2006

INFORMATION SYSTEM CONTEXTUAL DATA QUALITY: A CASE STUDY

Daniel Lee Davenport
University of Kentucky, dldave0@email.uky.edu

ABSTRACT OF DISSERTATION

Daniel Lee Davenport

The Graduate School

University of Kentucky

2006

INFORMATION SYSTEM CONTEXTUAL DATA QUALITY: A CASE STUDY

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Carol Martin Gatton College of Business and Economics at the University of Kentucky

By
Daniel Lee Davenport

Lexington, Kentucky

Director: Dr. Clyde W. Holsapple
Rosenthal Endowed Chair in Management Information Systems
Professor of Decision Science and Information Systems

Lexington, Kentucky

2006

ABSTRACT OF DISSERTATION

INFORMATION SYSTEM CONTEXTUAL DATA QUALITY: A CASE STUDY

This dissertation describes a case study comparing the effectiveness of two information systems that assess the quality of surgical care, the National Surgical Quality Improvement Program (NSQIP) and the University HealthSystem Consortium Clinical Database (UHCCD). For the comparison, it develops a framework for assessing contextual data quality (CDQ) from the decision maker's perspective. The differences in quality assessment systems to be studied are posited to be due to the differing contexts in which the data is encoded, transformed and managed impacting data quality for the purpose of surgical quality assessment.

Healthcare spending in the United States has risen faster than the rate of inflation for over a decade and currently stands at about fifteen percent of the Gross Domestic Product. This has brought enormous pressures on the healthcare industry to reduce costs while maintaining or improving quality. Numerous systems to measure healthcare quality have been, and are being, developed including the two being studied. A more precise understanding of the differences between these two systems' effectiveness in the assessment of surgical healthcare quality informs decisions nationally regarding hospital accreditation and quality-based reimbursements to hospitals.

The CDQ framework elaborated is also applicable to executive information systems, data warehouses, web portals, and other information systems that draw information from disparate systems. Decision makers are more frequently having data available from across functional and hierarchical areas within organizations and data quality issues have been identified in these systems unrelated to the system performance from which the data comes.

The propositions explored and substantiated here are that workgroup context influences data selection and definition, the data entry and encoding process, managerial control and feedback, and data transformation in information systems. These processes in turn influence contextual data quality relative to a particular decision model.

The study is a cross-sectional retrospective review of archival quality data gathered on 26,322 surgical patients at the University of Kentucky Hospital along with interviews of process owners in each system. The quality data include patient risk/severity factors and outcome data recorded in the National Surgery Quality Improvement Program (NSQIP) database and the University HealthSystem Consortium Clinical Database (UHCCD).

KEYWORDS: Contextual Data Quality, Data Quality, National Surgery Quality Improvement Program, University HealthSystem Consortium Clinical Database, Workgroup Context.

Daniel L. Davenport

February 10, 2006

INFORMATION SYSTEM CONTEXTUAL DATA QUALITY: A CASE STUDY

By

Daniel Lee Davenport

Dr. Clyde W. Holsapple

Director of Dissertation

Dr. Merl M. Hackbart

Director of Graduate Studies

February 10, 2006

RULES FOR THE USE OF DISSERTATIONS

DISSERTATION

Daniel Lee Davenport

The Graduate School
University of Kentucky
2006

INFORMATION SYSTEM CONTEXTUAL DATA QUALITY: A CASE STUDY

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Gatton College of Business and Economics at the University of Kentucky

By
Daniel Lee Davenport

Lexington, Kentucky

Director: Dr. Clyde W. Holsapple
Rosenthal Endowed Chair in Management Information Systems
Professor of Decision Science and Information Systems

Lexington, Kentucky

2006

This dissertation is dedicated to Sally, Katie, and my father.

ACKNOWLEDGEMENTS

I wish to acknowledge and thank Dr. Clyde W. Holsapple, Dissertation Director, and Dr. Robert M. Mentzer, Jr., Chair of the Department of Surgery at the University of Kentucky. Their scholarly excellence and leadership helped me throughout this project and continue to inspire me.

The following dissertation used data obtained by two dedicated and skilled nurse reviewers, Ms. Devauna Riley and Ms. Mary Beth Rice, as part of the National Surgical Quality Improvement Program. Data was also obtained from the Clinical Data Products Data Base maintained by the University HealthSystem Consortium.

Finally, my wife encouraged me to pursue a dream and has supported this work in innumerable ways.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

In 2002, the investigator reviewed five different reports regarding the quality of cardiac surgical care at an academic medical center. These reports were generated from large regional/national information systems that assessed surgical quality in terms of risk-adjusted mortality. Each report produced an assessment of the quality of the hospital's performance in terms of the number of cardiac patients who died versus an expected death rate based on statistical models in the systems. The assessments varied by a factor of more that two-to-one across the different systems. At the extremes, one report stated that significantly more patients died than would be expected versus another report that stated that slightly fewer patients died than expected (Personal Experience of the Investigator, 2002).

Understanding these report differences is important because: 1) they assess patient death, so the negative results of the worst report cannot be ignored, and 2) one of the reports is in support of a contract with a major payer representing millions of dollars of revenue for the hospital. Which, if any, of the reports accurately depicts the quality of cardiac care at this hospital? Why are the reports so different in their assessments? What should the hospital do regarding these diverse assessments of patient care quality? What should the CEO say to the press if the "bad" report shows up on the front page of the local paper?

This frustrating experience along with the growing national prominence of databases assessing surgical outcomes is one impetus for this research. It compares and contrasts two healthcare quality assessment systems, the National Surgical Quality Improvement Program (NSQIP) and the University HealthSystem Consortium Clinical Database (UHCCD). In doing so, it introduces, explores and elaborates a framework for the evaluation of contextual data quality (CDQ) in information and decision support systems.

The development of the CDQ framework is also motivated by reported data quality issues in ERP systems, Data Warehouses, Intra- and Extranet Web Portals, Executive Information Systems, and Decision Support Systems. These systems are more frequently using the power of networks to take information from legacy systems, once confined to the work community using them, and rapidly distributing it to others outside that community. A review of the data quality literature yields several studies (Koh and Watson, 1998, Ballou and Tayi, 1999, Yoon, Aiken and Guimaraes, 2000, Kumar and Palvia, 2001, Wixom and Watson, 2001, Reed and Catterall, 2005) finding data quality problems related to such systems, with a common theme being issues of context across systems and groups. Given the increasing mobility of data in the current networked age, it is important for decision makers to understand when data from well functioning systems is more or less fit for use for their particular decision.

The two systems studied in this case are well suited for stimulating and elaborating a theory of information system CDQ. There is no evidence that the two healthcare information systems in this case study are not functioning as designed. However, they are managed and used in different contexts. Comparing these two systems yields significant insights into CDQ, the goal of this study.

The study is an in-depth case review of a single site. Following the recommendation of Yin (1984, 2003) regarding case study, the guiding statements of the study are presented as propositions rather than theoretical hypotheses with hypothesis development being the outcome of the research. The study is exploratory rather than confirmatory and its value lies in the resulting new theoretical hypotheses. The case chosen fits well with criteria given by Yin (1984, 2003) for case study research and by Benbasat et al. (1987) for IS case study research in particular.

Chapter 2: Background and Significance

2.1 Healthcare Costs and Quality

Healthcare in the United States is a large business sector and has experienced significant growth in recent years even during the recent recession. Total healthcare expenditures in the United States increased from $1.05 trillion in 1997 to $1.55 trillion in 2002 and per capita expenditures increased from $3,517 to $4,695 (CMS, 2004). As a result, both federal and private payer groups have attempted to reduce costs. Commensurate with the pressures on cost containment has been an increased scrutiny and concern with quality. For example, 160 large insurers and Fortune 500 corporations who purchase and indemnify healthcare have joined together to form the Leapfrog Group whose goal is to “trigger giant leaps forward in the safety, quality and affordability of health care.” (Leapfrog Group, 2004)

2.2 Information Systems that Assess Healthcare Quality

In support of these concerns, numerous systems for measuring the quality of care at hospitals have been, and are being developed. These include, but are not limited to, standards established by the Leapfrog Group mentioned above, accreditation requirements imposed by the national Joint Commission for the Accreditation of Healthcare Organizations (JCAHO, 2004), Center for Medicare and Medicaid Services’ (CMS) standards for premium reimbursement, along with more targeted systems like the University HealthSystem Consortium Clinical Database (UHCCD) and the National Surgical Quality Improvement Program database (NSQIP). These last two are the subjects of this research as they have the same general approach of risk-adjusting outcomes in order to measure quality but derive their data quite differently.

These various systems that assess the quality of healthcare are important for their potential impacts on improving healthcare outcomes, regulatory compliance, payer contracts and reimbursement, as well as the public reputation and marketability of healthcare providers. In terms of improving patient outcomes, for example, the death and complication rates at the Veteran's Affairs hospitals performing major surgery have decreased during the last decade while the administration of the NSQIP has occurred. (Fink et al, 2002) As mentioned above, certain standards are mandated by JCAHO for a hospital to stay accredited. (JCAHO, 2004) Recently CMS, the major federal payer for healthcare, has started making premium payments for hospitals that meet certain quality criteria. In addition to financial impact from payers, regulatory compliance, and improved outcomes, the reputation and market viability of providers, both hospitals and clinicians, can be elevated or devastated by positive or negative reports respectively. As an example, early efforts at reporting cardiac surgery outcomes by surgeon in New York State resulted in several surgeons discontinuing their practice in that State. (Harlan, 2001) Given these impacts from quality reporting, a better understanding of the efficacy of information systems that assess quality is critical.

2.3 A Healthcare Quality Assessment Model (The Decision Model)

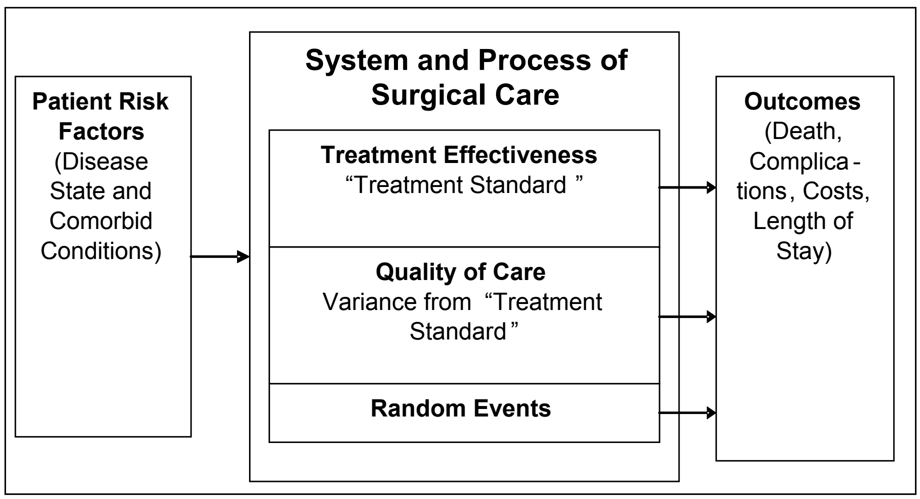
Quality in healthcare can be assessed in different ways. Three general areas of assessment are those using process, structural, or outcomes variables. (Donabedien, 2003) Process-based quality assessment measures whether a healthcare process known to be effective is implemented or not. An example would be the CMS core quality measure of whether or not patients who are smokers receive smoking cessation counseling while they are in the hospital. The percentage of smokers who do is compared to a national standard. In contrast, structural based assessments use measures such as the number of patients treated for a particular disease at a hospital or whether the ICUs are staffed by full time specialists called intensivists.

The NSQIP and UHCCD apply the third approach using outcome variables, and are based on Iezzoni's (Iezzoni, 2003) "algebra of effectiveness." This formula states that:

Healthcare Outcomes = *f* (intrinsic patient-related risk factors,
treatment effectiveness, quality of care, random chance)

Because outcomes depend on patient risk factors, systems that assess quality by measuring outcomes must also adjust for those factors and are called risk-adjustment systems. Figure 2.1 shows a block diagram illustrating the elements of the system model.

Figure 2.1 System Model of Factors Influencing Surgical Outcomes (derived from Iezzoni, 2003, Risk adjustment for measuring healthcare outcomes, HAP.)



Risk adjustment information systems normally apply this model in similar ways. They first identify an outcome of interest from a quality perspective, mortality in surgical patients for example. Next, they examine available patient-related risk factors that have been shown to influence the outcome. Patients with a history of heart disease for example may experience higher rates of complication and death after surgery unrelated to the quality of care provided them. Comparisons in the mortality rates for surgical patients at different hospitals must then control for the occurrence of heart disease in the respective hospitals' patient populations. Data is obtained for both the risk factors and outcomes across a statistically sufficient sample of patients and hospitals in order to construct regression models with which to compare the outcomes levels.

These models relate the patient risk factors to the outcome of interest. The modeling retains the population mean occurrence of outcomes for a particular patient risk profile and excludes variation related to individual sites. The resulting model is then used to estimate the risk of death or complication for each patient at a site. For mortality models, the individual patient estimates are summed up to predict the number of patient deaths at a particular site. This predicted number of deaths and complicated patients then becomes the standard for judging quality. It estimates the outcomes related to a "standard of care" as measured by the mean national death rate for a given set of patient risk profiles. A variance from this *de facto* "standard of care" is then due to one of three factors:

1) Random events,

2) The quality of care which includes variation from the standard process/system of care and error, or,

3) A risk factor unaccounted for in the model.

In order to assess the likelihood of variance in the observed to expected performance being due to random events, the model calculates confidence intervals for the estimates at each site. Observed values outside those intervals have a high confidence of being due to real differences caused by quality differences or other unknown factors rather than random events. The confidence intervals are strongly influenced by sample size and the ability for the model to distinguish random versus real differences is reduced for small samples. This has implications for assessing procedure-specific or surgeon-specific quality under this methodology. The models are also affected by the population from which they are drawn.

In terms of factors for which the model accounts, this becomes the area of interest for this study. Different systems use different types and numbers of factors based on the data they contain. Differences in assessments of the same hospital by different systems are then due to:

1. Inaccuracies in the data or errors in the modeling processes.

2. Differences in the populations from which the statistical models are derived.

3. Differences in the data recorded in the systems ability to estimate particular outcomes.

This latter cause is the primary focus of this study.

2.4 Application of the Case to Broader Information System Contextual Data Quality Issues

There are data quality problems associated with pulling data from systems designed for one purpose and using it for another. Executive Information Systems, Enterprise Resource Planning Systems, Data Warehouses, On-line Analytical Processing Systems, and Web Portals pull information from systems existing in different functional areas, managerial levels and locations, and deliver it to disparate users through a network connection. The ability to pull from diverse systems and deliver to diverse users is a challenge from a data and decision quality perspective. Users who own and regularly work with a particular dataset better understand its deficiencies and utilize that understanding when making decisions based on the data. The increased access to information by non-system owner users is part of the reason for an increased awareness of data quality issues. (Ballou & Tayi, 1999)

Koh and Watson (1998) analyzed data quality by surveying 85 organizations regarding executive information system (EIS) development and maintenance. Of the data quality issues they identified, the one most important and difficult for EIS managers was data standards. These managers reported that data standards are particularly challenging in EIS because of the "variety of data sources that cross functional boundaries and management hierarchies (p. 310)." They note that development of an EIS frequently uncovers many data compatibility and consistency problems that have gone unnoticed during the normal operation of the system. These "uncovered" data quality problems unrelated to normal operations of the system are also noted by Reed and Catterall (2005) in regards to CRM implementations.

Kumar & Palvia (2001) surveyed 48 firms regarding global EIS's. They reported that important issues impacting data management of EIS's were data integrity in feeder data sources, data security and data standards. They found that "business and IT staff in subsidiaries need to agree on common definitions of data entities and attributes (p. 160)." Inconsistencies in data among subsidiaries were common and a recognized problem.

Yoon and Aiken (2000) found similar data quality issues related to data definitions and proposed a new four-dimensional corporate data quality framework. Their dimensions were three common ones; the data value, the data representation, and the data model to which they added a fourth, the data architecture. Data architecture refers to metadata about data models held throughout the organization. It includes "information on relevant entities and attributes, such as their names, definitions, a purpose statement describing why the organization is maintaining information about this business concept, their sources, logical structures, value encoding, stewardship requirements, business rules, models associations, file designs, data uses, specifications, repositories, etc. (p. 6)" Their development of the data architecture dimension is in response to the "increasingly widespread requirement that users interact with multiple systems, and the need for developers to build more highly integrated system" in order to "coordinate data management activities in cross-functional system development and operations (p. 9)." In other words, the data architecture they propose seeks to build contextual information across the institution's diverse functional areas thereby improving data quality.

These studies highlight contextual data quality problems related to moving information out of the bounds of the system's work group owners to other users and decision makers. The information quality issues frequently do not arise within the functional work group primarily using the system, but only in transferring the information outside that group.

These data quality problems could be considered a system design problem, but a study by Wixom and Watson (2001) concludes differently. Their study of data warehouses surveyed 111 pairs of data warehousing managers (system managers) and data suppliers (analyst users) in order to investigate factors affecting data warehouse success. They separated the two outcomes of users' perceptions of *system quality* and *data quality*. Their results regarding users' perceptions of *system quality* supported the frequently cited positive impact of management support, a champion, allocated resources, and user participation in design. However, these factors, along with design team skills, source systems and development technology were not found to affect perceived *data quality* in data warehouses. They concluded, "*data quality* is best explained by factors not included in our model." In other words, the system can be functioning well and as designed yet still lead to poor data quality for certain users.

Two current trends confirm the importance of data quality in general to corporate America. The first is the number of articles in the business press on data quality (two recent examples are Redman 2005, and MarketWatch: Global Roundup, 2005). The second is the development of a market, since the start of this study, of IS vendors selling "data quality" software. Sales are estimated at $250 million to $300 million annually and growth is expected at 12% to 15% annually in the near future (Bailor, 2005).

This study posits that CDQ is one of the factors that significantly impacts system quality although it is not included in many system analysis models. In this case study of surgical quality assessment, both databases are nationally recognized with no evidence in the literature that the systems are malfunctioning. There is however criticism of the "fit" of the data from hospital and claims administrative systems for the assessment of clinical quality. (Jollis et al., 1993, Green and Wintfeld, 1993, Hannan et al., 1997, Davenport et al., 2005) This criticism leads to the first two propositions for this case study.

Proposition 1: The systems are not equally effective in risk-adjusting surgical outcomes.

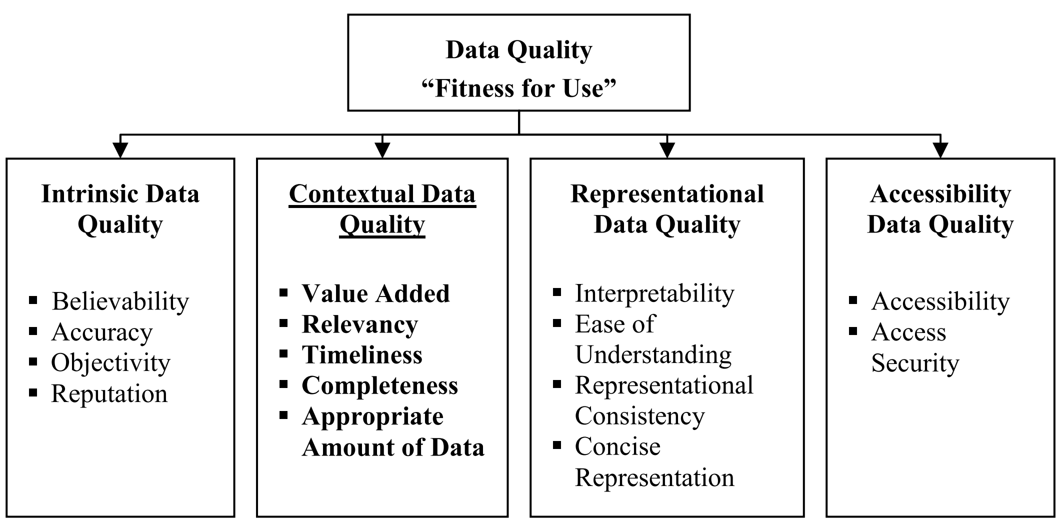
Proposition 2: Differences in the two information systems' effectiveness in risk-adjusting surgical outcomes are not due to system failure, but to differences in the workgroup context in which the data is derived.

2.5 Fitness for Use and the Contextual Quality of Data

In order to better understand contextual data quality issues that are not related to system failures, a user-centric definition of data quality is necessary. One perspective of data is that it is a "good" that is manufactured and then consumed by the user. (Wang & Strong, 1996) From this perspective high quality data are data that enable the user to effectively and efficiently make a decision or execute a task. In other words, quality data is data that are "fit for use" from the perspective of the objectives of the information consumer.

Wang & Strong's (1996) study of quality from the user's perspective resulted in 4 categories of 15 data quality sub-dimensions or fifteen different ways in which information can be more or less useful to the consumer. They are shown in Figure 2.2.

Figure 2.2 Data Quality Categories and Dimensions. (Wang & Strong, 1996)



The second of these four categories, contextual data quality, has historically been perceived as relevancy and timeliness (Holsapple & Whinston, 1996), the addition of completeness, value-added, and the amount of information are contributions of

Wang & Strong's study. Contextual data quality, from their perspective, differs from the other three categories in that it is dependent on the context of the user and therefore is not an information system attribute *per se* but a fit of the IS attributes to the information consumer's particular need. This category of data quality is of particular importance to systems that pull data from across functional and hierarchical boundaries. For instance, the individual dimensions form a significant portion of the E-Quality framework for web-based information put forth by Kim et al. (2005). We apply Wang and Strong's contextual data quality dimensions when analyzing this case. Application of this dimension takes the form of Proposition 3.

> Proposition 3: Differences in the two information systems' effectiveness in risk-adjusting surgical outcomes are due to differences in their contextual data quality dimensions of added value, relevancy, timeliness, completeness, and appropriateness of the amount of the data in the systems.

We define contextual data quality based on the above discussion as:

> *The fitness for a particular use of a dataset based on the context in which it was derived.*

2.6 Information System Context: Work Communities

In considering the different contexts that information crosses between groups, the concept of "community of practice" can be applied. From this perspective, work communities interact to create usable cognitive, social, physical, and system artifacts as they pursue common goals. These artifacts include language, routine, sensibilities, tools, stories, and styles and become the shared "repertoire" of a group. According to Wenger (2003), to be competent in a particular community is to have access to the common repertoire and to use it appropriately. Many of the elements of a common repertoire follow closely Newman's (2003) concept of knowledge artifacts. For Newman, artifacts (and in particular knowledge artifacts) are any human constructions and include both mental and physical components. From this perspective, the common repertoire of a community consists of its mutually developed and held knowledge artifacts used in support of its decisions and actions.

Information systems can be viewed as specific examples of knowledge artifacts with physical and cognitive components developed by a particular community in achieving its common goals. As such, information system context can be analyzed in light of community structure to give some indication of when information from the system may have less fitness for use for those outside the community. The community structure elements include the goals and common repertoire elements of language, routines, sensibilities, etc. mentioned above. The two systems compared in this study are embedded in different work community contexts.

2.6.1 The NSQIP Community

The NSQIP was initiated in 1991 by the National Veterans Administration (VA) as the National Veterans Affairs Surgical Risk Study in response to a congressional mandate to demonstrate the quality of care being delivered to veterans. It was designed from the beginning to measure surgical care quality. In 1994, the success of the program resulted in the Veterans Administration expanding it to include all veterans hospitals and it was renamed the National Surgical Quality Improvement Program. (Khuri et al., 1997) The program is managed by the Chief of Surgery inside the Department of Surgery in the various hospitals and data is collected by clinically trained nurse reviewers who report to the Chief of Surgery.

In 1999, the University of Kentucky Department of Surgery, along with two other academic medical centers, started a private sector pilot study applying the NSQIP to non-VA centers. This pilot study was effective in applying the NSQIP to the non-VA sector. (Fink et al., 2002) The Surgery Department at the University of Kentucky Hospital has submitted data on risk factors and outcomes for surgical patients since October 6, 1999.

2.6.2 The UHCCD Community

The University HealthSystem Consortium, (UHC) formed in 1984, is an alliance of academic health centers situated mainly in the United States. Its first major project was collective bargaining for the purchase of medical supplies. Its members are hospitals, and UHC provides its 90 full members and 120 associate members with a variety of resources aimed at improving performance levels in clinical, operational, and financial areas. The mission of the UHC is to advance knowledge, foster collaboration, and promote change to help members succeed in their respective markets (UHC, 2004).

The UHC's Clinical Database developed from its members' needs for reporting quality information. It pulls much of its data from the hospital administrative and cost accounting systems. In the hospitals, the data is managed by medical records or quality improvement personnel. Clerks on the various wards or trained medical record coders abstract the data from the written medical record.

In comparing the communities around these two databases there are two immediate differences. Historically one database was designed directly to assess quality while the other evolved in its initial use of primarily administrative data. The NSQIP is managed by surgeons primarily, while the UHCCD data is managed by financial or quality improvement managers. These differences are posited to impact the quality of the data through differences in the respective work community's context.

2.7 Workgroup Context Influences on IS Processes

In using information systems, work communities utilize ontologies embedded in their language to provide for shared discourse and understanding among the agents in the community. An ontology is a description of the concepts and relationships that can exist for an agent or a community of agents (Gruber, 1993). An information system is also a set of concepts and relationships defined by a particular agent or community of agents. Like ontologies, information systems are inherently a simplification of the real world events they represent.

Given any real world event, an information system can only capture and portray part of the reality and does so based on the ontology of the workgroup that creates and manages it. In this way workgroup ontology represents a significant portion of how workgroup context influences IS processes.

Data in an information system is extracted and coded from a real life event, frequently through a transaction database (i.e., a patient management database). Which data is captured and its associated meaning (data element definition) is a choice made by the initial system designer based on end-user input and then controlled by the system manager and the functional area manager in which the system is used. Managerial control and user practice may not agree with initial system design and may change over time. The data captured is a focused and limited view of the real life event, as seen through the lens of the ontology and context of the work community. This element of the system then becomes a target for analysis of data quality problems related to contextual data quality and results in propositions four and five.

> <u>Proposition 4:</u> The systems have different data elements, definitions, and encoding processes which reflect the context of the workgroup using them and affect contextual data quality.

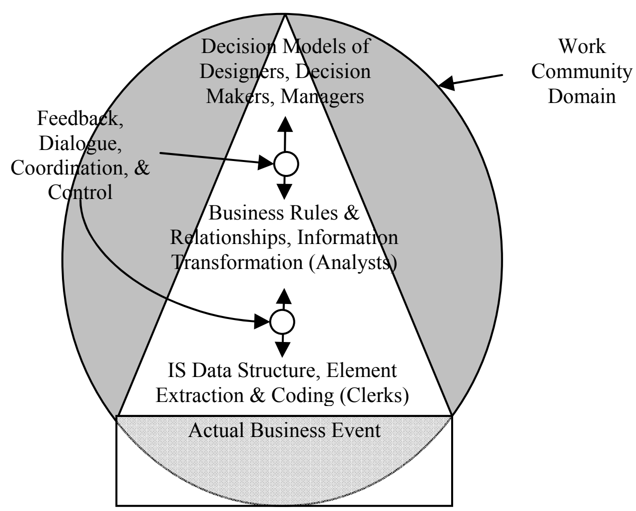
> <u>Proposition 5:</u> Managerial control differs in the two systems and affects contextual data quality.

These data are then often processed and transformed to yield structured information and insight into a particular problem of interest to a decision-maker, manager, or user (van Lohuizen, 1986, Holsapple & Whinston, 1996). The data elements, rules and relationships used to produce this new information are specific to the decision-maker's interests (Koustoukis, Mitra and Lucas, 1999, Davenport and Sena, 2003), again reflecting workgroup context, and further impacting contextual data quality. This results in proposition six for this case study.

Proposition 6: The systems have different data transformations which reflect workgroup context and affect contextual data quality.

Data extraction, managerial control, and data transformation all result in a particular focused and limited view of a real business event of interest to the system managers and users. This focusing of IS fitness for use, based on the designed and implemented business ontology, is represented by the triangular shape of the IS represented in figure 2.3. In Figure 2.3, the shaded oval represents the business domain of interest to the workgroup. The database structure and the business rules used to transform the data in the IS are the relationships; the data elements are the concepts of the work community's ontology.

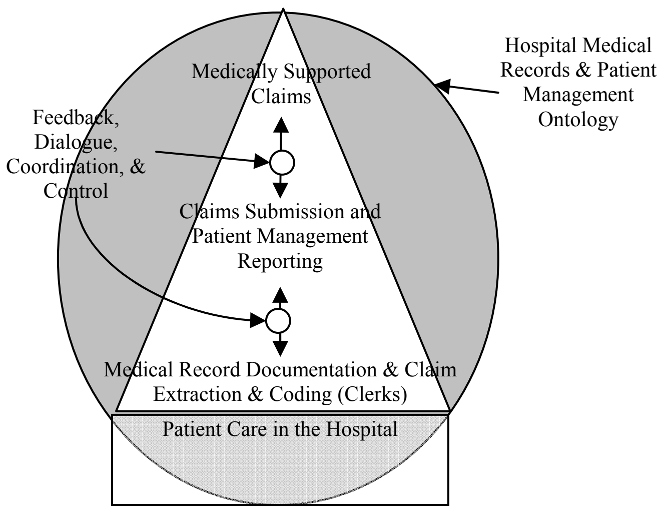
Figure 2.3 Information Systems Reflect Workgroup Ontologies



Users external to the workgroup context with different business ontologies but having an interest in the same real business event will experience a loss of contextual data quality simply due to their differing frame of reference or business context.

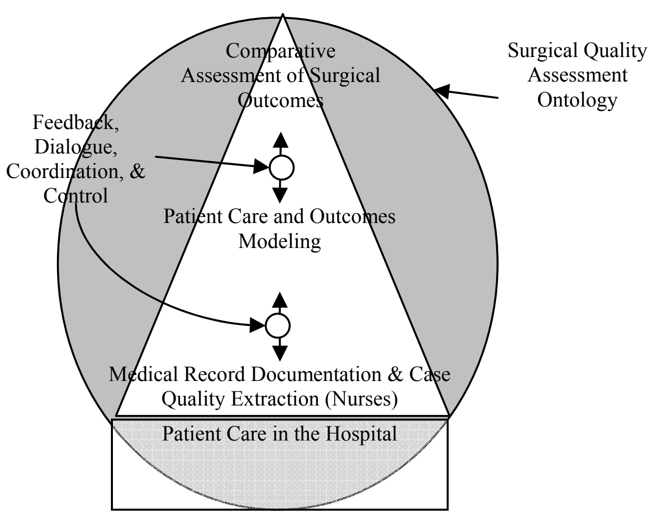
In this case study, the UHCCD pulls most of its data from the hospital claims and patient administration databases. The domain of the system is the administrative and financial reporting of the clinical event. The rules and relationships are primarily those of claims accounting with clinical elements. The manager for whom the system is primarily useful is the patient medical records and claims managers who must submit financial claims and provide audit substantiation of those claims to outside regulators. This system ontology is represented in Figure 2.4. Therefore, while the UHCCD contains much clinical data, it is data that was not originally coded and transformed for risk adjusting surgical outcomes.

Figure 2.4 A Hospital Claims and Medical Records Ontology



In contrast, the NSQIP database was designed specifically for the task of measuring factors that may influence surgical outcomes as well as measuring those outcomes. The NSQIP ontology is shown in Figure 2.5.

Figure 2.5 The NSQIP Surgical Risk Adjustment Ontology

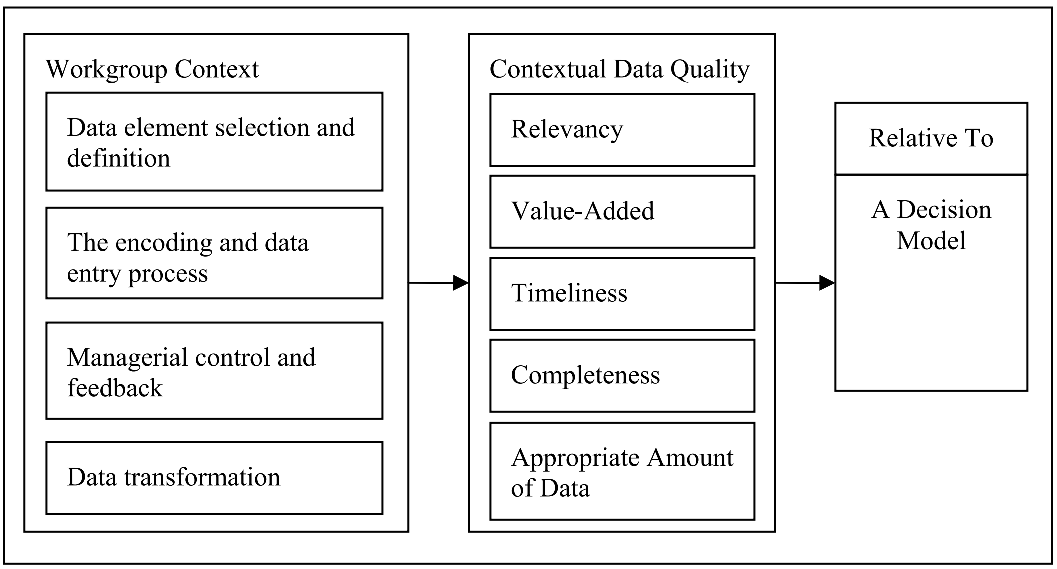


From an ontology view of information systems, the UHCCD data is expected to have less contextual data quality for the risk adjustment of surgical outcomes compared to the NSQIP.

2.8 Proposition Summary

The propositions stated above are summarized in Figure 2.6. Workgroup context influences data selection and definition, the data entry and encoding process, managerial control and feedback, and data transformation. These in turn influence contextual data quality relative to a particular decision model.

Figure 2.6 The IS Contextual Data Quality Model



Proposition 1: The systems are not equally effective in risk-adjusting surgical outcomes.

Proposition 2: Differences in the two information systems effectiveness in risk-adjusting surgical outcomes are not due to system failure, but to differences in the workgroup context in which the data is derived.

Proposition 3: Differences in the two information systems' effectiveness in risk-adjusting surgical outcomes are due to differences in their contextual data quality dimensions of added value, relevancy, timeliness, completeness, and appropriateness of the amount of the data in the systems.

Proposition 4: The systems have different data elements, definitions, and encoding processes which reflect the context of the workgroup using them and affect contextual data quality.

Proposition 5: Managerial control differs in the two systems and affects contextual data quality.

Proposition 6: The systems have different data transformations which reflect workgroup context and affect contextual data quality.

Chapter 3: Methods

3.1 Study Design

This study is a cross-sectional retrospective review of archival quality data gathered on major surgical patients as part of normal activities at the University of Kentucky Hospital. The quality data are numerous variables that measure the preoperative risk of poor outcomes and severity of illness along with the patient outcomes of mortality, complication, length of stay and costs. These data are taken from the National Surgery Quality Improvement Program (NSQIP) database and the University HealthSystem Consortium Clinical Database (UHCCD). The study also includes interviews of the system process owners of the various workgroups involved with a qualitative analysis of their responses.

3.2 Site Selection

The University of Kentucky Hospital (UKH) provides unique characteristics for the study of surgical outcome risk adjustment. It is a pilot site for expansion of the NSQIP from the Veteran's Administration Hospitals into the private sector starting in 1999-2000. As of 2004, there were about 20 non-VA hospitals nationwide using the NSQIP which has recently been targeted for national expansion by the American College of Surgeons. As an academic medical center, it also participates in the UHCCD and therefore is one of the few sites where a direct comparison of the two databases is possible. For the duration of the data collection, the primary investigator was employed at UKH making this a convenience sample that allows for detailed ongoing access to the systems and background information necessary for an in-depth case study (Yin, 1984/2003).

3.3 Study Population

The NSQIP database contains data on a random sample of surgical patients at UKH who underwent major surgery between October 1, 2001 and September 30, 2004. The population is further limited to patients 17 years old or greater on the General, Vascular, Neurological, Orthopaedic, Plastic and Thoracic Surgical services who received general, spinal, or epidural anesthesia. No specific distribution was sought with regard to age, gender, ethnicity, or race. No populations specifically identified as vulnerable were studied, although 17 year-old patients were included because of their inclusion in the NSQIP database. The UHCCD has information on all inpatients. This further limits the comparison sample by excluding patients who had surgery, but were discharged the same day and were not admitted to the hospital.

3.4 Subject Recruitment Methods:

All major surgery patients (in- or outpatient) were eligible for inclusion in the NSQIP database and selection occurred in the following manner according to the NSQIP protocol. Patients were randomly selected from the operating room schedule beginning the first day of an eight-day cycle established by NSQIP. The first consecutive 70 patients per eight-day cycle who met the NSQIP criteria for major surgery – those receiving general, spinal, or epidural anesthesia – were eligible for the program. This sampling methodology has historically resulted in an approximately 33% sample of the procedures on the services tracked. The inpatients were expected to have corresponding data in the UHCCD. Clerical and data entry error in the two systems were expected to create a small percentage of cases that would be unable to be matched. Individuals chosen for interviews were the managers and supervisors directly responsible for the data entry personnel for the two systems.

3.5 Informed Consent Process

Due to the large number of subjects included in this study (approximately 15,000 patients), the difficulties in locating many of them, and the minimal risk presented to the study participants, a waiver of informed consent was requested and granted by the Medical Internal Review Board of the University of Kentucky.

3.6 Interviews for Process Descriptions

The process owners for the data encoding and entry were interviewed in order to understand process, purpose and constraints related to the data. In each case, initial discussions led to a draft of a process description which was then reconfirmed with the interviewees. During the course of the analysis, particular questions regarding how purpose or process might impact interpretation of the results were referred back to these managers via phone and email.

Chapter 4: Measurements and Analysis

4.1 Qualitative Assessment of Workgroup Context

The process descriptions from the interviews and available documentation were reviewed for elements that appeared by inspection to derive from the work group context. Consideration was given to those that impacted data definition and selection, the data entry process, managerial feedback and control, and data transformation. These are treated as the precursors to differences in the two systems' contextual data quality.

4.2 Data Comparison and Contrast

How the two systems included cases into the model was explored through an analysis of the process of linking the records in the database. For instance, the NSQIP excludes patients under 17 years of age. The percentage of surgical patients excluded was calculated for this exclusion criterion and for all the others. Additionally, where variables from the two systems appeared by definition and description to measure the same quantity their levels of agreement across cases were measured and graphically analyzed as Venn diagrams. Where significant disagreement occurred, qualitative elements from the IS process interviews (workgroup context) were explored to explain the differences.

4.3 Data Transformation Analysis

The two systems were analyzed for how they transform the data in modeling quality assessment. Qualitative elements from the IS process interviews along with available literature describing the systems were analyzed to support or reject the proposition that data transformation reflected workgroup context and impacted contextual data quality.

4.4 Measurement of Outcomes

The outcomes of mortality, morbidity, cost and length of stay were measured using data from both systems. Mortality was measured by each of the two systems although the time frame differed slightly. Morbidity was expressly measured in the NSQIP where it is defined as one or more of the specific complications tracked. Potentially preventable complications (PPCs) are listed in the UHCCD but morbidity is not expressly defined. The same definition from the NSQIP, one or more complications, was applied to the UHCCD PPCs to obtain a UHCCD measure of morbidity. Costs are not available in the NSQIP but were available in the UHCCD. In the UHCCD costs are modeled using charges and a cost to charge coefficient. These modeled costs were compared to costs obtained from the hospital cost accounting system and used as the outcome. Finally, length of stay was calculated in the UHCCD but not in the NSQIP, although the NSQIP had admission and discharge dates so the calculation was readily made. Table 4.1 shows the outcomes that are immediately available in the two systems.

Table 4.1 Outcome Model Comparison

Outcome Model	**NSQIP**	**UHCCD**
Mortality	Yes	Yes
Morbidity	Yes	No
Costs	No	Yes
Length of Stay	No	Yes

4.5 Measurement of Contextual Data Quality Dimensions

The work of Wang and Strong in developing contextual data quality dimensions had not yet been implemented in a targeted study, so no prior measurement methodology is available. Given the quality assessment decision model described above, measurement of the five dimensions of contextual data quality related to the individual data elements was performed as follows and represents a new contribution of this study to the literature on data quality:

4.5.1 Relevancy

Relevancy is defined as statistically significant correlation or association between a particular data element and the four outcomes being studied. For interval data and outcomes, Pearson's ρ is used as the statistical test; for binary and interval data, Point Biserial Correlation; for binary and binary data, the Phi statistic; for ordinal and interval or ordinal and ordinal data, Kendall's T_{β}; for nominal and ordinal data Cramer's V; and for nominal and interval data ANOVA and Eta are performed.

4.5.2 Completeness

Completeness is defined as the strength of regression models of the combined variables from each system in predicting the four outcomes being studied. For the binary outcomes of morbidity and mortality, the c-index (equivalent to the area under the receiver operating characteristic curve, Harrell et al., 1984) is used, for cost and length of stay, adjusted R^2 is used. For mortality and morbidity, the receiver operating characteristic (ROC) curve is graphed for each model's estimates versus actual occurrence. The case estimates for each system are also summed across deciles of risk. The observed mortality rate of the patients within the individual deciles is graphically compared for each system.

The contextual data quality perspective assesses whether there is sufficient information in terms of breadth and depth of the domain coverage necessary for confident decision-making. This implementation of completeness differs from the normal IS usage, which is limited to evaluating the amount of missing data due to system and process failure. A system that is designed to provide information that it does not (due to failure) could lead to the contextual incompleteness, but in the case being studied, is not theorized to do so. This study focuses on the breadth and depth of domain coverage from the perspective of the decision model.

4.5.3 Value Added

Value-added is defined as the contribution of the individual data elements to predictive power in the multivariate models of the four outcomes being studied. In the linear regressions this is measured by the standardized coefficients. In the logistic regressions a ranking of the variables is calculated taking the coefficients multiplied by the standard deviation of the variables (Garson, 2005). Here, the logistic regression ranking values are referred to as ranking coefficients.

4.5.4 Timeliness

Timeliness is defined as the availability of the final transformed data to the decision maker within a quality improvement context. However, there are no direct time elements of the decision model used in this study with which to evaluate timeliness. Therefore timeliness is discussed qualitatively for the two systems but not measured quantitatively or compared.

4.5.5 Appropriate Amount of Data

The appropriate amount of data is measured as the statistical power based on sample size of the resulting models. Increasing the amount of data available for a decision model increases the confidence of the resulting decisions. Linking the records for direct comparison, however, results in exclusion of significant portions of both datasets. This limits the ability to directly compare statistical power and therefore the amount of data is discussed qualitatively, but not measured or compared.

Chapter 5: Results

5.1 UHCCD Data Extraction and Encoding Process

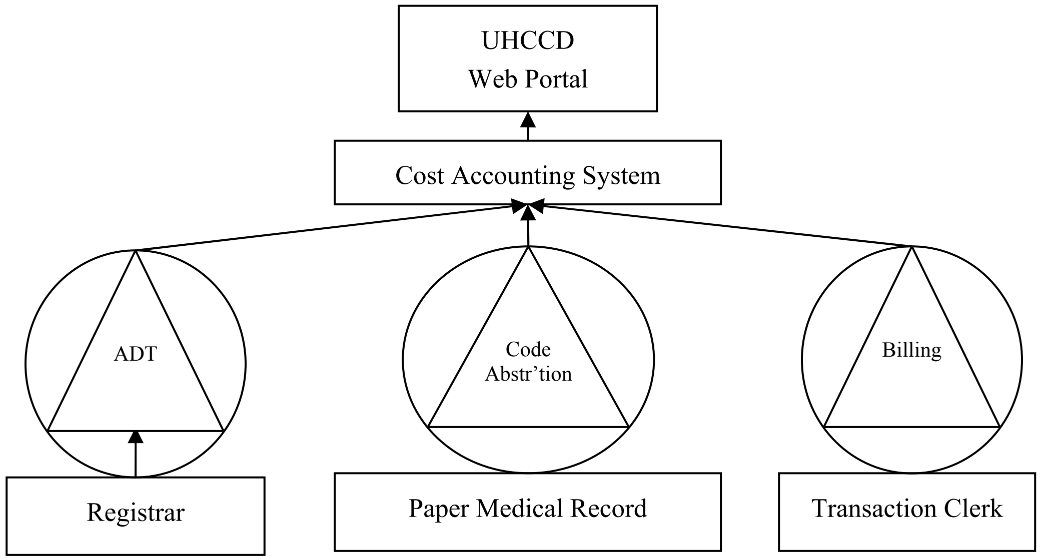
At UKMC, similarly to most hospitals, certified coders who are members of the medical records department review the medical record of virtually every patient after discharge and determine a single primary, and possibly multiple secondary, diagnoses (up to 16 in total are transmitted to the UHCCD). The primary diagnosis coded reflects the reason for admission to the hospital, not necessarily the eventual most acute condition of the patient. For example, a patient who is admitted for treatment of a urinary tract infection but is found to have cancer during the course of his/her stay will have the urinary tract infection listed as the primary diagnosis. A principal and possibly multiple secondary surgical procedures (up to 15 in total are transmitted to the UHCCD) are also coded.

These diagnoses and procedures are encoded using the International Classification of Disease-9th Revision Clinical Modification (ICD-9CM) coding system and entered into the SoftMed® (SoftMed Systems, Inc., Silver Spring, Maryland) abstracting system. There are no dates associated with the secondary diagnosis codes so exact sequencing of conditions and procedures is not possible from the coded data. The physician identifier associated with each procedure and the attesting physician identifier is also recorded along with the admitting service code. Fraud legislation requires a physician from the service who was primarily responsible for the management of the patient during the greatest part of their in-hospital stay attest to the veracity of the billing coding. This physician is encoded as the attesting physician. Additionally, avoidance of billing fraud leads to strict criteria about the documentation required to support a diagnosis code. In general, a documented statement by a physician, using a specific vocabulary, must exist to encode a diagnosis. This holds regardless of whether other medical record evidence in the form documentation from nurses or test results would contradict the diagnosis or would suggest additional diagnoses.

The medical records department then uses a vendor-provided software package (3M Coding and Reimbursement System) that classifies the patient into one of over 500 diagnostic related groups (DRGs). In most cases, the primary diagnosis and procedure determine the DRG. In some cases, secondary diagnoses that are considered comorbid conditions or complications (CCs) of the primary diagnosis and procedure change the DRG. These changes may increase reimbursement. The DRG codes are used by Medicare, Medicaid, and Blue-Cross/Blue-Shield, (together greater than 50% of UKMC's business by charge dollars) among other payers to calculate reimbursement for services. Thus the DRG-calculating algorithms are designed to maximize reimbursement. Finally, the medical records coders also extract the discharge status of the patient from the medical record. Discharge status is coded as the destination of patient at discharge or if expired. Discharge to a long term care facility like a nursing home can reduce the reimbursement to the hospital from some payers in some DRGs because the reimbursement is shared with the destination facility. Prior to October 1, 2005, twenty of the 500+ DRGs were affected by discharge status. (Interviews with process owners)

Added to the ICD-9 Codes and DRG are the patient medical record and encounter number, age, gender, race, if transferred from another health facility (and what type of facility), the primary and secondary insurer (or lack thereof), admission date, and discharge date, and status (destination or expired). This information is usually encoded by a registrar/clerk at the time of patient registration and discharge into the hospital admission, discharge, and transfer (ADT) system which is primarily used for patient management in the hospital. The information from the medical records system and the ADT system are then aggregated into the cost accounting system (TSI, Transition Systems International). This system adds charges from the various departments in the hospital and extracts all the data into a file that is uploaded to the UHCCD web portal on a quarterly basis (Interviews with process owners). The system described is summarized in Figure 5.1.

Figure 5.1 UHCCD System Architecture



5.2 UHCCD Managerial Control and Feedback

Process owners of the internal systems were interviewed regarding the focus of control efforts and monitoring of the performance of the staff encoding and entering data. In each case, the audits performed on a regular basis indicate what elements and processes of data encoding and entry are reinforced through measurement and feedback to the staff. The registration system is audited regularly with a focus on proper identification of the patient and the amount of accurate insurance information captured in support of billing.

The most significant data come from the coding abstraction system. The principal and secondary diagnoses and procedure codes in the system are audited on a yearly basis by federal auditors. A random sample of patient bills is selected and the medical charts reviewed to determine if sufficient physician medical documentation exists to support the submitted bills. The standard of documentation is based on legal judgments that have historically been applied in resolving fraud claims. The body of these rulings is discussed and applied by the national association for coders. However, no official complete manual of definitions exists regarding documentation sufficiency related to all ICD-9 codes. Given the potential for fraud litigation, coders are under the dual pressures to code sufficiently to maximize billing, but also conservatively in regards to sufficient physician documentation to justify the billing.

5.3 UHCCD Data Transformation Process

The UHCCD calculates three subsets of data based on the primary and secondary diagnoses and procedures coded. These are Comorbid Conditions, Potentially Preventable Complications, and All Patient Refined Diagnosis Related Groups (APR-DRGs). The APR-DRGs also have a secondary variable calculated called the Severity of Illness index. The comorbid conditions are diagnoses that are considered likely to be chronic conditions present in the patient at admission and represent an increase in risk of poorer outcomes or higher severity of illness in the patient. Some secondary diagnoses could be present at admission or develop in the hospital during the course of treatment and represent complications of the care delivered. These diagnoses are not considered for selection as comorbid conditions. The logic for the selection is based on work by Iezzoni et al. (1994 Oct).

Conversely, the potentially preventable complications are the secondary diagnosis codes that are considered unlikely to have been present at admission and more likely to have developed or been caused during the course of treatment. The algorithms making the selection were first developed by Iezzoni et al. (1994 Jul) under funding from the United States Agency for Healthcare Research and Quality. The UHCCD uses these PPCs to screen out patients in the cost and length of stay modeling process whose outcomes may have been caused partially by errors or poor quality care. Patients with PPCs are not included in the population from which the models are derived. The UHCCD does not develop models for risk adjusting morbidity from the PPCs, but does provide them to their sites with designations as complications (UHC, 2005).

The APR-DRGs are a regrouping and redefining of the DRGs with a focus on clinical severity and risk of mortality. The groupings are performed by proprietary software owned and licensed by 3M Corp. The program regroups the primary diagnoses and procedure codes into a smaller number of subgroups. For each DRG, it then looks at the secondary ICD-9 diagnoses and utilizes an 18 step algorithm separated into three phases to determine a Severity of Illness (SOI). This measure has four possible values which are minor, moderate, major, and extreme (Averill et al., 2003). A second variable, Risk of Mortality, is also calculated by the APR-DRG grouper software but is not used in the UHCCD to model mortality. (UHC, 2005) The APR-DRG SOI algorithm is quite complex, proprietary, and unique for each of the APR-DRGs. In general, specific relationships between the secondary diagnoses/procedures and the primary diagnosis and procedure which indicate more severe illness result in a higher SOI. Interaction between multiple secondary diagnoses, in particular from different organ and disease groups, increases SOI. Additionally, combinations of specific primary and secondary procedures result in a higher SOI.

One of the steps of note in the algorithms is that mechanical ventilation is considered a secondary procedure and prolonged mechanical ventilation increases SOI for some APR-DRGs. It is unclear from the available literature regarding the algorithms how other minor procedures performed in the ICU or at the bedside would impact SOI but may similarly increase the SOI level. Thus SOI may be increased by the presence of particular secondary diagnoses related to the primary diagnosis if it indicates increased severity and by minor procedures such as ventilation. These may be considered complications in the NSQIP dataset and represent a major challenge to the direct comparison of the two datasets in assessing outcomes. If the SOI includes significant information regarding what are considered complications, it will provide positive bias in the estimation of morbidity and in morbidity-related length of stay and costs. This is discussed in greater depth in the discussion section.

Once the comorbid conditions, PPCs, and SOI have been calculated, the UHCCD estimates mortality, costs and length of stay for each DRG using regression models of the same variables listed in Table A.1 in Appendix A. The incidence in the study population of each variable and their strengths of association with the four outcomes are reported. For some acute DRGs, notably cardiac procedures not included in this study, additional variables are used in the modeling process. If sufficient data is not available for effective modeling the mean rates of the outcomes by SOI are used rather than the regression model.

5.4 NSQIP Data Extraction and Encoding Process

In the NSQIP at UKMC, Nurse Reviewers who are registered nurses follow an established protocol for selecting and abstracting information regarding surgical patients. The program provides initial standardized training for the nurses and a user manual. The user manual describes the protocol in detail and it also lists detailed clinical definitions of the preoperative, intra-operative, and postoperative variables to be encoded by the nurse reviewer. The nurses randomized patient selection at UKMC by taking the first 70 major surgery patients on six services from the operating room schedule every eight days that matched the inclusion criteria. Using an eight-day cycle ensured a different daily operating room schedule was included as the majority of cases in consecutive cycles which randomized the service representation in the data.

The operating room (O.R.) scheduling information is transmitted electronically every eight days as an attached ASCII report. Included are the patient's name, gender, date of birth, and registration number, the date of surgery, the type of surgery (elective, urgent, emergent), the primary service performing the surgery, the anesthesia type, a list of up to six surgeon hospital identifiers, a list of up to 15 Current Procedural Terminology (CPT©, American Medical Association) procedure codes, and the time the patient entered the operating room. Procedures are normally CPT coded by the Attending Surgeon in the O.R. CPT codes exist for all physician services and are different from the ICD-9CM codes used for hospital coding of procedures. A hard copy of the O.R. log, a form documenting this information signed by the attending surgeon attesting to their presence, is also obtained weekly by the nurses.

This file is then read into an access database that automatically performs certain exclusions and leaves the subset for review by the nurses. The program automatically excludes patients under 17 years of age and all non-major surgical cases. “Major” surgery is defined by those procedures having general, epidural, or spinal anesthesia along with some monitored anesthetic (MAC) procedures. Additionally, some procedures that are very low risk as determined by the NSQIP are excluded by their CPT code. The primary diagnosis is listed on the O.R. log as well using ICD-9CD codes. At UKMC, patient data are then excluded for services other than General Surgery, Neurosurgery, Orthopaedics, Plastic Surgery, Thoracic Surgery, and Vascular Surgery.

The nurses obtain lists of admissions due to trauma from the trauma office in the hospital. They then exclude any cases performed during those admissions. They examine the various CPT codes to ensure that a case is not excluded when it has an excluded CPT code as the primary procedure but a more extensive included CPT code as a secondary procedure. The NSQIP methodology has been described in detail by Fink et al. (2002).

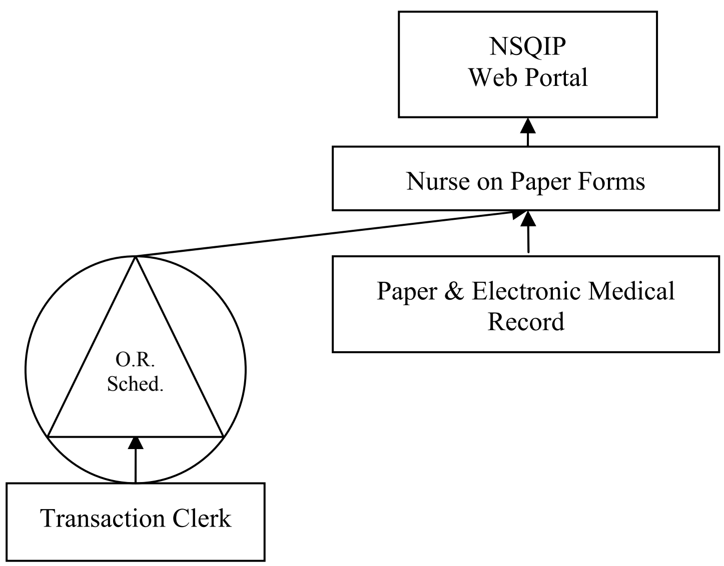
For each patient, the nurse reviewers examine the entire medical record looking for predefined clinical elements including 60 preoperative risk factors, 18 intraoperative factors, and 29 postoperative complications (including death) for 30 days postoperatively. Information after discharge is obtained through hospital and clinic medical document review as well as follow-up contact by letter and phone. These values are entered onto a paper form while being encoded and then directly into a web portal.

5.5 NSQIP Managerial Control and Feedback

The managerial controls related to the NSQIP nurse coordinators’ data encoding and entry consist of general supervision, volume reports, biannual on-line inter-rater reliability testing, conference calls coordinated by the national nurse coordinator, and annual site visits with a chart audit of 20 patients. The volume reports ensure that an adequate number of cases are being tracked by the nurse coordinators. They report the expected and actual number of cases submitted by the nurse coordinators by eight day cycle. The on-line inter-rater reliability testing consists of an email to the nurse reviewers with a sample medical record description of a particular patient. The nurses are then required to abstract the preoperative, intra-operative and postoperative variables for that patient. They submit the results back via email and they are scored and discussed on the conference call involving the other nurse coordinators and the national nurse coordinator. The conference calls also address common problems related to application of the definitions to specific patients, general information to the nurse reviewers regarding the program, and modifications to the protocol and variables as they occur. Finally, the annual site visits consist of an auditor encoding information from 20 surgical cases and then comparing it with the information submitted by the nurse reviewer. The auditor reviews the discrepancies with the nurse reviewer and reports the findings to the site program director.

The system diagram for the NSQIP data is shown in Figure 5.2.

Figure 5.2 NSQIP System Architecture



The resulting data variables in the NSQIP are shown in Table A.2 in Appendix A along with their correlations or strengths of association with the four outcomes.

5.6 NSQIP Data Transformation Process

The NSQIP dichotomize many of the lab variables into high or low values for purposes of modeling outcomes. The cutoff points are based on clinically accepted values. For each case, the CPT codes are used to calculate the Work Relative Value Units (WRVUs). WRVUs are a measure of the physician work jointly determined by the American Medical Association and the Center for Medicare and Medicaid Services (CMS) used for encoding physician procedural services. They are acquired nationally by the analysis center and are available online from the CMS (http://www.cms.hhs.gov/regulations-/pfs/2004/). They are applied to each of the CPT codes submitted for a particular operation and the maximum value is retained as a measure of the complexity of the operation. The preoperative factors and the WRVUs are then entered into separate regression models for each specialty. The preoperative and perioperative data points along with their association or correlation with each of the four outcomes are shown in Table A.2 in Appendix A.

5.7 Comparison of Patient Population

Surgical patients are identified differently in the two databases. In the UHCCD they are identified as any patient having a surgical DRG or an operating room charge to the patient account. In the NSQIP they are determined by the operating room schedule. Additionally the NSQIP excludes patients under 17 years old, procedures performed during an initial admission related to trauma, transplants, and minor procedures. Major operations are defined as procedures requiring general, epidural, spinal and some monitored anesthetic sedation (MAC) cases. Procedures considered minor are excluded via a list of exclusion Current Procedural Terminology (CPT®, ©American Medical Association) codes.

In addition to the definition as a surgical patient, the UHCCD classifies the patients as major or minor surgery based on ICD-9 coding of diagnoses and procedures (APR-DRG,© 3M Corp.). For purposes of a baseline to compare the domain coverage of the two databases, the counts of surgical patients on the included services from the operating room log are used. They are shown in Table 5.2 for the three years from October 1, 2001 to September 30, 2004. In total the Operating Room Scheduling System listed 26,322 cases on the included services over the time period. NSQIP had 9,742 cases or 37% and the UHCCD had

15,456 surgical admissions or 59% if each admission only had one case. This is not true for patients who return to the operating room during the same admission, but gives an approximation of the coverage in the database.

Table 5.2 A comparison of Counts of Surgical Patients by Service from the Operating Room Log, the NSQIP, and the UHCCD.

Specialty	**O.R. Log**	**NSQIP**	**UHCCD**
Gastroint. Surgery	2,274	[a]	829
Oncologic Surgery	2,672	[a]	738
Trauma	2,024	[a]	3,097
General Surgery Total	*6,970*	*3,329*	*4,664*
Neurosurgery	4,353	1,886	4,138
Orthopaedics	10,024	2,809	3,176
Plastic Surgery	3,008	738	920
General Thoracic	375	396	1,672[b]
Vascular Surgery	1,592	584	886
All Specialties	**26,322**	**9,742**	**15,456**

[a] NSQIP includes Gastrointestinal Surgery, Oncologic Surgery and Trauma Surgery services within General Surgery.
[b] UHCCD Cardiothoracic Service which includes cardiac cases in addition to General Thoracic cases.

5.8 Difference in Domain Coverage

The process of linking the two databases by medical record number and date of admission reveals the differences in domain coverage between the two datasets. The linking results in 4,618 operative cases occurring during 4,283 patient admissions. 397 cases were within 30 days of a prior case so are excluded to avoid ambiguity in assignment of outcomes. These deleted cases include 62 patient admissions where the case is secondary to a prior admission's case resulting in 4,221 unique case/admissions. The incongruence between the two databases and the total cases from the O.R. log results from:

a) the partial sampling methodology of the NSQIP, (24.0% of O.R. Log Cases Excluded)

b) the exclusion of minor and some low-risk procedures by the NSQIP, (15.1%)

c) the exclusion of Trauma cases by the NSQIP, (17.5%)

d) the exclusion of Minors (<17 yrs old) by the NSQIP, (6.8%)

e) the exclusion of secondary cases within 30 days by the NSQIP, (0.8%)

f) the exclusion of outpatient procedures by the UHCCD. (Estimated per NSQIP database at 54.8%)

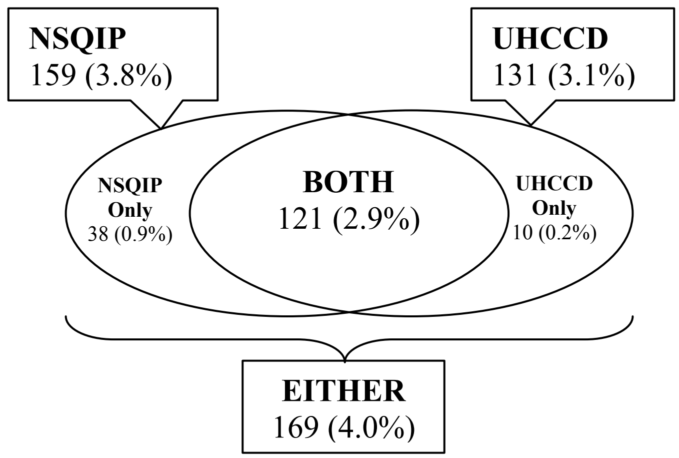
Total exclusions from NSQIP are 63.3% resulting in a 36.7% sampling rate of all cases performed on these services; exclusions from the UHCCD are estimated at 54.8% resulting in a 45.2% sampling. If the NSQIP and the UHCCD inclusion criteria were completely independent we would expect to have had 16.6% (36.7% x 45.2%) overlapping inclusion of the O.R. log cases. The data linked for 4,618 out of 26,322 (17.5%) of the O.R. log cases; 0.9% more than expected from independent exclusions. Because the exclusion of minor cases by the NSQIP would have correlated with more outpatient procedures, the exclusions are not completely independent and the slightly higher than predicted number of matches is reasonable.

5.9 Outcome Variable Definitions

5.9.1 Definition of Death in the Two Databases

The NSQIP defines death as any death occurring within 30 days of the surgery regardless of its potential relationship to the surgery or whether the patient is still in the hospital. The UHCCD, by contrast defines death as any death during the admission. Thus an in-hospital death may occur after 30 days and be included in the UHCCD but not in the NSQIP and a death may occur after discharge within 30 days of the procedure and be included in the NSQIP but not in the UHCCD. The overlap of the two definitions is shown by the Venn diagram in Figure 5.3. The NSQIP tracks the date of death if known after 30 days and this information is used to determine the cause of incongruence. In all cases, it is due to the definitional differences above. There is disagreement between death rates depending on whether death is defined narrowly as meeting both criteria (2.9% mortality rate), or broadly meeting either criteria (4.0% mortality rate). For the purposes of analysis, death in either database will be used.

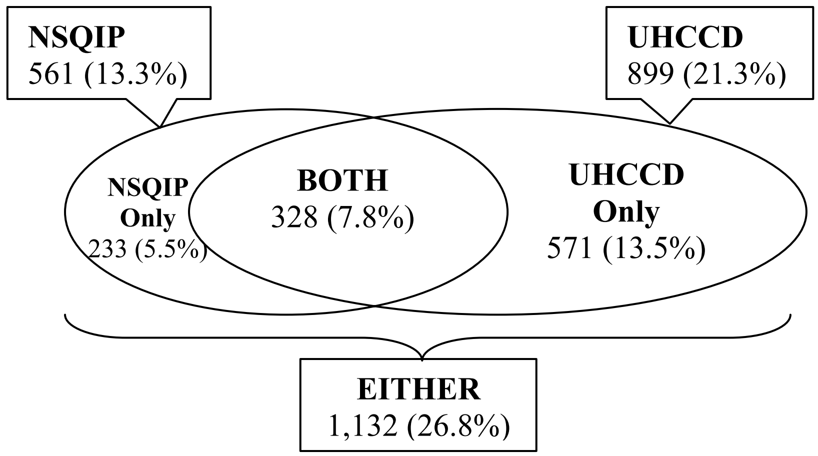
Figure 5.3 Overlap of Mortality: Occurrences and Rates Resulting from Differing Definitions



5.9.2 Definition of Morbidity in the Two Databases

Morbidity is defined in this study as a patient having one or more identified complications. The two systems differ in the number and types of complications identified. Unlike mortality, where the differences are relatively small in percentage and readily justifiable by the differing time periods, the differences in morbidity in the two databases are much greater and less justifiable. The differences in identified groups of morbid patients are shown in Figure 5.4.

Figure 5.4 Overlap of Morbidity: Occurrences and Rates Resulting from Differing Definitions



The two systems disagree, more than agree, on the labeling of patients who experienced complications. The UHCCD includes almost twice as many patients in this category as the NSQIP. Of the patients with complications listed only in the UHCCD, 346 out of 571 (61%) are in the non-specific categories of *Miscellaneous Complications*, or *Other Complications of Procedures*. The number of complications experienced by a patient recorded by the two system had a Pearson's correlation of 0.481 (P <0.001). Because of the number of different possible complications in both databases, a comparison by individual complications is shown in Table 5.3.

Table 5.3 Complication Comparisons

NSQIP Complication Description	NSQIP Only N (%)	Both N (%) [% Agreement]	UHCCD Only N (%)	UHCCD Potentially Preventable Complication Description
Any Complication	233 (5.5)	328 (7.8) [28.9]	571 (13.5)	Any Complication
Cardiac				
Cardiac Arrest	32 (0.8)	3 (0.1) [5.0]	24 (0.6)	Shock / Cardiorespiratory Arrest
Acute Myocardial Infarction	20 (0.5)	4 (0.1) [11.4]	11 (0.3)	PO Acute Myocardial Infarction
			4 (0.1)	PO Cardiac Abnormality Except AMI
Central Nervous System				
Stroke / CVA	19 (0.5)	0 (0.0) [0.0]	2 (0.0)	PO Stroke
Coma > 24 Hrs.	4 (0.1)	0 (0.0) [0.0]	3 (0.1)	PO Coma or Stupor
Peripheral Nerve Injury	1 (0.0)			
			7 (0.2)	Central or Peripheral Nervous System
Respiratory				
Pneumonia	138 (3.3)	26 (0.6) [13.4]	30 (0.7)	PO Pneumonia
Pulmonary Embolism	9 (0.2)	9 (0.2) [9.9]	73 (1.7)	Venous Thrombosis/ Pulm. Embol.
Unplanned Intubation	43 (1.0)	53 (1.3) [19.5]	176 (4.2)	PO Pulmonary Compromise
On Ventilator > 48 Hrs.	55 (1.3)	102 (2.4) [35.9]	127 (3.0)	PO Pulmonary Compromise
Unpl. Intub. Or On Vent. > 48 Hrs.	89 (2.1)	126 (3.0) [39.6]	103 (2.4)	PO Pulmonary Compromise
			20 (0.5)	Aspiration Pneumonia

Table 5.3 Complication Comparisons - Continued

NSQIP Complication Description	NSQIP Only N (%)	Both N (%) [% Agreement]	UHCCD Only N (%)	UHCCD Potentially Preventable Complication Description
Urinary Tract				
Progressive Renal Insufficiency	11 (0.3)			
Acute Renal Failure	23 (0.5)			
Urinary Tract Infection	121 (2.9)	0 (0.0) [0.0]	7 (0.2)	PO Urinary Tract Complication
Wound				
Superficial Surgical Site Infection	99 (2.3)	17 (0.4) [9.6]	79 (1.9)	Wound Infections
Deep Surgical Site Infection	19 (0.5)	4 (0.1) [3.6]	92 (2.2)	Wound Infections
Organ / Space SSI	13 (0.3)	2 (0.0) [1.9]	94 (2.2)	Wound Infections
Wound Dehiscence	85 (2.0)	17 (0.4) [10.4]	79 (1.9)	Wound Infections
Either of Four Wound Infections	146 (3.5)	29 (0.7) [11.9]	67 (1.6)	Wound Infections
			33 (0.8)	Cellulitis / Decubitus Ulcer
Other				
Graft/ Prosthesis/ Flap Failure	9 (0.2)	8 (0.2) [5.1]	139 (3.3)	Mech. Compl. Due to Device / Implant
DVT/ Thrombophlebitis	17 (0.4)	6 (0.1) [6.0]	76 (1.8)	Venous Thrombosis/ Pulm. Embol.
Systemic Sepsis	63 (1.5)	2 (0.0) [2.7]	9 (0.2)	Septicemia
Septic Shock				
Bleeding/ Transfusion	11 (0.3)	9 (0.2) [4.9]	165 (3.9)	PO Hemorrhage/Hematoma

Table 5.3 Complication Comparison - Continued

NSQIP Complication Description	NSQIP Only N (%)	Both N (%) [% Agreement]	UHCCD Only N (%)	UHCCD Potentially Preventable Complication Description
			7 (0.2)	PO GI Hemorrhage/Ulceration
			20 (0.5)	PO Infections (Not pneumonia/ wound)
			9 (0.2)	PO Phys./Metabolic Derangements
			92 (2.2)	Proc. Related Perforations or Lacerations
			247 (5.9)	Other Complications of Procedures
Return to O.R.	404 (9.7)	34 (0.8) [7.4]	20 (0.5)	Reopening of Surgical Site
			99 (2.3)	Miscellaneous Complications

The level of agreement between the two databases on similarly described complications is never greater than 40% and is less than 10% in 9 out of 11 complications. Because of this marked disagreement, we perform a detailed review of patient medical records for Postoperative Acute Myocardial Infarction (AMI) and Postoperative Pneumonia. AMI is chosen for review because of its clinical acuity which is theorized to not allow for ambiguity. Pneumonia is chosen because it does occur preoperatively, the clinical definition is more ambiguous, and timing confusion may play a role in the disagreement between databases.

5.9.3 Review of Postoperative Acute Myocardial Infarction and Pneumonia Complications

In the case of AMI, there are 35 instances recorded in either database, with agreement on only four. A nurse reviewed the medical chart and medical information systems regarding each of the 31 incongruous cases in order to verify the data and expose potential causes for the disparity of assessment. For three of the patients the chart was not readily available so no analysis was done. The results of the review of the 28 remaining patient records are shown below in Table 5.4.

Table 5.4 Reasons for Incongruent Coding Between the Two Systems of Postoperative Acute Myocardial Infarction as a Complication

No. of Patients	Reason for Incongruence in Coding
NSQIP-Only Recorded Postoperative AMI	
5 (18%)	Complication occurred after discharge so unavailable to hospital coders.
2 (7%)	Insufficient physician documentation to substantiate hospital coding.
13 (46%)	AMI was ICD-9 coded but not screened as a PPC in UHCCD.
UHCCD-Only Recorded Postoperative AMI	
7 (25%)	An AMI did occur and was hospital coded but did not meet NSQIP definition.
1 (4%)	No documentation of AMI existed; erroneously encoded.
28	**Total Cases Reviewed**

The most common cause of incongruence is the UHCCD complication screener not detecting AMI as a potentially preventable complication. This is likely due to the fact that it was related to the admitting diagnosis or because of the lack of verifiable timing. This is not known however. The second most significant cause of incongruence in AMI complication is due to the strictness of the clinical definition in the NSQIP (Q-wave AMI only) which excludes several of the milder AMIs which the hospital coders recorded. One error in hospital discharge coding was found and in two cases (7%) test results existed that supported the NSQIP AMI but physician documentation was considered poor by the reviewing nurse and therefore probably resulted in the non-coding by the hospital coders.

In the case of Postoperative Pneumonia there are 41 instances recorded in January through September 2004 in either database, with agreement on only eight. For two of the patients the chart was not readily available so no analysis was done. The results of the review of the 31 remaining patient records are shown in Table 5.5.

Table 5.5 Reasons for Incongruent Coding Between the Two Systems of Postoperative Pneumonia as a Complication

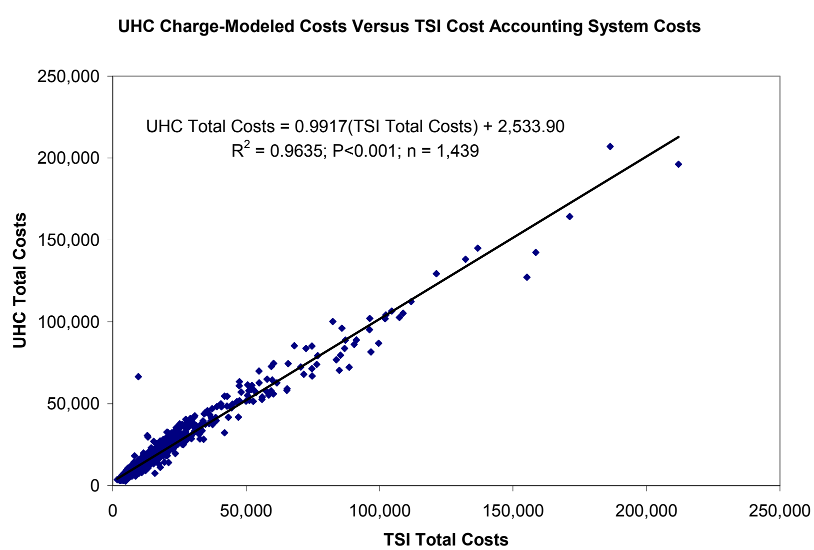
No. of Patients	Reason for Incongruence in Coding
NSQIP-Only Recorded Postoperative Pneumonia	
11 (35%)	Postoperative pneumonia was ICD-9 coded but not screened as a PPC in UHCCD.
4 (13%)	Pneumonia Occurred after discharge so unavailable to coder.
3 (10%)	Pneumonia Coded as Aspirate Pneumonia in UHCCD
2 (6%)	Insufficient physician documentation to substantiate hospital coding.
1 (3%)	Postoperative pneumonia was physician documented, coder missed.
UHCCD-Only Recorded Postoperative Pneumonia	
5 (16%)	Pneumonia occurred preoperatively or preadmission, confused timing.
2 (6%)	Treatment occurred for pneumonia but did not meet NSQIP definition.
2 (6%)	Postoperative pneumonia documented, NSQIP nurse missed.
1 (3%)	No documentation of pneumonia existed; erroneously encoded.
31	**Total Cases Reviewed**

The most common cause of incongruence is again the UHCCD complication screener not detecting Postoperative Pneumonia as a potentially preventable complication. The second most significant cause of incongruence, unlike AMI, is due to confusion regarding the timing of when the pneumonia occurred. The UHCCD incorrectly screened five (16%) preoperative pneumonias as postoperative pneumonias. The remaining incongruencies follow those of the AMI.

5.9.4 Definition of Costs

Inpatient hospital costs (hereafter costs) are modeled in the UHCCD based on charges and are not available in the NSQIP. Total costs from the hospital cost accounting system (TSI) are available from a prior study for the General and Vascular patients. When these were regressed against the modeled costs in the UHCCD, correlation is excellent at 0.99 (P<0.001). However, a scale increase of \$2,534 in the UHCCD modeled costs versus the cost accounting system costs is noted. The regression line and formula are shown in Figure 5.5. Based on the strong correlation, the UHCCD modeled costs are used in the analysis.

Figure 5.5 Comparison of UHCCD Total Costs and TSI Total Costs for 1,439 General and Vascular Surgery Patients



5.9.5 Definition of Length of Stay

Length of stay is available in the UHCCD dataset and is one of the modeled outcomes. It is not available directly in the NSQIP but readily calculated from the admit and discharge dates that are available. The correlation between the UHCCD and NSQIP lengths of stay is 0.97 (P<0.001). Because the UHCCD data is uploaded directly from the ADT system in the hospital whereas the NSQIP requires reentry by the nurses, it is deemed likely that the small differences are due to data entry error in the NSQIP dataset and the UHCCD LOS is used in the models.

5.10 Relevancy

Out of 40 variables tracked or calculated in the UHCCD and used in modeling, thirteen have no significant association with any of the four outcomes and are therefore considered irrelevant to the assessment of surgical quality leaving twenty-seven variables that are considered relevant. Fifteen variables are relevant to mortality, 16 to morbidity, 24 to length of stay, and 25 to costs.

Out of sixty four variables tracked and calculated by the NSQIP, three have no significant association with any of the four outcomes and are therefore considered irrelevant to the assessment of surgical quality leaving sixty one variables that are deemed relevant. Forty-eight of these are relevant to mortality, 45 to morbidity, 51 to length of stay, and 54 to costs.

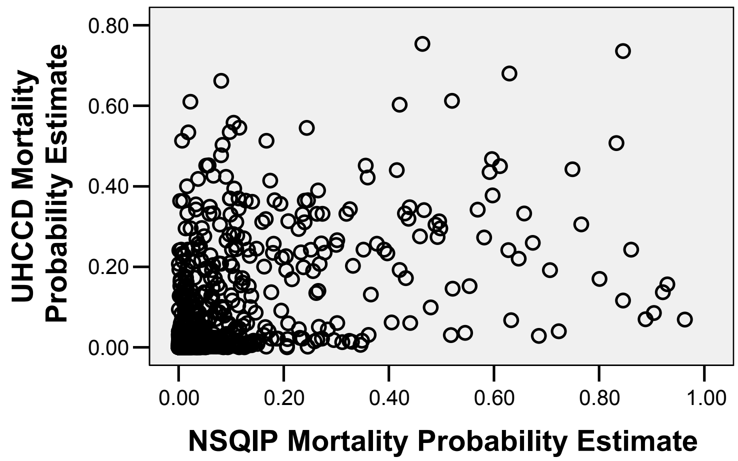
5.11 Completeness

Completeness is measured by the two datasets' total explanatory power relative to the four outcomes. This is measured by the C-indices resulting from the morbidity and mortality logistic regression models and the adjusted R^2 values from the costs and length of stay linear regression models.

5.11.1 Evaluation of Explanatory Power for Mortality

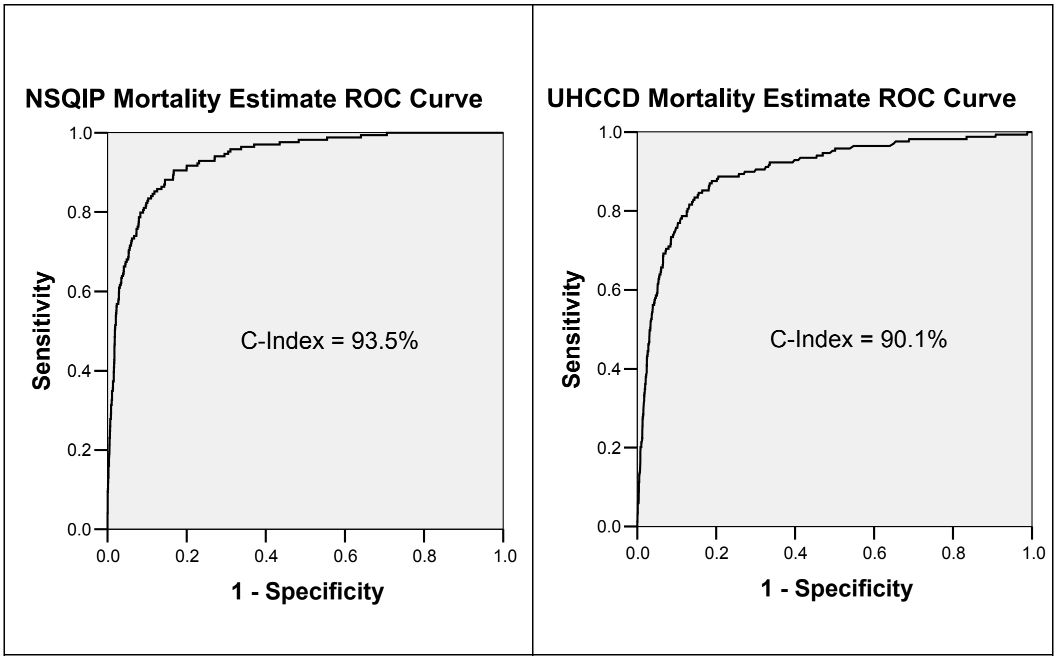
Estimates of mortality between the two models have a Pearson's correlation of 0.543 (P<0.001). Substantial agreement occurs between the two models at the low estimate level which includes most patients. The scatter plot (Figure 5.6) of the estimates shows greater dispersion however as the estimates increased.

Figure 5.6 Scatter Plot of UHCCD and NSQIP Mortality Probability Estimates



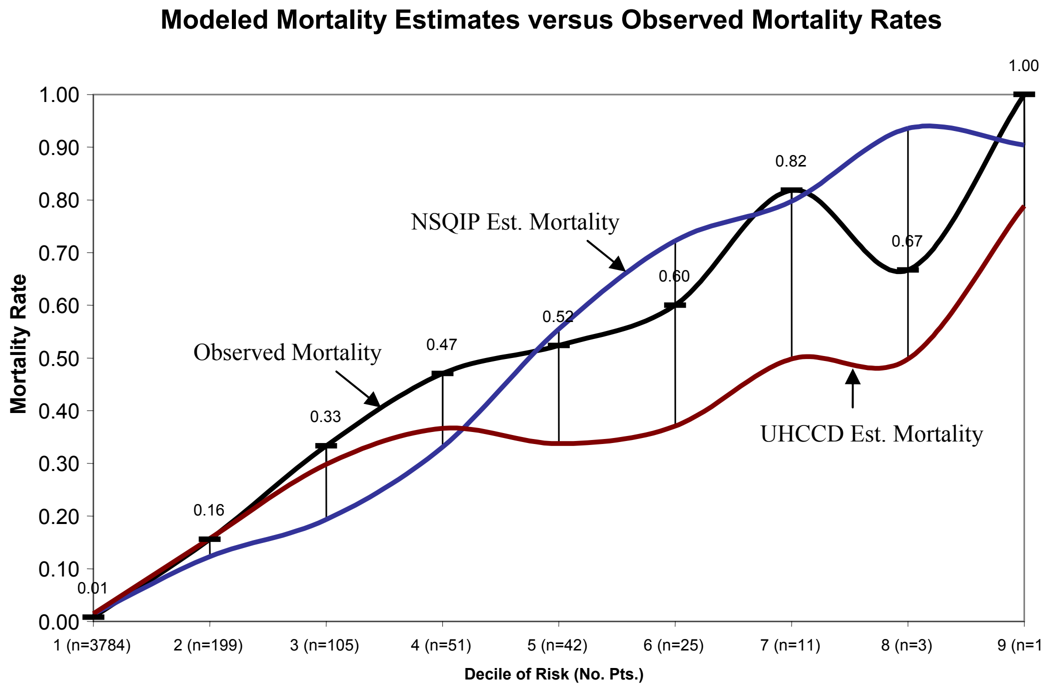
The c-indices for the two models in predicting mortality are shown in Figure 5.7 as the area under their respective ROC curves. The 3.4% increase in the c-index for NSQIP over the UHCCD is statistically significant (p=0.012) using the Hanley-McNeil method (Hanley & McNeil, 1983). Both models have "outstanding" calibration in predicting death as measured by a c-index greater than 0.90 (Hosmer and Lemeshow, 2000).

Figure 5.7 ROC Curves for Mortality Estimates (All Services)



Of note however, is the difference in estimation between the two models for high versus low risk patients. Figure 5.8 graphs the observed mortality rate along with the UHCCD and NSQIP estimated rates by decile of risk. A patient's decile of risk is determined by taking the average of the UHCCD and NSQIP estimates. In the 20% to 30% risk decile, the UHCCD estimated rate is closer to observed than the NSQIP, but in the greater than 50% deciles the NSQIP estimates match observed rates more closely. Both estimates are close to accurate in the lower than 10% risk patients, which includes the majority of the patients in the study.

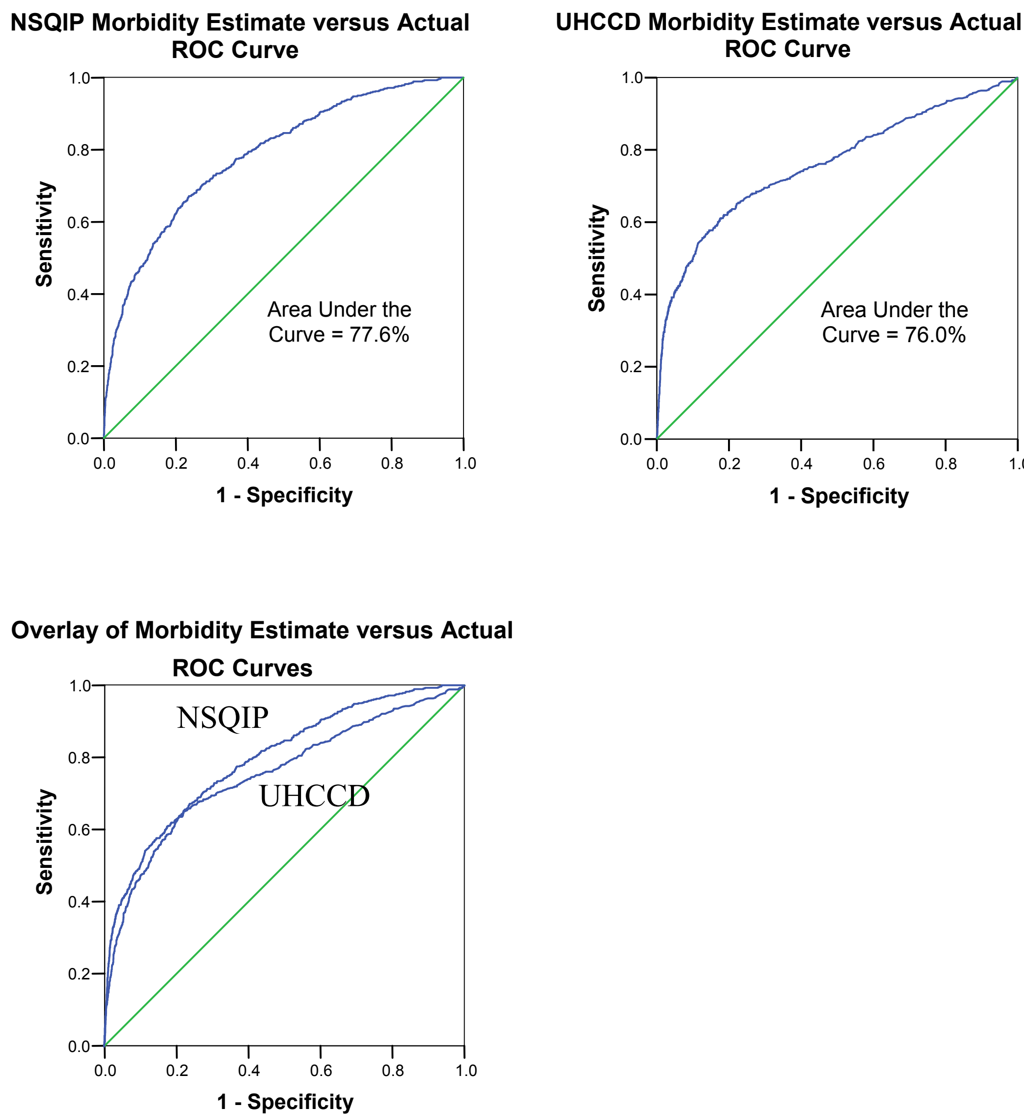
Figure 5.8 Observed Mortality Rates versus NSQIP and UHCCD Estimates by Decile of Risk



5.11.2 Evaluation of Explanatory Power for Morbidity

For purposes of modeling morbidity using logistic regression of the two datasets a decision must be made regarding identification of the morbid patients, given the marked incongruence between the two databases. Because of the number of non-specific "miscellaneous" PPCs coded in the UHCCD, the number of unidentified AMIs, and the wrongly timed pneumonias in the UHCCD from the chart review, the NSQIP morbid designation is chosen. Both datasets yield statistically significant models (P<0.001) which show "good" calibration in predicting morbidity as defined by a C-index greater than 0.70. The ROC curves for the estimates resulting from the logistic regressions of the two datasets versus morbidity are shown in Figure 5.9. The increase of the NSQIP over the UHCCD was not significant using the Hanley and McNeil method (p=0.07). The two curves are overlaid in the third panel and similar to the mortality results that were graphed, the sensitivity is better for the UHCCD estimates at the lower risks but better for the NSQIP at the higher risks.

Figure 5.9 ROC Curves for the NSQIP and UHCCD Morbidity Estimates



5.11.3 Evaluation of Explanatory Power for Costs and Length of Stay

The variation in costs and length of stay explained by the two datasets is shown in Table 5.6 as the adjusted R^2 values from the regression models. The

UHCCD data explain more of the variation in both outcomes than the NSQIP. The full results from the regressions are in Appendix B.

Table 5.6 Costs and Length of Stay Variation Explained by the Two Datasets

Outcome	NSQIP Adjusted R2	UHCCD Adjusted R^2
Costs[1]	0.349	0.457
Length of Stay[1]	0.411	0.426

[1] Costs and Length of Stay were transformed by taking the natural logarithm.

5.12 Value Added

A variable is considered to have added value to surgical quality assessment if it is significant in the multivariate regression model of one of the outcomes. The results for backwards stepwise logistic regressions of the NSQIP and UHCCD (p for variable entry 0.05, for exit 0.10) variables versus each of the four outcomes are available in Appendix B. They are ranked in descending order by the standardized coefficient for the linear regressions and by the product of the odds ratio and the variable standard deviation in the logistic regressions (Garson, 2005). The number of significant variables for each model is shown in Table 5.7. The NSQIP has more significant variables for each of the outcomes

Table 5.7 Added Value: The Number of Significant Variables from Backwards Stepwise Regression Models of the Two Datasets versus Each of the Four Outcomes.

Model	Mortality	Morbidity	Cost	Length of Stay
NSQIP	19	27	37	34
UHCCD	12	13	25	25

In each model, there are a few variables that have much higher standardized or ranking coefficients than the remaining variables. This occurs for both datasets, but is more striking for the UHCCD. In the UHCCD much more of the added value comes from the SOI Moderate, Major or Extreme classifications and Age than from the other variables.

5.12.1 Variables that add value in assessing mortality

The five most significant variables in the NSQIP mortality model are ASA Classes 3 to 5, Age, Dyspnea with Minimal Exertion, BMI (which is protective), and Work RVUs. There is a decrease in added value between the ASA Classes 3 to 5 and Age versus the rest of the variables which have about half of the ranking variable as the first four. The five most significant variables from the UHCCD mortality model are Major or Extreme SOI, Age, Transfer from an Acute Care Hospital, Deficiency Anemias, and Emergency Admission. There also is a decrease in added value between the Severity of Illness Major or Extreme, Age and the rest of the variables. This decrease is more striking than in the NSQIP, with the less important variables having about a third the ranking value as the most important variables. The UHCCD has fewer variables that add value in assessing mortality and the benefit is found more heavily in the top few variables, namely Major or Extreme SOI and age.

5.12.2 Variables that add value in assessing morbidity

The five most significant variables in the NSQIP morbidity model are Return to the O.R., Abnormal Bilirubin, Duration of the Operation, ASA Classes 3 to 5, and Work RVUs. The variation of ranking coefficients between the more numerous significant variables is less pronounced than in the mortality model indicating the multifaceted nature of complication in surgical patients. The five most significant variables in the UHCCD morbidity model are Age, Moderate, Major or Extreme SOI, Chronic Artery Disease, Coagulopathy, and Deficiency Anemias. As in the mortality model, the Extreme and Major SOI variables have substantially greater ranking coefficients than the other variables. Also as in the mortality models, the UHCCD has fewer variables than the NSQIP that add value in assessing mortality and the value is found more heavily in the top few variables.

5.12.3 Variables that add value in assessing costs

The five most significant variables in the NSQIP costs model are Duration of the Operation, Work RVUs, Preoperative Open Wound or Infection, ASA

Class 4, and a Hematocrit less than 38. Most hospital costs for surgical patients accrue in the O.R. (Davenport et al., 2005) so the importance of the perioperative variables from the O.R. is to be expected. In the NSQIP cost model there is a sharp decrease in the standardized coefficients between these variables and the preoperative risk variables. There is however, a greater number of variables in the cost model compared to the mortality and morbidity models. This is likely due to the increased statistical power from the continuous cost outcome available for all patients and also due to the multifaceted nature of cost drivers in the hospital. The five most significant variables in the UHCCD costs model are Moderate, Major or Extreme SOI, Emergency Admission, and Transfer from an Acute Care Hospital. As in the other models, the Extreme and Major SOI variables have significantly greater ranking coefficients than the other variables. Also as in the mortality models, the UHCCD has fewer variables that add value in assessing mortality and the value is found more heavily in the top few variables.

5.12.4 Variables that added value in assessing length of Stay

The results from the length of stay regressions are analogous to the costs results and are detailed in Appendix B.

5.13 Timeliness

As mentioned in the measures section, the decision model used in this case study does not contain timing information with which to evaluate timeliness in the two systems. In the UHCCD, data is uploaded for modeling on a quarterly basis and modeled results are available 3-6 months after discharge. The NSQIP data is reviewed starting 30 days after the operation and is uploaded over the course of the ensuing 30 to 60 days. However, modeling occurs on a biannual basis and so is available at 6 month intervals. Both systems provide retrospective assessment of quality several months after the fact.

5.14 Appropriate Amount of Data

As described above, the domain coverage of the two systems is quite different. Linking the records in order to provide direct comparison excludes large sections of each database. Applying power calculations, therefore, is not performed on the models for each system and no direct comparison is made. In its biannual report, the NSQIP does provide confidence intervals for the assessments. In its web reporting the UHCCD does note statistical outlier status in its models, but does not provide confidence intervals for the estimates.

Chapter 6: Discussion

6.1 Workgroup Context

Interviews with the process owners of the primary data encoding from the two systems reveal different workgroup contexts as expressed by primary purpose, history, different group identification and skill sets of the personnel involved, and by the vocabulary used in describing events. These differences impact the data element selection and definition, the encoding and data entry processes, the managerial control and feedback, and the data transformation that occurs in the two systems. In general, the NSQIP context derives from its surgeon designers with its perspective on the operation as the seminal event. The clinical factors and outcomes are the key variables measured and they are encoded based on strict clinical definitions applied by nurse reviewers who see themselves as clinicians. The UHCCD context by contrast, reflects its consortium of hospital designers with its perspective on the admission as the seminal event and the administrative and clinical factors related to the admission being the key variables. The hospital encoding staff do not consider themselves part of the clinical workgroup, but more related to the administrative workgroup supporting the billing functions of the hospital primarily and clinical functions secondarily. The interviews did not yield any mention of system malfunction resulting in support for Proposition 2.

> *Proposition 2: Differences in the two information systems effectiveness in risk-adjusting surgical outcomes are not due to system failure, but to differences in the workgroup context in which the data is derived.*

6.2 Data Element Definitions and Encoding Processes

The two workgroup contexts are expressed in the most fundamental definition of this comparison study, that is, “what constitutes a surgical patient?” In the UHCCD it is admitted patients with surgical DRGs. Practically, when the extract was made for this study, which patients had operating room charges during their care became the criteria. This excludes all outpatient surgeries; estimated at about half of surgical patients. In the NSQIP, included patients are those who had “major” surgery as defined by the type of anesthesia, a clinical factor unavailable to the UHCCD. The NSQIP was started in the VA hospitals which led to its exclusion of pediatric patients and trauma cases, about a quarter of surgical patients at UKMC but rare at VA hospitals.

In addition to differences of perspective on the definition of a surgical patient, the two systems code the surgical procedure itself differently. The UHCCD uses ICD-9 coding because it forms the basis for hospital billing and recording of clinical events. The NSQIP uses CPT codes used by physicians for billing and recording of clinical events. Neither of the organizations responsible for the respective coding systems provides an official crosswalk definition to allow for one-to-one mapping of CPT codes to ICD-9 codes or vice versa. Any targeted review of surgical procedures must choose one or the other, immediately imparting a measure of confusion and discomfort to the workgroup not normally using the codes. The differing workgroup ontologies related to these two systems clearly led to different perspectives on what defines a surgical patient and how to describe the surgical procedure itself.

These differences in definition continue throughout the systems and confirm the proposition that information systems are workgroup artifacts that reflect the workgroup context which includes repertoire and vocabulary. These findings in two datasets at the same institution covering the same acute clinical events highlight and confirm the potential contextual data quality issues related to data definition in other information systems proposed in this study. Proposition 4 is supported.

> *Proposition 4: The systems have different data elements, definitions, and encoding processes which reflect the context of the workgroup using them and affect contextual data quality.*

6.3 Managerial Control and Feedback

Managerial control and feedback are also closely tied to workgroup context and differ between the two systems. This is most clearly represented in the audits of the respective coders. In the NSQIP, audits and site reviews focus on correct application of the clinical definitions of risk factors and outcomes. A random sample of patients is selected and re-extracted by the auditor and the results compared to those encoded by the nurses. The UHCCD ICD-9 codes are audited as well, but by federal auditors who determine whether sufficient physician documentation exists to support submission of a bill to Medicare. This limits the coders to only events with unmistakable physician documentation. In this case, this limitation contributes to the under-reporting of postoperative AMI and postoperative pneumonia described earlier. This limitation may also contribute to the numerous complications listed as "other" or "miscellaneous" in the UHCCD. That is, an event occurred and was sufficiently documented to allow for ICD-9 coding but was not specific enough to be related directly to the surgical event.

In these two systems, managerial control and feedback result in direct impact on IS processes that result in differing contextual data quality. This substantiates the notion that it is not only system design, function, and data dictionaries that determine IS data quality, but ongoing influence of management on IS processes. The nature of the respective audits is particularly useful in this case for analyzing contextual data quality and therefore recommends itself as part of the "data architecture" metadata recommended by Yoon and Aiken (2000). Proposition 5 is supported.

> *Proposition 5: Managerial control differs in the two systems and affects contextual data quality.*

6.4 Data Transformation

The data transformation required for inputs into the quality assessment model also differs between the two systems and reflects workgroup context. The

NSQIP performs very little data transformation because the data elements are predefined to track specific information about preoperative risk, the nature of the operation, and surgical outcome, and, in particular, the timing of a clinical event relative to the operation. By contrast in the UHCCD, the hospital coding of secondary diagnoses does not include timing relative to the operation and, therefore, conservative assumptions and algorithms are applied to distinguish between comorbid conditions likely present at admission, and potentially preventable complications resulting from care. Limitations in these algorithms resulted in preoperative pneumonias labeled as postoperative and the mis-screening of both AMIs and postoperative pneumonias. Also in the UHCCD, the APR-DRG and SOI level calculated by the 3M coding grouper uses complex and proprietary algorithms based on combinations of diagnoses and procedure codes in order to maximize the information obtained from the secondary procedures and codes. Again, this transformation is required based on the limitations of the underlying data set.

These two systems differ markedly in their data transformations. In the UHCCD dataset the transformation is necessary given the structure of the underlying data. This transformation's high complexity, especially when the algorithm is proprietary, tends to obscure the contextual quality of the data. This raises a caution to decision makers using data that needs heavy transformation in order to support a particular decision and confirms this dissertation's proposition that data transformation does impact contextual data quality. Proposition 6 is supported.

> *Proposition 6: The systems have different data transformations which reflect workgroup context and affect contextual data quality.*

6.5 Identification of Complications

The influence of work group context on data selection and definition, managerial feedback and control, and data transformation along with the ensuing impact on contextual data quality is most clearly shown in this study by the marked disagreement regarding complications. Significant disagreement occurs even in complications that, on the surface, are described as the same. Results of the investigation of the almost complete disconnect between the two systems' assessments of postoperative AMI and pneumonia highlight these differences. The hospital ICD-9 codes for AMI include AMIs not defined as clinically acute enough for inclusion in the NSQIP – differences in data definition. In some cases, hospital ICD-9 coding does not occur when the clinical documentation clearly supports it but explicit physician documentation may be lacking – differences in managerial control. In a large percentage of the cases, hospital coding exists for the AMI and pneumonia, but it is not identified as a potentially preventable complication, most likely because of confounding with the principle diagnosis – limitations in data transformation. Indeed, analysis of the two systems does not yield a common understanding of what a complication is.

The differences in complications are so profound as to hinder a direct comparison of the two systems contextual data quality. Complications are important outcomes that impact the health status of the patient, as well as increasing costs and length of stay in the hospital. As noted in the results, the most significant predictive variable in the UHCCD for all of the four outcomes is the SOI level. The SOI assignment algorithm appears from the available documentation to include in some instances secondary diagnoses that are considered complications in the NSQIP and secondary procedures that are considered therapy resulting from complications, particularly ventilator dependence related to pulmonary compromise. Prolonged ventilator dependence is the single most common "complication" in the NSQIP. Because it may include information about complications, it is unsurprising then that SOI is the strongest single predictor of outcomes across the two systems. The significance of bias is incalculable however because of the complexity of the assignment algorithms and their unavailability for scrutiny.

6.6 Contextual Data Quality

This potential bias in what is the most significant predictor variable in the UHCCD puts in doubt the contextual quality comparisons made in the study. With that caveat, however, the following observations are made.

- The NSQIP has more relevant variables for assessing the four outcomes than the UHCCD.

- The NSQIP is more complete in terms of its ability to predict morbidity and mortality in surgical patients.

- The UHCCD is more complete in terms of its ability to predict costs and length of stay in surgical patients.

- The NSQIP has more variables that add value in assessing each of the four outcomes with a more even value distribution across variables than the UHCCD. However, the UHCCD variable SOI adds the most value in assessing the four outcomes.

In general then, the NSQIP does have higher contextual data quality which is clearly tied to the IS processes stemming from the workgroup context surrounding it. This confirms propositions one and three put forth in this study.

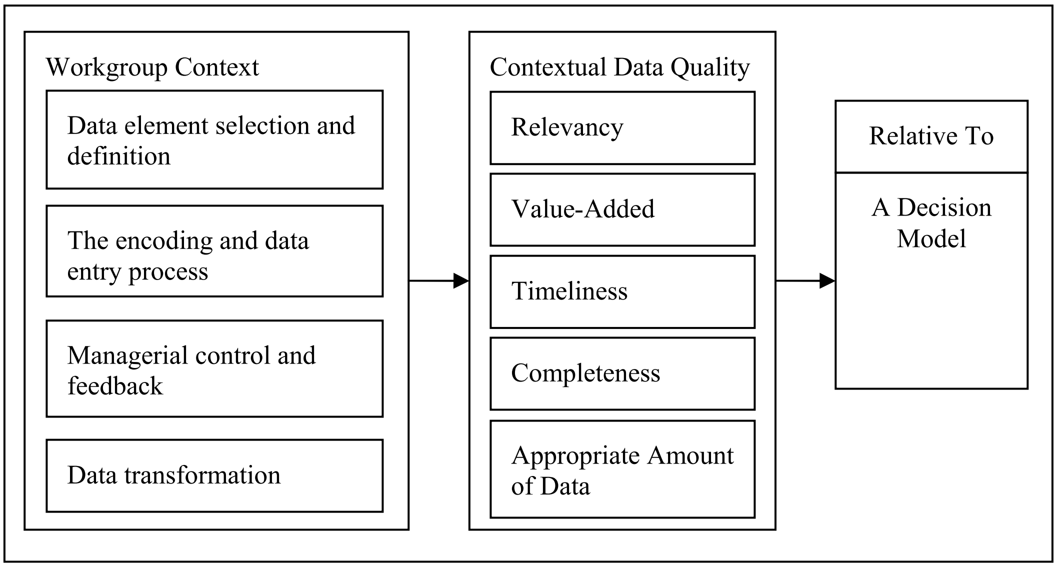
> *Proposition 1: The systems are not equally effective in risk-adjusting surgical outcomes.*
>
> *Proposition 3: Differences in the two information systems' effectiveness in risk-adjusting surgical outcomes are due to differences in their contextual data quality dimensions of added value, relevancy, timeliness, completeness, and appropriateness of the amount of the data in the systems.*

All the study propositions are confirmed in this study. They are represented in Figure 6.1, and suggest hypotheses in support of a theory of contextual data quality:

> Hypothesis1: Workgroup context influences the IS processes of data element selection and definition, the encoding and data entry process, managerial control and feedback, and data transformation.

Hypothesis 2: These IS processes influence contextual data quality relative to a particular decision model.

Figure 6.1 A Theory of IS Contextual Data Quality



6.7 Recommendations to IS Managers and Decision Makers

As recommended by Yoon and Aiken (2000), IS administrators and decision makers need meta-data about data contained in information systems drawn from across functional and hierarchical boundaries in the organization. The review of systems presented here provides some specific suggestions. First, understand the impact of context on data quality. Data quality issues may not arise in the normal operation of a particular system but significantly impact a use out of the context in which the data is derived. Second, the data dictionary, or face description of the data elements, is not sufficient to understand the contextual data quality related to a particular use. In these two systems, for instance, there are several complications appearing similar based on the data element description alone, but which have almost no overlap in what they actually measure.

If a quantitative comparison of CDQ is not feasible, there are elements of workgroup context that are shown in this study to influence CDQ with which to perform a qualitative assessment. These elements are the primary purpose and history of the data encoded in the system; the workgroup self-identification, skill set and vocabulary of the primary data encoders; and the audits and methods by which the encoders are evaluated by managers or outside agencies. Lastly, a practical rule is that if a dataset requires complex transformation to suit a particular need, it may have reduced CDQ relative to that use.

6.8 Recommendations to Those Assessing Surgical Quality

Both the NSQIP and UHCCD datasets showed CDQ dimensions of relevance, completeness, and value added in estimating the surgical outcomes of mortality, morbidity, length of stay and costs. As noted earlier, much of the CDQ for the UHCCD derives from the SOI variable calculated by the 3M© APR-DRG grouper. This variable appears to capture well the influence of combinations of primary and secondary diagnoses and procedures on surgical outcomes. From that perspective it effectively achieves its goal of adjusting overall hospital acuity based on the entire length of stay of the patient. In contrast, the NSQIP is a better prospective predictor of the outcomes of the patient because of its more robust clinical capturing of the preoperative physical and comorbid state of the patient. This is demonstrated in better prediction of outcomes for highly acute patients. When used prospectively therefore, it has the potential to more accurately identify high risk patients and to provide better understanding of the clinical conditions that might be more effectively managed to improve surgical quality.

The context of hospital coding for claims also obscures the identification of specific complications and their timing. The UHCCD SOI calculation appears to include procedures and diagnoses codes that could be considered complications of care, or treatment related to complications. This is likely why it is so strong in predicting complications, costs and length of stay in this analysis. For managing and reducing complications then, the NSQIP provides more useful information regarding the complication itself, and the clinical conditions preceding it.

The UHCCD does a better job at looking at what care was given to the patient throughout their stay and estimating the severity of illness. For purposes of comparing resource utilization then, it did well in this analysis. The NSQIP, on the other hand, does a better job at looking at the preoperative, and usually the preadmission, condition of the patient. Its dataset does a better job of estimating the impact of these conditions on clinical outcomes. For the purposes of risk-adjustment of mortality and morbidity then, the NSQIP is the better system.

The challenge, of course, that faces most hospital administrators is how to justify the costs related to the hiring of nurses to capture the NSQIP data; especially when the existing administrative dataset may do sufficiently well. The answer lies in the potential monetary benefit related to reducing complications and the potential market benefit of not being mislabeled as a hospital with poor risk-adjusted outcomes. The NSQIP's more accurate identification of complications and the clinical conditions preceding them may more effectively support process improvement efforts to reduce them. Complications have been shown to be costly in surgical patients (Davenport et al., 2005, Dimick et al., 2004) and their reduction has the dual benefit of improving patient health and reducing the costs of care. With respect to mislabeling, more information regarding hospital performance is being publicly reported nationally increasing the risk of market impact. The value of more accurate risk adjustment in the NSQIP lies in its ability to respond to inaccurate or less accurate ranking of a hospital in regards to mortality and morbidity. This value will only increase in the next decade. Indeed, national payers are at least anecdotally taking notice of the NSQIP methodology and are considering it for surgical risk-adjustment on the national level (Surgical Care Improvement Project, 2005). They also are struggling with cost, and, in addition to the two value propositions presented here are also faced with the ethical issue of being a national body that may mislabel a hospital or provider without using the best possible risk-adjustment methodology. This analysis clarifies the strengths and weaknesses of the two types of systems and contributes to debate.

Chapter 7: Contributions and Limitations

The major contributions of this study are twofold. The first is the demonstration of the contextual data quality differences between a clinical quality derived information system and an administrative system in the assessment of surgical quality. Understanding these differences will have impact nationally on decisions regarding hospital accreditation and on quality-based reimbursements to hospitals. Nationally and locally, a better understanding of surgical quality informs efforts to decrease surgical mortality and complication and increases the effectiveness of surgical care in improving patient health. The major limitation of the study in regard to this contribution is that it is a single hospital case, so its generalizability may be restricted. Further research comparing the systems' performance across multiple sites needs to be undertaken. This is only now beginning to be possible as the NSQIP is expanded nationally. This case example however, from a major academic health center, impacts the national debate and has no equivalent published in the literature.

The second contribution is the development of a theoretical framework for assessing contextual data quality in information systems. While contextual data quality problems have been noted in the IS literature, a method for analysis for quantifying and qualifying contextual data quality has not been developed. The case study executed is appropriate for the exploration and elaboration of new theory in this area. The resulting new theory based on the concepts of workgroup context and information systems as workgroup knowledge artifacts is a new tool for decision makers and system managers in assessing data quality. The case provides rich information needed to develop this theory, but is not able to confirm it. Generalizability and confirmation need further research through a larger study across multiple systems with different applications. The timeliness and appropriate amount of data dimensions of contextual data quality are unable to be measured in this case. They are included in the theoretical model, but with only qualitative support.

Lastly, this study examines the dimensions of contextual data quality only as they relate to workgroup context. It does not put these dimensions in relationship to other data quality dimensions, nor does it look at other outcomes such as IS cost. Further research is needed to integrate the concepts presented here into a broader IS data quality theory.

Table A.1 UHCCD Model Independent Variables, Correlations (P<0.01) and Strength of Associations (P<0.01) with Outcomes (n=4,221)

Pre-Operative Variable (Mean or Occur. Rate)	Mortality (4.0%)	NSQIP Morbidity (13.3%)	UHC Morbidity (21.3%)	Length of Stay (Median = 4.0)	Total Hospital Costs (Median = 12,549)
Demographic Risk Factors					
Age (51.7)	0.146	0.111	0.124	0.069	0.092
Gender=Female (49%)	-*	-	-	-	0.046
Minority (9%)	-	-	-	-	-
Emergency Admission (24%)	0.126	0.131	0.125	0.176	0.164
Transfer From Acute Care Hospital (8%)	0.197	0.127	0.171	0.207	0.228
Low Social Economic Status (22%)	-	-	-	0.068	0.041
Primary Diagnosis and Procedure					
DRG	0.284	0.419	0.503	0.562	0.641
APR-DRG	0.327	0.377	0.424	0.593	0.656
Severity of Illness	0.397	0.431	0.522	0.541	0.624

* "-" Indicates that no statistically significant association existed.

Table A.1 UHCCD Model Independent Variables, Correlations (P<0.01) and Strength of Associations (P<0.01) with Outcomes (n=4,221) - Continued

Pre-Operative Variable (Mean or Occur. Rate)	Mortality (4.0%)	NSQIP Morbidity (13.3%)	UHC Morbidity (21.3%)	Length of Stay (Median = 4.0)	Total Hospital Costs (Median = 12,549)
Comorbid Factors					
AIDS (0.2%)	-	-	-	-	-
Alcohol abuse (4.0%)	-	0.059	0.083	0.083	0.087
Cancer with poor prognosis (0.2%)	-	-	-	-	-
Chronic artery disease (1.5%)	-	-	-	-	-
Chronic blood loss/anemia (0.8%)	-	0.078	0.115	0.101	0.099
Chronic pulmonary disease (17.6%)	0.054	0.055	0.081	0.044	0.068
Chronic renal failure (0.1%)	-	-	-	-	-
Coagulopathy (3.0%)	0.185	0.199	0.223	0.293	0.369
Congestive Heart Failure (4.6%)	0.169	0.141	0.158	0.185	0.218
Deficiency Anemias (6.9%)	-	-	0.110	0.099	0.092
Dementia (0.0%)	-	-	-	-	-
Depression (5.5%)	-	-	-	-	-
Diabetes with CCs (1.9%)	-	-	0.056	0.043	-
Diabetes without CCs (12.2%)	-	-	0.040	-	-
Diabetes with end organ damage (0.0%)	-	-	-	-	-

Table A.1 UHCCD Model Independent Variables, Correlations (P<0.01) and Strength of Associations (P<0.01) with Outcomes (n=4,221) - Continued

Pre-Operative Variable (Mean or Occur. Rate)	Mortality (4.0%)	NSQIP Morbidity (13.3%)	UHC Morbidity (21.3%)	Length of Stay (Median = 4.0)	Total Hospital Costs (Median = 12,549)
Drug abuse (1.1%)	-	-	0.064	0.077	0.083
Fluid and electrolytic disorders (10.5%)	0.221	0.272	0.319	0.366	0.398
Functional impairment (0.3%)	-	-	-	-	-
Hypertension (35.9%)	-	-	-	-0.041	-0.043
Hypothyroidism (6.9%)	-	-	-	-	-
Liver disease (1.2%)	0.066	0.053	0.070	0.050	0.050
Lymphoma (0.5%)	-	-	-	-	0.043
Metastatic cancer (4.4%)	-	-	0.062	0.052	0.043
Nutritional deficiencies (0.2%)	0.051	-	0.050	0.077	0.060
Obesity (12.4%)	-0.047	-	-0.050	-0.052	-0.052
Other neurological disorders (3.8%)	0.060	0.067	0.065	0.116	0.107
Paralysis (3.0%)	-	0.041	0.040	0.068	0.055
Peptic ulcer dis. w/ bleeding (3.1%)	-	-	-	-	-
Peripheral vascular disease (5.6%)	0.045	0.045	0.121	0.055	0.069
Psychoses (1.5%)	-	-	-	-	-
Pulmonary circulation disease (0.4%)	-	-	-	-	-
Renal failure (3.2%)	0.114	0.068	0.152	0.142	0.173

Table A.1 UHCCD Model Independent Variables, Correlations (P<0.01) and Strength of Associations (P<0.01) with Outcomes (n=4,221) - Continued

Pre-Operative Variable (Mean or Occur. Rate)	Mortality (4.0%)	NSQIP Morbidity (13.3%)	UHC Morbidity (21.3%)	Length of Stay (Median = 4.0)	Total Hospital Costs (Median = 12,549)
Rheumatoid arthritis/ collagen vas (1.9%)	-	-	-	-	-
Severe chronic liver disease (0.1%)	0.063	-	-	-	-
Solid tumor w/o metastasis (7.1%)	-	-	-	-	-
Valvular disease (3.5%)	0.061	0.037	0.044	0.033	0.044
Weight loss (2.4%)	0.151	0.187	0.189	0.263	0.271

Table A.2 NSQIP Model Independent Variables, Correlations (P<0.01) and Strengths of Association (P<0.01) with Outcomes (n=4,221).

Pre-Operative Variable (Mean or Occur. Rate)	Mortality (4.0%)	NSQIP Morbidity (13.3%)	UHC Morbidity (21.3%)	Length of Stay (Median = 4.0)	Total Hospital Costs (Median = 12,549)
Demographic Risk Factors					
Age (51.7)	0.113	0.091	0.101	0.113	0.112
Gender=Female (49%)	-	-	-	-	-0.046
Minority (9%)	-	-	-	-	-
Transfer From Healthcare Facility (3.3%)	0.166	0.136	0.129	0.116	0.108
Cardiac Risk Factors					
Previous Cardiac Surgery (7.2%)	0.088	0.096	0.117	0.098	0.086
Previous PTCA (4.5%)	0.113	0.072	-	0.053	0.054
History of CHF (2.3%)	0.168	0.087	0.198	0.141	0.144
History of Angina (2.8%)	0.150	0.091	0.134	0.079	0.103
History of Myocardial Infarction (1.4%)	0.082	0.064	0.150	0.082	0.102
History of Hypertension (44.0%)	0.060	0.093	0.113	0.083	0.098

Table A.2 NSQIP Model Independent Variables, Correlations (P<0.01) and Strength of Associations (P<0.01) with Outcomes (n=4,221) - Continued

Pre-Operative Variable (Mean or Occur. Rate)	Mortality (4.0%)	NSQIP Morbidity (13.3%)	UHC Morbidity (21.3%)	Length of Stay (Median = 4.0)	Total Hospital Costs (Median = 12,549)
Central Nervous System Risk Factors					
Impaired Sensorium (4.7%)	0.240	0.183	0.176	0.173	0.196
Coma (0.4%)	0.135	0.124	0.081	0.057	0.076
Hemiplegia (4.9%)	0.083	0.093	0.060	0.061	0.082
History of TIA (3.7%)	-	-	0.041	-	0.033
CVA w/ Neurological Deficit (4.7%)	0.114	0.113	0.107	0.082	0.088
CVA w/o Neurological Deficit (2.5%)	-	0.057	0.067	0.062	0.063
CNS Tumor (4.4%)	-	-	-0.042	-	0.069
Hepatobiliary Risk Factors					
Esophageal Varices (0.1%)	-	-	0.059	-	-
Ascites (1.9%)	0.181	0.147	0.128	0.100	0.090
Pulmonary Risk Factors					
Dyspnea (w/ Min. Exert. 14.6%, At Rest 4.0%)	0.225	0.189	0.224	0.202	0.206
Ventilator Dependent > 48 Hrs. (3.5%)	0.308	0.231	0.289	0.164	0.217
History of COPD (12.3%)	0.122	0.124	0.142	0.162	0.146
Current Pneumonia (2.0%)	0.229	0.123	0.173	0.145	0.144

Table A.2 NSQIP Model Independent Variables, Correlations (P<0.01) and Strength of Associations (P<0.01) with Outcomes (n=4,221) - Continued

Pre-Operative Variable (Mean or Occur. Rate)	Mortality (4.0%)	NSQIP Morbidity (13.3%)	UHC Morbidity (21.3%)	Length of Stay (Median = 4.0)	Total Hospital Costs (Median = 12,549)
Nutritional / Immune / Other Risk Factors					
Diabetes (Orally Tr. 7.2%, Insul. 6.9%)	0.091	0.074	0.115	0.118	0.088
Disseminated Cancer (5.1%)	-	-	-	0.065	0.051
Open Wound or Infection (12.8%)	0.044	0.044	0.120	0.251	0.143
Steroid Use (5.9%)	0.088	0.065	0.072	0.055	0.056
Weight Loss > 10% (3.4%)	0.074	0.056	0.067	0..098	0.060
Bleeding Disorder (1.9%)	0.183	0.088	0.113	0.071	0.099
Transfusion > 4 Units (1.0%)	0.169	0.104	0.129	0.074	0.110
Chemotherapy (1.3%)	0.041	-	-	-	-
Radiotherapy (1.5%)	0.043	-	-	0.040	-
Sepsis (2.7%)	0.276	0.179	0.195	0.149	0.161
BMI (Mean = 28.7)	-0.052	-	-	-0.045	-
Renal Risk Factors					
Acute Renal Failure	0.130	0.099	0.140	0.107	0.103
On Dialysis	0.114	0.053	0.127	0.109	0.106
Vascular Risk Factors					
History of Peripheral Vascular Disease (4.1%)	0.043	-	0.085	0.105	0.074
History of Rest Pain / Gangrene (3.4%)	-	-	0.099	0.107	0.078

Table A.2 NSQIP Model Independent Variables, Correlations (P<0.01) and Strength of Associations (P<0.01) with Outcomes (n=4,221) - Continued

Pre-Operative Variable (% Obtained, Mean Result When Obtained)	Mortality (4.0%)	NSQIP Morbidity (13.3%)	UHC Morbidity (21.3%)	Length of Stay (Median = 4.0)	Total Hospital Costs (Median = 12,549)
Laboratory Values					
Alkaline Phosphatase (33%, 100)	-	-	-	0.069	-
Total Bilirubin (33%, 1.05)	0.130	0.129	0.076	-	0.050
Blood Urea Nitrogen (87%, 15.1)	0.105	0.084	0.100	0.030	0.046
Serum Creatinine (87%, 1.10)	0.086	0.074	0.111	-	0.043
Hematocrit (89%, 38.3)	-0.105	-0.091	-0.177	-0.262	-0.188
Platelet Count (88%, 285)	-0.089	-0.038	-	0.009	-0.030
Prothrombin Time (69%, 12.0)	0.140	0.100	0.111	0.131	0.059
Partial Thromboplastin Time (52%, 28.8)	0.075	-	0.077	0.093	0.062
Serum Glutamic Oxaloacetic Test (34%, 35.2)	0.074	-	-	-	-
Serum Sodium (87%, 138)	-	-0.038	-0.074	-0.141	-0.078
White Blood Count (88%, 9.5)	0.060	0.080	0.082	0.066	0.054
Serum Albumin (34%, 3.21)	-0.164	-0.124	-0.195	-0.293	-0.249
International Normalized Ratio (69%, 1.05)	0.157	0.105	0.133	0.167	0.126

Table A.2 NSQIP Model Independent Variables, Correlations (P<0.01) and Strength of Associations (P<0.01) with Outcomes (n=4,221) - Continued

Pre-Operative Variable (Mean or Occur. Rate)	Mortality (4.0%)	NSQIP Morbidity (13.3%)	UHC Morbidity (21.3%)	Length of Stay (Median = 4.0)	Total Hospital Costs (Median = 12,549)
General Risk Factors					
ASA Class (Median = 2)	0.248	0.236	0.304	0.300	0.300
Pack Years Smoked (Mean = 17.8)	-	-	0.050	0.063	0.051
Current Smoker (36.7%)	-	-	-	-	-
Alcohol > 2 drinks/day (3.0%)	-	-	-	0.038	0.042
DNR Status (0.3%)	0.061	-	-	-	-
Functional Status (Part. Dep. 11.0%, Tot. 6.9%)	0.258	0.214	0.258	0.285	0.264
Intraoperative Factors	**-**	**-**	**-**	**-**	**-**
Aneshesia Technique (98.6% General An.)	-	-	-	-	-
Surgical Specialty	0.125	0.151	0.231	0.221	0.204
Emergency Case (15.9%)	0.222	0.199	0.243	0.199	0.194
Wound Class (44.9% Not Clean)	0.105	0.088	0.139	0.201	.0136
CPT Codes	0.528	0.527	0.561	0.557	0.625
Max. Work RVUs (18.4)	0.051	0.058	0.055	0.049	0.124
Operative Time (Mean = 2.6 Hrs.)	-	0.061	0.053	0.056	0.143
PACU Time (Mean = 2.8 Hrs.)	-	0.043	0.042	-	0.043
Return to the O.R. (10.4%)	0.109	0.231	0.233	0.289	0.320

Table B.1 NSQIP Mortality Model Summary

Model Fit		
Chi-Square	Df	Sig.
562.277	15	0.000
Model Summary		
-2 Log likelihood	Cox & Snell R^2	Nagelkerke R^2
855.460	0.125	0.437

Variables in the Equation	*B*	Sig.	Odds Ratio	Variable S.D.	*B* x S.D.
ASA Class 3	1.541	0.000	4.67	0.487	0.750
ASA Class 4	2.362	0.000	10.61	0.293	0.691
Age	0.038	0.000	1.04	16.649	0.639
ASA Class 5	4.234	0.000	68.98	0.083	0.350
Dyspnea w/ Minimal Exertion	0.896	0.000	2.45	0.353	0.317
Body Mass Index	-0.042	0.009	0.96	7.173	-0.302
Maximum Work RVUs	0.032	0.000	1.03	9.321	0.300
Totally Dependent	1.101	0.000	3.01	0.253	0.279
Partially Dependent	0.702	0.004	2.02	0.313	0.220
Preoperative Sepsis	0.563	0.000	1.76	0.368	0.207
Bilirubin > 1.0	0.625	0.018	1.87	0.264	0.165
Radiotherapy	1.282	0.006	3.61	0.123	0.158
Alkaline Phosphatase > 125	0.648	0.024	1.91	0.234	0.152
Partial Thromboplastin Time > 35	0.734	0.012	2.08	0.192	0.141
Preoperative Pneumonia	0.998	0.003	2.71	0.140	0.140
Impaired Sensorium	0.603	0.022	1.83	0.212	0.128
Transfused > 4 Units	1.145	0.010	3.14	0.102	0.116
Ascites	0.773	0.036	2.17	0.138	0.107
History of Myocardial Infarction	0.788	0.042	2.20	0.119	0.094
Constant	-7.656	0.000	0.00		

Table B.2 NSQIP Morbidity Model Summary

Model Fit		
Chi-Square	Df	Sig.
302.814	27	0.000
Model Summary		
-2 Log likelihood	Cox & Snell R^2	Nagelkerke R^2
2108.299	0.080	0.165

Variables in the Equation	*B*	Sig.	Odds Ratio	Variable S.D.	*B* x S.D.
Duration of Operation (Hrs)	0.161	0.000	1.175	1.927	0.310
Bilirubin > 1.0	1.084	0.000	2.956	0.264	0.286
ASA Class 3	0.523	0.000	1.688	0.487	0.255
ASA Class 4	0.848	0.000	2.335	0.293	0.248
ASA Class 5	2.828	0.033	16.909	0.083	0.235
Totally Dependent	0.771	0.001	2.162	0.253	0.195
Preoperative Sepsis	0.517	0.038	1.678	0.368	0.190
White Blood Count > 11	0.457	0.001	1.579	0.416	0.190
Platelets > 150	-0.718	0.018	0.488	0.237	-0.170
Emergency Operation	0.446	0.010	1.562	0.365	0.163
Previous Cardiac Operation	0.602	0.002	1.826	0.259	0.156
Max. Work RVUs	0.016	0.035	1.016	9.321	0.147
Alkaline Phosphatase > 125	-0.610	0.030	0.543	0.234	-0.143
Wound Class	0.148	0.023	1.159	0.963	0.142
Age	0.008	0.038	1.008	16.649	0.137
Preoperative Pneumonia	0.959	0.021	2.608	0.140	0.134
General Anesthesia	1.075	0.145	2.931	0.119	0.128
History of COPD	0.387	0.016	1.472	0.328	0.127
Partially Dependent	0.385	0.029	1.470	0.313	0.121
Minority	-0.385	0.074	0.681	0.303	-0.117
Insulin Treated Diabetes	-0.460	0.049	0.631	0.254	-0.117
Serum Albumin	0.252	0.068	1.287	0.440	0.111
Variables in the Equation	*B*	Sig.	Odds Ratio	Variable S.D.	*B* x S.D.
Body Mass Index	0.015	0.054	1.015	7.173	0.109

Table B.2 NSQIP Morbidity Model Summary - Continued

Variables in the Equation	*B*	Sig.	Odds Ratio	Variable S.D.	*B* x S.D.
Serum Glutamic Oxaloacetic Test > 40	0.475	0.063	1.608	0.230	0.109
Hemiplegia	0.445	0.055	1.561	0.215	0.096
Radiotherapy	0.689	0.060	1.991	0.123	0.085
Constant	-6.705	0.000	0.001		

Table B.3 NSQIP Costs Model Summary

Model Summary			
R	R Square	Adjusted R Square	Std. Error of the Estimate
0.596	0.355	0.349	0.513

ANOVA	Sum of Squares	df	Mean Square	F	Sig.
Regression	522	32	16.32	62.04	0.000
Residual	950	3,611	0.263		
Total	1,472	3,643			

Variable	Unstandardized Coefficients	Standardized Coefficients	Sig.
(Constant)	8.724		0.000
Duration of Operation (Hrs)	0.100	0.282	0.000
Max. Work RVUs	0.014	0.193	0.000
Open Wound/Infection	0.258	0.132	0.000
ASA Class 4	0.307	0.111	0.000
Hematocrit < 38	0.143	0.108	0.000
Totally Dependent	0.322	0.106	0.000
Wound Class	0.071	0.105	0.000
Partially Dependent	0.194	0.093	0.000
ASA Class 3	0.100	0.076	0.000
Preoperative Pneumonia	0.443	0.072	0.000
Impaired Sensorium	0.268	0.064	0.000
Dyspnea w/ Minimal Exertion	0.109	0.060	0.000
Serum Sodium < 135	0.101	0.051	0.000
Age	0.002	0.050	0.001
Emergency Operation	0.094	0.047	0.002
Preoperative Sepsis	0.153	0.045	0.002
Dyspnea At Rest	0.180	0.039	0.006
Variable	Unstandardized Coefficients	Standardized Coefficients	Sig.
Partial Thromboplastin Time > 35	0.135	0.035	0.012

Table B.3 NSQIP Costs Model Summary - Continued

Variable	Unstandardized Coefficients	Standardized Coefficients	Sig.
Current Smoker	0.040	0.030	0.032
Transfused > 4 Units	0.338	0.029	0.033
Alcohol > 2/day	0.112	0.029	0.037
Disseminated Cancer	0.082	0.028	0.044
White Blood Count > 11	0.040	0.026	0.075
ASA Class 5	0.490	0.026	0.064
Bilirubin > 1.0	0.065	0.025	0.072
Radiotherapy	0.116	0.022	0.099
Orally Treated Diabetes	-0.069	-0.027	0.046
History of Rest Pain or Gangrene	-0.106	-0.029	0.043
Creatinine > 1.2	-0.067	-0.032	0.028
Female	-0.050	-0.039	0.005
Serum Albumin	-0.085	-0.053	0.000

Table B.4 NSQIP Length of Stay Model Summary

Model Summary			
R	R Square	Adjusted R Square	Std. Error of the Estimate
0.645	0.416	0.411	0.7351

ANOVA	Sum of Squares	df	Mean Square	F	Sig.
Regression	1392.0	34	40.882	75.660	0.000
Residual	1950.1	3609	0.540		
Total	3340.1	3643			

Variable	Unstandardized Coefficients	Standardized Coefficients	Sig.
(Constant)	-0.114		0.111
Wound Class	0.222	0.223	0.000
Duration of Operation (Hrs)	0.096	0.194	0.000
ASA Class 4	0.573	0.175	0.000
ASA Class 3	0.313	0.159	0.000
Open Wound / Infection	0.379	0.133	0.000
Hematocrit < 38	0.230	0.117	0.000
Max Work RVUs	0.012	0.115	0.000
Totally Dependent	0.372	0.098	0.000
Partially Dependent	0.287	0.094	0.000
Impaired Sensorium	0.373	0.082	0.000
ASA Class 2	0.125	0.065	0.018
Platelets > 400	0.182	0.058	0.000
Age	0.003	0.057	0.000
Emergency Operation	0.145	0.055	0.000
Dyspnea At Rest	0.250	0.051	0.001
Bilirubin > 1.0	0.180	0.049	0.000
Dyspnea w/ Minimal Exertion	0.122	0.045	0.002

Table B.4 NSQIP Length of Stay Model Summary - Continued

Variable	Unstandardized Coefficients	Standardized Coefficients	Sig.
Disseminated Cancer	0.190	0.044	0.001
Preoperative Pneumonia	0.271	0.040	0.004
Serum Sodium < 135	0.114	0.040	0.003
History of Chronic Obstructive Pulmonary Disease	0.115	0.039	0.006
History of Congestive Heart Failure	0.245	0.039	0.007
Acute Renal Failure	0.371	0.038	0.007
Partial Thromboplastin Time > 35	0.176	0.035	0.012
Serum Glutamic Oxaloacetic Test > 40	0.118	0.028	0.040
White Blood Count <= 4.5	0.130	0.028	0.031
Prothrombin Time >= 13	0.093	0.027	0.052
Alcohol > 2 drinks/day	0.127	0.023	0.080
History of Peripheral Vascular Disease	-0.120	-0.025	0.065
Blood Urea Nitrogen > 40	-0.151	-0.025	0.076
History of Hypertension	-0.053	-0.028	0.060
Hematocrit > 45	-0.119	-0.036	0.006
Central Nervous System Tumor	-0.214	-0.046	0.001
Preoperative Sepsis	-0.167	-0.064	0.000

Table B.5 UHCCD Mortality Model Summary

Model Fit		
Chi-Square	Df	Sig.
456.87	13	0.000
Model Summary		
-2 Log likelihood	Cox & Snell R^2	Nagelkerke R^2
909.780	0.105	0.372

Variables in the Equation	B	Sig.	Odds Ratio	Variable S.D.	B x S.D.
SOI Extreme	3.522	0.000	33.87	0.274	0.966
Age	0.041	0.000	1.04	16.555	0.678
SOI Major	1.775	0.000	5.90	0.359	0.637
Transfer from Acute Care Hospital	0.949	0.000	2.58	0.269	0.255
Deficiency Anemias	-0.605	0.048	0.55	0.253	-0.153
Emergency Admission	0.350	0.066	1.42	0.427	0.149
Functional Impairment	2.353	0.037	10.51	0.053	0.125
Diabetes with End Organ Damage	3.895	0.002	49.16	0.031	0.120
Chronic Renal Failure	3.160	0.029	23.58	0.034	0.109
Nutritional Deficiencies	2.572	0.019	13.09	0.041	0.105
Liver Disease	0.805	0.078	2.24	0.109	0.088
Severe Chronic Liver Disease	2.529	0.050	12.53	0.034	0.087
Constant	-7.313	0.000	0.00		

Table B.6 UHCCD Morbidity Model Summary

Model Fit		
Chi-Square	Df	Sig.
605.459	13	0.000
Model Summary		
-2 Log likelihood	Cox & Snell R^2	Nagelkerke R^2
2634.963	0.136	0.251

Variables in the Equation	B	Sig.	Odds Ratio	Variable S.D.	B x S.D.
Age	0.011	0.000	1.011	16.555	0.186
SOI Moderate	0.311	0.031	1.364	0.474	0.147
SOI Major	1.234	0.000	3.436	0.359	0.443
SOI Extreme	2.865	0.000	17.555	0.274	0.785
Chronic Artery Disease	1.350	0.000	3.858	0.120	0.162
Coagulopathy	0.382	0.094	1.466	0.170	0.065
Deficiency Anemias	-0.448	0.019	0.639	0.253	-0.113
Diabetes w/ CCs	-1.098	0.007	0.334	0.136	-0.149
Fluid and Electrolytic Disorders	0.525	0.000	1.690	0.307	0.161
Lymphoma	-1.851	0.091	0.157	0.070	-0.130
Nutritional Deficiencies	2.242	0.005	9.411	0.041	0.092
Obesity	0.269	0.083	1.309	0.329	0.088
Pulmonary Circulation Disease	-1.222	0.135	0.295	0.063	-0.077
Constant	-3.366	0.000	0.035		

Table B.7 UHCCD Costs Model Summary

Model Summary			
R	R Square	Adjusted R Square	Std. Error of the Estimate
0.678	0.460	0.457	0.538

ANOVA	Sum of Squares	df	Mean Square	F	Sig.
Regression	1,009.6	25	40.38	138.73	0.000
Residual	1,185.5	4,102	0.289		
Total	2,195.1	4,127			

Variable	Unstandardized Coefficients	Standardized Coefficients	Sig.
(Constant)	9.045		0.000
SOI Extreme	1.346	0.507	0.000
SOI Major	0.643	0.317	0.000
SOI Moderate	0.232	0.151	0.000
Nutritional Deficiencies	1.356	0.076	0.000
Coagulopathy	0.315	0.073	0.000
Fluid and Electrolytic Disorders	0.171	0.072	0.000
Emergency Admission	0.104	0.061	0.000
Transfer from an Acute Care Hospital	0.163	0.060	0.000
Age	0.002	0.051	0.000
Deficiency Anemias	0.139	0.048	0.000
Peripheral vascular disease	0.134	0.042	0.000
Functional Impairment	0.552	0.040	0.000
AIDS	-0.700	-0.039	0.001
Chronic Blood Loss Anemia	0.320	0.038	0.001
Diabetes w/CCs	-0.195	-0.036	0.002
Liver disease	-0.196	-0.029	0.012
Metastatic Cancer	0.099	0.028	0.017
Weight loss	0.124	0.026	0.038
Paralysis	0.103	0.024	0.039

Table B.7 UHCCD Costs Model Summary - Continued

Variable	Unstandardized Coefficients	Standardized Coefficients	Sig.
Diabetes w/o CCs	-0.052	-0.023	0.047
Hypothyroidism	-0.064	-0.022	0.058
Female	-0.031	-0.021	0.072
Pulmonary circulation disease	-0.238	-0.021	0.074
Cancer with Poor Prognosis	0.346	0.019	0.097
Depression	0.062	0.019	0.095

Table B.8 UHCCD Length of Stay Model Summary

Model Summary			
R	R Square	Adjusted R Square	Std. Error of the Estimate
0.655	0.430	0.426	0.7245

ANOVA	Sum of Squares	df	Mean Square	F	Sig.
Regression	1,621.1	25	64.84	123.55	0.000
Residual	2,153.0	4,102	0.525		
Total	3,774.1	4,127			

Variable	Unstandardized Coefficient	Standardized Coefficient	Sig.
(Constant)	9.045		0.000
SOI Extreme	1.346	0.507	0.000
SOI Major	0.643	0.317	0.000
SOI Moderate	0.232	0.151	0.000
Nutritional deficiencies	1.356	0.076	0.000
Coagulopathy	0.315	0.073	0.000
Fluid and Electrolytic Disorders	0.171	0.072	0.000
Emergency Admission	0.104	0.061	0.000
Transfer from an Acute Care Hospital	0.163	0.060	0.000
Age	0.002	0.051	0.000
Deficiency Anemias	0.139	0.048	0.000
Peripheral Vascular Disease	0.134	0.042	0.000
Functional Impairment	0.552	0.040	0.000
Chronic Blood Loss Anemia	0.320	0.038	0.001
Metastatic Cancer	0.099	0.028	0.017
Weight loss	0.124	0.026	0.038
Paralysis	0.103	0.024	0.039
Cancer with poor prognosis	0.346	0.019	0.097
Depression	0.062	0.019	0.095
Female	-0.031	-0.021	0.072
Pulmonary Circulation Disease	-0.238	-0.021	0.074
Hypothyroidism	-0.064	-0.022	0.058

Table B.8 UHCCD Length of Stay Model Summary - Continued

Variable	Unstandardized Coefficient	Standardized Coefficient	Sig.
Diabetes w/o CCs	-0.052	-0.023	0.047
Liver disease	-0.196	-0.029	0.012
Diabetes w/CCs	-0.195	-0.036	0.002
AIDS	-0.700	-0.039	0.001

References

Bailor, Coreen, (2005 Oct) “Data Quality: The Market”, *Customer Relationship Management Magazine*, Vol. 9(10), pp. 44-46.

Averill, R.F., et al., (2003) “All Patient Refined Diagnosis Related Groups (APR-DRGs): Methodology Overview” accessed at the Healthcare Cost and Utilization Project Website url = www.hcup-us.ahrq.gov/db/nation/nis/APR-DRGsV20MethodologyOverviewand-Bibliography.pdf on 11/4/2005.

Ballou, D.P., and Tayi, G.K., (Jan 1999) “Enhancing Data Quality in Data Warehouse Environments,” *Communications of the ACM*, Vol. 42(1), pp. 3-82.

Benbasat, I., Goldstein, D.K., and Mead, M. (September 1987) “The Case Research Strategy in Studies of Information Systems,” *MIS Quarterly*, Vol. 11(3), pp. 368-86.

CMS - Center for Medicare and Medicaid Services, (2004) “National Health Expenditure Tables” accessed at http://www.cms.hhs.gov/statistics/nhe/historical/ on 11/4/2004.

Davenport, D.L. and Sena, M. (2003) “Technologies for Knowledge Derivation: On-Line Analytical Processing,” in *Handbook on Knowledge Management*, (C. W. Holsapple, ed.), Vol. 2, Ch. 40., pp. 141-156, Springer-Verlag: Berlin.

Davenport, D.L., et al., (2005 Oct) “Preoperative Risk Factors and Surgical Complexity are More Predictive of Costs than Postoperative Complications: A Case Study Using the National Surgical Quality Improvement Program (NSQIP) Database,” *Annals of Surgery*, Vol. 242(2), pp. 463-471.

Dimick, J.B., Chen, S.L., Taheri, P.A., et al. (2004) “Hospital costs associated with surgical complications: a report from the Private-sector National Surgical Quality Improvement Program.” *Journal of the American College of Surgeons,* Vol. 199(4), pp. 531-537.

Donabedian, Avedis, (2003) *An Introduction to Quality Assurance in Health Care*, Oxford University Press: Oxford, p. 46.

Fink, A.S. et al., (2002) The National Surgical Quality Improvement Program in Non-Veterans Administration Hospitals: Initial Demonstration of Feasibility, *Annals of Surgery*, Vol. 236(3), pp. 344–354.

Green J., and Wintfeld N., (1993) “How accurate are hospital discharge data for evaluating effectiveness of care?” *Medical Care,* Vol. 31, pp. 719-31.

Gruber, T.R., (1993) “A Translation Approach to Portable Ontologies,” *Knowledge Acquisition*, Vol. 5(2), pp. 199-220.

Hanley, J. A. and McNeil, B. J., (1983) “A Method of Comparing the Areas Under Receiver Operating Characteristic Curves Derived from the Same Cases,” *Radiology*, Vol. 148, pp. 839-843.

Hannan E. L., Racz M. J., Jollis J. G., and Peterson E. D. (1997) “Using Medicare Claims Data to Assess Provider Quality for CABG Surgery: Does it Work Well Enough?” *Health Services Research,* Vol. 31, pp. 659-78.

Harlan, Bradley J., (March 2001) “Presidential Address: Statewide Reporting of Coronary Artery Surgery Results: A View from California”, *Journal of Thoracic and Cardiovascular Surgery*, Vol. 121(3), pp. 409-417.

Harrell, F. E., Jr., Lee, K. L., Califf, R. M., Pryor, D. B., and Rosati, R. A., (1984) “Regression Modeling Strategies for Improved Prognostic Prediction.” *Statistics in Medicine,* Vol. 3 (2), pp. 143-152.

Holsapple, Clyde W., and Whinston, Andrew B., (1996) *Decision Support Systems: A Knowledge-Based Approach*, Course Technology, Cambridge MA, pp. 105.

Hosmer, David W., and Lemeshow, Stanley, (2000) *Applied Logistic Regression* (2^{nd} Ed.), John Wiley and Sons, Inc.: New York.

Iezzoni L. I., Daley J., Heeren T., Foley S. M., Fisher E. S., Duncan C., Hughes J. S., Coffman G.A. (1994 Jul) "Identifying complications of care using administrative data." *Medical Care*, Vol. 32(7), pp. 700-715.

Garson, G. David, (2005) "Logistic Regression", in *PA 765 Statnotes: An Online Textbook*, (G. David Garson, ed.) accessed at http://www2.chass.ncsu.edu/garson/pa765/statnote.htm on 12/19/2005.

Iezzoni L.I., Heeren T., Foley S. M., Daley J., Hughes J., Coffman G. A. (1994 Oct) "Chronic conditions and risk of in-hospital death." *Health Services Research,* Vol. 29(4), pp. 435-460.

Iezonni L. I. (2003) Reasons for Risk Adjustment. (2003) In: Iezzoni LI, ed. "Risk Adjustment for Measuring Health Care Outcomes." Health Administration Press: Chicago.

JCAHO - Joint Commission for the Accreditation of Healthcare Organizations (2004) "Quality Check", accessed at http://www.jcaho.org/quality+check/ on 11/4/2004.

Jollis, J.G., Ancukiewicz, M., DeLong, E.R., Pryor, D.B., Muhlbaier, L.H., and Mark, D.B. (1993) "Discordance of Databases Designed for Claims Payment Versus Clinical Information Systems: Implications for Outcomes Research," *Annals of Internal Medicine*. Vol. 119, pp. 844-50.

Khuri, Shukri et al. (1997) "Risk Adjustment of the Postoperative Mortality Rate for the Comparative Assessment of the Quality of Surgical Care: Results of the National Veterans Affairs Surgical Risk Study." *Journal of the American College of Surgery*, Vol. 185, pp. 325-338.

Kim, Y. J., Kishore, R., and Sanders, G. L., (2005 Oct) "From DQ to EQ: Understanding Data Quality in the Context of E-Business Systems," Communications of the ACM, Vol. 48(10), pp. 75-81.

Koh, C.E., and Watson, H.J., (1998) "Data Management in Executive Information Systems," *Information & Management*, Vol. 33(6), pp. 301-312.

Koutsoukis, N., Mitra, G., and Lucas, C.A., (1999) "Adapting On-Line Analytical Processing For Decision Modeling: The Interaction Of Information And Decision Technologies," *Decision Support Systems*, Vol. 26, pp. 1-30.

Kumar, A. and Palvia, P., (2001) "Key Data Management Issues in a Global Executive Information System," *Industrial Management and Data Systems*, Vol. 1(3/4), pp. 153-164.

Leapfrog Group (2004) "Our Mission" accessed at http://www.leapfroggroup.org-/about_us/our_mission on 11/4/2004.

____________. (2005 Oct) "Company Spotlight: Pitney Bowes," *MarketWatch: Global Roundup*, Vol. 4(10), pp. 220-225.

Redman, Thomas C., (2005 Oct) "Data: A Sea Change Ahead," *Wall Street and Technology,* Vol. 23(10), pp. 58-59.

Reed, Andrea and Catterall, Miriam, (2005) "Invisible Data Quality Issues in a CRM Implementation," *Database Marketing and Customer Strategy Management*, Vol. 12(4), pp. 305-314.

Surgical Care Improvement Project (SCIP), "Preventing Surgical Complications," accessed at http://www.medqic.org/scip/scip_impsurgcare.html on March 11, 2006.

University HealthSystem Consortium (UHC), (2004) *About UHC*, Accessed at http:// www.uhc.edu/ results_list.asp?folder=WEB/ About_UHC/ on 10/22/2004.

University HealthSystem Consortium (UHC), (2005) "Risk Adjustment Methodology." Accessed at https://cimprod.uhc.edu/ pv/include/ document/ Main%20risk%20adjustment %20text%202005.DOC on November 10, 2005.

Van Lohuizen, C.W., (1986) "Knowledge Management and Policy Making," *Knowledge: Creation, Diffusion, and Utilization*, Vol. 8(1).

Wang, R.Y., and Strong, D.M., (Spring 1996) “Beyond Accuracy: What Data Quality Means to Data Consumers,” *Journal of Management Information Systems*, Vol. 12(4), pp. 5-34.

Wenger, Etienne (May 2000) “Communities of Practice and Social Learning Systems, *Organization*, Vol. 7(2), pp. 225-246.

Wixom, Barbara H. and Watson, Hugh J. (Mar 2001) “An Empirical Investigation of the Factors Affecting Data Warehousing Success,” *MIS Quarterly*, Vol. 25(1), pp. 17-38.

Yin, Robert K. (1984, 2003) *Case Study Research*, Sage Publications: London.

Yoon, Victoria Y. and Aiken, Peter (2000) Managing Organizational Data Resources: Quality Dimensions, *Information Resources Management Journal*, Vol. 13(3), pp. 5-14.

Vita

Daniel Lee Davenport

Born July 1, 1961 in Minneapolis, Minnesota

Education:

June 1996 *Master's in Business Administration*
Gatton College of Business and Economics,
University of Kentucky, Lexington, Kentucky

June 1988 *Diplôme Spécial*
Institute for the Study of Developing Countries,
Université Catholique de Louvain, Louvain-la-Neuve, Belgium

June 1984 *Bachelor's of Science in Mechanical Engineering*
Institute of Technology,
University of Minnesota, Minneapolis, Minnesota

Professional Positions Held:

Scientist II, Director of the Office of Decision Support, 2003-2006
Department of Surgery, University of Kentucky, Lexington, Kentucky

Associate Department Administrator, Director of the Office of Decision Support, 1997-2003
Department of Surgery, University of Kentucky, Lexington, Kentucky

Financial Analyst, 1995-1997
Department of Surgery, University of Kentucky, Lexington, Kentucky

Divisional Administrator, 1992-1995
Department of Surgery, University of Kentucky, Lexington, Kentucky

Assistant Director, 1988-1991
Foundation for the Development of the Sankuru, Kinshasa, Democratic Republic of the Congo

Publications:

Journal Articles

Davenport, D.L., Bowe, E.A., Henderson, W.G., Khuri, S.F., and Mentzer, R.M., Jr., (2006)"National Surgical Quality Improvement Program (NSQIP) Risk Factors Can Be Used To Validate American Society of Anesthesiologists Physical Status (ASA PS) Classification Levels," *Annals of Surgery*, In press for May, 2006.

Davenport, D.L., Henderson, W.G., Khuri, S.F., and Mentzer, R.M. Jr., (2005 October) "Preoperative Risk Factors and Surgical Complexity Are More Predictive of Costs Than Postoperative Complications: A Case Study Using the National Surgical Quality Improvement Program (NSQIP) Database," *Annals of Surgery*, 242(2), pp.463-471.

Book Chapters

Davenport, D.L. and Holsapple, C.W., (2005) "Knowledge Organizations," In David G. Schwartz (Ed.), *The Encyclopedia of Knowledge Management,* Hershey, PA: Idea Group.

Davenport, D.L. and Holsapple, C.W., (2005) "Social Capital Knowledge," In David G. Schwartz (Ed.), *The Encyclopedia of Knowledge Management*, Hershey, PA: Idea Group.

Davenport, D.L. and Sena, M., (2003) "Knowledge Derivation Technologies: Online Analytical Processing," In C. W. Holsapple (Ed.), *The Handbook of Knowledge Management*, Berlin/Heidelberg: Springer-Verlag.

Scholastic and Professional Honors:

Research Employee of the Year, University of Kentucky College of Medicine, 2003

Daniel L. Davenport

February 10, 2006