

Impact of Data Quality of Electronic Health Records on Job Characteristics of Physicians

¹Navneet Kaur Bajwa

²Harjot Singh

³Kalyan Kumar De

Abstract

Electronic health records (EHR) have become an integral part of most health care systems. It is promoted as a panacea to many problems in healthcare information management. This study examines the factors affecting data quality of EHR and their impact on job characteristics of physicians with respect to tertiary level healthcare services. Data were collected from 280 physicians of 12 multispecialty hospitals located in seven districts and the state capital in North India (Punjab) with the help of a structured questionnaire. Factor analysis and multiple regression techniques have been employed to analyze data. The study has revealed that factors affecting data quality of EHR are 'effectiveness', 'accessibility', 'accuracy' and 'timeliness'. It has been concluded that these data quality factors explain a positive variance in job characteristics of physicians using EHR technology.

Keywords: Data quality, EHR, India, job characteristics, multispecialty hospitals, physicians

1. Introduction

EHR, Electronic medical records (EMR) and paperless computer-based patient records have been variously defined. However, a consensus understanding is that an EHR is distinguished by its scope (Jeselon & Schoeffel, 2002). An EHR is a longitudinal record of a patient's health and healthcare – from cradle to grave (Department of Health, 2002). Other terms describe electronic records of periodic care, usually from one institution. Thus, an EHR is a permanent document which holds electronic information about a patient's lifelong physical, mental and social state. It also records diseases and other abnormal conditions which are detailed by healthcare professionals themselves, in primary, secondary and tertiary care (Boaden & Munir, 2001).

EHR data quality is often considered only within the narrow scope of data verification and validation. It, however, also concerns the equally critical aspect of assuring that EHR data are appropriate for use. Maintaining the quality of healthcare data is widely acknowledged critical and challenging to effective healthcare.

Various factors that can impede data quality are identified within US military health systems. These include inadequate management structures for ensuring complete, timely and accurate reporting of data; inadequate rules, training, and procedural guidelines for those involved in data

¹ Research Scholars, L. M. Thapar School of Management, Thapar University, India.

² Assistant Professor, L. M. Thapar School of Management, Thapar University, India.

³ Professor, L. M. Thapar School of Management, Thapar University, India.

collection; fragmentation and inconsistencies among services associated with data collection; and requirement for new management methods which utilize accurate and relevant data to support the managed care environment (Military Health Services, 2003). Thus, management and organizational factors, as well as effective technological mechanisms, affect the ability to maintain data quality.

Indian Healthcare Sector

India has a population of 1.3 billion residing in 29 states and seven union territories, out of which 2.25% are situated in Punjab, which consists of twenty-two districts and one union territory. Healthcare services in India have traditionally been provided by the government. In the early post independence period, Indian healthcare sector was suffering from a shortage of doctors and nurses, inadequacy of hospitals as well as a lack of medical equipments. In the 1980s, approximate 30% decline was witnessed in the use of both urban and rural public healthcare facilities (Kamra, Singh & De, 2015). Thereafter, this sector has achieved tremendous growth and has been growing faster than it was during the last few decades. It is expected to become US\$ 280 billion sector by the end of 2020 (Verma & Khandelwal, 2011).

2.Review of Literature

Data quality requirements can be found in many strategic documents of specific healthcare organizations. However, putting together such a list is instructive, in that it reveals the difficulty of deriving an exhaustive list as well as pitfalls of treating EHR data as structured data in a conventional manner.

Two important quality indicators i.e. completeness of medical record data and good communication between physicians, patients are identified (Mandl & Porter, 1999). However, metrics for these and other indicators have yet to be established. Thus, there is a need for an integrated and cohesive methodology which addresses both data entry and use. This methodology includes detailed and comprehensive requirement analysis of all data quality characteristics, with a focus on analysis of assurance measurements at the points of data entry, manipulation and retrieval. However, it must also address the important issue of ensuring that data are appropriate and in an appropriate form for the uses to which they are put.

There is evidence that it is often problematic to gain acceptance, by physicians, nurses and other healthcare professionals, with respect to EHR systems (Crownower & Rosenbaum, 2002). A primary reason cited is a general lack of trust in EHR systems with respect to critical issue of data quality, since possible errors may cause serious problems to patient health. Lieberman (2002) observes that, recently, intelligent interfaces have become increasingly necessary to cope with requirements for integration and interoperability between complex, inflexible and incompatible systems.

Since these problems also characterize EHR application, intelligent interfaces potentially have a role in allowing EHR systems to adapt to requirements of individual users and uses, and in

ensuring that data are collected and used in ways that are consistent with the task in hand, while applying necessary quality assurance procedures. It can help to gather data quickly, review historical data faster, and take decisions sooner with fewer possibilities for mistakes. Similarly, Scott (2001) believes that widespread adoption of intelligent interfaces will lead to significantly higher productivity with less stress for the computer users and fewer workplace injuries. Moreover agents can improve systems usability by performing tasks on behalf of the users, such as searching and filtering information from them and providing online assistance. Gil and Kim (2000) note that this technique aids ensuring completeness of data, by facilitating intelligent acquisition systems to help users understand how the new knowledge that their data provide relates to knowledge previously stored, and hence what additional knowledge is still missing and needs to be provided by the user.

3.Research Methodology

Objectives of the study are:

- 1) To identify the factors affecting data quality using EHR with respect to multispecialty hospitals.
- 2) To study the impact of data quality on job characteristics of physicians with respect to identified factors.

Job characteristics are a theory of work design providing a set of implementing principles for enriching jobs in organizational settings (Hackman & Oldham, 2005). The job characteristics theory proposes a model of five “core” job characteristics, namely, skill variety, task identity, task significance, autonomy, and feedback. Job characteristic inventory model has been used as a theoretical lens to understand the impact of data quality on physicians’ jobs.

In the present study, respondents have described their job characteristics in response to 25 statements on a five-point Likert scale ranging from 1 (strongly dissatisfied) to 5 (strongly satisfied). The study is exploratory-cum-descriptive in nature, as it endeavors to assess the relationship between data quality and job characteristics of physicians. Physicians in multispecialty hospitals, based in Punjab, comprise the population for the study. Table 1 shows the demographic profile of physicians.

Table 1: Summary of Demographic Profile of Physicians

Variable	Category	Number of Respondents (% age)
Age	20-30	66 (23.6)
	30-40	47 (16.8)
	40-50	36 (12.9)
	50-60	66 (23.6)

	60 or above	65	(23.2)
Gender	Male	161	(57.5)
	Female	119	(42.5)
Marital status	Married	203	(72.5)
	Unmarried	77	(27.5)
Education Qualification	MBBS	96	(34.3)
	MBBS-MD	91	(32.5)
	DM/MCH	93	(33.2)
Work area	Emergency Ward	5	(1.8)
	Intensive Care Unit	45	(16.1)
	Operation Theatre Personnel	56	(20.0)
	General Ward	77	(27.5)
	Out Patient Department	97	(34.6)
Designation	Professor	7	(2.5)
	Associate Professor	14	(5.0)
	Assistant Professor	3	(1.1)
	Senior Resident	143	(51.1)
	Junior Resident	113	(40.4)
Work experience in hospitals	0 - 2 years	105	(37.5)
	2 - 5 years	81	(28.9)
	5 - 10 years	77	(27.5)
	More than 10 years	17	(6.1)
Computer knowledge	Novice/Beginner	97	(34.6)
	Average	89	(31.8)
	Expert	94	(33.6)
Working hrs per week	Up to 40 hrs	7	(2.5)
	40 - 50 hrs	12	(4.3)
	50 - 60 hrs	110	(39.3)
	More than 60 hrs	151	(53.9)
Working hrs on EHR per week	Up to 5 hrs	68	(24.3)
	5 - 10 hrs	109	(38.9)
	10 - 15 hrs	79	(28.2)
	More than 15 hrs	24	(8.6)

Considering the awareness level of respondents, a questionnaire consisting of 40 questions has been administered to them. They have been requested to rate the importance of factors affecting data quality and job characteristics on a 5-point likert scale (1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree). The demographic variables sample has been categorized on the basis of various demographic variables according to understand the diverse behavior of respondents. On the basis of convenience sampling, a total of 350

questionnaires have been distributed and 280 fully-filled responses have been received. Data have been analyzed by applying principal component analysis and multiple regression analysis with the help of SPSS software, version 23.0.

4.Results

Principal Component Analysis

To summarize the data without losing any of the important information (Tabachnick & Fidell, 2007), principal component analysis (PCA) has been conducted. The Kaiser-Meyer Olkin (KMO) value >0.6 can be considered as adequate (Kaiser & Rice, 1974). From Table 2, it can be seen that KMO value (0.847) has been found to be acceptable, and seems to be indicative of a data set considered to be highly desirable for factor analysis (Kim & Mueller, 1978).

The most common method of factor analysis is the principal component analysis and the most common method of factor rotation is the varimax rotation (Kinnear & Gray, 2010). Thus, the factors affecting data quality have been tested using principal component factor analysis with varimax rotation. With principal component analysis, four factors have been extracted depending on Eigen values and variance explained as shown in Table 3. The factor structure has explained 65.74% of total variance, which exceeds the threshold level i.e., 60% commonly used in social sciences (Hair et al., 1995).

Table 2: Kaiser-Meyer-Olkin and Bartlett’s Test of Sphericity

Kaiser-Meyer-Olkin		0.847
Bartlett's test of sphericity	Approximate Chi-square	28584.230
	Degree of freedom	861
	Significance	.000

Table 3: Factors, Eigen Values and Variances Explained

Factors	Eigen values	Variance (%)	Cumulative (%)
Effectiveness	5.254	19.487	19.487
Accessibility	4.865	17.512	36.999
Accuracy	3.506	15.977	52.976
Timeliness	2.503	12.764	65.74

Naming of Factors

Effectiveness

It has been identified as the first factor affecting data quality from multispecialty healthcare service providers. Six data quality features which significantly load on this factor are 'High quality medicine than paper charts', 'Status of metadata described with the use of e-records', 'Data available using EHRs is inconsistent', 'EHRs provide consistency between items of multiple data from multiple sources', 'Complete view of records, images, tests and referrals is available using EHRs' and 'EHR shows the source and context of data'. Findings of various other studies have also supported these findings (Burton, Anderson & Kues, 2004).

Accessibility

It has been identified as the second factor affecting data quality. Five data quality features which significantly load on this factor are: 'Easy to use interfaces', 'Appropriate form of data available', 'Privacy from unauthorized users', 'Time of entry and retention is stated' and 'Instant data is available on all systems'. Other studies have also shown that these features are considerable determinants of data quality (Garrido et al., 2005).

Accuracy

The third factor affecting data quality has been named as 'accuracy'. Four quality features, namely, 'Easy and intuitive', 'Duplication of data is removed', 'Reduction of errors in documentation', 'Reduced wrong interpretation of data' and 'Less wrong prescription, tests and images using EHR' significantly load on this factor. Findings of other studies have also shown that accuracy of data is important determinants of data quality (Zolot, 1999).

Timeliness

It has been identified as the fourth factor affecting data quality. Four quality features which significantly load on this factor are 'Fast access to records', 'Ordering lab tests, referrals and imaging study is fast', 'Prescription of medicine, tests is fast' and 'Exchange of data between physicians of various departments is on real time basis'. Previous studies have also shown that these features are significant determinants of data quality (Overtveit et al., 2007).

Multiple Regression

To study the association between data quality improvements using EHR and job characteristics of physicians of hospital, stepwise multiple regression analysis has been carried out. Four factors obtained from factor analysis have been taken as independent variables and 'job characteristics' has been taken as dependent variable. The entered variables are significant at 5% level of significance.

Table 4(a) indicates significant correlations between data quality features and job characteristics; overall the data quality features are significantly correlated. Three factors namely, 'effectiveness' ($r = 0.573$, $p < .05$), 'accessibility' ($r = 0.670$, $p < .05$) and 'timeliness' ($r = 0.532$, $p < .05$) have a positive and significant correlation with job characteristics of physicians. The fourth factor namely, accuracy has positive but not significant correlation with the dependent variable.

Accessibility factor $R = .670$ accounts for 44.7% of variance in job characteristics as revealed by table 4(b), $R = 0.714^c$ (with respect to three significant factors those are: 'accessibility', 'effectiveness' and 'timeliness') signifies the correlation between job characteristics and

predicted by regression model. In terms of variability in job characteristics by the fitted model, this amounts to a proportion of $R^2 = 0.510$ which indicates the contribution to the three independent variable and determines the variance of job characteristics as explained by the three data quality factors. The three variables explain 51% of the variance in job characteristics. Durbin-Watson value which tests the serial correlation between errors is 1.996 which is close to 2 indicating that residuals are un-correlated (Field, 2005).

In table 4(c), we can clearly check ($F(1,278) = 226.481, p < 0.05$), and so hence conclude that ‘accessibility’ is strongly related to job characteristics of physicians.

Table 4(a): Summary of Correlation Analysis

Measures	E	AS	AC	T	JC
Effectiveness (E)	1				
Accessibility (AS)	.581	1			
Accuracy (AC)	.475	.600	1		
Timeliness (T)	.526	.605	.592	1	
Job Characteristics (JC)	.573	.670	.516	.532	1

Correlation is significant at 0.05 level

Table 4(b): Model Summary^d

Model	R	R Square	Adjusted R Square	Std. Error of Estimate	Durbin-Watson Value
1	.670 ^a	.449	.447	.55764	1.996
2	.707 ^b	.500	.496	.53235	
3	.714 ^c	.510	.505	.52774	

- a. Predictors: (Constant), Accessibility
- b. Predictors: (Constant), Accessibility, effectiveness
- c. Predictors: (Constant), Accessibility, effectiveness, timeliness
- d. Dependent Variable: Job characteristics

In table 4(d), the β values point to the degree by which each predictor affects the outcome. ‘Accessibility’ ($\beta = .449$), ‘effectiveness’ ($\beta = .242$) and ‘timeliness’ ($\beta = .133$) factors make major contribution to job characteristics as they have the large beta coefficients. *T-test* helps in measuring the significant contribution of each and every predictor to the model. The only factor which is not making significant contribution to the model is ‘accuracy’.

Accessibility → Job Characteristics	Supported
Effectiveness → Job Characteristics	Supported
Accuracy → Job Characteristics	Not Supported

Timeliness → Job Characteristics	Supported
----------------------------------	------------------

Supporting data quality factors on job characteristics

Table 4(c): ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	70.426	1	70.426	226.481	.000 ^b
	Residual	86.446	278	.261		
	Total	156.	279			
2	Regression	78.372	2	39.186	138.273	.000 ^c
	Residual	78.501	277	.283		
	Total	156.873	279			
3	Regression	80.003	3	26.668	95.751	.000 ^d
	Residual	76.869	276	.279		
	Total	156.873	279			

a. Dependent Variable: Job characteristics

b. Predictors: (Constant), Accessibility

c. Predictors: (Constant), Accessibility, effectiveness

d. Predictors: (Constant), Accessibility, effectiveness, timeliness

Table 4(d): Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.163	.177		6.580	.000
	Accessibility	.655	.044	.670	15.049	.000
2	(Constant)	.753	.186		4.053	.000
	Accessibility	.498	.051	.509	9.750	.000
	Effectiveness	.272	.051	.277	5.295	.000
3	(Constant)	.653	.189		3.462	.001
	Accessibility	.439	.056	.449	7.821	.000
	Effectiveness	.238	.053	.242	4.495	.000
	Timeliness	.122	.050	.133	2.420	.016

a Dependent Variable: Job characteristics

5. Conclusion

The present study has been undertaken with a primary objective to identify factors affecting data quality of EHR in multispecialty hospitals. To achieve this, factor analysis technique has been employed. It has been revealed that factors which affect data quality are 'effectiveness', 'accessibility', 'accuracy' and 'timeliness'.

Another objective is to study the impact of data quality factors on job characteristics of physicians on the basis of identified factors. Step-wise multiple regression has been performed to predict variance. 'Accessibility', 'Effectiveness' and 'Timeliness' in this order, have been found to have a positive and significant impact on job characteristics. 'Accuracy' has been found to have a positive but not significant impact.

6.Implications

The study corroborates that the physicians using EHR in hospitals work with full capacity and with enhanced possibility of realizing their full potential. EHR implemented data quality sets encourages physicians to be more efficient, creative and quick decision makers. Every hospital can benefit from IT transformation using EHR to record, monitor and analyze health records for clinical, research and development purposes. The significance of task performed by the physicians is enhanced using EHR high quality data. It leads to a better identification of tasks performed and greater variety in skills of the physicians as compared to use of manual records. A sense of autonomy and immediate feedback provides better follow-up of the patients. In addition, this study is expected to make a meaningful contribution to the literature on healthcare information technology. Researchers may also benefit from this study, while undertaking similar studies in future. For example, accessibility in terms of easy to use interfaces, appropriate form of data, privacy from unauthorized users, time of entry and retention and availability of instant data have maximum impact on job characteristics.

7.Limitations and future research

One limitation of this study is that it has been conducted in north India. Though the results concur with studies conducted in other countries, future research exercises can examine the factors affecting data quality of EHR in different parts of the world. Another limitation is that the study has been conducted with a limited number of respondents using convenience sampling method. Future research can look at increasing the sample size or using other methods of sampling to gain more insight into the factors affecting data quality. Additionally, the researchers can attempt to explore more factors in this regard and a EHR module wise analysis can also be carried out to further understand the factors which affect data quality and hence job characteristics of physicians in multispecialty hospitals.

8.References

1. Bajwa, N.K., Singh, H., & De, K.K. (2017). Critical Success Factors in Electronic Health Records (EHR) Implementation: An Exploratory Study in North India. *International Journal of Healthcare Information Systems and Informatics*, 12(2), 1-17.
2. Boaden, R., & Munir, S. (2001). Patient empowerment and the electronic health record. In *MEDINFO 2001: Proceedings of the 10th World Congress on Medical Informatics, Amsterdam* 663–5. London:IOS Press, IMIA.
3. Burton, L.C., Anderson, G.F., & Kues, I.W. (2004). Using Electronic Health Records to Help Coordinate care. *Milbank Quarterly*, 82(3), 457-81.
4. Crownower, A., & Rosenbaum, S. (2002). Designing user interfaces to maximize user acceptance of clinical information systems. 1998. <http://www.diamondbullet.com>.
5. Department of Health (2002). Information for Health: An Information Strategy for the Modern NHS 1998–2005, 1998. <http://www.doh.gov.uk>.
6. Field, A. (2005). *Discovering Statistics through SPSS*. Sage Publication: London.
7. Garrido, T., Jamieson, L., Zhou, Y., Wiesenthal, A., & Liang, L. (2005). Effect Of Electronic Health Records In Ambulatory Care: Retrospective, Serial, CrossSectional Study. *British Medical Journal*, 330(7491), 581-84.
8. Gil, Y., & Kim, J. (2000). User studies of an interdependency-based interface for acquiring problem-solving knowledge. In *Proceedings of the International Conference on Intelligent User Interface, Louisiana*.
9. Hackman, J.R., & Oldham, G.R. (2005). How job characteristics theory happened. *The Oxford handbook of management theory: The process of theory development*, 151-170.
10. Hair, J.F., Anderson, R.E., Tatham, R.L., & Black, W.C. (1995). *Multivariate Data Analysis*. New Jersey: Prentice-Hall.
11. Jeselon, P., & Schloeffel, P. (2002). ISO/TC 215 Ad Hoc Group report: standards requirements for the electronic health record and discharge/referral plans. <http://www.gpcg.org>.
12. Kaiser, H.F., & Rice, J. (1974). Educational and psychological measurements. *Little Jiffy Mark IV*, 34(1), 111-7.
13. Kamra, V., Singh, H., & De, K.K. (2015). Satisfaction Levels and Recommendation Intentions of Healthcare Consumers: An Exploratory Study Based on Multispecialty Hospitals in North India. *Indian Journal of Clinical Practice*, 26(5), 424-30.
14. Kim, J., & Mueller, C. (2010). *Introduction to factor analysis*. Beverly Hills: Sage Publications.
15. Kinnear, P.R., & Gray, C.D. (2010). *PASW statistics 17 made simple*. Hove, East Essex: Psychology Press.
16. Lieberman, H. (2002). Introduction to intelligent interfaces. <http://web.media.mit.edu>.
17. Mandl, K., & Porter, S. (1999). Data quality and the electronic medical record: a role for direct parental data entry. In *American Medical Informatics Association Three Year Cumulative Symposium Proceedings*.

18. Military Health Systems (2003). Data quality: mission needs statement for Data Quality Integrated Program Team (IPT). <http://www.tricare.osd.mil>.
19. Overtveit, J., Scott, T., Rundall, T.G., Shortell, S.M., & Brommels, M. (2007). Improving quality through effective implementation of IT in healthcare. *International Journal of Quality Health care*, 19(5), 259-66.
20. Scott, N. (2001). *Information Impacts Magazine*. <http://www.cisp.org>.
21. Tabachnick, B., & Fidell, L. (2007). *Using Multivariate Statistics*. Boston: Pearson.
22. Verma, D.S., & Khandelwal, U. (2011). Consumers preferences towards service industry: a factorial study of health care industry. *International Journal of Multidisciplinary Resources*, 1(8), 83-9.
23. Zolot, J.S. (1999). Computer-Based Patient Records. *American Journal of Nursing*, 99(12), 64-69.