DOI: 10.17705/1pais.12204

Implications for Practice

Our study also provide important practical implications. First, our specific finding (i.e. sufficient resources for EHR implementation facilitate EHR alignment to care processes and regulatory capability for EHR-enabled care processes), should be useful to decision makers at healthcare settings. In addition to improving the understanding and knowledge about EHR and awareness of data completeness, this study could draw the attention of decision makers to develop suitable and efficient strategies for resources allocation in EHR implementation, to ensure sufficient funding, time, and human resources that available to support relevant activities. It is worth pointing out that in China one of the unique challenges in healthcare service systems is uneven distribution of resources (Ye, 2018). Resources tend to be concentrated in large and developed cities and a great gap in the distribution of resources exists between urban and rural regions. To this end, smart healthcare is proposed to provide advices on a priority of health problems and help determine which level of healthcare institutions should visit through using EHR data (Xu, 2018). Therefore, governments should appropriately allocate resources and provide guidance to help healthcare settings upgrade their IT infrastructures for EHR implementation.

Second, our finding that regulatory capability for EHR-enabled care processes could be a determinant of data completeness in EHR, is of relevance to EHR program managers. This study may allow EHR program managers to recognize the role of this factor (such as establishing, implementing and improving regulations and procedures on data quality management and process management for EHR-enabled care processes), in achieving complete EHR data. Our empirical results also demonstrate that higher levels of EHR alignment to care processes could better facilitate EHR integration toward data completeness. While customization of EHR configuration at healthcare settings has been emphasized in EHR implementation, we note that a good communication between clinical staff and IT professionals for EHR implementation also facilitates EHR alignment to care processes. Hence, EHR program managers need to enhance communication with clinical staff about the benefits and difficulties of implementing EHR and place them in the role of addressing EHR data completeness.

Third, although the relationship hypothesized between data completeness in EHR and clinical staff's participation was not supported in this study, the clinical staff indicated a high level of participation in addressing EHR data completeness as shown in Table 4. In other words, the clinical staff in Chinese developed regions have realized that their knowledge and skills of using EHR, awareness of EHR data completeness, and focus on data management activities contribute to EHR data completeness. Our study thus suggests that an EHR-related training should be carried out at healthcare settings to assist clinical staff in use of EHR for achieving complete data. Furthermore, EHR vendors are also reminded that the degree to EHR integration can determine how good is the data in EHR. The EHR vendors must provide quality-assured EHR integration from ease of use, usefulness, and compatibility aspects, allowing users to achieve complete data aggregated from multiple sources. Hence, healthcare settings should carry out a rigorous qualification and selection process to select appropriate EHR vendors for best overcoming the technical problems related to EHR integration.

Limitations and Future Work

This study contributes to the field of data quality, while it also has several limitations. First, in the present work, we focused on a single data quality dimension (namely, data completeness), and we used data quality literature as a basis to construct the conceptual model. Future studies may consider the differences between information completeness and data completeness for reconstructing conceptual models of factors influencing data completeness. Furthermore, the weak eigenvalues for the dependent variables (e.g. "CSP" and "DC") suggest that other hidden constructs are likely to influence EHR data completeness that have not been examined and should be further investigated in the conceptual models.

Second, we only empirically evaluated the conceptual model focusing on the viewpoint of clinical staff in the developed regions of China, and the responses were collected from those who volunteered to answer our questionnaires that may have biased the results. Thus, future studies involving more clinical practitioners and regions or countries are encouraged to conduct a cross-country comparison.

Third, this study assumes that the more complete data is aggregated in EHR, the better it is, and thus could result in bias in our results. On the other hand, data completeness is defined for use and requirements in practice (De Feo & Juran, 2017). In other words, necessary data is available for use and requirements that is good enough to present completeness. Hence, future research should account for specific study context and redesign the measures for data completeness.

Lastly, since prior work (Tyree et al., 2006) reported that research use of insurance claims (e.g. billing data) encounters unique issues to address data quality. In this study we focused on missing data in EHR, and we only employed the conceptual model in the context of EHR. Thus, investigating the factors influencing the quality of billing data is not the focus of this study. The conceptual models of factors influencing the quality of billing data could be further proposed, based on relevant and specific literature.

Conclusion

In this work, we evaluated a conceptual model of factors influencing data completeness that extends the model of Xiao et al. (2009) in the EHR context. Different from Liu et al. (2020), we identified the factors influencing EHR data completeness by surveying the clinical practitioners in China, and these factors were: “resources”, “regulatory capability for EHR-enabled care processes”, “EHR alignment to care processes”, and “EHR integration” that help explain the phenomenon of addressing EHR data completeness. Furthermore, we found that some of the proposed hypotheses for relations between the included factors that were supported in prior studies have been found not significant in this study. The study has discussed the possible explanation of these differences. We have provided implications which are related to the factors that influence EHR data completeness for academics, decision makers at healthcare settings, EHR program managers, and EHR vendors. This work also reveals problem areas that require more efforts to deal with data completeness, in order to ensure quality-assured healthcare.

Acknowledgments

We would like to express our gratitude to our colleagues, Professor Lin Liu from Tsinghua University, and Dr Xiaohong Chen from East China Normal University, and Dr Bo Wei from Genowis Inc. for their valuable help in our data collection by distributing the survey in China.

References

Aanestad, M., Grisot, M., Hanseth, O., & Vassilakopoulou, P. (2017). *Information infrastructures within European health care.* Springer International Publishing: Cham, Switzerland.

Ai, S., Du, R., Straub, D.W., Maruping, L.M. & Miao, Y., 2019. Measuring creolization in IT outsourcing: Instrument development and validation. *International Journal of Information Management, 47*, 16-30.

Al-Hiyari, A., AL-Mashre, M. H. H., & Mat, N. K. N. (2013). Factors that affect accounting information system implementation and accounting information quality: A survey in University Utara Malaysia. *American Journal of Economics, 3*(1), 27-31.

American Academy of Family Physicians. (2017). Understanding Features & Functions of an EHR. Retrieved from https://www.aafp.org/practice-management/health-it/product/features-functions.html.

Baabdullah, A.M., Alalwan, A.A., Rana, N.P., Kizgin, H. & Patil, P. (2019). Consumer use of mobile banking (M-Banking) in Saudi Arabia: Towards an integrated model. *International Journal of Information Management, 44*, 38-52.

Bruland, P., Forster, C., Breil, B., Ständer, S., Dugas, M., & Fritz, F. (2014). Does single-source create an added value? Evaluating the impact of introducing x4T into the clinical routine on workflow modifications, data quality and cost–benefit. *International Journal of Medical Informatics, 83*(12), 915-928.

Calisir, F., & Calisir, F. (2004). The relation of interface usability characteristics, perceived usefulness, and perceived ease of use to end-user satisfaction with enterprise resource planning (ERP) systems. *Computers in Human Behavior, 20*(4), 505-515.

Carvalho, J. V., Rocha, Á., Vasconcelos, J., & Abreu, A. (2018). A health data analytics maturity model for hospitals information systems. *International Journal of Information Management, 46,* 278-285.

Cenfetelli, R.T. & Bassellier, G. (2009). Interpretation of formative measurement in information systems research, *MIS Quarterly, 33*(4), 689-707.

Chen, H. Y., & Boore, J. (2010). Translation and back‐translation in qualitative nursing research: Methodological review. *Journal of Clinical Nursing, 19*(1‐2), 234-239.

De Feo, J. A., & Juran, J. M. (2017). *Juran's Quality handbook: The complete guide to performance excellence*. New York: McGraw-Hill.

Estiri, H., Klann, J.G., Weiler, S.R., Alema-Mensah, E., Joseph Applegate, R., Lozinski, G., Patibandla, N., Wei, K., Adams, W.G., Natter, M.D. & Ofili, E.O. (2019). A federated EHR network data completeness tracking system. *Journal of the American Medical Informatics Association*, ocz014, https://doi.org/10.1093/jamia/ ocz014.

Foshay, N., & Kuziemsky, C. (2014). Towards an implementation framework for business intelligence in healthcare. *International Journal of Information Management, 34*(1), 20-27.

Fox, C., Levitin, A., & Redman, T. (1994). The notion of data and its quality dimensions. *Information Processing & Management, 30*(1), 9-19.

Freeze, R. D., & Raschke, R. L. (2007). An assessment of formative and reflective constructs in IS research. *Paper presented at the 15th European Conference on Information Systems*.

General Office of the State Council the People's Republic of China. (2015). *The National Planning Guideline for the Healthcare Service System (2015–2020)*. Retrieved from http://www.gov.cn/zhengce/content/2015-03/30/content_9560.htm

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, 186-192.

Gopalakrishna-Remani, V., Jones, R. P., & Camp, K. M. (2018). Levels of EMR Adoption in US Hospitals: An Empirical Examination of Absorptive Capacity, Institutional Pressures, Top Management Beliefs, and Participation. *Information Systems Frontiers*, 1-20.

Hair, J.F., Ringle, C.M. & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice, 19*(2), 139-152.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis*. (6th ed.). New Jersey: Pearson Prentice Hall.

Haskew, J., Rø, G., Saito, K., Turner, K., Odhiambo, G., Wamae, A., & Sugishita, T. (2015). Implementation of a cloud-based electronic medical record for maternal and child health in rural Kenya. *International Journal of Medical Informatics, 84*(5), 349-354.

Herzberg, S., Rahbar, K., Stegger, L., Schäfers, M., & Dugas, M. (2011). Concept and implementation of a computer-based reminder system to increase completeness in clinical documentation. *International Journal of Medical Informatics, 80*(5), 351-358.

Hoffer, D. N., Finelli, A., Chow, R., Liu, J., Truong, T., Lane, K., & Kurban, G. (2012). Structured electronic operative reporting: comparison with dictation in kidney cancer surgery. *International Journal of Medical Informatics, 81*(3), 182-191.

Hydari, M. Z., Telang, R., & Marella, W. M. (2018). Saving patient ryan—Can advanced electronic medical records make patient care safer? *Management Science*.

ISO 25000 Portal. (2019). *ISO/IEC 25012*. Retrieved from https://iso25000.com/index.php/en/iso-25000-standards/ iso-25012.

Jarvis, C.B., MacKenzie, S.B. & Podsakoff, P.M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research, *Journal of Consumer Research*, 30(2), 199-218.

Johnson, M., Hounkpatin, H., Fraser, S., Culliford, D., Uniacke, M. & Roderick, P. (2017). Using a linked database for epidemiology across the primary and secondary care divide: Acute kidney injury. *BMC Medical Informatics and Decision Making, 17*(1), 106.

Johnson, S. G., Speedie, S., Simon, G., Kumar, V., & Westra, B. L. (2015). *A data quality ontology for the secondary use of EHR data.* Paper presented at the *AMIA Annual Symposium Proceedings*.

Kelley, R. R., Mattingly, W. A., Wiemken, T. L., Khan, M., Coats, D., Curran, D., & Ramirez, J. (2015). Visual grids for managing data completeness in clinical research datasets. *Journal of Biomedical Informatics, 54*, 337-344.

Kokemueller, J. (2011). An empirical investigation of factors influencing data quality improvement success. Paper presented at the *17th of Americas Conference on Information Systems*.

Kumar, M., Gotz, D., Nutley, T. & Smith, J.B. (2018). Research gaps in routine health information system design barriers to data quality and use in low‐and middle‐income countries: A literature review. *The International Journal of Health Planning and Management*, *33*(1), e1-e9.

Kushniruk, A. & Nøhr, C. (2016). "Participatory design, user involvement and health IT evaluation," *Studies in Health Technology and Informatics*, *222*, 139-151.

Li, P., Xie, C., Pollard, T., Johnson, A. E. W., Cao, D., Kang, H., & Fan, Y. (2017). Promoting secondary analysis of electronic medical records in China: summary of the PLAGH-

MIT Critical Data Conference and Health Datathon. *JMIR Medical Informatics, 5*(4), e43.

Li, R., Chen, Y. & Moore, J.H., 2019. Integration of genetic and clinical information to improve imputation of data missing from electronic health records. *Journal of the American Medical Informatics Association*, ocz041, https://doi.org/10.1093/jamia/ocz041.

Lin, Y.K., Lin, M. & Chen, H. (2019). Do electronic health records affect quality of care? Evidence from the HITECH Act. *Information Systems Research*, 30(1), 306-318.

Liu, C., Talaei-Khoei, A., & Zowghi, D. (2018). *Theoretical Support for Enhancing Data Quality: Application in Electronic Medical Records.* Paper presented at the *24th Americas Conference on Information Systems*.

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, 9*(2), 75-100.

Liu, C., Zowghi, D., Talaei-Khoei, A., & Daniel, J. (2018). *Achieving Data Completeness in Electronic Medical Records: A Conceptual Model and Hypotheses Development.* Paper presented at the *51st Hawaii International Conference on System Sciences*.

Liu, C., Zowghi, D., & Talaei-Khoei, A. (2020). An empirical study of the antecedents of data completeness in electronic medical records. *International Journal of Information Management*, 50, 155-170.

Liu, Z. (2018). Health sector gets 'big data' boost. Retrieved from http://www.chinadaily.com.cn/a/201808/14/WS5b72107da310add14f38586c.html

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly, 35*(2), 293-334.

Mashoufi, M., Ayatollahi, H., & Khorasani-Zavareh, D. (2018). A review of data quality assessment in emergency medical services. *The Open Medical Informatics Journal, 12*, 19-32.

Masrom, M. & Rahimly, A. (2015). Overview of data security issues in hospital information systems. *Pacific Asia Journal of the Association for Information Systems*, 7(4), 51-66.

McCarthy, C., & Eastman, D. (2013). *Change management strategies for an effective EMR implementation*, HIMSS: Ohio, USA. Retrieved from:www.himss.org/content/files/Change Management.pdf

Muthee, V., Bochner, A.F., Kang'a, S., Owiso, G., Akhwale, W., Wanyee, S. & Puttkammer, N. (2018). Site readiness assessment preceding the implementation of a HIV care and treatment electronic medical record system in Kenya. *International Journal of Medical Informatics*, 109, 23-29.

National Health and Family Planning Commission of the People's Republic of China. (2018). "The number of medical and health institutions in China at the end of September 2018," Retrieved from

http://www.nhfpc.gov.cn/mohwsbwstjxxzx/s7967/201608/87343c7d63ce41ca8d8fddf7a5db66b7.shtml

National Bureau of Statistics of the People's Republic of China. (2017). Regional data. Retrieved from http://data.stats.gov.cn/easyquery.htm?cn=E0103

Nkanata, M. G., Makori, E. O., & Irura, G. (2018). Comparative analysis of hospital information management systems among healthcare workers in two selected hospitals in Kenya. *Library Philosophy and Practice*, 0_1.

Nord, G. D., Nord, J. H., & Xu, H. (2005). An investigation of the impact of organization size on data quality issues. *Journal of Database Management, 16*(3), 58-71.

Rahimi, A., Liaw, S.-T., Taggart, J., Ray, P., & Yu, H. (2014). Validating an ontology-based algorithm to identify patients with type 2 diabetes mellitus in electronic health records. *International journal of medical informatics, 83*(10), 768-778.

Berner E.S., Kasiraman R.K., Yu F., Ray M.N., & Houston T.K. (2005). Data quality in the outpatient setting: Impact on clinical decision support systems, in *Proceedings of American Medical Informatics Association Annual Symposium*, 41-45.

Sarker, S., Sarker, S., & Jana, D. (2009). Exploring work-life conflict in global software development (GSD) contexts: A survey of IT professionals based in India, in *Proceedings of the 30th International Conference on Information Systems*.

Shaw, N. (2014). The role of the professional association: A grounded theory study of Electronic Medical Records usage in Ontario, Canada. *International Journal of Information Management, 34*(2), 200-209.

Sherer, S. A., Meyerhoefer, C. D., Sheinberg, M., & Levick, D. (2015). Integrating commercial ambulatory electronic health records with hospital systems: An evolutionary process. *International Journal of Medical Informatics, 84*(9), 683-693.

Staff, M., Roberts, C., & March, L. (2016). The completeness of electronic medical record data for patients with Type 2 Diabetes in primary care and its implications for computer modelling of predicted clinical outcomes. *Primary Care Diabetes, 10*(5), 352-359.

Tee, S. W., Bowen, P. L., Doyle, P., Rohde, F. H., & Finance. (2007). Factors influencing organizations to improve data quality in their information systems. *Accounting & Finance, 47*(2), 335-355.

Thakurta, R., Urbach, N., & Basu, A. (2018). Understanding Technology Transition at the Individual Level. *Pacific Asia Journal of the Association for Information Systems, 10(3), 25-60*.

Tilly, R., Posegga, O., Fischbach, K., & Schoder, D. (2017). Towards a conceptualization of data and information quality in social information systems. *Business & Information Systems Engineering, 59*(1), 3-21.

Tyree, P. T., Lind, B. K., & Lafferty, W. E. (2006). Challenges of using medical insurance claims data for utilization analysis. *American Journal of Medical Quality, 21*(4), 269-275.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5-33.

Warsi, A., White, S., & McCulloch, P. (2002). Completeness of data entry in three cancer surgery databases. *European Journal of Surgical Oncology, 28*(8), 850-856.

Weiskopf, N. G., Hripcsak, G., Swaminathan, S., & Weng, C. (2013). Defining and measuring completeness of electronic health records for secondary use. *Journal of Biomedical Informatics, 46*(5), 830-836.

Weiskopf, N. G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research. *Journal of the American Medical Informatics Association*, *20*(1), 144-151.

Wixom, B. H., & Watson, H. J. (2001). An empirical investigation of the factors affecting data warehousing success. *MIS Quarterly*, 17-41.

Wrightson, W. (2010). A comparison of electronic and handwritten anaesthetic records for completeness of information. *Anaesthesia and intensive care, 38*(6), 1052.

https://aisel.aisnet.org/pajais/vol12/iss2/4
DOI: 10.17705/1pais.12204

Xiao, J.-h., Xie, K., & Wan, X.-w. (2009). *Factors influencing enterprise to improve data quality in information systems application - An empirical research on 185 enterprises through field study.* Paper presented at the *16th International Conference on Management Science and Engineering*.

Xu, C. (2018). Health officials: The way to break Chinese medical dilemma. Retrieved from http://news.medlive.cn/all/info-news/show-148896_97.html

Xu, H.-J., Koronius, A., & Brown, N. (2002). Managing data quality in accounting information systems. *IT-Based Management: Challenges and Solutions*, 277-299.

Xu, H.-J. (2013). *Factor analysis of critical success factors for data quality.* Paper presented at the *19th Americas Conference on Information Systems*.

Ye, X. (2018). The "three medical associations" platform was presented at the China family health conference, and science and technology drove the linkage of the three medical associations. Retrieved from http://www.xinhuanet.com/money/2018-12/17/c_1123864400.htm

Zellal, N., & Zaouia, A. (2015). An exploratory investigation of Factors Influencing Data Quality in Data Warehouse. Paper presented at *2015 Third World Conference on Complex Systems*.

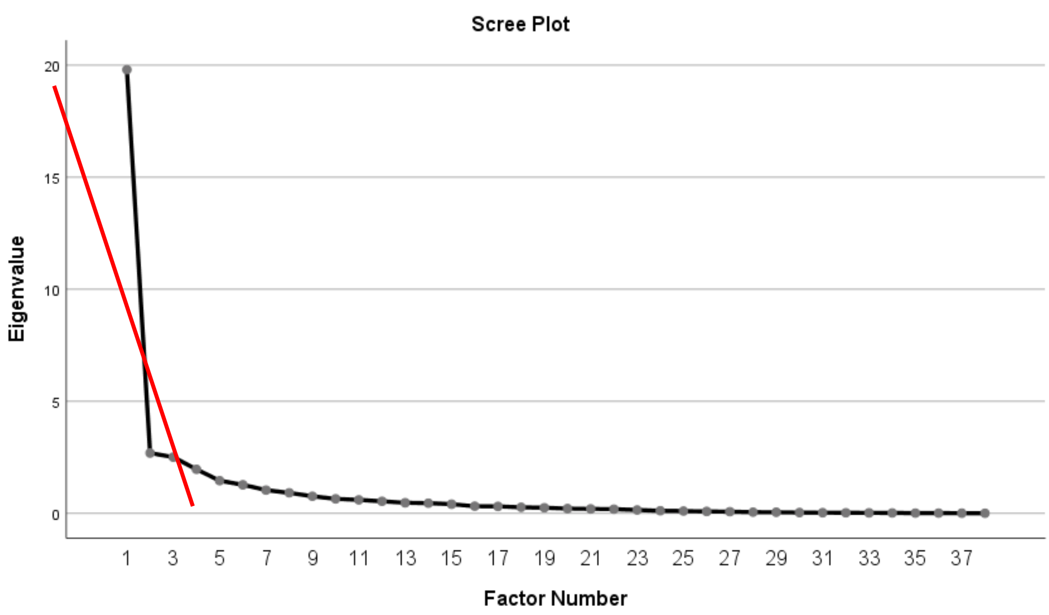
Appendix A. Survey Questions

Appendix A - Survey Questions			
Constructs	Measure Variables	Indicator	Questions in the survey
Clinic director's support for EHR implementation	Awareness	DAW1	Upper management has been aware of the importance of establishing an EHR system.
		DAW2	Upper management has been aware of the importance of achieving complete data in EHR.
	Attitude	DAT1	Upper management has encouraged establishing an EHR system.
		DAT2	Upper management has put achieving complete data in EHR in a priority.
	Competency	DCY1	Upper management understands the capability of an EHR system.
		DCY2	Upper management has the knowledge of EHR benefits.
		DCY3	Upper management has made effective and rapid responses to provide relevant resources to establish an EHR system.
Resources	Funding	FUN1	Establishment of the EHR system was adequately funded.
	Human resource	HR1	Establishment of the EHR system had sufficient human resources to get the work done.
	Time	TIM1	Establishment of the EHR system was given enough time for completion.
Regulatory capability of EHR-enabled care processes	Data quality management	DQM1	Data quality goals for establishing the EHR system have been defined.
		DQM2	Data quality policies for establishing the EHR system have been created.
		DQM3	Data quality standards for establishing the EHR system have been developed.
		DQM4	Data quality controls for establishing the EHR system have been conducted.
	Process management	PM1	Standardized data collection procedures in care processes have been formulated.
		PM2	Adequate training has been provided in the data collection of care processes.
		PM3	A system of measuring the quality of routinely collected data in care processes has been established.
EHR alignment to care processes	Customization	CUS1	The workflow of the EHR system aligns with the routine care processes.
	Communication	COM1	IT professionals are capable of addressing data requirements of clinical staff in the care processes after communication with clinical staff.
		COM2	Clinical staff can gain a better understanding of functions of the EHR system after communication with IT professionals.
		COM3	IT professionals and clinical staff work as a team while implementing the EHR system and addressing incomplete data.
Clinical staff's participation	Awareness	SAW1	I am aware that incomplete data in EHR could result in poor clinical decision making.
		SAW2	I am aware that incomplete data in EHR could increase the risk of harm to the patients.
		SAW3	I aware that incomplete data in EHR could result in loss of revenue for the organization.
	Attitude	SAT1	I am willing to follow instructions related to data management in care processes.
	Competency	SCY1	I have adequate knowledge and skills to record complete data when using the EHR system.
	Mental status	SMS1	I have the ability to concentrate on my tasks to record complete data when using the EHR system.
EHR Integration	Ease of use	EOU1	The EHR system is easy to use.
	Usefulness	UFN1	The EHR system is useful in capturing and tracking patients' demographics.
		UFN2	The EHR system is useful in capturing and tracking patients' medical histories.
		UFN3	The EHR system is useful in creating and tracking clinical documents and notes.

https://aisel.aisnet.org/pajais/vol12/iss2/4
DOI: 10.17705/1pais.12204

Appendix A - Survey Questions			
Constructs	Measure Variables	Indicator	Questions in the survey
		UFN4	The EHR system is useful in presenting and recording patient-specific care plans.
	Compatibility	CMP1	The EHR system frequently reports error messages.
		CMP2	The EHR system has the interface with external systems to capture clinical documents.
Data completeness in EHR	Documentation	DOC1	All observations made during a clinical encounter are recorded in the EHR system.
	Depth	DEP1	All required types of data (e.g. diagnoses and laboratory results) are available in the EHR system.
	Density	DEN1	All the values of patients' records in the EHR system are available as often as required.
	Predictive	PRE1	The data available in the EHR system is sufficient to make a clinical decision for a patient.

Appendix B. Scree Test's Results of This Study



Scree Plot for the Data Sample Collected from This Study

About the Authors

Caihua Liu is a post doctoral researcher and project-appointed associate research fellow at the School of Information Management, Sun Yat-Sen University, China. Caihua received her master and doctoral degrees in Information Systems from Hong Kong Baptist University and University of Technology Sydney, respectively. She was a visiting PhD researcher at the Enterprise of Things (EoT) Lab of University of Koblenz-Landau, Germany, in 2018. Her current research interests include data quality, electronic health records, Internet of Things, intelligent manufacturing, and mobile learning.

Didar Zowghi is a Professor of Software Engineering and Deputy Dean of Graduate Research School at University of Technology Sydney, Australia. Professor Zowghi's research focuses on improving the software development processes and the quality of their products. In particular, she is a world expert in in Requirements Engineering. She is Associate Editor of IEEE Software, Requirements Engineering Journal. She has supervised many PhD students, awarded prestigous resarch grants, and has published over 200 research articles in many conferences and journals. She has collaborated extensively with researchers in several top ranked Chinese Universities in the last decade and has co-authored research articles with over 90 different researchers from 30 countries.

Amir Talaei-Khoei is an Associate Professor of Information Systems at the Ansari College of Business in the University of Nevada, Reno (UNR) and a visiting scholar at the University of Technology Sydney (UTS). Prior to joining UNR, Amir spent almost five years in Australia as a faculty member. His research on healthcare analytics has been funded by international, federal and state agencies. His focus is on predictive analytics for health outcomes as well as abnormality analysis for public health surveillance and pandemic outbreaks. He has been consulted in several occasions for analysis of Ebola and COVID-19. He has received his PhD in Information Systems from the University of New South Wales (UNSW), Australia and holds MSc of Information Technology from Royal Institute of Technology, Sweden. Prior to academia, Dr. Talaei-Khoei worked in software engineering industry in Europe.

Zhi Jin is a professor of Computer Science and the deputy director of Key Lab of High Confidence Software Technologies (MoE) at Peking University. Her research work is primarily concerned with requirements engineering and knowledge-based software engineering. Recently, she pays more attention to the modeling of self-adaptive systems and the learning from both the natural language and the programming language. She has been principle investigator of over 15 national competitive grants including the chief scientist of a national basic research project (973 project) of the Ministry of Science and Technology of China, the project leader of three key projects of National Natural Science Foundation (NSF) of China and the project leader of key international collaborative project of NSF China. She is the co-author of four books and author/co-author over 150 journal and conference publications. Professor Jin is a Senior Member of IEEE and Fellow of China Computing Federation (CCF).