NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

A DATA WAREHOUSE ARCHITECTURE FOR DoD HEALTHCARE PERFORMANCE MEASUREMENTS

by

Jimmy E. Francis

September 1999

Thesis Co-Advisor(s): Daniel R. Dolk
Gregory Hildebrandt

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. **AGENCY USE ONLY** ***(Leave blank)***	2. **REPORT DATE** September 1999	3. **REPORT TYPE AND DATES COVERED** Master's Thesis

4. **TITLE AND SUBTITLE** : A DATA WAREHOUSE ARCHITECTURE FOR DoD HEALTHCARE PERFORMANCE MEASUREMENTS	5. **FUNDING NUMBERS**
6. **AUTHOR(S)** Jimmy E. Francis.	
7. **PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)** Naval Postgraduate School Monterey, CA 93943-5000	8. **PERFORMING ORGANIZATION REPORT NUMBER**
9. **SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)** N/A	10. **SPONSORING / MONITORING AGENCY REPORT NUMBER**

11. **SUPPLEMENTARY NOTES**
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. **DISTRIBUTION / AVAILABILITY STATEMENT** Approved for public release; distribution is unlimited.	12b. **DISTRIBUTION CODE**

13. **ABSTRACT** ***(maximum 200 words)***

In order to improve business practices and access to healthcare related information within the Department of Defense's Healthcare System, a Performance Metrics Driven Decision Support System (PMDSS) framework has been designed using Data Warehouse technology. As part of this performance metrics-driven framework, this thesis defines a methodology to design, develop, implement, and apply statistical analysis and data mining tools to a Data Warehouse of healthcare metrics. With the DoD healthcare system undergoing significant changes, the ability for senior healthcare officials to have access to a wide range of management information in one central data repository can greatly enhance their decision making capabilities. This framework can be used as the foundation for moving DoD healthcare into the 21st century.

The major challenges involved in designing a data warehouse using this methodology are to identify the critical data needed to generate the data warehouse schema. The relevant data suggested in this thesis are healthcare performance metrics. Therefore, a team should be assembled that represents each functional business area to identify the critical performance metrics. Once the data has been identified and collected, it is straightforward to build a DSS for tracking performance metrics. The DSS consists of decision metrics, Geographical Information System (GIS), World-Wide Web and Online Analytical Process (OLAP) tools, Data Mining and statistics and forecasting components. Finally, it is very important to select the right people and proper tools, technology, and equipment to construct a corporate data warehouse architecture using a PMDSS.

14. **SUBJECT TERMS** Performance Measures, Performance Metrics, Data Warehousing, Healthcare Data Warehouse, Performance Measurement, Metric Based Framework.			15. **NUMBER OF PAGES** 134
			16. **PRICE CODE**
17. **SECURITY CLASSIFICATION OF REPORT** Unclassified	18. **SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	19. **SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	20. **LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18 298-102

Approved for public release; distribution is unlimited

A DATA WAREHOUSE ARCHITECTURE FOR DoD HEALTHCARE PERFORMANCE MEASUREMENTS

Jimmy E. Francis
Lieutenant, United States Navy
B.S., Troy State University, 1990
M.S., Webster University, 1994

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL

September 1999

Author: Jimmy E. Francis
Jimmy E. Francis

Approved by: Daniel R. Dolk
Daniel R. Dolk, Thesis Co-Advisor

Gregory D. Hildebrandt
Gregory Hildebrandt, Thesis Co-Advisor

Dan C. Boger
Dan C. Boger, Chairman
Information Systems Academic Group

ABSTRACT

In order to improve business practices and access to healthcare related information within the Department of Defense's Healthcare System, a Performance Metrics Driven Decision Support System (PMDSS) framework has been designed using Data Warehouse technology. As part of this performance metrics-driven framework, this thesis defines a methodology to design, develop, implement, and apply statistical analysis and data mining tools to a Data Warehouse of healthcare metrics. With the DoD healthcare system undergoing significant changes, the ability for senior healthcare officials to have access to a wide range of management information in one central data repository can greatly enhance their decision making capabilities. This framework can be used as the foundation for moving DoD healthcare into the 21st century.

The major challenges involved in designing a data warehouse using this methodology are to identify the critical data needed to generate the data warehouse schema. The relevant data suggested in this thesis are healthcare performance metrics. Therefore, a team should be assembled that represents each functional business area to identify the critical performance metrics. Once the data has been identified and collected, it is straightforward to build a DSS for tracking performance metrics. The DSS consists of decision metrics, Geographical Information System (GIS), World-Wide Web and Online Analytical Process (OLAP) tools, Data Mining and statistics and forecasting components. Finally, it is very important to select the right people and proper tools, technology, and equipment to construct a corporate data warehouse architecture using a PMDSS.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENT

I would like to thank LCDR Mary Lacroix and the CEIS program office for assisting me with the data used in Appendix B to conduct the analysis portion of this thesis. I would also like to thank LCDR Ellen Roberts and LT Ruby Tennyson for their assistance. I would like to extend a special thanks to Dr Daniel R. Dolk and Dr Gregory Hildebrandt for their guidance, direction, and leadership throughout this entire thesis process.

I. INTRODUCTION

A. BACKGROUND

The Department of Defense (DoD) healthcare organization is one of the largest and most unique systems in America with its beneficiaries located worldwide. Prior to the Gramm-Rudman-Hollings Act, there were military bases and corresponding medical treatment facilities located around the world. The resulting base closures and reductions in military manpower, healthcare facilities, and resources are driving the military towards new ways to provide healthcare to its beneficiaries. DoD needs to design and develop a corporate database and an associated data warehouse to make patient care information be immediately available to healthcare provider or caregiver at the right time. A primary objective in this movement is to restructure the DoD healthcare system to improve access to patient care information, improve record tracking, and eliminate the islands of systems and databases. The functional advantages of such database or data warehouse would greatly enhance access to patient care information, streamline patient record tracking, reduce the amount of physical storage space needed, and establish a foundation for an integrated electronic healthcare delivery system. With unlimited availability to a patient's historical information, healthcare providers could vastly improve their quality of service.

Historical information could assist analysts in analyzing medical conditions, identifying disease patterns, and embarking healthcare trends in the DoD beneficiary population. Information stored in this database can also be used to design a preventive health care management initiative, based on the data warehouse and developing data mining for decision support.

A centralized database could also be used to create an electronic interface to support the sharing of information among private industry partners, Preferred Provider Networks, Healthcare Management Organizations (HMO's), Department of Veteran Affairs, other suppliers and facilities.

DoD healthcare faces several problems with its current information systems design. The following problems have been identified:

- Individuals must maintain a physical copy of their own medical records.

- Hospitals can't share information.
- Inability to view information from an enterprise perspective.
- Duplication of cost to maintain "islands" of systems.
- Inability to conduct analysis to support operational, tactical, and strategic decisions.

Technology and policy changes are forcing our private and public organizations to improve the flow of information to shareholders, partners, and competitors. Keeping up with current technology is no longer an option, but rather a requirement in order to survive in today's competitive environment. DoD must take the lead in designing a healthcare information system that supports "open" system standards in order to eliminate inefficiencies in its systems. DoD must become more effective and efficient, by identifying the information requirements and designing and creating a system that supports information sharing with its stakeholders and partners. This system should provide sufficient information that will eliminate redundancy and frustration, reduce data storage duplication, improve the quality of patient care, provide a vehicle to support strategic planning, and establish a medium to measure quality and support the growth of a competitive healthcare delivery system.

By creating a standard integrated database or data warehouse, information will be readily available to support the delivery of healthcare and help a road map to support strategic initiatives. The Composite Healthcare System (CHCS) is an excellent foundation for future designs of an integrated healthcare delivery system. Its main drawback is that information is only available to a specific region or facility, and a complete record of a patient's history is not available. For example, if an active duty member stationed at Bethesda, Maryland is transferred to Monterey, California, he must hand carry his medical records because access to the most basic medical information is not available electronically to another hospital. If the records are misplaced, there is no method for recovering the entire medical record. Individual hospitals retain only two years of electronic patient information, and prior to 1982, all information was handwritten. The information stored at the previous duty station is not available to current healthcare providers and caregivers. This loss of data continuity illustrates the dangers surrounding islands of systems. These individual systems store information that is only available to specific MTF's.

Another example demonstrates further difficulties with isolated systems. Personnel on vacation rarely carry their medical records. If they get in an accident and a medical emergency occurs, the treating physician cannot gain access to their past medical history in a timely manner. Access to healthcare information during a life-threatening situation is detrimental to healthcare providers and patients.

Another challenge arose when the Department of Defense was tasked with ensuring that active duty service members were administered the anthrax serum. To document that all personnel had been administered the complete anthrax serum, the Navy, Army, Air Force all decided to store this information in a service-specific system. Of course, when the DoD was required to create an integrated report providing vaccination status to senior military officials, each service was then faced with the problem of integrating its data. The question arises, "why not store all information in the same location, specifically in the CHCS?" Unfortunately, CHCS was not designed to track this type of information and modifications to accommodate this requirement would be very expensive. This is yet another example of the inefficiencies of storing information in systems that are not integrated. DoD healthcare professionals along with the Chief Information Officers (CIO) must work together with their private industry counterparts to streamline processes, improve the delivery of healthcare, and define methods to share information economically.

Many organizations are refining business practices using information technology to improve collection, delivery, access, and dissemination of information. DoD healthcare institutions will also have to re-engineer their business practices to survive the economic pressures placed on them by DoD initiatives and Congressional mandates. For DoD to negotiate contracts and agreements with partners and compete with industry and competitors, it needs to have the capability to analyze cost of patient care, provide status on health conditions, and demographic for all DoD beneficiaries. Therefore, DoD needs timely access to diagnosis, prescription, referrals, and other related healthcare costs affecting its healthcare budget. One solution is to develop a joint healthcare enterprise data warehouse, which not only stores information for all DoD healthcare facilities but also interacts with other public and private institutions. This data warehouse could be used to gain insight on past and present events, which dictates the cost of effective delivery of healthcare.

B. OBJECTIVES/RESEARCH QUESTIONS

The objective of this research is to design a data warehouse architecture to integrate DoD healthcare information for decision making. This is separate from the task of building an integrated corporate database for storing operational data; the data warehouse is analytically oriented rather than operationally focused. This research will suggest a design for a data warehouse framework based upon using performance metrics which can be used to integrate clinical, financial, logistical, human resources, and administrative data onto one platform. A data warehouse can dramatically assist in developing outcome and population based studies of managed healthcare products and services. To determine the degree which a data warehouse would improve the delivery of healthcare and assist senior DoD officials with the management of healthcare information, this research will:

1. Identify the current shortfalls in access to patient care information.
2. Document the current flow of patient care information.
3. Discuss how an organization can develop a data warehousing system based upon metrics.
4. Identify the information needed to be stored in the data warehouse.
5. Develop a data model that can effectively implement a healthcare performance management system.
6. Demonstrate how access to a healthcare performance management system can help senior DoD healthcare officials make strategic decisions.

C. METHODOLOGY

The approach is to identify initially the major any information shortfalls that exist within the current environment and document that environment. The next step is to define the logical process used to identify data requirements for a data warehouse environment. The process identifies the performance metrics needed in each business area to support their decision making. The metrics identified will be used as the foundation for determining the appropriate data elements to generate the data warehouse database(s). Prior to generating the database, a filtering and cleansing process must take place to validate data needed to support DoD healthcare decisions. Once the data is available in the data warehouse, a PMDSS model is used to define tools needed to explore and analyze the data. After the data warehouse architecture or framework has been established, a PMDSS a statistical analysis is conducted using SPSS to demonstrate some of the steps required to support data preparation and conduct an analysis on data stored in the data warehouse.

D. SCOPE AND LIMITATIONS

This thesis will identify different approaches to the design and development of a corporate data warehouse. A database structure will not actually be designed nor implemented. A high-level data model will be created to depict the information requirements and the processes involved in designing a corporate data warehouse. The information requirements will be determined using several different strategic approaches. The operational corporate database is not considered.

E. THESIS ORGANIZATION

Chapter II describes the problems of not having an integrated database which stores all healthcare related information and outlines the current environment. Chapter III describes the use of metrics and performance measures as the underlying basis of a data warehouse environment. Chapter IV documents the methodology for designing and developing a Corporate Database Warehouse. Chapter V explores the use of SPSS, a statistical analysis package, as an application tool to analyze data stored in the data warehouse. Conclusions and recommendations are given in the final chapter.

THIS PAGE INTENTIONALLY LEFT BLANK

II. CURRENT SYSTEM ENVIRONMENT

A. DESCRIPTION OF CURRENT ENVIRONMENT

In 1982, DoD Healthcare officials, in conjunction with a congressional mandate, sanctioned DoD Healthcare institutions to develop a computer information system to store, manage, and retrieve healthcare related information. This system came to be known as the CHCS, and is used by most DoD healthcare facilities. CHCS's purpose is to improve the tracking of patient related information gathered from its five modules Patient Administration, Laboratory, Pharmacy, X-ray, and Clinical related information.

1. Patient Administration Subsystem

The Patient Administration (PAD) module tracks administrative tasks within the Medical Treatment Facility including registration, admission, disposition, and transfer of patients. The PAD module is very important in collecting and monitoring patient's vital signs, bed occupancy rates, and managing the hospital census. The PAD module can also be used to create and edit ward locations, to search for and identify/merge duplicate patient records. The admission, disposition, and transfer clerks use this module to review pending actions, correct and cancel admission, disposition, and transfer transactions, place patients on absence status, review bed availability, reserve and release beds, process mass casualty events, and generate reports.

Within the PAD module is the Registration Output Menu, which is used to track medical records and the movement of patients. It also supports the basic file room functions associated with lending and retrieving of medical records as well as the advanced file room functions associated with supervisory activities. The inpatient and outpatient records functions are used to create and checkout outpatient records, flag records as missing, print labels, and process inquiries about a patient record.

The Advance Record Room Functions provide the capabilities of deleting records and creating Transfer Retire Lists of record to be removed from the active database. The automation record of patient's files operates very much as an automated library system where the flow and accountability of records resides on the database at all times. Finally, the Medical Services Accounting (MSA) functionality supports recording, collecting, and tracking money owed to or collected by the facility. This tracking includes charges for inpatient and outpatient services, dining hall collections, and charges to insurance companies. The PAD module is vital in managing the administration of each healthcare facility.

2. Laboratory Subsystem

The Laboratory module provides practical tools for assisting lab personnel. It integrates lab results into CHCS for use by health care providers (HCP), accounting, and administration. This module enables lab personnel to perform support tasks for government medical facilities such as collection management, order entry, management of samples shipped to other labs, result entry certification, quality control, and administrative support. It also improves the access to patient's lab results for Healthcare Providers (HCP) internal to the system. It also enables HCPs from the ward to enter orders, access test status and results, receive notification of problems running a test, and inquire into patient test history. The workload data from the lab module is used to create lab reports and provides Medical Expense and Performance Reporting System (MEPRS) workload counts for a single work element or a group of work elements. The lab module is vital to the management, control, and provides access to patient's lab results.

The lab module has the following features:

- Integrated data management that supports clinical pathology, anatomic pathology, microbiology as well a nuclear medicine, pulmonary, research, and respiration tests.
- Supports distributed order entry as well as result transmission from automated instruments to reduce transcription errors.
- System-generated work documents, labels, and reports to reduce clerical tasks performed by technicians.
- Generates management reports that encompass workload statistics, quality control, and result monitoring and task-related reports for lab personnel that encompass workload management, worksheets, and overdue procedures.

- Supports the management of test result monitoring through the flagging of results outside the normal range and of radical variances in results for a single patient.

- Maintains quality control.

3. Pharmacy Subsystem

The Pharmacy module is used to input and process inpatient and outpatient medication orders and prescriptions. Its function is to record and maintain formulary files, bulk and clinic issues, and to produce a variety of pharmacy-related reports. The Pharmacy module has seven main menus: Intravenous (IV), Narcotics, Unit Dose, Outpatient Pharmacy, Inpatient Order Entry, Formulary, and Support Functions. The Pharmacy module plays a major role in keeping track, controlling, and documenting access to prescription and control drugs.

Based on information documented at www-nmimc.med.navy.mil/chcs, the following lab functions are listed and explained below.

- Intravenous function supports processing current and future IV orders; mixing and dispensing sterile IV products; and creating IV recipes.

- Narcotics function supports dispensing and tracking of narcotics.

- Unit Dose function supports processing and distribution of current and future inpatient medications.

- Outpatient Pharmacy function supports the overall processing of outpatient orders including filling dispensing, refilling, renewing, holding, forwarding, reactivating, discontinuing, and canceling of prescriptions.

- Inpatient Order Entry function supports the overall processing of inpatient medication orders including order entry, maintenance, verification and printing, of lab and reports.

- Formulary function supports all aspects related to creating and maintaining a formulary and formulary items including adding a new drug, formulary maintenance, establishing minimum and maximum dosing parameters, entering and editing a drug authorization key, and creation of a formulary group.

- Support Functions supports entry and definition of bulk and clinic issues, creation of Ad Hoc labels, and stock item file maintenance.

4. Clinical Subsystem

Finally, the Clinical module provides the capability for physicians, nurses, and other HCP's to enter and modify clinical orders related to data used to document direct and indirect aspects of patient care. The data provided by all other CHCS functions are accessible through this module. Therefore, the clinical module functions as the hub of CHCS integration. Healthcare professionals use this module to develop patient care plans, enter new orders or modifying existing orders, review clinical results, and access online reference materials to support nursing procedures, patient instructions, laboratory tests, radiology procedures, ancillary/consult procedures, and formulary drug files. Clinical functions are also used to monitor the quality of patient care provided by an MTF, assisting the facility to meet standards established by designated licensing and accrediting bodies.

B. OPERATIONAL INFORMATION SYSTEMS

Furthermore, because of the parochialism involved in developing these systems, the data definitions relate only to the narrow departmental, functional, or business area view of the data. Users and analysts have traditionally used information delivery systems to view data. The same information delivery systems will play a major role in helping identify data to be stored in the data warehouse. The chief problem of data identification is that not all business areas define data the same. For instance, talk to administrators, and it is clear that an appointment is "booked" when a patient indicates intent to receive care from the healthcare provider. Talk to clinical personnel, and it is clear that an appointment is booked when it is scheduled for and the supplies and materials are ordered. Talk to an accounting clerk and it is clear that an account receivable is owed when the hospital receives payment for it. In fact, there may be literally weeks or months between the earliest interpretation of "booking an appointment" and the latest one. This lack of coherence within the organization on the meanings and derivations of the many terms used to describe the business. Given these varying interpretations, one can see how different business area managers would enact very different courses of action.

These information delivery systems listed below provide an overview of some of the main healthcare operational systems that will feed data to the data warehouse. These systems will be used as the foundation for identifying the data elements needed to create the data warehouse database design and the metadata database design. First, the Clinical Business System provides the capability for physicians, nurses, and allied health professionals to enter and modify clinical orders and related data, and to document direct and indirect aspects of patient care. Below are some of the of the Clinical Business Area Systems:

1. Ambulatory Data System (ADS)

ADS provide ambulatory data as a by-product to the health care delivery process. It was designed to capture patient encounters and ambulatory data with more elaborate detail than CHCS. It supports basic clinical and administrative purposes, including analyses for managed care, epidemiology studies, billing, and severity/acuity case mix analyses. It stores patient diagnostic and treatment data into a single record readily accessible to all authorized users to manage. It improves and enhances billing and reimbursements for TPOCS and documents care needed to support Residency Review Committee's (RRC's) and credentialing requirements.

2. Immunization Tracking System (ITS)

The Immunization Tracking System (ITS) is a central repository used to document immunizations for DoD beneficiaries.

3. Automated Central Tumor Registry System (ACTURS)

The Automated Central Tumor Registry System provides the medical community with the ability to register and track patient records, treatment, outcomes and other data for cancer patients. ACTURS is a single system designed to help medical facilities meet requirements for Cancer Programs set forth by the American College of Surgeons, as well as Service and State standards.

4. Composite Healthcare System (CHCS)

CHCS was designed to assist healthcare providers and administrators with the management and delivery of quality care to all DoD health care recipients. CHCS provides flexible medical data processing capabilities for DoD medical treatment facilities (MTFs). It is the primary system used throughout DoD healthcare facilities to support clinical management functions.

5. Defense Blood Standard System (DBSS)

DBSS was designed to provide blood management functions for every aspect of the Armed Services Blood Program. It manages blood donors both in-house and at the appropriate collection sties. It also provides patient and transfusion service management, testing, inventory and shipment of blood products. DBSS was designed for use at any location that deals with blood collection, processing and/or distribution.

6. Defense Eligibility Enrollment Reporting System (DEERS)

DEERS is a congressionally mandated program, which was designed to provide confirmation of eligibility for medical benefits.

Second, the Logistical Business Area Systems provides the capability for physicians, nurses, corpsman, and allied health professionals to order, track, and procure clinical supplies and equipment related data to the support of direct and indirect aspects of patient care. Below are some of the of the Logistical Business Area Systems:

7. Automated Procurement System (APS)

APS automates requisition input and tracking, procurement actions, receipt processing, and management information within the purchasing department at Medical Treatment Facilities. APS was also designed to support acquisitions and contracting of logistics at hospitals.

8. Central Processing and Distribution Systems (CPD)

CPD is used to keep track, manage, and capture financial accounting of stock room exchanges of surgical supplies. This system was designed to ensure healthcare providers receive surgical supplies timely. In addition, CPD allow comptrollers to charge departments their actual incurred expenses on a daily basis.

9. Medical Inventory Control System (MICS)

MICS automates the inventory control and financial management of Navy medical stock fund warehouses. MICS is an inventory control system designed to support medical treatment facilities directly supported by the Defense Business Operating Fund. Its primary functions are automating stock rooms, providing on-line logistic management reports, automatic or manual replenishment, electronic transfer of requisition and status information, and supply billing and financial inventory reporting.

10. Biomedical and Facilities System Project

The biomedical system is used to manage and track preventive and emergency maintenance services for biomedical equipment. It also documents resources expenditures, repair part inventories, and property budgetary accountability.

Next, the Human Resource and Financial Business Area Systems provides the capability for collection, tracking, and reporting of required provider data for the crendentialing office of physicians, nurses, and allied health professionals. Some of these systems are also used to help manage the financial aspect of each healthcare facility. Below are some of the of the Human Resource and Financial Business Area Systems:

11. Claims Processing System II (CPS II)

CPS II is a Bureau of Medicine and Surgery (BUMED) standard system. CPS II was designed to manage, track, and monitor non-naval health care services to active duty Navy and Marine Corps members.

12. Department of Defense Workload Management System for Nursing (DoD-WMSN)

The DoD-WMSN is a microcomputer-based patient classification and nurse staffing system operational in all Navy, and Air Force hospitals. It stores patient acuity category by occupied bed day, work center, actual scheduling of staffing by work centers, and patient demographics. This system automates the daily calculation of patient acuity classification and generates summary reports useful for staffing and manpower decisions at the hospital, Bureau of Medicine and Surgery (BUMED), and the Department of Defense (DoD).

13. Medical Expense and Performance Reporting System(MEPRS)

MEPRS is a management support tool that provides standardized reporting of expense, manpower, and workload data by the Department of Defense (DoD) medical and dental facilities at the medical treatment facility (MTF) level. MEPRS provides MTF Commanders with a system that contains sufficient cost detail to evaluate managed care alternatives (make, buy, or transfer), enhance third party reimbursements, and support the analysis of variations in patterns of treatment among providers.

14. Third Party Outpatient Collection System (TPOCS)

TPOCS was designed for identifying, recording, billing, and collecting reasonable costs for inpatient and outpatient care.

15. Centralized Credentials Quality Assurance System (CCQAS)

CCQAS is a worldwide tri-service credentialing information system designed to assist Medical Treatment Facilities (MTFs) in the collection, tracking, and reporting of required provider data to support Credentialing. CCQAS is critical for tracking and storing information about a provider's demographics, education, licenses, certification, affiliations, malpractice/insurance data, and adverse action reporting. CCQAS also provides MTF Commanders the ability to track provider's medical readiness training information.

16. Defense Medical Human Resource System (DMHRS)

DMHRS was designed to track scheduling, training, managing readiness, matching personnel with tasks, labor cost analysis and other human resources issues.

17. Expense Assignment System IV

EAS IV manages standardized workload, expense and manpower data for the Department of Defense (DoD) at the medical and dental treatment facility (M/DTF).

EAS IV provides MTF/DTF commanders with of day-to-day healthcare and resource management activities.

18. Standard Personnel Management System II(SPSS II)

SPSS II provides manpower, personnel, education and training, mobilization planning, and expense distribution functions for the Bureau of Medicine and Surgery (BUMED).

19. BUMED Manage Information Management System (BUMIS)

BUMIS is used to support Human Resources for the Medical Department at the Bureau of Medicine and Surgery. BUMIS provides major support for the BUMED Personnel and Analysis Department by tracking Officer Personnel and Billet Information of Medical Service Corps, Dental Corps, Nurse Corps, Medical Corps, Technical Nurses, and Physician Assistants. The database contains Information on current sub-specialties, education, assignments, location, special pay qualifications, etc.

Finally, the Executive Information System provides the capability for physicians, nurses, healthcare administrators, and allied health professionals to manage the overall operations of the healthcare facility using comparative data from other facilities.

20. Executive Information Systems (EIS)

EIS is a computer-based system designed to provide timely and accurate information for Navy Medical Department executives who make major decisions concerning the allocation of resources. EIS incorporates the standard performance measurements defined by BUMED, which will be used to indicate successful resource allocation, as well as identifying problem areas for resolution.

The systems listed above are by no means a comprehensive list of all the information systems used in DoD healthcare to support the delivery of patient care; but they are representative of the different types from which data will be extracted to create the data warehouse. In addition, these systems use varying database engines, schemes, data dictionaries, or standard data formats. These differing features create a complex integration environment. That is a common problem and challenge that many organizations face.

C. CHCS DATA-CENTRIC VIEW

The five modules, which define the CHCS system, are vital to management and administration of a healthcare facility. Figure 1 illustrates the integration of DoD healthcare systems and their interfaces used to make strategic decisions. It documents the interaction between and among systems used in DoD to capture healthcare information. Although most of the patient related information is stored in CHCS, there is limited access to it. The Corporate Executive Information System (CEIS) interfaces with CHCS to collect summary data to support strategic decision making. The Ambulatory Data System (ADS) interfaces with CHCS to collect ambulatory diagnostic and procedural coding data. The Third Party Outpatient Collection System (TPOCS) interfaces with CHCS to retrieve inpatient and outpatient billing information. TPOCS system access to CHCS is currently manual; data transferred from CHCS is re-entered into TPOCS.

The Expense Assignment System (EAS III) interfaces with CHCS to collect workload, expense, and manpower data. The Workload Assignment Module (WAM) is a new module being developed to interface with CHCS to automatically transfer information to the Expense Accounting System.

The systems listed below are being used to detect trends, utilization rates, establish benchmarks, and manage the health of the DoD healthcare population. The user interfaces for CHCS are easily usable providing the user has been trained on the system. CHCS can be considered the pioneer system that brought information management to DoD healthcare facilities. CHCS has a myriad of options and capabilities, that are being used to improve tracking, storing, and access to patient-related information. Since the development of CHCS in the early 80's, there have been significant improvements in the technology arena. Most technologies now support graphical user interfaces (GUI), window-based environments, and Worldwide Web connectivity.

There is a vast quantity of data being transferred across system interfaces. Unfortunately, CHCS is nearly 15 years old spanning three computer technology generations. As we enumerate below, there are severe shortcomings of the current information system technology which need to be addressed to bring the DoD healthcare environment into focus. In order for senior DoD healthcare officials to make sound decisions around providing the highest quality healthcare, they must be able to use this information, much more efficiently and effectively.

1. Corporate Executive Information System (CEIS)

CEIS is a computer-based system designed to provide timely and accurate information for Navy Medical Department executives who make major decisions concerning the allocation of resources. This system's design was based on the twelve business functions used to identify critical success factors to support the successful management of the Navy healthcare delivery system. CEIS incorporates the standard performance measurements defined by Bureau of Medicine and Surgery (BUMED), that will be used to indicate successful resource allocation, as well as identifying problem areas for resolution. CEIS also provides detailed information relating to CHAMPUS, Direct Care and DEERS information by catchment area or locale. CEIS uses a product called TRENDSTAR to receive and extract data from Integrated Databases monthly. CEIS also has Customer Education & Marketing, Professional Training and Education, Healthcare Planning, Medical Readiness, Provider/Network performance, Utilization and Quality Management, and Preventive Services/Health Promotion modules.

SYSTEM OVERLAPS/REDUNDANCIES

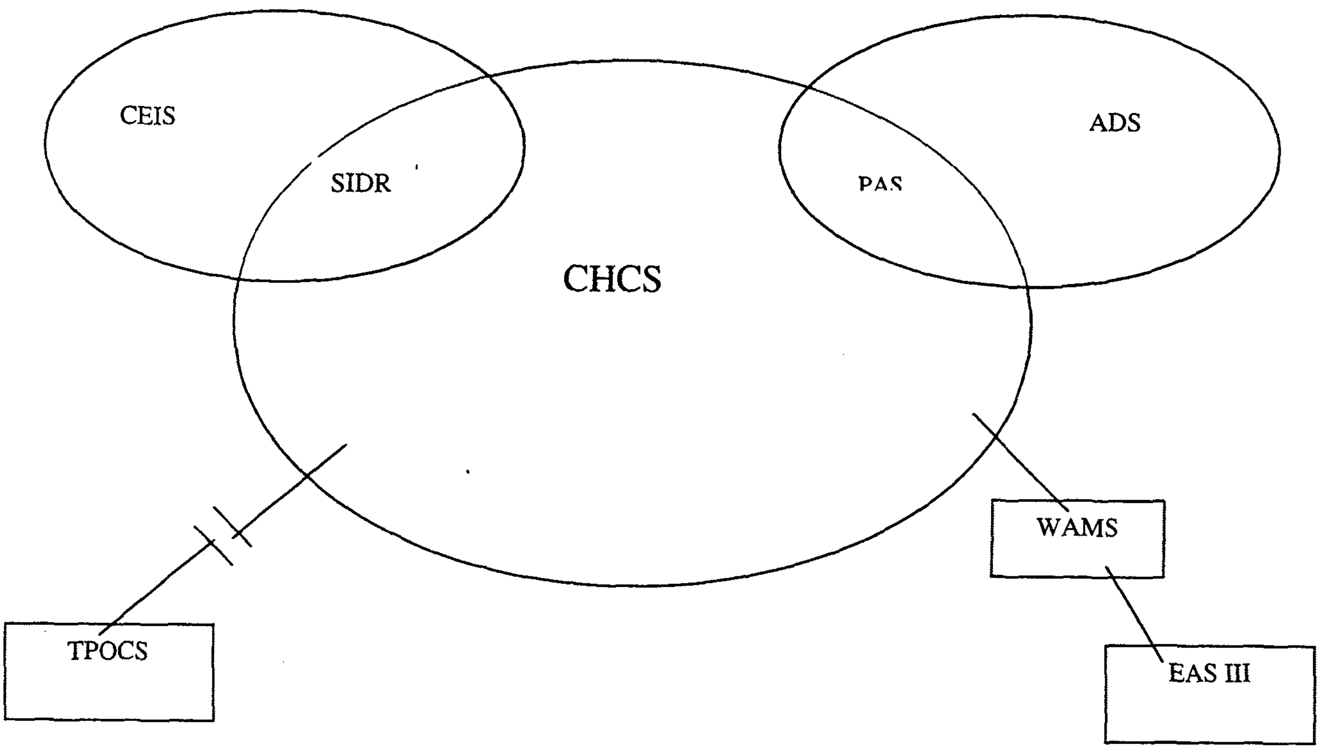


Figure 1 CHCS Data-Centric View

2. Ambulatory Data System (ADS)

ADS was designed to collect ambulatory diagnostic and procedural coding data as delineated by the International Classification of Disease, 9th edition (ICD-9) and Current Procedural Terminology (CPT) codes which CHCS does not capture. ADS allow the provider access to automated outpatient diagnosis and treatment information down to the individual patient level. This level of detail can then be used to assess actual disease rates of patient populations, evaluate the effectiveness of treatment protocols, project preventive healthcare requirements, identify patient panels, assist residency review committee evaluations, and so on.

ADS is based on a machine scanned form similar to "master bills" used in civilian medical practices. The lists are tied to specific appointment types in CHCS. In addition, ADS is used to collect patient demographics, insurance, and appointment information from CHCS and prepare an encounter form for each appointment to be verified by the patient prior to their visit. The current problem is that physicians feel they spend too much time completing and filling out bubble sheets to support ADS. This example documents shortcomings that exist in the current environment, which results in the collection of duplicate data thus adding administrative overhead. CHCS was not designed to provide the reporting and analysis capability of ADS; consequently, information is being stored in two systems. Data collected by ADS can be provided to support other Military Health Service System (MHSS) programs and projects. At present, automated extracts are passed to the TPOCS for local use, and a standard ambulatory data record is transmitted to a central database at Fort Detrick, MD and other interfaces are being developed to automate Medical Expense Performance and Reporting System (MEPRS) and Expense Assignment System (EAS) workload reports.

3. Third Party Outpatient Collection System (TPOCS)

TPOCS is used by all MTF's to implement an aggressive program to ensure that an effective mechanism is in place for identifying, recording, billing, and collecting reasonable costs for inpatient and outpatient care. The broken line in Figure 1 indicates that TPOCS is a manual interface. Therefore, data transferred to and from TPOCS must be manually transferred and is controlled by human interaction.

4. Expense Assignment System (EAS), Version IV

EAS IV is used as a management support tool that provides standardized reporting of workload, expense and manpower data by MTF's. EAS IV provides MTF/DTF commanders with a fully integrated component of day-to-day MTF healthcare and resources management activities. EAS IV determines costs, workload, and manpower at both the cost center and product level. It also will provide MTF commanders with a system that contains sufficient cost detail to evaluate managed care alternatives, identify fixed and variable cost, calculate MTF marginal costs for healthcare services, enhance third party reimbursement, and support the analysis of variations in patterns of treatment among providers.

5. Workload Assignment Module (WAM)

The Workload Assignment Module was designed to provide a mechanism within the CHCS that automates the process of reporting medical treatment facilities' workload in the Expense Accounting System. WAMS was developed to reduce the manual efforts of the staff hospital, become a single source for data entry, and create a direct link to CHCS to a job order number in the accounting system.

D. EXISTING DATABASE LIMITATIONS

The current CHCS environment presents data access challenges and limits healthcare providers ability to critical patient care information. Further, the system becomes outdated as industry and technology advance. As mentioned earlier, the operational databases storing patient care information are not readily shareable across sites. This inability to share information creates multiple problems particularly when a patient transfers from one facility to another. Similarly, when patients are referred to an outside physician or even a physician within the system, they are forced to hand-carry their medical records. Not only does this place an additional and unnecessary responsibility on the patients, but it is also a significant inconvenience.

CHCS was implemented and designed in a manner that demands each site bear the considerable expense of running and maintaining its own computer platform, including hardware, maintenance, software, life cycle, and support costs. While it is critical to have a computer-based information system, it is also important to identify and plan upgrades and modifications to meet the changing business needs of the organization. Administrative, fiscal, human resources, logistics, and clinical drivers demand information exchanges among physically isolated facilities; while access to eligibility, coverage, and co-payment information is critical to a doctor's office staff. These areas must be addressed in future developments.

There is a tremendous opportunity cost of not being able to share information among and between hospitals, physicians, caregivers, and healthcare management organizations (HMO) resulting in artificial limitations such as:

- Inability to access patient care information.
- Inability to streamline tracking of patient records outside the main hospital.
- Limited ability to control redundant data and the subsequent amount of storage needed.
- Proliferation of islands of systems and databases which currently do not interact with one another.
- Inability to create an electronic interface for private industry partners, Preferred Provider Networks, Healthcare Management Organizations, Department of Veteran Affairs, other suppliers and facilities.
- Difficulty in creating a foundation for designing an integrated electronic healthcare delivery system.
- Limited ability to assess adequately the quality of healthcare that is being delivered.

E. DEVELOPMENT TRENDS

Both private and public healthcare organizations are currently seeking ways to access, retrieve, modify, share, and protect healthcare information. With managed care and computerized patient records becoming the norm, sharing data and information will play a major role in the future delivery of healthcare. The trend for healthcare organizations is also changing rapidly toward regional systems and away from single, independent entities.

According to one study (Rauber, 1998), Healthcare Management Organizations are failing because the healthcare industry has not invested enough in critical information technology. HMOs must find ways to implement electronic data interchanges and electronic commerce. Health Services Corporation of America released an electronic catalog on the Internet with more than 600,000 products to support electronic purchasing. Their objective in adopting electronic purchasing is to reduce errors, provide fast service, and cut back on paperwork. However, several purchasing executives feel that electronic commerce has been slow because of technical difficulties, as well as the inertia of hospitals, vendors, and distributors.

The Veterans Hospital Alliance (VHA) is publishing medical information on the Internet as a part of a marketing campaign to boost consumer allegiance to its member hospitals and their affiliated physicians. The Internet site provides consumer healthcare information on commonly prescribed medical tests and drugs, cutting-edge treatments and the latest medical news. The VHA hopes this effort will counter sinking public confidence in healthcare providers, and to separate its hospitals from the pack. In a continuing push to computerize the front lines of healthcare delivery, the VHA has also launched a service that provides hospital physicians access to basic patient information through a secure Internet connection. VHA's physicians feel that the exchange of everyday clinical data between hospitals and physician practices is vital. The VHA has focused its technology development on ambulatory facilities and physician offices, which have become arenas for cost management and patient information in emerging healthcare systems. (Rauber, 1998, pp. 54)

The VHA has also launched three programs to create Internet-oriented services and capabilities:

- VHAseCURE.net is a members-only Internet style infrastructure, called an extranet, that runs on a private route within the network. It connects members with one another and with VHA. It also aids in developing intranets, the internal capacity necessary for a healthcare organization to participate in an internet-style network.

- WebWorx is a package of templates allowing healthcare systems to establish World Wide Web sites aimed at consumers in their service areas. The initiative includes a set of health content topics fed through an Internet connection.

- Hospital physician data exchange was developed to make a hospital's basic clinical results and patient-specific data routinely available to its physician staff. (Rauber, 1998, pp. 56)

The Department of Veteran Affairs (VA), like many activities, is encountering problems with the processing of claims and billing insurance providers. Due to Congressional budget cuts, the VA's mission has changed substantially; previously the VA had never faced the tasks of billing, filing claims, and accepting referrals from other organizations. Historically, the VA has always been fully funded by Congress to support the medical needs of the veteran population. With today's limited funding, the VA is now faced with the requirement to develop ways to share vital healthcare related information with outside hospitals, HMOs, physician offices, and other VA hospitals and clinics.

Information technology is thus being recognized as a vital resource in the healthcare environment, and private and public healthcare organizations are seeking new opportunities to share information utilizing the Internet, electronic commerce (EC), and electronic data interchange (EDI). Information technology is the most effective way to improve the delivery of healthcare, reduce cost, and move from managed care to managed health in the course of re-engineering healthcare administrative processes.

F. INABILITY TO CONDUCT HELATHCARE ANALYSIS

The result of not having full access to operational data prevents analysts from correctly identifying patterns and recognizing areas for improvement. The current process of manually extracting data from one individual system to another often results in data inconsistency, unnecessary delays, and reliability errors.

While measures of effectiveness and performance metrics can be used to improve the delivery of quality healthcare, each facility must measure the same data. Among the many indicators are types of treatments conducted, automated critical paths, percentage and amount of cost reductions, declines in diagnosis testing, number of patient visits, patient demographics, risk management issues and the reduction in pharmacy cost.

In the healthcare community, EDI is utilized among a number of trading partners, including hospitals, healthcare networks, and managed care organizations to determine enrollment and patient eligibility. However, the sharing of information and data are not the only functions to improve the quality of healthcare. In order for healthcare organizations to implement quality improvement initiatives, they must focus their attention on establishing performance measures, identify measure of effectiveness, and begin benchmarking competitors performance to support the re-engineering of healthcare. The next chapter will address the steps for designing a performance measurement system.

III. PERFORMANCE MEASURES

A. INTRODUCTION

Today's healthcare environment faces economic and regulatory changes as well as patients who demand improved and advanced services. These challenges necessitate the delivery of high quality services and the use of information systems to predict and contain costs. The requirements for these systems are driven by the need to acquire, store, and analyze a broad spectrum of clinical, administrative, logistical, and financial data in a timely, flexible, reliable and secure manner. Therefore, new database solutions, utilizing highly scaleable, portable and customizable technologies are more suitable to the current, demands of healthcare. There are also requirements to ensure that healthcare facilities collect and store relevant data needed to support these challenges.

The Federal Government has made a commitment to improving the quality of services provided by federal agencies by establishing a Presidential Award for Quality and the Award for Quality Improvement. The Federal Government believes that improvement measures are important because they support accountability to the taxpayers, incorporate better internal management of government programs, and improve executive and congressional decision-making. Support for the move towards improvement originated with President Clinton.

> As we move towards the 21st century, federal agencies are working to set clear goals, making certain employees understand how these goals are linked to their own jobs and measuring results. [My] administration seeks a government that gets results for the American people by recognizing the potential of every federal employee and empowering each worker to achieve his or her best. In seeking to meet and exceed customer expectations, we are continuing to build a government that delivers the highest quality products and service to the public (Presidential Quality Award Manual , 1998, pp. 2).

The commitment for quality was also supported by Vice President Al Gore who stated, "Every organization needs a clear and cohesive performance measurement framework that is understood by all levels of the organization and that supports objectives and the collection of results" (Presidential Quality Award Manual, 1998, pp. 2).

Organizations competing for the Presidential Award for Quality are required to demonstrate consistent application of quality approaches throughout the organization. They must exhibit sustained performance improvements in virtually all levels of the organizations for several years. To receive the Award for Quality Improvement, the organization must demonstrate a well-developed process approach supporting the businesses core competencies.

The DoD healthcare system is striving to implement its own quality improvements through the development of a performance measurement system. These measures will support the movement from a "fee for service" arrangement to a consistently higher quality "managed care" environment. A well-design performance measurement system will enable identification of the relevant performance information needed to support a Managed Healthcare System (MHS).

DoD healthcare leadership must find ways to implement cost control measures while expanding access to quality care. One way to improve these processes is to clearly define and monitor measures of performance, measures of effectiveness, and performance indicators that accurately reflect the status of healthcare delivery. The focus must shift to identification of the necessary steps to collect and analyze data to support strategic decision making.

A recommended procedure for specifying performance measures is to first identify overall strategic goals, and then to identify the relevant performance indicators that support each of the business areas. Problems can arise because each manager has a unique and often different, perspective depending upon their particular business area. For instance, a clinician's main concern is vastly different than a nurse's. An administrator, on the other hand, may be attuned to yet another area of concern with its own appropriate measures. Each profession concentrates on distinctly separate procedures within its own focus.

Finally, the most important concern of the patients is relief from their pain or illness and ensuring that they receive the best healthcare for the least cost. The bottom line is that physicians want the proper number of visits for the same illness, nurses want to ensure that the care plan which they prescribe properly nurses the patient back to good health, administrators want the healthcare facility to operate in an economically healthy fashion, and patients don't want to waste their time, money, or resources. The desired outcomes also differ from physicians to nurses to administrators. In an environment of sometimes conflicting perspectives, the objective remains to illicit appropriate performance measurement or metrics to support healthcare improvements.

B. PERFORMANCE MEASUREMENTS

According to the Joint Accreditation of Healthcare Organizations (JCAHO), the basic tenets of healthcare performance measurement were established half a century ago in the U.S. for use in the external accreditation of hospitals. This required hospitals to qualify various aspects of their clinical, financial, logistical, administrative and management systems. Today, the healthcare community has grown well beyond hospitals to include physician practices, nursing homes, public health organizations, insurance companies, large employers, and health care systems. This has expanded the audience for accountability to include managed care organizations, policymakers, payors, researchers, employers, and advocacy groups. To remain relevant, healthcare providers must employ a structured and focused approach in developing systems and tools to demonstrate that they are providing services that improve the quality of life of the populations they serve.

While such efforts demonstrate current quality, they also must work to improve future efficiency and value to achieve and maintain cost containment. Performance metrics help healthcare organizations focus on answering three very basic questions:

- Are we doing the right thing (strategic / innovation)?
- Are we doing the right thing well (operational / efficiency)?
- How can we do the right thing even better (tactical / improvement)?

Performance measures also help management track organizational performance to ensure it aligns with operational, tactical, and strategical objectives. According to Lord Kelvin, "When you can measure what you are speaking about, and express it in numbers, you know something about it...[otherwise] your knowledge is of a meager and unsatisfactory kind" (1824-1907). Neely [et al 1995] defines performance measurement *as* the process of quantifying the efficiency and effectiveness of action; performance measure as a metric used to quantify the efficiency and/or effectiveness of an action; and performance measurement system *as* an appropriate set of metrics implemented in an

Management Information System (MIS) or Decision Support System (DSS) (Neely, Gregory, and Platts, Vol 15, No 4 (1995), 80-116.).

DoD healthcare professionals are no different than any other industry. They would like to use performance measures to compare themselves to other healthcare facilities in the region, city, or area. Listed below are examples of some of the strategic questions that may be posed by a Chief Executive Officer (CEO) or Administrator of a large hospital:

- How many malpractice suits has this hospital had this year, what was their total cost, broken down by clinic, and physician, and what percent were they of the overall operating budget?
- What is the leading treatment, by clinic and total cost, compared to other hospitals with the same bed capacity?
- What is the hospital profit margin for this year compared to the last 5 years?
- What is the overall patient satisfaction level for this year compare to the last 5 years?
- How do the hospital's key business indicators stack up against its local, regional, national, and international competitors?

The above questions are not easy to answer, nor is it easy to collect data needed to support them. However, with a sound performance measurement system in place, leadership can gather the information required to support its strategic decisions. "...organizations achieve their goals, that is they perform, by satisfying their customers [patients] with greater efficiency and effectiveness than their competitors." (Neely,1995) In the healthcare environment, effectiveness is defined by quality patient care and efficiency is represented by economical use of resources.

According to Dennis Van Langen, Chief, Engineering Support Division, Defense Information System Agency, organizations have always had a wealth of metrics at the line manager level, but a shortfall exists with senior management. Figure 3-1 depicts the value of metrics to senior managers versus line managers.

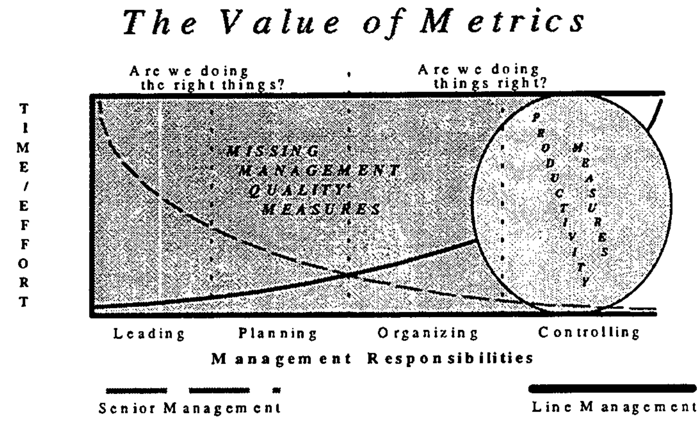


Figure 3-1 Value of Metrics

It illustrates that as management responsibilities move from leading, planning, and organizing to controlling, the value of metrics change. Senior managers rely heavily on metrics in their leadership and planning roles to support innovation and to determine if the organization is conducting the appropriate activities. Line managers, on the other hand, rely heavily on metrics in the day to day decision making at the operational level.

A vital point is that performance measures play a major role in establishing a framework to manage a quality organization. Organizations competing for the Malcolm Baldrige Award, International Standards Organization (ISO) 9000, or the Presidential Quality Award programs have to develop performance measures that demonstrate their productivity and effectiveness using a performance metrics approach. The awards fall to those organizations that measure values critical to their business objectives; those that measure everything without discernment are just measurement organizations.

C. DESIGNING A PERFORMANCE MEASUREMENT SYSTEM

The objectives of an appropriate performance measurement system are to promote focus and alignment, encourage communication, sustain a wide-ranging vision, and ensure that an organization's behavior is consistent with strategic goals. In addition, a balanced perspective on performance is essential to ensure that improvements in one area do not compromise performance in another. A comprehensive view of performance should incorporate all aspects of the organization's operations. Performance measurement systems succeed when they provide relevant facts and data about current performance, and show what needs to be improved, either immediately or in the future.

Based on experience, most Executive or Enterprise Information Systems were abandoned because they focused on narrow areas and were developed for top level managers. Executives need to see 'the big picture', concentrating on areas which need their attention. They should focus on measures that are aligned with strategic objectives and corporate goals. When organizations first set out to build a measurement system, there is often a tendency to incorporate measures for which data is readily available, however that may not be the appropriate help managers need to run the business. Instead, a structured process of measure identification should be followed. The process of defining a full scope measurement system consists of four main stages as described below.

1. Choosing a Quality Framework

There are a number of published frameworks which attempt to put measures into a pre-defined context, like the four perspectives of the Balanced Scorecard, or the seven pillars of the Baldrige Award. A group of senior DoD healthcare officials should be assembled to establish a framework to define the key strategic performance perspective to support the mission, vision, and goal statements. Most performance measures in the healthcare arena will be determined based on people, patients, clinical, logistics or finance.

2. Identifying Key Performance Areas

Within each business area, there should be a number of key business performance measures. In the clinical area, there may be various types of services, and associated key areas of performance, which are relevant. For example, one of the key performance measures might be to track the number of patients receiving a specific treatment. Each business area manager will similarly identify the key performance measures appropriate for his/her area. This process is repeated for each business area and the result is an overall context that can be used to help define, and subsequently align, performance measures.

3. Selecting Measures

The next step is to identify those specific measures or metrics that support the achievement of the business goals and strategies. Normally, this will be accomplished through a series of facilitated workshops that will include several steps:

a. *Definition of Metrics*

Metrics themselves have *primary* attributes which include definition, computational procedure, thresholds, periodicity, scale level, and associated "drilldown" dimensions, as well as *information system* attributes such as data sources and report distribution profiles that allow the metrics to be implemented in a data warehouse environment (Dolk, 4):

Definition. The metric definition is a text description fully describing the semantics of a metric and the organizational values and goals that it supports.

Computational Procedure. Defines how the decision metric is calculated or derived.

Dimension/Units. The unit of measurement used to represent the metric e.g., Cost/Unit Cost in dollars.

Thresholds. Decision metrics have upper and lower limits placed on them to gauge or monitor performance in a particular area of an organization. Thresholds can be manually prescribed, or statistically determined as in the case of quality control metrics. They can be displayed as simple visual cues; for instance, a patient's *WaitingTime* may be displayed as green, yellow, or red. They can also be depicted on a map depending upon whether it's above, within, or below one standard deviation for the healthcare industry. Each decision maker may assign different thresholds to each metric based on their respective business needs.

Periodicity. The frequency in which a metric is measured. A metric may be measured hourly, daily, weekly, monthly, or yearly depending how often it changes. For example, emergency rooms visits may be measured every hour or even every minute whereas hospital supply data may be measured on a daily, weekly or even monthly basis. The periodicity is the minimum time granularity for which a metric may be represented.

Scale level. The numeric value assigned to a metric. Metrics are assigned a scale level using ordinal, cardinal, interval and sometime ratios. The scale may be ordinal for example in a multiple attribute utility model such as the 100 most desirable hospitals to receive care in America, in which case the value provides ordering information. The scale may be cardinal in which case the differences between quantities can be precisely determined.

Drilldown Dimensions. Each metric has affiliated attributes that comprise an n-dimensional space for calculating the metric at various levels of aggregation. Two of the most common drilldown dimensions are location and time (Dolk, 4).

Data Sources: Source database(s) for retrieving all attributes necessary to calculate the metric and all attributes associated with drilldown dimensions. The data sources provide the basis for the underlying data warehouse.

Report Distribution Profile: This is the information delivery mechanism of the associated data warehouse describing who receives what reports at what levels of aggregation at which frequency for which metrics.

b. Create and Compile a List of Performance Measures

A comprehensive list of performance measures covering all aspects of an organization's performance should be established. Listed below are some of the frequently used measures in a healthcare organization.

- Total amount spent on surgeries.
- Percentage of hospital budget spent on training.
- Number of emergency visits per month.
- Percentage of hospital budget spent on salaries.
- Percentage of patient return visits.
- Measures that capture patient satisfaction.

c. Streamline the list of Performance Measures

The next task is to refine the performance measure list, filtering out those that are inappropriate. All performance measures should be critically reviewed to determine if they support the decision making process, represent the interest of each business area, and clearly illustrate the key issues of time, cost, patient care, and improve the delivery of quality healthcare. Organizations should reject measures which:

Do not initiate action, where a change in reported performance does not cause individuals or teams to act differently.

Do not provide good information, where it is unclear exactly what is being measured and how results should be interpreted.

Do not support the achievement of the strategic goals and objectives of the organization.

Do not have a goal tender, a person responsible for reporting, monitoring, collecting, and managing its progress.

Do have currently available data.

4. Alignment of Measures with Business Objectives

Once a list of key measures has been generated, these measures should be aligned with corresponding performance areas and business objectives. As an organization proceeds through this process, additional groupings of measures may be required. Due to differing significance, distributing or ranking them using weighting factors should differentiate measures. Each measure should also have an owner whose objective is to communicate to other stakeholders and managers the current status and implications associated with each metric. The goal tender is responsible for the accountability of the metric, but the owner is responsible for ensuring the data is accurate and up-to-date. Measurement owners should periodically provide updated commentaries that inform others about their indicators.

5. Examples of Healthcare Metrics

When an organization defines their performance landscape, it will include the business areas that should be measured. Some areas will have missing or inaccurate data and others may require new ways of capturing and extracting data from existing systems. Another important consideration is to design a reporting system with flexibility enough to allow easy integration of performance measures when accurate data becomes available. There are several metrics lists in Appendix A that represent real world applications used in healthcare organizations and recommended by the Joint Commission of Accreditation for Hospitals. These metrics have been revised using the attributes described in the Definition of Metrics section of this chapter. The target thresholds, periodicity, and drill dimensions are educated guesses to demonstrate the metric concept.

D. WHAT ROLE DO METRICS PLAY IN IMPROVING THE DELIVERY OF HEALTHCARE?

Government organizations of all types are experiencing dramatic change with respect to their strategic objectives. Many are moving from the traditional methods of operational efficiency to one of quality management with a focus on the customer, patient, continual improvement of processes, team approach, and data-driven processes using accurate and meaningful data. (Koehler and Pankowski, 1996, pp. 1). Figure 3-2 depicts the relationship between performance metrics and organizational mission, goals, and objectives. The foundation for executing the strategic plan is based on the ability to develop sound metrics.

The new quality initiatives require a different perspective on current and future patients. Spoiled by decades of success, during which patients merely accepted the services offered, healthcare providers must now grasp the requirements to satisfy patients needs. Providers must also educate patients, for if they do not, competitors will. The patient is the new reference point and priorities become easier to establish. In a quality-oriented healthcare organization, most work becomes focused on improving the delivery of patient care.

For any organization, there are two sets of customers: external and internal. The external customer is the end user of products or services, whereas the internal customer receives a product or service from another unit within the company. For example, in a hospital, a physician's external customer is the patient and the internal customer is the nurse, whose performance is affected by whether or not orders are written legibly. Once the order is filled, the physician then becomes the nurse's customer. The orders to an internal customer directly affect the care given to an external customer or patient.

Benchmarking is a method of comparing internal process results to those of a recognized leader, and identifying areas for improvements. If an organization can determine its position with respect to its competition, it may be possible to improve internal processes to match or exceed competitors' standards. They can also gain a better feel for the metrics, which are important to monitor. This step provides the data that may become an impetus to change and new initiatives.

Continual process improvement requires data collection and analysis, and successful teams that understand when and how to use quality tools. The collection, analysis, and use of quality data become the backbone for developing a quality organization. (Walton, 1991, pp. 20-21)

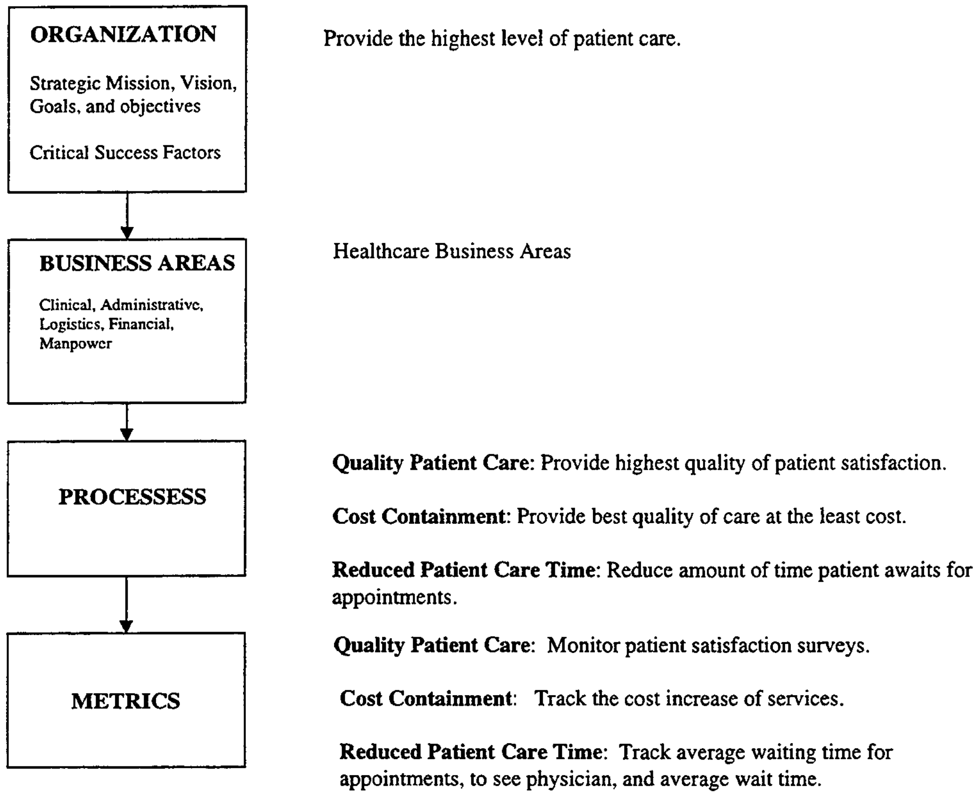


Figure 3-2 Organizations Goal, Business Areas, Processes, and Metrics

A well-designed performance measurement system will correspondingly support the efforts to improve the quality of healthcare. The metrics chosen to evaluate quality of patient care can help identify trends, patterns, and measure patient satisfaction. Performance metrics can also be used to measure how accurately physicians, nurses, and caregivers assess, diagnose, and evaluate patient care. They also provide management with tools for planning and leading mission accomplishments. Metrics that track improvements in both management and business areas enhance the linkage between strategic, tactical, and operational plans.

A specific example of a metric may be a nurse's care plan for bringing a patient back to health. A care plan consists of the five steps a nurse takes to execute physician orders to achieve desired outcomes in a cost effective fashion: Assessment, Diagnosis, Planning, Implementation, and Evaluation. Nurses take deliberate steps to maximize efficiency and attain long-term beneficial results. A simple metric might be a patient's evaluation of the quality of their care or the amount of care provided by the nurse. The patient could be required to complete a simple questionnaire using a scale from 1 to 5. This information could later be used to improve the nursing care plans and provide feedback to the nurses and management. Patient evaluations could be implemented through nonattributional and strictly confidential methods. The information gathered establishes a baseline metric value to compare to industry standards, and thus determine whether improvements are needed. If the nurse does not develop the proper care plan, feedback from the patient can be used to improve it. This example illustrates the importance of getting the main stakeholder, in this case, the patient involved in improving the quality of healthcare.

E. CONCEPTUAL VIEW OF QUALITY PERFORMANCE HIERARCHY MODEL

During an interview, a physician at the Naval Postgraduate School stated that physicians rely heavily upon visualization, instinct, past experiences, medical history, knowledge, and other symptoms to make a diagnosis of the condition of a patient. For example, a patient comes to the hospital complaining of abdominal pains. The first step for the physician would be to make an objective assessment using a visual inspection, taking the patient's vital signs to determine the cause of the pain, and then conducting a physical examination of the abdomen. If the physical examination reveals an enlarged gall bladder, for example, the physician usually orders a x-ray to aid in making the diagnosis. Physicians use every available resource from x-rays to visual inspections to support their diagnosis. Their decisions, evaluations and assumptions are based on an internal analytical assessment of their environment. Physicians are accustomed to negotiating face to face because they rely heavily on visual clues that tell them their patients thoughts and feelings. These business relationships are based on a wealth of visual clues exchanged between the patient and physician during an office visit.

In a like manner, Berelowitz introduces a metrics visualization view of performance that claims to present a more dynamic and instantly meaningful view of the most pertinent performance information about an organization. A hexagon is used with each surface representing an aspect of a healthcare organization (Figure 3-3).

Most large organizations have many facets that could be represented, but in reality, most important organizational strategies can be represented on a simple polygon. According to the Military Healthcare System, Figure 3-3 represents the most important performance information regarding the delivery of quality healthcare. One way of using this decision aid would simply be to rotate the hexagon, and on the basis of a simple color change, say green, yellow or red for example, determine if an organization was meeting, exceeding or failing to execute a strategic goals.

Extending the polygraphical representation to a healthcare organization, assume that behind each surface of the healthcare model there exists another series of polygraphs that graphically explains performance at a more detailed level. Now the polygraph becomes a multi-dimensional view with an immediate analog representation. Without knowing any specific information about patient care, a manager can immediately, through the visual color-coded representation, gain some understanding of all processes affecting patient care relating to the conceptual model. Of course, hidden behind all the boxes are data; data that one can view in a textual format, once the non-performing areas have been highlighted. Figure 3-4 displays the analog representation with the full visual effect. The representation of each measure is depicted in this model using a border scheme that can easily be observed in a graphical representation, allowing a manager to focus on areas needing attention.

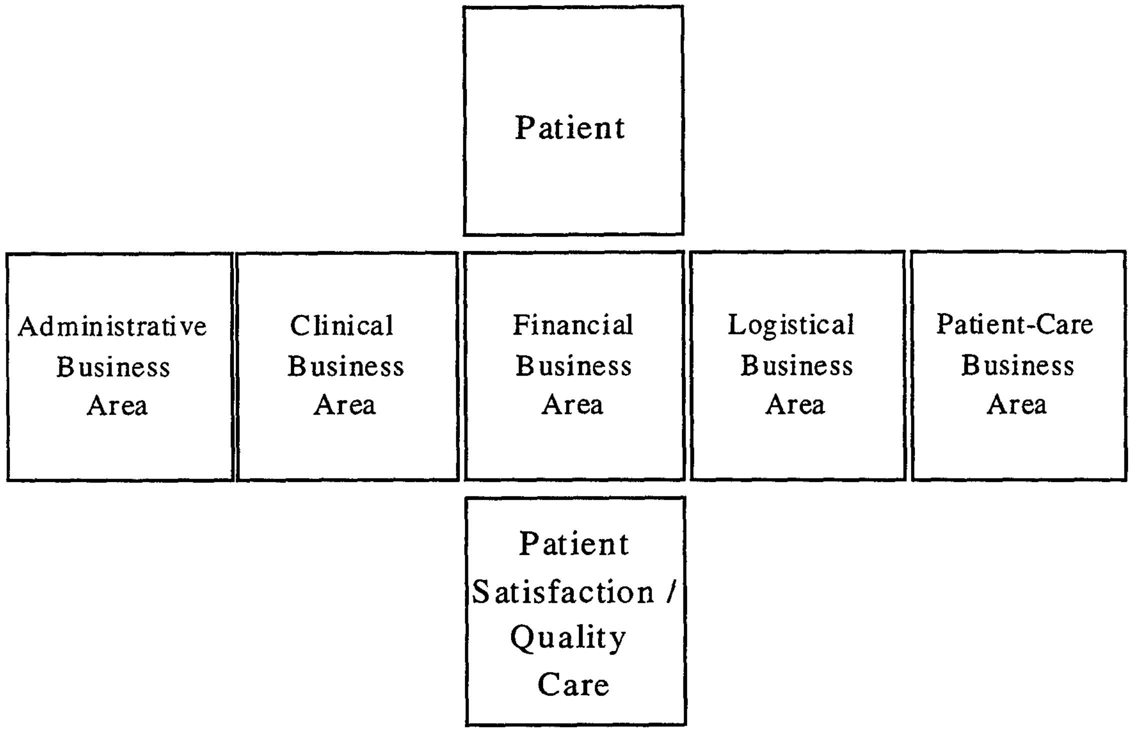


Figure 3-3 Healthcare Conceptual Model

The grey box with shadowed borders represent within standards. The double bordered white box represents a cautionary measure and the bold dotted white box represents below standards. The keys to business success are obviously drawn from many sources, but there are a few pieces of vital information that are crucial to measuring ongoing performance. By refining this information so that it is discernible and visual, managers can focus their attention on vital indicators. Each business area can use this approach to identify and manage the performance measurements that are important to determining their own efficiency and effectiveness. This kind of visualization and exploration tool coupled with the appropriate time-oriented healthcare data is essential for effective decision making.

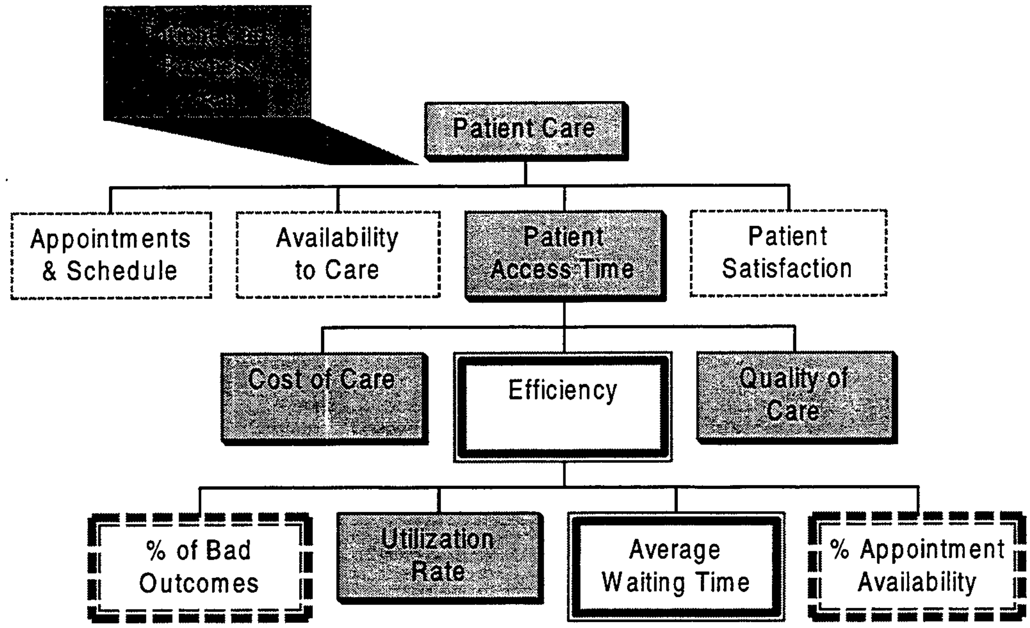


Figure 3-4 Visual Performance Model

The graphical presentation of information and performance measurement is important, and display capabilities should be available at aggregate levels of an organization. Managers in the organization may each view information differently at their respective levels. Figure 3-5 is an illustration of how data is pushed up the pyramid and information flows down. The visual performance model identifies problems areas, tracks quality and service, and depicts achievement of organizational objectives. Listed in Appendix A are some performance measures used by actual healthcare organizations to measure their performance.

Top level mangers focus their attentions on the strategic mission, vision, and goals of the organization, whereas middle managers concentrate most of their attention on specific business areas. Finally, line managers focus their attention on the day-to-day operations. It is very important when using a performance metrics approach that both managers and employees understand their roles and responsibilities within the organization and how their particular business area and others contribute to the overall strategic plan.

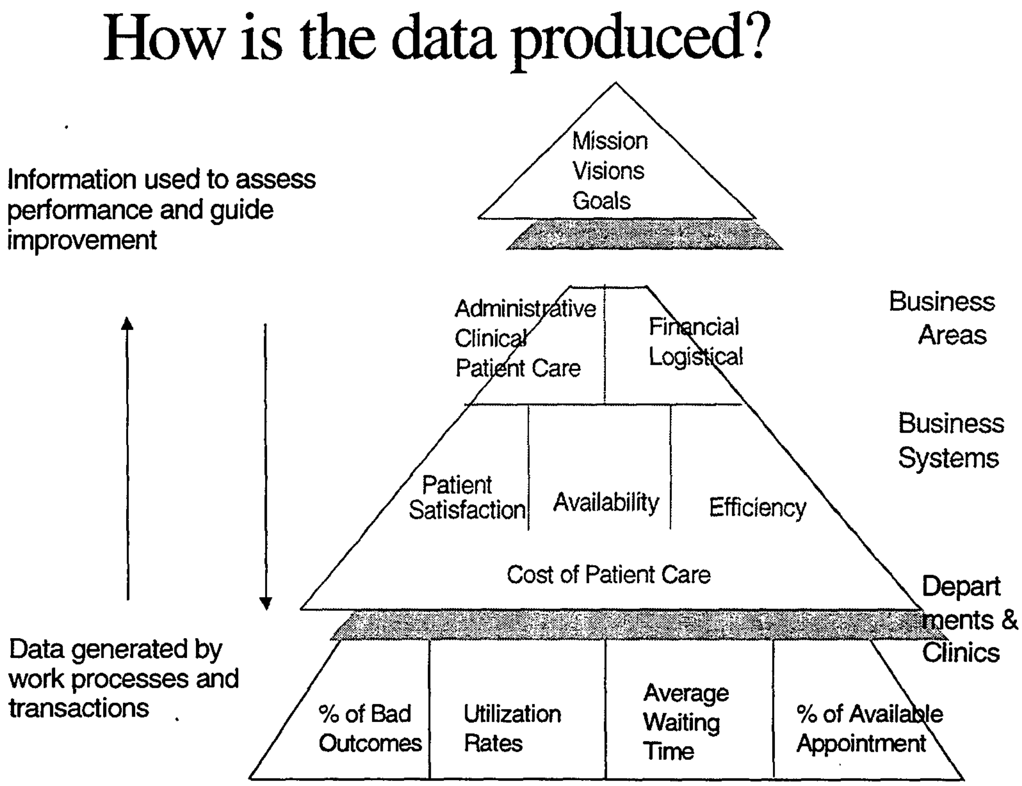


Figure 3-5 Healthcare Data Produced

F. SUMMARY

Performance measures are integral to identifying and managing a quality organization, and to designing a supporting data warehouse. They define and identify functions for each business area to measure and track. The alignment of performance measures with corporate goals and objectives results in information use that can lead to real improvements. Multiple performance measures are used to monitor resources or functions, which play a major role in supporting the operational, tactical, and strategical goals. Performance measures are considered by most to be the pulse of the organization. Many managers may only be concerned with performance measures relevant to their specific area; thus, upper-level strategic decisions must be made by leaders concerned with the bigger picture involving external environment effects, innovation, and appropriate corporate focus.

Within the DoD healthcare system, the functions that drive cost, quality, patient care service, and patient satisfaction are usually a good starting point for defining the basic performance measures. Analyzing such data requires that it be contained in an information system that will allow complex manipulation. Data warehousing technology is now available for storing large volumes of data to support management's decision making.

The framework and models discussed in this chapter provide the foundation one needs to understand how to design a data warehouse and establish quality improvements within an organization. The next chapter will discuss in detail the process of defining a data warehouse which using metrics to identify data requirements.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. PLANNING A HEALTHCARE DATA WAREHOUSE

A. INTRODUCTION

A basic fact about the Information Age is that we are surrounded by essentially limitless amounts of data and information. As a result, information managers must discover ways to collect, analyze, understand, and act on this information. It is reasonable to expect that organizations will continue to receive a super-abundant amount of information; however, information that is not used in the proper context does not realize its true value. For organizations to effectively utilize information, they must design and develop systems or products that will render information useful at strategical, tactical, and operational levels to support decision making and business goals. At the very least, organizations planning to utilize a data warehouse have recognized the need to view data across multiple business areas to get an integrated perspective of the entire enterprise. According to Berson and Smith, the latest technologies emerging in the client-server arena are focused on filtering unnecessary data and presenting valuable information in a user-friendly, intuitive, and easy to understand way. Among these technologies are data warehousing, metadata repositories, on-line analytical processing (OLAP) and data mining. Data warehousing has emerged as a technology by which management can migrate to a single comprehensive view of the state of the organization. (Berson & Smith, 1997, pp. 17)

Creation of a data warehouse represents an organizational commitment towards integrating, collecting and disseminating key performance information. A data warehouse is neither a decision support system (DSS) nor an executive information system (EIS); however, it is an integrated data platform, which may play an important role in the development or design of a DSS or EIS. A data warehouse can be used to improve the productivity of corporate decision-makers through consolidation, conversion, transformation, and the integration of operational data to create a consistent view of the organization. Data warehousing is a blend of technologies aimed at effective integration of operational databases for strategic use. According to W.H. Inmon, the founder of data warehousing, "A data warehouse is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management decisions" (1996, pp. 73). It is important to note that the data warehouse environment is not a product; it is an architectural construct that provides users with current and historical decision support information (Berson & Smith, 1997, pp.76). The information stored in a data warehouse will enable organizations to share, explore, and discover important business trends from current and historical data.

DoD healthcare, like many other organizations has previously designed operational database systems to meet mission-critical requirements and to support strategic decision making. Many systems have not been successful because managers and decision-makers have not understood how tactical and operational information differs from strategic information. Strategic managers prefer to have summary rather than operational information, and do not realize that a metrics driven approach is required to properly identify data needed to generate information to support decision makers at different levels. Hospitals and Healthcare Facilities thus have a high demand for systems that deliver quality information. Therefore, efficiency is no longer the only key to business success; flexibility and responsiveness have taken its place. A few years from now most organizations will be conducting business primarily over the Internet and those that harness the power of information will possess the competitive advantage over their competitors and partners. One of the keys will be the design and development of a central data repository that not only stores information about themselves but also about their competitors and the overall industry.

B. STRATEGY FOR DESIGNING THE DATA WAREHOUSE

The focus of this work centers on a healthcare performance management system that integrates the Malcolm Baldridge Framework, a strategic management component, and the identification of sound performance metrics to support the development and design of a data warehouse as described in Figure 4–1. The Baldridge Framework has been used by many organizations to establish their strategic plans and establish corresponding goals and objectives. It facilitates determination of measures of effectiveness needed to support the strategic plans. The Malcolm Baldridge framework uses seven categories to measure the efficiency, effective, and quality of an organization: Leadership, Strategic Planning, Customer Focus, Information and Analysis, Human Resources, Process Management, and Business Results. (Presidential Quality Award Program, 1998, pp. 11) The Leadership category examines senior leader's personal leadership and involvement in creating and sustaining values, organization directions, performance expectations, customer focus, and a leadership system that promotes performance excellence. Strategic Planning helps organizations sets strategic directions and defines key action plans. Customer Focus establishes a roadmap for identifying customer requirements, preferences, expectations, markets, and relationships. Information and Analysis measures how effective management is at using data and information to support key organizational processes and the organizations performance management system. Human Resources explores how organizations enable employees to develop and utilize their full potential aligned with overall mission objectives. Process Management examines how key processes are designed, effectively managed, and improved to achieve better performance. Finally, Business Results examines performance and improvement in customer satisfaction, overall financial program performance, human resources, supplier, partners and operational performance (Presidential Quality Award Program, 1996 , pp. 23-40).

Strategy management is a "commitment to undertake one set of actions rather than another" (Strickland and Thompson,1992, pp. 29). An organization's strategy consists of the moves and approaches devised by management to produce successful organizational performance. Strategies guide an organization through its business and determine how it will achieve its target objectives. Without a defined strategy, there is no established course to follow, no roadmap to manage by and no cohesive action plan to produce the intended results. A strategic management approach helps management craft and implement strategies to support core business functions. According to Strickland and Thompson, the process of strategy-making and strategy-implementing consist of five interrelated components (1992, pp. 30-44).

- Developing and forming a vision of where the organization needs to be heading (mission).
- Converting the mission into specific performance objectives.
- Crafting a strategy to achieve target performance.
- Implementing and executing a chosen strategy efficiently and effectively.
- Evaluating performance, reviewing the situation, and initiating corrective adjustments in mission, objectives, strategy, or implementation in light of actual experience, changing conditions, new ideas, and new opportunities.

INTEGRATING THE THREE HIERARCHIES

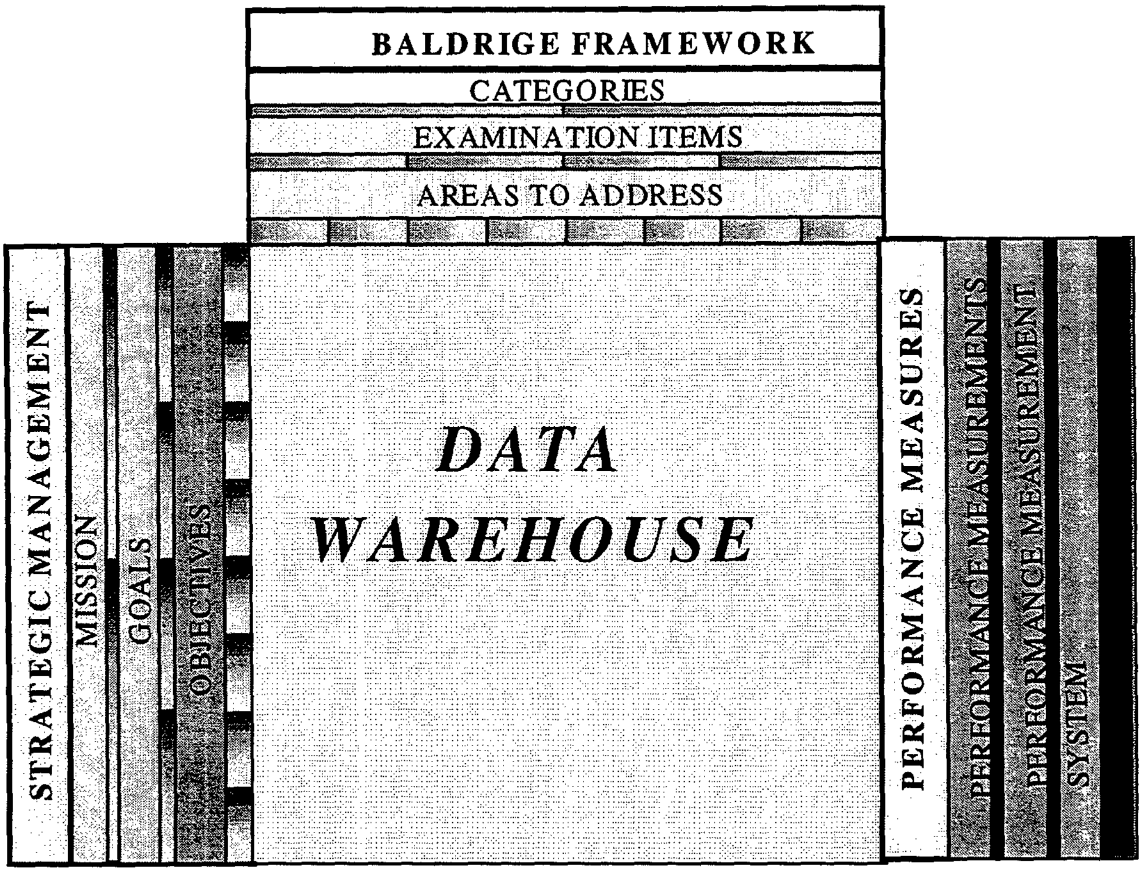


Figure 4-1 Integration of Strategic Goals, Baldridge Award, and Data Warehouse Environment

An organization's strategic action plan is a dynamic construct that undergoes continuous review, refinement, enhancement, and occasional major revisions. Today's managers must think strategically about their organization's position and the impact of changing conditions. They should monitor the internal and external situations closely enough to institute strategic changes when appropriate. The strategic plan coupled with a quality framework establishes a foundation to support the identification of proper performance measures.

As discussed in the previous chapter, a performance measurement system will support design and development of a data warehouse framework. Performance measures become the "glue" that supports decision-making of senior officials and helps them determine the strategic direction of the organization. An organization that lacks clear-cut direction because of vague or undemanding objectives, or one with a flawed strategy is an organization whose performance suffers. Having critical data in the hands of the right managers can help overcome these obstacles.

Figure 4-2 details the next piece in the model, the Data Warehouse. Many organizations are committing considerable human, technical, and financial resources to building and using data warehouses. Their primary purpose is to provide easy access to historical data that can be used with decision support applications, such as management reporting, queries, executive information systems, statistical analysis, OLAP, and data mining tools.

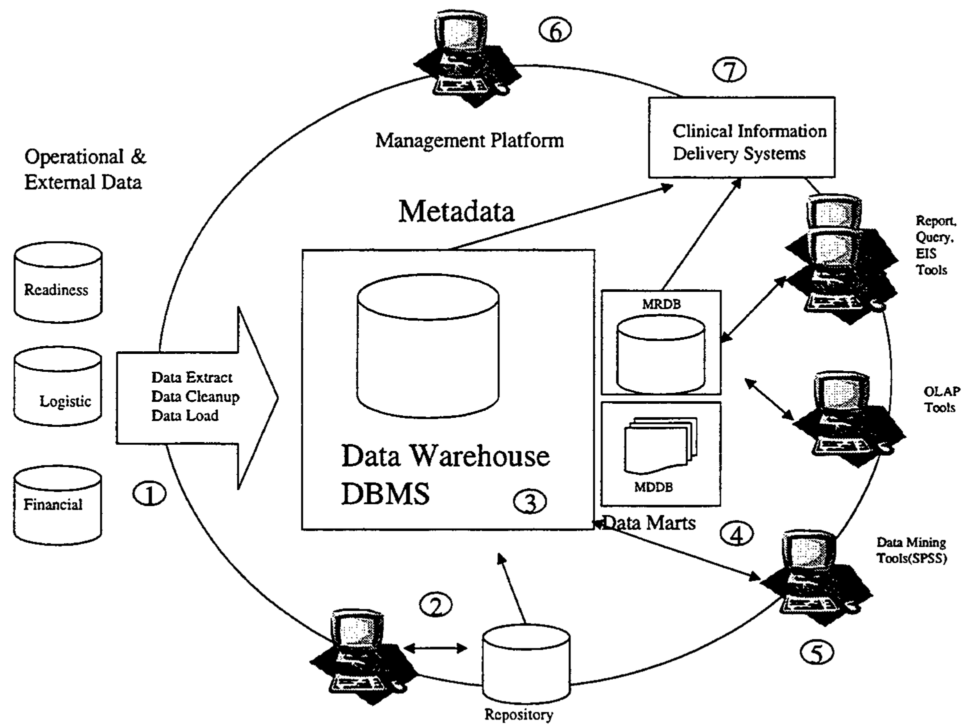


Figure 4-2 Healthcare Data Warehouse Environment After Ref. Berson and Smith

There are several functions involved in supporting a data warehouse environment. They include extracting data from legacy systems and other data sources; cleansing, scrubbing, and preparing data for decision support; maintaining data in appropriate data stores; accessing and analyzing data using a variety of end user tools; and mining data for significant relationships. These functions exist to help management and decision makers transform raw data into information that help identify key trends and predictable events, enhance management understanding of the macro view and enables focused reactions to those events, such as redefining and reengineering business processes. (Berson and Smith, 1997) A prerequisite to enabling management to accomplish these tasks that is accessible, current, flexible and available in a format usable by the requester.

Enterprises are constantly seeking competitive advantages. It has become a popular adage that information is the key to determining how to gain such an advantage. The difficulties lie in interpreting mountains of raw data that legacy transaction systems are collecting and processing. All organizations are literally in the midst of a volcanic eruption of data. Data warehousing is one alternative to coping with this overload that allows organizations to develop an architecture that can predict and adapt to the environment.

C. CHARACTERISTICS OF A DATA WAREHOUSE ARCHITECTURE

The data warehouse is a contemporary database construct that addresses the growing need of businesses for enterprise-wide data access. It provides decision-support processing at the strategic level by extracting data from operational database sources and making it available without affecting day-to-day operations. It is not software-specific and can be used in any computing environment. "There are several major issues related to the design and construction of the data warehouse, including granularity of data, partitioning data, metadata, lack of creditability of decision support systems (DSS) data, the system of record, migration, and more" (Inmon, 1996).

Many organizations today are large, complex entities with numerous systems developed over the last thirty years. They often possess many departments and a lack of standards. These attributes make system integration exceptionally difficult. Data Warehouses differ in significant ways from their more traditional databases. The following characteristics of most data warehousing installations are critical to the performance and success of any system (Sperley, 1999, pp.72).

1. Large Databases. Databases are large because data warehouses often store data on all transactions by all customers, or in healthcare case, all treatments for all patients. Further, the nature of analytical processes requires raw operational data to derive trends, averages, standard deviations, etc. Large databases possess

problems such as large amounts of time and CPU power to scan data in response to specific queries. The time and horsepower required to load and refresh the warehouse grows rapidly and it subsequently becomes difficult to maintain its currency and relevance to the organization. As aggregations, summaries, trends, and averages increase the amount of time required to refresh the existing tables with additional data can become prohibitive. Many organizations underestimate the database requirements of their systems. It is very important to conduct an analysis up front to ensure one's system can expand to meet unanticipated demands. This is one of the driving factors to using a metric-driven approach. Hospitals have a vast amount of data stored that do not improve or enhance their decision-making capabilities.

2. Cross Functional Analysis. Analysts will need to combine information from one view of data with information from other views. This will require the use of complex multi-way joins on the data tables. When combined with large databases where the tables themselves are large, the problem is compounded making these operations tedious on computer resources.

3. Ad Hoc Queries. Unlike traditional reporting structures in operational systems where the database queries are defined prior to the system's construction, the data warehouse is developed to satisfy an unstructured query model. Many models can be built to determine the results of each potential response, each based on assumptions regarding behaviors patients, customers, suppliers, costs, distribution models, and so on. The data warehouse can be used to conduct many iterations and permutations of queries into the demographics of patients, the costs associated with providing patient care, the revenues and profits generated, identify critical pathways and trends in the healthcare environment, etc. The ability to execute Ad Hoc queries on a data warehouse system will add significant stress and require a large amount of CPU and I/O resources.

4. Exposure to Many People in the Organization. Today's management philosophies are much different than they were a decade ago when Executive Information Systems first surfaced. In fact, the very name denoted the exclusivity of audience for these systems. It was assumed that only the top executives in an organization were involved in decisions requiring ad hoc analysis and summary information from a cross functional perspective. Today there are various levels of system and business analysts and mid level managers who need of DSS. This has become even more prevalent as organizations become flatter. Responsibility and decision making has been pushed down the organizational pyramid into the hands of analysts and mid level managers. Thus, demand for access to information has vastly increased. As data warehousing implementations grow and expand into the organization it is important that a DSS is set up to support these demands.

5. Increased User Sophistication. Users experience a learning curve in acclimating to any new technology. As they become more familiar with the data structures, query languages, and data analysis functions, they will develop increasingly sophisticated models for analyzing their particular areas. This will mean more complex joins, more full table scans, more sophisticated mathematical computations, and, in general, more demand on the system in terms of horsepower and I/O capacity. Therefore, hardware and software systems must have the capacity and capability to ensure performance.

6. Use of Multidimensional Databases. In an effort to improve response times for data warehousing systems, an option is to develop multidimensional databases or star schemas in relational databases. In these structures, the different parameters that describe data are given as dimensions of the database in time, number of patients, number of hospitals which provides a specific service, TRICARE regions, demographics of the patients, etc. It is very easy for multidimensional databases to become quite large, simply because of the multiplicative effect of carrying many dimensions. With 100 diagnosis being performed by 50 physicians, each of which is represented in five geographic regions tracking the number of diagnosis over the last three years by user demographics which may include five age groups, five income levels, and four levels of education, there are over 90 million facts. In fact, a recently published study indicates that for each additional dimension added, the raw data for each table increased by a factor of almost two to one (Thomsen, 1999).

This explosion in data can cause a data warehouse I/O subsystem a great deal of stress especially when dealing with multiple users. It may also create devastating problems in refreshing and uploading of data. In order to meet demands, systems must have exceedingly high performance characteristics and must be able to scale up to meet increase demands.

D. DATA WAREHOUSE ENVIRONMENT

According to Berson and Smith, the data warehouse development life cycle consists of a number of steps that should be performed to ensure that an optimized data warehouse is designed and the data in the data warehouse is accurate and trustworthy. Each step is depicted in Figure 4-2 and is further described in this section. (1996)

1. Data Transformation. This process involves the movement of data from operational source databases to the data warehouse. The data transformation process includes data access, validation, cleansing, consistency, and

conversion. Data access is the process of entering a database to store or retrieve data while data cleansing is the process of manipulating data extracted from operational systems. These processes are very important due to differences in data format compatibility between the operational systems. Data consistency is the result of using a repository to capture and manage data as it changes so that data warehouse can be continually updated. Finally, data conversion is the process of changing data from one physical environment to another, making any necessary alterations to move data from one electronic medium or operational system to another. The quality of the data warehouse contents is determined by how well these processes are executed.

2. Metadata Management. Metadata describe the types of information stored, how it is encoded, how it relates to the business, keeps track of the logical structure of data within the warehouse, the logical structure of the relevant operational source databases from which data are derived, and the transformation rules for data migration (Berson and Smith, 1996). Metadata synchronizes data elements from different operational systems and describes the characteristics of stored data. Therefore, changes to source databases must be captured and synchronized with the data warehouse.

3. Database Engine. A data warehouse requires additional complex and efficient indexing above and beyond what a normal Relational Database Management System (RDBMS) engine provides. A database engine is responsible for providing data integrity, enforcing and supporting concurrency, maintaining security, and supporting backup and recovery methods.

4. OLAP and Data Mining Tools. Originally introduced in 1994 in a paper by E. F. Codd, " OLAP is a decision support counterpart to On Line Transaction Processing" (Thomsen, 1999). OLAP allows non-programmers to derive information and business intelligence from a data warehouse. OLAP was developed so that non-programmers could run complex queries on data without learning complicated SQL commands. In particular, OLAP allows multidimensional views and analysis of data for decision support processes. OLAP tools run from a user's Web browser.

5. Information Delivery System. There are a large number of information delivery systems that transport processed data to the user. Each system provides analysts different views of data and allows them to use OLAP tools to construct different drill downs and rollup views of data stored in the data warehouse. Many users use the data warehouse to schedule the delivery of periodical reports much like the daily newspaper.

6. Data Warehouse Administration and Maintenance. Data administration and maintenance is significantly more complicated than that of a conventional database administrator since administrators not only have to keep abreast of the data warehouse database design, but also the changes that take place in the entire database environment. The data administration and maintenance processes ensure the integrity and currency of the data are maintained. The size of the database may also add an additional challenge in the areas of backup and recovery as data warehouses are frequently in the multi-gigabyte or terabyte range (Kroenke, 1988).

The data warehouse's ability to deliver reports from different business areas rapidly and in different presentation formats makes it a powerful tool. Although the overhead of building and maintaining a data warehouse is high, the capability to provide timely, near real time management information is superb. This long time goal of the information system (IS) community has previously been unattainable.

E. HEALTHCARE DATA WAREHOUSE MODEL

1. Data Transformation

Data transformation is the starting point of data analysis and is directly related to data quality. Acquiring quality raw data, combining and integrating data to make useful information, and then analyzing information and making high quality decisions supports good business practices. The crucial issue is knowing which data to use to create useful information.

For the Military Healthcare System (MHS), the data warehouse would be designed to draw data from eleven regional data marts that integrate the healthcare source databases. This data would be loaded from the five business areas: Clinical, Financial, Logistics, Administration and Human Resources and other systems. In addition, the data may contain a large number of anomalies and inconsistencies, including different abbreviations, formats for dates, field length, types, etc. The data loading process focuses on extracting data from operational databases, transforming it to comply with standards, and applying it to a completed data warehouse and MHS. Therefore, data cleansing and extraction tools will play a major role in transitioning data from the subject area data marts and loading it into the data warehouse database. Data cleansing, extraction, and loading are an ongoing process, which may be minimized as interface design become maturer. Data is extracted at regular intervals from existing systems and placed in the data warehouse. It is also summarized to allow management to look at trends although it is still available for detailed drill down data access and analysis. More importantly, if the data loading process is not successful, the data in the data warehouse will be deemed inaccurate and therefore valueless for decision support.

There is a tremendous challenge identifying all data sources, filtering out data, keeping it up-to-date, making comparisons, and creating achives and log files. Figure 4-2 represents an example of the possible data inconsistencies that strengthens the cases for establishing a systematic data cleansing and extraction process to support the data warehouse design. Despite all the efforts to eliminate data anomalies, there will still be situations or cases where data is missing. Missing data is the result of nondisclosure of private information or the method of collection used by different business areas. Policies and procedures must be established to handle the problem of missing data, since it can result in inaccurate and inconsistent data analysis and queries. For example, consider two laboratory technicians responsible for drawing blood from patients; one documents the hemoglobin blood count, and the other does not. This is a very simple example of how missing data can occur. Different clinical areas may have different standards and approaches for defining the length of a patient's stay in the hospital; and finally each of the eleven regions may have different ways of structuring their regional data marts. These kinds of inconsistencies require a well thought out and thorough data transformation process.

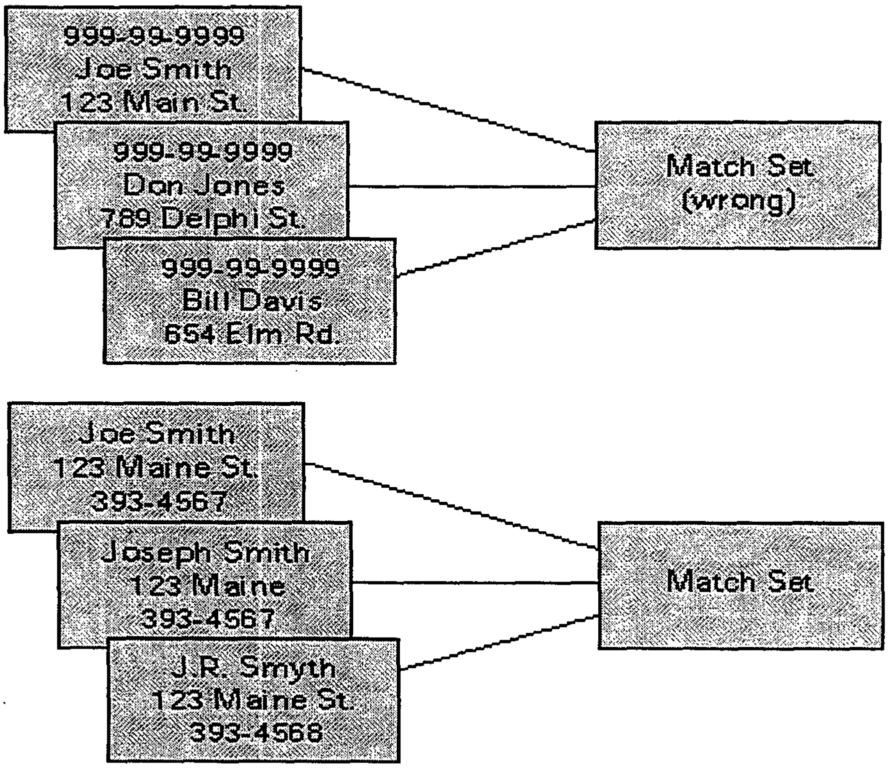


Figure 4-3 Example of Data Inconsistencies

2. Metadata Management

Metadata is data about data. Metadata management represents the backbone of the data warehouse environment. It documents data element, data type, attribute/property, range/domain, and process/method descriptions. The data repository environment potentially encompasses all corporate metadata resources: database catalogs, data dictionaries, and navigation services. Metadata is stored in a data dictionary and repository and insulates the data warehouse from changes in the schema of operational systems. Metadata must be time stamped and synchronized as it includes data element definitions, validation routines, algorithms for calculating derived or summary data, and more (Berson and Smith, 1996). Without maintaining metadata history, the business changes affecting data in the data warehouse would be lost, and the passage of time would steadily erode the value of the historical data. Metadata management is the process of consolidating, relating and synchronizing data elements with the same or similar meaning from different systems. Metadata synchronization joins these differing elements together in the data warehouse to allow for easier access. However, in many organizations, the same data element may be used by different business areas to represent different things.

Metadata is fast emerging as the single most troublesome issue in creating and managing data warehouses. It refers to information about the data captured and loaded into the warehouse. If users of a data warehouse want to know exactly how the data is stored, which source it was captured from, and how often it is captured, metadata will provide that information. There is no single factor which alienates users from data warehouses faster or more permanently than the lack of confidence instilled in users when they believe that data stored in the warehouse cannot be trusted, either because they cannot understand what it means, or because it is not properly managed.

Metadata management offers a consolidated view of the metadata from all warehouse components through an interface acceptable to both technical and business users in the performance of their jobs. One of the major issues today is that there is no single place for an organization to capture all of the knowledge it uncovers. While many IS professionals are familiar with the metadata about the legacy system and understand the warehouse, they are uncomfortable with the data transformation process. If done properly, a well managed set of metadata details can be re-used to generate new data marts, and, in some cases, queries against new or existing data marts or warehouses.

There are a large number of information systems used in healthcare to store data. CEIS, ADS, and CHCS store a significant amount of information about the management of each healthcare facility. If an analyst builds a query or uses OLAP tools to generate a report from the data warehouse monthly, weekly, or daily using data stored in CEIS, ADS, or CHCS, the meta data management process is responsible for ensuring the database schemas from each system remains in the same format so that the query processes successfully. Suppose a senior healthcare official is interesting is determining the total cost of inpatient and outpatient care for the hospital by DRG type. This would require data to be retrieved from ADS and CHCS. Therefore, the data elements used from each of these systems must remain synchronized to ensure valid data is returned. Meta data management has sole responsibility for referential and data integrity.

3. Database Engine

All data in a data warehouse or data mart is stored in a database. The primary role of a database engine is to provide data integrity, support concurrency, and backup and recovery. The DBMS is simply the physical framework within which data resides, and the choice of one framework over another is typically a result of tradeoffs between speed and flexibility of access to the data.

The most common database engine tools are relational data base management systems (RDBMS) such as Oracle, Informix, or Sybase; a multi-dimensional DBMS (MDBMS) such as Arbor/Essbase, or a proprietary RDBMS such as Red Brick; or, finally, some hybrid such as object-oriented database management systems. Objects are usually used by organizations that need real-time data because of their ability to access data quickly. Vendors such as Oracle are exploring the opportunity to utilize objects as a hybrid along with RDBMS. The differences between RDBMS and OODBMS are that programmers are more familiar with relational tables, the technology is easier to use, and there are many relational products on the market. DoD healthcare should utilize RDBMS tools for non real-time applications and use OODBMS to support real-time healthcare requirements such as telemedicine and multimedia.

The best plan is to select a database engine that supports the integration of both RDBMS and OODBMS. There are many other requirements to selecting a database engine. Each technology will create different challenges when designing a data warehousing architecture.

4. OLAP and Data Mining Tools

OLAP is used to help verify or falsify a hypothesis. OLAP is a category of business software that gives users access to analytical content such as time series, trend analysis views, and summary level information, as well as insight into data organized into multiple dimensions. According to Thomsen (1999), many vendors take three approaches to delivering OLAP applications. The first uses a multidimensional database server to store data. Data is stored as a cube that lets the user peel off pieces of it while keeping certain dimensions of the data constant. In the second approach, data from relational database engines is retrieved in a multidimensional fashion. Metadata defines where the data resides in the relational database. The OLAP server uses this metadata dynamically to generate the SQL statements necessary to retrieve the data as the user requests it. Users see a multidimensional view of data that is stored in relational tables. The last approach, less common and somewhat of a hybrid, uses a multidimensional database server as middleware to access data stored in a relational database. This process, commonly referred to as "reach-through," lets a multidimensional server provide users with the detailed transaction data that contributes to the summary totals stored in a multidimensional server (Thomsen, 1999).

OLAP software falls into two major categories: servers and clients. OLAP servers can be relational or multidimensional, but all provide a multidimensional view of the data. Examples of OLAP products that support multidimensional databases, multidimensional queries, and basic queries include BrioQuery, Business Objects, Cognos Powerplay, RedBrick, and Orbit. OLAP clients support heterogeneous data and allow end users to see data at different levels on the same display. OLAP provides analysts the tools needed to enable them to look at data, discover business relationships and develop new business metrics. With Web OLAP, updates to applications are a matter of updating a single site, simplifying ongoing maintenance as well. One of the key benefits to Web OLAP is that it makes it easier to broaden the audience for access to data. OLAP can also be used to inform decision-makers about future actions.

OLAP allows the user to view data using multiple dimensions, which may have multiple attributes. For example, consider that the Department Head from the OBGYN clinic who is interested in determining how much it cost to deliver a baby at hospital X? The answer to this question could be answered generally using a very simple query. However, on the other hand, suppose the he is interested in comparing the total cost of delivering a baby by hospital, region, and year. The cost to deliver a baby is now being used as a metric that is qualified by dimensions of facility, geography, and time. OLAP tools allow the user to a query using multi-dimensions that may have a number of different attributes. Region, state, or city in this case could break down the dimension geography. One important function to understand about data stored in multi-deminsion is that dimensions are not physically represented in the data warehouse. That is, attributes are always associated with dimensions because they are stored in the columns in the data warehouse database. Therefore, queries are developed using attributes such as region='Southeast', state='Georgia', or city='Atlanta'.

Data mining is used to help generate a hypothesis. It is a decision support process, which is used by users to search for patterns, trends, and relationships in data that may result in information. It allows users to look in the data warehouse without a predetermined idea or hypothesis in mind hoping to find interesting discoveries. OLAP and data mining tools are complementary. Data mining will be discussed in more detail in Chapter V.

5. Information Delivery System

Information Delivery Systems are no more than reports created using structured queries from OLAP tools and developed by end users to execute on the data warehouse. These queries are usually pre-determined, processed on a daily, weekly, or monthly basis, and are needed by the users to support decisions or metrics.

For example, consider a healthcare administrator for hospital X who wants a report that provides him/her information pertaining to the cost of surgeries taken place in the operating room by region, state, and diagnosis code. A query could be constructed using OLAP tools to deliver this information to the desktop of the user daily, weekly, or monthly.

Reports that are traditionally needed daily are the number of patients seen in the emergency room, intensive care unit, cardiology clinic, X-Rays conducted and the hospital census. Monthly each department or business area has a shopping list of reports they need to manage their perspective areas. These reports range from the total amount of supplies on hand to the amount of the budget spent on salaries. The ability to request and schedule reports to meet each business managers needs is vital to the management of any organization. These reports are usually instrumental in helping forecast, predict, and determine resource allocations and requirements to support day-to-day operations. Information is critical to success of a healthcare organization. Inability to properly plan for emergency situations in a healthcare facility could result in the loss of a life. For example, a patient enters the emergency room suffering from asthma and the physician needs to administer an inhaler. Later, he discovers that there are no inhalers in stock; this could result in the death of the patient or extreme pain and suffering until an inhaler is procured. Reports could be generated to establish special levels on supplies and other medical logistics to automatically create requisitions to procure inhalers basis on usage consumption and supplier availability. The number of reports or queries requested by a user are unlimited; however, each requires a significant amount of resources.

6. Data Administration and Maintenance

Data administration and maintenance are critical concerns when supporting a data warehouse environment. Data administrators manage the quality of corporate data and their chief responsibilities are identifying, administering, and establishing data management standards, policies and procedures, and coordinating the approval of data models. There are two critical challenges for database administrators; the first is preparation for potentially disastrous events that may affect the database system. These are events that could cause the database system to be inoperative for some period. The second is to anticipate and adequately prepare for requests that require changes in either the database system or the database system procedures. (Kroenke, 1988)

Database administrators must establish guidelines for managing corporate data resources. The database administrator reviews data models created during business analysis and system development ensuring models conform to data management standards, policies and procedures. Administration of a data warehouse has more complex responsibility that of an operational database that only contains information related to a specific business area. The database administrator not only has the task of design, development, administration, and documentation of the database structure and functional dependence of data items; but they also have to ensure that data retrieved from other information systems remain in the same format to support data transfer and extraction processes. For example, suppose the database administrator of the clinical business area data system elects to change the field type or size of the field "length of stay". If this data field is used in the data warehouse to determine the average length of stay for patients in the hospital, this data field must be updated in the data warehouse database schema. This may sound trivial but something this small could conceivably ripple into a lot of work for a data warehouse administrator. The worst case scenario for this example is that the erroneous data field is never detected and the information that analysts provide to senior healthcare officials is inaccurate. For this reason the data warehouse database administrator must be aware of the environment surrounding the source databases as well as the data warehouse itself. Again, the maintenance and administration of a data warehouse are very cumbersome and it is a demanding task that entails not only data administrator skills but also personnel and project management skills.

G. BENEFITS OF A DATA WAREHOUSE

Healthcare organizations must have management processes in place for monitoring and controlling the organization while, at the same time, decentralizing decision making in order to react to competitive changes and take advantage of unexpected opportunities. One central element that supports this balance between control and flexibility is shared knowledge. Such knowledge, derived from both internal and external data sources, is converted to information that can be readily interpreted. If knowledge and intellectual capital are becoming the key drivers of competitive advantage, then the intelligent organization is the one that can modularize, standardize and broadly share its knowledge both internally and, in many cases, externally. These are the healthcare institutions that will transform their knowledge bases in the context of their changing business strategy. Information systems play a role in creating and distributing that knowledge. Specifically, the data warehouse, a central repository of subject-oriented data originating from the organization's transactions systems and external data sources, becomes a critical information system. The successful implementation of a data warehouse can have a significant effect in fostering a culture of knowledge sharing. A healthcare data warehouses, utilized correctly, can provide insights into patients, customers, physicians, nurses, and healthcare providers behaviors and performance and profitability and cost structures of the organization. Also, in order to thrive or even survive virtually every healthcare organization must hone its ability to measure and monitor the quality and cost of care with accurate, timely, integrated information. A well-designed data warehouse delivers this business value.

Consider a healthcare example. In the past, DOD healthcare was primarily provided by Military Treatment Facilities (MTF's) whereas no cost was charged to the beneficiary. Now with the closing of military installation and dowsizing, DOD healthcare has had to negotiate contracts with local healthcare providers to provide healthcare to beneficiaries not located near military healthcare facilities. This has forced DOD healthcare to pursue cost effective avenues to provide healthcare. The government would like to provide better or at least compatible healthcare as previous. With healthcare reforms, mergers, acquisitions, and insurance premiums escalating, the costs relative to potential competitors has skyrocketed. Today's demand for healthcare information is driven by capitation, managed care economics, increased sophistication of purchasers, and a new emphasis on consumerism. Hospitals, clinics, and healthcare institutions must leverage data stored in the data warehouse by augmenting existing internal cost data with competitors' cost data and developing revenue projection scenarios. A "best practice"' cost analysis and profit projection analysis by diagnoses type, illness category, and beneficiaries age. Healthcare institutions that have not been accountable for profitability in the past, now have the knowledge to make potential cost cutting trade-offs depending on potential pricing changes. Further, they are in a position to be proactive about advocating capital investment decisions, which will make their healthcare contracts more competitive than the competition. Significant knowledge in the form of competitive data and sample analyses which directly support of the DOD healthcare strategic direction can be shared broadly throughout the organization. The effect is to raise the bar of knowledge and prepare the organization to react quickly. Significant value can be added by the implementation of a healthcare data warehouse.

A healthcare data warehouse could provides the following benefits:

- support decisions based on facts and knowledge not guesses,
- supports modeling of managed care contracts and services,
- supports retrospective and predictive analysis,
- provides fast, flexible and insightful real-time information,
- delivers reports to different business areas,
- allows an organization to manage information as an asset,
- enhances decision making,
- positions the organization to react quickly to potential business opportunities,
- supports the discovery of business and knowledge intelligence.

These are the primary characteristics or benefits that distinguish successful organizations in this information age. The most successful data warehouses are often developed in industries undergoing significant change. The DOD healthcare industry is a prime candidate due to cost increases in managed care arena and the closing of military installations.

A data warehouse can also help healthcare organizations locate the right information, standardize information to support presentations, reports, and graphs for decision making. A Data Warehouse can address problems classified as retrospective and predictive analysis. Respective analysis focuses on issues of past and present events while analysis focuses on forecasting certain events or behaviors based on historical information. Many organizations are have been faced with the inability to discover valuable and often previously unknown information that remains hidden in stored data. This prevents organizations from transforming data into knowledge and wisdom. (Berson and Smith, 1996, pp.11)

Many organizations have spent hundreds of thousands of dollars on consultants to study their business practices in an attempt to get various and diverse business areas to agree on what are the most important information requirements. These studies are rarely successful because organizations seldom reach an agreement between different departments on their information needs. The data warehousing architecture coupled with the three-tiered hierarchy, if properly designed can be used to identify information needs for the entire organization.

H. SUMMARY

As the data warehouse market develops, many different philosophies and methodologies evolve, some requiring extensive resources and some with quick fix solutions. Moving forward, it is apparent that no single architecture is globally applicable. To move the data warehouse technology into the next stage of its lifecycle, organizations must examine their information needs, data source structures and its business goals. Some organizations will choose to design an enterprise data warehouse, and some will select smaller data mart solutions. Many will conclude that multitiered architecture will bring the best results and the most flexibility for the future. Once requirements are gathered and a strategy is outlined, the organization should build an information infrastructure data map using a metrics approach. Although data mapping is time consuming, if the process starts with one business area at a time, adding mapping and planning as more areas are added, the payoff will be many times the effort. The easiest, quickest solution may be to build from the bottom up, creating data marts to meet specific business area needs, but with the forethought to remain in line with the corporate data warehouse strategy in mind supporting data standardization. This will allow for early, quick successes to provide departmental information to end users without derailing the enterprise-wide information effort. The data warehouse multitiered architecture has the flexibility and scalability to support both short-term and long-term objectives. With the implementation of a data warehouse, an organization can deliver better information to its end users and enable informed decisions based on accurate, timely, and precise corporate data. The next chapter will demonstrate the use of a standard analytical tool on a sample of data stored in a healthcare data warehouse.

THIS PAGE INTENTIONALLY LEFT BLANK

V. APPLICATIONS OF A DATA WAREHOUSE

A. INTRODUCTION

The focus of this chapter is on the use of statistical and analytical tools to support decision-making in DoD Healthcare. Raw data has no meaning if it cannot be transformed into useable information to support the decision making process. Once a data warehouse is constructed, the next step is to select user friendly and powerful statistical analytical and data mining tool(s) to gain knowledge, intelligence, and insight. Healthcare professionals are very interested in information that will improve quality and reduce cost of patient care. Two types of applications using a data warehouse will be discussed in this chapter: a performance measurement decision support system using a geographical information system, and statistical analysis using SPSS to demonstrate how a healthcare analyst might apply it to a data warehouse to verify a hypothesized model. The performance metric decision support system (PMDSS) defines components decision-makers can use to explore, identify, and analyze healthcare data. The PMDSS's main functions are to create an integrated model that analysts can use to access data, identify trends patterns and relationships which may result in relevant information to support good sound decision-making in healthcare organizations.

Healthcare professionals typically collect and analyze immense quantities of data. They collect data on patient records, drug trial results, outbreaks of specific illnesses, and ways in which they might be related to other factors. Many organizations are overwhelmed by data, with the attendant risk of becoming simultaneously data rich and information poor. Statistical analysis and data mining tools in conjunction with the data warehousing technology can help convert masses of data into information which managers need to make more effective and efficient decisions. SPSS, OLAP, Data mining and other statistical analysis tools can be used to identify trends, patterns, and indicators using the data stored in the data warehouse.

The components of the PMDSS allow organizations to use information as an asset and weapon. The PMDSS is also designed to help gain, knowledge, insight, and intelligence about operational data to counteract competitors. Once healthcare performance metrics have been identified, data warehouses will be used to store, update, and retrieve information to support healthcare managers in the decision making process.

The data warehouse is the central component that consolidates healthcare related information on a single platform.

B. PERFORMANCE MEASURE DECISION SUPPORT SYSTEM (PMDSS)

Each component in Figure 5-1 has a specific function that works together to comprise this system. The decision metric component is vital to identifying the raw data required to create the data warehouse schema. Functional business area managers can use decision metrics to support the organization's strategic plan, mission, vision, goals and objectives. The foundation for decision metrics will originate from the performance measurement initiatives defined in Chapter III. If an organization elects to develop a data warehouse, they will need to establish a process to identify the data needed to support business goals and objectives. Otherwise, they are prone to selecting all available data that will result in the storage and processing of large amounts of unnecessary data. Large data warehouses require a substantial amount of resources that escalate equipment, application and system software costs. Therefore, decision metrics play a major role supporting the development and design of an efficient performance metric driven data warehouse.

The data warehouse architecture relies heavily on the decision metrics process to identify data relevant to the organization. Appendix A provides examples of healthcare metrics. Data warehousing technology allows organizations to stay competitive and make more efficient and effective business decisions by conducting analyses on data both internal and external to the organization. Data in a data warehouse is usually stored in a time series format. This allows analysts, end users, and managers to retrieve certain data based on specific time periods to conduct analysis.

A Geographic Information System (GIS) is an information system designed to work with data referenced by spatial or geographic coordinates. It is used to capture, store, check, integrate, manipulate, analyze and display data that is spatially referenced to the earth. A GIS is a database system with specific capabilities for spatially-referenced data, as well as a set of operations for working [analysis] with the data (Star and Estes, 1990). GIS provides a range of image and data processing services to organizations that wish to visualize their data, and communicate ideas, using a geo-spatial context. GIS's are a subset of data visualization with the key distinction being that the graphics displayed are maps. In the context of a PMDSS, the GIS serves as the primary user interface whereby users can see selected metrics displayed on a map, and then drill down on specific sites which are of interest.

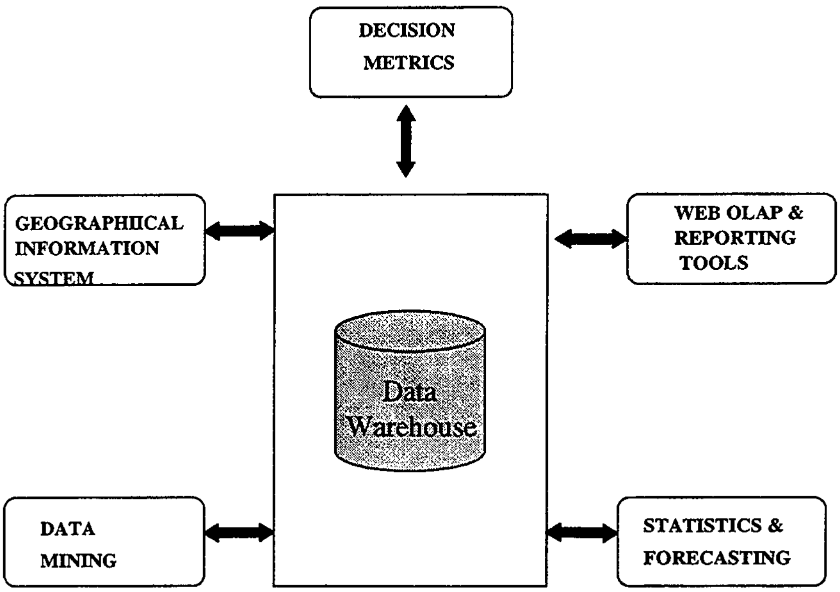


Figure 5-1 Performance Measure Decision Support System

World-Wide Web and On-line Analytical Processing (OLAP) tools are also important to the PMDSS system. OLAP uses multidimensional views of aggregate data to provide quick access to strategic information for further analysis. OLAP transforms raw data so that it reflects the real dimensionality of the organization as understood by the user. OLAP tools were designed so that non-programmers can run very complex queries without having to learn a complex query language. OLAP provides executives, analysts, and managers with valuable information via a "slice, dice and rotate" method of end user data access, augmenting or replacing the more complicated relational query. This slice and dice method gives the user consistently fast access to a wide variety of views of data organized by key selection criteria that match the real dimensions of the organization. OLAP performs multidimensional analysis of enterprise data including complex calculations, trend analysis and modeling. It enables end-users to perform ad hoc analysis of data in multiple dimensions, thereby giving them the insight and understanding they need for better decision making (Thomsen, 1999). A typical OLAP calculation is more complex than simply summing data, for example: What would be the effect on patient care costs to healthcare facilities if the medical supply prices went up by 10% percent and healthcare providers costs went up by 5%?

The World Wide Web is a pipeline that enables users to retrieve, access, and analyze data utilizing GIS, data mining, OLAP, and statistical analysis tools, all running from the user's Web browser. Each component of the PMDSS can be deployed from a Web browser, facilitating the ease of access. The PMDSS is a thin client application platform which provides easy access to each decision support tool.

Generally, statistics provide summary information used to determine the degree of relationship between two variables or whether the two are significantly different. Forecasting, on the other hand, is a subset of statistics that allows users and analysts to make predictions about the future based on patterns that occurred in the past. Statistics and forecasting can be used to create strategies for exploring data, drawing conclusions from the data, and graphically present results in a manner that managers, executives, analysts and users can use to make more effective and efficient decisions. Prior to the recent development of new tools and technologies such as OLAP, data mining, data warehousing and GIS's, statistics and forecasting were the primary tools used by managers to support decision making. The foundation of each of these tools is formed using statistics and forecasting methods.

Data mining is used to identify the variables needing research. "Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques." (Gartner Group, 1999) The amount of data and information that MIS professionals, analysts, and users are asked to process and analyze has grown astronomically in recent years. Data mining products can be invaluable tools, which provide efficient and effective ways to focus attention on data that is most relevant.

C. HOW PMDSS WORKS

The field of decision support is changing rapidly. From primarily a personal support tool, DSS' s are becoming a shared commodity across the organization. With the advent of intranets and internets, decision makers can utilize high performance analytical tools around the world. Traditionally, DSS's focuses on providing information interactively to support specific types of decisions by individual managers, analysts, and end users. They were also used to support the roles and functions of management at different levels of the organization that is, managers at each level and business area of the organization have different data requirements and information needs. For example, decisions made at the operational level tend to be more structured, those at the tactical level more semistructured, and those at the strategic level more unstructured. Therefore, each DSS must be designed to produce a variety of information products to meet the changing decision needs of managers at different levels and business areas of an organization. DSS's allow managers to make better decisions because they have more accurate information at their fingertips. A DSS uses hardware, software, data, models, and people resources to provide interactive decision support for managers.

As depicted in Figure 5-1 a PMDSS consists of several different components. The decision metrics module has the functions of helping managers, analysts, and end users filter and identify relevant data needed to develop the data warehouse. The data warehouse stores data in the proper format, create links to legacy databases and applications, ensures records are updated, maintains database schema, and maintains links and pointers to other databases. Once the data warehouse is functioning properly, a map user interface such as GIS can overlay it to provide analysts or end users a graphical display or depiction of the data. Geographical displays allow analysts or end users to visually detect areas that are within or outside of standards. It also allows users to compare information from different geographical areas of the United States using a mapping agent.

A Geographic Information System (GIS) is a special category of a DSS that integrates computer graphics and geographic databases with other DSS features. A GIS can be used to construct and display maps and other graphic displays that support decisions affecting the geographic distribution of people and other resources in an environment. A GIS geocodes rows of data to specific points on a map. GIS's also provide thematic mapping features that allow for comparison of data across different geographical areas such as a nation, region, or hospital using color-coding or symbols (red, yellow, or green) to compare and determine if information at other areas are within standards or not. Not only can a GIS provide analysts a meaningful view of information, it can help them identify areas with substandard performance measures, and allows analysts, end users, or managers to zoom in on data to make comparisons by region, district, state, catchment area, hospitals, etc. In addition, the drilldown feature allows analysts or end users the ability to detect problems and search for events that may have triggered them.

A GIS can use information from many different sources and in many different forms. The primary requirement for the source data is that the locations for the variables are known. The location of variables are usually annotated by x, y, and z coordinates which represent longitude, latitude, and elevation, or by such systems as ZIP codes or highway mile markers. Any variable that can be located spatially can be fed into a GIS. GIS technology is an expansion of cartographic science and can be used to enhance the efficiency and analytic power of traditional mapping. It can also be useful in helping the healthcare community understand the process of healthcare delivery.

To demonstrate the benefits of a PMDSS, take for example, an analyst who is interested in discovering how many patients have contracted pneumonia after ventilation is applied. Using Figure 5-2, the first step would be to use the GIS to drill down using metric number 3, pneumonia after ventilation to determine the numerical statistic for each hospital, region, or state using a report, chart, or table. Once the data has been retrieved from the warehouse, an analyst can compare the numbers to other hospitals in the area, region, or state using the U.S. Map feature in GIS. This information can be very vital in the problem identification stages of an analysis. In this example, the GIS system uses color-codes such as (red, yellow, or green) to determine if certain statistics fall within or below industry standards.

Figure 5-2, for example, shows that the central region (cross-hatched in this example) falls below the acceptable threshold for the "Pneumonia after Ventilation" metric. Thus, the user might choose to drill down in this region (Figure 5-3) to see where the source of this disparity lies. At this lower level, specific medical facilities are shown along with their specific performance vis-à-vis the metric. The user can quickly identify under performing sites and then click on specific sites for even more detailed data. This process continues until the user is satisfied that she/he has an understanding of what underlying causes(s) of the under performance is (are). A similar process can be undertaken by trending the metric over a specified number of periods (six months, say), and seeing where potential trouble spots may occur in the future.

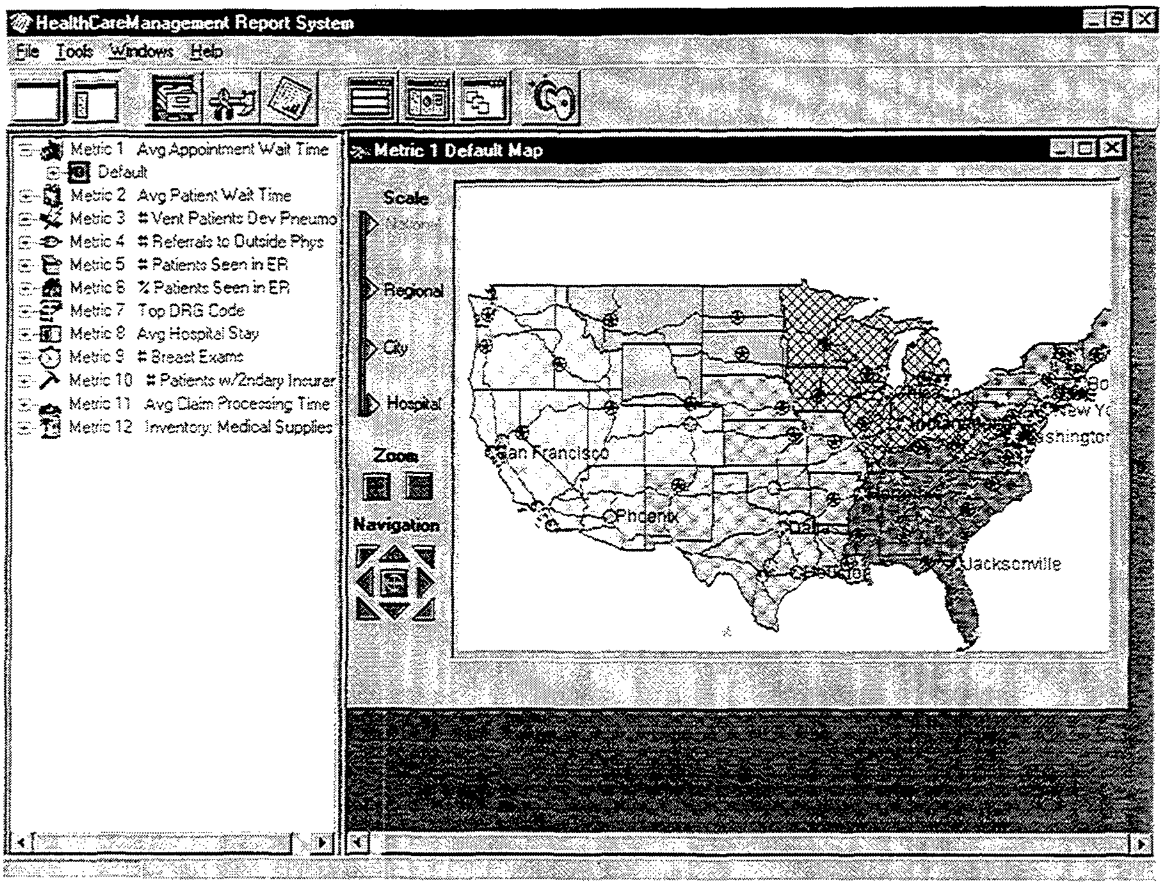


Figure 5-2 Top Level View of GIS

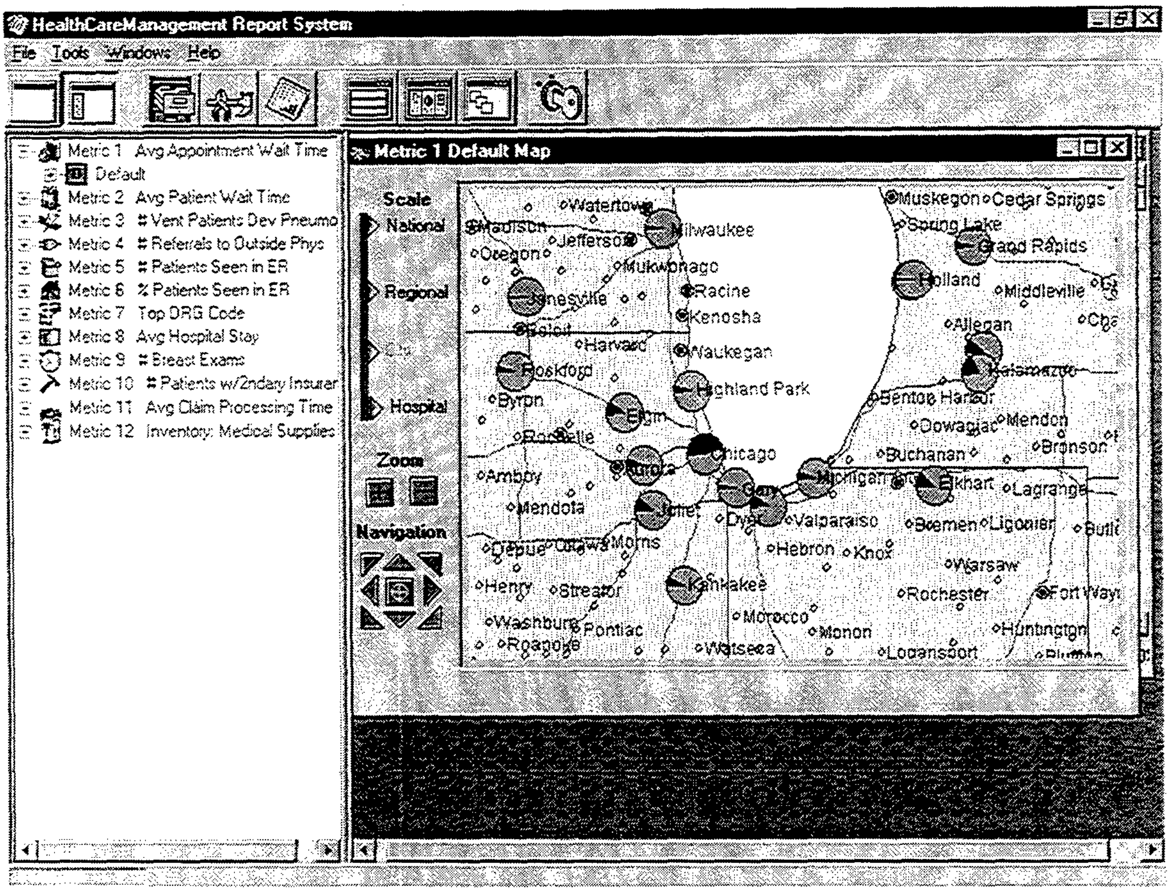


Figure 5-3 Drilldown View of GIS

Once a problem is identified, it is likely that a more detailed analysis will be appropriate to conduct problem diagnoses and analyses to determine a more definite cause for ventilated patients getting pneumonia. This is where data mining and OLAP tools may be more useful. GIS technology allows users to drill down further but does not provide the capability to conduct a detailed analysis of the problem.

The World-Wide Web and On-line Analytical Processing (OLAP) tools allow analysts, managers, and end users the ability to view data at different levels and eliminate access barriers to information. Analyst can use OLAP tools to look at data, discover business relationships and develop new business metrics. The combination of the World-Wide Web and OLAP tools simplifies updating, maintenance, and access to information. The World-Wide Web can also be used to publish reports on the Internet for people to print out or view as they wish and make it easier for others to gain access to data. Data mining allows analysts the ability to explore, analyze, identify patterns, and define relationships between and among variables to create a hypothesis.

Finally, statistics and forecasting allows a manager, an analyst, or end user to explore, analyze, graph, and draw inferences about data and predict certain future events. The second half of this chapter will discuss the use of data mining and OLAP on the data stored in the data warehouse.

D. BENEFITS OF THE PMDSS

A PMDSS allow analysts, end users, and managers to identify metrics, gain access to enterprise-wide data, and drill down using tools to assist in data analysis and exploration. Without the proper decision metrics, organizations are apt to store all available data in the data warehouse, which could lead to information overloading. Organizations can use the performance metrics systems strategy defined in Chapter III to identify the important decision metrics, which will lead to appropriate data elements to be stored in the data warehouse. A PMDSS stores multiple years of tactical performance information that answers "who?" or "what?" questions about past events, for example, how to gain new revenue or reduce cost of patient care, provides immediate access to information, integrates data across and even outside of the organization, and provides users the freedom to create reports from one central location (Inmon, 1996).

The GIS component of the PMDSS is a visual display that highlights the problem identification process. A GIS can also be used to compare data from hospital to hospital, region to region, or state to state. It uses a very simple user-friendly color-coding scheme (red, yellow, or green) which can quickly and intuitively identify problem areas.

Data mining and statistical analysis tools can be used to further analyze the problems identified from the PMDSS and to predict when problems may arise in the future. Data mining uses pattern-recognition technologies, ratios, and influences in data sets to help healthcare organizations understand their patient base better and make decisions that are ultimately more effective. It can uncover hypotheses which statistics and analysis tools can only subsequently verify or falsify. Data mining can also be instrumental in supporting data-driven decisions based on the historical business patterns that organizations accumulate daily as they interact with their customers. For example, data mining can tell one about the factors influencing the number of ventilated patients contracting pneumonia in a hospital, region or state. This information could lead to saving lives and eliminating pain from a very severe illness.

Finally, statistics and forecasting are used to identify and utilize information hidden in data using query generators and data interpretation systems, such as SPSS or SAS, the traditional tools used by analysts. Traditionally, analysts have used statistical and forecasting tools to formulate theories and hypotheses. This has been considered the manual, user-driven, top-down approach to data analysis. In contrast, data mining is a data-driven, self-organizing, bottom-up approach to data analysis, whereas statistics are user or verification driven. Data mining extends statistical approaches by automatically generating a portfolio of hypotheses which can subsequently be verified or falsified using conventional statistical techniques. Data mining has major advantages over statistics when the scale of databases increase in size, simply because manual approaches to data analysis are rendered impractical and do not scale very well as data sets increase in size and complexity. The next section will discuss in significantly more detail how to apply a statistical analysis application to a health care data warehouse.

E. WHAT ARE STATISTICS AND WHY USE THEM?

Statistics is a general method of reasoning from data. It is a basic approach shared by people in today's society to draw conclusions and make decisions in business and life. Statisticians and analysts use statistics to communicate effectively about a wide range of topics from patient care to quality service to operational efficiency. It is also used as a way to "reason effectively about data and chance in everyday life" (Stephenson, Rogness, Ritchie, and Stephenson, 1999). The goal of statistical analysis is to test hypotheses and gain insight through numbers. There are four important aspects of statistics: developing good data, strategies for exploring data, drawing conclusions from the data, and presenting results (Stephenson, Rogness, Ritchie, and Stephenson, 1999).

David Moore, in his book Statistics: Concepts and Controversies say: "Statistics are used to face the variability and uncertainty of the world directly. Statistical reasoning can produce data whose usefulness is not destroyed by variation and uncertainty. More important, statistical reasoning allows us to say just how uncertain our conclusions are." (Moore, 1997, pp. 17)

Because of the over abundance of data and inferences made from variables, conclusions drawn from data are uncertain. For example, as one measures patient satisfaction, attitudes, and characteristics, the results could differ because people and things vary or sometimes the measurement system itself introduces variation into the measures.

A data warehouse allows analysts to use statistical tools to analyze a subset, sample, segment, or the entire database. Statistics have been used to help hospitals, healthcare institutions, and other healthcare organizations get an accurate picture of the quality of patient care provided by a physician or healthcare provider by randomly surveying a subset of patients rather than the entire population without bias. There is a tremendous cost saving to conducting a sample rather than surveying the entire patient population.

It also allows analysts to experiment by looking at explanatory and response variables to isolate and understand the impact of changes in one explanatory variable on a response variable. A physician might, for example, try altering the dosage of a drug on a patient to determine if there are severe changes in the patient health condition due to the amount of dosage of medication administered. The results may contribute to a reduction in amount of medication administered or just completely terminate it.

Statistics deals with cases, variables, and values that are individual objects being measured, specific properties being measured, and assigned number measuring the property an object (Moore, 1999, pp. 23). A physician's performance is usually determined by how effective he/she cures his patients, patient's complaints, and number of patients return visits; however, each variable has a different measure. All measurements vary and one goal of analysis is to understand if a measurement variation is random or biased in a specific direction. Understanding data is essential to effective data analysis, designing a data warehouse, using statistical analysis tools, and utilizing OLAP and data mining concepts. To draw meaningful conclusions from data, it must actually measure characteristic that analysts are interested in understanding.

Data exploration is usually the first step in data analysis; it is used to help identify overall patterns and extreme exceptions to patterns. Many analysts start the data exploration process by using descriptive analysis such as the mean, median, mode, standard deviation, variance, percentages, distributions, and counts. Histograms, dot plots, boxplots, line charts, and other graphs are usually used to examine and visually display descriptive analysis results to help identify patterns, trends, and relationships which may be hidden in the raw data. Some analysts explore data by looking at each variable separately to determine the basic counts, percentages, and proportions of measures at each level.

Data exploration is important for understanding the quality of data stored in a data warehouse or repository to begin looking for areas to query for information. It can also be vital in the investigative process that helps analysts identify missing data, data inefficiencies, and for highlighting any inherent problems with data. The data exploration process prevents analysts from spending time and effort on unsupported analyses and from drawing conclusions that are not supported by the data. The goal of data exploration is to start a successful path towards full data discovery before drawing conclusions from the data. Essentially, exploratory data analysis gives analysts a "feel" for the data and helps them uncover possible directions which the analysis may go (Moore, 1999, pp. 26).

Once an analyst has hypothesized a model with an interesting relationship, statistical tools can be used to refine data to understand strengths of relationships and factors that cause them to exist. Appendix B is used to present an illustrative analysis in order to make inferences about a healthcare issue. Tables, graphs, cross tabulations, and a linear regression model are used to describe associations, which may exist among the dependent and independent variables.

Statistics is not a pure science; therefore, it is very difficult to move from the association of variables to a causal explanation. Statisticians are very careful about concluding that associations are due to causation. In fact, once they have properly designed a model that demonstrates a causal link, they are still very careful to avoid suggesting a causal association exists, purely because statistics is based on the likelihood that an association or relationship is coincidental.

In sum, basic understanding of statistical techniques allow analyst and users to learn more about data, discover interesting relationships, better use scarce resources and give credibility to their ideas. Statistics are the method by which analysts describe relationships, draw conclusions, and state the level of uncertainty associated with conclusions.

F. STATISTICAL ANALYSIS TOOLS

The SPSS statistical tool is being integrated into the OLAP component of the CEIS architecture. Figure 5-4 below illustrates how the Corporate Executive Information business office, a business system component used by senior DoD healthcare officials, plans to integrate source data into a data warehouse. Also, listed are the data sources, interfaces, and tools used by analysts and users to examine and transform data into meaningful information to support healthcare decision making. The information stored in a data warehouse can either be exported to a data mart or manipulated using OLAP tools such as SPSS, Business objects, or data mining to produce canned reports or detailed information supporting business goals and objectives.

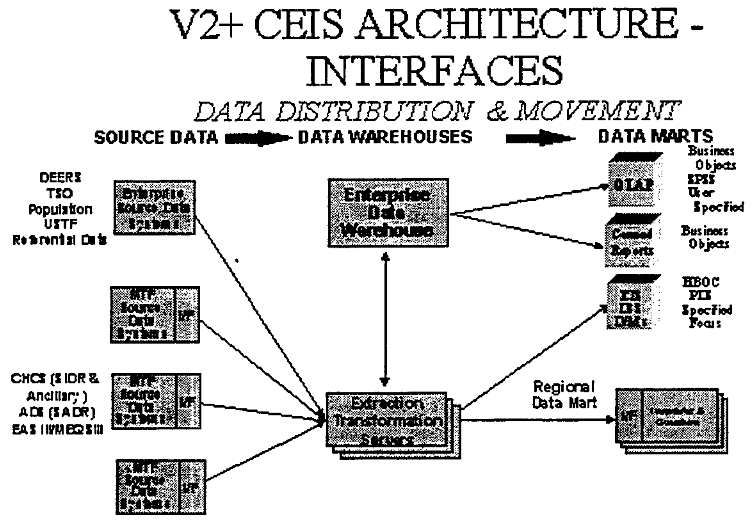


Figure 5-4 CEIS Architecture Interfaces

As seen in Figure 5-4, SPSS is one of the statistical analysis tool used by analysts to conduct analysis on data stored in the enterprise data warehouse or data marts in the Corporate Executive Information Systems (CEIS) architecture. During the early stages of CEIS production, SPSS was the only statistical analysis tool used and many interfaces were created to support it. Designers later determined that additional tools were also useful due to the size of the data warehouse, Business Objects. The illustrative analysis listed in Appendix B using SPSS is not intend to be a thorough analysis of this data set, but an illustrative demonstration of how SPSS can be used to analyze healthcare data. The author's intentions are merely to take this data and demonstrate some of the steps involved in transforming data into useful information by applying some of the features of SPSS using statistical techniques.

It is very important to collect data properly in order to apply different statistical techniques. If data is not properly collected and formatted, SPSS allows analysts to manipulate data types or recode variables in data sets or to other fields to support statistical procedures. It also helps users and analysts see the distribution of the data and identify outliers, calculate medians, etc. SPSS also allows end users to share the power of statistics with colleagues and customers by distributing multidimensional reports as interactive pivot tables and graphs via e-mail, Web, intranet or internet.

SPSS could be used to answer some of the following healthcare related questions listed below.

- Is there a relationship between the demographic characteristics of a given area and the drugs prescribed to its population?
- Are certain services in higher demand in specific areas and why?
- How will the closure of an important facility such as a day care center for the elderly or a General Practitioner's (GPs) surgery affect the local population?
- Do the services that are being provided match the needs of the local population?
- Where is the best place to locate a new facility in order to ensure the maximum number of patients live within a short walk or drive of it?
- What is the number of referrals from each local authority area?
- What is the total and average number of days spent in hospital?

The data set used in the illustration listed in Appendix B has been retrieved from a healthcare data warehouse. The data was extracted using a product called Business Objects, which exported data to an Excel Spreadsheet to be transferred to the SPSS Data Editor. This is just one method of transferring data into the SPSS Data Editor. The sole purpose of this demonstration is to familiarize one with some of necessary steps needed to prepare data for the analysis process. There are many steps involved in the data preparation stage. Some include recoding variables, identification of missing data, running frequency counts to gain insights about data, running crosstabs to determine if relationships exist, and conducting inferences about data. The illustration listed in

Appendix B is just a tip of the "iceberg" of the steps necessary to applied SPSS or any other statistical tool to a data set.

G. WHAT IS DATA MINING?

Data mining can be used to help identify dependent and independent variables needed to design models. Data mining also uncovers the hidden meaning and relationships in massive amounts of data stored in the data warehouse (Groth, 1998, pp. 40) In short, the value of a data warehouse lies in the information that can be derived from its data through the mining, analysis, and exploration process. Data mining can reveal relationships in data which analysts might not be sufficiently skilled to detect. Since its creation in the 18th century, statistics have provided the mathematical tools and analytical techniques for dealing with large amounts of data. Today, as we are confronted with increasingly large volumes of data, statistics are, more than ever, a critical component of data mining and refining toolkit that facilitates making effective business decisions.

Statistical analysis is the preferred weapon of many successful businesses today. It is an essential tool for mining and refining data into useful information. Healthcare organizations that effectively employ statistical analysis usually increase revenues, cut costs, improve operating efficiency, and improve patient satisfaction. Analysts also use these tools to more accurately identify problems, opportunities, and understand their causes so that they can quickly act to eliminate threats or capitalize on opportunities.

Research reveals that data mining is synonymous with knowledge discovery, the process of discovering non-obvious patterns in data that can help with marketing, forecasting, identifying trends, and other relative information about an organization and its environment. Consequently, successful data mining requires both business knowledge and analytical ability. Therefore, statistics and data mining are complementary. They allow analysts the ability to create models and makes inferences about data.

Data mining tools can be used to further analyze the cause of patients contracting pneumonia after the use of a ventilator. They can also be used to compare information based on historical data that hospitals or medical facilities accumulate daily as they interact with their patients. Data mining uses ratios and patterns to support one's intuitions about factors influencing and variations affecting the number of ventilate patients contracting pneumonia in a hospital. Data mining tools can be used to support the Joint Commission on Accreditation of Hospital Association (JCAHO) belief that etiology of nosocomial pneumonia in the incubated patient is related to oropharyngeal and gastric colonization with microorganisms, microaspiration, and impaired lung defense mechanisms. They believe the primary factor associated with ventilator associated pneumonia is the length of time (days) the patient is ventilated. Their belief is supported by a strong positive correlation between nosocomial infection rate and proper management of devices used in caring for patients. Data mining tools can be used to explore and analyze patterns of long extended time periods which patients have been ventilated and resulted in pneumonia. The results may reveal that pneumonia occurred after ventilation at admission, 48 hours later, or after discharge or removal. The results of such a study could result in the discovery or identification of new methods or techniques preventing the occurrence of pneumonia in ventilated patients.

Data mining and statistical analysis tools can lead analysts to conclusions supported by quantitative analyses. There objective is to look for interesting patterns and associations and dig down into them to draw conclusions or make predictions. Prior to the development of data mining and statistical analysis tools, analysts began by building simple models for predicting defaults and making them more sophisticated over time using a top down approach. Data mining and statistical analysis tools such as OLAP, SPSS, and Business Objects have made data analysis and data exploration easier.

H. SUMMARY

The value of data increases with the use of statistics, statistical analysis and data mining tools. These tools improve an analyst's ability to discover important facts about data. Statistical analysis, data mining, and GIS tools translate data into meaningful information and provide deeper insights into data. They are also integral parts of the decision support process. Analysts and users can turn insights into better decision-making and develop smarter, more effective policies. Therefore, statistical and analytical tools are essential for effective use of data stored in a data warehouse. Powerful, flexible, usable statistical tools are required in a data warehousing environment. There is no reason for saving and storing data beyond using it to make more informed business decisions. These tools allow analysts and users the ability to explore data, discover interesting hypotheses, and create models for further examination. They also allow analysts and users to test whether data is truly reflecting what they think is being measured. Statistical analysis and data mining tools fits comfortably with the approach of getting started with a data warehouse rather than spending years specifying the ever-changing data requirements before beginning to bringing data elements together.

Analysts cannot find the new exciting "nuggets" of information without some data exploration, asking questions, examining relationships, and so forth (Cabena, Hadjinian, Stadler, Verheees, Zanasi, 1998). To continuously improve the profitability and operations of an organization, the maxim "you get what you measure" always holds (Adriaans and Zantinge, 1996, pp. 27). If you don't put systems in place to measure and evaluate associations that are important to the success of a business, how will organizations know if they are succeeding? GIS, Data Mining, and statistical analysis tools provide flexibility to make data, exploration, and ad hoc analysis appear easy to the novice. They provide features required to repeat an analysis yearly, quarterly, monthly, daily, or hourly and to compare results over time. These tools help analysts measure relationships and determine how changes in one variable affect others. As you can detect from this discussion, a data warehouse has no value unless appropriate statistical and analytical tools are selected to provide analysts the ability to analyze data. Data analysis allows analysts to gain insights, knowledge, and intelligence about data stored in the data warehouse. These are a few reasons why organizations are making an investment in statistical analysis tools and data warehouse technology.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

This research established a framework and process to build a metrics-based Healthcare data warehouse. The objective was to identify a process to design and develop a healthcare data warehouse framework that could be used to integrate healthcare information. In order to determine if a corporate data warehouse would improve the delivery of healthcare or assist senior DoD officials with the management of healthcare information, the following research steps were taken:

- Identify the current shortfalls in access to patient care information.
- Document the current flows of patient care information.
- Discuss how an organization can develop a data warehousing system based upon metrics.
- Identify the information needed to be stored in the data warehouse.
- Develop a data model that can effectively implement a healthcare performance management system.
- Demonstrate how access to a healthcare performance management system can help senior DoD healthcare officials make strategic decisions.

DoD healthcare faces several disadvantages by not having information relating to their healthcare beneficiary population stored in a central data repository. Some of the disadvantages include limited access to patient care information, inability to streamline patient record tracking, increased amount of physical storage space needed, inability to conduct a thorough data analysis and inability to establish a foundation for an integrated electronic healthcare delivery system. These challenges are major impediments to identifying ways to reduce cost and improving the delivery of quality healthcare. With limited access to patient's healthcare records and historical healthcare information, healthcare providers and senior DoD officials are unable to identify disease patterns, trends, and relationships which might prevent the spread of a disease or discovery of a cure for a uncontrollable illness.

The DoD healthcare information systems environment is predominantly designed using decentralized systems. This design strategy makes it very difficult to identify the flow of information distributed to healthcare professionals to support decision making. As documented in the Chapter II, there are various independent systems used in DoD healthcare to support different business requirements. Therefore, each functional business area manager is responsible for maintaining and updating data related to his/her specific areas. Based on interviews and discussions with individuals in the healthcare arena, the current environment supports some of their information needs, however there is a serious problem with information being scattered across several different systems. Healthcare personne interviewed at Lee Moore Hospital in California, Davis Grahman in Fairfield, CA, and Presidio of Monterey Clinic, voiced strong opinions that improvements need to take place in the current information technology environment. The central theme of most complaints was too many data sources exist to easily answer senior healthcare officials' questions.

The design and development of a healthcare data warehouse is essential for providing access to information, and developing outcome and population based studies of managed healthcare products and services. This research describes a corporate healthcare data warehouse based upon performance metrics that can serve as a foundation for a PMDSS as well as for extensive data mining and statistical analysis. A high-level data model was used to depict a strategy for identifying information requirements and defining processes involved in designing a performance metric decision support system to analyze data stored in the corporate healthcare data warehouse. The strategy for designing the data warehouse centered on selecting and identifying critical information needed to support the mission, vision, goals, objectives, and the strategic plan. A three-tiered approach was used to ensure that data selected would represent all DoD healthcare stakeholders needs. First, an organization must craft and develop a strategic plan to guide its direction. Second, it is important that a quality model be used to ensure that organizations are building their business strategies on a sound foundation. The Malcolm Baldrige model was selected because it has been used by many healthcare organizations and it identifies seven categories for measuring the efficiency, effectiveness, and the quality of an organization. The third component of the model addresses the importance of creating a performance measurement system. A performance measurement system identifies the critical data elements needed from each operational system to generate the data warehouse schema and help establish a framework to manage a quality organization. Finally, the sum of all these components determines what should be stored in the data warehouse.

After the data warehouse architecture was established, SPSS was used to conduct an illustrative analysis on a healthcare data set from the financial business area to demonstrate the potential utility of the data warehouse. There are several steps, which must take place prior to the analysts having access to the data in the SPSS active window. The following steps are listed below:

1. Data must be extracted from the data warehouse.
2. Some type of application program must be run against the data to transform it into a format readable by the data warehouse database.
3. A data cleansing process must take place to ensure it is accurate.
4. Data is transferred to SPSS so that the analyst can apply queries on the data.

The benefits of the metrics-based data warehouse include: access to data in central location, ability to apply OLAP tools to generate reports and explore the data, use of data mining tools to identify trends, patterns, and relationships in healthcare data, and application of statistical analysis tools to verify/falsify hypothesized models. The data warehouse can also support comparative studies by nation, region, city, clinic, hospital, and time. The information stored in this data warehouse repository can be used in many ways to improve the delivery of healthcare and support in the decision making process. The implementation of such a technology is definitely a commitment to move toward a quality healthcare environment.

In conclusion, we believe healthcare organizations that adopt a metrics-driven data identification methodology will become more efficient. Listed below are the five critical elements included in this methodology:

1. Establishing a sound performance measurement system.
2. Identifying a quality framework.
3. Aligning the data warehouse with the business objectives.

4. Creating a performance metric driven data warehouse architecture.

5. Utilizing statistical analysis and data mining tools to conduct data analyses.

By adopting these principles in developing a data warehouse, healthcare organizations can potentially improve access to information, knowledge, insight, and even return on their investment. The data warehouse plays a central role in the acquisition and dissemination of knowledge through the organization. Decision making is enhanced and the healthcare organization is positioned to react more quickly to business challenges and opportunities. The successful implementation of a data warehouse can have a significant effect in fostering a culture of knowledge sharing in the DoD healthcare environment.

B. RECOMMENDATIONS

This research describes only the conceptual model needed to support a quality healthcare environment. In order to support this healthcare data warehouse framework, the first step should be to design a distributed database environment that will capture the different data elements needed to create the data warehouse environment or architecture. Each functional business area should be required to develop a data mart that consolidates all pertinent information needed to support their prospective business area using a performance metric driven approach and a data warehousing strategy. These data marts would store data needed to manage each business area and have the same data dictionary and schemas as the database warehouse to eliminate anamolies in the data transformation process. The data warehouse architecture depends completely on this data.

DoD healthcare is in the process of moving towards a quality care prospective. Most of the data needed is already being captured in some type of medium. The goal now is to consolidate and gain access to the data needed to support this new approach to DoD healthcare. Once this framework is reviewed by senior DoD healthcare officials and modified to support their needs, a program manger must be identified to champion the project. After a program manger is selected, a team of healthcare professionals, internal and external to the organization, needs to be assembled to define a strategy for making this project a success.

The data warehouse architecture supports not only senior DoD healthcare officials but also tactical and critical operational level healthcare individuals such as physicians, caregivers, and nurses. One physician interviewed believes that the data warehouse could be instrumental in helping build a physicians' decision support system to support physician diagnosis and to map out the critical pathways to effective treatment. Further research would be needed to support this requirement because not only will information be real-time but also secure and integrated with local patients' vital data and global physicians' information such as MEDLINE. The integrative role of a data warehouse opens many doors to new healthcare applications that previously could not be considered within the confines of the current stovepiped healthcare systems.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. HEALTHCARE METRICS

EXAMPLES OF METRIC(S) USED TO BUILD THE HEALTHCARE DATA WAREHOUSE

The metrics lists below are real world applications used in healthcare organizations and recommended by the Joint Commission of Accreditation for Hospitals (JCAHO) located at www.jcaho.org. These metrics have been revised using the attributes defined in the definition of metrics section of this paper found in Chapter III. The target thresholds, periodicity, drill dimensions were selected based on the knowledge of the author. The metrics founded in this appendix are just examples to demonstrate how to develop them. This information is being collected to identify if there are a better way to improve these processes.

Business Area: **Clinical Performance Pneumonia Site Infection**

Computational Procedure: (No. of Ventilated inpatients who develop pneumonia) / (Total no. of inpatients).

Target Thresold:

Acceptable performance - <= 2 percent
Cautionary performance - 2 percent < x <= 10 percent
Unacceptable performance - > 10 percent

Focus of Measure: Surveillance and prevention of ventilator associated pneumonia.
Rationale: This indicator is an important measure of nosocomial infections. The etiology of nosocomial pneumonia in the incubated patient is related to oropharyngeal and gastric colonization with microorganisms, microaspiration, and impaired lung defense mechanisms. The primary factor associated with ventilator-associated pneumonia is the length of time (days) the patient is ventilated. A strong positive correlation between this nosocomial infection rate and proper management of the device in caring for the patient has been demonstrated.

Dimensions/ Units: % / None

Data Element(s): A. Signs and symptoms of pneumonia
B. ICU where pneumonia originated

Corresponding Data Source(s): A-B. Medical record

Included Populations: 1. For hospitals with an ICU(s): by ICU(s) type, identify those patients who have signs and symptoms of pneumonia; identify the ICU where the pneumonia originated. 2. For hospitals without an ICU: identify those patients who have signs and symptoms of pneumonia.

Excluded Populations: 1. For hospitals with an ICU(s): pneumonia present or incubating at the time of ventilator initiation; at the time of ICU admission; or that develops 48 hours or more after discharge from ICU(s) or after removal from the ventilator. 2. For hospitals without an ICU: pneumonia present or incubating at the time of ventilator initiation; at the time of admission; or that develops 48 hours or more after discharge or after removal from the ventilator.

Inpatient (ICU/non-ICU) ventilator days.

Data Source Description
Data Element(s): A. Admission date B. Discharge date C. ICD-9-CM procedure codes D. Patient location while on ventilator E. Total non-ICU ventilator days F. Total ICU ventilator days, specified by type of ICU

Corresponding Data Source(s): A-D. Medical record; billing data; UB-92 E-F. Data source identified by participating health care organization

Included Populations: 1-2. All patients with a length of stay 48 hours or greater who are ventilated as defined by the presence of an ICD-9-CM procedure code, aggregated by total ventilator days by location of care when ventilated

Excluded Populations: 1-2. Patients ventilated exclusively during surgical procedures where the ventilator is discontinued prior to discharge from the postoperative recovery room; patients with a length of stay less than 48 hours

Periodicity: Monthly
Scale Level: Scalar

Drill Dimensions: by clinic, region, hospital, patient gender, patient race, time.

Report Distribution Profile: clinics, region, healthcare administrators, and head physicians.

Submitting Organization Joint Commission on Accreditation of Healthcare Organizations (Joint Commission), Indicator Measurement System (IMSystem®) Profile confirmed March 1996
Original Performance Measure Source/Developer Joint Commission 0100-00041

***Business Area*: Clinical Performance Mortality**

Computational Procedure: (No. of inpatients who die in the hospital) /
(Overall no. of inpatients)

Target Threshold:
Acceptable performance - <= 1 percent.
Cautionary performance - 1 percent < x <= 3 percent.
Unacceptable performance - > 3 percent.

Focus of Measure: Mortality and prehospitalization patient evaluation and education, treatment selection, monitoring, and timely clinical intervention.
Rationale Inpatient mortality is a performance indicator that can help a facility evaluate outcomes related to the inpatient episode. Crude mortality rates do not take into consideration the severity of illness of the patient, nor do they identify specific clinical areas in which the hospital and medical staff may initiate corrective action. Sometimes, these mortality rates are inappropriately used as a proxy for the quality of care delivered. The hospitals' Collaborative Approach to Resource Effectiveness (CARE) Program risk and severity adjusts inpatient mortality, and reports this data by DRG, groupings of DRGs, procedure or clinical service. Hospitals can display these rates over time, identify acceptable ranges of variation, and continuously work to reduce variation as appropriate. This performance assessment indicator (PAI) provides specific information by which the medical staff can monitor important clinical and utilization processes and outcomes, and hospitals can monitor appropriate management support and governance of processes and outcomes.
The ultimate goal of tracking this indicator is to focus resources for continuous improvement and monitor progress over time. This indicator is merely a screening mechanism or starting point for further analysis. The comparison of this indicator rate to appropriate peer group aggregate rates will enable hospitals and their medical staff to identify areas where further investigation may be warranted.

Dimensions/Units: % /None.

Data Source Descriptions:

Data Element(s): A. DRGs B. Patient disposition expired

Corresponding Data Source(s): A-B. UB-92

Included Populations: All discharges where patient disposition status is coded as expired

Excluded Populations: None

Total number of inpatient discharges.

Data Element(s): Discharges

Corresponding Data Source(s): UB-92

Included Populations: All discharges

Excluded Populations: Records with invalid or unknown disposition are not used in the calculation; records with DRGs 385-391, 456, 468-470 and 477 are not used in the calculation

Additional Information Global measures are used for screening purposes. Organizations need to consider getting down to the level of clinical services or DRG in order to benchmark. Forty-one clinical services have been defined and are available upon request from the submitting organization.

Periodicity: Monthly.

Scale Level: Scalar

Drill Dimensions: by clinic, region, hospital, % of patients, ICD-9 code, time.

Report Distribution Profile: clinics, region, healthcare administrators, and head physicians, diagnosis code.

Submitting Organization Georgia Hospital Association Profile confirmed March 1996
Original Performance Measure Source/Developer
Collaborative Approach to Resource Effectiveness (CARE) Program 0105-00083

Business Area: Clinical Performance Asthma

Computational Procedure: (No of Asthma: pediatric discharges)/ (Total No. of inpatients).

Target Threshold:

Acceptable performance - <=1 percent
Cautionary performance - 1 percent < x <= 5 percent
Unacceptable performance- > 5 percent

Focus of Measure: Ambulatory care sensitive (ACS) admissions for pediatric asthma.

Rationale: Adequate and timely preventive, primary, and outpatient ambulatory care may avoid hospitalizations for such chronic conditions as asthma, congestive heart failure, and diabetes. Similarly, effective outpatient treatment of acute conditions--such as pneumonia and cellulitis--may avoid complications requiring hospitalization. Such avoidable hospitalizations are known as access or ACS admissions.

Access/ACS indicators derived from an inpatient discharge database can be used as a proxy measure that hospitals and their medical staff can use as a screening mechanism to identify areas where further investigation may be warranted regarding access to primary and preventive health care services. This information would be useful to a hospital and its medical staff in negotiations with business, employers and insurers, and in the development of community partnerships that examine the delivery system for health care services for a community. Opportunities may exist to collaborate with business to provide health care and/or diagnosis-specific education and prevention programs, or develop local solutions that ensure access to health care. Hospitals and medical staff, in reviewing their community health system, can act as agents for change in their communities in addition to the delivery of the traditional range of health care services.

Dimension/ Units: % / None

Numerator Statement Non-maternal, non-neonatal pediatric discharges less than 18 years old with asthma.

Data Element(s): A. ICD-9-CM diagnosis codes.
B. UB-92's

Corresponding Data Source(s): Medical Records

Included Populations: Non-maternal, non-neonatal pediatric discharges less than 18 years old with asthma. ICD-9-CM diagnosis code 493.xx (asthma).

Excluded Populations: None

Denominator Statement All non-maternal, non-neonatal discharges less than 18 years old.

Data Source Description

Data Element(s): Discharge DRG

Corresponding Data Source(s): UB-92

Included Populations: All non-maternal, non-neonatal discharges less than 18 years old. The following age groups are recommended to facilitate the analysis: infant, 29 days to younger than age one; preschool child, ages one-five years; elementary child, ages six-12 years; teenage adolescents, ages 13-17 years.

Excluded Populations: Discharges 18 years of age or older

Periodicity: Daily

Scale Level: Scalar

Drill Down Dimensions: by clinic, hospital, region, patient gender, patient race, patient age, time, patient status.

Report Distribution Profile: clinic, hospital, region, Nursing Administrator, Medical Administrator, and Pediatric Physician.

Submitting Organization Georgia Hospital Association Profile confirmed March 1996
Original Performance Measure Source/Developer Agency for Health Care Policy and Research (AHCPR) 0105-00072

Business Area: Clinical Performance Asthma

Computational Procedure: (No of Asthma: adults discharges)/ (Total No. of inpatients).

Target Threshold:

Acceptable performance - <=3 percent
Cautionary performance - 3 percent < x <= 10 percent
Unacceptable performance- > 10 percent

Focus of Measure: Ambulatory care sensitive (ACS) admissions for adult asthma.
Rationale Adequate and timely preventive, primary, and outpatient ambulatory care may avoid hospitalizations for such chronic conditions as asthma, congestive heart failure, and diabetes. Similarly, effective outpatient treatment of acute conditions--such as pneumonia and cellulitis--may avoid complications requiring hospitalization. Such avoidable hospitalizations are known as access or ACS admissions.
Access/ACS indicators derived from an inpatient discharge database can be used as a proxy measure that hospitals and their medical staff can use as a screening mechanism to identify areas where further investigation may be warranted regarding access to primary and preventive health care services. This information would be useful to a hospital and its medical staff in negotiations with business, employers and insurers, and in the development of community partnerships that examine the delivery system for health care services for a community. Opportunities may exist to collaborate with business to provide health care and/or diagnosis-specific education and prevention programs, or develops local solutions that ensure access to health care. Hospitals and medical staff, in reviewing their community health system, can act as agents for change in their communities in addition to the delivery of the traditional range of health care services.

Dimensions/ Units: % / None

Data Element(s): A. ICD-9-CM diagnosis codes
B. UB-92's

Corresponding Data Source(s): Medical Records.

Included Populations: Nonmaternal discharges 18 years of age or older. ICD-9-CM diagnosis codes 493.xx (asthma).
Excluded Populations: None

Data Source Description

Data Element(s): Discharge DRG

Corresponding Data Source(s): UB-92; Medical Records; ICD-9-CM procedure codes.

Included Populations: All nonmaternal discharges 18 years of age or older. The following age groups are recommended to facilitate the analysis: young adults, ages 18-44; middle age adults, ages 45-64; elderly adults, ages 65-84; very elderly adults, ages 85 and older.

Excluded Populations: Discharges less than 18 years of age

Periodicity: Daily

Scale Level: Scalar

Drill Down Dimensions: by clinic, hospital, region, patient gender, patient race, patient age, time, patient status.

Report Distribution Profile: clinic, hospital, region, Nursing Administrator, Medical Administrator, and Pediatric Physician.

Submitting Organization Georgia Hospital Association Profile confirmed March 1996 Original Performance Measure Source/Developer Agency for Health Care Policy and Research (AHCPR) 0105-00073

Business Area: Clinical Performance Asthma

Definition: The number Asthma inpatient admissions rate.

Computational Procedure: (No of Asthma: inpatient admission)/ (Total No. of inpatients).

Target Threshold:

Acceptable performance - <=2 percent
Cautionary performance - 2 percent < x <= 8 percent
Unacceptable performance- > 8 percent

Focus of Measure: Ambulatory care for asthma.
Rationale: Approximately 10 million people in the United States are affected by asthma: Of those affected, 2.5 million are children. Overall, asthma is more common in children. Several studies indicate that the incidence of asthma is rapidly rising (33% increase between 1979 and 1987 [Public Health Service, 1990]). Health care costs related to asthma were estimated at $6.2 billion in 1990. This measure addresses the success or failure of outpatient management of asthma.

Numerator Statement 1. (Admissions) Number of members in the denominator with one or more inpatient discharges with a principal diagnosis of asthma (ICD-9-CM diagnosis code 493.xx). 2. (Readmissions) Number of members in the denominator with two or more inpatient discharges with a principal diagnosis of asthma (ICD-9-CM diagnosis code 493.xx).

Dimensions/ Units: % / None

Data Element(s): A. Member ID
B. Diagnosis
C. Discharge date

Corresponding Data Source(s): A-C. Administrative data, Medical record.

Included Populations: Not applicable

Excluded Populations: None

Data Source Description

The number of members, ages 2-19 years and those 20-39 years as of December 31 of the reporting period, who were members as of December 31 of the reporting year, who were continuously enrolled during the reporting year, and who were identified as asthmatic through pharmacy data and/or a face-to-face encounter with a diagnosis of asthma (ICD-9-CM diagnosis codes 493.xx).

Data Element(s): A. Member ID B. Member age (date of birth) C. Enrollment eligibility D. Payer group (for example, commercial) E. Product type (for example, HMO, POS, PPO)

Corresponding Data Source(s): A-E. Administrative data, Medical records

Included Populations: Members who have had one or more prescriptions for cromolyn sodium or aerosol corticosteroid; or one prescription for a bronchodilator and at least one prescription for cromolyn sodium or aerosol corticosteroid; or two prescriptions for a bronchodilator; or two prescriptions for theophylline; and/or two face-to-face encounters in an ambulatory setting; or one face-to-face encounter in an inpatient or emergency room setting with a diagnosis of asthma (ICD-9-CM diagnosis codes 493.xx)

Excluded Populations: None

Periodicity: Daily, Weekly, and Monthly.

Scale Level: Scalar

Drill Down Dimensions: by clinic, hospital, region, patient gender, patient race, patient age, time, patient status.

Report Distribution Profile: clinic, hospital, region, Nursing Administrator, Medical Administrator, and Pediatric Physician.

Submitting Organization National Committee for Quality Assurance (NCQA), HEDIS 2.5
Profile confirmed March 1996
Original Performance Measure Source/Developer NCQA 0121-00282

Business Area: Clinical Performance Breast Cancer, Colorectal Cancer, and Lung Cancer

Computational Procedure: (No. of Cancer Patients: undergoing resection for primary cancer of the lung, colon/rectum, or female breast for whom a surgical pathology consultation is present in medical records) / (Total No. of inpatients).

Target Threshold:

Acceptable performance - <=2 percent
Cautionary performance - 2 percent < x <= 8 percent
Unacceptable performance- > 8 percent

Focus of Measure: Availability of data for diagnosis and staging.
Rationale: The process of communicating pathologic information is of great importance for both prognosis and clinical decisions regarding the treatment of patients with lung, colon/rectum, or female breast cancer.

Dimension / Units: % / None

Data Element(s): Surgical pathology consultations report present

Corresponding Data Source(s): A. Medical record
B. Billing/claims data;
C. Tumor registry

Included Populations: All patients who have a surgical pathology report present in the medical record within 30 days of discharge

Excluded Populations: None

Data Source Description

Patients undergoing resection for primary cancer of the lung, colon/rectum, or female breast.

Data Element(s): A. Admission date B. Discharge date C. ICD-9-CM principal and other diagnosis codes D. ICD-9-CM procedure codes

Corresponding Data Source(s): A-D. Medical record; billing data; UB-92

Included Populations: All inpatients who have a diagnosis of primary cancer of the lung, colon/rectum, or female breast as defined by the presence of a related ICD-9-CM diagnosis code and who undergo surgical resection for the related cancer as defined by the presence of a corresponding ICD-9-CM procedure code, subcategorized by histological type; in situ or invasive type noted; lymph node examination; pTN noted; size of tumor; and status of margins. ICD-9-CM principal and other diagnosis codes and ICD-9-CM procedure code tables available upon request from the submitting organization.

Excluded Populations: None

Periodicity: Daily, Weekly, and Monthly.

Scale Level: Scalar

Drill Down Dimensions: by clinic, hospital, region, patient gender, patient race, patient age, time, patient status.

Report Distribution Profile: clinic, hospital, region, Nursing Administrator, Medical Administrator, and Physician, Pathology Department, OBGYN.

Submitting Organization Joint Commission on Accreditation of Healthcare Organizations (Joint Commission), Indicator Measurement System (IMSystem®) Profile confirmed February 1996
Original Performance Measure Source/Developer Joint Commission 0100-00018

Business Area: Clinical Performance Health Status Breast Cancer Health Maintenance-- Adult

The number of Breast examination conducted. Females age 40 and above should have yearly breast examinations.

Computational Procedure: (No of Female Patients: Breast examination conducted) / (Total No. of Female Patients).

Target Threshold:

Acceptable performance - >= 70 percent
Cautionary performance - 50 percent < x <= 69 percent
Unacceptable performance- < 50 percent

Focus of Measure: Prevention.
Rationale: The effectiveness of clinical examination of the breast combined with mammography has been shown by several studies.

Dimension / Units: % / None

Data Element(s): Breast exam present

Corresponding Data Source(s): Medical record

Included Populations: Not applicable

Excluded Populations: None

Data Source Description

Data Element(s): A. Date of birth B. Gender

Corresponding Data Source(s): A-B. Medical record; administrative data

Included Populations: Not applicable

Excluded Populations: Male patients; female patients younger than 18 years of age

Periodicity: Daily, Weekly, and Monthly.

Scale Level: Scalar

Drill Down Dimensions: by clinic, hospital, region, patient gender, patient race, patient age, time, patient status.

Report Distribution Profile: clinic, hospital, region, Nursing Administrator, Medical Administrator, and Physician, OBGYN, Women's Health Clinic.

Submitting Organization Harvard School of Public Health Profile confirmed March 1996 Original Performance Measure Source/Developer A Project to Develop and Evaluate Methods to Promote Ambulatory Care Quality (DEMPAQ) 0124-00299

Business Area: Clinical Performance Cervical Cancer

Cervical cancer screening.

Computational Procedure: (No of Female Patients: Cervical cancer screening conducted) / (Total No. of Female Patients).

Target Threshold:

> Acceptable performance - >= 70 percent
> Cautionary performance - 50 percent < x <= 69 percent
> Unacceptable performance- < 50 percent

Focus of Measure: Cervical cancer screening; reduce mortality due to cervical cancer. Rationale: Approximately 1,300 new cases of cervical cancer are diagnosed annually; the primary setting test is the Pap test (US Preventive Services Task Force, 1989). A number of organizations (for example, American College of Obstetricians and Gynecologists, American Medical Association, American Cancer Society) have common screening guidelines for cervical cancer. Recommendations include annual Pap smears for all women who are or have been sexually active or have reached age 18. If three or more annual Pap smears have been normal, the interval may be increased from one year to three years at the physician's discretion.

Numerator Statement 1. As determined by claims/encounter data, the number of women in the denominator who had one (or more) Pap tests during the preceding three calendar years (that is, during the reporting year or the preceding two years). 2. As determined in the medical record review, the number of women in the sample who had one or more Pap tests during the preceding three calendar years (that is, during the reporting year or the preceding two years). 3. As determined by administrative data or in the medical record review, the number of women in the sample who had one or more Pap tests during the preceding three calendar years (that is, during the reporting year or the preceding two years).

Dimensions / Units: % / None

Data Element(s): A. Member ID
B. Procedure code
C. Date of service

Corresponding Data Source(s): 1. Administrative data (A-C) 2. Medical record (A-C) 3. Administrative data and/or medical record (A-C)

Included Populations: 1. A woman is identified as having a Pap test if she has a claim/encounter that meets any of the specified codes: CPT-4 code 88150, 88151, 88155, 88156, or 88157; or revenue code 923; or revenue code 300 or 310 in conjunction with one of the following cervical related ICD-9-CM diagnosis codes: 180.x, 233.1, 622.x, 795.0, 795.1, V72.3, or V76.2; or ICD-9-CM procedure code 91.46. 2. At a minimum, documentation will include the date the test was performed, signature or initials of the responsible physician, and the result or finding. 3. If the medical record is used, documentation will include, at a minimum, the date the test was performed, signature or initials of the responsible physician, and the result or finding.

Excluded Populations: None

Number of women, ages 21-64 as of December 31 of the reporting period, who were members of the plan as of December 31 of the reporting year, and who were continuously enrolled during the reporting year and the preceding two years. 2. A random sample of 384 women members ages 21-64 as of December 31 of the reporting period, who were members of the plan as of December 31 of the reporting year, and who were continuously enrolled during the reporting year and the preceding two years. 3. A random sample of 384 women members ages 21-64 as of December 31 of the reporting period, who were members of the plan as of December 31 of the reporting year, and who were continuously enrolled during the reporting year and the preceding two years.

Data Source Description

Data Element(s): A. Member ID B. Member age (date of birth) C. Enrollment eligibility D. Payer group (for example, commercial) E. Product type (for example, HMO, POS, PPO)

Corresponding Data Source(s): A-E. Administrative data

Included Populations: Not applicable

Excluded Populations: None

Periodicity: Daily, Weekly, and Monthly.

Scale Level: Scalar

Drill Down Dimensions: by clinic, hospital, region, patient gender, patient race, patient age, time, patient status.

Report Distribution Profile: clinic, hospital, region, Nursing Administrator, Medical Administrator, and Physician, OBGYN, Women's Health Clinic.

Additional Information 1. Administrative data specification 2. Medical record review specification 3. Hybrid method specification
Submitting Organization National Committee for Quality Assurance (NCQA), HEDIS 2.5
Profile confirmed March 1996
Original Performance Measure Source/Developer NCQA 0121-00281

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. HEALTHCARE DATA ANALYSIS USING SPSS

This appendix contains data printouts from SPSS statistical package version 8.0. The printouts include tables, graphs, and statistical test results to support a descriptive hypothesis model. A multi-linear regression model was used to determine the variables that affect the number of bed days for a patient. A hypothesized model was specified using total bed days as the dependent variable and the patient's age, sex, total government authorized bed days, and DRG type as independent variables. The model was used to predict the affects the independent variables have on the dependent variable (total bed days) and determine if those changes have significant positively or negatively affects on the dependent variables (total bed days). These tables and graphs are provided as documentation.

Overview

The purpose of Appendix B is to illustrate how healthcare data might be analyzed. It is not intended as a full analysis of the available healthcare data. Hopefully, it will provide suggestions as to how one might proceed when conducting a complete analysis. With the explosion of information, healthcare analysts, managers, and end users need statistical analysis tools that will assist them in identifying patterns and trends in data. The statistical analysis tool SPSS is used to perform this task because the Corporate Executive Information System (CEIS) office selected it to be one of the key analytical tools used to retrieve data from the data warehouse and conduct statistical analyses. As shown in Figure 4-4 in Chapter IV, interfaces have been constructed in the CEIS architecture so that SPSS can easily retrieve data from the healthcare data warehouse.

Figure B-1 is a representation of the data used in this appendix, which is taken from a healthcare data warehouse. Figure B-2 illustrates the process that takes place to transfer the data from the data warehouse to the SBSS Data Editor.

The data used in this appendix is collected from the financial business area portion of the data warehouse which includes information such as number of claims processed, total bed days authorized and actually used per hospital visit, number of days to process a claim, Diagnosis Related Group (DRG) information, and demographic information such as age, sex, race, etc related to the processing of claims.

HEALTHCARE DATA WAREHOUSE

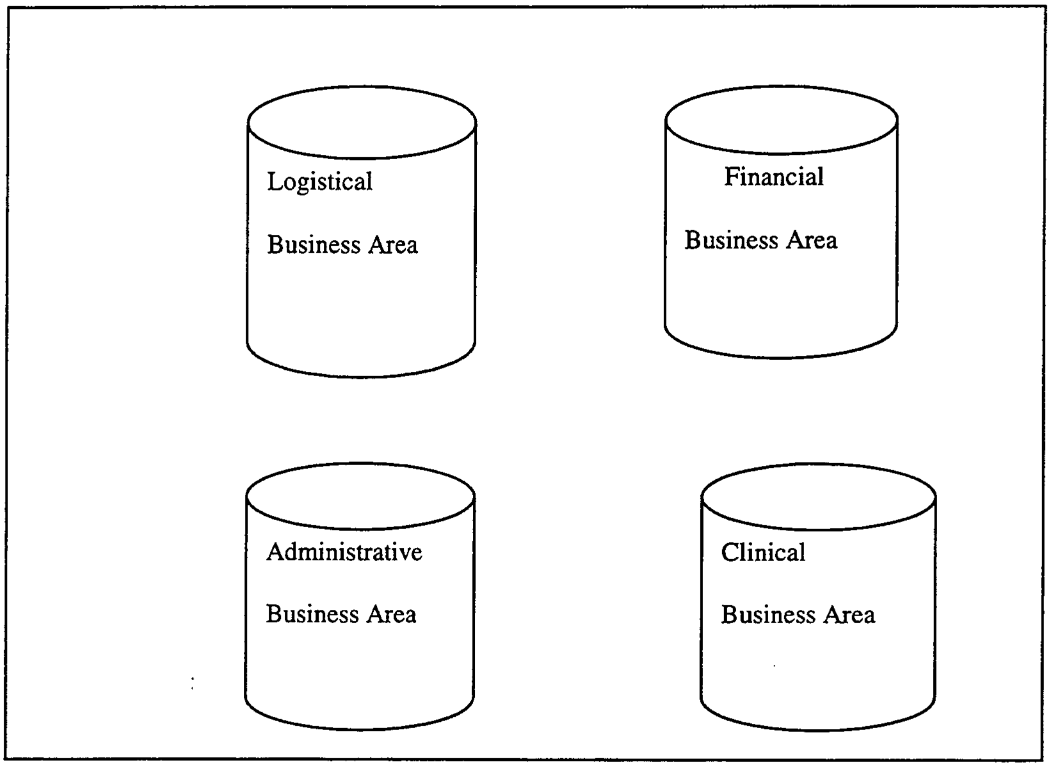


Figure B-1 Healthcare Data Warehouse Logical View

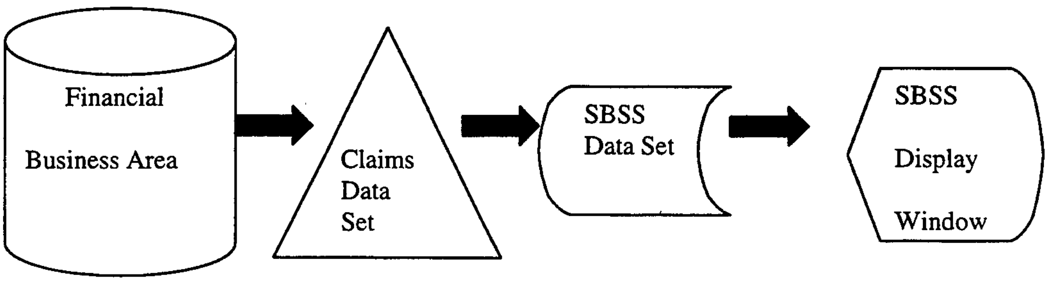


Figure B-2 Data Extraction

Data Exploration

The analyst's first step is to become familiar with the data. It is necessary to determine, based on knowledge and experience, what types of statistical relationships can be developed. Next, one would determine which data variables in the data set are relevant to support the analytical issue being investigated. One central variable in this data set is the total number of bed days that a patient spends in the hospital. Total bed days actually spent in the hospital play a major factor in determining the cost of a patients' healthcare. Therefore, I have hypothesized a statistical regression model in which total bed days is the dependent variable and the patient's age, sex, diagnosis related group code (DRG), and total authorized bed days are the independent variables. The regression model would look similar to the following format.

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + \hat{\beta}_5 X_{5i}$$

$\hat{Y}_i$ represents the predicted value dependent variable total bed days for the ith patient; the independent variables X1i thru X5i represent patient's age, sex, DRG code, government total authorized bed days and DRG type (medical, surgery, or unknown). The model hypothesized that the dependent variable $\hat{Y}_i$ is related to the independent variables. The $\hat{\beta}_0$ equals the estimated intercept of the regression model and the estimated coefficients $(\hat{\beta}_1, \hat{\beta}_{2,}, \hat{\beta}_3, \hat{\beta}_4, and \beta_5)$ describe how the dependent variable predicted value changes when there is a unit change in the relevant independent variable.

One of the main problems typically encountered is that the variables in the original database are not in the format needed to conduct an analysis. If this occurs, recoding of the variables may be required. In certain instances when continuous variables are provided in the data set, recoding to numeric or categorical variables is appropriate to execute specific statistical functions. Listed below in Figure 7-3 is a depiction of the claims data set prior to the recoding of some of the variable and the data set after recoding has taken place. As one can see none of the original variables has been removed, but additional variables have been added to the data set.

DATA SET BEFORE RECODING

DDS	Dependent Number
BIRTH_DA	Birth Date
HCP_NAME	Health Care Provider Name
HCP_SPEC	HCP Specialty Description
CHAMPUS	CHAMPUS Provider Indicator
MDC_DESC	Medical Description
DRG_CD	Diagnosis Related Grouping Code
DRG_DESC	DRG Description
DRG_GROU	DRG Grouper Name
DRG_TYPE	DRG Type
RESIDENC	Residence State
RES_ZIPC	Residence Zip
RES_REGI	Residence Region
RES_CTY	Residence Country
RES_CATC	Residence Catchment Area
SPON_PAY	Sponsor Pay Grade
MEDICARE	Medicare Eligibility
MARITAL_	Marital Status
BRANCH_O	Branch of Service
RACE	Race
GENDER_C	Gender Code
AGE	Age
WEEKDAY	Day of the week
CAL_DATE	Calendar Date
MONTH	Month
CAL_QTR	Calendar Quarter
FISC_QTR	Fiscal Quarter
CAL_YR	Calendar Quarter
FISC_YR	Fiscal Quarter
CLINIC_N	Clinic Name
DMIS_MEP	DMIS Medical Expense Code
MEPRS1_D	DMIS Medical Expense Code Description
REGION_D	Regional Description
BRANCH_S	Branch of Service
MAJOR_CO	Major Command Description
PARENT_D	Parent DMIS or CIVILIAN Name
ZIP_CODE	Zip Code
CONUS_IN	CONUS_Indicator
FACILITY	Facility Country
FAC_STAT	Facility State
AMT_OHI_	Amount OHI Allowed
AMTOHP	Amount OHI _Paid
AMT_PAID	Amount Paid by Govt Cont
AMT_PAYM	Amount Payment Reduction

DATA SET AFTER RECODING

DDS	Dependent Number
BIRTH_DA	Birth Date
HCP_NAME	Health Care Provider Name
HCP_SPEC	HCP Specialty Description
CHAMPUS	CHAMPUS Provider Indicator
MDC_DESC	Medical Description
DRG_CD	Diagnosis Related Grouping Code
DRG_DESC	DRG Description
DRG_GROU	DRG Grouper Name
DRG_TYPE	DRG Type
RESIDENC	Residence State
RES_ZIPC	Residence Zip
RES_REGI	Residence Region
RES_CTY	Residence Country
RES_CATC	Residence Catchment Area
SPON_PAY	Sponsor Pay Grade
MEDICARE	Medicare Eligibility
MARITAL_	Marital Status
BRANCH_O	Branch of Service
RACE	Race
GENDER_C	Gender Code
AGE	Age
WEEKDAY	Day of the week
CAL_DATE	Calendar Date
MONTH	Month
CAL_QTR	Calendar Quarter
FISC_QTR	Fiscal Quarter
CAL_YR	Calendar Quarter
FISC_YR	Fiscal Quarter
CLINIC_N	Clinic Name
DMIS_MEP	DMIS Medical Expense Code
MEPRS1_D	DMIS Medical Expense Code Description
REGION_D	Regional Description
BRANCH_S	Branch of Service
MAJOR_CO	Major Command Description
PARENT_D	Parent DMIS or CIVILIAN Name
ZIP_CODE	Zip Code
CONUS_IN	CONUS_Indicator
FACILITY	Facility Country
FAC_STAT	Facility State
AMT_OHI_	Amount OHI Allowed
AMTOHP	Amount OHI _Paid
AMT_PAID	Amount Paid by Govt Cont
AMT_PAYM	Amount Payment Reduction

AMT_PAYM	Amount Payment Reduction
CLAIM_PR	Claim Process Days
GOVT_AUT	Govt Authorized Bed Days
NAS_ISSU	NAS Issued Claims
NUMBER_O	Number of Payment Reduction Days
NBR_SERV	Number of Services
PATIENT_	Patient Coinsurance Amount
PTCOPMT	Patient Co-payment Amount
PTTOTPMT	Patient Total Payment
PAYMENT_	Payment Amount TPL
TOTAL_AD	Total Admissions
TOTAL_AM	Total Amount Allowed
TOTAMTBI	Total Amount Billed
TOTAL_BE	Total Bed Days
TOTAL_CL	Total Claims
TOTAL_PA	Total Patients

AMT_PAYM	Amount Payment Reduction
CLAIM_PR	Claim Process Days
GOVT_AUT	Govt Authorized Bed Days
NAS_ISSU	NAS Issued Claims
NUMBER_O	Number of Payment Reduction Days
NBR_SERV	Number of Services
PATIENT_	Patient Coinsurance Amount
PTCOPMT	Patient Co-payment Amount
PTTOTPMT	Patient Total Payment
PAYMENT_	Payment Amount TPL
TOTAL_AD	Total Admissions
TOTAL_AM	Total Amount Allowed
TOTAMTBI	Total Amount Billed
TOTAL_BE	Total Bed Days
TOTAL_CL	Total Claims
TOTAL_PA	Total Patients
ADRG_CD	Aggregate DRG Code
GTOT_BE	Grouped Total Bed Days
CLAIMPRG	Claims Process Day Range
AGERANGE	Age Breakdown By Range
RC_DRGCD	Recoded DRG Code
- alc_drug	Alcohol/Drug recoded DRG 433-437;900-901
- burns	Burns recoded DRG 456-460
- dis_diso	Disease & Disorder recoded DRG 302-369,394-9
- endo_nut	Endocrine Nutritional recoded DRG 286-301
- fact_Inf	Factors Influence recoded DRG 461-467
- Human_Im	Human Immunodefiecy recoded DRG 489-490
- Inf_Para	Infections & Parasitic recoded DRG 415-423
- Inj_Pois	Injuries & Poisioning recoded DRG 440-455
- Md_diso	Mental Disease & Disorder recoded DRG 425-31
- Mul_Sigt	Multiple Significant Trauma recoded DRG 484-7
- Myl_diso	Myeloproliferative recoded DRG 401-414,473
- Other	Others recoded DRG 468-9,477-483
- Newb_neo	Newborn & Neonates recoded DRG 600-36,391
- Preg_chd	Pregancy & Childbirth recoded DRG 370-384
RC_RESCY	Recoded Resident City
RC_GENDE	Recoded Gender Variable
GPAUT_BE	Group Authorized Govt Bed Days
RC_Drgtv	Recooded DRG Type (MED, SURG, UNK)

Figure B-3 Claims Data Set Before and After Recoding

Take for example the age variable. Based on the descriptive frequency count applied to the age variable, the subject's age ranges from 0 to 95 years old. Using the recoding option, the analyst can conduct an analysis on the Age_Grp variable, which has been recoded to, represents the age variable in eight categories that are assigned nominal values. After recoding has taken place, the age variable falls into the following categories under the variable Age_Grp. The following represents the nominal value and associated age categories after recoding has taken place.

Nominal Value	Age Category
0	Patient is less than 12 months old.
1	Patient is 1 through 12 years old.
2	Patient is 13 through 18 years old.
3	Patient is 19 through 24 years old.
4	Patient is 25 through 39 years old.
5	Patient is 40 through 55 years old.
6	Patient is 56 through 64 years old.
7	Patient is 65 or older.

Before the regression analysis is performed, it is advisable to conduct a data analysis of the relevant variables. Descriptive statistics are quite helpful for understanding the data. Therefore, the following measures should be included: number of cases, mean, median, minimum, maximum, range, standard deviation, and variance. This process is considered the data exploratory analysis phase. The next step in the data analysis might be to generate graphs for the dependent and independent variables in the model. Listed below are several examples of bar charts and cross tabulation tables depicting information relevant to several variables affecting the model.

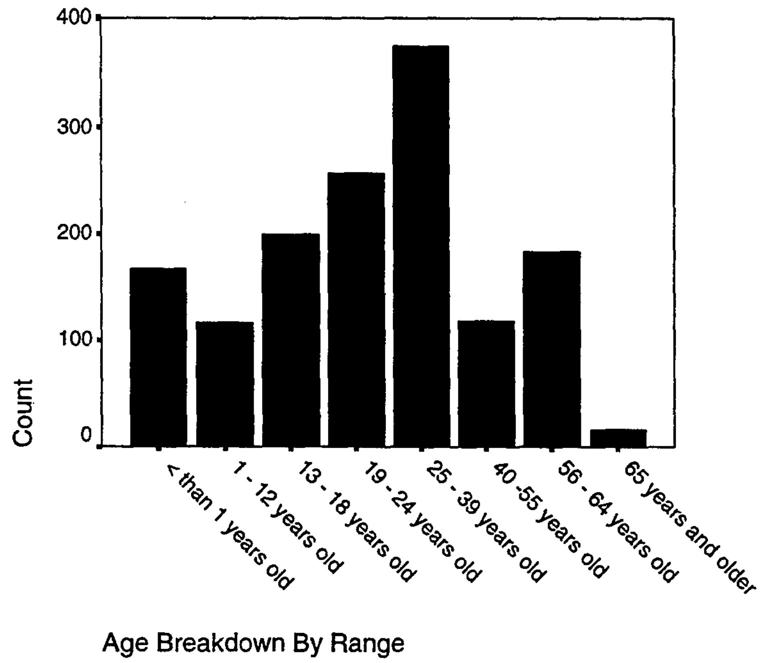
DATA DISTRIBUTION

Description of AGE distribution of Claimants

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
AGE	1431	95	0	95	26.91	18.57	344.762
Valid N (listwise)	1431						

Bar Chart of Patient's AGE Breakdown by Category



As one can see from the descriptive statistics, the ages of the subjects range from 0 to 95 with a mean of 26.91 and a standard deviation of 18.57 years old. This alerts the analyst that the subjects in the data set are fairly young. The bar chart shows the number of individuals in each age category.

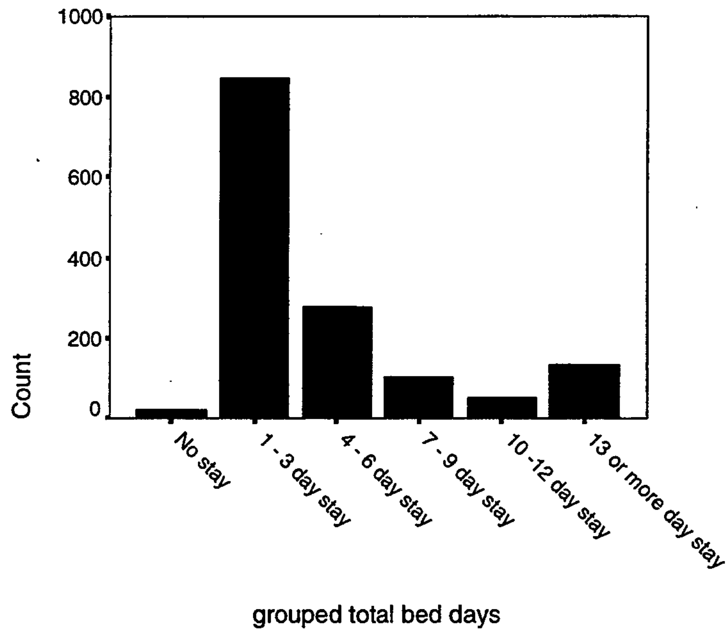
The descriptive statistics and bar charts for total bed days of a patient's hospital stay are displayed next.

Description of TOTAL BED DAYS of a Patient's Hospital Stay distribution of Claimants

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Total_Bed_Days	1431	188.00	.00	188.00	5.3634	9.3020	86.528
Valid N (listwise)	1431						

Bar Chart of Total Bed Days of a Patient's Hospital Stay after Recoding



As one can see from the descriptive statistics, the total bed days of the subjects range from 0 to 188 with a mean of 5.3634 and a standard deviation of 9.3020 bed days. This tells the analyst that the subjects in the data set do not spend, on the average, a long period in the hospital. In addition, based on the information presented in the bar chart after recoding of the variables, over 80.1 % of the patients had total bed days of 6 days or less. The most frequent time spent in the hospital is 1 to 3 days.

Total bed days of a patient's hospital stay have been categorized using the following nominal values.

Nominal Value	Hospital Stay Category
0	Patient did not stay in the hospital.
1	Patient stayed 1 - 3 days in the hospital.
2	Patient stayed 4 - 6 days in the hospital.
3	Patient stayed 7 - 9 days in the hospital.
4	Patient stayed 10 -12 days in the hospital.
5	Patient stayed 13 or more days in the hospital.

The cross tabulation procedure forms two-way tables and measures of the degree between two categorical variables. The cross tabulation table below represents total bed days versus age.

Crosstabulation of Group Total Bed Days and Age

grouped total bed days * Age Breakdown By Range Crosstabulation

Count

		Age Breakdown By Range								
		.00	1.00	2.00	3.00	4.00	5.00	6.00	7.00	Total
grouped total bed days	.00		1			2	5	11	3	22
	1.00	96	82	82	190	248	62	81	5	846
	2.00	34	15	38	39	75	29	46	2	278
	3.00	8	7	28	15	18	10	14	2	102
	4.00	4	4	15	5	7	6	8	1	50
	5.00	25	8	36	8	24	6	23	3	133
Total		167	117	199	257	374	118	183	16	1431

Pearson Chi-Square Value= 197.121 Significance Level Value =.000

Looking at the Chi-Square Test table the observed chi-square value is 197.121 for total bed days and age by range and a significance level value less than .05. This means that there is a significant association between the two variables.

We now show the crosstab table for total bed days and gender. For the gender variable a value of 0 is assigned to males and 1 to females.

Crosstabulation of Group Total Bed Days and Gender

grouped total bed days * Recoded Gender Variable Crosstabulation

Count

		Recoded Gender Variable		
		.00	1.00	Total
grouped total bed days	.00	9	13	22
	1.00	664	182	846
	2.00	188	90	278
	3.00	63	39	102
	4.00	30	20	50
	5.00	67	66	133
Total		1021	410	1431

Pearson Chi-Square Value= 69.295 Significance Level Value =.000

The observed chi-square value is 69.295 for grouped total bed days and gender and a significance level value less than .0005. This means that again there is a significant association between the two variables.

Multiple Linear Regression Model Results

Linear regression analysis is used to examine the relationship between a dependent variable and one or more independent variables. In most realistic situations, predicting the values of a dependent variable requires more than a single independent variable. A multi-linear regression model will be used to estimate the strength and significance of the association between the independent variables used in this analysis and total bed days.

As discussed above, it is hypothesized that total bed days depends on the patient's age, sex, government authorized bed days, and DRG type. The model above is used to predict how the independent variables affect the dependent variable (total bed days) and determines if those changes have significant positively or negatively affects on the dependent variables. The results obtained are contained in the following tables.

COEFFICIENTS

Model	COEFFICIENTS B	Std. Error	t	Sig.
1 (Constant)	.735	.304	2.416	.016
AGE	6.89E-03	.005	1.326	.185
Govt_Authorized_ Bed_Days	.942	.010	92.356	.000
Recoded Gender Variable	.848	.208	4.069	.000
Recoded DRG Type (MED =1, SURG=0)	.349	.258	1.351	.177

a. Dependent Variable: Total_Bed_Days
b. $R^2 = .864$
c. N=1431

The intercept of the fitted regression equation = .735. The coefficient for AGE = .000689, Government Authorized Bed Days = .942, Recoded Gender = .848, and Recoded DRG type = .349 under the " Coefficient " column. These coefficients describe how the dependent variable is predicted to change when there is a unit change in an independent variable. Therefore, the equation of the least squares regression line is

$$\hat{Y}_i = .735 + .000689X_{1i} + .942X_{2i} + .848X_{3i} + .349X_{4i}$$

As an example, the coefficient of Government Authorized Bed Days equal to .942 means that a unit changes in Government Authorized Bed Days is associated with a .942 increase in predicted total bed days, other variables held constant. Each variable in this regression model is positive. In addition, a t-statistic values greater than 1.96 (with an associated significance [Sig.] of less that .05) indicates that the coefficient of a variable is statistically significant. Most of the variation to total bed days is explained in government authorized bed days and gender variables. The variables AGE and DRG type do not have statistically significant coefficients.

SUMMARY

This preliminary analysis is very illustrative, and a great deal of further analysis can be done with this data set. However, the intention was to use this illustration to demonstrate the use of SPSS on a data set and familiarize one with the steps taken by analysts in order to start the analytical process. It was hypothesized that the dependent variable total bed days could be predicted using a patient's age, gender, DRG code, government authorized bed days, and DRG type. The results were analyzed and it appears that since government authorized bed days has such a high t-statistic value that additional analysis should be undertaken to examine the determinants of government authorized bed days. This, however, would take us beyond the intended scope of this Appendix.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Berelowitz, Roy. " Making Data Visual," URL: http://www.pbviews.com/perform/articles/metrics_visualization.html/, (25 Oct 1998).

Berson, Alex and Smith, Stephen J. "Data Warehousing, Data Mining, and OLAP" McGraw-Hill New York, San Francisco, Washington, D.C., 1997.

Cabena, Hadjinian, Stadler, Verheees, Zanasi. " Discovering Data Mining from Concept to Implementation, Prentice Hall PTR, Upper Saddle River, New Jersey, 1998.

Composite Healthcare System (CHCS), URL: http://www-nmimc.med.navy.mil/chcs.html/, (21 May 1997).

Dolk, Daniel. NPRDC Model Integration DSS Report September, 1998.

Groth Robert. " Data Mining: A Hands-On Approach for Business Professionals". Prentice Hall PTR, Upper Saddle River, New Jersey 1998, pp. 40- 50.

Inmon W.H. "Building the Data Warehouse" Wiley Computer Publishing, John Wiley and Sons, Inc, New York and Singopore 1996.

Janah, Monue "Integrated System Directives" June 1, 1998, pp. 46-58.

Joint Commission of the Accrediation of Healthcare, URL: http://www.jcaho.org/, (22 Feb 1998).

Kendall, Kenneth E., and Kendall, Julie E. " System Analysis and Design" Prentice Hall, Englewood Cliffs, New Jersey 07632, 1992.

Koehler, Jerry W. and Pankowski, Joseph M. " Continual Improvement in Government: Tools and Methods" St. Lucie Press, Delray Beach, Florida 33483, 1996.

Kongstevedt, Peter R. " Essential of Managed Healthcare." Aspen Publication. Aspen Publishers, Inc. Gathersburg, Maryland 1997, pp. 30-40.

Kroenke, David. " Database Processing, Fundamentals, Design, and Implementation, 6th edtion, Prentice Hall, Upper Saddle River, New Jersey 07458, 1988.

Macleod,G.K., " An overview of Managed Care: In the Managed Healthcare Handbook, 2d ed., ed. P.R.Kongstrevdt (Gaithersburg, MD.:Aspen,1993), pp. 3-11.

Moore, David. " Statistics: Concepts and Controversies" Fourth Edition, Purdue University, W.H. Freeman, 1997.

[Neely et al 1995] Neely, Gregory, Platts. Performance measurement system design: A literature review and research agenda. Int'l Journal of Opns & Prodn Mgmt, Vol 15, No 4 (1995), pp. 80-116.

Phelps, Charles E. " Health Economics" Harper Collins Publishers Inc. 1992.

Rauber, Chris. Evolution or Extinction Modern Healthcare October 19,1998, pp. 54-67.

Sperley, Eric "The Enterprise Data Warehouse: Planning, Building, and Implementation." Prentice Hall PTR, Upper Saddle River, New Jersey 07458, 1999, pp. 73- 271.

Star J. and Estes, J "The Geographic Information System: An Introduction." Prentice-Hall, Inc., Englewoods Cliffs, 1990.

Stephenson Paul L., Rogness, Neal T., Ritchie, Justine M., and Stephenson, Patricia A.B. Introduction to the Practice of Statistics. 1999.

Strickland, A.J. and Thompson, Arthur A. " Strategic Management: Concepts and Cases." Irwin, Homewood, Ill 60430, 1992.

The President's Quality Award Program, "1999 Application and Information" United States Office of Personnel Management, PQAP-1 April 1998.

Thomsen, Erik. "OLAP solutions: Building Multidimensional Information Systems". Wiley Computer Publishing John Wiley and Son, Inc. New York and Weinbeim 1999, pp. 7.

Walton, Mary. "Deming Management at Work". Perigee Books, The Putnam Publishing Group, New York, NY 10016, 1991.

BIBLIOGRAPHY

Alexander Ted and Bischoff, Joyce. " Data Warehouse Practical Advice from the Experts" Prentice Hall, Upper Saddle River, New Jersey 1997.

Adriaans Pieter and Zantinge Dolf. " Data Mining " Addison –Wesley. Harlow, England and Reading, Massachusetts 1996. pp. 26-29.

Brackett Michael H. " The Data Warehouse Challenge: Taming Data Chaos" Wiley Computer Publishing, John Wiley and Sons, Inc. New York, 1996.

Devlin, Barry. " Data Warehouse from Architecture to Implementation" Addison-Wesley, 1997.

Hammergren Tom " Data Warehousing: Building the Corporate Knowledge" International Thomson, Computer Press, London and Detroit Michigan, 1996.

Kaplan, Robert S. and Norton, David P. " The Balanced Scorecard: Translasting strategy into action" Harvard Business Press, Boston, Massachusetts 1996.

Keen, Peter G.W. "Shaping the Future: Business Design through Information Technology." Harvard Business School Press, 1991.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center 2
   8725 John J. Kingman Road, Ste 0944
   Fort Belvoir, VA 22060-6218

2. Dudley Knox Library 2
   Naval Postgraduate School
   411 Dyer Road
   Monterey, California 93943-5101

3. Professor Daniel Dolk, Code IS/Dk 1
   Naval Postgraduate School
   Monterey, CA 93943

4. Professor Gregory Hildebrandt, Code SM/Hg 1
   Naval Postgraduate School
   Monterey, CA 93943

5. Professor D. C. Boger, Code IS 1
   Naval Postgraduate School
   Monterey, CA 93943