

# Empirical Evaluation of the Influence of EMR Alignment to Care Processes on Data Completeness

Caihua Liu  
University of Technology Sydney  
[Caihua.Liu@uts.edu.au](mailto:Caihua.Liu@uts.edu.au)

Didar Zowghi  
University of Technology Sydney  
[Didar.Zowghi@uts.edu.au](mailto:Didar.Zowghi@uts.edu.au)

Amir Talaei-Khoei  
University of Nevada, Reno  
University of Technology Sydney  
[atalaeikhoei@unr.edu](mailto:atalaeikhoei@unr.edu)

## Abstract

*Data completeness is an important dimension of data quality in electronic medical records (EMR). There are many constructs that influence data completeness in EMR. In this paper, we investigate three of these constructs: Clinical staff participation, EMR integration, and EMR alignment to care processes. We use these constructs from related studies as theoretical support to propose a conceptual model of factors influencing data completeness in EMR. The conceptual model is empirically validated using a survey with clinical staff participants. The results reveal that a high level of clinical staff's participation influences the data completeness in EMR. Furthermore, the alignment of EMR to the care processes has an impact on the data completeness in EMR.*

## 1. Introduction

Recent development of literature proposes that data management capability has an impact on organizational performance. For instance, Mithas et al. [1] defined data management capability as the ability of organizations of utilizing IT in business processes to address data quality that in return could affect their performance. However, how data management capability contributes to data quality has not been explained by Mithas et al. [1]. Liu et al. [39] developed a conceptual model based on Xiao et al.'s model [2] that involves a process of business-IT alignment in data management for addressing data quality, in the context of electronic medical records (EMR). However, the empirical evidence that can account for the influence of this alignment on data completeness as a dimension of data quality in the EMR context was missing in [39]. Although a set of factors influencing data completeness in EMR and the hypotheses of the relationships between these factors were proposed in [39], the relationship between 'EMR integration' and 'clinical staff's participation' toward data completeness

as well as the mediating effects of 'EMR integration' and 'clinical staff's participation' on the relationships between 'EMR alignment to care processes' and 'data completeness in EMR' were not investigated in [39]. By exploring and extending data management literature, this study thus investigates whether EMR alignment to care processes influences data completeness in EMR by EMR integration and clinical staff's participation, addressing the aforementioned missing points in [39], from the perspective of clinical practitioners.

As an example of the undefined process on how data management capability could help achieve high-quality data, this study has been motivated by the case of a system developed in 2015. The system was expected to improve diabetes management for patients across 36 medical clinics by the analysis of daily generated EMR data. An analytics infrastructure, linking the EMR data of the diabetes patients across these clinics, was implemented and a centralized statistical engine carried out a weekly predictive analysis. The physicians then could take advantages of the system that provided a recommendation based on the centralized analysis, to predict which patients were at risk of developing diabetes. However, the case failed to deliver its objectives due to incomplete EMR data being entered at clinics.

Data completeness, as a dimension of data quality, describes the extent to which all required data is available for a given task [3]. Problems associated to completeness can have negative economic and social effects in healthcare. For example, California fined Kaiser Permanente \$2.5 million because this healthcare consortium turned over missing data on patient care to the state's Medicaid program [8]. It was reported that having incomplete data on patients' records has posed the greatest threat to patient care in the United States [5]. Hence, addressing EMR data completeness should be treated as a high priority.

There is a number of empirical studies that highlight the influence of data management on addressing data quality [6, 7]. For instance, Xiao et al. [2] defined data management in this context as a complex phenomenon of business-IT alignment, integration of information systems (IS) and staff participation. They stated that

the influence of business-IT alignment on data quality is mediated by IS integration and staff participation. The development of IT and the demands of organizations have enabled technical architectures to embed into a complete range of business processes in manufacturing cores during the past two decades [35]. Due to the unconvincing return on investment in the short term but large investment in technology, the alignment of EMR to care processes for data management is falling behind in healthcare [37]. In the above-mentioned case of diabetes management by analyzing routinely collected EMR data, the technical infrastructure was available, however, the data management failed to deliver high-quality data in EMR. Both academic and practical gaps described above show that the influence of data management capability on data completeness in EMR needs to be further explored. The current study thus aims to investigate whether EMR alignment to care processes influences data completeness by EMR integration and clinical staff's participation, and to answer the following research question:

**Research Question:** Is data completeness in EMR influenced by EMR alignment to care processes mediated by EMR integration and clinical staff's participation?

In this research, we use Xiao et al.'s model [2] as a starting point to theorize a conceptual model to evaluate the influence of 'EMR alignment to care process' on 'data completeness in EMR'. This influence is then empirically investigated based on the viewpoints elicited from clinical practitioners in northern Nevada.

The present work thus contributes to the literature by developing a conceptual model to evaluate relationships between EMR alignment to care processes and EMR data completeness, assisting in the tasks of explaining how data management capability contributes to data completeness as a dimension of data quality in the EMR context. Furthermore, this study could draw attention from clinical practitioners to data completeness in EMR that is influenced by the alignment of EMR to care processes mediated by clinical staff's participation as well as the combination of EMR integration and clinical staff's participation.

The remainder of this paper is organized as follows: Section 2 reviews the literature and positions the present work in the body of related studies; Section 3 develops the hypotheses; Section 4 gives the method of constructs operationalization, data collection, and data analysis; Section 5 presents the results; Section 6 discusses the findings and Section 7 concludes the paper.

## 2. Literature review

### 2.1. Definition of data completeness

The definition of quality, as noted in De Feo and Juran's quality handbook, is stated both as "fitness for

use" and as "conformance to specification" [9]. This generic understanding of quality could help compare data quality and information quality conceptually [10]. For example, Tilly et al. [10] attempted to differentiate the two from definition of data and information: data is objective to present a phenomenon unrelated to an IS, while information is subjective to put the data into context using an IS that users can understand. Thus, data quality is defined as data's "fitness for use" and can be described by a set of dimensions (e.g. accuracy and completeness) [3]. In this study, we are interested in one dimension of data quality, completeness. The research of data quality can be used as the theoretical support to propose a conceptual model for evaluating the influence of EMR alignment to care processes on data completeness in EMR.

In the EMR context, completeness is divided into four perspectives: "*Documentation: a record contains all observations made about a patient; Breadth: a record contains all desired types of data; Density: a record contains a specified number or frequency of data points over time; and Predictive: a record contains sufficient information to predict a phenomenon of interest*" [4]. We also utilize these four perspectives to better understand and measure data completeness as opposed to data quality in this study (see our measures<sup>1</sup>). Data completeness problems can pertain to missing values in a record and can also refer to missing records for an entity [11]. In this study, we are concerned with missing values in EMR due to data misrepresentation made by individual clinical staff. Although we acknowledge the theoretical differences between data and information, we do not differentiate between data completeness and information completeness here, in order to simplify the presentation and improve the readability of the work.

### 2.2. Theoretical support of factors influencing data completeness as a dimension of data quality

As mentioned, we could take advantages of related studies to construct the conceptual model in our study context. Prior work has conducted a preliminary survey on the models of factors influencing data quality during the period from 2000 to 2017 and summarized three significant characteristics of constructing these models [12] as follows. First, since establishing the relationships between the factors influencing data quality could help explain underlying mechanisms to achieve high-quality data, this should be considered at first when studying the factors that influence data quality. The four models identified in the survey have established the relationships between the included factors [2, 6, 7, 13], thus serving

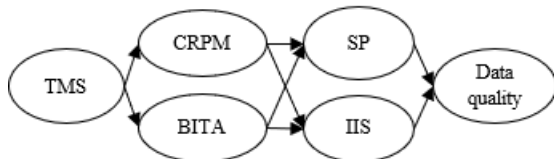
---

<sup>1</sup> <https://www.dropbox.com/s/d8dtqojal8pwoz0/Online%20measures.docx?dl=0>

as a starting point to develop a conceptual model of factors influencing data completeness in EMR. Second, the efforts from both top managers and staff members in an organization contribute to data quality management and improvement (e.g. [2, 13]). Accordingly, both top management support (at a top management level) and staff's participation (at an individual level) should be considered as the factors that influence data quality. Among these four models, Kokemueller [6] and Tee et al. [7] tested the hypotheses of the relationships between the factors, however, staff participation was not included in their models. Last, the most frequently used factors that influence data quality in these models include human, managerial and technical factors.

Based on this analysis, Al-Hiyari et al. [13] and Xiao et al. [2] developed a more inclusive conceptual model in the context of factors influencing data quality. However, Al-Hiyari et al. [13] did not use the data from practice to test their model that may impact the validity of the results.

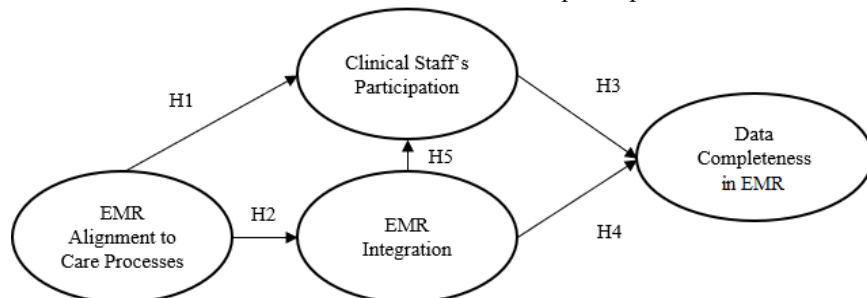
There are five constructs in Xiao et al.'s model [2], including top management support (TMS), capability on the regulation and process management (CRPM), business-IT alignment (BITA), staff participation (SP), and integration of IS (IIS). Figure 1 depicts the constructs together with significant relations between these constructs in Xiao et al.'s model [2]. Because this research aims to investigate whether EMR alignment to care processes influences data completeness in EMR, Xiao et al.'s model [2] best fits us to evaluate this influence from the perspective of clinical practitioners.



**Figure 1. The model of factors influencing data quality management [2]**

### 3. Hypothesis Development

Based on the model of factors influencing data quality management [2], we construct a conceptual model to assess the influence of EMR alignment to care processes on data completeness in EMR. Moreover, the



**Figure 2. The conceptual model being examined in this study**

literature of quality management, EMR, data quality, and data completeness in electronic records has enabled us to hypothesize the relationships between these constructs. Figure 2 presents the proposed conceptual model.

We now make connections between our conceptual model and three significant characteristics of constructing models of factors influencing data quality [12]. First, we hypothesize the relationship between 'clinical staff's participation' and 'data completeness in EMR' as well as the relationship between 'EMR integration' and 'data completeness in EMR', establishing the relations between data completeness in EMR and its factors. Further, we also propose the relationships between the factors influencing data completeness in EMR in the conceptual model, considering (1) the influence of 'EMR alignment to care processes' on 'clinical staff's participation' and 'EMR integration'; and (2) the influence of 'EMR integration' on 'clinical staff's participation'. Second, we discuss the influence of clinical staff's participation in addressing EMR data completeness, including the factor that influences data completeness in EMR at an individual level. Third, our conceptual model includes human, managerial, and technical factors that influence data completeness in EMR. Among these factors, human factor refers to clinical staff; managerial factor addresses EMR alignment to care process, and technical factor concerns EMR integration. Table 1 lists definition of the four constructs proposed in the conceptual model.

#### 3.1 Relationship between EMR alignment to care processes and clinical staff's participation

When IT and business are closely tied, organizations and individuals can better understand the benefits of IT in the business [14]. Understanding the relationships between business and IT improves staff' ability to apply IS [2]. Similarly, EMR customization in care processes along with a good communication about EMR between IT and clinical staff could allow staff members to better understand EMR benefits and apply EMR in care process for addressing data completeness [15, 16]. We hypothesize:

**Hypothesis (H1):** A high level of EMR alignment to care processes is positively associated with clinical staff's participation.

**Table 1. Definition of the constructs extracted from literature**

Construct	Definition	Literature Support
EMR alignment to care processes	EMR are embedded into care processes	Adapted from [2, 15]
Clinical staff's participation	clinical staff understand the importance of addressing data completeness in EMR and make commitment to achieving complete data when using EMR	Adapted from [2]
EMR integration	EMR are integrated across multiple clinical systems and repositories, providing integrated data for different use	Adapted from [2, 28]
Data completeness in EMR	the extent to which required data is available from four perspectives including documentation, breadth, density, and predictive	Adapted from [4]

### 3.2 Relationship between EMR alignment to care processes and EMR integration

Without coordination and communication of the entire organization, IS integration cannot be conducted [2]. In the EMR context, a good communication between IT and clinical staff contributes to addressing mismatch problems between workflow of EMR and care processes, considering possibilities of using EMR data during care delivery [15, 17]. Furthermore, EMR customization in care processes could well aggregate required data from multiple sources for different use [18-20]. We hypothesize:

**Hypothesis (H2):** A high level of EMR alignment to care processes is positively associated with EMR integration.

### 3.3 Relationship between clinical staff's participation and data completeness in EMR

As noted in Kelley et al. [21], if staff members lack knowledge and skills about data entry, they could generate incomplete data during care delivery. Additionally, if staff members are not aware of the importance of data completeness in EMR, data delays or omission errors might occur at point of data entry due to human carelessness [22] and time limitation [22, 23]. We hypothesize:

**Hypothesis (H3):** A high level of clinical staff's participation is positively associated with data completeness in EMR.

### 3.4 Relationship between EMR integration and data completeness in EMR

The integration of data is one of the most common challenges for ensuring data quality in data management [24]. The level of integration technologies implemented and applied in organizations could determine how good is the data aggregated from multiple systems [2], because data errors could occur if clinical systems and repositories are poorly integrated that decrease data quality in EMR [25]. We hypothesize:

**Hypothesis (H4):** A high level of EMR integration is positively associated with data completeness in EMR.

### 3.5 Relationship between EMR integration and

### clinical staff's participation

EMR allow users to achieve the integrated data from different sources simultaneously for decision making [26]. A high level of EMR integration could be better accepted by users [20] and promote further use of EMR in care processes [18]. Hence, clinical staff are likely to participate in the tasks related to addressing data completeness in using EMR. We hypothesize:

**Hypothesis (H5):** A high level of EMR integration is positively associated with clinical staff's participation.

### 3.6 Mediating role of clinical staff's participation in addressing data completeness in EMR

The combination of H1 and H3 proposes a mediating effect of clinical staff's participation on the relationship between EMR alignment to care processes and EMR data completeness. A good communication with clinical staff could help IT professionals better capture users' requirements and customize EMR configuration at clinics in order to provide the corresponding data products for users [17]. The effective data practices then could motivate clinical staff to use EMR in care processes [15, 16], to address data completeness in EMR. We hypothesize:

**Hypothesis (H6):** Clinical staff's participation mediates the influence of EMR alignment to care processes on data completeness in EMR.

### 3.7 Mediating role of EMR integration in addressing data completeness in EMR

The combination of H2 and H4 proposes a mediating effect of EMR integration on the relationship between EMR alignment to care processes and data completeness in EMR. If IT professionals improve their understanding of data needs and correct data errors by communication with users, high-quality data aggregated from different systems could have a better chance to address during IS integration [2]. Furthermore, EMR customization in care processes could help organize and collect required data for clinical decision making [18-20]. We hypothesize:

**Hypothesis (H7):** EMR integration mediates the influence of EMR alignment to care processes on data completeness in EMR.

### 3.8 Mediating Role of EMR integration and

## clinical staff's participation in addressing data completeness in EMR

The combination of H2, H5, and H3 posits a mediating effect of EMR integration and clinical staff's participation on the relationship between EMR alignment to care processes and data completeness in EMR. The procedures of innovative IT (e.g. EMR) integrated into work processes could better consider opportunities for use of EMR data [15]. This high level of EMR integration then could facilitate use of EMR during care delivery [15, 19], and thus, data completeness could be better addressed in EMR. We hypothesize:

**Hypothesis (H8):** The combination of EMR integration and clinical staff's participation mediates the influence of EMR alignment to care processes on data completeness in EMR.

Our conceptual model does not hypothesize the influence of clinical staff's participation on alignment of EMR to care processes, because we consider the communication between clinical and IT staff as an instance to assess EMR alignment to care processes as suggested by Xiao et al. [2], which implies that clinical staff involves in this process.

## 4. Research design

This section outlines our research design of a questionnaire survey to assess the influence of 'EMR alignment to care processes' on 'data completeness in EMR', entailing operationalization and measurement of constructs, data collection, and data analysis.

### 4.1 Operationalization and measurement of constructs

Our conceptual model links 'EMR alignment to care processes' with 'data completeness in EMR' through 'EMR integration' and 'clinical staff's participation'. We operationalized the included four constructs based on their measures in a survey of clinical practitioners and developed the initial survey questions following the process suggested by MacKenzie et al. [29]. The measured variables for each construct were identified and adapted from prior literature as shown in our measures<sup>1</sup> that contains the sources of the measure variables and the survey questions used in this study.

The instrument used in the survey includes two parts. The first part contains a set of questions about the demographic information of participants and their practices. The second part involves the measurement of our constructs based on a 5-point Likert scale ranging from 1 ("Strongly disagree") to 5 ("Strongly agree") as advised by Xiao et al. [2]. To reduce marginal error, some survey questions were reversely arranged [2, 30].

These adapted survey questions were further reviewed by academics outside the research groups with specific expertise in data quality, EMR, information systems, software engineering, and survey construction, and pre-tested by physicians at clinics in a pilot study before formal delivery. These multiple iterations of instrument development assisted in establishing the content validity for the questionnaire.

### 4.2 Research site and data collection

The data analyzed in this study was collected in a medical informatics workshop from March 13 to April 11, 2018 that was held to improve data practices of using EMR at clinics, involving 96 clinical practitioners from various clinical centers across northern Nevada. One of the researchers of this study who works in Nevada and can approach these clinical practitioners advertised and collected the paper-based questionnaires at the workshop. Before the participants answer the questionnaire, an introduction of the purposes of the survey together with the instructions about filling in the questionnaire were given to them. The actual survey questions are available online [38].

### 4.3 Data Analysis

**4.3.1 Sample Size.** The unit of analysis is each response for our survey questionnaire. The data provided was de-identified, and only the data from responses was arranged, imported, coded, and edited using IBM SPSS Statistics V25.0 for data analysis. A total of 96 responses were received. We eliminated 20 invalid responses that had more than one answer to one question and/or missing answers and remained 76 valid questionnaires for further analysis. Table 2 presents the demographic characteristics of the respondents in this survey.

**4.3.2 Control Variable.** The responses were collected from those clinical practitioners who are interested in the topics related to EMR, and therefore, this could result in bias in our sample. Furthermore, because all practitioners were invited to answer our survey questions based in northern Nevada, USA and they are subject to the same regulations under the State health system and the same EPIC EMR system, there are few differences between their experience of using EMR that should be controlled in the analysis. However, some of them were clinical staff (i.e. physicians and nurses), and others were administration staff that may result in different answers for our survey questions. Given the influence of job position on perceptive data quality in [7], job position was considered as a control variable in the analysis, and therefore, we focused on the responses

**Table 2. Demographics of respondents**

Characteristic	Frequency	Percent	Characteristic	Frequency	Percent
Gender			Final degree		
Male	42	55.3	Undergraduate	31	40.8
Female	34	44.7	Postgraduate	45	59.2
Age			Job position		
20 - 39 years	32	42.1	Physician	26	34.2
>= 40 years	44	57.9	Nurse	28	36.8
Years of experience			Administration staff	22	29.0
< 5 years	43	56.6	Total	76	100.0
>= 5 years	33	43.4			
Total	76	100.0			

form clinical staff in this study. That is to say, the sample size of the participants used for the questionnaire assessment is 54.

## 5. Results

We tested the conceptual model using Partial Least Squares (PLS) approach, since this approach is appropriate for the context of small sample size and formative constructs [30, 31] in this study. By using the tool smartPLS 3.0, we firstly conducted the reliability and validity test for the constructs (latent variables) to assess the measurement model, and then tested the structural model as advised by prior work [29, 30], before testing the mediations.

### 5.1 Measurement model

The test of the measurement model includes the examination of (a) internal consistency and reliability as well as (b) convergent and discriminant validity of the constructs. We firstly conducted factors analysis for the constructs included in our conceptual model. According to Hair et al. [32], only variables with a factor loading greater than 0.5 are chosen in the measurement model. Based on the results of factor analysis (these results are available via the link in the footnote<sup>2</sup>), we thus removed three measures with a factor loading less than 0.5 (i.e. ‘SAW1’, ‘UFN1’, and ‘SAT’).

Regarding (a), because reflective and formative measures should be treated differently and for reflective measures they are caused by the latent construct, their internal consistency and reliability can be tested using Cronbach’s  $\alpha$  coefficient [33]. In our conceptual model, reflective constructs are ‘Clinical Staff’s Participation’ and ‘Data Completeness in EMR’. Table 3 shows that the Cronbach’s  $\alpha$  coefficient of the constructs was well above the recommended threshold of 0.6 [32], which indicates an adequate level of internal consistency. However, for formative measures they are representing instances in which the indicators cause the latent construct, and thus internal consistency (reliability testing) is not applicable for formative constructs [33].

In this study, formative constructs entail ‘EMR Alignment to Care Processes’ and ‘EMR Integration’. We thus only examined the relevance of the measures on EMR alignment to care processes and EMR integration to the conceptual model by reviewing relevant literature as advised by Wixom and Watson [30].

Regarding (b), convergent validity is adequate when (1) factor loading for each measured variable upon the relevant latent variable is greater than 0.5 and (2) constructs have an Average Variance Extracted (AVE) of at least 0.5 [32]. Given the results by factor analysis and Table 3, the measurement model was well above the recommended thresholds. For discriminant validity, the square root of AVE of the latent variable should be greater than the correlation coefficient of other latent variables[32]. Table 3 also reports the correlations between all pairs of questions measuring these constructs computed as advised by Ai et al. [34], indicating an adequate discriminant validity between each latent variable.

### 5.2 Structural model

The test of the structural model includes the examination for the path coefficients and  $R^2$  values. Path coefficient explains the strengths of the relationships among the variables, while  $R^2$  value presents the extent to which the variance in the dependent variable can be explained from the independent variable. Both path coefficient (loading and significance) and  $R^2$  value address the quality of the model [30]. The results of the structural model are shown in Figure 3.

The values of the coefficient of determination ( $R^2$ ) extracted from dependent constructs were relatively encouraging, with the range from 0.222 to 0.739. When ‘Clinical Staff’s Participation’ combined with ‘EMR Integration’, they explained 51.8% of the variance of ‘Data Completeness in EMR’. However, the weak  $R$  squared value in ‘EMR Integration’ suggests that other hidden factors influencing this construct not included in the conceptual model should be explored in future studies.

Furthermore, power analysis plays an important role in interpreting the explanatory power of a model [27]. A post hoc power analysis was therefore conducted to determine whether or not our sample size explained the

<sup>2</sup> The results of factors analysis are available at <https://www.dropbox.com/s/iqycl1746huzl6u/Factor%20analysis.docx?dl=0>

**Table 3. Validity results of the constructs**

Construct	Cronbach's $\alpha$ coefficient	Composite Reliability	AVE	EACP	CSP	EIN	DC
EACP				<b>1.000</b>			
CSP	0.728	0.843	0.651	0.723	<b>1.000</b>		
EIN				0.478	0.750	<b>1.000</b>	
DC	0.668	0.801	0.508	0.592	0.707	0.599	<b>1.000</b>

proposed conceptual model with sufficient power, using the Interaction software package (see www.danielsoper.com). The results of power analysis show that our study had sufficient power with 0.99 (above 0.80), presenting adequate power of our sample size for testing the conceptual model.

**The results of this study accept H1:** The alignment of EMR to care processes improved clinical staff's participation (with the path coefficient of 0.377 and the significance is at the  $p < 0.001$  level). See Figure 3. As such, a higher level of EMR alignment to care processes is expected to result in better participation in the tasks related to addressing data completeness in EMR, from the clinical staff's perspective in northern Nevada. When EMR are embedded into care processes, this facilitates clinical staff to further use EMR in care process for addressing data completeness. This finding is consistent with the implications suggested by previous work [15, 16].

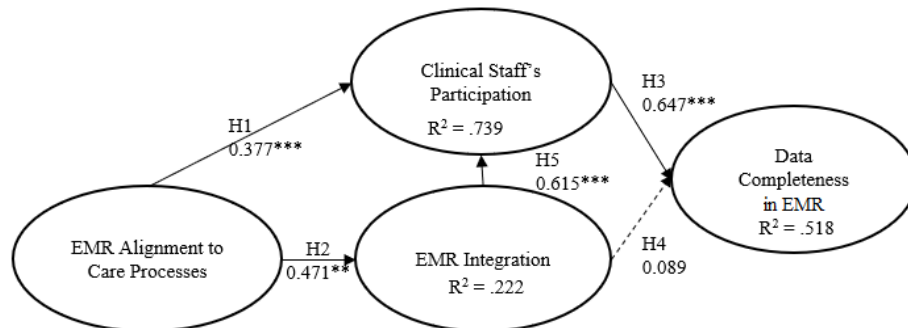
**The results of this study accept H2:** The hypothesis of relationship between alignment of EMR to care processes and EMR integration is supported in this study (with the path coefficient of 0.471 and the significance level is below the level of 0.01). See Figure 3. In this light, a higher level of EMR alignment to care processes is positively associated with EMR integration for addressing data completeness in EMR. For example, if clinics address EMR customization in care processes and/or communication between IT and clinical staff for EMR, EMR could be better integrated with multiple clinical systems and repositories in order to well prepare required high-quality data for users. This finding is similar with the results reported by prior studies [18-20].

**The results of this study accept H3:** Clinical staff's participation contributed to data completeness in EMR (with the path coefficient of 0.647 and the

significance is at the  $p < 0.001$  level). See Figure 3. As we expected, a higher level of clinical staff's participation is positively associated with better data completeness in EMR. If clinical staff are aware of importance of data completeness in EMR, they would like to focus on recording and/or reporting tasks, resulting in effective data practices to preserve data completeness in EMR. Furthermore, clinical staff are willing to improve their understanding and knowledge of using EMR and further their ability to achieve complete EMR data. Our results confirm these findings from prior studies [21, 22], which stated that clinical staff's participation could ensure data completeness in EMR, leading to quality-assured services delivered to patients.

**The results of this study reject H4:** EMR integration failed to influence data completeness in EMR (with the path coefficient of 0.089 and the significance is at the  $p > 0.05$  level). See Figure 3. This outcome may be due to errors frequently reported by EMR systems based on the participants' answers, that may reduce the influence of EMR integration on data completeness in EMR. Prior work indicated that integration technologies (e.g. data warehouses) to some degree could determine how much high-quality data can be aggregated from different sources [2, 6], while poor systems integration could result in data quality problems (e.g. incomplete data)[25]. Further verification of the answers to the survey questions related to EMR integration should be sought from the respondents (e.g. interviews).

**The results of this study accept H5:** Our results support the hypothesis for the causal relationship between clinical staff's participation and EMR integration (with the path coefficient of 0.615 and the significance is at the  $p < 0.001$  level). See Figure 3. As such, a higher level of EMR integration is positively associated with a higher level of clinical staff's participation in addressing data completeness in EMR. When EMR address ease



Note that “\*” presents  $p < 0.05$ , “\*\*” presents  $p < 0.01$ , and “\*\*\*” presents  $p < 0.001$ .

**Figure 3. Results for the structural model**

of use, usefulness, and compatibility with multiple systems, clinical staff who have experienced effective practices in using EMR would rely more on EMR to address data completeness in care processes. This finding is similar with previous studies [18, 20].

### 5.3 Hypotheses for the mediations

H6, H7, and H8 suggest that the indirect effect of EMR alignment to care processes on data completeness in EMR is mediated by clinical staff's participation and EMR integration. We examined this indirect effect by bootstrapping. Table 4 presents the results for testing the mediations between EMR alignment to care processes and EMR data completeness.

It can be observed in Table 4 that the relationship between EMR alignment to care processes and data completeness in EMR was mediated by (1) clinical staff's participation as well as (2) the combination of EMR integration and clinical staff's participation, but was not mediated by EMR integration.

**Table 5. Mediation results**

No	Mediation path	t-Stat	p value	Hypothesis result
H6	EACP → CSP → DC	3.396	0.001	Supported
H7	EACP → EIN → DC	0.535	0.593	Not Supported
H8	EACP → EIN → CSP → DC	2.605	0.009	Supported

## 6. Discussion

Our study is motivated by a case of analytics infrastructure that failed to deliver its objectives across a set of linked clinics in Australia due to incomplete data entered in EMR. We examined the influence of 'EMR alignment to care processes' on 'data completeness in EMR', by using a conceptual model drawn on the literature of quality management, EMR, data quality, and data completeness in electronic records. Two constructs (i.e. clinical staff's participation and EMR integration) were used to understand the relationships between EMR alignment to care processes and data completeness in EMR. Examination of the conceptual model assisted in answering our research question:

**Research Question:** Is data completeness in EMR influenced by their alignment to care processes mediated by EMR integration and clinical staff's participation?

**Findings:** Yes, EMR alignment to care processes does influence data completeness in EMR. Specifically, data completeness in EMR was influenced by EMR alignment to care processes mediated by clinical staff's participation as well as the combination of EMR integration and clinical staff's participation, while the mediation of EMR integration between EMR alignment to care processes and data completeness in EMR was not supported by this study.

Although Liu et al. [39] have proposed the factors that influence data completeness in EMR and developed the hypotheses of the relationships between these factors, our study differs from [39] in at least three ways. First, the relationship between 'EMR integration' and 'clinical staff's participation' proposed in the present work was not included in [39]. Second, the mediating effects of 'EMR integration' and 'clinical staff's participation' on the relationships between 'EMR alignment to care processes' and 'data completeness in EMR' investigated in our work were not studied in [39]. Last, our conceptual model empirically evaluated can add substantial evidence to support our research findings and implications on addressing data completeness in EMR while [39] did not validate their conceptual model.

### 6.1 Academic implications

Our research contributes to the literature in at least three ways. First, we establish a conceptual model to evaluate the influence of EMR alignment to care process on data completeness in EMR, addressing how data management capability influences data quality that has not been explained by Mithas et al. [1], within our study context. This conceptual model links data management capability with data completeness as a dimension of data quality, by explaining data management as a complex phenomenon of EMR alignment to care processes, clinical staff's participation, and EMR integration. Although this study is limited to the context of EMR, the proposed conceptual model can be reexamined or extended to better understand the factors influencing data completeness in multiple applications.

Second, we show how communication between IT and clinical staff for EMR contributes to alignment of EMR to care processes. Previous studies highlighted the role of EMR customization in alignment of EMR to care processes [18, 19]. Our research further reveals that EMR alignment to care processes involves a good communication between IT and clinical staff for EMR. Hence, alignment of EMR to care processes could be achieved by customization of EMR configuration at clinics and/or by communication between IT and clinical staff for EMR.

Last, we draw attention to the relationships between data completeness in EMR and alignment of EMR to care process. Specifically, we add to the literature that data completeness in EMR is significantly influenced by EMR alignment to care processes mediated by clinical staff's participation and the combination of EMR integration and clinical staff's participation from the clinical staff's perspective, highlighting the effect of clinical staff's participation in addressing EMR data completeness. This finding recasts clinical staff in a significant enabler to achieve complete data when using EMR, which offers interesting possibilities for extending



the measures of clinical staff's participation or exploring other potential factors that influence clinical staff's participation in addressing EMR data completeness.

## 6.2 Practical implications

Our study informs that workflow of EMR embedded into care processes facilitates clinical staff's participation in addressing data completeness in EMR and contributes to integrating complete EMR data from different sources. We further demonstrate that a good communication between IT and clinic staff for EMR could be an alternative way to address the EMR alignment to care processes that have not received much attention as our participants reported a relatively low level of communication with IT professionals about EMR. These findings could allow EMR program managers to realize the role of clinical staff in EMR alignment to care processes toward data completeness in EMR. If clinical staff could have a good understanding about technical features and business value of EMR and provide their feedback for EMR, data completeness in EMR could have a better chance to be addressed. Thus, EMR program managers should hold EMR-related activities (e.g. seminars) to engage clinical staff in data management.

Furthermore, the findings of this study are similar with Kelley et al. [21] and Warsi et al. [22], which stated that clinical staff's participation influences EMR data completeness. In addition to a high level of awareness and competency required to address data quality as advised by Xiao et al. [2], this study reveals that individual mental status has an impact on achieving complete data when using EMR. If clinical staff have a fresh mind and concentrate on their tasks related data management in care processes, they are likely to have better EMR data completeness. Therefore, an effective training for improving the knowledge and skills of addressing EMR data completeness and the awareness of the importance of data completeness in care processes, contributes to addressing data completeness in EMR.

We also reminder EMR vendors that the outcomes of EMR integration could determine (1) the extent to which all required data is aggregated from different systems and repositories; and (2) the extent of which clinical staff members participate in the EMR-related initiatives. Although this study did not support the hypothesis regarding the influence of EMR integration on data completeness in EMR, previous studies showed that the success of systems integration help achieve high-quality data [2, 6]. We noted that from the perspective of our participants, the EMR system frequently reported error messages. This may have an impact on our results on the correlation between EMR integration and data completeness in EMR. Therefore, EMR vendors must address the issues of EMR integration with a high level

of ease of use, usefulness, and compatibility, to provide high-quality data for users.

For the motivational case of this study, if EMR are embedded into care processes and clinical staff are placed in the role of addressing data completeness in EMR, clinics will have more complete data, based on the lessons learned from northern Nevada. Furthermore, our empirical findings may allow practitioners at clinics to identify which of these three areas should receive more efforts to address data completeness in EMR.

## 6.3 Limitations and future work

Although this study makes significant contributions, the current research has some limitations. First, this work is only interested in data completeness as a dimension of data quality, and our choice of constructs was restricted by our adherence to data quality studies as our theoretical lens. Hence, researchers are suggested to take the differences between data completeness and information completeness into account and redesign the present work in their study context. Second, given the importance of EMR alignment to care processes for addressing EMR data completeness in this study, we suggest researchers to add more variables and survey questions for measuring this construct. Third, the dataset used in this work is limited from one state in USA (i.e. Nevada) that could have an impact on the generalization of our findings for any other country or even other states in USA, being subject to a locality bias [40]. To overcome a locality bias, future study is encouraged to involve more clinical staff and across EMR systems and geographies, to better understand these factors. Last, the present work only includes three constructs to address the factors influencing data completeness in EMR. Another topic of interest that we are currently pursuing is to include more factors that influence EMR data completeness in the conceptual model (see [39]).

## 7. Conclusion

In this study, by empirically examining a conceptual model drawn on [2], we reveal that the influence of EMR data completeness is influenced by EMR alignment to care processes mediated by clinical staff's participation as well as the combination of EMR integration and clinical staff's participation. We also provide practical implications on enhancing data management capability from these three areas to address EMR data completeness.

## 8. References

- [1] S. Mithas, N. Ramasubbu, and V. Sambamurthy, "How information management capability influences firm performance,"

- MIS Quarterly*, vol. 35, no. 1, pp. 237-256, 2011.
- [2] J.-h. Xiao, K. Xie, and X.-w. Wan, "Factors influencing enterprise to improve data quality in information systems application—An empirical research on 185 enterprises through field study," in *ICMSE Proc*, 2009, pp. 23-33.
- [3] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *Journal of Management Information Systems*, vol. 12, no. 4, pp. 5-33, 1996.
- [4] N. G. Weiskopf, G. Hripcsak, S. Swaminathan, and C. Weng, "Defining and measuring completeness of electronic health records for secondary use," *Journal of Biomedical Informatics*, vol. 46, no. 5, pp. 830-836, 2013.
- [5] R. Philips. (2017, Nov 16). Patient safety in critical condition - Top threats point back to incomplete data. Retrieved May 30, 2019, from: [https://www.usa.philips.com/cdam/b2bhc/master/featuredetail/pm-deepdive/pm-group-page/PCMS\\_survey\\_project\\_report\\_UPDATED.pdf](https://www.usa.philips.com/cdam/b2bhc/master/featuredetail/pm-deepdive/pm-group-page/PCMS_survey_project_report_UPDATED.pdf).
- [6] J. Kokemueller, "An empirical investigation of factors influencing data quality improvement success," in *AMCIS Proc*, 2011, pp. 154-163.
- [7] S. W. Tee, P. L. Bowen, P. Doyle, F. H. Rohde, and Finance, "Factors influencing organizations to improve data quality in their information systems," *Accounting & Finance*, vol. 47, no. 2, pp. 335-355, 2007.
- [8] C. Terhune. (2017, January 26). California Fines Kaiser Permanente \$2.5 Million Over Missing Medicaid Data. Kaiser Health News. Retrieved May 16, 2019, from: <http://khn.org/news/california-fines-kaiser-permanente-2-5-million-over-missing-medicaid-data>
- [9] J. A. De Feo and J. M. Juran, *Juran's Quality handbook: The complete guide to performance excellence*. New York: McGraw-Hill, 2017.
- [10] R. Tilly, O. Posegga, K. Fischbach, and D. Schoder, "Towards a conceptualization of data and information quality in social information systems," *Business & Information Systems Engineering*, vol. 59, no. 1, pp. 3-21, 2017.
- [11] Y.-N. Liu, J.-Z. Li, and Z.-N. Zou, "Determining the real data completeness of a relational dataset," *Journal of Computer Science Technology*, vol. 31, no. 4, pp. 720-740, 2016.
- [12] C. Liu, A. Talaei-Khoei, and D. Zowghi, "Theoretical support for enhancing data quality: Application in electronic medical records," in *AMCIS Proc*, 2018, pp. 1-10.
- [13] A. Al-Hiyari, M. H. H. AL-Mashre, and N. K. N. Mat, "Factors that affect accounting information system implementation and accounting information quality: A survey in University Utara Malaysia," *American Journal of Economics*, vol. 3, no. 1, pp. 27-31, 2013.
- [14] J. E. Gerow, J. B. Thatcher, and V. Grover, "Six types of IT-business strategic alignment: an investigation of the constructs and their measurement," *European Journal of Information Systems*, vol. 24, no. 5, pp. 465-491, 2015.
- [15] S. Herzberg, K. Rahbar, L. Stegger, M. Schäfers, and M. Dugas, "Concept and implementation of a computer-based reminder system to increase completeness in clinical documentation," *International Journal of Medical Informatics*, vol. 80, no. 5, pp. 351-358, 2011.
- [16] P. Bruland, C. Forster, B. Breil, S. Ständer, M. Dugas, and F. Fritz, "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*, vol. 83, no. 12, pp. 915-928, 2014.
- [17] J. F. Cohen, E. Coleman, and M. J. Kangethe, "An importance-performance analysis of hospital information system attributes: A nurses' perspective," *International Journal of Medical Informatics*, vol. 86, pp. 82-90, 2016.
- [18] S. A. Sherer, C. D. Meyerhoefer, M. Sheinberg, and D. Levick, "Integrating commercial ambulatory electronic health records with hospital systems: An evolutionary process," *International Journal of Medical Informatics*, vol. 84, no. 9, pp. 683-693, 2015.
- [19] M. C. Moon, R. Hills, and G. Demiris, "Understanding optimization processes of electronic health records (EHR) in select leading hospitals: A qualitative study," *Journal of Innovation in Health Informatics*, vol. 25, no. 2, pp. 109-125, 2018.
- [20] O. Nov and W. Schechter, "Dispositional resistance to change and hospital physicians' use of electronic medical records: A multidimensional perspective," *Journal of the American Society for Information Science and Technology*, vol. 63, no. 4, pp. 648-656, 2012.
- [21] R. R. Kelley et al., "Visual grids for managing data completeness in clinical research datasets," *Journal of Biomedical Informatics*, vol. 54, pp. 337-344, 2015.
- [22] A. Warsi, S. White, and P. McCulloch, "Completeness of data entry in three cancer surgery databases," *European Journal of Surgical Oncology*, vol. 28, no. 8, pp. 850-856, 2002.
- [23] M. Staff, C. Roberts, and L. March, "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*, vol. 10, no. 5, pp. 352-359, 2016.
- [24] J. Lismont, J. Vanthienen, B. Baesens, and W. Lemahieu, "Defining analytics maturity indicators: A survey approach," *International Journal of Information Management*, vol. 37, no. 3, pp. 114-124, 2017.
- [25] J. V. Carvalho, A. Rocha, J. Vasconcelos, and A. Abreu, "A health data analytics maturity model for hospitals information systems," *International Journal of Information Management*, vol. 46, pp. 278-285, 2019.
- [26] M. Barcellos Almeida, F. Farinelli, and Technology, "Ontologies for the representation of electronic medical records: The obstetric and neonatal ontology," *Journal of the Association for Information Science and Technology*, vol. 68, no. 11, pp. 2529-2542, 2017.
- [27] J. F. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, 1988. NJ: Lawrence Earlbaum Associates.
- [28] M. Aanestad, M. Grisot, O. Hanseth, and P. Vassilakopoulou, *Information infrastructures within European health care*. Cham: Springer International Publishing, 2017.
- [29] S. B. MacKenzie, P. M. Podsakoff, and N. P. Podsakoff, "Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques," *MIS Quarterly*, vol. 35, no. 2, pp. 293-334, 2011.
- [30] B. H. Wixom and H. J. Watson, "An empirical investigation of the factors affecting data warehousing success," *MIS Quarterly*, pp. 17-41, 2001.
- [31] J. F. Hair, G. T. M. Hult, C. M. Ringle, and M. Sarstedt, *A primer on partial least squares structural equation modeling*, 2nd ed. Thousand Oaks: Sage Publications, 2017.
- [32] J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, and R. L. Tatham, *Multivariate data analysis*. Upper Saddle River NJ: Pearson Prentice Hall, 2006.
- [33] R. D. Freeze and R. L. Raschke, "An assessment of formative and reflective constructs in IS research," in *ECIS*, 2007, pp. 1481-1492.
- [34] S. Ai, R. Du, D. W. Straub, L. M. Maruping, and Y. Miao, "Measuring creolization in IT outsourcing: Instrument development and validation," *International Journal of Information Management*, vol. 47, pp. 16-30, 2019.
- [35] T. Huang, and A. Yokota, "Inventing a business-ERP alignment assessment model through three Japanese companies," *Business Process Management Journal*.
- [36] C. Liu, A. Talaei-Khoei, D. Zowghi, and J. Daniel, "Data completeness in healthcare: A literature survey," *Pacific Asia Journal of the Association for Information Systems*, vol. 9, no. 2, pp. 75-100, 2017.
- [37] M.G. Nkanata, E.O. Makori, and G. Irura, "Comparative analysis of hospital information management systems among healthcare workers in two selected hospitals in Kenya," *Library Philosophy and Practice*, vol. 2056, pp. 1-19.
- [38] C. Liu, D. Zowghi, and A. Talaei-Khoei. (2018, Mar 11). Information sheet and consent form: Factors influencing data completeness in electronic medical records. Retrieved Aug 23, 2019, from: [https://www.dropbox.com/s/c5kx1m96yqf5im5/Survey%20questions\\_Formal%20delivery.docx?dl=0](https://www.dropbox.com/s/c5kx1m96yqf5im5/Survey%20questions_Formal%20delivery.docx?dl=0)
- [39] C. Liu, D. Zowghi, A. Talaei-Khoei, and J. Daniel, "Achieving data completeness in electronic medical records: A conceptual model and hypotheses development," in *HICSS Proc*, 2018, pp. 2824-2833.
- [40] V. Hamuryudan, I. Fresko, H. Direskeneli, M.J. Tenant, S. Yurdakul, T. Akoglu, and H. Yazıcı, "Evaluation of the Turkish translation of a disease activity form for Behcet's syndrome," *Rheumatology*, vol. 38, no. 8, pp.734-736, 1999.