Regis University
ePublications at Regis University

All Regis University Theses

Spring 2010

The Use of Olap Reporting Technology to Improve Patient Care Services Decision Making Within the University Health Center

Carlos M. Gonzalez
Regis University

Abstract

The purpose of this paper is to demonstrate that it is feasible for the student health center to leverage existing clinical data in a data warehouse by using OLAP reporting in order to improve patient care and health care services decision making. Historically, University health care centers have relied mainly on operational data sources for critical health care decision making. These sources of data do not contain enough information to allow health officials to recognize trends or predict how future changes in health care services might vastly improve overall heath care. Four proof of concept artifacts are constructed through design science research methodology, and a feasibility study is presented to build the case for the adoption of OLAP reporting technology. The study concludes that it is feasible to implement an OLAP reporting infrastructure at the student health center if physicians, clinical staff, and administration clearly define the need for the new technology, develop proper data extraction, loading, and transformation strategy, and adequately provide project management and data warehouse design towards the implementation of the solution.

Acknowledgements

I thank the Lord for His strength and grace during the development of this project from conception to completion. I am also gratefully thankful to my wife and my two daughters for their patience and support during the development and implementation of this project. My most sincere thank to my parents and sisters for their support and prayers for the completion of my project. I am also thankful for the instruction, advice, and support provided by Professor Shari Plantz-Masters program advisor, and Professor Darl Kuhn content advisor.

Table of Contents

A THESIS i
Abstract ii
Acknowledgements iii
Table of Contents iv
List of Figures v
List of Tables vi
List of Appendices vii
Chapter 1 – Introduction 1
1.1 Project Background 11
1.2 Research Ontology definition and underlying Epistemology 15
1.3 Project Barriers and/or Constraints 15
1.4 Project Contributions to the field of Study 16
1.5 Project scope 17
Chapter 2 – Review of Literature and Research 18
2.1 Literature Review Strategy 18
2.2 Body of Literature 19
2.3 Literature Review Synthesis 19
2.4 Literature Comparison 30
2.5 Literature Findings and Conclusions 34
Chapter 3 – Methodology 37
Chapter 4 – Project Analysis and Results 46
4.1 Challenge Question #1 – Health Education Programs 47
4.2 Challenge Question #2 – Immunizations Compliance 55
4.3 Challenge Question #3: - Health Services Performance 62
4.4 Challenge Question #4: - Episode of Care Data Mart 68
4.5 Baskerville et al (2008) - Design Science Research Risk Management 78
4.6 OLAP Reporting for the Student Health Center: Feasibility Study 82
Chapter 5 – Project History 100
Chapter 6 – Conclusions 101
References 106
References - Continued 107
References - Continued 108
Annotated Bibliography 149
Glossary 179

List of Figures

Figure 1 - SHC Conceptual Architecture 10
Figure 2 - SHC OLAP Reporting Project Processes Data Flow Diagram 40
Figure 3 - Peffers et al (2007) Design Science Research Method Process Model 44
Figure 4 - DSRM Process Model with Integrated Risk Management Framework 45
Figure 5 - Health Ed. Programs Performance Conceptual Schema 53
Figure 6 - Health Ed. Programs Logical Design 54
Figure 7 - Immunization Compliance Conceptual Schema 60
Figure 8 - Immunization Compliance Logical Design 61
Figure 9 – Health Service Performance Conceptual Schema 66
Figure 10 – Health Service Performance Logical Model 67
Figure 11 – Episode of Care Conceptual model 74
Figure 12 – Episode of Care Logical model 75
Figure 13 - Student Health Center Data Mart Architecture 105

List of Tables

Table 1 – Literature Review Strategy 18
Table 2 - Risks to consider based on Baskerville et al Design Science Risk Framework 22
Table 3 – Adapted Golfarelli and Rizzi "What vs. how in testing" table 28
Table 4 – Adaptation of Golfarelli and Rizzi Testing Activities 29
Table 5 – Research Project Work Break Down structure 41
Table 6 - Design and Development Approach 43
Table 7 - Artifact #1 - Requirement Specification Attributes 52
Table 8 - Artifact #2 - Requirement Specification Attributes 57
Table 9 - Artifact #3 - Requirement Specification Attributes 65
Table 10 - Artifact #4 - Requirement Specification Attributes 71
Table 11 – Artifact Evaluation Results Summary – Part 1 76
Table 12 – Artifacts Evaluation Results Summary – Part 2 77
Table 13 - Student Health Center Technology Infrastructure 85
Table 14 - Hwang and Xu Data Warehouse Critical Implementation Factors 86
Table 15 - Hwang and Xu System Success Variables Table 87
Table 16 - Hwang and Xu Implementation Factors Feasibility 88

List of Appendices

Appendix A - Some of the Research Rigorous Methods Used 109
Appendix B – Analysis of Challenge Question #1 110
Appendix C – Analysis of Challenge Question #2 117
Appendix D – Analysis of Challenge Question #3 120
Appendix E – Analysis of Challenge Question #4 123
Appendix F - Restriction Levels 128
Appendix G – Health Education Artifact Functional Dependencies 129
Appendix H – Vaccination Artifact Functional Dependencies 130
Appendix I – Health Services Artifact Functional Dependencies 131
Appendix J – Episode of Care Artifact Functional Dependencies 132
Appendix K – Health Education Programs Artifact Evaluation Queries 133
Appendix L – Immunization Compliance Artifact Evaluation Queries 136
Appendix M – Health Services Performance Artifact Evaluation Queries 141
Appendix N - Episode of Care Artifact Evaluation Queries 143
Appendix O – Sample Health Services Performance Artifact Evaluation Reports 146
Appendix P – Sample Microsoft Excel Pivot Table Report (Oracle Database Data) 147
Appendix Q - Sample Health Services Performance HTML Report (Oracle Reporting Services) 148

Chapter 1 – Introduction

The use of business intelligence technologies in the healthcare industry is one of the common research topics found in the literature along with decision support systems, OLAP technologies, and data warehousing. Health care in general seems to be embracing business intelligence through the use of data warehousing and data analysis tools (OLAP and data mining) as enabling technologies to help improve quality, and efficiency of clinical practice (Ledbetter, C.S, & Morgan, M.W, 2001) for planning (Hristovski, D., Rogaè, M., & Markota, M., 2000; Tremblay, M.C., Fuller, R., Berndt, & D., Studnicki, J., 2007), and to improve financial and operational performance (Glaser, J., & Stone, J., 2008). The literature suggests that the use of OLAP technology and data warehousing in the health care industry can lead to higher efficiencies (Tremblay et al, 2007; Gordon, B. D., & Asplin, B. R., 2004).

Several studies in the literature about data warehouse and decision making technologies in health care, strongly support the use of OLAP for data analysis (clinical and operational) and decision support processes (Tremblay et al, 2007; Gordon & Aspilin, 2004). The health care related literature also suggests that leveraging patient historical data (clinical and non-clinical) can be beneficial to understand trends on the different areas of patient care and health care services (Gordon & Asplin, 2004; Tremblay et al, 2007; Ledbetter & Morgan, 2001).

However, while it is common to find some level of automation in the student health center (practice management systems and/or electronic medical record systems), based on the literature review conducted for this study, the University student health center does not appear to be an environment where data warehouses and OLAP technologies are used for many obvious reasons. However, the literature review did not provide any evidence suggesting that the student health center is an environment using or not using data warehouses and/or OLAP technology for decision making support.

In order to support this generalization, several factors were identified as potential inhibitors for the adoption of business intelligence technologies in the University student health center. Some of the factors to consider are: limited funds, lack of IT staff trained and experienced on business intelligence, no experience or knowledge about data warehousing and OLAP technologies, and the lack of understanding on how a data warehouse and OLAP tools can help leverage existing data generated by the different health care related areas of service within the student health center.

The literature suggests that the need to make better informed decisions within the health care industry is often triggered by financial, quality, and strategic challenges. For instance, Canel and Fletcher (2001) use the results of a student health center quality of service study to make investment decisions for performance improvement sake and to minimize unnecessary spending (Canel, C., & Anderson Fletcher, E. A., 2001). Eilers (2004), present the case of a quality improvement initiative at a student health center resulting from a patient satisfaction survey showing that students rated waiting time lowest in the list of categories indicating their dissatisfaction (Eilers, G.M., 2004). On a different study regarding quality of service Kenagy, Berwick, and Shore (1999) point out the importance of placing the focus of quality of service on the patient (Kenagy, J.W., Berwick, D.M., & Shore, M.F., 1999).

This project seeks to support the idea that the University student health center just like any other healthcare organization can leverage existing healthcare data (clinical and operational) to support decision making processes. This study also seeks to demonstrate how multidimensional data can be analyzed in order to identify trends that can help improve patient care and health care services.

This study is aimed at determining if it is feasible for the student health center to implement an OLAP reporting solution useful for leveraging clinical data in order to improve decision making processes related to patient care and health care services. The goal of this paper is to present a proof of concept study, and a feasibility study to make the business case for the adoption of OLAP reporting technology. Four hypothetical problems or challenge questions are presented as examples of the possible decision making challenges faced by the student health center. The challenge questions are analyzed and the conclusions and/or recommendations are used as input to the Design Science Research process.

A Design Science Research (DSR) method is used to develop the artifact solutions relevant to the given challenge problems demonstrating through well-executed evaluation methods the utility, quality, and efficacy of the design (Hevner, A. R., March, S. T., Park, J., & Ram, S., 2004). This study adapts Peffers, Tuure, Rothenberger, and Chatterjee, (2007) nominal design-science research model, and makes use of Hevner et al (2004) design-science research guidelines to analyze the literature, and to conduct the design activities (Peffers, K., Tuure, T., Rothenberger, M. A., & Chatterjee, S., 2007; Hevner et al, 2004).

Furthermore, this study applies the risk management framework proposed by Baskerville, Pries-Heje, and Venable (2008) to identify, assess, prioritize, and mitigate potential risks inherent to DSR (Baskerville, R., Pries-Heje, J., & Venable, J., 2008).

Design science research is different from regular design in its “clear identification of a contribution to the archival knowledge base of foundations and methodologies.” (Hevner et al, 2004)

The four artifacts developed in this study represent relevant technology based solutions in the form of data marts that can be used and/or enhanced by the student health center physicians, clinical staff, and management to service the decision making needs of the organization. The utility of the artifacts is demonstrated by the execution of queries and the creation of reports. Pivot tables were also created as a way to demonstrate the use of Microsoft Excel as an analysis tool. The design of each of the artifacts was kept simple but effective as a way to encourage design enhancements as next evolution of the project.

The DSR process is followed by the feasibility assessment for the proposed technology. The feasibility study covers areas like organizational feasibility, technology feasibility, and cost feasibility. Specifically this study points out that one of the critical success factors for the OLAP reporting implementation is the hiring and/or contracting of qualified staff for the roles of project manager and data warehouse database administrator role. The feasibility study also presents a discussion of several implementation tools like Microsoft SQL Server 2005 Reporting services, Oracle BI Tools, and Microsoft Excel.

According to Connolly and Begg (2005), a data warehouse is “A subject-oriented, time-variant, and non-volatile collection of data in support of management’s decision-making process.” (Connoly, T., & Begg, C., 2005, 1151)

For the purposes of this study, the subject-oriented data will be any of the resulting objects of analysis, or subject areas from the analysis of the challenge questions. A subject area could be patient diagnosis and respective treatment, health care service performance, and/or other specific business subject critical to the business. Also, the time-variant characteristic of the data warehouse is represented in this study by the use of time (i.e. year, month, and/or day) as a way to represent data specific to the subject of analysis. A time dimension is used to provide a retrospective view of the data and to allow for hierarchical analysis of that data (Ponniah, P., 2001; Hüsemann, B., Lechtenbörger, J., & Vossen, G., 2000; Perdesen, T. B., & Jensen, C. S., 1998). The non-volatile characteristic of the data warehouse refers to the collection of clinical and non-clinical data that will be historically preserved for analysis from the different multidimensional perspectives.

Online Analytical Processing (OLAP) is described by Connolly and Begg (2005), as the technology "that uses a multi-dimensional view of aggregated data" to provide access to strategic information for analysis (Connolly & Begg, 2005, 1205). The multidimensional view and aggregation of data is made possible by the use of OLAP tools (Vassiliadis and Sellis, 1999). According to Gorla (2003), some of the common capabilities found in an OLAP system are multi-dimensionality, aggregation, drill-down and roll-up, and slicing and dicing (Gorla, N., 2003, p. 112). As a result "On-Line Analytical Processing (OLAP) has emerged as a valuable tool for the analysis; navigation and reporting of hierarchically organized data from data warehouses." (Oliveira, R., Bernardino, J., 2006)

According to Jarke, Lenzerini, Vassiliou, and Vassiliadis, OLAP multidimensional models are known as OLAP cubes (Jarke, M., Lenzerini, M., Vassiliou, Y., Vassiliadis, P., 2003, 88). The cube represents data as an array of cells with each dimension of the array representing a dimension, and cells or content of the array are the measures (Vassiliadis & Sellis, 1999; Connolly & Begg, 2005, p. 1209). The physical implementation of the OLAP cube can be accomplished through the use of multidimensional OLAP (MOLAP) or relational OLAP (ROLAP) architecture. According to Gorla (2003), the two are different because in MOLAP the data is stored in multidimensional arrays, and in ROLAP the data is aggregated and stored along with relational databases (Gorla, 2003, p. 112).

The logical representation of the OLAP cube in the ROLAP architecture can be accomplished through two available logical structures, the star schema, the snowflake schema, or the starflake schema (Connolly & Begg, 2005, p. 1185). According to Connolly and Begg (2005), a star schema is "A logical structure that has a fact table containing factual data in the center, surrounded by dimension tables containing reference data (which can be denormalized)." (Connolly & Begg, 2005, p. 1183)

According to Ponniah (2001), the star schema "is simply a relational model with a one-to-many relationship between each dimension table and the fact table." Ponniah also points out that the star schema is not a normalized model since the dimension tables are purposely denormalized making it different to the relational schemas used in Online Transaction Processing (OLTP) systems (Ponniah, 2001, p. 220). Some of the advantages of the star schema are the structural simplicity making it easy for users to understand, allowing for optimized navigation, easy adaptation to changes, and suitable for query processing among others (Ponniah, 2001, p. 220; Connolly & Begg, 2005, p. 1185). However, Wang, Zhang, and Ramanathan (2005) point out that the star schema seems to be insufficient for modeling the semantics of complex data spaces such as the clinical data space (Wang, Zhang & Ramanathan, 2005, p. 8).

This study makes use of the ROLAP architecture for the physical implementation of the solution artifacts. According to Vassiliadis and Sellis, the ROLAP architecture has two advantages: it can be easily integrated into other existing relational information systems, and relational data can be stored more efficiently than multidimensional data (Vassiliadis & Sellis, 1999). Another advantage of the use of the ROLAP architecture is that it is supported by commercially available database management systems like Microsoft SQL Server 2005 and Oracle database 11g or 10g. Since both of these database management systems are commonly found in Academic environments, they make their adoption convenient for developing business intelligence solutions based on ROLAP.

The resulting multidimensional models designed as star schemas include dimensions, fact tables and measures representing specific objects of analysis in order to demonstrate the analysis capabilities provided by OLAP. The historical patient data can be analyzed throughout time for a specific disease, treatment, and/or vaccination requirement (dimension) at a specific level of detail provided by each of the dimensional hierarchies in the OLAP cubes.

This project does not take into consideration the integration of both the OLTP and OLAP environments as one data structure for the analysis of real-time data. As pointed out by Conn (2005), real-time data analysis can help organizations improve decision making and business intelligence. Furthermore, Conn has indicated that the problems business face in regards to real-time data analysis is that "OLAP queries are not real-time queries because of the refresh cycle of data into the OLAP data repository." (Conn, S., 2005, p. 515)

Conn (2005) suggests that since operational data is moved into the OLAP data repository after being integrated and transformed through a complicated and time consuming process the disadvantage imposed by the OLAP data repository "is that data may not be (relatively) recent enough to qualify as real-time data for business intelligence purposes." (Conn, S., 2005, p. 516)

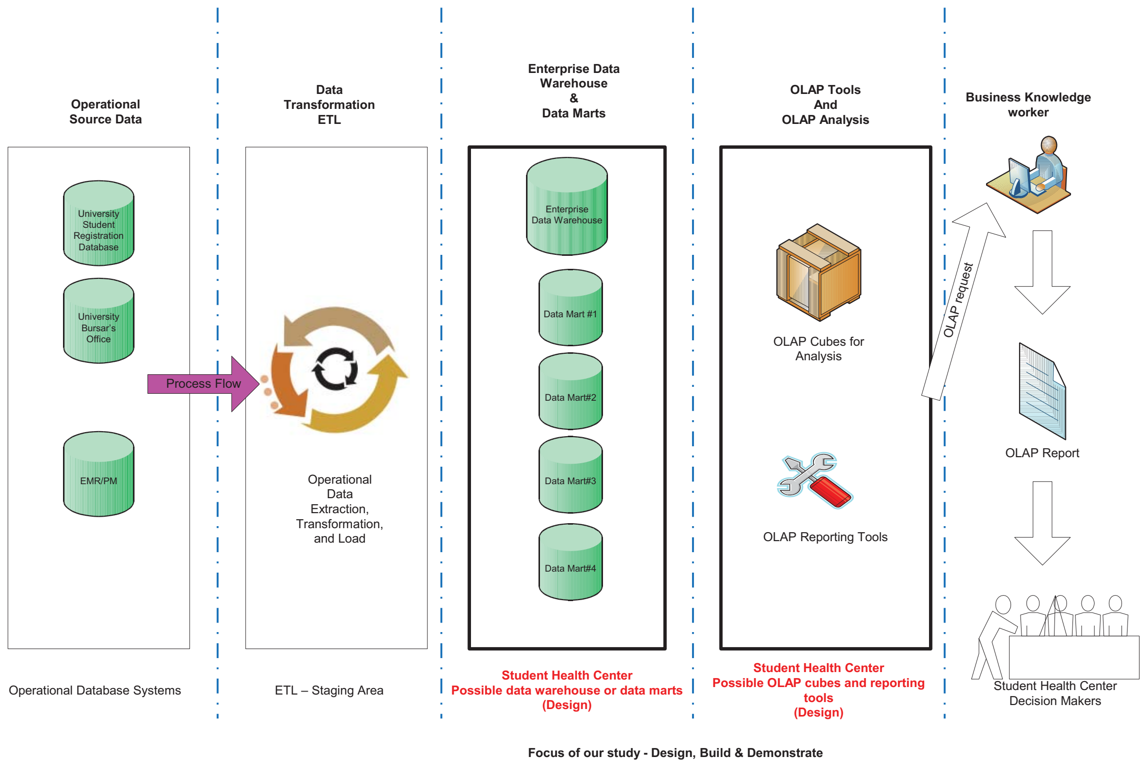
While the integration of the student health center EMR OLTP environment and an OLAP environment would be ideal, in this case it is not feasible since it would be too complex to integrate logical and physical structures, and to obtain adequate performance from the ROLAP system. Therefore, this project follows a traditional or conventional OLAP and data warehouse implementation approach that rely on the data extraction, transformation and loading (ETL) process to move the data into the respective data marts.

The rest of this paper is organized in the following sections. In chapter 1, an introduction and project background is presented. In chapter 2, the literature review and analysis strategy used in the project is presented. The following chapter describes the design-science research methodology for developing knowledge through design and evaluation. Figure 2 shows a project process data flow diagram illustrating the inputs and outputs between project processes. Table 5 shows the work break down structure listing the sequence of work activities to be executed in order to create the project deliverables.

Furthermore, chapter 4 presents a description of the DSR process as used in the study (Peffers et al, 2007), and the related project risks identified from Baskerville, Pries-

Heje, and Venable (2008) risk management framework (Baskerville, R., Pries-Heje, J., & Venable, J., 2008). The figures showing the requirement attributes, conceptual models, and logical design for each artifact are also illustrated in chapter 4. The feasibility study is also presented in chapter 4 including the assessment of the organizational, technological, and cost feasibility of the proposed technology solution. Finally, chapter 5 presents the project history discussion, and chapter 6 discusses the next evolution of the project.

Figure 1 - SHC Conceptual Architecture



1.1 Project Background

The primary goal of this project is to build the business case for the development and implementation of a business intelligence infrastructure for the student health center through the design and development of several data marts used to integrate data from source systems to be used for the analysis of a specific subject area within the healthcare organization. The utility of the proposed prototypes is demonstrated by the ability to produce OLAP reports from each multidimensional model addressing the specific object of analysis. This should ultimately support organizational decision making directed to improve quality of service in the student health center institution.

However, while there is evidence in the literature of the benefits of implementing a data warehouse in combination with OLAP tools within the context of healthcare (Tremblay et al, 2007), in order for this study to be successful on the adoption of this new technology, strong evidence of its feasibility must be presented to the stakeholders and decision makers. Unfortunately, the initial secondary research conducted to narrow down the focus of this study did not find any prior research on OLAP and/or decision making systems within the University student health center.

The lack of research within the context of the student health center motivated the idea of formulating several hypothetical, but relevant tractable questions like: what kind of decision making problems is the student health center currently facing that might benefit from the use of OLAP reporting? What would be the best way to leverage the existing patient data or clinical data to help management make decisions that could potentially improve the quality of service provided by the student health center?

The review of the secondary research literature, Internet searches, and the American College of Health Association website helped identify some of the types of information related challenges a University student health center might face that could be addressed by implementing OLAP reporting capabilities (American College Health Association, 2009). As stated by Canel and Fletcher (2001), quality management is an important issue for any health care institution, but more so for the university health clinics. Many of the University student health centers operate as an ancillary service part of the University system. The effectiveness of the services and programs offered by the student health center can be questioned at any time based on the need to make decisions about funds allocation and service performance improvements (Canel & Fletcher, 2001).

For example, the effectiveness of the substance abuse education program can be questioned by parents, University administration, and/or even the media based on the issues related to alcohol abuse. The literature suggests that alcohol consumption and abuse in college institutions is a problem. According to Ehrlich, Haque, Swisher-McClure, and Helmkamp (2006), student alcohol consumption "results in deaths, injuries, crimes, and sexual assaults." (Ehrlich, P. F., Haque, A., Swisher-McClure, S., & Helmkamp, J., 2006)

Allowing management and knowledge workers to extract data from a data warehouse or data mart to perform analytical data functions can help generating reports showing the level of effectiveness of the program. A report can be generated to answer the question of what percentage of students showed GPA improvement upon their completion of the alcohol abuse program. In order to provide such report, data about the completion of the alcohol abuse program from the health education department of the student health center will need to be correlated to student GPA data during the specific term. This study does not claim that this question can be answered and/or that it is easy to be answered. However, providing historical evidence of the level of influence of the alcohol abuse course on student's academic performance can help management promote informed decisions about health education services.

Student immunization is another area of service that can benefit from the use of data warehousing and OLAP reporting. International students are required to provide immunization records as an acceptance requirement to the University. The student health center is responsible for the compliance of student immunization requirements, and ultimately the prevention and mitigation of a pandemic threat or outbreak in campus. Data from both the student health center and the bursar's office can be used to generate an ad-hoc report showing the total number of students from China, Japan, Korea, and Vietnam (or from other countries), that did not provide proof of immunization for measles, mumps, rubella, and hepatitis B at the time of registration during a span of seven years. This kind of report can help the University student health center management to monitor potential pandemic threats within the student population and to develop mitigation plans.

Another important factor of service within the student health center is that of cost of services in relationship to patient visits. For example, management may need to make a decision on cost increase to cover operational costs, but the student board believes students should not have to pay for certain services. The student health center administration is then requested to provide data that supports the increase of cost for some services and the reduction of cost for other services.

Finally, the student health center clinical data can be leveraged by physicians and clinical staff through the use of an episode of care data mart. Mehta, Suzuki, Glick, and Schulman (1999) discuss the definition of an episode of care for diabetic foot ulcer based on the analysis of claim data (Mehta S, Suzuki S., Glick H., & Schulman K., 1999, p. 1110).

According to Mehta et al (1999), the literature defines an episode of care as the period initiated after the patient is presented with a diagnosis of a clinical condition and concludes when the condition is resolved (Mehta et al, 1999). Mehta et al, point out that with knowledge of the time course of an illness, physicians can develop management programs for protocols, diseases, and can assess the costs and outcomes of alternative treatment strategies, specifically for the management of patients with diabetes. (Mehta et al, 1999)

Wall, Stromberg, Pothoff, & Kane (2004), point out that the literature defines an episode of care as “a sequence or cluster of health care services related to a particular condition or disease”. According to Wall et al, the definition or construction of episode of cares “from the health care utilization records facilitates the investigation of health outcomes research at the population level.” (Wall, M.M., Stromberg, K.D., Pothoff, S., & Kane, R.L., 2004)

Parmanto, Scotch, and Ahmad (2005) present the use of a multidimensional data warehouse of healthcare rehabilitation outcomes intended to support various outcome analyses of outpatient rehabilitation therapies (Parmanto, B., Scotch, M., & Ahmad, S., 2005, p. 3). Parmanto et al (2005) point out that the outcome analysis supported by the design "has the potential to reduce the length of patients' episodes of care, increase the quality of care, and lead to better health-related outcomes." (Parmanto et al, 2005, p. 7)

Therefore, the conclusions and design approach presented by Parmanto et al (2005) deserve to be considered in the development of a multidimensional model for an episode of care for general diagnosis and treatment outcomes like the one proposed in this study.

1.2 Research Ontology definition and underlying Epistemology

This research study will be conducted as constructivist epistemology to develop and implement a solution to the problem statement in the form of a construct or prototype. The expected outcome of the research should be the identification of the most suitable construction and evaluation methods for the artifact to be developed into a full production system at a later time.

1.3 Project Barriers and/or Constraints

The University student health center can be viewed as an environment of constant activity and data flow. Interruption of clinical services could impact patient care negatively, and could lead to significant adverse events. The implementation of any technology in support of clinical or administrative decision making processes must not present or pose any functional and technical risk to the existing practice management and electronic medical records systems.

Based on the fact stated above, several barriers and constraints can be identified as affecting the development of this study. First, the literature supports the fact that clinical data is more complex than traditional business data when it comes to multidimensional modeling (Pedersen & Jensen, 1998). This study might be limited by the level of complexity addressed in the proposed designs in comparison to the levels required by the different data management groups and/or stakeholders in the student health center.

Also, it is possible that the implementation of the solution data marts could be limited by the Extraction-Transformation-Loading (ETL) workflow process as it may introduce or present unexpected challenges on "schema/data translation and integration" from data sources to the proposed data marts (Rahm & Hai Do, 2000).

The conceptual and logical models presented in this study are constrained by what the stakeholders (physicians and management) could consider to be relevant to them based on their respective data analysis needs. This study is using hypothetical questions to make a business case which might not be relevant to physicians or inadequate for the analysis of the proposed business subjects.

1.4 Project Contributions to the field of Study

The primary contribution of this study is the presentation of four solution artifacts, and the discussion of a feasibility study for the introduction of OLAP reporting technology for decision making support. This study should be considered relevant to the area of business intelligence in healthcare since it is focused on the decision making needs of the University student health center institution that ultimately serves the student population in the Higher Education sector. Other contributions to the field of study are listed below:

- Use of design-science research methodology and guidelines to conduct design and evaluation iterations.

- Prototype a solution in an area of need as a precursor of a full implementation initiative.
- Research and synthesize data warehouse theoretical perspectives to give the readers a theoretical foundation for project.
- Present the use of Baskerville et al (2008) design-science research risk management framework to assess and respond to inherent design-science research risks (Baskerville et al, 2008).

1.5 Project scope

The problem statement in question is narrowly focused by addressing specifically the needs of the area of application in this case the University student health center, and the domain of technology studied, data warehousing and OLAP technologies. In particular, the focus is on the review of the literature related to data warehousing, and OLAP reporting platform used in the context of student health care services that has not been mentioned in the literature as needing business intelligence technologies to improve decision making process in order to improve service quality. Therefore, the scope of this study is focused on the design and construction of four artifacts useful to solve the hypothetical business problems presented in the introduction of this paper. Furthermore, the solution artifacts must reflect the use of rigorous design foundations, and/or design methodologies, and should demonstrate effectiveness in solving the business problems.

Chapter 2 – Review of Literature and Research

2.1 Literature Review Strategy

Table 1 illustrates the literature review approach used in this project, which derives from Jourdan, Rainer, and Marshall (2006) a three phase literature review and analysis on Business Intelligence literature (Jourdan, Z., Rainer, R. K., & Marshall, T. E., 2006).

Table 1 – Literature Review Strategy

Phase 1 Literature Gathering	Phase 2 Literature Selection and Review	Phase 3 Literature Analysis and Synthesis
Literature Gathering: based on research topic and area of application on the specific industry context (i.e. Healthcare & University Student Health Center)	Literature Selection and Review: criteria based on seven design science guidelines presented by Hevner., March, Park, and Ram (Hevner, A. R., March, S. T., Park, J., & Ram, S., 2004)	Analysis and Synthesis of relevant literature
▪ OLAP tools ▪ Multidimensional databases ▪ Multidimensional modeling ▪ Data warehouse in Health care ▪ Data warehouse design methods ▪ Data warehouse implementation methods	▪ Is this article relevant to the study? ▪ Are there any suitable design and/or evaluation methods? ▪ Is the literature proposing any evaluation methods that can help demonstrating the utility, quality, and/or efficacy of the designed artifact?	▪ Contrast Healthcare and Non-Healthcare literature ▪ Compare selected data warehouse design approaches found in the literature ▪ Synthesis of findings, conclusions, and contributions

(Jourdan et al, 2006)

All secondary research literature selected for this study was chosen from peered review sources on the topics of data warehouse, OLAP design, and implementation methods within the context of health care services or specifically the student health center. This study also selected design-science research literature in order to better understand the use of design science research use within the information systems domain. The results of the literature review and analysis should provide theoretical input into the design, development, and evaluation of the solution artifact. The conclusions from the literature analysis are also considered to be part of the contributions to the research community.

2.2 Body of Literature

The body of Literature is comprised of healthcare focused papers and none healthcare related papers from journals, proceedings, and other types of peer-reviewed articles. The Annotated Bibliography section lists the comprehensive bibliography used in this study.

2.3 Literature Review Synthesis

Design Science Literature:

The design science research methodology used in this study is based on the works of Hevner, March, Park, and Ram (2004) which describes the design science paradigm within the context of Information Systems by using a conceptual framework and present guidelines for "conducting and evaluating good design-science research." (Hevner, A. R., March, S. T., Park, J., & Ram, S., 2004, p. 77)

Also, this study makes use of Peffers, Tuure, Rothenberger, and Chatterjee, (2007) design science research "nominal process" in order to present and evaluate the solutions proposed in this project (Peffers, K., Tuure, T., Rothenberger, M. A., & Chatterjee, S., 2007). Finally, this study incorporates the risk framework presented by Baskerville, Pries-Heje, and Venable (2008) in order to assess the inherent design science risks as part of the rigorous evaluation methods characteristic of design science research (Baskerville, R., Pries-Heje, J., & Venable, J., 2008).

According to Hevner et al (2004), "contribution arises from utility". Without utility, development efforts can only be recorded as failed designs. The design science literature is clear on this regard. According to Hevner et al (2004), if the newly constructed artifact does not solve the problem (search, implementability), it has no utility. If utility is not demonstrated (evaluation), then there is not basis upon which to accept the claims that it provides any contribution (contribution). If the problem, the artifact, and its utility are not presented in a manner such that the implications for research and practice are clear, then publication in the IS literature is not appropriate (communication). (Hevner et al, 2004, p. 91)

Hevner et al (2004), also point out that if an existing artifact is adequate to solve the given problem, the creation of a new artifact is unnecessary and irrelevant (Hevner et al, 2004, p. 91). The review and analysis of the chosen literature is also focused on finding cases of successful data warehouse and OLAP technologies implementations specific to health care or preferably the student health center organization.

However, as pointed out by Basekerville et al (2008), there are some inherent risks involved in using design-science research (Baskerville et al, 2008). The design science research risk assessment framework developed by Baskerville et al (2008), is used in this study in combination with Peffers et al (2007) design science research method (Peffers et al, 2007) to identify risks related to the design of each artifact and to propose risk mitigations. Table 2 lists the applicable design science research risks from Baskerville et al approach relevant to this project. A discussion about the risk assessment conducted for this study is presented at the end of chapter 4.

Table 2 - Risks to consider based on Baskerville et al (2008) Design Science Risk Framework

A	Business Needs (Problem Analysis and Choice)
A-1	Selection of a problem that lacks significance
A-4	Poor understanding of the problem to be solved
A-5	Solving the wrong problem
A-6	Poor/vague definition/statement of problem to be solved
A-7	Inappropriate choice or definition of a problem according to a solution at hand
A-8	Inappropriate formulation of the problem
B	**Applicable Knowledge (Retrieved from the Body of Recorded Human Knowledge)**
B-1	Ignorance or lack of knowledge of existing research relevant to the problem understanding
B-2	Ignorance or lack of knowledge of existing design science research into solution technologies for solving the problem
C	**Develop/Build (Develop Theory/Knowledge and Build an Instantiation)**
C-2	Development of a hypothetical (untried) solution which is **ineffective in solving the problem**
C-3	Development of a hypothetical (untried) solution which is **inefficient in solving the problem**
C-5	Development of a hypothetical (untried) solution which cannot be taught to or understood by those who are intended to use it
C-6	Development of a hypothetical (untried) solution which is difficult or impossible to get adopted by those who are intended to use it
C-7	Development of a hypothetical (untried) solution which causes new problems that make the outcomes of the solution more trouble than the original problem
D	**Justify/Evaluate (Justify Theory/Knowledge and Evaluate an Instantiation)**
	The above risks of untried solutions may be reduced through justification (or possibly falsification) of an IS Design Theory (ISDT, Walls et al. 1992) and the evaluation of instantiations of the solution. However, evaluation itself carries risks of making errors, resulting In possible type I (false positive) or type II (false negative) errors (Baskerville et al. 2007).
E	**Applications (of Knowledge to Business and Organizational Problem Situations)**
	Once a new solution has been published and promoted to the public, especially if it doesn't work well or at all, but also even if it actually can work effectively, there are a number of other risks:
E-1	Implementation in practice of a solution does not work effectively, efficiently, and/or Efficaciously
E-2	Misunderstanding the appropriate context for and limitations of the solution
E-3	Misunderstanding how to make use of (implement) the solution
E-4	Inappropriate handling of adoption, diffusion, and organizational change
F	**Additions (to the Knowledge Base of Recorded Human Knowledge)**
	The risks in this area are primarily to the researcher, but also to others engaged in the publication Process and even other researchers and eventually the public at large. Risks include:
F-1	Inability to publish or present research results
F-2	Publication of low significance research
F-3	Publication of incorrect research

(Baskerville et al, 2008)

Healthcare Related Literature:

The healthcare literature reviewed for this study is characterized by a clear emphasis on the natural complexity of clinical data as opposed to business data (Pedersen & Jensen, 1998). Specifically, what makes the data warehousing in healthcare and/or in any of the medical related disciplines and sciences supporting healthcare, is its data. According to Wang et al (2005), business processes are “logically simple and temporally stable, biology has very complex research methodologies and a huge fast-growing body of background knowledge. The task of capturing, modeling and encoding some of the biological knowledge for a data warehouse appears to be a great challenge.” (Wang et al, 2005)

However, the literature also presents strong evidence of the use of data warehousing, OLAP and data mining in the healthcare domain. Ewen, Medsker, Dussterhoft, Levan-Shultz, Smith, and Gottschall (1998), describe the process of developing the business case for a data warehouse for a non profit health care organization (Ewen, Medsker, Dusterhoft, Levan-Shultz, Smith, & Gottschall, 1998). The case study described by Ewen, Medsker, Dusterhoft et al, clearly presented the need for a data warehouse, identified key business areas in need of decision support technology, defined and selected object of analysis, selected “business sponsors”, established cost justification, and determined a group of achievable project goals (Ewen, Medsker, Dusterhoft et al, 1998). The case study presented by Ewen, Medsker, Dusterhoft et al, is relevant to this study because it describes a systematic approach for the implementation of a health care data warehouse that can be used for the student health center (Ewen, Medsker, Dusterhoft et al, 1998).

According to Ledbetter and Morgan (2001), the need to use of data warehousing, OLAP, and data mining to analyze and mine clinical data is motivated by the need to leverage existing clinical data for decision support and to improve quality of care (Ledbetter & Morgan, 2001), (Hristovski et al, 2000). According to Tremblay et al (2007), OLAP analytical tools can help knowledge workers become more efficient in gather data needed for decision making. Unfortunately, the literature also suggests that healthcare data warehouse design methods are behind compared to industries outside of healthcare (Parmanto et al, 2005, p. 2).

The healthcare related literature presented different cases of specific data warehouse implementations and design approaches in support of healthcare related decision making processes (Parmanto et al, 2005), (Bréant, C., Thurler, G., Borst, F., & Geissbuhler, A., 2005), (Berndt, D. J., & Hevner, A. R., 2000), (Verma & Harper, 2001).

Pedersen and Jensen (1998) compare the characteristics of conventional data warehouses with those required for a clinical data warehouse. According to Pedersen and Jensen (1998), the design of a clinical data warehouse requires the use of complex modeling constructs capable of handling the case of a patient with multiple diagnostics, which is not “easily possible using a conventional multidimensional model.” (Pedersen & Jensen, 1998)

Furthermore, Song, Rowen, Medsker, and Ewen (2001) present a study on six different approaches to many-to-many relationships between a dimension table and a fact table based on the relationship of a diagnostic dimension and a billable patient encounter fact table (Song, I-Y., Rowen, W., Medsker, C., & Ewen, E., 2001, p. 6-2). The different approaches analyzed by Song et al (2001) addressing the case of the billable patient encounter fact table and the diagnosis dimension can be viewed as another example of the level of complexity involved in the analysis of health care data.

Finally, the work of Parmanto et al (2005) emphasizes on the need to define adequate levels of grain to allow for "multiple levels of analysis" (Parmanto et al, 2005). The OLAP cube for healthcare rehabilitation data presented by Parmanto et al (2005) makes use of three different levels of grain allowing for multiple levels of analysis. Parmanto et al (2005) suggest that the level of analysis offered by traditional business models is not appropriate for healthcare outcome research. Therefore, multidimensional models used for healthcare outcome analysis must provide the appropriate level of grain required by physicians and/or researchers. (Parmanto et al, 2005)

Non-Healthcare Related Literature:

The non-healthcare related literature is characterized by a strong emphasis on conceptual data modeling methods and techniques. The literature offered an abundance of options on conceptual data warehouse design methods (Hüsemann, B., Lechtenbörger, J., & Vossen, G., 2000; Serrano, M., Trujillo, J., Coral, C., & Piattini, M., 2007; Peralta, V., & Ruggia, R., 2003). In particular, Rizzi, Abelló, Lechtenbörger, and Trujillo (2006), present an overview of outstanding issues in data warehousing focused on modeling and design with the intent to analyze what outstanding research challenges remain (Rizzi, S. & Abelló, A. & Lechtenbörger, J. & Trujillo, J., 2006). According to Rizzi et al (2006), the literature has presented the conceptual modeling of a data warehouse from the point of view of multidimensional and ETL modeling. However, according to Rizzi et al (2006), some important issues like standardization, modeling security, and mining-aware design still outstanding.

In regards to logical modeling, Rizzi et al (2006) suggest that once the conceptual modeling phase has been completed, the logical modeling follows with the purpose of transforming "the conceptual schemata into the logical schemata that can be optimized for and implemented on a chosen target system." (Rizzi et al, 2006, p. 5)

However, despite the advances in the area of multidimensional modeling following the implementation of a relational or multidimensional structure, Rizzi et al (2006), point out that future research should address the challenges of semantic gaps between advanced conceptual models and data cube implementation structures, in addition to the challenge transforming conceptual ETL schemata into logical models (Rizzi et al, 2006, p. 5).

Furthermore, different data warehouse design and implementation methods were found in the literature providing great insight about architectural options (Sen & Sinha, 2005; Gutiérrez & Marotta, 2000; Conn, S., 2005; Karayannidis, N., Vassiliadis, P., Tsois, A., & Sellis, T., 2001). The work of Sen and Sinha (2005), presented a discussion about the different data warehouse architectures (Enterprise, Data Mart, Hub-and-spoke Data Mart, Enterprise with Operational Data Store, and Distributed Data Warehouse) along with a visual representation of each of them. Among the many contributions made by Sen and Sinha, the discussion about the different data warehouse architectures deserves the attention of the designer interested in a blueprint conducive for "communication, planning, maintenance, learning, and resuse." (Sen & Sinha, 2005, p. 80)

Also, Sen and Sinha (2005), briefly discuss data warehouse implementation strategies by contrasting Inmon's and Kimball et al approach (Sen & Sinha, 2005, p. 81).

According to Sen and Sinha, Oracle and Microsoft SQL Server core technology vendor-based data warehouse design methodologies, support a data warehouse architecture based on data marts, and enterprise data warehouse (Sen & Sinha, 2005, p. 81). Both, Oracle and Microsoft SQL Server have been used in the literature as implementation platforms for healthcare and non-healthcare related data warehouses (Verma & Harper, 2001; Dell'Aquila, C., Di Tria, F., Lefons, E., & Tangorra, F., 2008).

Finally, several papers in the body of literature have addressed the issues of data warehouse quality, data quality, and data warehouse testing by presenting different approaches relevant to this study (Marotta, A., Piedrabuena, F., & Abelló, A., 2006; Serrano, M., Trujillo, J., Coral, C., & Piattini, M., 2007). However, the work of Golfarelli and Rizzi (Golfarelli, M., & Rizzi, S., 2009) presented a comprehensive approach to data mart testing relevant to this study, characterized by the following features:

a. Focus testing on design phase to “reduce the impact of error correction.”

b. Data mart testing activities classified in terms of what is tested and how it is tested.

c. Adoption of a reference design methodology that relates tightly to the proposed testing activities.

d. Aims to relate testing activities to quality metrics to allow for quantitative assessment.

According to Golfarelli and Rizzi (2009), testing design quality is almost as important as testing data quality. Golfarelli and Rizzi (2009) suggest that the meaning of design quality testing is that user requirements are well represented by the conceptual schema, and that both the conceptual and logical schemata are well built. In Table 3, an adaptation of Golfarelli and Rizzi (2009) is presented based on the "What vs. how in testing" approach used to illustrate the testing process of the design phases relevant to this study (*).

Table 3 – Adapted Golfarelli and Rizzi "What vs. how in testing" table

How	What: Analysis & Design			Implementation		
	*Conceptual Schema	*Logical Schema	*Physical Schema	ETL Procedures	Database	Front-end
Functional	R	R	R	R		R
Usability	R	R	R			R
Performance		R	R		R	R
Stress			R		R	R
Recovery			R		R	
Security			R		R	R
Regression	R	R	R		R	R

R (Required) - Test applies to specific design model, construct or artifact;
(Golfarelli & Rizzi, 2009, p. 20)

According to Golfarelli and Rizzi (2009), verifiable and quantifiable conditions must be available for passing each testing activity. The definition of specific metrics for a data warehouse test should be based on the following main phases:

a. Identify metric goals and quality criteria.
b. Adopt a formal definition of the metrics.
c. Theoretically validate the metrics and assess the metrics correctness and applicability.
d. Empirically validate the metrics. Understand the metrics.
e. Evolve metrics definitions and thresholds to adapt to new projects.

Golfarelli and Rizzi (2009), describe in detail each of the testing phases and point out what and how each phase should be tested (Golfarelli & Rizzi, 2009). Finally, Table 4 lists an adaptation of the testing activities proposed by Golfarelli and Rizzi (2009) that were further used to evaluate the solution artifacts (Golfarelli & Rizzi, 2009, p. 20-24; Hevner et al, 2004).

Table 4 – Adaptation of Golfarelli & Rizzi Testing Activities

		What	How
1	Conceptual Schema Test		
	Fact test	Test for user requirements	Verify that requirements are supported by conceptual schema
	Conformity test	Assesses “how well conformed hierarchies have been designed.”	Test all data mart dimensions
2	Logical Schema Test		
	Star test	Verify the correct formulation of SQL queries against preliminary workload; Priority should be given to: many-to-many associations or cross-dimensional attributes; complex aggregation schemes; and non-standard temporal scenarios	Verify through SQL queries
	Usability	Test fact and dimension tables	Verify through SQL queries; Could use simple metrics based on the number of fact and dimension tables to “capture schema understandability”
3	ETL Procedures	Tests that “ETL procedures extract, clean, transform, and load the data.”	Test for data loading correctness, hierarchies management, use of correct aggregations
4	Database	Aims at “checking the database performances using either standard (performance test) or heavy (stress test workloads).”	Performance – response time Data security
5	Front-End	Check correctness and usability of reports	Compare reports, involve users on test process

(Golfarelli & Rizzi, 2009, p. 20-24)

2.4 Literature Comparison

The relevant work on data warehousing and OLAP systems found in the body of literature related to both the healthcare and non-healthcare domains, present existing research challenges, design approaches, and a strong emphasis on the need to develop new modeling methods for complex data, quality, and security (Rizzi et al, 2006; Wang et al, 2005).

The literature analysis suggests that healthcare data warehouse design methods are behind compared to industries outside of healthcare (Parmento et al, 2). However, Pedersen and Jensen (1998) among other researchers, have extensively researched, evaluated, and contributed to the work of body specific to the clinical domain in order to help clinicians their data for “quality improvement and research.” (Pedersen and Jensen, 1998; Pedersen & Jensen, 1999; Pedersen & Jensen, 2001)

Bréant et al (2005), suggests that healthcare data warehouse design must provide an adequate level of grain in order to make possible the analysis of complex clinical data (Bréant et al, 2005, p. 170).

Furthermore, the healthcare literature suggests that there are mature commercial data warehouse solutions specific to healthcare available (Akhtar et al, 2005), (Ledbetter & Morgan, 2001) and (Pedersen et al, 1998). According to Akhtar et al (2005), participants in their study identified seven motivating factors that would lead to favor commercial products over in-house developed data warehouses. Akhtar et al (2005) point out that these motivating factors reflect the preference shown by the healthcare organizations based on the "perceived benefits", including cost, scope, and flexibility among others (Akhtar et al, 2005, p. 23).

The use of data marts as an architectural approach to support integration and evolution of the data warehouse design is supported in both the healthcare and non-healthcare literature. The use of data marts provides a flexible approach to the integration of new data marts into the enterprise level data warehouse schema. In addition to data marts, the literature presents the use of conformed dimensions as a way to share dimension tables among different fact tables providing a way to link dimensionally related data marts (Bréant et al, 2005, p. 174). The works of Sahama and Croll (2007), and Bréant et al (2005) make use of data marts for their particular data warehouse architectures to leverage data from different departmental perspectives, and to allow for future integration of other sets of data for analysis (Sahama & Croll, 2007), (Bréant et al, 2005). Also, Bréant et al (2005) data warehouse design is based on several data marts inter connected through conformed dimensions tables with each fact table describing patient medical data like encounters, laboratory results, diagnoses, and procedures (Bréant et al, 2005). This approach is also known as a fact constellation where the dimensional model is comprised of a more than one fact table sharing one or more conformed dimension tables (Connolly & Begg, 2004). Palaniappan and Sook Ling (2008) made use of the same design approach for their clinical decision support system using OLAP and data mining (Palaniappan & Sook Ling, 2008).

Ledbetter et al (2001), suggests that to leverage existing clinical data, the clinical data warehouse design should be focused on "data useful for retrospective aggregate analysis." (Ledbetter et al, 2001). However, Conn (2005), points out that "the importance of generating real-time business intelligence is that it is a building block to achieve better business process management and true business process optimization." (Conn, S., 2005, p. 515)

Vassiliadis and Sellis (1999), compare different multidimensional data cube models for OLAP applications by categorizing the research work as commercial tools and academic efforts. Vassiliadis and Sellis (1999), further divide the academic efforts category into two classes, the relational extensions and the cube oriented approaches to OLAP logical modeling. Vassiliadis and Sellis (1999) comparison of the various cube models contributes to the field of study by providing a survey, comparison, and summary of OLAP of the different multidimensional models available in the research body to understand the related terminology and semantics. Vassiliadis and Sellis (1999), also elaborate on some of the commercial products and technologies available like ROLAP and MOLAP architectures, which are of particular interest and align with the rest of the body of literature (Vassiliadis and Sellis, 1999).

The ROLAP architecture is just a multidimensional interface to relational data. On the other hand, the ROLAP architecture has two advantages: (a) it can be easily integrated into other existing relational information systems, and (b) relational data can be stored more efficiently than multidimensional data.

According to Tomic (2006), Relational OLAP (ROLAP) accesses data stored in a relational data warehouse to allow for data analyses based on the OLAP capabilities provided directly to the relational database, i.e. data warehouse. Tomic (2006) points out that ROLAP "is a three tier, client/server architecture where a database layer utilizes relational databases for data storage, access and retrieval processes." Furthermore, Tomic

(2006) indicates that the application logic layer is the ROLAP engine, which executes the multidimensional reports from multiple end users. Finally, Tomic (2006) points out that the ROLAP engine integrates with a variety of presentation layers, through which users perform OLAP analyses. (Tomic, 2006)

The literature also supports the idea of using OLAP tools as a way to empower knowledge workers with the tools they need to perform analysis and presentation of data at different levels of abstraction (Tremblay et al, 2007), which in some cases is more useful when integrated into larger repository of health information (Gordon & Asplin, 2004). The student health center could realize the benefits of integrating different data marts as an enterprise data warehouse to analyze different perspectives of patient care.

Dell'Aquila, Di Tria, Lefons, and Tangorra (2008), present a case study focused on the evaluation of known commercial business intelligence tools based on an existing criteria introduced by the Gartner group using the function point metric methodology for analysis of the features characterizing the application (Dell'Aquila et al, 2008). The experimental results analysis conducted by Dell’Aquila et al (2008) to analyze the functional complexity of Microsoft SQL Server 2005, Oracle Discoverer, and Microstrategy business intelligence platforms, show that the Microstrategy’s business intelligence platform has high functional complexity due to its object oriented design (Dell'Aquila et al, 2008). This study in particular is relevant to this project in that it provides an analysis of three known business intelligence platforms and their functional complexity. Furthermore, the work of Dell'Aquila et al (2008) provides an applicable example of the functional aspects characterizing Business Intelligence tools. This study adapts the functional complexity score table developed by Dell’Aquila et al (2008) to present the prototypes evaluation results (Dell'Aquila et al, 2008, p. 620). Table 11 and Table 12 show the artifact evaluation results.

2.5 Literature Findings and Conclusions

The comprehensive literature analysis presented several design theories and implementation methodologies relevant to this project. The literature offered different methods for the definition of system requirements from data and user perspectives (Rizzi et al, 2006), practical conceptual modeling techniques (Hüsemann et al, 2000), considerations for the development of logical design (Peralta & Ruggia, 2003), and the development of the physical design of the data warehouse and OLAP tools solution (Sen & Sinha, 2005), (Sahama & Croll, 2007).

The conceptual design methodology proposed by Hüsemann et al (2000), offered a practical approach for the development of the conceptual models for each solution artifact, and the graphical representation of attributes, dimensions, and facts. Rizzi et al (2006), suggests that conceptual modeling is fundamental for the construction of a database or in this case a data warehouse that is "well-documented and fully satisfies the user requirements" (Rizzi et al, 2006).

The work of Bréant et al (2005), makes use of a star schema data model showing three key design aspects, granularity, database model and architecture, and lifecycle of the data base (scalability) (Bréant et al, 2005). The design presented by Bréant et al (2005), uses an elementary fact to standardize the instantiated facts, inter connected through conformed dimension tables. The conformed tables used in this implementation are for patient, episode of care, medical services, and medical unit dimensions connected to four other fact tables. The design work presented by Bréant et al (2005), provides relevant design considerations applicable to the design of the solution artifacts required in this project.

Palaniappan and Sook Ling (2008), present a similar design for diabetes, heart, and liver disorder training databases using a conformed patient dimension and a conformed time dimension (Palaniappan & Sook Ling, 2008). These two approaches can be easily adapted to the design of a solution for the student health center project. Specifically the artifacts designed and built by Bréant et al (2005), and Palaniappan and Sook Ling (2008), provide logical and physical structures that match some of the requirements addressed in this study. Also, the discussions presented by both Bréant et al (2005), and Palaniappan and Sook Ling (2008) help us understand the use of the artifact in the context of health care (Bréant et al, 2005), (Palaniappan & Sook Ling, 2008).

The use of data marts in some of the works presented in the literature offered great flexibility when dealing with different departmental business facts of measure. This is clearly presented in the works of Sen and Sinha (2005), Sahama and Croll (2007), and the work of Bréant et al (2005), (Sen & Sinha, 2005), (Sahama & Croll, 2007), (Bréant et al, 2005). The healthcare related literature presented several data warehouse implementations based on SAS, and Oracle commercial data warehouse and business intelligence products (Sahama & Croll, 2007).

In regards to OLAP tools, Oracle and Microsoft SQL Server were presented in the literature as two vendor-based data warehousing and OLAP reporting solution products commonly used in all industries and business areas (Sen & Sinha, 2005), (Verma & Herpa, 2001), (Dell'Aquila et al, 2008). Clearly these are two well known business intelligence platforms that are commonly found in academic environments.

Dell'Aquila et al (2008) examine several business intelligence platforms, specifically Microsoft SQL Server 2005, Oracle Discoverer, and Microstrategy using a software measurement method designed to analyze functional complexity. The results of the study show that Microstrategy's business intelligence platform has high functional complexity due to its object oriented design. Dell'Aquila et al (2008) present the final experimental results from a study comparing three leading business intelligence tools based on information delivery, integration, and analysis capabilities.

While the experimental results show that Microstrategy allows for the execution of more complex tasks due to its object oriented nature, the comparison of the other two products provides significant insight about the capabilities of both the Oracle and SQL Server BI platforms relevant to designers.

Dell'Aquila et al (2008) work provides a relevant example of the evaluation of functional aspects characterizing Business Intelligence tools. This study adapts elements from the work of Dell'Aquila et al (2008) and Gorla (2003) to present the student health center artifacts evaluation results. Table 11 shows the artifact evaluation results (Part 1) based on: utility, ease of use, accuracy, simplicity of structure, and relevant output. Table 12 shows the artifact evaluation results (Part 2) based on report generation and ad-hoc querying functionality (Dell'Aquila et al, 2008; Gorla, 2003).

Chapter 3 – Methodology

This study follows the design-science methodology for developing knowledge through the design, building and evaluating of an artifact intended to solve the problem or issue addressed in the problem statement, drawing base theories from an existing body of knowledge (Hevner & March, 2004).

This design based research project, seeks to follow the iterative design-science research process to guide the artifact development and evaluation process. This research study adapts the design-science research methodology model process proposed by Peffers et al (2007), to establish the artifact prototype build and evaluation process. Peffers et al (2007) developed their design-science model from a synthesis of various design science research works developed based on congruencies among the different approaches. (Peffers et al, 2007, p.52).

As pointed out by Henver et al (2004), “design is inherently an iterative and incremental activity,” the objective of the iterative process is to rigorously evaluate the artifact and to provide feedback about its quality to the construction phase. According to Hevner et al (2004), the designed artifact is considered complete and effective when it satisfies the requirements of the problem it was intended to address at the problem awareness phase (Hevner et al, 2004, p.85).

Hevner et al (2004), point out that the artifacts are innovations purposely used to “define ideas, practices, technical capabilities, and products” used to analyze, design, and implement information systems, instead of a “full-grown” information system product of a project. (Hevner et al, 2004, p.83).

This study is not seeking to develop a complete information system solution, but to build only a prototype as a proof of concept seeking to demonstrate the potential benefits of data warehousing and OLAP reporting to improve existing decision making processes ultimately leading to improvements in patient care services.

The design-science research guidelines presented by Hevner et al (2004) were useful in the literature review and analysis process, and helped guide the requirements definition process. According to Hevner et al (2004), the guidelines are established to help the researcher “understand the requirements for effective design-science research.” (Hevner et al, 2004, p. 82)

The design science research guidelines provide the IT thesis researcher with the criteria that should lead the project process to produce an artifact designed to address the stated problem, produce a relevant solution through a "rigorously" evaluated process (Peffers et al, 2007).

The focus of the literature review in this study is on identifying methodologies and rigorous methods for both construction and evaluation of data warehouse and OLAP implementations. This study will be made tractable and manageable by breaking down the problem statement into three independent research sub-problems:

a. What design aspects can be derived from the literature and the different data warehouse design theoretical perspectives to help us develop a rigorous approach for the student health center data warehouse prototype?
b. Can the resulting prototype and the OLAP reporting tool answer the challenge questions presented earlier in the study?

c. How feasible would be for the student health center to develop, and implement a data warehouse system from the prototype and design recommendations?

In order to answer the three independent research sub-problems three independent processes are executed. First, the literature review and analysis process is conducted. This process includes the review, analysis, and conclusions from the literature relevant to the research project. Table 1 shows the literature review strategy, and some of the relevant considerations included in the analysis process. Second, an adapted version of the design science research method (DSRM) proposed by Peffers et al (2007) is executed to produce the solution artifacts and to share knowledge relevant to the research community.

However, the DSRM is preceded by the analysis of all four challenge questions by briefly discussing purpose, motivation, object of analysis, key stakeholders, proposed value to stakeholders, and issues and/or limitations. The conclusions from the literature review and the analysis of the challenge questions, and the research study assumptions are used as an input to the design science process model adapted from Peffers et al (2007) (Peffers et al, 2007). Finally, a feasibility study is conducted to determine if it would be feasible for the student health center to implement a data warehouse from the proposed prototypes. A diagram illustrating the key processes executed in this study to answer the research sub-problems is shown in Figure 2.

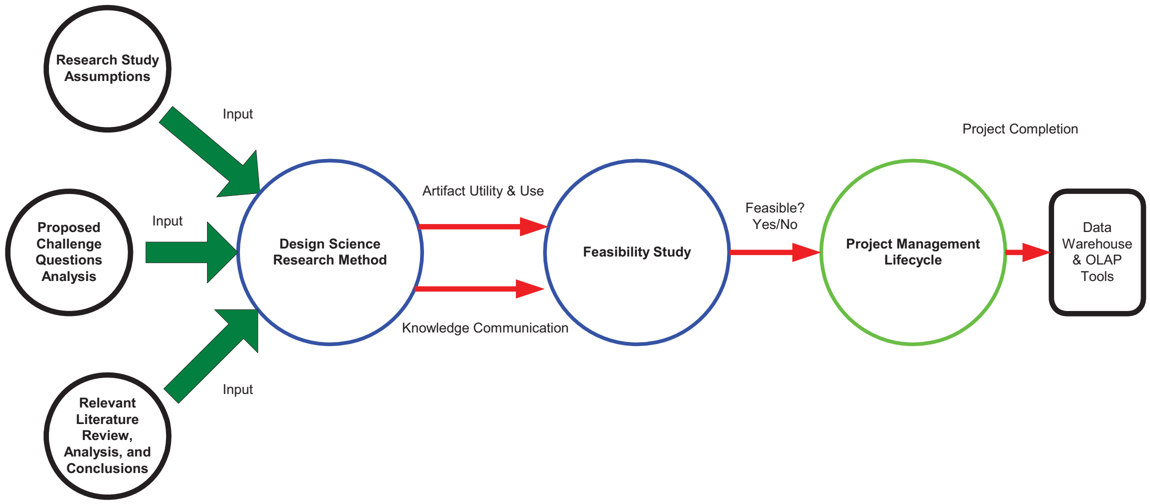


Figure 2 - SHC OLAP Reporting Project Processes Data Flow Diagram

Table 5 shows the work break down structure developed for this project.

Table 5 – Research Project Work Break Down structure

(HIGH LEVEL WBS)

Student Health Center Data Marts Prototypes & Feasibility Study

I. Preparation Stage

1. **Research Study Assumptions**
   1.1. The Student Health Center Organization
   1.2. Existing IT Infrastructure
   1.3. Information Needs
   1.4. Conclusions (Input to Design Science Process)
2. **Literature Review, Analysis, and Conclusions**
   2.1. Literature Gathering
   2.2. Literature Selection and Review
   2.3. Make Recommendations (Input to Design Science Research Process)
3. **Challenge Questions Analysis**
   3.1. Challenge Question Assessment
   3.2. Address Issues and/or Limitations
   3.3. Make Recommendations (Input to Design Science Research Process)

II. Design & Implementation Stage

1. **Design Science Research Process - Problem Identification**
   1.1. Define Research Problem
   1.2. Justify the value of the solution
2. **Design Science Research Process - Problem Solution Proposal**
   2.1. Define objectives of the solution
   2.2. Define new solution's efficacy
3. **Design Science Research Process - Design & Development**
   3.1. Define design specific assumptions
   3.2. Define and choose the healthcare process to model
   3.3. Define system requirements (data driven and user driven)
   3.4. Conceptual model development
   3.5. Develop logical design
   3.6. Develop physical design
4. **Design Science Research Process - Demonstration**
   4.1. Test artifact functionality (demonstration only - Use Microsoft Access 2003 for rapid prototype)
   4.2 Define how to use the artifact to solve the given problem
5. **Design Science Research Process - Evaluation**
   5.1. Test artifact functionality (Evaluate functionality)
   5.2 Document artifact test results
   5.3 Present conclusions from design

Table 5 – Research Project Work Break Down structure Continued

III. Feasibility & Study Closure Stage

1. **Feasibility Study**
   1.1. Review and Assessment of Existing Technical Infrastructure (Hypothetical Case)
   1.2. Review and Assessment of BI (Data Warehousing & OLAP) for the student health center
      1.2.1 Organizational Feasibility
      1.2.2 Technical Feasibility
      1.2.3 Operational Feasibility
   1.3. Recommendations (Architecture and Implementation options)

The design and development phase of the design science research process will be based on the works of Parmanto et al (Parmanto et al, 2005), Hüsemann et al (Hüsemann et al, 2000), and Rizzi et al (Rizzi et al, 2006). Parmanto et al (2005), present a data warehouse design approach for healthcare outcome research derived from Kimball and Ross's method (Parmanto et al, 2005). Hüsemann et al (2000), present a conceptual modeling approach compatible with traditional database design (Hüsemann et al, 2000). Finally, Rizzi et al (2006), present an overview of the state of research in data warehouse focused on modeling and design outstanding issues. Table 6 shows the tasks related to the artifact design approach derived from the works mentioned above.

Each artifact is designed based on the assumptions presented at the introduction of the study. Furthermore, each artifact design takes into consideration the design methodologies and/or approaches found in the literature. Finally, the artifact design is also based on the assumptions and/or conclusions resulting from the analysis of the challenge questions.

Table 6 - Design and Development Approach

	Design Process
3.1	**Design Assumptions**
3.1.1	Use research study assumptions.
3.1.2	Use relevant literature findings.
3.1.3	Use challenge questions assumptions.
3.2	**Define and choose the healthcare process to model**
2.1	Determine process to model from challenge questions analysis:
2.2	Determine if the literature presents any similar cases that match the challenge question.
2.3	Determine if the operational schema is representing any of the business processes in question
3.3	**Define Requirements (Data & User)**
3.3.1	Define data requirements: Analyze challenge questions, and related literature
3.3.2	Define user requirements: Consider assumptions, challenge questions analysis, and any other relevant literature
4	**Conceptual model development**
4.1	Context definition of measures (Hüsemann et al, 6-6)
4.2	Dimensional hierarchy design (Hüsemann et al, 6-6)
4.3	Definition of Summarizability Constraints (Hüsemann, et al, 6-10)
4.3	Conceptual model for proposed artifact
5	**Develop Logical designs**
5.1	Develop star schemas (Use Microsoft Access 2003)
6	**Physical design**
6.1	Use Microsoft Access 2003 to develop an initial instance of the artifact before moving to development on any of the platforms below.
6.2	Use both Microsoft SQL Server 2005 and Oracle 10g Enterprise Edition

This study uses the data warehouse design approach presented by Parmanto et al (2005) which was derived from Kimball and Ross's method (Parmanto et al, 2005). This study also makes use of the conceptual modeling approach presented by Hüsemann et al (2000) (Hüsemann et al, 2000).

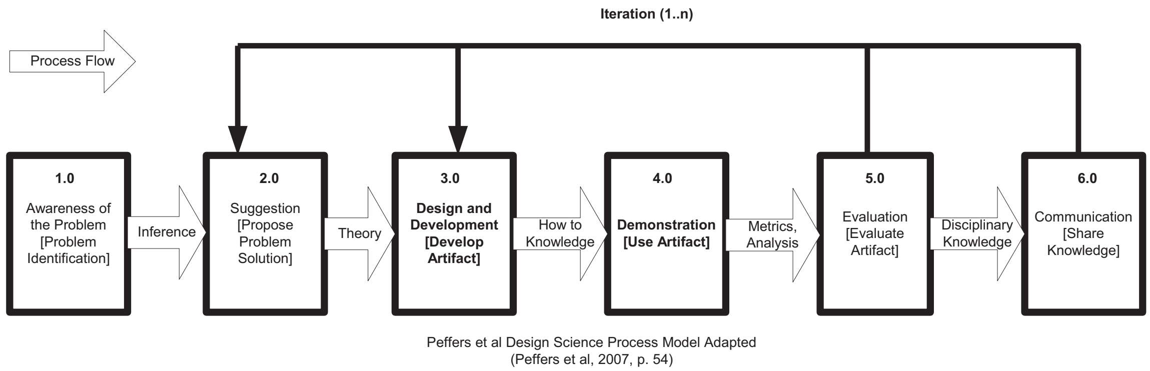


Figure 3 - Peffers et al (2007) Design Science Research Method Process Model

The process shown in Figure 3 is the result of Peffers et al (2007), analysis and synthesis of different perspectives on design science. The process consists of six activities in a nominal sequence as illustrated in Figure 3 including the sequence numbers added to suit the project needs. This study also makes use of Baskerville et al design science research risks evaluation framework – Adapted (Baskerville et al, 2008) as illustrated on Figure 4, integrated to Peffers et al (2007) nominal process model. According to Baskerville et al, there are six areas of potential risk in design science research. Baskerville et al, list the following areas:

- Business Needs (Problem Analysis and Choice)
- Applicable Knowledge (Retrieved from the Body of Recorded Human Knowledge)
- Develop/Build (Develop Theory/Knowledge and Build an Instantiation)
- Justify/Evaluate (Justify Theory/Knowledge and Evaluate an Instantiation)
- Applications (of Knowledge to Business and Organizational Problem Situations)
- Additions (to the Knowledge Base of Recorded Human Knowledge)

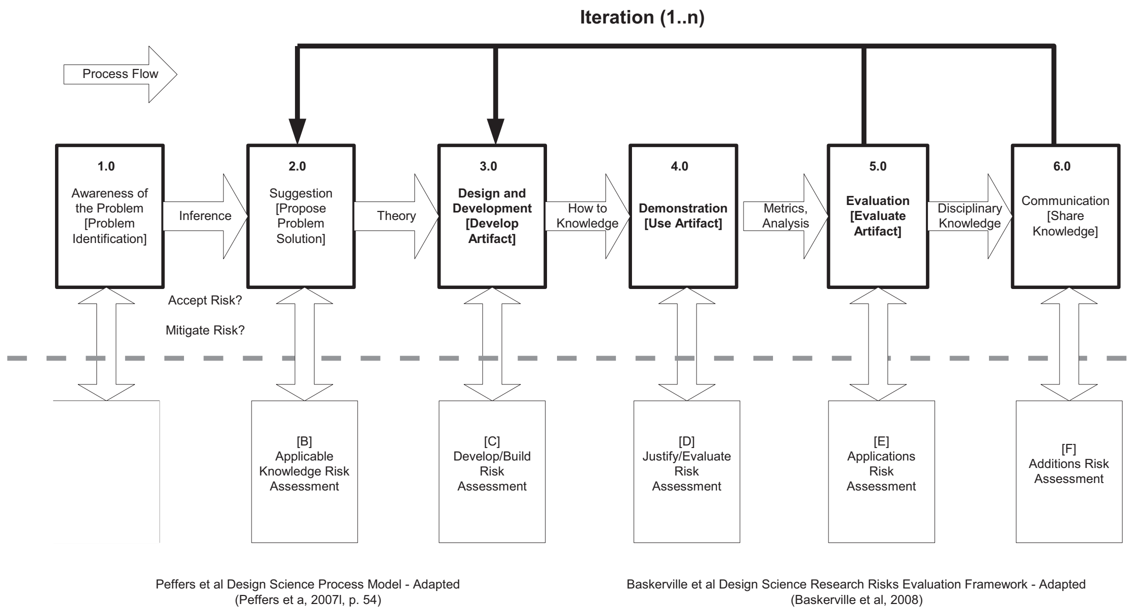


Figure 4 - DSRM Process Model with Integrated Risk Management Framework

Baskerville et al (2008), provide an example of how their risk assessment was used for a project, and identify the top five risks along with consequence, probability, and treatment. The risk assessment is applied to each of the design science research processes, in an effort to assess and mitigate the most significant risks.

Chapter 4 – Project Analysis and Results

In this chapter, the use of the design science research process methodology (DSRM) proposed by Peffers et al (2007) is presented, to describe the design, implementation, and evaluation of the solution prototypes. Secondly, the design-science research risk management framework presented by Baskerville et al (2008) is used to help identifying potential risks resulting from the design-science research methodology process (Baskerville et al, 2008). Finally, an assessment of the feasibility of implementing OLAP reporting at the student health center is presented including several implementation options, recommendations, and concluding remarks.

In this study, the artifact design process was influenced by the contributions listed on Appendix A showing the "rigorous methods" most suitable for the construction and evaluation of the solution artifacts (Hevner et al, 2004, p. 87). The different theories and/or methodologies resulting from the literature analysis and summarized on Appendix A, influenced the design decisions of each of the solution artifacts. The summary of the research contributions is what design science research considers "rigorous methods" resulting from the different data warehouse design and implementation theories found in the literature (sub-problem question #1). In addition to the "rigorous methods", recommendations from the challenge question analysis are also used to influence each design. Appendix B, C, D, and E show the analysis of the four challenges questions including a subset of the data from the operational systems required to develop the answer to the respective question, and comments about the possible value to the stakeholder. Finally, a summary of the artifact evaluation results is shown in Table 11 and Table 12.

4.1 Challenge Question #1 – Health Education Programs

Problem-Centered Approach:

Due to the many cases of alcohol abuse among students (Ehrlich et al, 2006), the effectiveness of the health education programs can be questioned. This study argues that historical data about the health education programs like the alcohol abuse course could be leveraged if is analyzed in combination with student's academic performance. This particular challenge question is concerned with the availability of data to show the level of influence exerted by the health care education programs on the student's academic performance, and student health habits.

Problem Identification and Motivation:

The problem presented in this section is: what percentage of students showed GPA improvement after the successful completion of the alcohol abuse program? The challenge described here is how to find evidence of GPA improvement after the successful completion of the alcohol abuse program in order to provide some auditable evidence of the efficacy of the health education programs. This question is motivated by the idea that linking student academic performance data to health education programs data could serve as a monitoring tool for measuring program quality and/or effectiveness. The conceptual aim of the problem presented here is that the availability of supportive evidence in favor of the effectiveness of the health education programs should translate into further support from the University administration.

Objective of the Solution:

The objective of the solution to the hypothetical question is to develop a tool for trend analysis on health education courses data. The tool should help in the development of reports showing the relationship between health education programs and student academic performance (GPA).

Design and Development:

The artifact designed to answer this particular challenge question required considerable analysis of the approach and it took several design iterations to eventually arrive to a useful, realistic, and effective final conceptual, logical and physical model comprised of one fact table including measures related to health education courses evaluation surveys used to aggregate values for trend analysis. Table 7 lists the attributes required for the model and Appendix G lists the functional dependencies between terminal dimensions and measures, and the summarizability for the fact schema. The conceptual model representing the health care education programs is shown in Figure 5. Finally, the resulting multidimensional structure is shown in Figure 6 based on a star schema. The education course related measures are obtained through the use of course evaluation questionnaires to gather information about student's expectations, course rating, and level of influence on health habits and academic performance. Several of the health related research studies analyzed in this project helped to identify the use of surveys as a mechanism to obtain feedback from the patient population about the quality of service and service effectiveness (Wall et al, 2004; Erlich et al, 2006; Eilers, G., 2004). As a result of the effective utilization of surveys in those studies, the design presented here relies on the availability of health education course evaluation data targeting specific areas like student academic performance and student health habits.

Demonstration:

This study used the demonstration process to determine some level of utility to demonstrate the utility of the health education programs artifact a physical model was developed in Microsoft Access 2003 to test some of the queries and to determine if changes or modifications to the current design were necessary. As stated before several design iterations took place leading to new versions of the artifact.

Evaluation:

The evaluation of the health education program performance artifact was intended to demonstrate its utility, quality, and efficacy. The utility of the artifact was demonstrated by generating reports (See Appendix K) showing data relevant to answer the challenge question (What percentage of students showed GPA improvement after the successful completion of the alcohol abuse program?). The quality of the artifact was evaluated based on the ease of use, and simplicity of design. The efficacy of the artifact was demonstrated by the use of the report to provide evidence of a connection between the health education program and the student academic performance. The report generated by the artifact showed the alcohol abuse program, academic year, and number of students claiming academic performance improvement after the completion of the course.

A better way to evaluate the artifact would've been to show the stakeholders how to use the tool and to involve them in the evaluation of the results. However, that was not the case. Therefore, the artifact evaluation in this study was conducted from a utility and functionality perspective. The works of Gorla and Dell'Aquila et al (Gorla, 2003; Dell'Aquila et al, 2008) contributed to the design of the evaluation process used in this study, and helped develop a tabular summary of the evaluation results as shown in Table 11 and 12.

Communication:

Several aspects of the design and evaluation of the health education program artifact can be shared with stakeholders and/or researchers. In order to report the number of cases of academic performance improvement after the successful completion of the alcohol abuse course, the design had to make use of a course evaluation survey as a data collection mechanism. The use of the course evaluation survey allows for the volunteer disclosure of the academic performance improvement data needed.

However, is important to share knowledge about any possible limitations of the proposed design. One of the limitations to consider is that the use of a simple yes or no field to acknowledge GPA improvement after the completion of the alcohol abuse course might not be a reliable way to acquire the necessary data to correlate course completion and academic performance.

Another important factor to consider is the limitation imposed by the time it takes to process the collected course evaluation data, and the availability of the data for analysis. Therefore, the stakeholders will need to analyze the proposed design and/or find a more effective way to provide the data required for academic performance improvement analysis based on successful course completion.

Contribution:

The health education programs data mart was primarily evaluated on the capability of report generation. The challenge question health education courses OLAP cube for course performance linked to student academic performance through voluntary disclosure survey. According to Hevner et al (2004), the design science research guideline 3 requires that the utility, quality, and efficacy of a design artifact be rigorously demonstrated via well executed evaluation methods. This artifact was used effectively to collect data in a consistent and automated fashion from disparate local health-care organizations. The main contribution of this design is the integration between health education course and student academic performance data.

Table 7 - Artifact #1 - Requirement Specification Attributes

Requirements Specification for Health Education Programs Performance Cube - DSQ105

Attribute	Description	M	D	O
TimeID	Time key	no	yes	no
Month	time aggregation	no	yes	no
Year	time aggregation	no	yes	no
TermID	Term key	no	yes	no
TermCode	Term code	no	yes	yes
TermStart	Term start date	no	yes	no
TermEnd	Term end date	no	yes	no
AcademicYear	Academic year (Time aggregation)	no	yes	no
HealthEdProgID	Health education program key	no	yes	no
HealthEdProgName	Health education program name	no	yes	no
HEdProgSession	Health education program session	no	yes	no
HEdProgStartDate	Health education program start date	no	yes	no
HEdProgEndDate	Health education program end date	no	yes	no
HEdProgDescription	Health education program description	no	yes	yes
HEdEnrollmentStatus	Health education program status	yes	no	no
HEdEnrollmentJustification	Enrollment justification	yes	no	no
HEdCourseEvaluationFrmCompletion	Completion of student evaluation form	yes	no	no
ImprovedAcademicPerformance	Did the student achieve better academic performance after completing the health education course?	yes	no	no
ImprovedHealthHabits	Did the student health habits change?	yes	no	no
StudentID	Student key	no	yes	no
StudentEnrollmentStatus	Student enrollment status flag	no	yes	no
StudentCountry	Student country of origin	no	yes	no
StudentGender	Student sex	no	yes	no

Description and categorization of relevant attributes as measure, dimension, or optional attribute. According to Hüsemann et al (2000), an optional property attribute (O) is an attribute that does not have to be specified for each element of the corresponding dimension level and therefore may contain NULL values.

(Hüsemann, et al, 2000, p. 6-7)

Figure 5 - Health Ed. Programs Performance Conceptual Schema

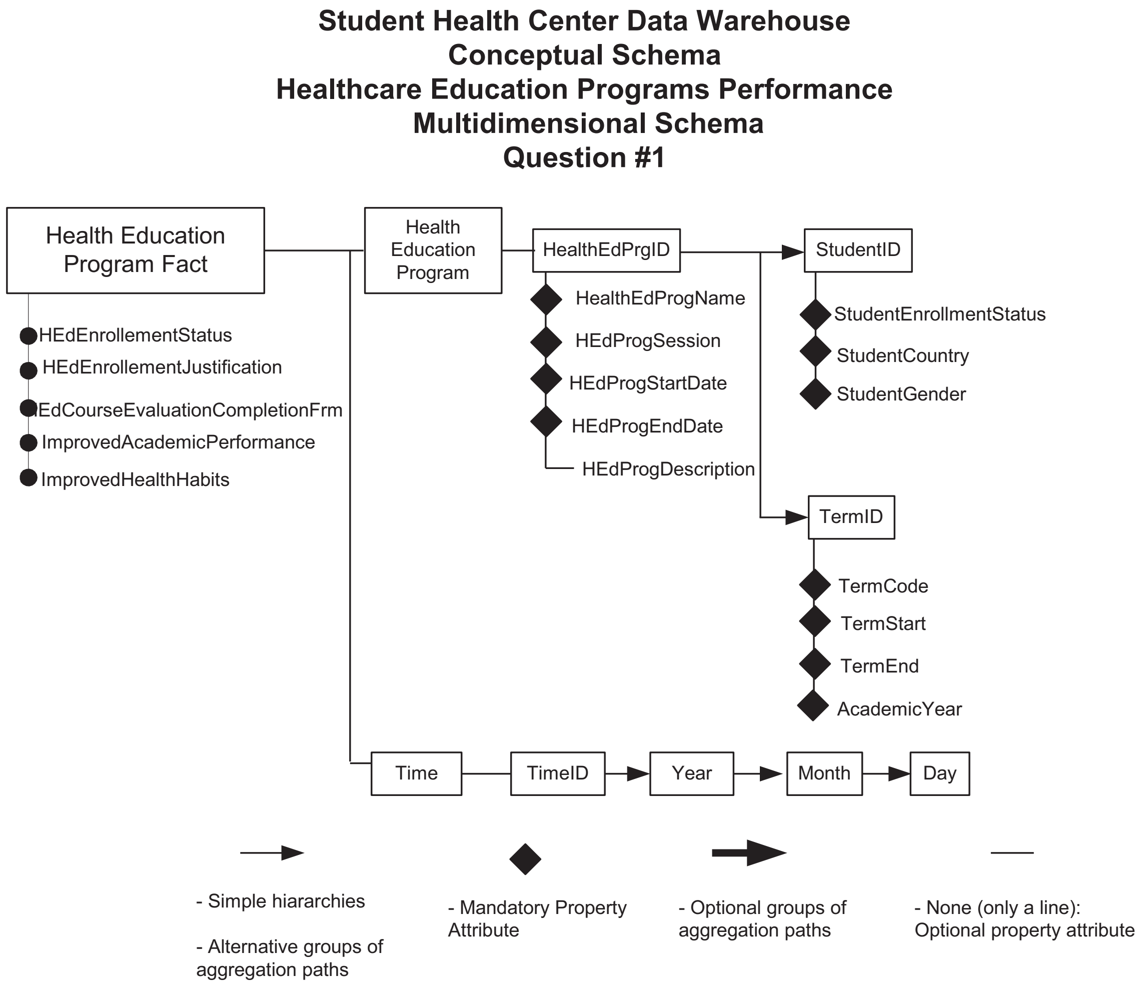
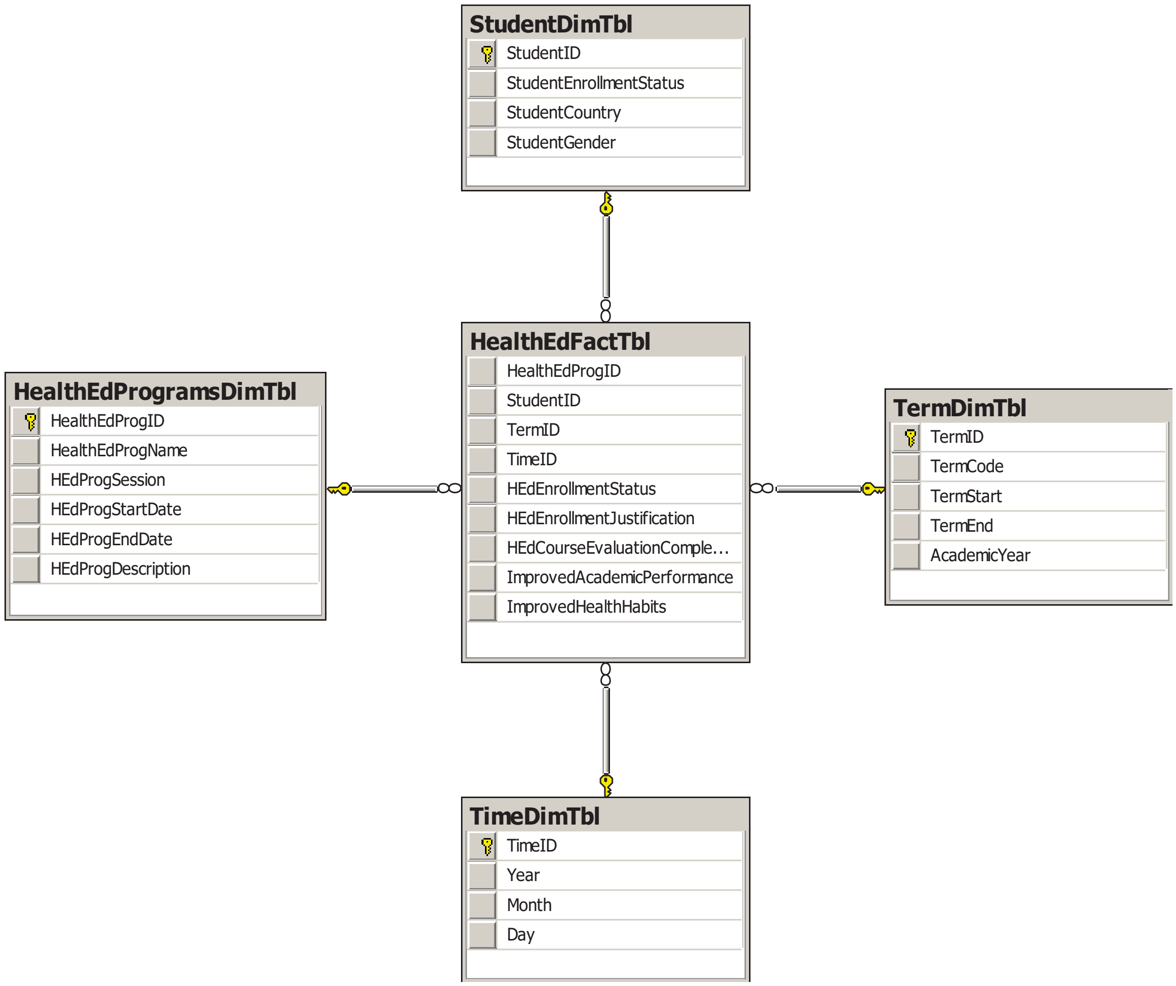


Figure 6 - Health Ed. Programs Logical Design



4.2 Challenge Question #2 – Immunizations Compliance

Problem Centered Approach:

The student health center is responsible for tracking student compliance with immunization requirements as mandated by the respective state's health department. A historical view on immunization data could help management and/or clinical staff in identifying trends that can potentially lead to service improvements. Specifically, the analysis of immunization exceptions based on country of origin can be used to improve areas of focus for preventive treatment in order to manage the risk of pandemic.

Problem Identification and Motivation:

The problem addressed by this challenge question is the ability to show the number of international students that have failed to present proof of vaccinations for the current and previous years. The lack of data showing trends on international student immunization requirements compliance does not contribute to the overall monitoring of pandemic risk. This study argues, that immunization requirements compliance data should be analyzed to reveal trends that could ultimately impact the student population. This hypothetical question is motivated by the realization of the potential risk of pandemic outbreak in any higher education institution in the United States due to the number of foreign students enrolled not meeting the vaccination requirements established by the institution.

Objective of the Solution:

The objective of the artifact is to provide a multidimensional schema for analysis of historical immunization data extracted from the electronic medical record system. The trend analysis on immunization compliance data can help management make decisions about the effectiveness and efficacy of the immunization program to better protect student health. Also, the use of trend analysis can provide evidence of the effectiveness of the immunization program.

Design and Development:

The literature analysis provided several examples of data marts using patient as a dimension, but none of the papers found addressed immunizations or vaccinations in any way. Table 8 shows the requirements specification attributes for the immunizations compliance multidimensional model. The object of analysis represented by this model is the vaccination compliance as shown in Figure 7 as the conceptual model.

The logical model shown in figure 8 represents the design of an artifact modeled as a fact schema comprised of a vaccination fact table including the vaccination compliance measure as the object of analysis, a vaccination dimension table with vaccination name as “mandatory property attribute”, time, patient, student, and the provider dimension all with their respective attributes (Hüsemann et al, 2000, p. 6-7). The grain of the vaccination fact table is vaccination requirement compliance (see star schema model figure 8).

Demonstration:

After developing the vaccination compliance conceptual and logical schemas, the physical model was developed using Microsoft Access 2003 as the proof-of-concept prototype. Several queries were developed and tested to demonstrate the artifact's utility to solve the problem. Appendix E shows the outputs of the immunization data mart queries.

Table 8 - Artifact #2 - Requirement Specification Attributes

Requirements Specification for Student Immunization Compliance Cube - DSQ201

Attribute	Description	M	D	O
TimeID	Time key	no	yes	no
Year	time aggregation	no	yes	no
Month	time aggregation	no	yes	no
Day	time aggregation	no	yes	no
VaccinationID	Vaccination key	no	yes	no
VaccineName	Name of vaccination	no	yes	no
VaccineDose	Vaccination dosage required	no	yes	no
VaccinatorName	Name of staff providing the vaccination	no	yes	yes
VaccinationRequired	Required vaccination flag	no	yes	no
VaccinationReqCompliance	Marks patient or student record as meeting all vaccination requirements (Yes/No)	yes	no	no
PatientID	Patient key	no	yes	no
PatientCountry	Patient country of origin	no	yes	no
PatientAge	Patient age	no	yes	no
PatientGender	Petient sex	no	yes	no
ProviderID	Providers key	no	yes	no
ProviderFName	Provider's first name	no	yes	yes
ProviderLName	Provider's last name	no	yes	yes
ProviderProfTitle	Provider's professional title	no	yes	yes
StudentID	Student key	no	yes	no
StudentEnrollmentStatus	Student enrollment status	no	yes	no
StudentCountry	Student country of origin	no	yes	no
StudentGender	Student sex	no	yes	No

Description and categorization of relevant attributes as measure, dimension, or optional attribute. According to Hüsemann et al (2000), an optional property attribute (O) is an attribute that does not have to be specified for each element of the corresponding dimension level and therefore may contain NULL values.

(Hüsemann, et al, 2000, p. 6-7)

Evaluation:

The evaluation of this artifact began by running a query to answer to the challenge question (See Appendix L). The first query was to calculate the number of students from China, Japan, Vietnam, and Korea that were in compliance with the vaccination requirements. Furthermore, a second query to find students not in compliance with the immunization requirements was executed successfully. The immunizations compliance artifact produced accurate answers to the ad-hoc queries generated using Microsoft Access 2003, Microsoft SQL Server 2005, and Oracle 11g database were also accurate. A summary of the evaluation process results for each of the artifacts is shown in Table 11 and 12.

Communication:

As stated in the analysis of the challenge question #2, the problem and the artifact presented here aims to demonstrate how leveraging historical student immunization data could benefit management and clinical staff on analyzing vaccination compliance trends. The utility and efficacy of this artifact should be communicated to providers in charge of administering vaccines, and those in charge of tracking immunization records. These stakeholders can help on implementing and/or improving the design of the proposed artifact. The knowledge generated from the stakeholder's contributions can be further communicated to researchers through research publications.

Contribution:

The immunizations compliance data mart can be used as a tool to help the immunizations department on analyzing compliance trends in order to reinforce population groups of the importance of meeting immunization requirements. The data mart serves as a simple tool that can be used to assess program effectiveness, and to report on immunization requirements compliance trends. The data mart presented here can be enhanced to include other relevant measures in addition to immunization compliance measure.

Figure 7 - Immunization Compliance Conceptual Schema

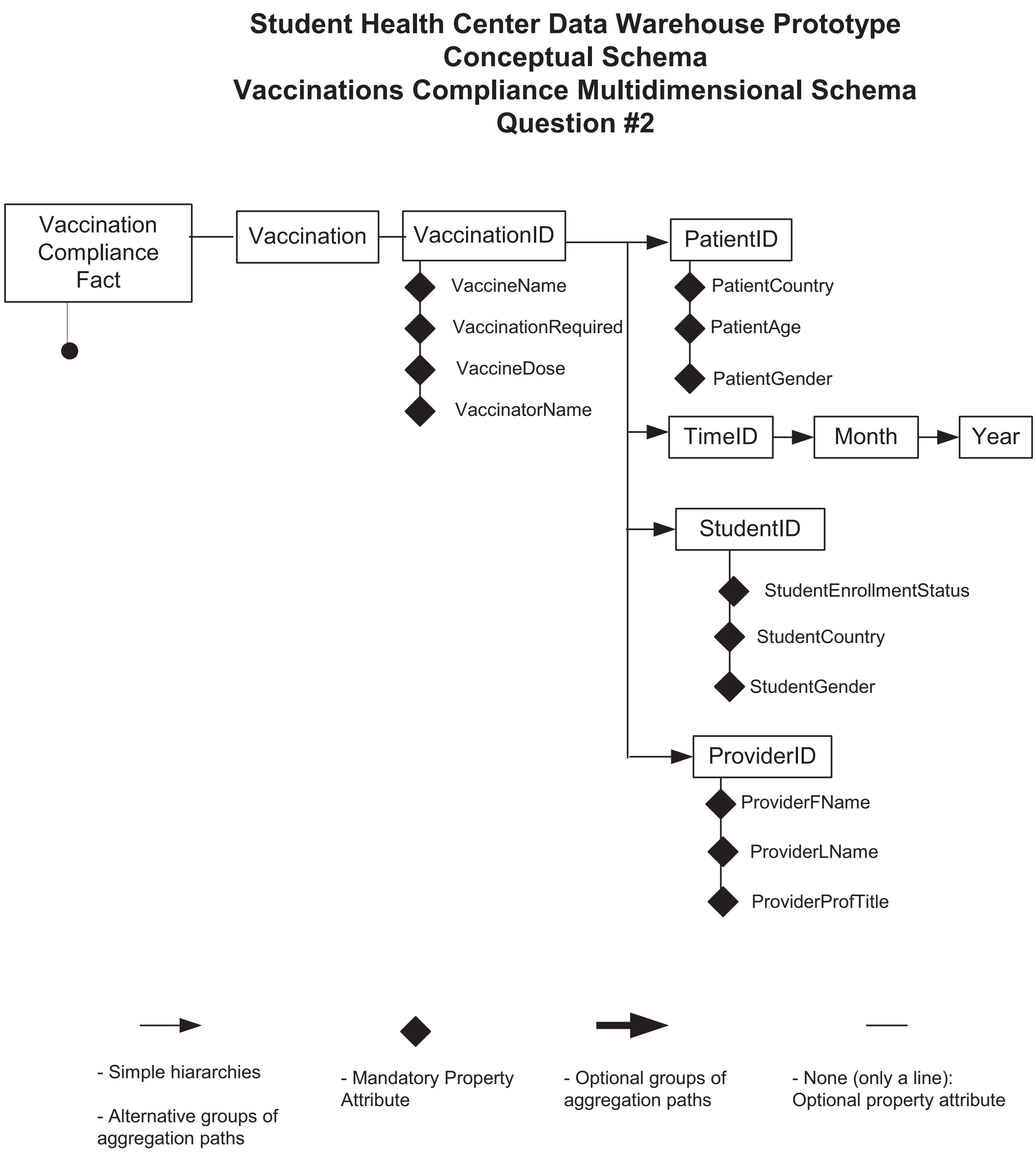
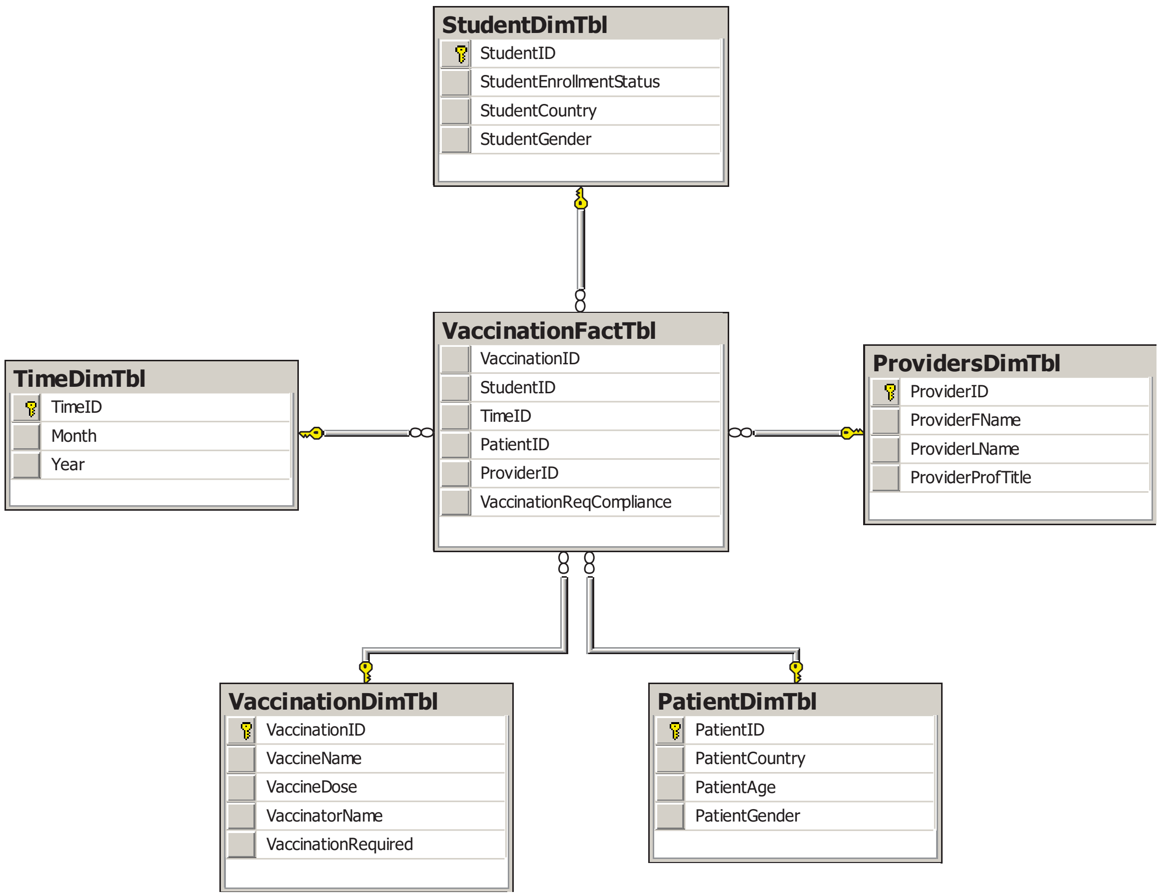


Figure 8 - Immunization Compliance Logical Design

Logical schema for the immunization compliance multidimensional model
(Microsoft SQL Server 2005 schema diagram)



4.3 Challenge Question #3: - Health Services Performance

Problem-Centered Approach:

Due to the increase on health care costs, the student health center administration is in need to make changes to the health services costs. However, relevant historical data is not readily available to provide evidence to help justifying necessary changes. Therefore, they have requested the provision of data related to the utilization of services and costs.

Problem Identification and Motivation:

In this hypothetical question, the student health center management is requesting data that supports the cost increase of health services, and data showing statistics of health services utilization. The idea motivating the challenge question is that the health care services historical data can be analyzed for trends on health care service utilization, and cost through time. The question is also motivated by the assumption that the student health center to often struggle to balance quality of service with the cost of operating a clinic. The need to allocate funds in support of services more frequently sought by students, and health services critical to student population health requires that the analysis of trends on healthcare service requests. The challenge posed by this question is the design of an artifact capable of expressing cost, revenue, and time data on health care services in a multidimensional structure. Also, the ability to easily retrieve cost and utilization data spanning a number of years will be critical to develop a solution to this problem.

Objective of the Solution:

Specifically, this artifact needs to be useful in showing health service utilization, generate reports showing cost versus demand, show health service utilization. The objective of the artifact is to motivate the discussion about how to utilize the tool for healthcare services performance assessment.

Design and Development:

This artifact provides a historical perspective on the number of visits (ServicedPatients), number of scheduled appointments (ServiceRequests), service revenue generated per month (ServiceRevenue), and the cost of rendering a particular health care service (ServiceOperCost). All of the previously mentioned measures are pre-calculations stored in the fact table. These measures can be used to contrast service revenue with service operating cost, and can also be contrasted with service utilization as a way to measure service performance (revenue and utilization) and effectiveness. The design makes use of a very basic cost and revenue structure used in order to keep design simple and flexible. Table 9 shows the requirement attributes for the health education programs performance multidimensional model showing the measures and dimensions used in the model. The functional dependencies between the fact table ServiceFactTbl, and the ServiceRevenue, ServiceOperCost, ServicedPatients, and ServiceRequests measures related to the terminal dimensions TimeID and ServiceID are shown in Appendix I. Definition of the restriction levels applicable to all measures identified as aggregation paths in the fact schema are also shown in Appendix I.

Figure 9 shows the conceptual model for the health care services conceptual schema, and Figure 10 shows the resulting logical model for the health service performance.

Demonstration:

The health services artifact utility was demonstrated by the execution of queries showing service name, aggregation of serviced patients, service cost, and service revenue based on year.

Evaluation:

The challenge question did not provide enough information to develop solid evaluation criteria for this artifact. The utility, quality, and efficacy of the artifact is demonstrated through the generation of a report and the ad-hoc queries executed against the model. The artifact showed relevant data useful for the analysis of health services showing pre-aggregated monthly totals for each health service showing cost, revenue, number of students serviced, female and male students serviced aggregations. A summary of the evaluation process results for each of the artifacts is shown in Table 11 and 12.

Communication:

The health services performance artifact demonstrated its utility by showing data relevant to health care services performance like: revenue, cost, serviced patients, and service requests. The presentation of this artifact is intended to encourage future design enhancements. However, the proposed design also carries a number of limitations worth noting.

Contributions:

The health services performance artifact makes use of pre-aggregated measures to contrast service utilization, and cost versus revenue providing data useful on making decisions about what services to continue or cancel. This hypothetical problem and solution can help management, physicians, and directors to contrast the actual number of patients serviced (ServicedPatients) with the number of service requests (ServiceRequests) to analyze why students were not serviced whether they walked in or scheduled an actual appointment.

Table 9 - Artifact #3 - Requirement Specification Attributes

Requirements Specification for Health Services Performance Cube - DSQ301

Attribute	Description	M	D	O
TimeID	Time key	no	yes	No
Month	time aggregation	no	yes	No
Year	time aggregation	no	yes	No
ServiceID	Service key	no	yes	No
ServiceName	Health service name	no	yes	No
ServiceDescription	Health service description	no	yes	yes
ServiceCategory	Health service category	no	yes	yes
ServiceRevenue	Health service revenue (monthly figure)	yes	no	No
ServiceCost	Health service operating cost (monthly figure)	yes	no	No
ServicedPatients	Health service number of patients serviced	yes	no	No
ServiceRequests	Health service number of requests for service or scheduled visits	yes	no	No

Description and categorization of relevant attributes as measure, dimension, or optional attribute. According to Hüsemann et al (2000), an optional property attribute (O) is an attribute that does not have to be specified for each element of the corresponding dimension level and therefore may contain NULL values.

(Hüsemann, et al, 2000, p. 6-7)

Figure 9 – Health Service Performance Conceptual Schema

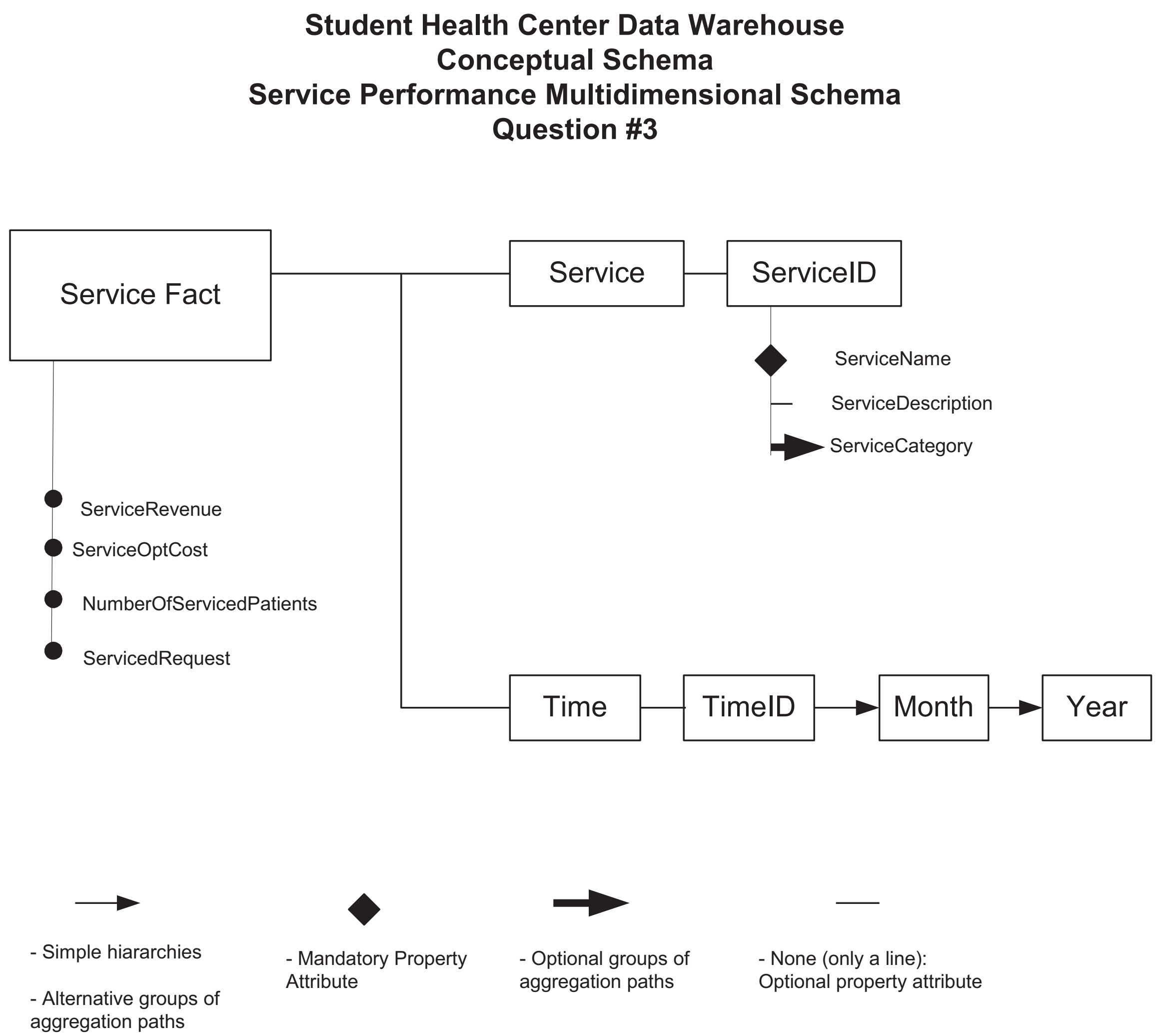
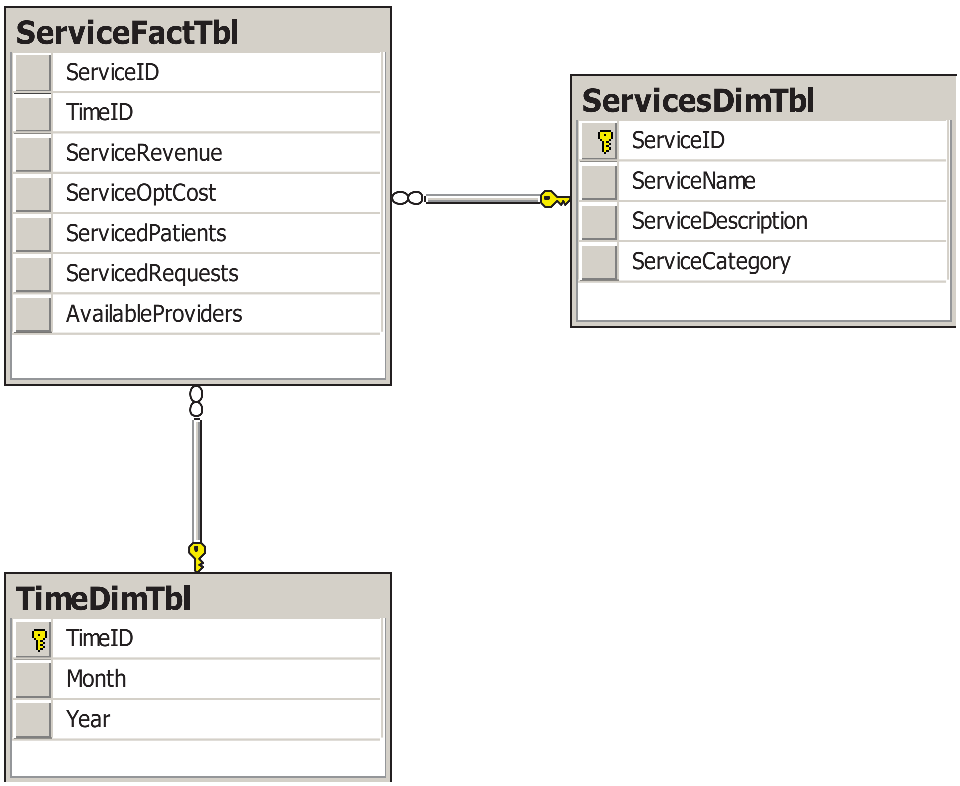


Figure 10 – Health Service Performance Logical Model

Logical Schema for health care services multidimensional model
(Microsoft SQL Server 2005 schema diagram)



4.4 Challenge Question #4: - Episode of Care Data Mart

Since patient care is the object of service, the student health center should be able to leverage existing patient data in support of decision making processes at different levels of the organization. The use of an episode of care for population disease-treatment outcomes analysis can provide a "mechanism to measure the effectiveness of care in treating the health problem." (Wall et al, 2004; Ehrlich et al, 2006). This study argues that the episode of care data mart would be an effective tool for physicians to analyze patient population based on diagnosis and/or treatment.

Problem-Centered Approach:

In this study, the need to improve the quality and effectiveness of clinical care at the student health center is used to motivate the multidimensional analysis of diagnosis and/or treatment data in order to understand episode of care outcomes on different student populations (Wall et al, 2004; Parmanto et al, 2004). The literature shows that episode of care can help in the analysis of healthcare related outcomes (Parmanto et al, 2004; Wall et al, 2004). However, the literature also shows that healthcare outcome analysis and/or the use of episode of care approaches for disease/treatment analysis are both complex in nature (Parmanto et al, 2004; Wall et al, 2004). How can the student health center leverage historical patient data related to a specific diagnosis and treatment episode? In this section the intent is to present a multidimensional model derived from the healthcare literature for the analysis of disease and treatment outcomes.

Problem Identification and Motivation:

In order to convince management and/or physicians of the benefits of leveraging clinical historical data, an episode of care data mart is used as a tool for outcome analysis. The problem identified here is the lack of automation on patient data analysis related to diagnosis and treatment episodes.

This initiative is motivated by the literature findings on the use of an outcome data warehouse for episodes of care for patient treatment (Parmanto et al, 2004), and the use of episodes of care for the analysis of alcoholism treatment episodes (Wall et al, 2004). Both studies point out the benefits of using an episode of care for outcome analysis.

Objective of the Solution:

The objective of the solution is to present the student health center physicians and clinical researchers with a multidimensional model in the form of a data mart to be used for analyzing episodes of treatment for a particular diagnosis.

Design and Development:

Two different proof-of-concept prototypes were developed for this challenge problem. Several different iterations from design to demonstration were necessary for each prototype version. The design presented by Parmanto et al (2005) was chosen to develop the initial prototype for the stakeholders. Also, the diagnosis dimension design proposed by Song et al (2001) was adopted since the primary diagnosis and the secondary diagnosis were in the same dimension table (Parmanto et al, 2005; Song et al, 2001). Table 10 shows the list of attributes required for the episode of care multidimensional model and whether the attribute is a dimension, measure, or an option. Figure 11 shows the episode of care conceptual model illustrating dimensional hierarchies and possible aggregation paths.

The resulting artifact follows a star schema design as illustrated in Figure 13 including patient, diagnosis, treatment, date, and services dimensions. The fact table stores several pre-calculations resulting from the data extraction, transformation, and loading form the clinical transaction system (EMR) based on the encounters for a specific condition or disease. The pre-calculated values include the number of encounters required for the completion of the patient's episode of care, and the episode start date, the episode end date, and the minimum number of encounters for the given condition and its related treatment.

Demonstration:

To demonstrate the utility and efficacy of this artifact, an initial prototype was developed using Microsoft Access 2003 designed based on a single fact table linked to diagnosis, patient, date, and treatment dimensions. The design for this artifact followed the method proposed by Parmanto et al (2005) comprised of: a single fact table, and dimension tables like patient, diagnosis, date, and clinic among others. The approach proposed by Song et al (2001), for a single diagnosis dimension table including primary and secondary diagnosis was also used for this design. The utility of the first version of the Microsoft Access 2003 artifact was tested by the successful execution of queries before developing a new version based on the SQL Server 2005 database management system. However, for the sake of simplicity, a decision was made not to develop another version of the prototype using a diagnosis group table to provide a way to query diagnosis data based on group or family (Pedersen and Jensen, 4).

Table 10 - Artifact #4 - Requirement Specification Attributes

Requirements Specification for Health Episode of Care Cube - DSQ401-001

Attribute	Description	M	D	O
TimeID	Time key	no	yes	no
Year	time aggregation	no	yes	no
Month	time aggregation	no	yes	no
Day	time aggregation	no	yes	no
Date	time aggregation - EpisodeEndDate	no	yes	no
PatientID	Patient key	no	yes	no
PatientGender	Student gender	no	yes	no
PatientAge	Student Age	no	yes	no
PatientCountry	Student country	no	yes	no
ServiceID	Service key	no	yes	no
ServiceName	Name of health service	no	yes	no
ServiceDescription	Health service description	no	yes	yes
DiagnosisID	Diagnosis key	no	yes	no
Primary_Diagnosis	Primary diagnosis name	no	yes	no
Primary_Diagnosis_Desc	Primary diagnosis description	no	yes	no
Primary_Diagnosis_Code	Primary diagnosis ICD-9 code	no	yes	no
Secondary_Diagnosis1	Secondary diagnosis [1] name	no	yes	no
Secondary_Diagnosis1_Desc	Secondary diagnosis [1] description	no	yes	yes
ProviderID	Provider key	no	yes	no
ProviderFName	Provider's first name	no	yes	yes
ProviderLName	Provider's last name	no	yes	yes
ProviderProfTitle	Provider's professional title	no	yes	yes
TreatmentID	Treatment key	no	yes	no
TreatmentDescription	Treatment description	no	yes	no
TreatmentCategory	Treatment category	no	yes	no
EpisodeID	Episode key	no	yes	no
NumberOfEncounters	Number of visits required to complete episode	yes	no	no
EpisodeStartDate	Episode start date - Diagnosis date	yes	no	no
MinNumberOfEncounters	Minimum number of encounters	yes	no	no

Description and categorization of relevant attributes as measure, dimension, or optional attribute.
According to Hüsemann et al (2000), an optional property attribute (O) is an attribute that does not have to be specified for each element of the corresponding dimension
level and therefore may contain NULL values.

(Hüsemann, et al, 2000, p. 6-7)

Evaluation:

As for the other artifacts, the episode of care data mart was evaluated based on its functionality and the objectives defined at the beginning of the research. The evaluation of the artifact was accomplished by creating and formatting reports on episode of care outcomes for disease and treatment measures and, by executing some OLAP queries. The utility of the artifact was demonstrated by testing and observing the functionality of the artifact. The artifact produced relevant data to answer the answer challenge question. Also, the artifact allowed for easy use, simplicity, accuracy of relevant data. A summary of the evaluation process results for each of the artifacts is shown in Table 11 and 12.

Communication:

The knowledge generated from the design, development, demonstration, and evaluation of the episode of care artifact will be relevant to the stakeholders in order to visualize and find a suitable context within the clinical domain for the use of the presented artifact. The issues, limitations, and risks surrounding the design of this artifact are also important to stakeholders for future evolution of this project.

Contributions:

The episode of care data mart can be viewed as a model for physicians to evaluate, test, and change according to their specific needs in regards to diagnosis and/or treatment. The objective of the artifact solution is to demonstrate an untried (production) tool that offers a multidimensional model for outcome analysis that follows the work of Parmanto et al (2005) design method for outcomes analysis, and Song et al (2001), for a single diagnosis dimension table including primary and secondary diagnosis. Also, the primary diagnosis attributed used in the episode of care artifact was designed to be the only field related to a diagnosis code family.

Figure 11 – Episode of Care Conceptual model

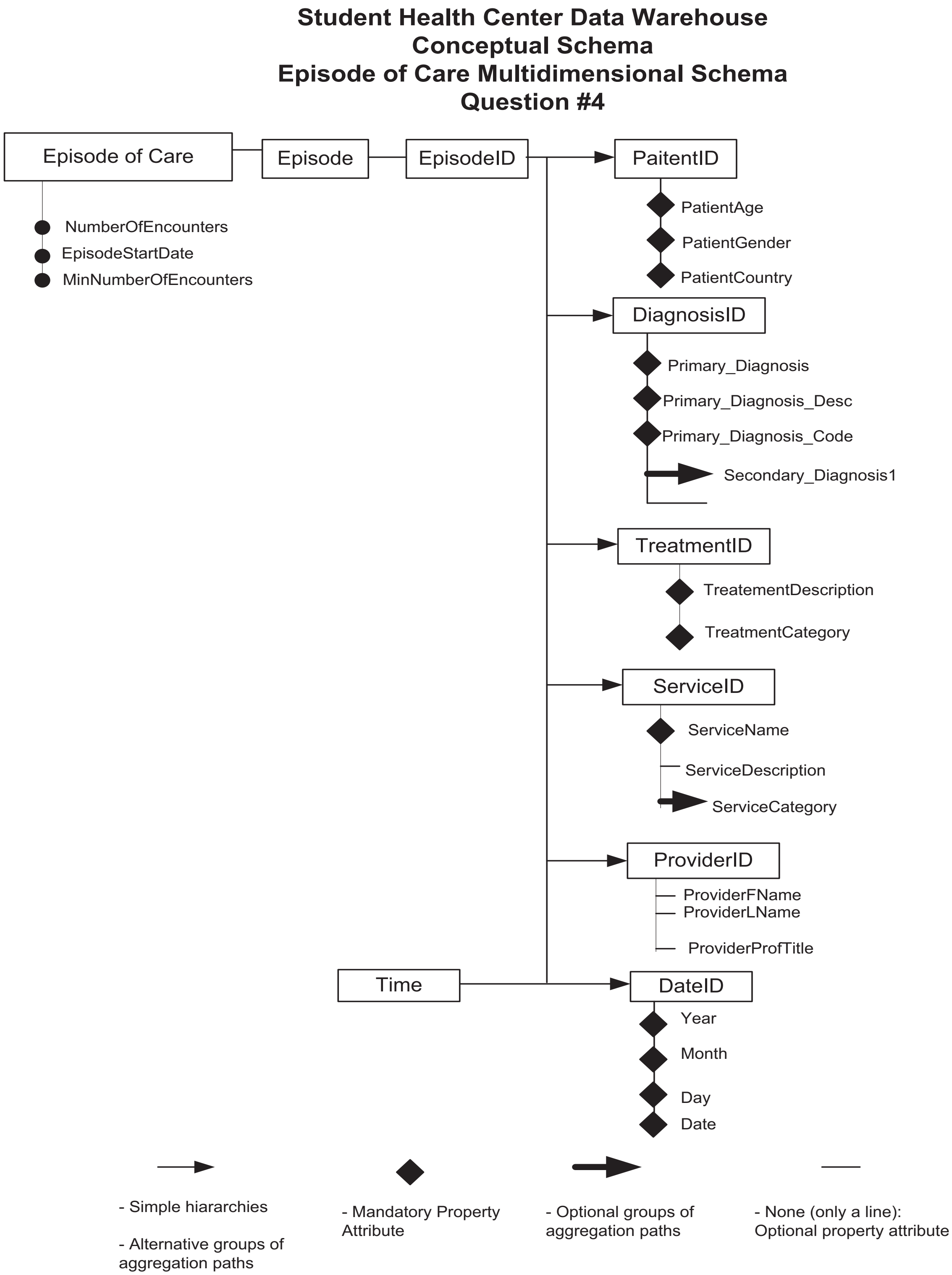


Figure 12 – Episode of Care Logical model

Logical Schema for the episode of care multidimensional model
(Microsoft SQL Server 2005 schema diagram)

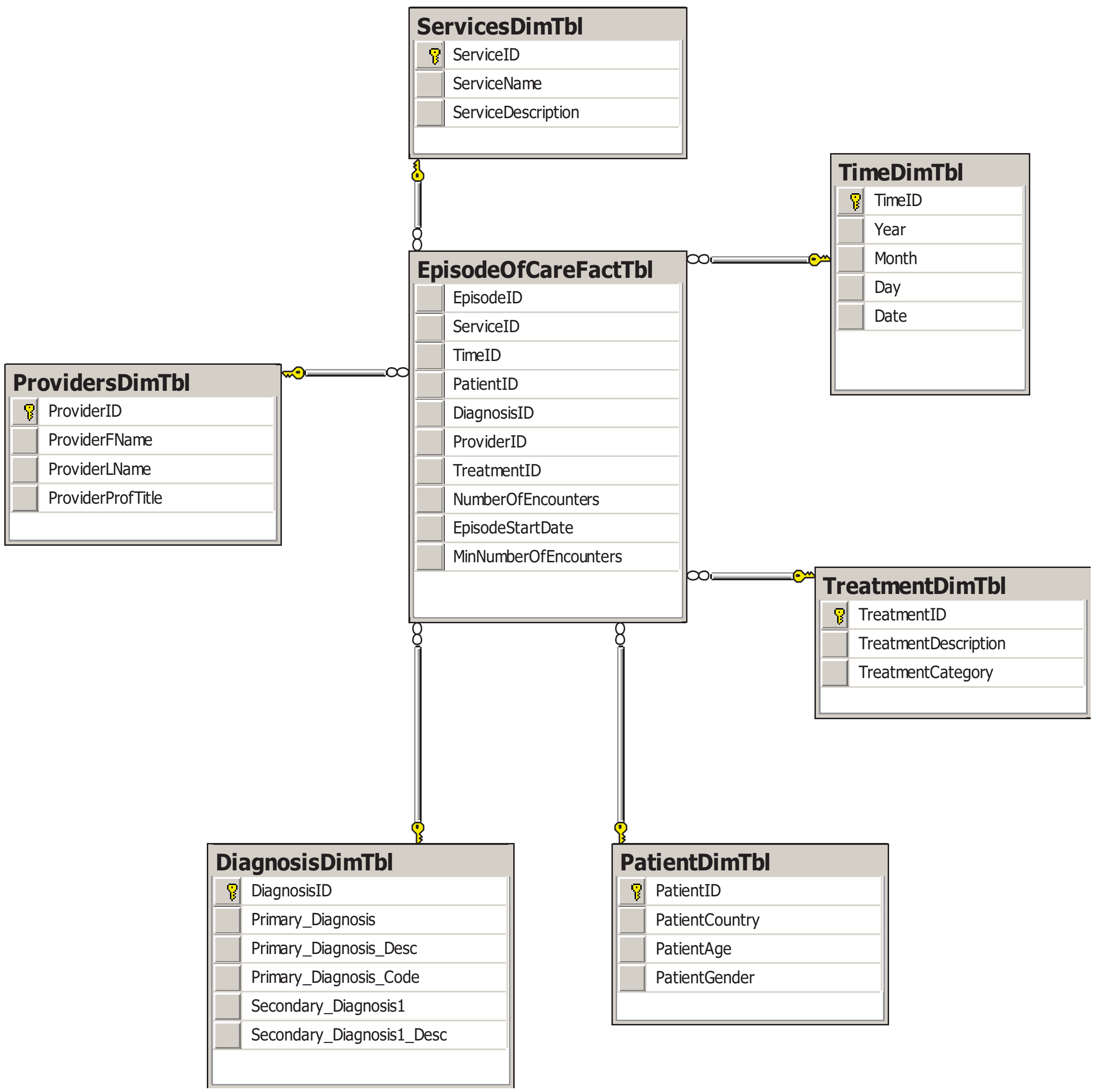


Table 11 – Artifact Evaluation Results Summary – Part 1

	Challenge Question	Question 1	Question 2	Question 3	Question 4
		What percentage of students showed GPA improvement after the successful completion of the alcohol abuse program?	Show the total number of students from China, Japan, Korea, and Vietnam (or from other countries), that did not provide proof of immunization.	Provide data to support the increase or reduction of health services cost.	How to leverage historical clinical data to ultimately improve quality of clinical care?
1.0	Objectives	Artifact 1	Artifact 2	Artifact 3	Artifact 4
1.1	Did the artifact answer challenge question?	The artifact demonstrated to be effective on generating data to answer the challenge question.	The artifact was capable of calculating the answer to the challenge question.	The health services artifact presents data on the number of patients serviced, and the cost of rendering the particular service. This can be contrasted to the revenues, and possibly be related to the student's needs and preferences of service.	The artifact successfully executed several queries to answer the challenge question.
1.2	Ease of Use	The artifact was perceived to be easy to use.	Relatively easy to use for us.	The artifact (star schema and one fact table) was easy to use.	Simple multidimensional schema comprised of one fact table including number of encounters, episode start date, and minimum number of encounters.
1.3	Accurate	Yes, the artifact gave the accurate answer to the challenge question.	Yes, the artifact gave accurate answers to the queries.	Aggregations were accurate.	The aggregations were accurate.
1.4	Simple	This artifact is based on a single star schema including one fact table and several dimension tables.	The immunizations compliance data mart represents a simple design including one fact table, and vaccination dimension, patient, student, providers, and time dimensions.	Simple multidimensional schema (star schema).	Simple multidimensional schema. Primary and secondary diagnosis part of diagnosis dimension.
1.5	Relevant Data	This artifact contains measures relevant to student academic performance, and health habits.	The immunizations compliance data mart hosts vaccination compliance data. This data is relevant to the immunization department and student registration.	Artifact shows relevant pre-aggregated data like number of scheduled appointments, service revenue generated per month, service operating costs stored in the services fact table.	This artifact provides relevant data related to episode of care outcome for disease specific diagnosis and treatment.

(Dell'Aquila et al, 2008) (Gorla, 2003)

Table 12 – Artifacts Evaluation Results Summary – Part 2

	Challenge Question	Question 1	Question 2	Question 3	Question 4
		What percentage of students showed GPA improvement after the successful completion of the alcohol abuse program?	Show the total number of students from China, Japan, Korea, and Vietnam (or from other countries), that did not provide proof of immunization.	Provide data that supports the increase of cost for some services and the reduction of cost for other services.	How to leverage historical clinical data to ultimately improve quality of clinical care?
2.0	Functionality	Artifact 1	Artifact 2	Artifact 3	Artifact 4
2.1	Report Generation	Reports were generated using any of the available OLAP tools (MS SQL Server 2005 Analysis Services & Reporting Services; Oracle 11g Rel. 1 & Oracle Business Intelligence Tools - Report Builder).	Successfully generated reports using available OLAP tools.	Successfully generated reports using available OLAP tools.	Successfully generated reports using available OLAP tools.
2.2	Ad-hoc Queries	The perception was that the artifact was easy to query.	Ad-hoc queries were executed successfully based on knowledge of logical schema.	Ad-hoc queries were executed successfully based on knowledge of logical schema.	Ad-hoc queries were executed successfully based on knowledge of logical schema.

(Dell'Aquila et al, 2008; Gorla, 2003)

4.5 Baskerville et al (2008) - Design Science Research Risk Management

The risk assessment framework proposed by Baskerville et al (2008), is used in this study as a guideline to assess the risks inherent to the design science research methodology presented in this project. The assessment questions proposed by Baskerville et al (2008) are applied to the sequence of activities or phases from the design science methodology used in this study (Baskerville et al, 2008; Peffers et al, 2007, p. 49). The risk assessment framework proposed by Baskerville et al (2008) serves as a tool available to the designer for the assessment of the design science phases based on a rigorous method of evaluation. The results of the assessment can be ultimately communicated to stakeholders, researchers, and readers by describing any of the possible limitations, and/or shortcomings of the design. Since, this design science research project is intended to be used as prelude to a data warehousing project, it is recommended that the results of the risk assessment be taken into consideration for the next evolution of the project.

The risk assessment proposed by Baskerville et al (2008), is used in this study by identifying risks posing a high probability (HP), and mid probability (MP), that could potentially result on severe impact (SI) to the respective activity or phase of the design science research process.

In the category of “Business Needs (Problem Analysis and Choice)”, the identified risks were: “Selection of a problem that lacks significance”, “Poor understanding of the problem to be solved”, “Poor/vague definition/statement of problem to be solved”, and “Solving the wrong problem” as highly probable and resulting on severe impact (Baskerville et al, 2008). So, if any of the problems presented in this study are not significant and/or relevant to the stakeholder's needs, then any possible solution has no value. Furthermore, if the problem and the solution do not reflect a clear understanding of the problem we are trying to solve, and the identification of the problem is wrong, then any efforts to develop a solution hold no value to the stakeholders.

The risks of the category of "Applicable Knowledge (Retrieved from the Body of Recorded Human Knowledge)" are all related to the "Ignorance or lack of knowledge of existing research relevant to the problem understanding" (Baskerville et al, 2008). This area of risk addresses the lack of knowledge based on the lack of research specific to the defined problem, and the lack of understanding about the problem from the researcher's perspective. The risk identified above is relevant to this study because no specific research was found in the literature analysis addressing any of the proposed hypothetical problems. As a result, this study has been developed based on a limited number studies.

However, this study did find some relevant literature to support the formulation of the proposed hypothetical problems, and the development of their respective solution artifacts. The analysis of the challenge questions discusses the motivations, issues and limitations of the problem and possible solution, and also presents any supportive evidence found in the literature demonstrating the level of knowledge available about each problem.

In respect to the design and development of the proposed solutions, the "Develop/Build (Develop Theory/Knowledge and Build an Instantiation)" category use useful on identifying risks of untried hypothetical solutions. The specific severe impact risks identified were “Development of a hypothetical (untried) solution which is ineffective in solving the problem”, “Development of a hypothetical (untried) solution which cannot be taught to or understood by those who are intended to use it”, “Development of a hypothetical (untried) solution which is difficult or impossible to get adopted by those who are intended to use it”, “Development of a hypothetical (untried) solution which causes new problems that make the outcomes of the solution more trouble than the original problem” (Baskerville et al, 2008).

This study did not find evidence of the application artifacts like those designed and developed in this project. Therefore, the artifacts presented in this study are considered untried solutions. Baskerville et al (2008) clearly describe the potential risks related to an untried solution as being ineffective, inefficient, and/or one that could be hard to adopt or understand by the intended users (Baskerville et al, 2008).

Another significant risk related to the “Develop/Build (Develop Theory/Knowledge and Build an Instantiation)” category is that of new problems resulting from the implementation of the proposed solution adding more challenges than the ones being addressed (Baskerville et al, 2008).

Some of the risks identified above can be mitigated during project planning and execution. Also, while the artifacts presented in this project are considered untried solutions their design has been developed based on the relevant literature and the utility test through the evaluation queries and reporting capabilities. The feedback from the stakeholders will also provide insight on the utility and efficacy of the artifacts.

Finally, this study identifies the “Applications (of Knowledge to Business and Organizational Problem Situations) category proposed by Baskerville et al (2008), including the risks of the implementation of an ineffective solution in practice, the use of a solution in the wrong context and the “Inappropriate handling of adoption, diffusion, and organizational change” as relevant to this project (Baskerville et al, 2008).

The risks identified here can be addressed during the stakeholder’s evaluation. The implementation of an ineffective solution should not take place if stakeholders and the IT implementation team use "rigorous methods" to evaluate the problem definitions, problem solutions, and the stakeholder expectations.

The risks presented in this section are in alignment with Hevner et al (2004), design science concepts about the artifact’s utility. According to Hevner et al (2004), “If utility is not demonstrated (evaluation), then there is no basis upon which to accept the claims that it provides any contribution (contribution).” Also, Hevner et al (2004), point out that “If the new artifact is complicated to use, and users can't get the information they need in a timely fashion, then the artifact is not adequate for them.” (Hevner et al, 2004, p. 91)

It is equally important, that the design of a new artifact be relevant. According to Hevner et al (2004), “If existing artifacts are adequate, then design-science research that creates the new artifact is unnecessary (it is irrelevant).” The goal of this project is to present relevant artifacts effective and efficient to meet the given business needs. The literature review did not present any cases of similar artifacts that could render the proposed prototypes useless.

4.6 OLAP Reporting for the Student Health Center: Feasibility Study

Introduction:

The purpose of this section is to assess the feasibility of implementing a data warehouse and OLAP reporting solution to support decision making processes in the context of the University student health center.

According to Alter (2002) a feasibility study "is a user-oriented overview of the proposed information system's purpose and feasibility." Alter (2002), points out that the feasibility study usually covers the economical, technical, and organizational perspectives of the information system project (Alter, 2002, p. 481).

The objective here is to analyze the cost, technical, and organizational perspectives required for the implementation of the new technology, and to list some of the feasible implementation options. The following section briefly presents a hypothetical case of the student health center technology infrastructure.

Existing Technology Infrastructure (Hypothetical Case):

For the purposes of this study, the student health center technology infrastructure is comprised of the following components:

- Data:
    - Patient Health Information (PHI) – Includes patient demographics, clinical information like patient chart, lab orders and results among other information.
    - Patient financial information – Patient insurance information, billing, and payment information.

- Student information – Includes registration data from registrar's office from a batch upload every semester.

- IT Infrastructure:
  - Internal Microsoft Windows Active Directory domain separate from Campus IT.
  - Two Microsoft Windows Server 2008 domain controllers and DNS servers forwarding requests to Campus DNS servers.
  - Electronic Medical Record (EMR) – This is a critical system used daily by at least 100 users.
  - Laboratory Information System (LIS) – This is a critical system used daily by Laboratory staff to process laboratory tests requests from practitioners and physicians. This system delivers reports and interfaces with the EMR system to complete billing and charges operations.
  - Microsoft Office 2003 Professional – Word and Excel used extensively by all level users. Excel can be considered as the best known tool for data analysis and aggregation.
  - Other non-integrated databases and/or applications – Other applications used by different departments like facilities, pharmacy, and human resources (PeopleSoft from Campus IT).
- Users (Three levels):

  The list below shows the three levels of computer skills were identified among the student health center users.

  - Basic – These are users with basic computer skills knowledge.

- Intermediate – These are users with experience and knowledge of common desktop applications like Microsoft Office suite, Adobe reader, Microsoft Windows operating systems like Windows XP and features like explorer for file management, and other commonly used business applications.
- Advanced – This group of users include those who own a business process and have used Microsoft Office Excel extensively for data management needs. These users are also familiar with databases like Microsoft Access and other business applications. This group also includes users with decision making responsibilities.

Table 13 shows a summary of the student health center's existing applications, description of use, hosted data, and relevance to business operations.

Table 13 - Student Health Center Technology Infrastructure

<table>
<tr><td colspan="4">Electronic Medical Records system:</td></tr>
<tr><td>Architecture</td><td>Use</td><td>Data</td><td>Relevance</td></tr>
<tr><td>Microsoft Windows Server 2003 Enterprise; SQL Server 2005; Client server system; Used by at least 100 users daily</td><td>Patient record tracking, clinical charting</td><td>At least 8 years worth of patient data is stored in this system</td><td>Data from this system must be extracted, transformed, and loaded into the data warehouse</td></tr>
<tr><td colspan="4">Student Registration system (Oracle based system):</td></tr>
<tr><td>Architecture</td><td>Use</td><td>Data</td><td>Relevance</td></tr>
<tr><td>Oracle Enterprise</td><td>Student registration</td><td>Student registration data</td><td>Student enrollment and academic performance data hosted by this system</td></tr>
<tr><td colspan="4">Laboratory Information system:</td></tr>
<tr><td>Architecture</td><td>Use</td><td>Data</td><td>Relevance</td></tr>
<tr><td>Microsoft Windows Server 2003 Enterprise; SQL Server 2005; Client server system; Used by at least 15 users daily</td><td>Patient test results tracking and delivery, lab billing, patient reports, and quality control</td><td>Interfaces with EMR; Patient test data, demographics, insurance and other reference laboratories</td><td>Physicians and clinical staff may request data to be uploaded into data warehouse from Lab information system</td></tr>
<tr><td colspan="4">Financial system (PeopleSoft):</td></tr>
<tr><td>Architecture</td><td>Use</td><td>Data</td><td>Relevance</td></tr>
<tr><td>This is the University's financial system shared with departments. Access to this system is restricted to employees of the finance and accounting group within the student health center</td><td>This system is used for payroll, human resources, and finance – accounts payable, and others</td><td>Financial & personnel data</td><td>This system hosts and manages data relevant to payroll, human resources, and finance operations</td></tr>
</table>

Organizational Feasibility:

According to Alter (2002), the organizational feasibility involves the assessment of whether the new technology has enough support from the organization to be implemented successfully, or whether it brings an excessive amount of change and/or the change is too rapid to be handled by the organization (Alter, 2002, p. 481).

Hwang and Xu (2007) identified eleven factors from the literature analysis that contribute to the success or failure of any data warehouses implementation project (Hwang and Xu, 2007, p. 4). Table 14 shows the list of implementation factors proposed by Hwang and Xu (2007), and what is being measured. The first three factors listed in Table 14 are identified as critical implementation factors related to organizational feasibility. In order to measure if the new technology has enough support from the organization, the designers and project manager must ensure that the business needs have been clearly defined, top management support has been obtained, and that an adequate level of user involvement and/or participation has been achieved.

Table 14 - Hwang and Xu (2007) Data Warehouse Critical Implementation Factors

Implementation Factors	Measure
• Clearly defined business needs • Top management support • User involvement/participation	According to Hwang and Xu (2007) these factors “measure the operational aspect of the project.”
• Source data quality • Proper development technology • Adequate IS staff & consultants • Project management (teamwork)	According to Hwang and Xu (2007) these factors “measure the availability of technical resources and expertise for the project.”
• Practical implementation schedule • Proper planning/scoping of project	According to Hwang and Xu (2007), these factors “measure how reasonable the time allowed for development of a data warehouse is.”
• Adequate funding • Measurable business benefits	According to Hwang and Xu (2007), these factors “measure the economic aspect of the project.”

(Hwang & Xu, 2007)

Table 15 lists eight success variables proposed by Hwang and Xu (2007) derived from previous data warehouse studies with the intent to answer their research question of "Which implementation factors have an effect on which system success variables?" (Hwang and Xu, 2007, p. 6)

Table 15 - Hwang and Xu (2007) System Success Variables Table

System Success Variables	Measure
• Easy to use • Speedy information retrieval	According to Hwang and Xu (2007), these two variables "measure the quality of the system developed."
• More information • Better quality information	According to Hwang and Xu (2007), these two variables "measure the benefits reflected in the output of a system—information."
• Improved productivity • Better decisions	According to Hwang and Xu (2007), these two variables "measure benefits resulted from the use of information by individual decision makers."
• Improved business processes • Increased competitive position	According to Hwang and Xu (2007), these two variables "measure benefits accrued at the organizational level."

(Hwang & Xu, 2007)

The implementation factors proposed by Hwang and Xu (2007), and previously shown in Table 14, are used in this study to describe the organizational feasibility related to the implementation of the data marts and OLAP tools for the student health center. Table 16 shows the assessment of the feasibility of each factor. These implementation factors can be arranged in a logical order to be presented to the student health center leadership as a group of factors necessary to build the business case.

Table 16 - Hwang and Xu (2007) Implementation Factors Feasibility

Implementation Factors	Feasibility
• Clearly defined business needs	This should be possible after presenting the use and utility of the prototypes based on the challenge questions.
• Top management support	This is feasible and is considered a critical success factor.
• User involvement/participation	This is a critical success factor. Users should be involved in the design, planning, and implementation of the solution to the chosen subject area. This needs to be addressed at the beginning of the project.
• Source data quality	To successfully extract and clean source data to populate the data marts, both IT designers and physicians will need to address specific data quality problems through a data cleaning approach.
• Proper development technology	This is also feasible since the development technology used here is based on known and tried development products (Oracle Database 10g or 11g R2; Microsoft SQL Server 2005 or 2008).
• Adequate IS staff & consultants	The realization of this factor presents a challenge. However, several options exist like consulting, partnering with other University departments, and/or event.
• Project management (teamwork)	This is feasible if the student health center appoints an experienced IT project manager whether is under contract or a member of the IT staff.
• Practical implementation schedule	This is feasible. This is negotiated and established during project scope and planning phases.
• Proper planning/scoping of project	This is feasible as long as project manager, SHC management, and physicians are involved on initiative, and are willing to cooperate.
• Adequate funding	Feasibility in this area is determined by Finance director and SHC leadership.
• Measurable business benefits	It is feasible for the project team to define ways to measure the expected business benefits for the given application as long as there is agreement among stakeholders.

(Hwang & Xu, 2007)

Technical Feasibility:

The technical feasibility of this project is dependent on the availability of the technical expertise necessary to implement and maintain the data warehouse infrastructure and the OLAP tools, the availability of the features required, and the demonstration of the efficacy of the artifact(s) (Alter, 2002, p. 481).

The critical components of this implementation are the source data, the data extraction, transformation, and loading process, the data warehouse architecture, and the data analysis tools or OLAP tools.

Operational Data Source:

The student health center will realize the benefit of all patient care data (clinical and transactional) stored in the multiple sources like the Electronic Medical Records (EMR) system, and in the Practice management system. Other, source data exists in the form of Microsoft Excel spreadsheets, Microsoft Access databases, and text documents.

The number of heterogeneous data sources mentioned above and the quality of the data they host will have an impact on the level and complexity of data cleansing required prior to the insertion into the data warehouse or data marts (Rahm & Hong, 2000).

Data Extraction, Transformation, and Loading (ETL) Mechanism:

The assumption in this study is that the data extraction, transformation, and loading from the operational systems will be addressed at a later time after management makes the decision to adopt the new technology and initiate the data warehouse and OLAP tools solution project. However, the ETL process is extremely important for the implementation and maintenance of the student health center’s data warehouse project.

Data Warehouse Architecture:

The conceptual architecture shown in Figure 2 is comprised of the operational source (EMR and Student Registration system) layer, the data transformation (ETL) staging area layer, the data warehouse layer, and the data analysis (OLAP tools) layer.

Several of the cases presented in the literature made use of a data mart architecture as a way to meet the needs of departments, to provide flexibility in the design, and to provide an architecture that will be easier to evolve with time (Sahama & Croll, 2007, p. 228), (Bréant et al, 2005). In this study, a data mart architectural approach is used to build each of the proposed prototypes. Four separate data marts were designed, implemented, and evaluated in this study to meet the requirements established at the beginning of the project. Each data mart represents a separate cube implemented in a relational OLAP server (ROLAP). One of the key advantages of selected architecture is that each data mart will allow each department to address their business analysis needs as its own object of analysis, and still allowing for the possibility of integrating all the separate (conformed) data marts into an enterprise data warehouse (Sen & Sinha, 2007, p. 81).

The student health center can use any of the known commercial relational database management system (RDBMS) like Oracle 10g or Microsoft SQL Server 2005 as the Relational OLAP (ROLAP) engine. However, as stated before, the ease of use of the system is a critical success factor for this initiative. In a business intelligence systems comparative analysis, Dell'Aquila et al (2008), show that Oracle Discoverer obtained better score on the creating reports task under the reporting capability than Microsoft SQL Server 2005 (Dell'Aquila et al, 2008).

The student health center can use either Microsoft SQL Server 2005 with Analysis Services, and Reporting Services, or Oracle database 10g with Oracle BI Tools (Discoverer).

Data Analysis (OLAP tools):

The OLAP tools are one of the critical components of the overall architecture that could decide the success of the project. The OLAP tools will extract the decision making data from the data warehouse, or from the proposed data marts. In this study, Microsoft Access 2003 has been used as a tool to demonstrate the utility and efficacy of the artifacts prior to their evaluation using Microsoft SQL Server 2005 and Oracle 11g.

Technical Expertise:

As pointed out by Hwang and Xu (2007), the availability of adequate and qualified IS staff and/or consultants to help plan and implement the project are critical success factors (Hwang and Xu, 2007, p. 6). For this particular project two important skill sets and roles are necessary. See below for details.

- Project Management role:

  The project manager's role is critical for the success of this initiative. The project manager responsibilities can be fulfilled by a consultant or a member of the IT department with project management training and/or experience. The project manager responsibilities include the development of the project charter, project scope statement, and the project management plan. The project manager needs to assemble the project team, and work out all schedules. As stated by Hwang and Xu (2007), the proper planning and scoping of the project is essential. This will be one of the most important and critical task of the project (Hwang & Xu, 2007, p. 6).

- Data Warehousing and Business Intelligence tools analyst or designer:

  As mentioned before, this role is critical to the success of the proposed project initiative. However, this position could be challenging to fill since the size of the organization, and possibly the limited availability of funds to compensate this role could impose some limitations. However, if this role is combined with the responsibilities of a DBA, it is more likely to be filled faster than for an exclusive data warehousing and/or business intelligence position. An internet search on Dice career site http://www.dice.com for data warehousing jobs, showed a great variety of titles for the position described above. Some of the critical responsibilities to consider for this role are:

  - Be member or part of project team
  - Review DSR project designs in light of user's perspective, and identify business and data needs
  - Translate user and data requirements into conceptual, logical, and physical multidimensional data design
  - Develop ETL data mappings, and scripts
  - Develop a data source analysis strategy that includes managing quality, integrity, and security of data

Proposed Solution Options:

Option#1 - Complete Microsoft based solution (Excel 2003)

This solution makes use of Microsoft Excel 2003 as the analysis tool and Microsoft SQL Server 2005 Standard, Microsoft SQL Server 2005 Analysis Services to provide the software technical infrastructure for the proposed problem solution. This solution stores the multidimensional data in MS SQL Server 2005 as the core of the ROLAP architecture.

PivotTable reports can be created from Microsoft Excel 2003, and Online Analytical Processing (OLAP) source data from the MS SQL Server 2005 database.

Option#1 - Advantages:

- SHC users are familiar with Excel spreadsheets
- Low cost solution
- Easy deployment and maintenance
- Interface with EMR easier to develop since both share same DBMS platform

Option#1 - Disadvantages:

- Pivot tables won't provide the same functionality found in OLAP tools from both commercial and open source products
- While Microsoft Excel 2003 can save and export documents in different formats, it does not save as PDF format. Microsoft Reporting Services supports saving reports as PDF format.

Option#2 - Complete Microsoft based solution (SQL Server 2005)

This solution makes use of Microsoft SQL Server 2005 Standard, Microsoft SQL Server 2005 Analysis Services, and Microsoft Reporting Services to construct the multidimensional data cubes, the different data marts, and to generate the different queries and/or reports used in this project.

Option#2 - Advantages:

- Single vendor solution – Expected to provide easier integration with Office suite
- Still considered a low cost solution (SQL Server 2005 Standard or Enterprise provides the Analysis Services option and the Reporting Services option
- Single vendor product simplifies IT technical skills required
- Interface with EMR easier to develop since both share same DBMS platform
- SQL Server Reporting Services enables users to easy create ad-hoc reports and to export them in different formats (CSV-comma-separated value format; TIFF file—Saves the report in Tagged Image File Format; Acrobat (PDF) file—Saves the report in Acrobat Portable Document Format; Web archive—Saves the report in MIME HTML format (mhtml); Excel—Saves the report as a Microsoft Excel spreadsheet.

Option#2 - Disadvantages:

- Requires Microsoft SQL Server 2005 Analysis Services and Reporting Services training and/or experience
- Microsoft Windows Server based solution, can't run on Linux. This limits the student health center on platform options.

Option#3 – Complete Oracle Database & BI Tools solution:

The Oracle OLAP option enables Oracle Database 10g or 11g to store multidimensional data. Oracle offers a business intelligence suite of products that complement the OLAP option for the Oracle Database 10g or 11g mentioned above. The

Oracle business intelligence line of products includes Oracle Discover Desktop and Administrator (Dell'Aquila et al, 2008), and the OracleBI Spreadsheet Add-In.

Option#3 Advantages:

- Oracle BI Spreadsheet Add-In enables users to display and navigate Oracle OLAP data directly from within Microsoft Excel
- Strong BI platform – based on Dell Aquila's functional Complexity of BI platforms
- Can use wizards to create ad-hoc reports

Option#3 Disadvantages:

- Requires Oracle database and business intelligence tools training and/or experience
- Could turn out to be an expensive solution
- Need to develop interface to existing SQL Server 2005 EMR system product
- Due to the complexity of Oracle as a database platform, organizations using Oracle tend to hire experienced and expensive database professionals to perform necessary maintenance, backups and restores, and upgrade operations

Cost Feasibility:

The cost feasibility is used to address the questions related to cost of implementing the proposed data marts. However, for the purposes of this study no pricing information was provided since each academic institution pricing agreement is different. The purpose of this section is to emphasize that the lower cost of software and hardware based on academic pricing provides higher education institutions and departments with an affordable mean to obtain the tools and systems required to meet their needs. The student health center should be able to benefit from these types of agreements for computer software and hardware. This section lists the components of each of the proposed solution options including both Microsoft and Oracle databases, and their respective OLAP solution tools.

Option #1 & #2 – Complete Microsoft based solution

- Microsoft SQL Server 2005 Enterprise – 1 license & media kit
- Microsoft SQL Server 2005 CALs – At least 5 licenses
- Database server like DELL 2950III Dual QuadCore server with 16GB RAM, and 4TB (1TB hard drives) – 1 server minimum to be used as a database server
- Database server like DELL 2950III Dual QuadCore server with 16GB RAM, and 4TB (1TB hard drives) – 1 server minimum to be used as a web/application server
- Storage Area Network (SAN) – This could be storage space leased from Central IT for a monthly fee or purchase of a small SAN for the student health center. DELL offers a SAN solution for small business that could meet the needs of the student health center for an affordable cost
- Microsoft Windows Server 2003 or 2008 Standard or Enterprise Edition 64 bit
- Microsoft Office 2003 or 2007 Professional

Option #3 – Complete Oracle Product based solution

- Oracle Database 10g or 11g Standard or Enterprise server – 1 license includes 5 clients minimum
- Oracle BI Discoverer Desktop and/or Oracle BI Spreadsheet Add-In (Client tool)
- Database server like DELL 2950III Dual QuadCore server with 16GB RAM, and 4TB (1TB hard drives) – 1 server minimum to be used as a database server

- Database server like DELL 2950III Dual QuadCore server with 16GB RAM, and 4TB (1TB hard drives) – 1 server minimum to be used as a web/application server
- Storage Area Network (SAN) – This could be storage space leased from Central IT for a monthly fee or purchase of a small SAN for the student health center. DELL offers a SAN solution for small business that could meet the needs of the student health center for an affordable cost
- Microsoft Windows Server 2003 or 2008 Standard or Enterprise Edition 64 bit
- Microsoft Office 2003 or 2007 Professional

Since academic pricing allows higher education institutions to acquire hardware and software at competitive discount rates, it should be feasible for the student health center to acquire any of the software and hardware components required for the proposed solutions.

Recommendations and Conclusions:

This section presents the final recommendations and conclusions from the feasibility study.

Recommendations:

1. Acquire essential human resources (Project manager & Data warehouse-DBA). This could be accomplished by enhancing the designs presented here to support the request for research grants to support the hiring of the essential clinical data analyst staff as part of the IT group or part of the clinical staff. Also, an agreement can be reached between the graduate the faculty and/or the computer information systems department of the University to allow graduate students to participate in the design and support

of the proposed technology. Finally, the student health center administration can establish a service level agreement with the centralized campus IT group for the services needed to support the acquisition of the new technology.

2. Assemble a project team to evaluate this project proposal including the artifacts and their respective designs, evaluation criteria, and benefits.
3. Review and evaluate with project team and/or stakeholders the need for OLAP reports and data warehousing as technologies to benefit decision support operations.
4. "Identify Key Business Questions" or areas of analysis (Ewen et al, 1998, p. 48).
5. Evaluate existing data sources for relevant data, content, and structure (Ewen et al, 1998, p. 48).
6. Ensure to include enough storage or the acquisition of a Storage Area Network (SAN) solution in the data mart architecture.

Conclusions:

1. The implementation of an OLAP reporting technology infrastructure for the student health center is a challenging undertaking, but a feasible one. The use of a data mart architecture approach as shown in Figure 13 is feasible and convenient for the implementation of the artifacts proposed in this study.
2. Ewen et al (1998), point out that in their project they elected to focus on a single subject area with the intention to provide a foundation for additional subject areas (Ewen et al, 1998, p. 52). In this project four prototypes have

been presented based on four multidimensional models representing four hypothetical subject areas relevant to the student health center organization. In order to implement the proposed prototypes, the student health center can follow the approach presented by Ewen et al (1998), by prioritizing the implementation of the subject area artifacts one at a time to ensure project control and success (Ewen et al, 1998, p. 52).

3. The formulation of a data cleaning approach for this project should be treated as a critical aspect of the overall data warehouse initiative. According to Rahm and Hai Do (2000), general data cleaning approaches involves several phases like data analysis, definition of transformation workflows and mapping rules, verification, transformation, and backflow of cleaned data (Rahm & Hai Do, 2000, p. 5). This study recommends, that both Physicians and IT staff work as a team to develop an effective and efficient data cleaning strategy.

4. Finally, the student health center has the option of acquiring a commercial clinical data warehouse system that can be customized to meet the data analysis needs of management and/or physicians. In a paper presented by Akhtar et al (2005), present the results of a survey comparing in-house versus commercial clinical data warehouse, showing that 80 percent of the respondents representing hospital and/or clinics preferred a commercial solution based on factors like cost, quality, and scope (Akhtar et al, 2005, p. 25). The student health center administration should consider a commercial clinical data warehouse solution.

Chapter 5 – Project History

The original proposal for this project was focused on the development of an enterprise data warehouse for the student health center. However, it became apparent that it was more effective and efficient to approach the project as separate artifact designs in order to clearly demonstrate the object of analysis for each individual problem. Also, the development of individual prototypes to solve each proposed problem could give the stakeholders several options to choose from as separate subprojects in a staged implementation approach.

The use of a design-science research methodology required the use of several relevant concepts and “rigorous design” methods found in the literature to design and build the solution artifacts. Design-science research allowed us to iterate through the phases of design, construction, and demonstration in order to arrive to an acceptable and usable artifact. An attempt was made to track design iterations as a way to control and manage changes. The risk framework proposed by Baskerville et al (2008) was used in this study to provide an objective assessment of the risks related to each design phases and to propose possible risk mitigation options.

Chapter 6 – Conclusions

Project Challenges:

The analysis of the hypothetical questions presented in this study proved to be challenging. Each challenge question had to be analyzed prior to the start of the design science research process in order to determine the design requirements and assumptions. The analysis of the first challenge question triggered a thorough review of the design assumptions prior to the development of the conceptual design. For this particular question, several ideas derived from the literature review helped redefine the assumptions and solution design approach. Specifically, a second analysis of the first challenge question and the literature motivated the idea of using course evaluations as a method for collecting data about the student academic performance and health habits in a volunteer basis. This approach would then mitigate the possible limitation of the proposed design that was originally based on querying student academic performance data that is deemed confidential.

The episode of care challenge question presented an unexpected level of complexity in regards to the subject of analysis. Further analysis of the literature helped us understand some of the applications of episode of care analysis for a specific disease or type of treatment. However, it is clear from the literature that physicians and/or researchers must define how the episode of care will be used for the analysis of patient disease and/or treatment. While the evaluation process successfully demonstrated the utility of the episode of care artifact, a more effective evaluation should be provided by the physicians and/or researchers with aim to effectively validate the utility of the artifact for a particular subject of analysis. Appendices B, C, D, and E present a description of the analysis process for each of the hypothetical questions.

Another challenging aspect of this study was that of data quality. Data hosted by each data mart must meet some quality standards before is inserted and as it is analyzed. For example, in order for the data in the episode of care data mart to be relevant to the stakeholders, every primary diagnosis had to be properly matched to the appropriate treatment and diagnosis code. Through the testing and evaluation process, data quality had to be validated to ensure the accuracy of the expected output. The need to validate the quality of data also reinforces the fact that practitioners, clinicians and physicians need to be involved in the data quality assessment process.

Finally, the selection of an OLAP tool presented several challenges. The original intent was to use an open source tool like Pentaho Mondrian OLAP server (http://mondrian.pentaho.org) to generate OLAP reports. However, any efforts to setup Pentaho Mondrian were not effective. As a result, the decision was made to continue the project using Microsoft Reporting services and Oracle Business Intelligence tools like the Add-In and Discoverer.

Lessons Learned:

This project had set ambitious goals for the design and implementation of four untried solutions within the context of the student health center. While the artifact solutions presented in this study seem to be simple, each multidimensional model represents a subject of analysis requiring a depth of knowledge in the specific subject areas in order to conduct the design, evaluation, and effective use of the proposed artifacts.

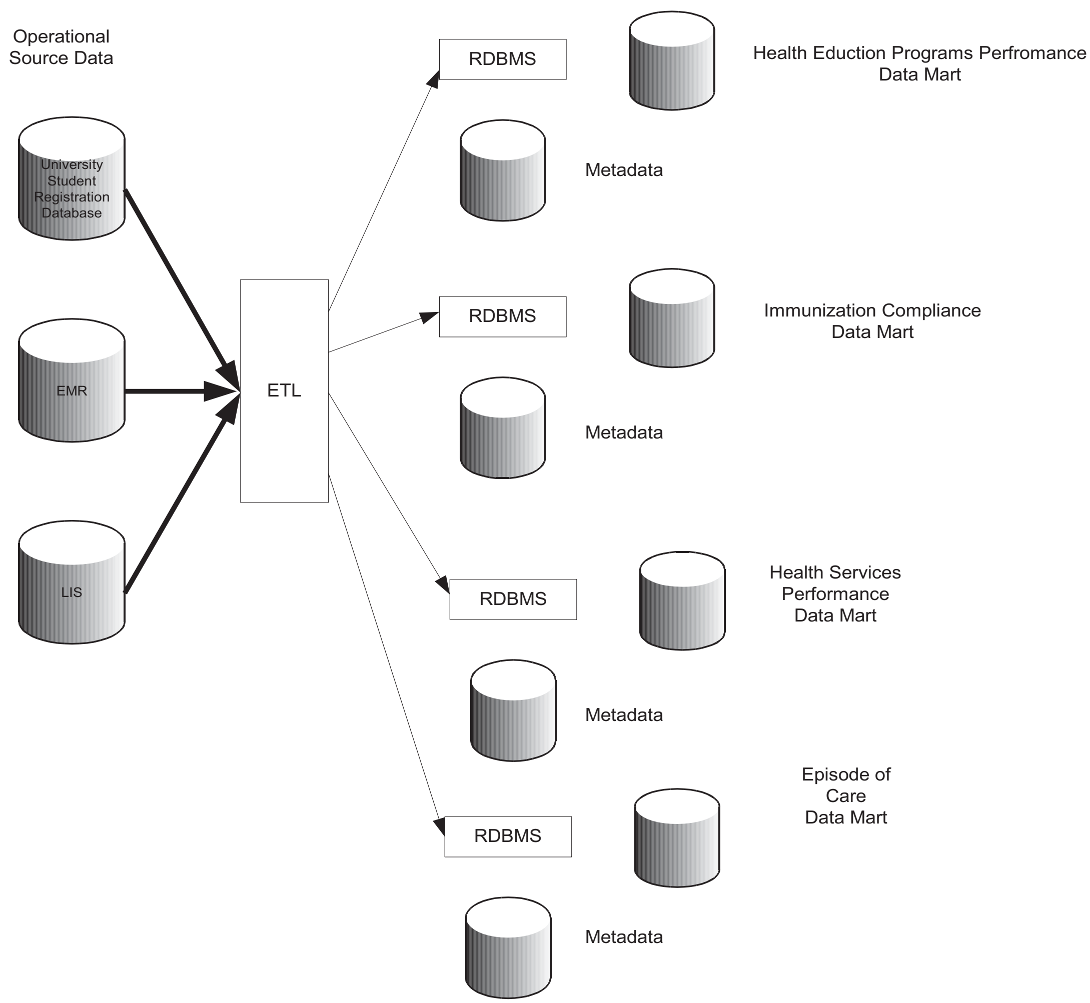
The analysis of the health care related literature and the process of designing the solution artifacts revealed the importance of including physicians as domain experts in the selection of the subject of analysis, artifact design, and the potential use of the solution. However, despite the limited subject area knowledge, this study has presented enough evidence to support the idea that the student health center can leverage existing patient data through the use of OLAP reporting technology to support decision making processes towards better patient services and patient care.

Next Evolution of the Project:

Each of the artifacts designed and developed in this project, was tested to demonstrate utility, quality, and efficacy through rigorous and "well-executed evaluation methods." However, the evaluation methods applied in this study should not be considered as sufficient to satisfy the needs of the health care professionals. It is recommended that the artifacts developed in this study undergo further evaluation from the perspective of physicians and clinicians in order to validate the utility of each artifact (Hevner et al, 2004, p. 83) (Tremblay et al, 2007).

The next evolution of this project should be the analysis and evaluation of the artifacts, project, and the feasibility study by the student health center administration and physicians. The insight provided by these users about the particular artifact designs and the proposed value to the student health center is considered part of the rigorous evaluation methods that characterize design-science research. The expectation is that the results from the evaluation of the solution artifacts and from the project in general will serve as a motivation to the student health center administration for the initiation of a project for the implementation of OLAP reporting technology.

Figure 13 - Student Health Center Data Mart Architecture



(Sen & Sinha, 2005, p. 80)

References

American Psychological Association. (2002). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author. ISBN: 1-55798-791-2.

Hacker, D. (2003). *A Writer's Reference* (5th ed.). Boston, MA: Bedford/St. Martin's. ISBN: 0-312-397767-4 soft.

Leedy, P. D., & Ormrod, J. E. (2005). *Practical Research: Planning and Design* (8th ed.). Upper Saddle River, NJ: Pearson Education.

Ponniah, P. (2001). *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*, Wiley-Interscience Publication.

Connoly, T., Begg, C. (2005). *Database Systems: A Practical Approach to Design, Implementation, and Management* (4th Ed.), Addison Wesley.

Rizzi, S., “Conceptual modeling solutions for the data warehouse”, in: R. Wrembel, C. Koncilia (Eds.), *Data Warehouses and OLAP: Concepts, Architectures and Solutions*, Idea Group, 2006.

Jarke, M., Lenzerini, M., Vassiliou, Y., Vassiliadis, P. (2003). *Fundamentals of Data Warehouses*, Springer.

Connoly, T., Begg, C. (2005). *Database Systems: A Practical Approach to Design, Implementation, and Management* (4th Ed.), Addison Wesley.

Kimball, R., & Ross, M. (2002). *The Data Warehouse Toolkit* (2nd Ed.) Wiley Computer Publishing.

References - Continued

Alter, S., (2002). *Information Systems: The Foundation Of E-Business (4th Ed.)*, Prentice Hall/Pearson Education.

American College Health Association - National College Health Assessment (2008). *Reference Group Executive Summary Fall 2008*. Retrieved May 15, 2009 from http://www.acha.org/info_resources/hc2010.cfm

American College Health Association (2009). ACHA Position Paper: College Health and National Health Care Reform. Retrieved May 11, 2009 from http://www.acha.org/College_Health_and_Health_Care_Reform.pdf

Project Management Institute, (2004). *PMBOK: A Guide to the Project Management Body of Knowledge*. Project Management Institute, Newtown, PA.

Pentaho (2009). *Pentaho Analysis Services: Mondrian Project*. Retrieved May 1, 2009 from http://mondrian.pentaho.org/

University of Berkeley Student Health Services. (2005). *Frequently Asked Questions*. Retrieved February, 15, 2009 from http://www.ocf.berkeley.edu/~health/faq.html#q8

Oracle (2003). *Oracle OLAP Application Developer's Guide 10g Release 1 (10.1)*. Retrieved April, 23, 2009 from http://www.oracle.com/technology/documentation/olap.html

Microsoft Corporation (2009). Retrieved March, 12, 2009 from http://www.microsoft.com/sql

References - Continued

Microstrategy (2009). *Business Intelligence Products and Services*. Retrieved July 2, 2009 from http://www.microstrategy.com

Appendix A - Some of the Research Rigorous Methods Used

Theory	Contribution	Construction/Implementation	Evaluation
Conceptual modeling approach (Hüsemann et al, 2000)	This conceptual design method can be used to guide the design of the conceptual models. The visual representations presented in this study can help the design process.	Used the some of the ideas and concepts proposed by Husemann et al (2000) approach to develop the conceptual models for this study.	The conceptual models generated from the conceptual modeling process will be evaluated after the test of the physical implementation of the model.
Design Framework for a multidimensional database for a healthcare outcomes data warehouse (Parmanto et al, 2005)	According to Parmanto et al (2005), their proposed design process "can be used as a blue print for the development of a data warehouse for healthcare decision support."	Developed Episode of care artifact based on Parmanto et al (2005) model and theory.	Used Parmanto et al (2005) model to develop a prototype for a disease/treatment episode of care artifact for the student health center.
Approach to many-to-many relationships between dimension and fact tables (Song et al, 2001)	The analysis of many-to-many relationships, along with recommendations for a solution (Song et al, 2001).	Used Song et al (2001), Method C-2: One-to-Many relationship between dimension and fact tables, for Episode of care data mart (Song et al, 2001, p. 6-9).	Evaluation based on ability to query multidimensional model. Physician's input will greatly enhance this area.
Comparison of Data Warehouse Architectures (Sen & Sinha, 2005)	Sen and Sinha (2005) provide two possible data warehouse architectures to be considered in this project: Enterprise data warehouse and data mart architectures.	Chose to construct individual data marts as represented by Sen and Sinha (2005) in their study. The literature suggests the use of data marts as a flexible architectural option (Sen & Sinha, 2005).	Created separate artifacts modeled as separate data marts for a specific object of analysis.
Clinical Data Warehouse implementation (Sahama & Croll, 2007)	Discussion of a clinical data warehouse implementation approach.	This study was influential in the selection of a data mart architecture (Sahama & Croll, 2007).	Supports use of data marts.
Feasibility study and comparison of two groups of students in regards to alcohol problems. Seek to lay the groundwork for "developing an effective secondary prevention program for college drinking." (Ehrlich et al, 2006, p. 280)	This is study describes the issue of alcohol abuse in US colleges. Ehrlich et al (2006), make use of a questionnaire to collect data about alcohol use disorders (Ehrlich et al, 2006, p. 287).	Implemented multidimensional model measures based on health education course evaluation questionnaire for both academic performance improvement and health improvements as a result of the completion of the course.	The artifact was able to link or relate health education course data with student academic performance data voluntarily released by students through the course evaluation.

Appendix B – Analysis of Challenge Question #1

What percentage of students showed GPA improvement after the successful completion of the alcohol abuse program?

Hypothetical Question #1:

What percentage of students enrolled in the alcohol abuse program showed GPA improvement until graduation after the successful completion of the program?

Student1

	Year					
Term-GPA	2008	2007	2006	2005	2004	2003
Spring		3.45	3	3.35	2.75	
Summer		3.6	3.3		2.85	
Fall		3.75	3.4	3.45	2.95	2.65

Student1
Degree Completed:
Graduation Term: Fall 2007
Final GPA: 3.75

Alcohol abuse program completed: 12/10/2003

Student 1 Enrolment date: 8/01/2003

…

Percentage of students showed GPA improvement every term until graduation subsequent to the completion of the alcohol abuse program.

	YEAR				
	2008	2007	2006	2005	2004
Alcohol abuse program	90%	80%	76%	90%	80%
Non-Alcohol abuse program	20%	10%	5%	7%	3%

Purpose:

This question aims to provide evidence that demonstrates the positive impact of the alcohol abuse program on student academic performance.

Motivation:

This question is motivated by the need to find data that supports the effectiveness of the student health education programs demonstrated through the student's academic performance.

Object of analysis:

This question suggests that the object of analysis is the student's academic performance (TermGPA) after the successful completion of the alcohol abuse program. In essence, this study is seeking to establish a connection between health education programs and student academic performance.

Required data:

It should be fairly simple to obtain a report on student academic performance on each term until graduation from the registrar's database system. The report showing the student's academic performance (TermGPA) could look like the example in the Sample Data sets below.

Key Stakeholders:

The following roles or group of users can be considered to be key stakeholders: Student Health center director, Clinical services director and Counseling and behavioral services director and/or staff (Psychiatric and Psychology).

Proposed Value to Stakeholders:

Addressing this question provides an opportunity find auditable evidence of the effectiveness of the student health education programs in student academic performance, which is one of the key organizational goals.

Sample Data Sets:

University Student Registration System Database query – sample data

StudentID	StudentSSN	TermID	TermGPA
12345678	453121488	FL2003	**2.65**
12345678	453121488	SP2004	**2.75**
12345678	453121488	SM2004	**2.85**
12345678	453121488	FL2004	**2.95**
12345678	453121488	SP2005	**3.3**
12345678	453121488	SM2005	**3.35**
12345678	453121488	FL2005	**3.45**
12345678	453121488	SP2006	**3**
12345678	453121488	SM2006	**3.3**
12345678	453121488	FL2006	**3.4**
12345678	453121488	SP2007	**3.45**
12345678	453121488	SM2007	**3.6**
12345678	453121488	FL2007	**3.75**

Student 1 (12345678) Term GPA report based on enrolment date: 8/01/2003, Fall term 2003 and degree completed on the Fall 2007 with final GPA of 3.75.

Student Health Center Health Education program – sample data

PatientID	PatientSSN	HealthEdProgID	HealthEdPrgName
12345678	453121488	HED1001-03	Alcohol Abuse program – 03
12347911	160122048	HED1001-04	Alcohol Abuse program – 04
12349056	102129632	HED1001-05	Alcohol Abuse program – 05
12350201	102963212	HED1001-06	Alcohol Abuse program – 06
12351346	102963216	HED1002-04	Impact of Drugs in Academic Performance – 204
12352491	453099525	HED1002-06	Impact of Drugs in Academic Performance – 206

Student Health Center Health Education program – sample data - Continued

HEdProgSession	HEdProgStartDate	HEdProgEndDate
HED1001-03	03-Oct-03	10-Dec-03
HED1001-04	03-Feb-04	10-Apr-04
HED1001-05	03-Oct-05	10-Dec-05
HED1001-06	03-Oct-06	10-Dec-06
HED1002-04	15-Apr-04	15-May-04
HED1002-06	15-Apr-06	15-May-06

However, as a standard confidentiality requirement, student grades are not released to anybody unless authorized by the Registrar with the consent of the student. This is commonly accomplished through an official transcript.

On the other hand, the student health center should not have a problem generating a report of all students that successfully completed the alcohol abuse program for a given date range. However, since this data is also considered patient data, some restrictions (HIPAA) may apply.

Based on the confidential nature of the data required to answer this question, it is important to realize that this particular question presents a real challenge for this project and it might be difficult to develop an efficient solution to the proposed problem. Therefore, further analysis of this challenge question will be necessary to develop a significant and/or realistic problem definition and problem solution. The section below describes the analysis in detail:

In order to answer this challenge question, the following data is required:

- Select all students or patients that were enrolled in the alcohol abuse program in the fall term of 2003 (Enrolment date: 8/01/2003), and that successfully completed the class (Class completion: 12/10/2003), and that subsequently graduated (Graduation Term: Fall 2007).

- From the result set above, all students that achieved a higher term GPA following the completion of the alcohol abuse program must be selected. That means, each subsequent term GPA must be compared with the fall term GPA.
- Then, number of students can be represented as a percentage.

Issues and/or Limitations:

a. The solution to the given question requires the extraction and manipulation of confidential data requiring the implementation of security measures, the approval of confidentiality agreements between the student health center and the registrar's office. Most likely not possible to share student performance data with anybody.

b. Another issue related to the challenge question is whether or not the academic performance improvement is measured based on the case of progressive GPA improvement or immediately after the completion of the alcohol abuse program instead the extraction of that data will require the use of a script, stored procedure or other data manipulation process. This will impact query design and ETL processes.

c. Another approach to matching student academic performance with health education program data is to ask students if they have experienced academic performance improvement after the completion of the course. This needs to be a voluntary release of information from the student.

Recommendations:

a. The confidential nature of the student academic information presents a challenge to the availability of the necessary operational data required to answer this challenge question. In order to obtain student academic information or grades, students must authorize the release of that information. Apparently there two possible options for acquiring the student academic performance data. One option is to request a data dump after the consent of each student to release term GPA information, after the successful completion of the alcohol abuse program. However, information release authorization process does not make this particular option feasible from an implementation perspective since the authorization must occur before the release of data in a batch mode. The second option is the use of a hard copy report or an electronic report for each student from the Registrar's office to the student health center with the consent of each student after the completion of the course. However, in contrast with the previous approach, this option releases term GPA data on an individual basis. Therefore, data needs to be entered manually or uploaded from the electronic report for each student.

b. A more realistic approach to capture trends on the impact of the alcohol abuse program on student academic performance would be to use a course evaluation form as the instrument to capture the students' voluntary feed back about academic performance after the completion of the course. For example, any of the questions below can be used to collect the required information about academic performance:

a. Did completion of this course help you achieving better academic performance within the same term? Academic year?

b. Did you achieve better academic performance (Term GPA) after the completion of this course? (Yes/No) Would you say that this course contributed to your achievement? (Yes/No). Explain how:____

c. Use Baskerville et al (2008), DSR framework to analyze risks of designing an artifact to solve the given challenge problem.

Appendix C – Analysis of Challenge Question #2

Show the total number of students from China, Japan, Korea, and Vietnam (or from other countries), that did not provide proof of immunization for measles, mumps, rubella, and hepatitis B at the time of registration during a span of seven years.

Purpose:

This question aims to motivate management to understand the benefits of leveraging historical student immunization requirement data towards the implementation of stricter compliance measures as a proactive approach to protect the student population's health.

Motivation:

This hypothetical question is motivated by the realization of the potential risk of pandemic outbreak in any higher education institution in the United States due to the number of foreign students enrolled not meeting the vaccination requirements established by the institution.

Object of analysis:

The object of analysis on this question is the compliance with student vaccination requirements (Number of exceptions of vaccination compliance) based on national origin. Understanding the trend of vaccination compliance among students of different countries might help on improving communication with students in regards to the importance of meeting immunization requirements established by the student health center and the state health agencies.

Required Data:

This hypothetical question requires the use of student or patient demographic data, specifically patient country and proof of vaccination. The necessary data to formulate a query to answer this question can be obtained from the student health center's practice management system and/or the electronic medical record system (EMR).

Key Stakeholders:

The following roles or group of users are considered to be key stakeholders: Student Health center director, and Clinical services director.

Proposed Value to Stakeholders:

This study proposes that the information provided by the analysis of vaccination compliance data be used to improve compliance monitoring activities that ultimately will help protecting student health, and provide management with evidence of the effectiveness of the immunization requirements program.

Issues and/or Limitations:

a. Students can question how data produced by this report can improve the quality of service rendered by the student health center on their behalf. The question might be perceived more as a tool to enforce standards rather than improving existing services.
b. Action taken based on inaccurate data outputs can lead stakeholders to serious consequences.
c. Immunization reporting criteria might not be significant to stakeholders.

Recommendations:

- Develop simple star schema model to demonstrate how to leverage existing immunization data.
- Use Baskerville et al DSR framework to analyze risks of proposed challenge problem.
- Communicate findings of artifact design and evaluation.

Appendix D – Analysis of Challenge Question #3

Provide data that supports the increase of cost for some services and the reduction of cost for other services.

Purpose:

For the purposes of this study, this question is intended to generate data related to services financial performance in order to analyze cost of service versus usage, and cost service versus revenue.

Motivation:

This hypothetical question is motivated by the need to prioritize the allocation of funds to areas of service that will potentially add value to patients (Canel & Fletcher, 2001, p. 260), (Kenagy et al, 1999, p. 664). While the literature did not provide any specific examples of how funds for health care services are allocated and prioritized (methods or approaches), this study proposes that comparing the number of patients serviced and the cost of supporting a particular service could help management make a decision on whether to keep or cancel a health care service.

Object of analysis:

This question does not provide enough information to help identify data required to support the cost increase of some services, and the cost reduction of others. However, for the purposes of demonstrating the use of multidimensional data modeling to measure performance of healthcare services, this study assumes that the analysis of the service fact can be measured in cost, revenue, and number of student serviced in a monthly or yearly basis. This analysis should provide management with data showing what services have been generating the most revenue based on the number of patients serviced. Also, the analysis should help management contrast revenue versus cost of service. Furthermore, services not deemed critical to the student population based on usage, and cost can be outsourced and/or cancelled. However, this is only a discussion example of a hypothetical approach to a more complex problem.

Required Data:

The data required to answer this hypothetical question should be extracted, transformed, and loaded from the Electronic Medical Record system to provide the number of students or patients serviced each month by service area, the total operating cost per service area, and the total revenue per service area.

Key Stakeholders:

The following roles or group of users are considered to be key stakeholders for this particular subject of analysis: Student Health center director, Clinical services director, and Finance manager.

Proposed Value to Stakeholders:

The facts being measured for this particular question provide a historical perspective on the number of patients serviced, and the cost of rendering the particular service. This can be contrasted to the revenues, and possibly related to the student's needs and preferences of service.

Issues and/or Limitations:

a. The initial secondary literature research did not provide evidence that support the approach to solve the challenge question and/or a standard method to guide the selection of relevant data.

b. However, the results of an Internet search led us to Berkley's web site (http://www.ocf.berkeley.edu/~health/faq.html#q8) where there is evidence that health services utilization data is used by some student health center organizations as a factor on cost of services decision (University of Berkeley Student Health Services, 2005).

Recommendations:

a. Develop simple star schema model to analyze cost, revenue, and health services utilization data.

b. Use Baskerville et al DSR framework to analyze risks of proposed challenge problem.

c. Ensure findings are communicated.

Appendix E – Analysis of Challenge Question #4

In order to demonstrate the benefit of OLAP reporting in support of decision making processes, this study presents the use of an episode of care data mart in addition to health care education program performance, immunization compliance, and health care services performance data marts.

Purpose:

The purpose of this initiative is to design and build a proof-of-concept artifact to analyze patient episode of care data extracted from the OLTP system based on a number of visits related to a primary diagnosis and a treatment.

Motivation:

The need for this artifact is motivated by the strong message presented in the literature suggesting that clinicians and/or physicians should be able to leverage historical clinical data to ultimately improve quality of clinical care (Ledbetter & Morgan, 2001), (Perdesen & Jensen, 1998).

Ledbetter and Morgan (2001), point out that clinicians, administrators, and researchers seeking to practice evidence-based medicine (outcome focused healthcare), require the aggregate analysis of clinical data for retrospective population-based studies (Ledbetter & Morgan, 2001, p. 121).

The paper presented by Parmanto et al (2005), describes a multidimensional database design for a data warehouse of healthcare rehabilitation outcomes to support various outcome analyses of outpatient rehabilitation therapies. The data warehouse designed in this study provides a tool for researchers to researchers to "create subsets of the total population and compare them statistically." (Parmanto et al, 2005, p. 3)

According to Parmanto et al (2005), patients are referred to the center by physicians based on a diagnosis or particular disability. The object of analysis in the artifact presented by Parmanto et al (2005) is the episode of care which consists of three to ten visits per week, and once the treatment is completed the patient's therapy ends. (Parmanto et al, 2005, p. 3)

According to Parmanto et al (2005), outcomes measured for healthcare processes can be divided into three grains (or levels) of information:

1. The episode-of-care outcome for general health measures
2. The episode-of-care outcome for disease-specific measures
3. Detailed outcome and treatment analysis for individual visits

The intention of this study is to construct an episode of care artifact to demonstrate how physicians can benefit from the analysis of episode of care data related to diseases and treatment.

According to Mehta, Suzuki, Glick, and Schulman (1999), the literature defines an episode of care as the period that starts after the patient is diagnosed with a clinical condition and ends when the condition has been resolved (Mehta et al, 1999).

An episode of care can be defined starting from the disease diagnosis and ending with the successful completion of the treatment as suggested by the works of Parmanto et al (2005) and Mehta et al (1999). The design of the artifact for the episode of care should be comprised of patient, diagnosis, treatment, date, and services dimensions. The fact table should include measures for number for the number of patient visits required for the completion of the patient's episode of care, and the time to complete the episode in terms of hours (Parmanto et al, 2005; Mehta et al, 1999).

Object of analysis:

The proposed object of analysis for this particular problem is the student's episode of care including but not limited to diagnosis, treatments, and measures like number of visits.

Required data:

The episode of care data mart requires data from the electronic medical record system (EMR), specifically patient, providers, diagnosis, and treatment data.

The application extracting, transforming, and loading data into the data mart must add an episode of care ID for every record entered into the fact table to identify and track each episode of care.

Key Stakeholders:

Physicians and clinical staff

Proposed Value to Stakeholders:

The physicians and practitioners can derive value from the episode of care artifact by analyzing historical data for "episode-of-care outcome for general health measures", "disease-specific measures", "detailed outcome and treatment analysis for individual visits" (Parmanto et al, 2005, p. 3).

Issues and/or Limitations:

a. Mapping data sources to an episode of care star schema attributes could be more complex than expected. This could render the proposed design useless.

b. Physicians may see the need to analyze patient laboratory result data in addition to the proposed dimensions.

c. The design is not including any insurance or billing information that could potentially be used by management to make critical funds allocation decisions.

d. The ICD-9 code: Updates to the code might be required.

e. Diagnosis dimension approach limitations: The literature analysis provided four different approaches for relating the diagnosis dimension to the fact table.

  - Pedersen and Jensen (Pedersen & Jensen, 1998): Proposed diagnosis group, diagnosis family, and diagnosis (low level); Use a grouping table with diagnosis valid dates.
  - Song et al (Song et al, 2001): Analyze different approaches but chose one-to-many relationship between dimension and fact tables.
  - Kimball and Ross (Kimball & Ross, 2002, p. 262): Use many-to-many join to diagnosis group bridge table.
  - Parmanto et al (Parmanto et al, 2005): Episode of care fact table relates to a diagnosis dimension including a column for ICD9 codes.

Each of the approaches listed above offers different advantages and disadvantages that could help meet the needs of the Physicians. However, for simplicity sake, the design will remain as presented.

a. No ICD-9 code attribute used for secondary diagnosis, this limits queries to only a name.
b. Physicians may not find any value proposed measures. However, other measures can be added if necessary.

Recommendations:

a. Develop an episode of care data mart assuming that the object of analysis will be the patient's episode of care from first visit to the last visit to follow up on treatment effectiveness.
b. Use Baskerville et al (2008), design science research risk management framework to analyze risks resulting from the use of design science research methodology.
c. Use Song et al (2001) denormalized non-positional diagnosis dimension without flag attributes "Method C-2: One-to-Many Relationship between Dimension and Fact tables" (Song et al, 2001, p. 6-9)

Appendix F - Restriction Levels

Classification of Restriction Levels

Restriction level	Applicable aggregate functions	Description
1	{SUM, AVG, MIN, MAX, STDDEV, VAR, COUNT}	"Given a pair (m;d) of a measure m and a dimension level d, we associate restriction level 1, if all aggregate functions may be applied to roll-up m from dimension level d to every functionally dependent higher level."
2	{AVG, MIN, MAX, STDDEV, VAR, COUNT}	"all aggregate functions but the SUM-Operator are applicable"
3	{COUNT}	"represents the highest limitation, where aggregation is still possible, but only in terms of counting"
4	{ }	"no aggregation function is permissible"

Classification of Restriction levels - levels of allowed aggregate functions (Husemann, et al, 6-10)

Appendix G – Health Education Artifact Functional Dependencies

Functional Dependencies between Terminal Dimension Levels and Measures

Fact schema	Measure	Dimension	Terminal dimension level
HealthEdFactTbl	HEdEnrollmentStatus HEdEnrollmentJustification HEdCourseEvaluationFrmCompletion ImprovedAcademicPerformance ImprovedHealthHabits	HealthEdProgramsDimTbl	HealthEdPrgID
		Term	TermID

(Hüsseman et al, 2000, p. 6-10)

Summarizability Appendix for Health Education Programs Fact Schema

Fact schema	Measure	Dimension levels	Restriction level
Health Ed Program Facts	HEdEnrollmentStatus	HealthEdPrgID TermID	1
	HEdEnrollmentJustification	HealthEdPrgID TermID	3
	HEdCourseEvaluationFrmCompletion	HealthEdPrgID TermID	1
	ImprovedAcademicPerformance	HealthEdPrgID TermID	1
	ImprovedHealthHabits	HealthEdPrgID TermID	1

(Hüsseman et al, 2000, p. 6-10)

Appendix H – Vaccination Artifact Functional Dependencies

Functional Dependencies between Terminal Dimension Levels and Measures

Fact schema	Measure	Dimension	Terminal dimension level
VaccinationFactTbl	VaccinationReqCompliance	VaccinationDimTbl TimeDimTbl	VaccinationID TimeID

(Hüsseman et al, 2000, p. 6-10)

Summarizability Appendix for Immunization Compliance Fact Schema

Fact schema	Measure	Dimension levels	Restriction level
Vaccination Facts	VaccinationReqCompliance	VaccinationID TimeID	1

(Hüsseman et al, 2000, p. 6-10)

Appendix I – Health Services Artifact Functional Dependencies

Functional Dependencies between Terminal Dimension Levels and Measures

Fact schema	Measure	Dimension	Terminal dimension level
ServiceFactTbl	ServiceRevenue, ServiceOptCost, ServicedPatients, ServiceRequests	Service	ServiceID
		Time	TimeID

(Hüsseman et al, 2000, p. 6-10)

Summarizability Appendix for Services Fact Schema

Fact schema	Measure	Dimension levels	Restriction level
Service facts	ServiceRevenue	ServiceID	1
		Year	2
		Month	2
	ServiceOptCost	ServiceID	1
		Year	2
		Month	2
	ServicedPatients	ServiceID	1
		Year	2
		Month	2
	ServiceRequests	ServiceID	1
		Year	2
		Month	2

(Hüsseman et al, 2000, p. 6-10)

Appendix J – Episode of Care Artifact Functional Dependencies

Functional Dependencies between Terminal Dimension Levels and Measures

Fact schema	Measure	Dimension	Terminal dimension level
EpisodeOfCareFactTbl	NumberOfEncounters, EpisodeStartDate, MinNumberOfEncounters	PatientDimTbl	PatientID
		DiagnosisDimTbl	DiagnosisID
		TreatmentDimTbl	TreatmentID
		DateDimTbl	Date
		ServicesDimTbl	ServiceID
		ProviderDimTbl	ProviderID

(Husemann et al, 2000, p. 6-10)

Summarizability Appendix for the Episode of Care Fact Schema

Fact schema	Measure	Dimension levels	Restriction level
Episode of care facts	NumberOfEncounters	EpisodeID	1
		Date	2
	EpisodeStartDate	EpisodeID	1
		Date	2
	MinNumberOfEncounters	EpisodeID	1
		Date	2

(Hüsseman et al, 2000, p. 6-10)

Appendix K – Health Education Programs Artifact Evaluation Queries

Q1.1 - What percentage of students showed GPA improvement after the successful completion of the alcohol abuse program?

First, we count all students enrolled in the Alcohol abuse course of 2003, having successfully completed the course with no academic performance improvement. See the SQL query below and the respective output.

```
SELECT HFT.HealthEdProgID, HEdPd.HealthEdProgName, TDT.Year,
COUNT(HFT.StudentID) AS 'No_Academic_Performance_Improvement'
FROM HealthEdProgramsDimTbl AS HEdPd INNER JOIN HealthEdFactTbl AS HFT
ON (HEdPd.HealthEdProgID = HFT.HealthEdProgID)
    INNER JOIN StudentDimTbl AS STD ON (STD.StudentID = HFT.StudentID)
INNER JOIN TimeDimTbl AS TDT ON (TDT.TimeID = HFT.TimeID)
WHERE  (HEdPd.HealthEdProgName ='Alcohol Abuse program - 03') AND
(TDT.Year = 2003) AND HFT.ImprovedAcademicPerformance = 0 AND
HFT.HEdEnrollmentStatus = 'Completed'
GROUP BY HFT.HealthEdProgID, HEdPd.HealthEdProgName, TDT.Year WITH
ROLLUP
ORDER BY HFT.HealthEdProgID, HEdPd.HealthEdProgName, TDT.Year
```

HealthEdProgID	HealthEdProgName	Year	No_Academic_Performance_Improvement
NULL	NULL	NULL	15
HED100103	NULL	NULL	15
HED100103	Alcohol Abuse program – 03	NULL	15
HED100103	Alcohol Abuse program – 03	2003	15

Second, we count all students enrolled in the Alcohol abuse course of 2003, having successfully completed the course, and had reported academic performance improvement after the course. See the SQL query below and the respective output.

```
SELECT HFT.HealthEdProgID, HEdPd.HealthEdProgName, TDT.AcademicYear,
COUNT(HFT.StudentID) AS 'Improved_Academic_Performance'
FROM HealthEdProgramsDimTbl AS HEdPd INNER JOIN HealthEdFactTbl AS HFT
ON (HEdPd.HealthEdProgID = HFT.HealthEdProgID)
    INNER JOIN StudentDimTbl AS STD ON (STD.StudentID = HFT.StudentID)
INNER JOIN TermsDimTbl AS TDT ON (TDT.TermID = HFT.TermID)
WHERE  (HEdPd.HealthEdProgName ='Alcohol Abuse program - 03') AND
(TDT.AcademicYear = 2003) AND HFT.ImprovedAcademicPerformance = 1 AND
HFT.HEdEnrollmentStatus = 'Completed'
GROUP BY HFT.HealthEdProgID, HEdPd.HealthEdProgName, TDT.AcademicYear
WITH ROLLUP
ORDER BY HFT.HealthEdProgID, HEdPd.HealthEdProgName, TDT.AcademicYear
```

Health Education Programs Performance Artifact Evaluation Queries - Continuation

HealthEdProgID	HealthEdProgName	AcademicYear	Improved_Academic_Performance
NULL	NULL	NULL	5
HED100103	NULL	NULL	5
HED100103	Alcohol Abuse program - 03	NULL	5
HED100103	Alcohol Abuse program - 03	2003	5

So, if we divide the number of students reporting “Improved_Academic_Performance” by the number of students that reported “No_Academic_Performance_Improvement” we obtain 0.33 %.

Q1.2 - How many students from Spain completed the alcohol abuse course?

```
SELECT HFT.HealthEdProgID, HEdPd.HealthEdProgName, TDT.AcademicYear,
COUNT(HFT.StudentID) AS 'No_Academic_Performance_Improvement'
FROM HealthEdProgramsDimTbl AS HEdPd INNER JOIN HealthEdFactTbl AS HFT
ON (HEdPd.HealthEdProgID = HFT.HealthEdProgID)
    INNER JOIN StudentDimTbl AS STD ON (STD.StudentID = HFT.StudentID)
INNER JOIN TimeDimTbl AS TDT ON (TDT.TermID = HFT.TermID)
WHERE (STD.StudentCountry = 'Spain') AND (HFT.HEdEnrollmentStatus =
'Completed') AND (HFT.ImprovedAcademicPerformance = 0)
GROUP BY HFT.HealthEdProgID, HEdPd.HealthEdProgName, TDT.AcademicYear
WITH ROLLUP
ORDER BY HFT.HealthEdProgID, HEdPd.HealthEdProgName, TDT.AcademicYear
```

HealthEdProgID	HealthEdProgName	AcademicYear	No_Academic_Performance_Improvement
NULL	NULL	NULL	1
HED100106	NULL	NULL	1
HED100106	Alcohol Abuse program - 06	NULL	1
HED100106	Alcohol Abuse program - 06	2006	1

Health Education Programs Performance Artifact Evaluation Queries - Continuation

Q1.3 - How many female students completed the alcohol abuse course each year?

```
SELECT HFT.HealthEdProgID, HEdPd.HealthEdProgName, TDT.AcademicYear,
COUNT(HFT.StudentID) AS 'Number_of_students'
FROM HealthEdProgramsDimTbl AS HEdPd INNER JOIN HealthEdFactTbl AS HFT
ON (HEdPd.HealthEdProgID = HFT.HealthEdProgID)
    INNER JOIN StudentDimTbl AS STD ON (STD.StudentID = HFT.StudentID)
INNER JOIN TimeDimTbl AS TDT ON (TDT.TermID = HFT.TermID)
WHERE (HEdPd.HealthEdProgName LIKE 'Alcohol Abuse program %') AND
(STD.StudentGender = 'F') AND HFT.HEdEnrollmentStatus = 'Completed'
GROUP BY HFT.HealthEdProgID, HEdPd.HealthEdProgName, TDT.AcademicYear
WITH ROLLUP
ORDER BY HFT.HealthEdProgID, HEdPd.HealthEdProgName, TDT.AcademicYear
```

HealthEdProgID	HealthEdProgName	AcademicYear	Number_of_students
NULL	NULL	NULL	44
HED100103	NULL	NULL	11
HED100103	Alcohol Abuse program - 03	NULL	11
HED100103	Alcohol Abuse program - 03	2003	11
HED100104	NULL	NULL	3
HED100104	Alcohol Abuse program - 04	NULL	3
HED100104	Alcohol Abuse program - 04	2004	3
HED100105	NULL	NULL	2
HED100105	Alcohol Abuse program - 05	NULL	2
HED100105	Alcohol Abuse program - 05	2005	2
HED100106	NULL	NULL	17
HED100106	Alcohol Abuse program - 06	NULL	17
HED100106	Alcohol Abuse program - 06	2005	2
HED100106	Alcohol Abuse program - 06	2006	15
HED100107	NULL	NULL	11
HED100107	Alcohol Abuse program - 07	NULL	11
HED100107	Alcohol Abuse program - 07	2007	11

Appendix L – Immunization Compliance Artifact Evaluation Queries

Q2.1 - Total number of students from China, Japan, Korea, and Vietnam that did not provide proof of immunization for measles, mumps, rubella, and hepatitis B at the time of registration during a span of seven years.

Q2.1.1 - Basic Query of students requiring immunization (China, Vietnam,Korea, Japan)

```
SELECT TimeDimTbl.Year, StudentDimTbl.StudentCountry,
VaccinationDimTbl.VaccineName, COUNT(VaccinationFactTbl.StudentID) AS
StudentsRequiringImmunization
FROM VaccinationDimTbl INNER JOIN (TimeDimTbl INNER JOIN (StudentDimTbl
INNER JOIN VaccinationFactTbl ON StudentDimTbl.StudentID =
VaccinationFactTbl.StudentID) ON TimeDimTbl.TimeID =
VaccinationFactTbl.TimeID) ON VaccinationDimTbl.VaccinationID =
VaccinationFactTbl.VaccinationID
WHERE StudentDimTbl.StudentCountry IN
('China','Vietnam','Korea','Japan')
        GROUP BY TimeDimTbl.Year, StudentDimTbl.StudentCountry,
                     VaccinationDimTbl.VaccineName
```

Year	StudentCountry	VaccineName	StudentsRequiringImmunization
2003	China	Measles	2
2003	China	Mumps	1
2003	Japan	Hepatitis B	1
2003	Japan	Rubella	1
2003	Korea	Rubella	1
2003	Vietnam	Hepatitis B	1
2004	China	Hepatitis B	2
2004	Japan	Rubella	1
2005	China	Mumps	1
2005	Vietnam	Mumps	1

Immunization Compliance Artifact Evaluation Queries - Continuation

Q2.1.2 - Query showing students complying with vaccination requirements

```
SELECT TimeDimTbl.Year, StudentDimTbl.StudentCountry,
VaccinationDimTbl.VaccineName, COUNT(VaccinationFactTbl.StudentID) AS
In_Compliance
FROM VaccinationDimTbl INNER JOIN (TimeDimTbl INNER JOIN (StudentDimTbl
INNER JOIN VaccinationFactTbl ON StudentDimTbl.StudentID =
VaccinationFactTbl.StudentID) ON TimeDimTbl.TimeID =
VaccinationFactTbl.TimeID) ON VaccinationDimTbl.VaccinationID =
VaccinationFactTbl.VaccinationID
WHERE StudentDimTbl.StudentCountry IN
('China','Vietnam','Korea','Japan') AND
(VaccinationFactTbl.VaccinationReqCompliance=1)
GROUP BY TimeDimTbl.Year, StudentDimTbl.StudentCountry,
VaccinationDimTbl.VaccineName
```

Year	StudentCountry	VaccineName	In_Compliance
2003	China	Measles	2
2003	Japan	Hepatitis B	1
2003	Japan	Rubella	1
2003	Vietnam	Hepatitis B	1
2004	China	Hepatitis B	2
2004	Japan	Rubella	1
2005	Vietnam	Mumps	1

Query showing students not in compliance with vaccination requirements

```
-- Query#3
-- Basic Query (China, Vietnam, Korea, Japan)
-- No compliance with immunization requirements
SELECT TimeDimTbl.Year, StudentDimTbl.StudentCountry,
VaccinationDimTbl.VaccineName, COUNT(VaccinationFactTbl.StudentID) AS
No_Compliance
FROM VaccinationDimTbl INNER JOIN (TimeDimTbl INNER JOIN (StudentDimTbl
INNER JOIN VaccinationFactTbl ON StudentDimTbl.StudentID =
VaccinationFactTbl.StudentID) ON TimeDimTbl.TimeID =
VaccinationFactTbl.TimeID) ON VaccinationDimTbl.VaccinationID =
VaccinationFactTbl.VaccinationID
WHERE StudentDimTbl.StudentCountry IN
('China','Vietnam','Korea','Japan') AND
(VaccinationFactTbl.VaccinationReqCompliance=0)
GROUP BY TimeDimTbl.Year, StudentDimTbl.StudentCountry,
VaccinationDimTbl.VaccineName
```

Year	StudentCountry	VaccineName	No_Compliance
2003	China	Mumps	1
2003	Korea	Rubella	1
2005	China	Mumps	1

Immunization Compliance Artifact Evaluation Queries - Continuation

Q2.2 – Total number of US students that did not provide proof of validations.

```
SELECT TimeDimTbl.Year, PatientDimTbl.PatientCountry,
VaccinationDimTbl.VaccineName, COUNT(VaccinationFactTbl.PatientID) AS
Total_Exceptions
FROM VaccinationDimTbl INNER JOIN (TimeDimTbl INNER JOIN (PatientDimTbl
INNER JOIN VaccinationFactTbl ON PatientDimTbl.PatientID =
VaccinationFactTbl.PatientID) ON TimeDimTbl.TimeID =
VaccinationFactTbl.TimeID) ON VaccinationDimTbl.VaccinationID =
VaccinationFactTbl.VaccinationID
WHERE (((VaccinationFactTbl.VaccinationReqCompliance)=0)) AND
(PatientDimTbl.PatientCountry = 'US')
GROUP BY TimeDimTbl.Year, PatientDimTbl.PatientCountry,
VaccinationDimTbl.VaccineName
WITH ROLLUP;
```

Year	PatientCountry	VaccineName	Total_Exceptions
2003	US	Hepatatis A	1
2003	US	Mumps	4
2003	US	Rubella	3
2003	US	NULL	8
2003	NULL	NULL	8
2004	US	H1N1 (Swine Flu)	1
2004	US	Hepatitis B	1
2004	US	NULL	2
2004	NULL	NULL	2
2005	US	Hepatitis B	2
2005	US	NULL	2
2005	NULL	NULL	2
2008	US	Hepatitis B	1
2008	US	NULL	1
2008	NULL	NULL	1
NULL	NULL	NULL	13

Immunization Compliance Artifact Evaluation Queries - Continuation

Q2.3 – Query to show the number of H1N1 female students per country.

```
-- Q2.3 - Show the number of female students per country that got the
H1N1 vaccine each month of the year 2008.

SELECT TimeDimTbl.Year, TimeDimTbl.Month, PatientDimTbl.PatientCountry,
VaccinationDimTbl.VaccineName, COUNT(VaccinationFactTbl.PatientID) AS
In_Compliance
FROM VaccinationDimTbl INNER JOIN (TimeDimTbl INNER JOIN (PatientDimTbl
INNER JOIN VaccinationFactTbl ON PatientDimTbl.PatientID =
VaccinationFactTbl.PatientID) ON TimeDimTbl.TimeID =
VaccinationFactTbl.TimeID) ON VaccinationDimTbl.VaccinationID =
VaccinationFactTbl.VaccinationID
INNER JOIN ProvidersDimTbl ON VaccinationFactTbl.ProviderID =
ProvidersDimTbl.ProviderID
WHERE (((VaccinationFactTbl.VaccinationReqCompliance)=1)) AND
(VaccinationDimTbl.VaccineName = 'H1N1 (Swine Flu)') AND
(PatientDimTbl.PatientGender = 'F') AND TimeDimTbl.Year = 2008
GROUP BY TimeDimTbl.Year, TimeDimTbl.Month,
PatientDimTbl.PatientCountry, VaccinationDimTbl.VaccineName
WITH ROLLUP;
```

Year	Month	PatientCountry	VaccineName	In_Compliance
2008	3	US	H1N1 (Swine Flu)	1
2008	3	US	NULL	1
2008	3	NULL	NULL	1
2008	5	US	H1N1 (Swine Flu)	1
2008	5	US	NULL	1
2008	5	NULL	NULL	1
2008	NULL	NULL	NULL	2
NULL	NULL	NULL	NULL	2

Appendix M – Health Services Performance Artifact Evaluation Queries

Q3.1 - Show the number of patients that received contraceptive services broken down by gender per year.

```
SELECT [ServicesDimTbl].servicename,
SUM([ServiceFactTbl].ServicedPatients) AS Patients_Serviced,
SUM([ServiceFactTbl].ServiceRequests) AS Service_Requests,
SUM([ServiceFactTbl].ServicedFemPatients) AS Female,
SUM([ServiceFactTbl].ServicedMalePatients) AS Male,[TimeDimTbl].Year
FROM [ServicesDimTbl] INNER JOIN [ServiceFactTbl] ON
[ServicesDimTbl].serviceID = [ServiceFactTbl].serviceID INNER JOIN
[TimeDimTbl]
ON [ServiceFactTbl].TimeID = [TimeDimTbl].TimeID
WHERE [ServicesDimTbl].servicename = 'Contraceptive services'
GROUP BY [ServicesDimTbl].servicename, [TimeDimTbl].Year WITH CUBE
```

servicename	Patients_Serviced	Service_Requests	Female	Male	Year
Contraceptive services	243	243	171	72	2006
Contraceptive services	243	243	171	72	NULL
NULL	243	243	171	72	NULL
NULL	243	243	171	72	2006

Q3.2 - Show the number of HIV screening cases by gender per month:

```
SELECT [ServicesDimTbl].servicename,
SUM([ServiceFactTbl].ServicedPatients) AS Patients_Serviced,
SUM([ServiceFactTbl].ServiceRequests) AS Service_Requests,
SUM([ServiceFactTbl].ServicedFemPatients) AS Female,
SUM([ServiceFactTbl].ServicedMalePatients) AS Male,[TimeDimTbl].Month
FROM [ServicesDimTbl] INNER JOIN [ServiceFactTbl] ON
[ServicesDimTbl].serviceID = [ServiceFactTbl].serviceID INNER JOIN
[TimeDimTbl]
ON [ServiceFactTbl].TimeID = [TimeDimTbl].TimeID
WHERE [ServicesDimTbl].servicename = 'HIV Screening services'
GROUP BY [ServicesDimTbl].servicename, [TimeDimTbl].Month WITH CUBE
```

Health Services Performance Artifact Evaluation Queries - Continuation

servicename	Patients_Serviced	Service_Requests	Female	Male	Month
HIV Screening Services	20	20	14	6	1
HIV Screening Services	30	30	24	6	2
HIV Screening Services	20	20	14	6	3
HIV Screening Services	21	21	15	6	4
HIV Screening Services	20	20	14	6	5
HIV Screening Services	30	30	24	6	6
HIV Screening Services	20	20	14	6	7
HIV Screening Services	21	21	15	6	8
HIV Screening Services	20	20	14	6	9
HIV Screening Services	30	30	24	6	10
HIV Screening Services	20	20	14	6	11
HIV Screening Services	21	21	15	6	12
HIV Screening Services	273	273	201	72	NULL
NULL	273	273	201	72	NULL
NULL	20	20	14	6	1
NULL	30	30	24	6	2
NULL	20	20	14	6	3
NULL	21	21	15	6	4
NULL	20	20	14	6	5
NULL	30	30	24	6	6
NULL	20	20	14	6	7
NULL	21	21	15	6	8
NULL	20	20	14	6	9
NULL	30	30	24	6	10
NULL	20	20	14	6	11
NULL	21	21	15	6	12

Appendix N - Episode of Care Artifact Evaluation Queries

Q4.1 - Show all primary diagnosis related to sexually transmitted diseases cases per year and month for female students.

```
SELECT TimeDimTbl.Year, TimeDimTbl.Month,
DiagnosisDimTbl.Primary_Diagnosis,
COUNT(EpisodeOfCareFactTbl.DiagnosisID) AS NumberOfDiagnosis
FROM PatientDimTbl INNER JOIN (DiagnosisDimTbl INNER JOIN (TimeDimTbl
INNER JOIN EpisodeOfCareFactTbl ON
TimeDimTbl.TimeID=EpisodeOfCareFactTbl.TimeID) ON
DiagnosisDimTbl.DiagnosisID=EpisodeOfCareFactTbl.DiagnosisID) ON
PatientDimTbl.PatientID=EpisodeOfCareFactTbl.PatientID
WHERE (((EpisodeOfCareFactTbl.DiagnosisID)=DiagnosisDimTbl.DiagnosisID)
And (EpisodeOfCareFactTbl.TimeID=TimeDimTbl.TimeID))
AND EpisodeOfCareFactTbl.DiagnosisID IN (7005, 7011, 7004) AND
PatientDimTbl.PatientGender = 'F'
GROUP BY TimeDimTbl.Year, TimeDimTbl.Month,
DiagnosisDimTbl.Primary_Diagnosis
WITH ROLLUP;
```

Year	Month	Primary_Diagnosis	NumberOfDiagnosis
2003	2	Encounters for testing for HIV	1
2003	2	Herpes Simplex NOS	1
2003	2	HIV	3
2003	2	NULL	5
2003	NULL	NULL	5
2004	2	Encounters for testing for HIV	1
2004	2	Herpes Simplex NOS	1
2004	2	NULL	2
2004	NULL	NULL	2
2005	2	Encounters for testing for HIV	1
2005	2	Herpes Simplex NOS	1
2005	2	NULL	2
2005	NULL	NULL	2
2007	2	Encounters for testing for HIV	1
2007	2	NULL	1
2007	3	Encounters for testing for HIV	1
2007	3	Herpes Simplex NOS	2
2007	3	NULL	3
2007	NULL	NULL	4
2008	2	Encounters for testing for HIV	1
2008	2	Herpes Simplex NOS	1
2008	2	HIV	1
2008	2	NULL	3
2008	3	Herpes Simplex NOS	1
2008	3	HIV	1

Episode of Care Artifact Evaluation Queries – Continuation

2008	3	NULL	2
2008	NULL	NULL	5
NULL	NULL	NULL	18

Q4.2 - Show the number of HIV diagnosis for every year for male students.

```
SELECT DiagnosisDimTbl.Primary_Diagnosis,
COUNT(EpisodeOfCareFactTbl.DiagnosisID) AS TotalNumberOfDiagnosis,
TimeDimTbl.Year
FROM PatientDimTbl INNER JOIN (DiagnosisDimTbl INNER JOIN (TimeDimTbl
INNER JOIN EpisodeOfCareFactTbl ON
TimeDimTbl.TimeID=EpisodeOfCareFactTbl.TimeID) ON
DiagnosisDimTbl.DiagnosisID=EpisodeOfCareFactTbl.DiagnosisID) ON
PatientDimTbl.PatientID=EpisodeOfCareFactTbl.PatientID
WHERE DiagnosisDimTbl.DiagnosisID = 7004 AND
PatientDimTbl.PatientGender = 'M'
GROUP BY DiagnosisDimTbl.Primary_Diagnosis, TimeDimTbl.Year
WITH CUBE;
```

Primary_Diagnosis	TotalNumberOfDiagnosis	Year
HIV	1	2007
HIV	1	2008
HIV	2	NULL
NULL	2	NULL
NULL	1	2007
NULL	1	2008

Q4.3 - Show the number of encounters for HIV testing on female patients per year for each month for years 2006, 2007, 2008.

```
SELECT TimeDimTbl.Year, TimeDimTbl.Month,
DiagnosisDimTbl.Primary_Diagnosis,
COUNT(EpisodeOfCareFactTbl.DiagnosisID) AS TotalNumberOfDiagnosis
FROM PatientDimTbl INNER JOIN (DiagnosisDimTbl INNER JOIN (TimeDimTbl
INNER JOIN EpisodeOfCareFactTbl ON
TimeDimTbl.TimeID=EpisodeOfCareFactTbl.TimeID) ON
DiagnosisDimTbl.DiagnosisID=EpisodeOfCareFactTbl.DiagnosisID) ON
PatientDimTbl.PatientID=EpisodeOfCareFactTbl.PatientID
WHERE DiagnosisDimTbl.DiagnosisID = 7005 AND
PatientDimTbl.PatientGender = 'F' AND TimeDimTbl.Year IN (2006, 2007,
2008)
GROUP BY TimeDimTbl.Year, TimeDimTbl.Month,
DiagnosisDimTbl.Primary_Diagnosis
WITH CUBE;
```

Episode of Care Artifact Evaluation Queries – Continuation

Primary_Diagnosis	TotalNumberOfDiagnosis
Appendicitis	6
Chickenpox, NOS	5
Chronic Obstructive Pulmonary Disease [COPD] and Asthma	6
Encounters for testing for HIV	9
H1N1 (Swine Flu)	9
Hepatitis-Infectious	5
Herpes Simplex NOS	10
HIV	7
Knee	6
Meningitis	9
Rocky mountain spotted fever	6
Sacrum Coccyx SCI	7
Sinusitis, Chronic (Sinus Infection)	8
Viral conjunctivitis	7
NULL	100

Appendix O – Sample Health Services Performance Artifact Evaluation Reports

Services Utilization Report

	2003	2004	2005	2006	2007	2008
Contraceptive services				243		
Dental services		331				
Immunization services	311					
Men's Services					283	
Nutrition counseling services			281			
Optometry Services						273

Health Services Revenue Summary

	2003	2004	2005	2006	2007	2008
Contraceptive services				111800		
Dental services		116000				
Immunization services	113800					
Men's Services					120700	
Nutrition counseling services			107800			
Optometry Services						113700

Appendix P – Sample Microsoft Excel Pivot Table Report (Oracle Database Data)

Sum of Serviceoptcost	Year SUM					
Servicename	2003	2004	2005	2006	2008	Grand Total
Contraceptive services				$65,800.00		$65,800.00
Dental services		$77,600.00				$77,600.00
Immunization services	$69,600.00					$69,600.00
Nutrition counseling services			$69,600.00			$69,600.00
Optometry Services					$70,800.00	$70,800.00
Grand Total	$69,600.00	$77,600.00	$69,600.00	$65,800.00	$70,800.00	$353,400.00

Sum of Serviceoptcost	Year SUM			
Servicename	2003	2004	2005	Grand Total
Dental services		$77,600.00		$77,600.00
Immunization services	$69,600.00			$69,600.00
Nutrition counseling services			$69,600.00	$69,600.00
Grand Total	$69,600.00	$77,600.00	$69,600.00	$216,800.00

Appendix Q - Sample Health Services Performance HTML Report (Oracle Reporting Services)

SERVICENAME	YEAR	Operating Cost	Service Revenue
Contraceptive services	2006	$65800	$111800
Contraceptive services		$65800	$111800
Dental services	2004	$77600	$116000
Dental services		$77600	$116000
Immunization services	2003	$69600	$113800
Immunization services		$69600	$113800
Nutrition counseling services	2005	$69600	$107800
Nutrition counseling services		$69600	$107800
Optometry Services	2008	$70800	$113700
Optometry Services		$70800	$113700
	2003	$69600	$113800
	2004	$77600	$116000
	2005	$69600	$107800
	2006	$65800	$111800
	2008	$70800	$113700
		$353400	$563100

Annotated Bibliography

Glaser, J., & Stone, J. (2008, February). Effective Use of Business Intelligence. Healthcare Financial Management, 62(2):68-72.

Glaser and Stone claim that effective business intelligence by managing and leveraging the organization's data can help enhance financial and operational performance. The authors present five preliminary steps that lead to effective BI efforts: Establish business needs and value. Obtain buy-in from managers. Create an end-to-end vision. Establish BI governance. Implement specific roles for managing data quality. Glaser and Stone provide a valuable advice in regards the implementation of BI in their paper. They point out that it "is not practical to implement BI technologies and disciplines across the organization all at once." Some insight about common mistakes healthcare organizations make when implementing BI are presented which are applicable to the student health center.

Ledbetter, C.S., & Morgan, M.W. (Summer 2001). Towards Best Practice: Leveraging the Electronic Patient Record as a Clinical Data Warehouse. Journal of Healthcare Information Management, 5, 2.

According to Ledbetter and Morgan, multidimensional information can help healthcare improve clinical practice and the respective evaluation methods. Their research work describes how outcome-focused healthcare organizations can support best practices by using a clinical data warehouse ROLAP based system that contains clinical processes result data integrated to the electronic patient record (EPR) designed to automate clinical and administrative processes. The authors compare and contrast both the electronic patient record system (EPR) and the clinical decision (CDS) support system requirements and functionalities, and point out the importance of the clinical data warehouse component of their business intelligence solution. The work of Ledbetter and Morgan contributes to the domain of healthcare business intelligence by presenting a proof of concept case study of the clinical data warehouse solution. Key concepts of OLTP and OLAP integration are reviewed in this paper.

Parmanto, B., Scotch, M., & Ahmad, S. (2005, Summer). A Framework for Designing a Healthcare Outcome Data Warehouse. Perspectives in Health Information Management 2, 3.

Parmanto et al proposed a variation of the dimensional modeling technique in the context of rehabilitation services and implemented a data warehouse that provides three levels of grain (scope) making possible different levels of analysis. The paper presents a case study of a multidimensional database design for a data warehouse of healthcare rehabilitation outcomes at the University of Pittsburgh Medical Center. According to Parmanto et al (2005), the multidimensional database design can be used as a blueprint for the development of a data warehouse for healthcare decision support.

Hristovski, D., Rogaè, M., & Markota, M. (2000). Using Data Warehousing and OLAP in Public Health Care. Journal of American Medical Informatics Association, 369-73.

This paper presents the implementation process of an outpatient data warehouse and OLAP to generate reports previously produced with traditional statistical tools. Hristovski et al point out that data warehousing and OLAP are suitable for the domain of public health, and can enable “new analytical possibilities in

addition to the traditional statistical approaches." This paper contributes to support the idea of the use of OLAP technology in the healthcare information management. This paper does not elaborate only on OLAP as a data analysis tool to generate reports. No other architectures are considered in this work.

Pedersen, T.B., & Jensen, C.S. (1998). Clinical Data Warehousing – A Survey, Proceedings of the MEDICON 98 conference.

Pedersen and Jensen survey and elaborate on the developments on clinical data warehousing. The authors discuss how the clinical data warehouse can be used in a clinical setting, and point out that the data warehouse provides "an explorative way of working with the data" in comparison with an "ordinary" clinical decision support system (CDSS). This work presents a list some of the requirements a clinical data warehouse should meet in order to add value to the business. Pedersen and Jensen point out, that clinicians will eventually realize the potential benefits of using data warehouses to leverage existing data in order to improve quality and research.

Tremblay, M. C., Fuller, R., Berndt, & D., Studnicki, J. (2007, August). Doing more with more information: Changing healthcare planning with OLAP tools. Decision Support Systems, 43, 4, 1305-1320.

Tremblay et al present a qualitative field study demonstrating how OLAP decision support capabilities helped knowledge workers expand and enhanced their tasks. Health planners carried out activities to acquire, create, and disseminate knowledge to decision makers from different data sources, used hard copy reports from the existing data warehouse system, and data from other

agencies (Tremblay et al, 1312). In this paper OLAP functionality is introduced in a health planning agency in Florida to provide data analysis and presentation. Tremblay et al make a valuable contribution to the study of OLAP tool implementation within healthcare by describing how the users embrace the new functionality.

Canel, C., & Anderson Fletcher, E. A. (2001). An analysis of service quality at a student health center. International Journal of Health Care Quality Assurance, 14/6, 260-267.

Canel and Fletcher provide an analysis of the quality of service provided by a University student health center as the result of a student survey. Canel and Fletcher point out that quality management is an important issue for both the academic community and the health care practitioners. Canel and Fletcher also point out that health care administrators within the University health center environment are not only challenged with contain costs, but also with not letting the quality of care degrade. According to Canel and Fletcher, research shows that patients can provide valuable information about health care service delivery expectations and perceptions. Canel and Fletcher also point that that Parasuraman et al “suggest that management policies and processes can affect customer satisfaction”, and they further suggest the existence of gaps between management and customer perceptions in regards to health care service performance and expectations. The work presented by Canel and Fletcher provides an excellent example about student’s perceptions in regards to quality of service in the student health center.

Perdesen, T. B., & Jensen, C. S. (1998). Research Issues in Clinical Data Warehousing. Proceedings of SSDBM'98, Jul. 1-3.

Pedersen and Jensen present an excellent discussion about research challenges faced by the data base research community originated from clinical applications. Some of the challenges presented by Pedersen and Jensen include the need for complex-data modeling features, advanced temporal support, advanced classification structures, continuously valued data, dimensionally reduced data, and the integration of very complex data. Pedersen and Jensen explain the issue of the diagnosis dimension hierarchy and "correct aggregation in the case of non-strict hierarchies." This should be a consideration in this project since the goal is to develop an episode of care data mart including patient diagnosis data. This work helps understanding the characteristics and features that make a clinical data warehouse (CDW) different than a typical data warehouse.

Bréant, C., Thurler, G., Borst, F., & Geissbuhler, A. (2005). Design of a Multi-Dimensional Database for the Archimed Data Warehouse. Studies in health technology and informatics, 116, 169-174.

Bréant et al, describe and discuss three design aspects relevant to conception of the Archimed data warehouse intended to facilitate the access to patient medical data in an integrated and coherent fashion with the purpose to support data analysis, and data mining (Bréant et al, 169). Bréant et al, describe the granularity, model and architecture, and the life cycle of the database. The hospital data warehouse solution presented in this study is achieved by progressively integrating different data marts designed based on a format called the standardized elementary fact that allows for the same level of detail for facts from different domains. The data warehouse design implemented in this study interconnects the data marts through conformed dimension tables. Another important aspect Archimed data warehouse design presented by Bréant et al is that of the evolution. The work of Bréant et al is relevant to this project because it provides an example of a clinical data warehouse implementation through data marts and the use of different fact tables, and capable of handling the evolution of the system.

Berndt, D. J., & Hevner, A. R. (2000). Hospital Discharge Transactions: A Data Warehouse Component. Proceedings of the 33rd Hawaii International Conference on System Sciences.

The work of Berndt and Hevner describes the implementation of the CATCH project based on the CATCH methodology. Berndt et al discuss the design, methodology, data quality, and the twin-star data staging for “data transformation, quality checks, and simple reports” (Berndt et al, 61). The data warehouse solution implemented was designed to aggregate data from heterogeneous data sources into a staging area for transformation and data quality procedures before moving into a permanent data structure. Also, a simplified example of the hospital discharge star schema is shown in this paper illustrating some dimensions relevant to this study. The “twin star data staging” design approach presented in this offers great possibilities to the development of an ETL process for the student health center data warehouse. This work is relevant to this study in that it provides an implementation case study within the health care domain.

Akhtar, M.U., Dunn, K., & Smith, J.W. (2005). Commercial clinical data warehouses: from wave of the past to the state of the art, Journal of Healthcare Information Management, 19, 20–26.

Akhtar et al, review the history of commercial clinical data warehousing and present the results of a survey about current clinical data warehousing trends. According to Akhtar et al, both the analysis of the historical perspective and the survey results support the fact that healthcare organizations have reduced efforts towards an in-house clinical data warehouse to favor commercial products. This overview of clinical data warehouses provides insight about in-house versus commercial clinical data warehouse adoption considerations, trends, and vendors. The administration of the student health center and physicians might be interested on knowing what commercial clinical data warehouse products are available and why they have been adopted by health care institutions. It could be feasible for the student health center to adopt a commercial clinical data warehouse product as long as their analysis needs are met.

Pedersen T. B., & Jensen C. (1998). Clinical Data Warehousing - A Survey. Proceedings of the MEDICON 98 conference.

Pedersen and Jensen describe some initiatives for clinical data warehousing from a non-exhaustive list of some commercial data warehouse products and clinical data warehouse projects, and provide a comparison criteria. This paper of what clinical data warehouse systems vendors were available in the market at the time of the writing of the paper. However, from Pedersen and Jensen comparison we can learn that some of the systems presented are for “drug development” (Oracle

Clinical & SAS) and "disease studies". Any of those systems could be an expensive solution for the student health center. This study discusses important features that a clinical data warehouse should support (billing, contracts, demographic data, diagnoses, and procedures), and the requirements for data analysis. A discussion about the importance of supporting different levels of data analysis including patient level, group level, and the healthcare enterprise level is presented. These are important factors to consider in the artifact design process.

Gordon, B. D., Asplin, B. R. (2004, November). Using Online Analytical Processing to Manage Emergency Department Operations. ACAD EMERG MED, 11, 11.

According to Gordon and Asplin, OLAP software summarizes data from different data sources to allow managers, providers, and researchers "to examine patterns and trends in operations and patient flow." The authors present a discussion on information and data needs within an emergency department operations workflow that has similar needs like the ones often encounter by management and staff at the student health center. In their discussion Gordon and Asplin mention the issue of ED operation throughput and list some factors that might contribute to operational delays (staffing, available treatment areas, and length of time for clinical care operations among others). Gordon and Asplin also discuss some of the challenges of implementing an OLAP solution. According to Gordon and Asplin, resource limitations, lack of understanding about levels of data integration, lack of data sharing, and data cleaning as barriers to the implementation or adoption of OLAP in the context of ED.

Verma, R., & Harper, J., (2001). Life Cycle of a Data Warehousing Project in Healthcare. Journal of Healthcare Information Management.

Verma and Harper discuss the construction and implementation of a decision support system based on a data mart architecture by integrating different data sources, and providing analysis capabilities for management and other decision makers. Verma and Harper point out that the system was developed using Oracle PL/SQL, Oracle Discoverer, and data source included file from Microsoft Excel and Microsoft Access databases. In their discussion, Verma and Harper also point out that part of the motivation for the initiative was the need to improve the existing reporting capabilities. In addition to the reporting capabilities, the need to integrate different islands of information was also a contributing factor on initiating the project. This case study presents different implementation aspects relevant to this study. Specifically, Verma and Harper point out that a feasibility study was conducted at an early stages of the initiative that helped on prioritizing the subject areas in need of analytical capabilities (Verma and Herpa, 111).

Palaniappan, S., & Sook Ling, C. (2008, September). Clinical Decision Support Using OLAP with Data Mining. IJCSNS International Journal of Computer Science and Network Security, 8, 9.

Palaniappan and Ling present a prototype that integrates an OLAP and data mining clinical decision support model for analysis of patient data. The objective of the prototype presented in this study is to design for diabetes, heart, and liver disorder training databases. Palaniappan and Ling contrast OLAP and data mining and demonstrate how an integrated approach provides advanced decision support

not available or possible with individual OLAP or data mining features. The work of Palaniappan and Ling is relevant to this study in that we have a case study we can follow including the logical and physical structures, the discussion about the multidimensional model used, the research questions, and the implementation of the solution.

Song, I-Y., Rowen, W., Medsker, C., & Ewen, E. (2001, June 4). An Analysis of Many-to-Many Relationships Between Fact and Dimension Tables in Dimensional Modeling. Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2001), Switzerland.

Song et al, present different approaches to many-to-many relationships between fact and dimension tables illustrated by a healthcare diagnosis billing-encounter schema. The diagnosis dimension problem is that of a patient having more than one diagnosis for each billable encounter. Song et al study analyzes previous works to determine the best approach to handle many-to-many relationships, discuss their advantages and disadvantages. The authors propose two ad-hoc methods that maintain a star schema structure by de-normalizing the dimension to avoid many-to-many relationships. This study intends to use Song et al study to analyze the design options, and to address the need for the episode of care multidimensional model used to answer the given hypothetical question.

Pedersen, T.B., & Jensen, C.S. (1999). Multidimensional Data Modeling for Complex Data. In Proceedings of 15th Int. Conf. on Data Engineering (ICDE}, IEEE Computer Society, 336-345.

Pedersen and Jensen present a discussion about existing multidimensional data models and a series of unmet requirements for OLAP applications. The results of a survey presented in this study reveal the lack of “support for many-to-many relationships between facts and dimensions, built-in support for handling change and time, and support for uncertainty as well as different levels of granularity in the data.” Pedersen and Jensen make use of a case study on patients diagnosis related to their place of residence to investigate environmental or lifestyle factors in order to present the features that a multidimensional data model should have to meet complex data requirements (Pedersen & Jensen, 1998).

Ewen, E. F., Medsker, C. E., & Dusterhoft, L. E. (1998). Data Warehousing in an Integrated Health System; Building the Business Case. DOLAP ’98, Washington, DC.

Ewen et al present the approach taken for the implementation of a data warehouse in a healthcare setting with the aim to reach better integration levels among functional units in order to cope with a rapidly evolving environment. Ewen et al, discuss how the need for a data warehouse was established or defined, how the key business areas were identified, and describe the results from the data sources inventory conducted for the project. The work of Ewen et al discuss how an OLAP prototype was used to present inpatient statistics, and emergency department utilization, and show data aggregations representing high volume utilization of diseases posing “high cost risks.” The work of Ewen et al, presents a relevant example of how to approach a data warehouse project. This paper presents important points to consider in this project.

Berndt, D.J., Fisher, J.W., Hevner, A.R., & Studnicki, J. (2001, December). Healthcare Data Warehousing and Quality Assurance. Computer, 34, 12, 56-65.

Berndt et al, present an overview of the research and development decisions made in constructing the CATCH healthcare data warehouse, focused on the topics of data staging and quality assurance. Berndt et al discuss the design, data aspects, the catch methodology, data quality and error handling. The discussion about the CATCH data warehouse design describes the star schema used for the multidimensional structure, the three levels of granularity provided by the structure that allows for different levels of data reporting, and the use of the data warehouse intended for reporting and research needs. The work of Berndt et al is relevant to this research project in that it provides an actual implementation case study that could be used to design and evaluate artifact solutions. Among many of the concepts discussed in this paper we can identify as relevant the discussion about the CATCH methodology as an approach for data format translation and integration to help us understand how a staging area (staging star) can be used as component of an enterprise data warehouse architecture.

Sahama, T.R., & Croll, P.R., (2007, January). A Data Warehouse Architecture for Clinical Data Warehousing. ACSW '07: Proceedings of the fifth Australasian symposium on ACSW frontiers - Volume 68 Publisher: Australian Computer Society, Inc.

Sahama and Croll present a practical solution to the implementation of a Clinical Data Warehouse. Sahama and Croll implement some of the data warehousing methodologies presented by Sen and Sinha and discuss why some of the options were not acceptable for their particular project. Sahama and Croll describe their approach to build a clinical data warehouse for oncology patients using SAS© Warehouse Administrator and the issues related to experimented with different combinations of Sen and Sinha data warehouse architectures. Sahama and Croll decided to implement an Enterprise data warehouse with an operational data store architecture and a distributed data warehouse architecture.

Wang, L., Zhang, A., & Ramanathan, M. (2005). BioStar models of clinical and genomic data for biomedical data warehouse design. Int. Journal of Bioinformatics Research and Applications.

Wang et al, present the BioStar clinical and genomic data approach for use in the Biological domain. Wang et al, compare the major characteristics of clinical and genomic data with those found in business data, and list the different requirements between them. According to Wang et al, "the structure of clinical and genomic data is very complex and fast evolving," due to the complexity of biological research and the pace of experimental advances. Specifically this means that existing entity types in both clinical and gene data can be defined as dimensions, and new ones can be added as necessary. The discussion and example provided in this study about the multidimensional models for the clinical data space are relevant to this project. Also, the discussion about the application of existing multidimensional models specific to the clinical data space is relevant to the design options. According to Wang et al, the star schemas do not appear to be sufficient or adequate for modeling the semantics of complex data spaces like the clinical and gene data spaces. Wang et al, also compare the characteristics of clinical and genomic data when compared to business data. The work of Wang et al will help enhance the understanding of how multidimensional data models are used in the clinical context.

Kenagy, J.W., Berwick, D.M., Shore, M.F. (1999). Service Quality in Health Care. JAMA, Feb. 17, 281, 7.

This paper presents an evaluation of health care service quality principles through the analysis of a routine encounter. This paper presents the evaluation of health care from the patient's perspective about the healthcare services. Kenagy et al, analyze a healthcare encounter from a service quality perspective, and make recommendations focused on the patient, instead of clinicians or institutions. Some of the recommendations from this study are: focus service quality on the patient; redesign processes to be more efficient (eliminate steps that don't add value to patient experience); hire for success, seek commitment and great attitude; develop a service oriented leadership standard; "3 Rs of service quality" and patient/results combination. This work supports our idea that health care services should be focused on the patient. Measuring patient satisfaction by use, retention, and satisfaction can help the student health center improve services and quality of care.

Mehta S, Suzuki S., Glick H., & Schulman K. (1999). Determining an episode of care using claims data: diabetic foot ulcer. Diabetes Care. 22(7):1110-1115.

This study is about the assessment or analysis of episodes of care for diabetic foot ulcer patients using claims data to determine the duration of the episode. Specifically, this study compares two patient data sources two assess diabetes patient inpatient cost for foot complications. This study provides an excellent example of the use of episode of care as a tool to analyze different diseases and treatment strategies, specifically for the management of patients with diabetes. This study is relevant to our study specifically to provide information related to the definition of episodes of care in order to design the episode of care multidimensional model for the student health center.

Eilers, G. (2004). Improving patient satisfaction with waiting time. Journal of American College Health, 53, 41-43.

This study presents a discussion of the strategies implemented in a student health center to improve patient satisfaction with waiting time. Based on the results of a patient satisfaction survey, the author describes how the student health center staff used a quality improvement process (QI) to improve wait times in the clinic. The author points out that the overall quality of service is determined by the system. Based on another study Eilers points out that decision making "is based on data and principles, using a multidisciplinary team process." (Eilers, G.M., 2004) This study is relevant to our project in that it provides evidence that wait times in the student health center environment is a factor to be considered when assessing quality of service. The artifacts we intend to build and evaluate in our project should help on the analysis of data fundamental to decision making in support of quality improvements.

Ehrlich, P. F., Haque, A., Swisher-McClure, S., & Helmkamp, J. (2006). Screening and brief intervention for alcohol problems in a university student health clinic. Journal of American College Health, 54(5), 279-287.

The work of Erlich et al is about a study conducted to determine the feasibility of the SHC (student health center) to introduce a campus screening and brief intervention program for alcohol, and “to determine whether the patients seen in the SHC differ in terms of prevalence and severity of alcohol-related problems compared with students previously seen in the ED.” The results of study contrasting drinking characteristics of the students seen in ED with those seen at the SHC based on the study protocol support the idea that the SHC is a feasible location for SBI alcohol intervention and that “risk profiles of SHC patients are very similar to those in the ED.” This study is relevant to our research in that it provides evidence of how important it is for the SHC to integrate alcohol abuse programs and to monitor their effectiveness within the student population. The work of Ehrlich can be used by the SHC along with our proposal to develop specific measures for the effectiveness of the alcohol intervention programs.

Wall MM, Stromberg KD, Pothoff S., & Kane, RL. (2004). Alcoholism treatment episodes validly defined using mental health care utilization records. Journal of Clinical Epidemiology. 57(4):373-380.

Wall et al, describe a method for testing a number of definitions of an episode of care for alcoholism treatment from health care utilization records with the aim to “describe statistical methods for assessing the validity of such episode definitions.” In this study the treatment episodes are defined on a minimum number of alcoholism encounters and the length of the “clear zone” which can be though of as the “cluster of months with no encounters”. According to Wall et al the “Episode definitions based on utilization data facilitate the comparison of

health outcomes across clinical sites and across time because definitions of treatment may vary spatially and temporally." Wall et al point out that researchers can benefit from the relative ease in which different episodes of care definitions can be created to allow for group selection based on episode definition factor importance. This study argues that "the episode of care provides an important tool for studying treatment outcomes across clinical setting or clinical management region or for comparing pre-/post-episode behaviors." This study supports the idea of using the episode of care as a tool to measure treatment effectiveness.

Rizzi, S. & Abelló, A. & Lechtenbörger, J. & Trujillo, J. (2006, November 10). Research in Data Warehouse Modeling and Design: Dead or Alive. Proceedings of the 9th ACM international workshop on Data warehousing and OLAP. Arlington, Virginia, USA, ACM Press, 3-10.

Discussion and review of Data Warehouse modeling and design research open issues. Rizzi et al present a concise but detailed discussion of both modeling and design concepts for Data Warehouses. Rizzi et al, discuss areas like modeling security, semantic gap, schema evolution, and quality metrics among others. This work is relevant to our project in that it provides an overview of the concepts surrounding the modeling and design of data warehouses.

Gutiérrez, A., & Marotta, A. (2000, November). An Overview of Data Warehouse Design Techniques. Reporte Técnico INCO-01-09. In Co - Pedeciba, Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay.

Gutiérrez and Marotta present a brief overview of data warehouse design approaches. According to Gutiérrez and Marotta, the literature, "presents two different approaches for Relational DW design: one that applies dimensional modeling techniques, and another that bases mainly in the concept of materialized views." This study is in agreement with Gutiérrez and Marotta in that existing data warehouse design work consists of techniques and design patterns applied to a specific domain area.

Hüsemann, B., Lechtenbörger, J., & Vossen, G. (2000). Conceptual data warehouse design. In Proceedings DMDW, 3–9.

Present a method for conceptual data warehouse design compatible with traditional database design. The authors show "how to systematically derive a conceptual warehouse schema" that is in "generalized multidimensional normal form." (Hüsemann et al, 6-1) Authors claim three contributions are made through their work being: the establishment of guidelines for distinction between a dimension level and property attribute, the presentation of a graphical formalism for conceptual data warehouse design, and show how to obtain the generalized multidimensional normal form (GMNF) for a data warehouse schema design (Hüsemann et al, 6-2). Provide a table listing attributes in the categories of measure or dimension resulting from the analysis of the ER schema (Hüsemann et al, 6-6).

Serrano, M., Trujillo, J., Coral, C., & Piattini, M.(2007). Metrics for data warehouse conceptual models understandability. Inf. Software Technology, 49(8), 851-870.

Serrano et al, present a set of validated metrics defined to measure the understandability (a quality sub-characteristic) of conceptual models for data warehouses, and present their theoretical validation to assure their correct

definition. According to Serrano et al, the validated metrics will be useful in measuring "the understandability and the efficiency of designers and users in working with the schemas." Serrano et al also point out that the focus of their study is on "the star level metrics as the star schema is the main issue of a DW conceptual model." While this study is relevant to our search for rigorous methods since the authors propose an approach (formal measures) to guide designers "to reduce subjectivity and bias in evaluation," we will make reference to Serrano et al work but instead we will assess for our needs the utility of other quality evaluation methods.

Conn, S.S. (2005, April). OLTP and OLAP data integration: a review of feasible implementation methods and architectures for real time data analysis. South West Conference, IEEE Proceedings, 8-10, 515 – 520.

This paper reviews methods and architectures for OLTP and OLAP integrated environments in support of real time data analysis. Conn addresses the problems related to OLAP environments in contrast with OTLP operational systems and the need to build real-time data analysis systems. Four major approaches to real-time data analysis data warehousing architectures are identified by Conn. The author also points out that OLTP and OLAP integration is increasingly feasible through network zones and materialized views. The concept of ROLAP and DOLAP discussed by Conn are relevant to our proposed research topic.

Sen, A., & Sinha, A.P., (2005, March). A Comparison of Data Warehousing Methodologies. Communications of the ACM, 48, 3.

Sen and Sinha analyze and compare fifteen different data warehousing methodologies based on a common set of attributes including: core competency, requirements modeling, data modeling, and support for normalization/Denormalizion, Architecture Design Philosophy, Implementation strategy, metadata management, query design, scalability, and change management. Present a visual representation of five different types of data warehouse architectures. Analyze and compare different data warehousing methodologies. Present a set of attributes that comprise the essential features of any data warehousing methodology (Sen and Sinha, 82). This will be useful to the development of the student health center DW architecture by using the visual representation of the different architectures provided by Sen and Sinha (Sen and Sinha, 80).

Karayannidis, N., Vassiliadis, P., Tsois, A., & Sellis, T. (2001, November). ERATOSTHENES: Design and Architecture of an OLAP System. Proceedings of the 8th Panhellenic Conference on Informatics, 207-216.

Karayannidis et al, review the basic characteristics of an OLAP system, and discuss the requirements and design issues for three models of an OLAP system. Karayanidis et al present design choices for a prototype, and review conceptual, logical, and physical models. The authors elaborate on the architecture requirements and design issues of the ERATOSTHENES project. This study provides examples of the design concepts and addresses the future need to fully implement the ERATOSTHENES system.

Hwang, M. I., & Xu, H. (2007). The Effect of Implementation Factors on Data Warehousing Success: An Exploratory Study. Journal of Information, Information Technology, and Organizations, 2.

Hwang and Xu examine different data warehousing studies about successful implementation factors and develop a research model for data warehousing. Hwang and Xu consolidated the success factors from prior studies and derived a list of eight variables of success of data warehousing implementation (Hwang and Xu, 4). Hwang and Xu's study is significant to our work by providing not only a model for data warehousing success, but also by presenting the consolidated data warehouse implementation factors to be considered in any data warehousing project.

Chaudhuri, S., & Dayal, U. (1997). An Overview of Data Warehousing and OLAP Technology. Proceedings of ACM SIGMOD, 65–74.

Chaudhuri and Dayal present an overview of data warehousing and OLAP technologies elaborating and describing tools for extraction, cleansing, and loading of data, multidimensional OLAP models, front end tools for data analysis, and tools for metadata management. Chaudhuri and Dayal provide additional information to consider about data warehouse servers and server architectures. Chaudhuri and Dayal also briefly describe the use of ROLAP servers and comment about their strengths and weaknesses.

Phipps, C., & Davis, K.C. (2002). Automating Data Warehouse Conceptual Schema Design and Evaluation. DMDW'02.

Phipps and Davis present an algorithm used to derive data warehouse conceptual schemas versions originated from OLTP or operational schemas. In addition, Phipps and Davis provide an algorithm for evaluating conceptual schemas using user queries. Phipps and Davis discussions about conceptual model selection and conceptual schema creation could be of special interest to researchers based on the comparison and summarization of various models.

Vassiliadis, P., & Sellis, T. (1999, December). A Survey of Logical Models for OLAP Databases. SIGMOD Record 28(4).

Vassiliadis and Sellis present a comparison of different multidimensional data cube models for OLAP applications by categorizing the research work as commercial tools and academic efforts. Vassiliadis and Sellis proceed to further dived the academic efforts category into two classes, the relational extensions and the cube oriented approaches to OLAP logical modeling. Vassiliadis and Sellis work of comparing the various cube models contributes to the field of study by providing a survey, comparison, and summary of OLAP of the different multidimensional models available in the research body to understand the related terminology and semantics. Vassiliadis and Sellis also elaborate on some of the commercial products and technologies available like ROLAP and MOLAP architectures, which are of particular interest and align with the rest of our body of literature.

Peralta, V., & Ruggia, R. (2003). Using Design Guidelines to Improve Data Warehouse Logical Design. Proceedings of the Int. Workshop on Design and Management of Data Warehouses, colocated with VLDB.

The work of Peralta and Ruggia presents a set of guidelines for dealing with data warehouse conceptual to logical design schema problem, and the use of a “formalism” to identify implementation requirements not considered in the conceptual model stage. According to Peralta and Ruggia, their work makes the contribution of formalism that should help express implementation related guidelines in a simplified manner while producing data warehouse relational schemas through a semi-automated process. However, Peralta and Ruggia point out that their design guidelines should help dealing with some specific design problems, but they could be extended or be used along with other available proposals to compensate their design guidelines.

Lujan-Mora, S., & Trujillo, J. (2004, November). Physical Modeling of Data Warehouses using UML. DOLAP’04, 12-14, 2004.

Lujan-Mora and Trujillo present a Unified Modeling Language (UML) based method for modeling the physical design of a data warehouse by adapting the component and deployment diagrams. Lujan-Mora and Trujillo elaborate their study on the need to for techniques to model data warehouse physical design at early stages for the data warehouse project. Lujan-Mora and Trujillo’s discussion on the topic provides an overview of a data warehouse design framework, advantages of their approach, and conclusions including future work. Lujan-Mora and Trujillo claim that their approach allows the designer to encompass all the data warehouse phases. They provide valuable examples of the use of their approach.

Marotta, A., Piedrabuena, F., & Abelló, A. (2006). Managing quality properties in a ROLAP environment. In Proceedings of CAiSE, 27–141.

Marotta et al provide a technique to evaluate the quality of retrieved data in a distributed data mart architecture using multidimensional queries. Marotta et al present some quality properties, and the formulas used for calculating the property values of the query results. This study provides a discussion on concepts of quality that can be applied to other data warehouse and OLAP architectures.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. MIS Quarterly, 28(1), 75-105.

The work of Hevner and March presents a conceptual framework and a set of six guidelines intended to help develop the understanding, execution, and evaluation of design science research. According to Hevner and March, the primary goal of this paper "is to inform the community of IS researchers and practitioners of how to conduct, evaluate, and present design science research." The work of Hevner and March is relevant to our research study in that it describes the design science within the context of IS by using the proposed conceptual framework and present guidelines for "conducting and evaluating good design-science research." This study will make use Hevner and March (2004) guidelines as part of the approach to guide the process of developing the solution artifacts, to solve the given business problems.

Peffers, K., Tuure, T., Rothenberger, M. A., & Chatterjee, S. (2007, Winter). A Design Science Research Methodology. Journal of Management Information Systems, 24, 3, 45-77.

Peffers et al (2007) present and demonstrate the use of a design science research methodology (DSRM) based on research principles, practices, and procedures with the fundamental aim to remain consistent with prior literature, provide a “nominal process” model for design science, and to provide a “mental model” for presenting and evaluating design science research in information systems. Four cases are presented and demonstrated in this study on how the DSRM is used. The work of Peffers et al (2007) provides a relevant approach to design science that can be used as “road-map” to achieve the objectives of this project. This study will make use Peffer et al (2007) design science research methodology process as the adopted research method.

Baskerville, R., Pries-Heje, J., & Venable, J. (2008, May). Evaluation Risks in Design Science Research: A Framework. In Proceedings of the Third International Conference on Design Science Research in Information Systems and Technology.

Baskerville et al present a framework for evaluating risk in Design Science Research (DSR) based on the process framework developed by Hevner et al (Hevner et al, 2004) through which six potential risk areas are identified (Baskerville et al, 2008). The risk management approach presented by Baskerville et al (2008) should be used in this study to assess and minimize design-science research risks.

Jourdan, Z., Rainer, R. K., & Marshall, T. E. (2006). Business Intelligence: An Analysis of the Literature. Information Systems Mangement, 25, 121-131.

Jourdan et al conducted a literature review and analysis based on three phases choosing research articles in the topic area from “general, mainstream journals,

rather than specialized journals." In this paper the authors review BI methodologies, and the topics being addressed in the BI research, in order to identify gaps in the research with the intent to propose their own BI research agenda. In their study, Jourdan et al analyze the BI literature and found that several BI research methodologies were not well represented or were completely excluded. Jourdan et al, also identified additional subject areas in need of further research. The work of Jourdan et al provides a great example on how to conduct the analysis and review of a body of literature.

Thomsen, C., & Bach Pedersen, T. (2005). A Survey of Open Source Tools for Business Intelligence. In Proceedings of DaWaK'05.

This paper is a survey of open source BI tools and the capabilities they support. In this paper, discuss the features of Extract Transform Load (ETL) tools, database management systems (DBMSs), On Line Analytical Processing (OLAP) servers, and OLAP clients. According to the authors, there are "mature and powerful tools in all these categories." However, according to Thomsen et al, these open source tools still don't match the features found in commercial tools. This work is relevant to our study because we can perform our own evaluation of the tools recommended by the authors and assess their utility for our student health center.

Dell'Aquila, C., Di Tria, F., Lefons, E., & Tangorra, F. (2008). Evaluating Business Intelligence Platforms: a case study. Proceedings of the 7th WSEAS International Conference on Artificial intelligence, knowledge engineering and data bases, Cambridge, UK, 558-564.

Dell'Aquila et al, examine several Business Intelligence Platforms, specifically Microsoft SQL Server 2005, Oracle Discoverer, and Microstrategy using a software measurement method designed to analyze functional complexity. At the completion of the study the results show that Microstrategy's Business Intelligence platform has high functional complexity due to its object oriented design. The study conducted by Dell'Aquila et al present the final experimental results scores comparing the three leading business intelligence tools in a table format based on area (information delivery, integration, analysis), capabilities, and task. This study in particular is relevant to this project because it provides an analysis of three known Business Intelligence platforms and their functional complexity, but also it provides an applicable example of the functional aspects characterizing Business Intelligence tools.

Gorla, N. (2003, November). Features to Consider in A Data Warehousing System. Communications of the ACM, 46, 11.

Discussion of how OLAP features impact the perceived easy of use (PEU) and the perceived usefulness (PU) of the OLAP implementation. Also, the authors make suggestions about the appropriate use of ROLAP and MOLAP systems within specific contexts. The authors make use of a questionnaire-based survey to capture OLAP users' perceptions about PEU and PU. According to Gorla, their study "findings indicate MOLAP tools make the data warehouse system easy to use but not useful; ROLAP tools make the data warehouse useful but not easy to use." Gorla makes some suggestions to improve the design of data warehouse with OLAP. Among the suggestions made by Gorla we find one to be absolutely

critical for the student health center staff. Is that of making ROLAP user-friendly. The work of Gorla is relevant to our study in that it presents a discussion contrasting OLAP features, and ROLAP versus MOLAP based on perceived ease of use (PEU) and perceived usefulness (PU). The ideas and suggestions presented by Gorla can be used in this study for the evaluation of the proposed artifact solutions.

Tomic, D. (2006). Business intelligence in managerial accounting. SEE Journal.

This paper reviews and evaluates the possibilities of using business intelligence (BI) in managerial accounting. An excellent discussion about data warehouses, OLAP tools, distinctions between MOLAP and ROLAP, and the importance of BI in managerial accounting. The authors present a case study about the development of a financial analysis data mart and the activities required to complete the implementation of the new system. In this case study Oracle Data Warehouse Builder 10g and Oracle Discoverer Desktop are used to implement the new system.

Oliveira, R., Bernardino, J. (2006). Building OLAP Tools Over Large Databases. Proceedings of IADIS Virtual Multi Conference.

Oliveira and Bernardino, present an OLAP tool prototype containing features recommended on the implementation of on-line analytical processing (OLAP) tools used for data analysis. The prototype was evaluated on a data warehouse implemented in a relational repository using the star schema but this prototype can be implemented on any database system. The authors recommend the most

important features that an OLAP tool must include. The authors analyzed two OLAP tools, Discoverer Desktop from Oracle 9i, and JPivot 1.3.0.

Rahm, E., Hai Do, H., “Data Cleaning: Problems and Current Approaches”, Bulletin of the Technical Committee on Data Engineering, 23, 4, 2000.

Rahm et al, discuss data cleaning problems and approaches. Rahm et al, present a classification of data quality problems by differentiating between single-source and multi-source, and each one with schema level and/or instance level problems. The authors discuss each of the data cleaning phases like data analysis, transformation, mapping rules, verification, and backflow of cleaned data. Rahm et al, also describe some of the tools used for data transformation and data cleaning tasks. This work provides an excellent overview of data quality problems originated from the integration of heterogeneous operational data sources. This study is relevant to our project because it provides an approach to data cleaning and a description of data cleaning tools. This work did not provide any examples or discussions specific to clinical data cleaning and/or quality.

Golfarelli, M., Rizzi, S. (2009, November). A comprehensive approach to data warehouse testing. DOLAP '09: Proceeding of the ACM twelfth international workshop on Data warehousing and OLAP, 17-24.

In this paper Golfarelli and Rizzi propose a comprehensive approach to data mart testing by presenting different testing activities related to the phases of data mart design based on “what is tested and how it is tested.” Golfarelli and Rizzi describe the items to be tested and discuss each one of them. The data mart testing approach proposed by Golfarelli and Rizzi is characterized by a focus on testing design phases to “reduce the impact of error correction”, data mart testing activities classified in terms of what is tested and how it is tested, adoption of a reference design methodology that relates tightly to the proposed testing activities, and the objective to relate testing activities to quality metrics to allow for quantitative assessment. The data mart testing approach presented by Golfarelli and Rizzi can be considered a rigorous method applicable to this project since it provides a design quality testing approach.

Glossary

1. University student health center – Usually the department within the University institution with the mission of providing health services to students and promoting good health habits.
2. Design-science research - According to Hevner et al, design science research "addresses important unsolved problems in unique or innovative ways or solved problems in more effective or efficient ways." Also, Hevner et al point out that design science research is distinguished from routine design in that it makes a "clear identification of a contribution to the archival knowledge base of foundations and methodologies." (Hevner et al, 2004)
3. Data Warehousing – "A subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision-making process." (Connolly and Begg, 1151)
4. Data Mart – A data structure designed to meet the data analysis needs of a department, organizational unit, and/or a business function (Connolly and Begg, 1151), (Sahama and Croll, 2007).
5. Online Analytical Processing (OLAP) – According to Connolly and Begg, Online Analytical Processing known as OLAP is "The dynamic synthesis, analysis, and consolidation of large volumes of multi-dimensional data." (Connolly and Begg, 1205)

6. Fact table – According to Sen and Sinha, a fact table is “a specialized relation with a multi-attribute key and contains attributes whose values are generally numeric and additive.” (Sen and Sinha, 2005)
7. Dimension table – According to Sen and Sinha, a dimension table “has a single attribute primary key (usually surrogate) that corresponds exactly to one of the attributes of the multi-attribute key of the fact table.” (Sen and Sinha, 2005)
8. Star schema - “A logical structure that has a fact table containing factual data in the center, surrounded by dimension tables containing reference data (which can be denormalized).” (Connolly and Begg, 1183)
9. Conceptual modeling – According to Rizzi, conceptual modeling "provides a higher level of abstraction in describing the warehousing process and architecture in all its aspects, aimed at achieving independence of implementation issues." (Rizzi, 2006)
10. Multidimensional modeling – According to Rizzi (2006) multidimensional modeling "intuitively represents data under the metaphor of a cube whose cells correspond to events that occurred in the business domain" (Rizzi, 2006)