Patient-centric Data Warehouse Design

An Empirical Study Applied in Diabetes care

Lichin Chen
Graduate Institute of Biomedical Electronics and Bioinformatics
National Taiwan University
Taipei, Taiwan
e-mail: d98945012@ntu.edu.tw

Chiou-Shiang Wang, I-Ching Wang, Hui-Chu Yu
Department of Nursing
National Taiwan University Hospital, National Taiwan University
Taipei, Taiwan

Hui-Yu Peng, Hui-Chuen Chen
Department of Dietetics
National Taiwan University Hospital, National Taiwan University
Taipei, Taiwan

Chia-Hsiun Chang, Tien-Jyun Chang, Yi-Der Jiang, Lee-Ming Chuang
Department of Internal Medicine
National Taiwan University Hospital, National Taiwan University
Taipei, Taiwan

Feipei Lai
Graduate Institute of Biomedical Electronics and Bioinformatics,
Department of Computer Science and Information Engineering,
Department of Electrical Engineering
National Taiwan University
Taipei, Taiwan

***Abstract* — There are many healthcare settings dedicated for different diseases and treatments. As the information of the patients is collected in different healthcare settings, there is an opportunity to reveal further knowledge and understanding of the patients when the information is integrated. However, the services and evaluation are often isolated. The aim of this research is to design a data warehouse based on a patient-centric vision and to integrate the data to enable inner knowledge exploration. The design is initially applied in the case of diabetes. This study initially integrated the data of regular visits and telehealthcare program of diabetes care, and the data of 61 patients, who participates in both service settings, are collected. The results show that the data inputted by the patients outside the medical institutes are reliable and do represent patient's conditions. The potential value of the data warehouse is promising, and it is valuable to integrate the data across healthcare settings based on a patient-centric vision.**

Data warehouse; diabetes; diabetes self-management education; data integration.

I. INTRODUCTION

Diabetes is a chronic disease, which refers to a person who has high blood sugar, either because the pancreas does not produce enough insulin, or because cells do not respond to the insulin that is produced [1]. Diabetes control relies heavily on patient self-management and life style adjustment. The treatments of diabetes are aimed to postpone and prevent the development of complications, including medications, regular screening, self-management skill education, and long-term follow-ups. Traditionally, patients visit the medical institutes every three month for laboratory tests, diabetes self-management education (DSME), and receive complication screening annually. The data are commonly collected and recorded in information systems and the patient outcomes are evaluated based on the three-month interval data. Recently, telehealthcare has become popular that is designed to enhance self-care behaviors for diabetic patients outside the medical institutes. It combines the information communication technology and the commercialized biometric sensor devices to address disease management at a distance and facilitates longitudinal health status monitoring [2]-[5]. Patients are to record their daily activities on the online self-management information system, and monitor their glucose and other vital signs with the glucometer or other biometric sensors. The online self-management information system is often integrated with the biometric sensors, which enable data uploaded automatically. Patients often record their data on a daily or weekly time interval. Nevertheless, the services mentioned above are usually isolated and rarely connected to one another, and the information systems and the databases are commonly scattered. When it comes to evaluate patient's performances and outcomes, the two services are commonly done individually.

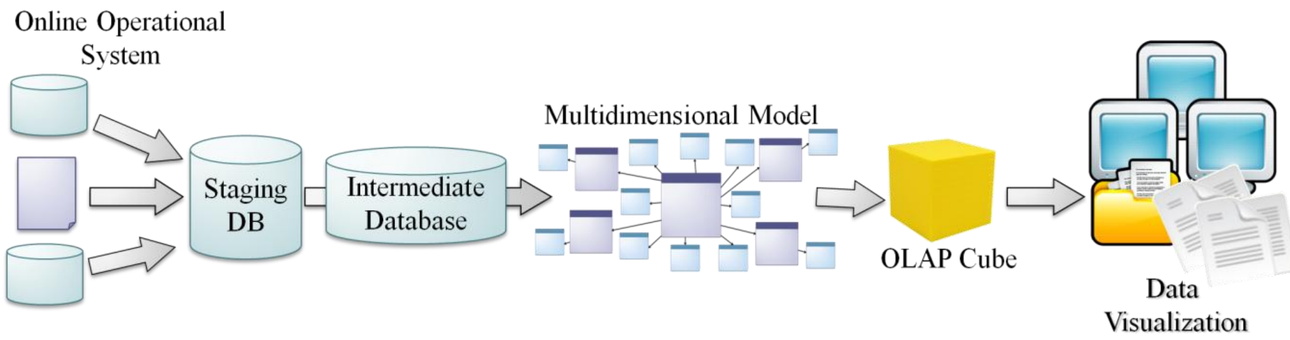

Figure 1 Process flow of data warehousing

However, integrating data from different perspectives offers an opportunity to further explore and understand patient conditions. Meanwhile, data warehousing has become prevalent in the healthcare industry because of the provision of manipulating large quantities of data and the decisions made based on the data [6]. It represents the process of data centralization, and duplicates the data from the online operational systems and organizes them into the analytical data structures. Data warehousing has the potential to integrate various data among applications, support massive data analysis, and improve the ability of future data mining, big data manipulation, and knowledge discovery [6]-[11]. The aim of this article is to design a data warehouse and offers an opportunity to integrate patient data and further explore patient conditions.

This article will first start with three hypotheses to evaluate patient condition and further introduce the data processing flow of the presented data warehouse in the method section. Based on the hypotheses, the result section presents the evaluation outcomes of patient conditions, and the result is interpreted in the discussion section, and concludes in the conclusion section.

II. METHODS

The aim of this research is to design a data warehouse, which is based on a patient-centric perspective, and further to explore the knowledge of diabetes through the integration of the data from regular visits and telehealthcare program of diabetes care. This study is applied in an educational hospital in northern Taiwan, which has highlighted the values of health information technology, and developed information systems for many clinical practices.

The hospital has implemented a shared care program for diabetes, which supports the patients with regular physician visits, DSME, laboratory tests, and routine follow-ups, and has additionally provided a telehealthcare program recently. The telehealthcare program allows the patients to interact with an online information system and upload daily glucose through an off-the-shelf, 3G glucometer [12]. Patients may participate in one of the programs, shared care program and telehealthcare program or both. The information generated from the services of diabetes patients is shown in Table I. It can be observed that the data are integrated from three healthcare settings and three heterogeneous databases.

After the construction of the data warehouse, this research initiates three hypotheses to evaluate patient's conditions.

1. In order to validate the reliability of the data given by the patients themselves. It is meaningful to find out how does daily glucoses correspond to the HbA_1c tests results?
2. Does telehealthcare program strengthen the skill of patients in choosing appropriate food ingredient after participating telehealthcare program?
3. Does telehealthcare program induce more performances in the frequency of self-monitoring of blood glucose (SMBG)?

This research focused in analyzing the data of twenty-one-month duration, from September, 2011 to June, 2013, and compared patient's skills evaluated by the certificated diabetes educators (CDEs). The daily glucose monitoring data are grouped based on a three-month interval, which matched to the corresponding HbA_1c results. The skill evaluations discussed here consist of the eating behavior and the performing of SMBG. The eating behaviors of patients are evaluated through four aspects of food ingredient intake, including fiber, good fat, high fat, and sodium & desserts, the questionnaires are shown in Table II. Each question scores one point for achieving good behavior, and become the scoring of each ingredient intake. The frequencies of patients performing SMBG were recorded to see the effect of telehealthcare on patients performing SMBG. The frequency is calculated into weekly frequency and daily frequency. The first skill evaluation (T1) acts as the baseline of the self-care ability of patients, and the baseline is compared to the second (T2), and third (T3) evaluations if available.

A. Process of data warehousing

The data warehousing centralizes the data from the online operational systems and integrates them based on the analytical requirements. The data warehouse represents as a central data repository, and also consists of various kinds of

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-335-3

defined analysis dimension tables, which enables the reusability among different analyses. The data warehouse applications commonly consist of three stages of data processing, including data transcription stage, data manipulation stage, and data visualization stage, as shown in Fig. 1. The data transcription stage is done through the use of Extract, Transform and Load (ETL) tools to duplicate the data from the online systems to a staging database in its original format. This is to ensure that the work of data warehousing does not interfere with the online systems.

TABLE I. INFORMATION IN DIABETES TREATMENT.

Information	Service	Source System	Database	Frequency
Diagnosis	Outpatient setting	Outpatient information system	Oracle	Every 3 month
Medication	Outpatient setting	Outpatient information system	Oracle	Every 3 month
Laboratory results	Outpatient setting	Laboratory information system	Oracle	Every 3 month
Diabetes education and assessment	Shared care program	Disease management information system	Sybase	Every 3 month
Self-management records	Telehealthcare	Online self-management information system	SQL server	Daily
Glucose and other vital signs monitoring	Telehealthcare	Online self-management information system	SQL server	Daily

```
      <property name="Virtual" value="0" vartype="11" />
      <property name="VisibleAP" value="0" vartype="3" />
    </ddsxmlobj>
  </layoutobject>
  <connector lineroutestyle="MSDDS.Rectilinear" sourceid="39" destid="37" sourceattach
    <point x="2640" y="-2950" />
    <point x="2640" y="-2494" />
    <point x="2747" y="-2494" />
    <point x="2747" y="-2037" />
  </connector>
</ddscontrol>
<ddscontrol controlprogid="DdsShapes.DdsObjectManagedBridge.2" tooltip="Station" left=
  <control>
    <ddsxmlobjectstreaminitwrapper binary="00080000401a0000fe080000" />
  </control>
  <layoutobject>
    <ddsxmlobj>
      <property name="LogicalObject" value="FALLDOWN_DIM_FD_EVENTDEPT" vartype="8" />
    </ddsxmlobj>
  </layoutobject>
  <shape groupshapeid="0" groupnode="0" />
</ddscontrol>
<ddscontrol controlprogid="MSDDS.Polyline" left="508" top="8208" logicalid="122" contr
  <control>
    <ddsxmlobj>
      <polyline endtypedst="6" endtypesrc="3" usercolor="0" linestyle="0" linerender="
    </ddsxmlobj>
  </control>
  <layoutobject>
    <ddsxmlobj>
      <property name="LogicalObject" value="dataSet.Relations[DIM_FD_EVENTDEPT-FACT_FD
      <property name="Virtual" value="0" vartype="11" />
      <property name="VisibleAP" value="0" vartype="3" />
    </ddsxmlobj>
  </layoutobject>
```


Figure 2 Example of XMLA

During data manipulation stage, the data are integrated, calculated, and manipulated in the intermediate database based on the topic and analysis requirements, which is known as the data marts. It is restructured into a Multi-dimensional Data Model (MDM) format based on the analytical requirements. Based on the MDM, the data are aggregated into an XML for Analysis (XMLA) format through on-line analytical processing (OLAP). The XMLA is also known as the OLAP data cube (shown as Fig. 2), which is designed to pre-calculate all the defined numerical values to facilitate instantaneous data queries and multi-factor comparisons [13]-[14].

Finally, the data visualization stage simply visualizes the data cube through a viewer. The users are free to drag the analyzing aspects into the analyzing column to obtain instantaneous queries and explore the data freely. The numbers of analyzing aspect are not limited. The MDM is usually interpreted into the structure of star schemas or snowflake schemas, which consists of a fact table in the middle and numerous dimension tables surrounding it. Each dimension represents a defined analysis aspect. Among the dimensions, few of the characters consisted of multiple-choice items that could refer to more than one option, such as medication. The star schema is suitable for singular items, and the snowflake schema is able to aggregate data in a normalized way. The relationships, attributes, and hierarchies between dimensions and fact tables are also defined during this stage.

TABLE II. QUESTIONNAIRES OF THE SKILL EVALUATIONS.

Ingredient	Questionnaire
Fiber	At least 2 meals consist of grain crops and rice each day.
	At least 1.5 bowl of vegetable a day.
	Eating fruit every day.
Food fat	Not taking in fat meat weekly.
	Not using pig fat at home.
	Not drinking complete fat milk.
	Not having saturated, trans fat and cholesterol-rich foods twice a week.
High fat	Not having deep fry more than 3 times a week.
	Not having 7 meals or more outside every week.
Sodium & desserts	Not having 2 times or more unplanned biscuits or dessert weekly.
	Not having pickled food 3 times or more weekly.

TABLE III. DEMOGRAPHIC INFORMATION OF PATIENTS.

		n	%
Diabetes Type	Type I	19	31.1
	Type II	42	68.9
Age	< 40	14	23.0
	40 ~ 65	38	62.3
	> 65	9	14.8
Gender	Male	30	49.2
	Female	31	50.8
Insulin	Yes	45	73.8
	No	16	26.2
Oral hypoglycemic agents	Yes	40	65.6
	No	21	34.4

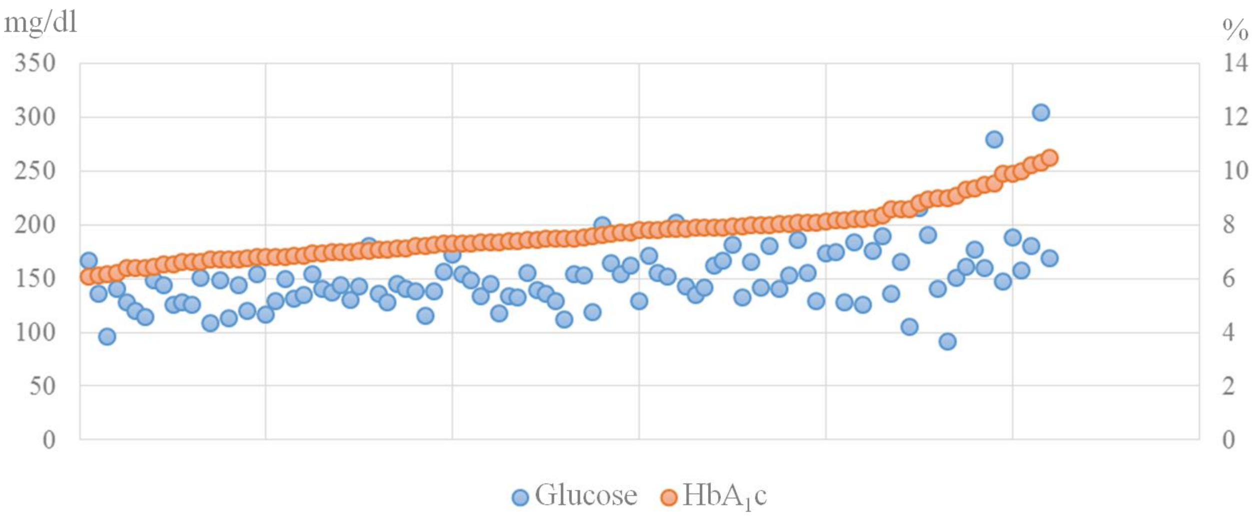

Figure 3 Relationship between daily glucose and HbA_1c

III. RESULTS

A total of sixty-one patients participate in both service settings, and Table III shows the demographic information of the patients recruited. Fig. 3 illustrates the correlation of the daily glucose and the HbA_1c results, and the patients are arranged and listed according to their HbA_1c level. It can be observed that the two values are highly related and the daily glucose recorded by the patients does represent their outcomes.

TABLE IV. SCORES OF HEALTHY EATING.

	Ingredient (Score)	Mean	SD
T1 (n = 60)	Fiber score (3)	1.40	0.89
	Good fat score (4)	1.79	1.61
	High fat score (2)	0.69	0.74
	Sodium score (2)	0.51	0.75
T2 (n = 26)	Fiber score (3)	1.65	1.09
	Good fat score (4)	1.04	1.51
	High fat score (2)	0.38	0.64
	Sodium score (2)	0.27	0.60
T3 (n = 7)	Fiber score (3)	1.29	1.11
	Good fat score (4)	1.71	1.70
	High fat score (2)	0.57	0.79
	Sodium score (2)	0.71	0.95
T4 (n = 2)	Fiber score (3)	2.00	1.41
	Good fat score (4)	1.00	1.41
	High fat score (2)	0.50	0.71
	Sodium score (2)	0.50	0.71

Table IV shows the food ingredient score of the patients. Ninety-eight percent (n=60) of the patients were educated with healthy eating, but only 43% (n=26) of the patient were educated the second time, and even less were educated the third time. Comparing to T1, fiber score has increased in T2 and T4, good fat score and high fat score have decreased in T2, T3, and T4, and sodium score has decreased in T2 and T4. Table V shows the frequency of patients performing SMBG. 100% (n=61) of the patients were educated with SMBG, and 95% (n=58) of the patients were educated the second time. The frequency of SMBG has increased in T2 and T4.

TABLE V. SCORES OF BLOOD GLUCOSE MONITORING.

		Mean	SD
T1 (n = 61)	Weekly	7.42	9.62
	Daily	1.06	1.37
T2 (n = 58)	Weekly	8.34	10.13
	Daily	1.19	1.45
T3 (n = 38)	Weekly	7.30	11.07
	Daily	1.04	1.58
T4 (n = 8)	Weekly	8.25	5.87
	Daily	1.18	0.84
T5 (n = 2)	Weekly	3.50	4.95
	Daily	0.50	0.71

IV. DISCUSSION

The reuse of data and data warehousing has become more prevalent due to the large quantities of data stored and the amount of decisions based on the data. The data warehouse offers the opportunity of obtaining better information, which results in better quality of care. The data

 ISBN: 978-1-61208-335-3

warehouse supports dynamic comparison among multiple factors and provides instantaneous query, and is a powerful tool in the development of protocols for treatments and further knowledge exploration of diseases. The cleaning and the verifying of data require a lot of efforts. When the original design of the online information system offers free text input to provide the convenience of data input, it is common to obtain data that are easy for human to interpret but difficult for the machine to recognize, such as inputting “Y” for yes and the entire word “YES” means the same, but requires extra definitions for the system to recognize and identify them as the same meaning.

The result shows that while there are still debates on the reliability of the data that are inputted by the patients outside the medical institutes, the result in this study shows that the data do correspond to the results of the blood tests, and implies that the data are reliable. The eating behavior and the performing of SMBG are basic skills in self-management for diabetes patients. The result shows an unstable variation but a moderately positive outcome for patients in the behavior of healthy eating and SMBG performing. Generally, fiber score has increased and the other scores have decreased. Less desirable outcomes appear in T3 for eating behaviors and SMBG performing, the causes of such condition require further study.

Missing data and incomplete records are common when collecting data from online information systems. Meanwhile, there are seven skills to be enhanced in self-care behaviors. It is unlikely for CDEs to educate individual skill repeatedly or over three times in a short period of time, and it is difficult see the changes of the patients without repeated evaluation. Also, the evaluations were done by four different CDEs, and the results may differ from one to another, which has become a limitation of this study. More work is needed to further explore the behavior and outcome of patients. It would be promising in revealing more information by adding the laboratory results into the analysis.

V. Conclusion

As the services and treatments of healthcare become more and more advanced, and the involvement of the information communication technology increases, the integration of healthcare services has become essential in future service development.

The integration the data across healthcare settings based on a patient-centric vision is valuable and is supportive in the development of the research field of integrated healthcare. This study initially explored the integration of the data of diabetes care and applied in the validation of three hypotheses about patient conditions. The result shows that daily glucoses do correspond to the HbA_1c tests results, and the data inputted by the patients are reliable. The result shows a moderately positive encouragement for patients in the behavior of healthy eating and SMBG performing. However, the causes of the changes require further study. More work is required to specify the role of data warehouse in the healthcare industry and patient care.

References

[1] L. Chen, H. C. Yu, H. C. Li, Y. V. Wang, H. J. Chen, I. C. Wang, C. S. Wang, H. Y. Peng, Y. L. Hsu, C. H. Chen, L. M. Chuang, H. C. Lee, Y. Chung, and F. Lai, “An architecture model for multiple disease management information systems,” Journal of Medical Systems, 37(2): pp. 9931, 2013.

[2] H. Y. Chiu and C. M. Chen, “Telenursing: The Integration of Information Technology and Community Health Nursing,” Yuan-Yuan Nursing, 4 (2), pp. 5-10, 2010.

[3] L. Heinemann, “Measuring glucose concentrations: daily practice, current and future developments,” Journal of Diabetes Science and Technology, 2 (4), pp. 710-717, 2008.

[4] E. H. Wagner, “The role of patient care teams in chronic disease management,” British Medical Journal, 320 (7234), pp. 569-572, 2000.

[5] S. Y. Liu, J.L. Hsiao, J. L. Shen, and H. Q. Li, “A Research for Information Integration in Case Management System of Diabetes Mellitus,” The Journal of Nursing, 4 (2), pp. 169-179, 2006.

[6] M. Silver, T. Sakata, H. C. Su, C. Herman, S. B. Dolins, and M. J. O Shea, “Case study: how to apply data mining techniques in a healthcare data warehouse,” Journal of Healthcare Information Management, 15 (2), pp. 155-164, 2001.

[7] T. T. Lee, C. Y. Liu, Y. H. Kuo, M. E. Mills, J. G. Fong, and C. Hung, “Application of data mining to the identification of critical factors in patient falls using a web-based reporting system,” International Journal of Medical Informatics, 80 (2), pp. 141-150, 2011.

[8] M. de Mul, P. Alons, P. van der Velde, I. Konings, J. Bakker, and J. Hazelzet, “Development of a clinical data warehouse from an intensive care clinical information system,” Computer methods and programs in biomedicine, 105, pp. 22-30, 2012.

[9] M. J. Ball, C. Weaver, and P. A. Abbott, “Enabling technologies promise to revitalize the role of nursing in an era of patient safety,” International Journal of Medical Informatics, 69 (1), pp. 29-38, 2003.

[10] M. C. Tremblay, R. Fuller, D. Berndt, and J. Studnicki, “Doing more with more information: Changing healthcare planning with OLAP tools,” Decision Support Systems, 43 (4), pp. 1305-1320, 2007.

[11] J. H. Lubowitz and P. A. Smith, “Current Concepts in Clinical Research: Web-Based, Automated, Arthroscopic Surgery Prospective Database Registry,” Arthroscopy: The Journal of Arthroscopic & Related Surgery, pp. 425-428, 2011.

[12] L. Chen, L. M. Chuang, C. H. Chang, C. S. Wang, I. C. Wang, Y. Chung, H. Y. Peng, H. C. Chen, Y. L. Hsu, Y. S. Lin, H. J. Chen, T. J. Chang, Y. D. Jiang, H. C. Lee, C. T. Tan, H. L. Chang, and F. Lai, “Evaluating Self-Management Behaviors of Diabetic Patients in a Telehealthcare Program: Longitudinal Study Over 18 Months,” Journal of medical Internet research, 15 (12), pp. e266, 2013.

[13] M. C. Tremblay, R. Fuller, D. Berndt, and J. Studnicki, "Doing more with more information: Changing healthcare planning with OLAP tools," Decision Support Systems, vol. 43, pp. 1305-1320, 2007.

[14] N. Prat and J. Akoka, “A UML-based data warehouse design method,” Decision Support Systems, 42, pp. 1449-1473, 2006.