Data Warehousing in an Integrated Health System; Building the Business Case

Edward F. Ewen, MD
Director of Disease Management
Christiana Care Health System
Ewen.e@christianacare.org

Carl E. Medsker, MS
Research Engineer
Drexel University
Medsker@compuserve.com

Laura E. Dusterhoft, RN
Project Coordinator
Christiana Care Health System
Dusterhoft.l@christianacare.org

Kelly Levan-Shultz, RN, BSN
Levan-shultz.k@christianacare.org

James L. Smith, BS
Jsmith9871@aol.com

Milton A. Gottschall, BS
Gottschall.m@ christianacare.org

1. ABSTRACT

Healthcare presents unique challenges for the architect of a data warehouse. Integrated health systems are shifting its focus away from the acute care setting and moving towards cross-continuum care management. Improving healthcare quality while reducing costs requires the elimination of unnecessary variation in the care process. This paper describes the lessons learned during the business case development for the project. Topics include establishing the need for a data warehouse, understanding data warehousing in healthcare, justifying the cost of a data warehouse, building the team, and setting achievable goals.

1.1 Keywords

Data warehouse, healthcare, integrated health system, integrated delivery system, decision support system. On-line analytical processing, OLAP, IHS, IDS, DSS.

2. INTRODUCTION

Christiana Care is a community based not-for-profit healthcare organization dedicated to improving the health of all individuals in the communities we serve. To achieve this mission, we strive to provide quality services in a caring and cost-competitive manner. Our constant goal is to achieve the best possible clinical outcomes and service excellence while preserving individual dignity, comfort and convenience.

DOLAP '98 Washington DC USA

Service Statistics for Fiscal Year 1997	
Licensed-beds	1,040
Inpatient admissions	37,223
Births	6,204
Emergency department visits	109,468
Extended care pavilion admissions	387
Health center visits	65,818
Outpatient visits	300,917
Surgical procedures	39,999
Radiologic procedures	181,195
Inpatient days (excluding newborns)	214,579
Home health visits	511,733
Health plan covered lives	~30,000
Average length of stay	5.7 days

Table 1.

Christiana Care Health System reorganized as an integrated delivery system (IDS) in July 1997 from the former Medical Center of Delaware. The IDS is comprised of Wilmington Hospital, Christiana Hospital, Riverside Health Care Center, a preventative medicine and rehabilitation facility, an insurance health plan, 10 primary care physician practices, an imaging center, occupational and physical therapy centers, and a 600-member physician organization. The health plan covers approximately 30,000 lives enrolled in both global risk and more traditional insurance products.

Services offered by Christiana Care include acute inpatient care, long term care, home health care, infusion services, at-home medical care and monitoring for high-risk obstetrical patients, durable medical equipment and wellness centers in local high schools and community wellness programs. Service statistics are summarized in Table 1. Despite the volatility of the regional market, Christiana Care reported an operating gain for FY1997. This net-gain in operating revenue occurred despite the insurance division posting an operating loss.

2.1 Establishing the Need for a Data Warehouse

Over the last several years there has been increasing pressure from health maintenance organizations (HMO) and employers in the Delaware region to reduce the total cost of healthcare and to reduce payment for physician services. Strong managed care penetration in Pennsylvania has resulted in Philadelphia providers attempting aggressive incursions into the Delaware market place.

This competition led to a change in strategic direction resulting in the formation of the new Christiana Care IDS. By more closely aligning physicians, hospitals, and home health agency services; Christiana Care is better positioned to manage care across the continuum of health, offer superior clinical products and lower the overall costs while continuously improving clinical outcomes. The insurance division also provides the opportunity to develop innovative risk sharing arrangements across the continuum of care. See Figure 1.

In order to plan for the challenges of the rapidly evolving healthcare environment, Christiana Care conducted-a business needs and organizational assessment. An opportunity to integrate information across previously separate functional units was identified. Senior management commissioned a national consulting group to further explore this opportunity. Data warehousing was identified as an enabling technology that could provide Christiana Care with strategic advantages.

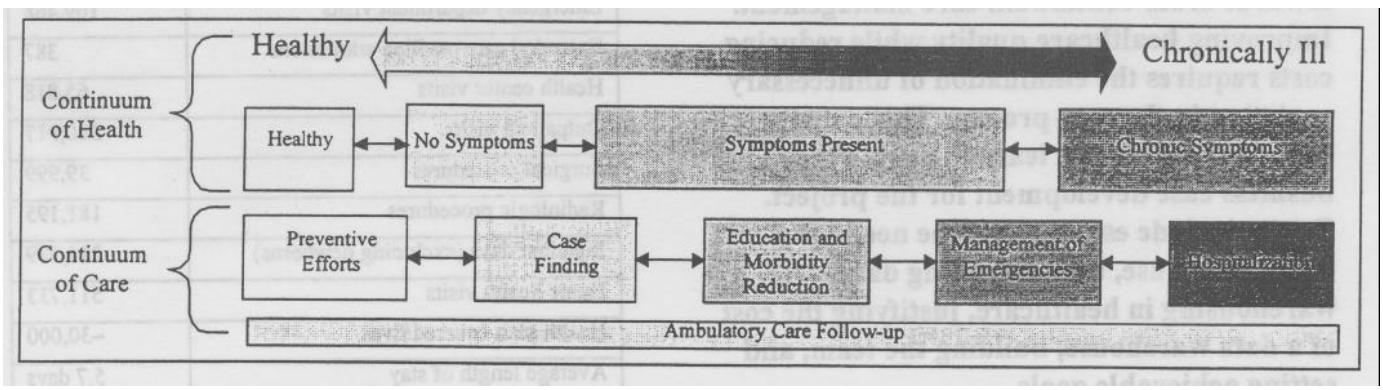


Figure 1. Continuum of health compared to the continuum of care.

2.2 Identifying Key Business Questions

Twenty-eight key business managers were interviewed regarding their priority information needs for decision support. Several crucial areas of concern were identified during this activity that may be best categorized as business specific interests, data accessibility, corporate culture, leadership, and process inefficiencies. Specifically, the perception that too much time is spent gathering data and too little time is spent analyzing the information was repeatedly articulated. See Figure 2.

The interview team quickly discerned a significant risk that the data warehouse could be mistakenly seen as a global solution for all expressed areas of concern. In fact, a data warehouse is only able to affect data accessibility and consistency issues. Greater access to information will not change leadership, corporate culture or the basis of healthcare.

Sample survey questions are excerpted below:

1. List the top three decisions you need to make.
2. What reports or tools do you require to make these decisions?
3. What are the strengths/weaknesses of the current reports/tools?
4. What is the time/quality/dollar impact of making good/bad decisions based on current the information?
5. What information/calculations/aggregates would enhance your decision-making capability?
6. How likely would you be to use an on line system to analyze this information?
7. In what way would your decision-making process change if you could run ad hoc queries of the data?
8. What operational changes would be needed to take advantage of this information?
9. What would be the key benefits of having this information available?

2.3 Inventory and Analysis of Data Sources

Concurrent to the management interviews, a data catalog was created to determine the appropriate candidate data sources to populate the data warehouse. This catalog contains detailed information on the content and structure of the 312 individual applications that are used in the IDS. The primary goal of the inventory was to compare current data sources to the perceived information needs. The results of the data source inventory and

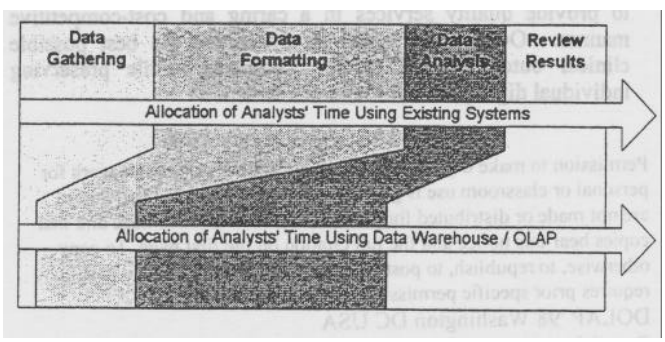


Figure 2. Allocation of analysts' time.

management interviews revealed the following key challenges:

- Data about the same subject is spread across multiple applications
- Some applications contain empty data structures
- Systems are not integrated to update/transfer/load data automatically which leads to fragmentation and data inconsistency
- Multiple and incompatible data structures make it difficult and sometimes impossible to combine similar data
- Data from one system is frequently printed and reentered, compounding data inconsistency problems

2.4 Defining Candidate Subject Areas

A list of potential subject areas was developed and prioritized based on the following:

- Benefit – the quantitative and qualitative benefits to the health system achieved by implementing a subject area with respect to patient satisfaction, clinical outcome and operational efficiency
- Data gap – the difference between the data needed to implement a particular subject area and the quality and the quantity of data available
- Complexity – the amount of effort required to create an effective design for a particular subject area
- Implementation risk – the relative ease of implementing a particular subject area in terms of organizational readiness and ability to act, number of required system interfaces, time frame and breadth versus functional depth

2.5 Subject Area Selection

Generally, it is best to select a topic with the highest potential benefit and which has the lowest risk factors. Sometimes business considerations will override an apparently easy decision. You must also keep in mind that if there is no compelling business need, the value of the data is zero.

Christiana Care senior management named a steering committee charged with the responsibility of choosing the first subject area. In order to identify the best choice, the list of key dimensions for the potential subject areas were rated for expected benefits, data gap, complexity and implementation risk to derive an overall score.

Based on this scorecard, subject areas were analyzed and ranked. When the top three subject areas were further scrutinized, two seemingly better candidates did not withstand business case analysis. It was discovered that these topics suffered from insurmountable data gaps and lacked organizational readiness for change. The committee then considered the third ranked subject area, physician activity reporting. Because physicians are ultimately responsibility for the quality of care and significantly impact on its cost, physician practice patterns are a logical control point. See Figure 3.

Following agreed upon guidelines, the steering committee

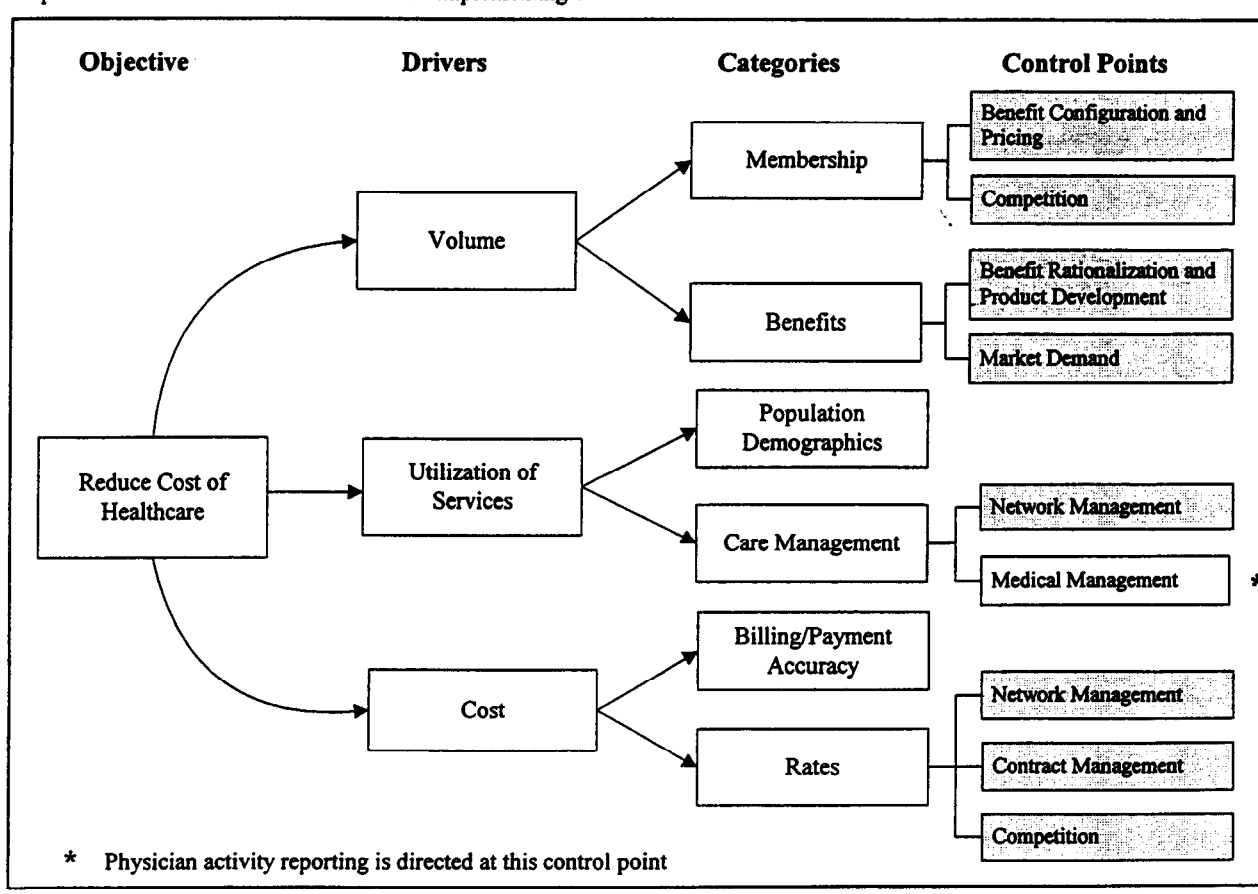


Figure 3. Cost model for an integrated delivery system.

confirmed the selection of physician activity reporting as the first subject area of the data warehouse. However, this decision did not guarantee enthusiasm or sustainable support. To be successful, support for this initiative must come from the highest level of management.

2.6 Selection of Business Sponsors

In established data warehouses, the business sponsor typically proposes the subject area based on a critical business need. In our situation and healthcare in general, there are multiple groups with vested interests in the topic of care management. Co-sponsorship from the physician organization and patient care management was accepted because the two sponsors share a strong common goal of improving clinical outcomes by decreasing needless variation in pratice patterns.

3. UNDERSTANDING DATA WAREHOUSING IN HEALTHCARE

Decision support applications based on operational transaction systems already existed within individual business units of the IDS. Management justifiably asked why existing systems were not sufficient. To answer this question, an elucidation of the risks, benefits and costs associated with a data warehouse were required. First, we considered our risks:

- Few people understood the differences between operational and decision support systems.
- It was difficult to conceptualize the added benefit of data integration.
- The target user groups were not accustomed to an interactive data interface since the standard had been static reporting.
- There was a concern that it would be difficult to sustain support throughout the life of the project

Next, we considered the potential benefits:

- The data warehouse will provide demonstrable data integration resulting in one version of the truth across the IDS.
- On-line analytical processing systems (OLAP) will open new decision support perspectives for executives through a dynamic and easy to use interface.
- Using data pre-packaged for analysis would allow business experts to locate answers quicker and solve problems sooner.
- Better understanding of populations and providers will improve efficiency of care management programs
- Routine reporting will no longer require the advanced technical skills of the IT department
- Metadata, information about where the data came from and how it was transformed will be accessible to users
- Confidence in data quality will increase
- Benchmarks will be established
- The newly formed physician organization will be able to develop of an information-based culture

This led to the finally to the final factor – cost. Data warehouse projects are expensive. Low-end estimates start at $1 million and run as high as $5 million just to get the project off the ground. To help you along, we have included a sample budget which we promptly learned was too optimistic. See Table 2.

Capital Requirements	
Consultants	Dollars
Data Architect, modeler and developer	$ 202,800
DBA and E/T/L developer	$ 101,400
E/T/L and SQL developer	$ 202,800
Consultant Subtotal	$ 507,000
Hardware	
Database Server	$ 90,097
Workstations (2)	$ 8,000
Server Backup	$ 15,400
Hardware Subtotal	$ 113,497
Software	
Warehouse Database Server	$ 3,200
Extract, Transform, Load Software	$ 75,000
Software for source data quality analysis	$ 1,400
Hypercube query and report software	$ 1,395
Web server and user interface software	$ 42,000
Software Maintenance fees	$ 12,300
Software Subtotal	$ 135,295
Facilities Requirements	
Cubicles and Office Furniture	$ 12,000
Meeting Furniture and Fixtures	$ 2,130
Network installation	$ 5,500
Telephone communications hardware	$ 800
FAX machine	$ 900
Facilities Subtotal	$ 21,330
Total Projected Capital Budget	$ 777,121

Table 2.

Given the significant risk and cost involved in implementing a data warehouse, it was necessary to educate management in some basic principles.

3.1 Operational Transaction Systems Versus Decision Support

Although healthcare is well acquainted with operational systems, there is little industry experience with enterprise-wide decision support. There was a misconception that local on-line transaction processing (OLTP) systems already provided enterprise-wide decision support. The following paragraphs summarize the differences between the two concepts.

Operational systems are designed for the efficient storage of data and rapid processing of individual transactions or small sets of records. Normalized relational databases are optimized for managing the integrity of data for insert, update, and delete operations and store data across multiple tables. Statistical analysis programs used for decision support require data to be combined into flat records reduced to the lowest granularity necessary for the query. The different characteristics of operational data versus statistical data for decision support are outlined in Table3.

Decision-makers use software tools that can broadly be divided into reports, data mining/statistics, and data browsing

categories. Reports can be developed to summarize well-understood business processes and issues. Data mining and data browsing enable the analyst to explore large data sets to find business trends and opportunities. Data mining is a largely machine driven process where numerous algorithms are used to summarize, model, and cluster data. Data browsing, called OLAP provides the analyst with integrated data and an efficient interface for rapidly manipulating views and levels of aggregation.

Characteristic	Operational Data	Decision Support Data
Business goal	Tactical	Strategic
Purpose	Business operations; customer service	Strategic planning; business reorganization
Users/audience	Front line employees; customers	Managers, executives, analysts
Focus	Specific customer; department	Product, line of business, customer profiles
Outputs	Orders; reports	Graphs; stats; models; forecasts
Grain	Individual transactions	Aggregates of transactions
Character	Dynamic, continually updated	Nonvolatile, read-only snapshot
Format	Fixed structure, variable contents	Flexible structure, multidimensional
Refresh rates	Real time; minute-to-minute	Weekly; monthly, quarterly
Retention time	< 90 days	Years
Data structures	Normalized tables	Single, flat data sets; Star schema
Data needs	Small sets of rows and tables	All records from multi-table joins
Data coding	Textual descriptions preferred	Mostly require numeric data coding
Data manipulations	Insert, edit, delete, retrieve individual records or small sets	Summarize, aggregate, cross-tab,

Table 3.

3.2 Gaining Political Endorsement

The subject area was selected but endorsement remained tentative. The business analysis team feared that the lack of understanding would lead to waning enthusiasm before the project was completed. The dual goal of the presentation was to foster understanding of data warehousing among senior management and emphasize the importance of integrating data sources to support the new strategic direction of the IDS.

3.3 Prototyping With OLAP

OLAP is based on the simple to understand and frequently used cross tabulation summarization. The crosstab compares two variables along some third, measurement variable. The distinct values of one variable form the columns and the distinct values of the other variable form the rows while the third, or measurement, variable is summarized in each cell of the matrix. A standard data analysis can include many individual crosstabs as part of the process of understanding relationships in the data. A dimensional matrix is known as an OLAP cube. Converting large data sets to a crosstabs places a heavy demand on computing resources and is not optimized in our current environment.

The power of this approach was realized when Dr. Ewen created and demonstrated an OLAP cube that contained data about real healthcare concerns. Data presented statistics on inpatient and emergency department utilization and focused on the cost of uninsured encounters. The data were also segregated by disease categories that are chronic and lend themselves to care management techniques. Asthma, CHF, diabetes, and HIV typically represent high volume utilization and pose high cost risks.

The ability to reduce the time between asking a question and getting an answer was compressed from days or weeks to seconds or minutes. Our target audience was able to explore the available data until all possible aspects were exhausted. Audience participation was enthusiastic and a number of business opportunities were immediately identified. An important revelation was the high percentage of uninsured admissions and emergency department visits seen in this population. It was proposed that better care management would reduce the impact of the cost of caring for the indigent population since it is cheaper to deliver ambulatory care than it is to deliver acute care.

3.4 Other Tools Considered

Statistical software like SPSS 8.0 and SAS 6.0 perform crosstabs, but are not widely installed, require training and statistical experience, and run as client processes. Since they run as clients, the data required in the crosstab is copied to the user workstation and runs slowly due to insufficient resources. Microsoft Access 7.0 includes a crosstab query that is easy to use, but is a client process and suffers from the same data movement and resource limitations as with the current configuration of SPSS and SAS. SQL Server 6.5 improves the resource limitation and has more computational power, but does not support the crosstab query. Custom, complex SQL code may be written to simulate a crosstab, which most analysts are not trained to write. OLAP software can address all these limitations by providing rapid, interactive crosstab views of large data sets as server-side processes.

3.5 OLAP and Healthcare

All businesses can be organized along sets of hierarchical categories but medicine is unique because numerous aspects of care intertwine and overlap. To manage this complexity, we need to group data into episodes of care that may be composed of multiple encounters with various providers. Clinical users of the data warehouse will need to view data that yields information about the entire patient care process.

OLAP tools model and implement the hierarchies of a business and combine them with crosstab summaries. The data browser allows the user to combine and recombine pairs of categories as the columns and rows of the crosstab and then select from a set of measures to summarize in the matrix. Furthermore, the user

can drill down and drill up on the category hierarchies to expand or aggregate the data. This is a powerful method of information presentation that allows the user to rapidly iterate through a series of question in a fraction of the time allowed by standard, code based statistics and query tools.

Several approaches to creating OLAP crosstab data structures exist, including storage as relational tables, creation of proprietary multi-dimensional data structures, and "virtual" crosstabs generated on demand. Commercial software is available as either a client-process or a server-process and most now provide Web access. A server-side implementation with Web access is less expensive than buying multiple client licenses and leverages server resources for faster performance.

4. JUSTIFYING THE COST

Next we constructed a business case for the selected subject area based on improving patient care while reducing the cost of delivering care. Since it is difficult to quantify the return on investment (ROI) for wellness initiatives, we created several scenarios using conservative to moderate assumptions based on the previous process improvement experience of the team. Then we estimated costs over 5 years and projected the time until ROI is realized. We worked up a number of scenarios and graphed the break-even points for the project.

By affecting changes in physician practice patterns, it was proposed that the subsequent shift in utilization to the ambulatory care setting would result in cost avoidance. Based on the previous process improvement experience of the business analysis team, conservative to moderate assumptions were made regarding the potential reductions. See Table 4.

Projected Cumulative Reductions in Resource Utilization:					
	Yr1	Yr2	Yr3	Yr4	Yr5
Scenario 1					
Inpatient days reduce by	0.5%	1%	2%	3%	3%
ED visits reduce by	0.5%	1%	2%	3%	4%
Total for Scenario			**20%**		
Scenario 2					
Inpatient days reduce by	0.5%	1%	2%	4%	5%
ED visits reduce by	1%	2%	3%	5%	6%
Total for Scenario			**29.5%**		
Scenario 3					
Inpatient days reduce by	0.5%	1%	2%	4%	6%
ED visits reduce by	1%	3%	5%	7%	9%
Total for Scenario			**38.5%**		

Table 4.

The projected percentages were then multiplied by estimated facility costs using FY 1997 service statistics. Facility costs avoided were equated to dollars saved and graphed against the total cost of the data warehouse project. The total cost is the sum of development costs (see sample capital budget) and operating costs for the first five years of the project. For scenario 1, we start to see a return on investment after year five. For scenario 2 and scenario 3, ROI begins in year four. See Figure 4.

5. BUILDING THE TEAM

The most important aspect of staffing the data warehouse is to create a solid business focus for the project. One survey found that data warehouses run by IT departments were significantly more likely to fail than those commissioned and led by the business. IT personnel are essential but cannot be the primary drivers of the project.

We have learned that the real strength of our team is having people who understand the healthcare process as well as the technology behind data warehousing. A brief description of our team's composition follows:

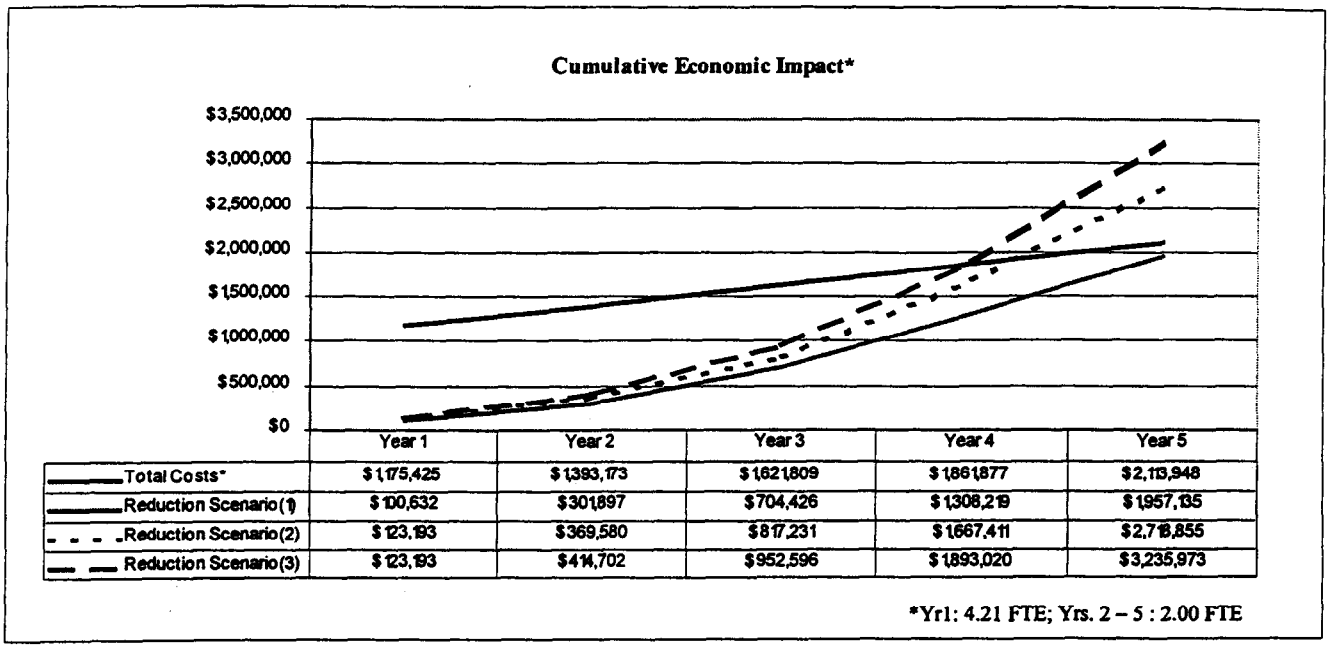
The project manager is a practicing physician with an extensive understanding of information systems. The project coordinator is a nurse/systems analyst with years of experience in data analysis for healthcare corporations. The third member of the team is also a nurse/systems analyst who brings experience with databases and WEB development. Rounding out the team are two consultants experienced in healthcare systems development who function in the roles of data warehouse architect, data modeler, SQL developer and data transformation tool specialist. A database administrator and additional systems analysts as needed also support our project. The subject area sponsors have dedicated resources to the business process team and look to our IT department for support of the network infrastructure.

We realize we have fewer people than is typically recommended for a data warehouse project. We are fortunate to have team members who assume multiple roles and adjust to the dynamic requirements of the risk-based iterative development cycle. Hire flexible people.

6. ESTABLISHING ACHIEVABLE GOALS

Our goal is to establish a process that will allow us to serve an expanding constituency without significantly increasing our investment. We have elected to focus our initial efforts on a single subject area with the expectation of it being the foundation for future iterations of the lifecycle. Even with careful planning, parallel timelines, reusable documentation and methodology, we realize we have set ourselves to an ambitious task. We have learned enough to know that a "data warehouse in 90 days or less" and "data warehouse in-a-box" are not realistic. Resist the urge to jump on the bandwagon. It will take a minimum of nine to twelve months before the data warehouse will be ready to render decision support. Guidelines to planning:

- Explicitly state all assumptions and risks up front so the steering committee can make informed decisions
- Provide regular decision points whether to continue
- Meet weekly to review risks and critical path
- Deal with the toughest problems early to minimize impact
- Establish up front and then refine a set of metrics to determine the utility of the warehouse
- If these measures do not support continuation of a subject area, then eliminate maintenance of that area



	Year 1	Year 2	Year 3	Year 4	Year 5
Total Costs*	$1,175,425	$1,393,173	$1,621,809	$1,861,877	$2,113,948
Reduction Scenario(1)	$100,632	$301,897	$704,426	$1,308,219	$1,957,135
Reduction Scenario(2)	$123,193	$369,580	$817,231	$1,667,411	$2,718,855
Reduction Scenario(3)	$123,193	$414,702	$952,596	$1,893,020	$3,235,973

*Yr1: 4.21 FTE; Yrs. 2 – 5 : 2.00 FTE

Figure 4. Time until return on investment is achieved.

7. SUMMARY: MEASURING SUCCESS

If your data warehouse is not driven by a significant and legitimate business need is not worth the investment. The "build it and they will come" approach is technology looking for a solution, and is responsible for the 70% of failed data warehouse efforts. A data warehouse is a business tool and its true value can only be determined by the business. We have compiled the following critical metrics for the physician activity reporting subject area:

- Did the project reduce practice pattern variation?
- Did resource utilization decline as predicted in the economic impact statement?
- Was the information used to design disease management programs?
- Was evidence-based case management implemented?
- Were cost saving opportunities identified and realized?

In the past, many attempts to deliver decision support technology to users met with limited success because these efforts were grounded on the assumption that the delivery of data enables its use. Although this was done with the best of intentions, it presumes that a given business area understands its purpose, needs, and processes from a system perspective well enough to define useful metrics. It further assumes that the business area is prepared as an organization to accept, analyze, communicate, and take action on the information. For a data warehouse to be successful the organization must be ready from a cultural perspective to openly consider opportunities, develop creative solutions, and commit to the change often required for action to occur.

8. REFERENCES

[1] Abston, K.C., Pryor, T.A., and Haug, P.J. Inducing practice guidelines from a hospital database. Proceedings of AMIA (Nashville TN, October 1997)

[2] Gilbreath, R. E. Health care data repositories: components and a model. The Journal of the Healthcare Information and Management Systems Society, Vol. 9, No. 1, pp63-73.

[3] Hackney, D. (1997). Understanding and Implementing Successful Data Marts. Reading, MA. Addison-Wesley Developers Press.

[4] Kimball, R. et al. (1998). The Data Warehouse Lifecycle Toolkit. New York: John Wiley & Sons Inc.

[5] Kimball, R. (1996). The Data Warehouse Toolkit. New York: John Wiley & Sons Inc.

[6] Meyer, D. and Cannon, C. (1998). Building a Better Data Warehouse. Upper Saddle River, NJ: Prentice Hall PTR.

[7] Niederman, F. (1997). Data warehousing at an urban hospital. Jrnl of Data Warehousing, Vol. 2, No. 4, pp2-12.

[8] Prather, J.C. et al. Medical data mining: knowledge discovery in a clinical data warehouse. Proceedings of AMIA (Nashville TN, October 1997)

[9] Vowler, J. (1997). Has data warehousing had its day? http/ www.computerweekly.co.uk/

[10] Wang, P. et al. The web enabled IHC enterprise data warehouse for clinical process improvement and outcomes measurement. Proceedings of AMIA (Nashville TN, October 1997)