Evaluation of a Data Warehouse in an Academic Health Sciences Center

Jane R. Schubart, MBA, MS and Jonathan S. Einbinder, MD, MPH
Department of Health Evaluation Sciences, University of Virginia School of Medicine, Charlottesville, Virginia

ABSTRACT

A data warehouse can provide significant benefits to a health care organization if successfully designed and implemented. The Clinical Data Repository (CDR) at the University of Virginia Health Sciences Center improves access to needed data for clinical research and effective decision making at many levels of the organization. We conducted an evaluation of the CDR using a survey questionnaire and interviews of key executive leaders. Our results suggest factors that influence the initial decision to use an information resource, examine the impact of communication channels, and highlight key issues that determine the continued use and ultimate success of a healthcare data warehouse.

INTRODUCTION

A data warehouse can potentially provide enormous benefit to a healthcare organization in clinical research, quality improvement, and decision support by enabling quick and efficient access to information from legacy systems and linkage to departmental databases. Research shows that the key factors for successful data warehouse implementation are organizational in nature. Management support and adequate resources are most important because these address political resistance.

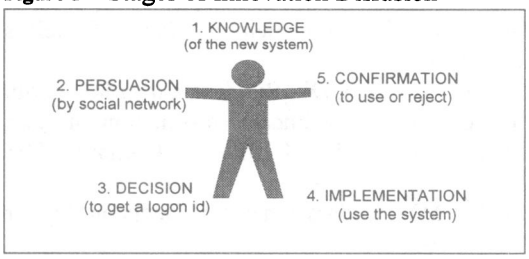
A 1997 survey revealed that the main reason for building a data warehouse is to improve the quality of information in the organization.[16] Yet, even information systems designed to meet obvious organizational needs can fail if developers and administrators neglect to consider the cultural and social impacts of their implementation.[3]

Theoretical Framework

There has been increasing interest in testing and modifying diffusion of innovations (DOI) theory with respect to information technology.[1] Everett Rogers contributed much of the theory of diffusion research.[11] Figure 1 illustrates the five stages he identified through which individuals pass when exposed to new ideas or technologies.

Our research builds on earlier DOI studies by examining the development, adoption, and consequences of a data warehouse in a healthcare setting. Ash has extended the DOI model by examining innovation and organizational attributes that predict infusion and diffusion of healthcare information systems.[1,2] Using the innovation diffusion framework, we focused on an information system that is in the early stages of diffusion, thus allowing us to give timely and very practical feedback to its development team.

Figure 1 Stages of Innovation Diffusion



The Clinical Data Repository

The Clinical Data Repository (CDR) is a frequently updated relational data warehouse that provides users with direct access to detailed, flexible, and rapid retrospective views of clinical, administrative, and financial patient data for the University of Virginia Health System. Authorized users can access the CDR through a standard web browser and can view or download data to their personal computers for further analysis.[13] Its purpose is "to meet the challenge of providing a way for anyone with a need to know – at every level of the organization – access to accurate and timely data necessary to support effective decision making, clinical research, and process improvement."[10]

The CDR is a voluntary system with a diverse user group that includes teaching faculty, administrators, clinicians, researchers, and analysts. Since 1997, when the CDR system went live, more than 150 individuals have obtained a logon id. Thus far, it has been a low cost project, and enhancements continue to be added to make it a more useful tool. Improving the interface, enriching the data dictionary, and meeting users' needs are priorities.

Objectives

We undertook evaluation of the CDR to examine its functionality and understand its adoption at the University of Virginia. We addressed practical questions such as whether the system works

1091—8280/99/$5.00

technically as designed from a user perspective and whether it produces desired results. We used a framework for our analysis that is based on Rogers' diffusion of innovation theory.

We tested the following hypotheses:

1. Initial CDR usage is predicted by an individual's proficiency in pertinent computer applications, familiarity with standard coding conventions, and an understanding of how the UVa Health System records and retrieves data.
2. An individual's initial decision to use the CDR is influenced by communication channels and social networks.
3. Continued use of the CDR is explained by the perceived attributes of the CDR.

METHODS

Study Population

We chose to study users of the CDR, where "users" are defined as faculty and staff with logon id's, as of September 1998, who "hands-on" construct queries and extract data. Students and those involved in the CDR's development were excluded. The result was a sample of 65 users. Fifty-three completed surveys were returned (response rate of 82%).

Survey Questionnaire

A survey was given to all faculty and staff who have a CDR logon id. A questionnaire was developed, building on tools with documented reliability and validity. As a starting point, instruments were collected from various disciplines including information systems research, medical informatics, organizational and management theory, and the social sciences.[3,6,8] The questionnaire contains 39 questions and takes 10-15 minutes to complete.

Specific methods to control non-response included personalization, a mixed mode approach, a simple questionnaire format, multiple contacts, and confidentiality.[5,9,12] We used a combination of electronic mail pre-notice, Web-based questionnaire, and a paper booklet version.

The perceived attributes of an innovation are an important explanation of the rate of its adoption. Rogers states that from 49 to 87 percent of the variance in rate of adoption is explained by the five attributes described in Figure 2.[11] We constructed our questionnaire to address the CDR's functionality in the context of these 5 attributes. We included questions about information content, ease of use, output format, data accuracy, cost, timeliness, work style, and visibility of results.

Figure 2 Rogers' Innovation Attributes

Relative advantage	Degree to which an innovation is perceived as being better than the idea it supercedes.
Compatibility	Consistency with existing values, past experiences, and needs.
Complexity	Degree to which perceived as relatively difficult to understand and use.
Trialability	Extent to which the technology may be experimented with on a limited basis.
Observability	Extent to which the results of the innovation are visible to others.

Executive Interviews

We interviewed 12 key individuals to better understand the organizational factors that influence diffusion. Individuals whose involvement included conceptualization of the project, its development, and funding were selected, including the Medical Schools' Dean, the Chief of Staff, CFO, CIO, and various medical directors, and administrators.

Our questions addressed the CDR's strengths and weaknesses, key success factors and obstacles to implementation, the institutional need that it was intended to address, and the appropriate institutional management and oversight. The interviews were analyzed using QSR NUD*IST 4 (Qualitative Solutions and Research Pty Ltd., Australia).

Statistical Analysis

We used S-PLUS, Version 4.5 (MathSoft, Inc.) for statistical analysis. Variable clustering analysis was performed prior to fitting a binary logistic regression model.[7] The independent variables correspond to a 5-point Likert scale (1 = Expert to 5 = Not at All Familiar). The model computes a Somers' D_{xy} rank correlation between the predictor variables and a binary outcome.[15] When $D_{xy} = 0$, the model is making random predictions; when $D_{xy} = 1$, the predictions are perfectly discriminating. The accuracy of the model was assessed using a "bootstrapping" validation procedure.[7]

RESULTS

Hypothesis 1

We are interested in understanding the factors that influence the *implementation stage* of diffusion. "Implementation" was defined as the initial usage of the CDR. Responses to the question, "Have you personally ever submitted a query using the CDR?" constituted the dependent variable. Sixteen (31%) of the users answered "No" and 36 (69%) answered "Yes", revealing that although all of the Respondents had made a *decision* to use the CDR, by obtaining a logon id, not all had *implemented* it.

Nine questions address the knowledge and skills hypothesized to correlate with the decision to use the CDR. We constructed 3 independent variables that are intuitively appealing because they represent measures of related, but distinct, skill sets: 1) knowledge of ICD-9, CPT, and DRG coding, 2) knowledge of spreadsheet and database software, and 3) knowledge our hospital's patient financial data and cost accounting systems.

Figure 3 Results of Binary Logistic Regression Model to Predict Initial Usage of the CDR

LR X^2	d.f.	p-value	D_{xy}	R^2
14	3	0.0028	0.6	0.34
Variable			X^2	P-value
Knowledge of coding conventions			4.9	0.026
Knowledge of computer software			5.7	0.017
Knowledge of patient financial data			7.9	0.005
TOTAL			9.3	0.026

This model has a D_{xy} statistic of 0.6, indicating moderate utility in predicting individual responses. The log likelihood ratio is a test of overall association for whether any variables included are associated with initial use of the CDR. The LR chi-square test value in this model is 14 with a p-value of 0.0028. All three variables in the model are significantly associated with CDR initial usage. Approximately 5/6 of the model was maintained in the validation procedure, resulting in a "corrected" R^2 of 0.24 which is very acceptable for a binary outcome.

Hypothesis 2

The "persuasion" stage of Rogers' diffusion model was tested with a logistic regression model using responses to the question, "Overall have you heard more positive or negative comments about the CDR?" Our results showed a statistically insignificant relationship between the social networks and the decision to actually use the CDR. We also asked users whether they were influenced to use the CDR by another person, either praising or criticizing it. Fourteen (26%) indicated that they were influenced to use the CDR, and noted that this influence was important in their decision.. Only two Respondents indicated that they were influenced not to use the CDR.

Hypothesis 3

The "confirmation" stage of the diffusion process was assessed by asking users whether they had stopped using the CDR. Twenty-four individuals had stopped using the CDR; 18 said that they had not stopped. We quantified the ability of each of Rogers' constructs to predict the continued use of the CDR by users who have used the resource at least once.

Figure 4
Innovation Attributes and Predicted Continued CDR Usage

Rogers' Construct	Somers' D_{xy}	n
Relative Advantage	-0.053	15
Compatibility	0.52	40
Complexity	0.59	38
Observability	-0.079	39
Trialability	0.27	42
Reliability*	0.33	41

* added construct to Rogers' model

Other Findings

Training

Our survey results confirm the need for training. Three of the 16 Respondents who had not used the CDR listed "not enough training or support" as their major reason. Five Respondents who have discontinued using the CDR cited lack of training as the main reason. Even current users did not rate training highly. 54% disagreed with the statement, "I received sufficient orientation and training to learn how to use the CDR." The need for more training was often mentioned in the responses to the open-ended question. Conversely, user-support provided by the developers was rated very highly.

The "Time" Factor

"Not enough time to try it" was the most often cited reason for not using the CDR. Eight Respondents who have never submitted a query cited lack of time as the major reason. This result points to the time constraints of health care professionals as an important consideration for design of both the information resource and the training programs.

Data Quality

When asked about their satisfaction with the accuracy of the data provided by the CDR, 75% of the Respondents were satisfied. These results were similar to overall satisfaction with data provided by the UVa Health System from all sources.

Encryption

To protect confidentiality, patient and physician identifiers have been encrypted. Unencrypted subsets of data are made available with special approval. 44% reported that encryption posed a problem for them, suggesting that policy be reviewed to better accommodate users' needs for unencrypted data.

Other Computer Uses

Respondents were asked to check the main ways they use a computer from a list of 12 types of use,

including database applications, desktop publishing, and statistical analysis. 51% use the computer in 7 or more ways and 87% use computers in 5 or more ways. E-mail (98%) and word-processing (94%) were the most popular choices. Respondents use a variety of other computerized sources of data including online clinical data and reports generated by the computing department. Manual paper chart review is also used extensively.

Overall Satisfaction and Future Use

Several questions address expected future use and success of the CDR. 74% of Respondents believe that the CDR will be an important resource at the University of Virginia. The usefulness of the CDR was measured on a 10-point Likert scale. 47% of Respondents who have "hands-on" experience rated the CDR *as it is today* with a score of "7" or higher. Overall, Respondents were enthusiastic about the concept of a data warehouse and supported the continued development of the CDR.

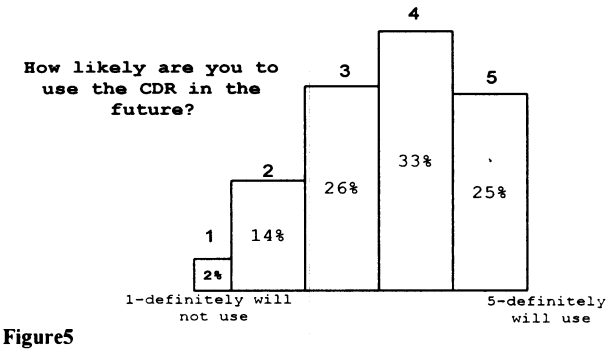


Figure5

Executive Interviews

The most noteworthy observation from the executive management interviews concerns the organizational culture.

> "Until there is even a secondary carrot for getting [CDR type] data, it [will be] a curious group of people who want to use it, some who have the ability to inquire other sources. That it hasn't caught on is not in any way negative toward either the idea or the implementation. The culture has not changed enough that the output of the CDR is perceived as an integral part of every physician's practice self evaluation. We haven't begun to use it in rewards and incentives."

DISCUSSION

Measuring Success

We undertook this evaluation of the CDR to better understand its adoption at the University of Virginia, using Rogers' diffusion of innovation theoretical model. The word "successful" has important implications for measuring diffusion. Criteria for what success might look like includes: high user acceptance, high usage, productivity, technical adequacy, and system flexibility.[4] Level of system use can indicate user satisfaction, especially when system use is discretionary.[14] Frequency of use can affect attitudes toward the system.[3] The logic behind usage as a measure is that a system cannot be a success if it is not used. The CDR users often reported that their usage of the system was somewhat "seasonal". They might use the resource extensively for a project and then not use it at all for several months. This was confirmed by the usage data obtained from the CDR system; usage data did not correlate with satisfaction or expected future use.

We believe that use of a data warehouse is quite different from use of a clinical system. The latter tends to be used daily in patient care, whereas a data warehouse might be used very sporadically. Because the appropriate measure of success is contingent upon the particular situation, our survey focused on expected future usage and perceived effectiveness as measures of success.[8]

Explaining Usage

Rogers' diffusion model adds a framework to better understand how individuals gain knowledge about information systems, the role of "champions" and opinion leaders in persuading them to try a voluntary system, and the importance of perceptions. We now suspect that "persuasion", defined as the influence of positive or negative comments, might relate more to the *decision* to obtain a logon id rather than to the implementation or actual use of the system. A limitation in our study is that we did not survey individuals who were aware of the CDR, had a potential need to use it, but did not get a logon id.

The *implementation* stage is operationalized as the action of submitting a query. We were not surprised that initial use was best explained by the user's proficiency in pertinent computer applications, familiarity with standard coding conventions, and an understanding of how the data are recorded and retrieved at our institution. The types of other computer applications used and the frequency of their use has implications for designing user interfaces and for training.

The lack of association between "relative advantage" and continued use of the CDR is an important finding. Because the CDR was built by linking existing legacy systems and departmental databases together, it does not contain any new data. Slack states that "There is no a priori reason for clinicians

to use computers, when they offer no information that cannot be obtained somewhere else."[14] Perhaps, this is the single, most important advantage of a data warehouse in a healthcare setting, suggesting that improved access (direct, flexible, and rapid) to data is not enough to attract users if the information is available elsewhere. This finding points to enriching the CDR with data not otherwise available as the highest priority. An example is adding data captured by physiological monitoring devices.

CONCLUSIONS

We have shown that diffusion of innovations theory has relevance to understanding the success of a data warehouse system and tested this model in a health care setting. This research identified characteristics of users that impact the initial decision to use a new, voluntary information resource. Compatibility with an individual's work style and skills was strongly associated with satisfaction and continued use of the CDR. Although, users did not rate the "current" CDR highly in terms of "relative advantage" compared to other sources of information, users were enthusiastic about the concept of a data warehouse and the future prospects of the CDR at the University of Virginia.

Second, the importance of organizational culture and the need for data was illuminated by several of the management interviews. Although our user survey did not specifically address the need for data, we now hypothesize that when the organization is truly data-driven and data accountable, that individuals will demand the richness of information housed in a data warehouse. Our conclusions have very practical implications for the future development and focus of the CDR. These observations also suggest areas for future research and inquiry as the CDR evolves.

REFERENCES

1. Ash JS. Factors affecting the diffusion of the computer-based patient record. AMIA Annual Symposium, 1997.
2. Ash JS. Organizational factors that influence information technology diffusion in academic health sciences centers. Journal of the American Medical Informatics. 1997;4:102-111.
3. Aydin CE. Survey methods for assessing social impacts of computers in health care organizations. In: Anderson JG, Aydin CE, Jay SJ (eds). Evaluating Health Care Information Systems: Methods and Applications. Chapter 4. Thousand Oaks, CA: Sage Publications, 1994.
4. Chin H, McClure P. Evaluating a comprehensive outpatient clinical information systems: a case study and model for system evaluation. AMIA Abstract and Presentation, 1995.
5. Dillman DA. Mail and Telephone Survey: The Total Design Method. New York: John Wiley & Sons, 1978.
6. Doll WJ, Torkzadeh G. The measurement of end-user computing satisfaction. MIS Quarterly. 1998;12: 259-74.
7. Harrell, Jr FE. Predicting Outcomes: Applied Survival Analysis and Logistic Regression. University of Virginia, August 12, 1997.
8. Moore GC, Benbasat I. Development of an instrument to measure perceptions of adopting an information technology innovation. Information System Research. 1991;2(3):192-222.
9. Mullin PA. Innovations and insights in the design of web-based surveys. Paper presented at the American Association for Public Opinion Research Conference, St. Louis, MO, 1998.
10. Reynolds R, Knaus R. Clinical data repository enhancements. Internal Memorandum. University of Virginia. January 30, 1998.
11. Rogers E. Diffusion of Innovations. Third Edition. New York: The Free Press, 1983.
12. Schaefer DR, Dillman DA. Development of a standard e-mail methodology: results of an experiment. Paper presented at the 53rd Annual Conference of the American Association for Public Opinion Research, St. Louis, MO, 1998.
13. Scully KW, Pates RD, Desper GS, Connors AF, Harrell FE, Pieper KS, Hannan RL, Reynolds RE. Development of an enterprise-wide clinical data repository merging multiple legacy databases. AMIA Annual Symposium, 1997.
14. Slack WV. Assessing the clinician's use of computers. Editorial. MD Computing. 1993;10(8):357-60.
15. Somers RH. A new asymmetric measure of association for ordinal variables. American Sociological Review. 1962;27:799-811. In: Harrell FE. Predicting Outcomes: Applied Survival Analysis and Logistic Regression. University of Virginia, August 12, 1997.
16. Watson H, Haley B. Managerial considerations. Communications of the ACM. 1998;41(9):32-7.

ACKNOWLEDGEMENTS

This work was supported by an Applied Medical Informatics Fellowship 1F38LM00068-01 from the National Library of Medicine.