STATISTICAL PROCESS CONTROL AS QUANTITATIVE METHOD TO MONITOR AND IMPROVE MEDICAL QUALITY

by

Kevin Edward Driesen

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PSYCHOLOGY

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2004

The University of Arizona ®
Graduate College

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Kevin Edward Driesen

entitled Statistical Process Control as Quantitative Method to Monitor and Improve Medical Quality

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

Lee Sechrest	4/28/04 date
[signature]	4/28/04 date
[signature]	4/28/04 date
Michael D Lebowitz	4/28/04 date
[signature] (T. SANTIAGO)	4/28/04 date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

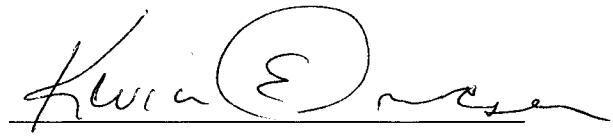
Lee Sechrest 4/28/04

Dissertation Director: date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: 

ACKNOWLEDGEMENTS

The five faculty members who served on the Dissertation Committee were very generous with their time and knowledge. I am extremely grateful to:

Aurelio Figueredo, Ph.D.	Department of Psychology
Mike Lebowitz, Ph.D.	Epidemiology Graduate Program
John Ramberg, Ph.D.	Department of Systems and Industrial Engineering
José Santiago, MD	Department of Psychology
Lee Sechrest, Ph.D.	Department of Psychology

To the many instructors, students, colleagues, and co-workers who made it possible to complete this dissertation, thank you very much.

Lee Sechrest, José Santiago, and Mike Lebowitz have provided intellectual and professional guidance spanning a period of years. Their teaching has transformed my work, and that of many others. Bernard of Chartres spoke of them when he said... *"We are like dwarfs on the shoulders of giants, so that we can see more than they, and things at a greater distance, not by virtue of any sharpness of sight on our part, or any physical distinction, but because we are carried high and raised up by their giant size."*

Northwest Analytical, Inc. provided an evaluation copy of the NWA Quality Analyst software that was used to carry out and validate some of the analysis.

Finally, I'd like to express appreciation to Sharon Knutson and to Dr. Farrokh Alemi, at George Mason University, for sharing their expertise.

DEDICATION

This work is dedicated to my wife, Maria Antonieta Driesen Ortiz, and children, Frederick Alexander, Edward Ramon, and Anna Maria del Sol. A great deal of effort was spent pursuing the doctorate; we can now enjoy, even more, the pursuit of family excellence.

Frederick Emile and Virginia Adele Driesen, have always provided encouragement. It is all thanks to you, Mom and Dad.

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued

TABLE OF CONTENTS - Continued

LIST OF TABLES

LIST OF TABLES - Continued

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS - Continued

ABSTRACT

Statistical Process Control (SPC) methods, developed in industrial settings, are increasingly being generalized to medical service environments. Of special interest is the control chart, a graphic and statistical procedure used to monitor and control variation. This dissertation evaluates the validity of the control chart model to improve medical quality. The research design combines descriptive and causal comparative (ex-post facto) methods to address the principal research question, *How is the control chart model related to medical quality?* Hospital data were used for patients diagnosed with Community Acquired Pneumonia (CAP). During the initial research phase, five medical quality "events" assumed to affect CAP medical quality indicators were pre-specified by hospital staff. The impact of each event was then evaluated using control charts constructed for CAP quality indicators.

Descriptive analysis was undertaken to determine whether data violated the statistical assumptions underlying the control chart model. Then, variable and attribute control charts were constructed to determine whether special cause signals occurred in association with the pre-specified events. Alternative methods were used to calibrate charts to different conditions. Sensitivity was computed as the proportion of event-sensitive signals.

The descriptive analysis of CAP indicators uncovered "messy," and somewhat complex, data structure. The CAP indicators were marginally stable showing trend, seasonal cycles, skew, sampling variation and autocorrelation.

Study results need to be interpreted with the knowledge that few events were evaluated, and that the effect sizes associated with events were small. The charts applied to the CAP indicators showed limited sensitivity; for three chart-types (i.e. XmR, Xbar, and P-charts), there were more false alarms than event-associated signals. Conforming to expectation, larger sample size increased chart sensitivity. The application of Jaehn Decision Rules led to increases in both sensitivity and false alarm. Increasing subgroup frequency from month, to week samples, increased chart sensitivity, but also increased data instability and autocorrelation. Contrary to expectation, the application of hybrid charting techniques (EWMA and CUSUM) did not increase chart sensitivity.

Study findings support the conclusion that control charts provide valuable insight into medical variation. However, design issues, data character, and causal logic provide conditions to the interpretation of control charts.

CHAPTER 1. - INTRODUCTION

Quality assessment first emerged as a medical sub-specialty in the mid-1960s (Brook, 2000). Interest in medical quality has increased exponentially since mid-1990s, leading the Institute of Medicine (IOM) to publish reports on accessibility to care (2001 and 2001), effective use and safety of care (1999 and 1999), and medical quality (IOM, 2000 and 2001). The 1999 Report, *To Err is Human*, has been widely cited for its critical analysis of patient safety, providing estimates of adverse hospital events (2.9 - 3.7% of hospitalizations) and iatrogenic treatment (44,000 - 98,000 annual deaths). The 2001 Report, *Crossing the Quality Chasm*, concluded, "Quality problems are everywhere, affecting many patients. Between the health care we have and the care we could have lies not just a gap, but a chasm."

Today, the debate on how to best address medical quality continues: should it be addressed by educating consumers, improving clinical information systems, implementing organization-level interventions, imposing regulatory structures, providing economic incentives, or changing the medical care delivery system? This study reviews an organizational-level approach to improving medical quality through the application of continuous quality improvement methods.

Medical Quality

Conceptual and operational difficulties have traditionally challenged efforts to improve medical quality. Until recently, the paucity of operational definitions made quality

difficult to measure. Frequently, definitions were conceptual and vague, and quality was a matter of "doing the right thing, and doing it well" (Joint Commission on the Accreditation of Healthcare Organizations, JCAHO) or "delivering the right care, to the right patient, at the right time, in the right way" (Agency for Health Care Policy and Research, AHCPR). One widely cited definition of medical quality was "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current medical knowledge" (Institute of Medicine, 2001). Such definitions make good policy statements but are insufficient guides to measurement. Definitions that focus on the macro level and that use broad conceptual terms are more like goal statements than operational definitions. Table 1. (following page) summarizes medical quality definitions used during the past decade.

The American Medical Association definition introduces the idea of evidence-based practice that places patients at the center of treatment. The IOM definition considers quality from the perspective of both patient and practitioner and views medical care utilization in terms of overuse, underuse, and misuse. Service overuse addresses those medical practice and utilization factors that contribute to both poor quality and excessive cost. Service underuse addresses decision-making and structural conditions (e.g. insurance status or geographic isolation) that impact utilization. Finally, misuse is primarily directed at decision-making that leads to inappropriate use of services (emergency care) or may reflect poor application of technical care. Misuse is most directly associated with concerns of patient safety and adverse events.

Table 1. Medical Quality Definitions

Organization	*Definition / Characteristics*
Institute of Medicine (IOM)	Quality of care is the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge. ♦ **Underuse**: Failure to provide a health care service when it would have produced a favorable outcome for the patient. ♦ **Overuse**: A health care service is provided under circumstances in which its potential for harm exceeds its possible benefits. ♦ **Misuse**: An appropriate service has been selected, but a preventable complication occurs, and the patient does not receive the full potential benefit of the service.
American Medical Association (AMA)	The provision of medical services that: ♦ Produces optimal improvement in the patient's health. ♦ Emphasizes promotion of health and the prevention of disease. ♦ Is provided in a timely manner. ♦ Seeks to involve the patient in decision-making. ♦ Is scientifically based. ♦ Is provided with sensitivity and concern for patient. ♦ Makes sufficient use of technology. ♦ Is sufficiently documented to allow for continuity of care and benefits.
Donabedian	♦ ***Efficacy***: ability of care to improve health. ♦ ***Effectiveness***: degree to which improvement is realized. ♦ ***Efficiency***: greatest improvement at lowest cost. ♦ ***Optimality***: balancing costs and benefits. ♦ ***Acceptability***: conformity to patient preferences: cost, communication, etc. ♦ ***Legitimacy***: Conformity to social preferences (concerning all above). ♦ ***Equity***: Fairness in the distribution of care.
Joint Commission for the Accreditation of Healthcare Organizations (JCAHO)	♦ Doing the right thing, and doing it well.
Agency for Health Care Policy and Research (AHCPR)	♦ Patient can access timely care, receive accurate and understandable information about risks and benefits, are protected from unsafe health care services and products, and receive understandable and reliable information. ♦ Delivery of right care, to the right patient, at the right time, in the right way. ♦ Both patients and clinicians have their rights respected.

During the past decade, medical quality has increasingly been defined in terms of reduced variation and supported by advances in information technology. Specific quality indicators now address every clinical procedure and most service environments. Conclusions from an Expert Panel recently emphasized that the "quality of health care can be precisely defined and measured with a degree of scientific accuracy comparable with that of most measures used in clinical medicine" (National Roundtable, 1998).

A new emphasis has emerged creating another challenge, that of uniformity. An increasing number of benchmarks and clinical guidelines are being promulgated, frequently resulting in inconsistency, contradiction, and turnover. The increase in clinical measures, benchmarks, and guidelines has made it difficult to determine their validity and generalizability. In some areas, the volume of information has become a burden.

Today, medical quality measures typically address the perception of experience, clinical processes, and/or medical outcomes from multiple perspectives. Three types of performance indicators are frequently used to evaluate quality (mortality rates, complications, readmission rates, postoperative infection rates, patient satisfaction and functional performance, etc.); clinical resource utilization (length of stay, medication ordering, laboratory, etc.); and administration (revenue, expenditures, etc.).

Although progress has been made in developing metrics, the measurement of medical quality continues to present fundamental challenges. First, there is a high degree of

variation related to every component of medical care; practice and disease conditions are diverse, as are practitioners, patients, payers, guidelines, technologies, and organizational environments. Second, medical care reflects complex causal processes; a medical treatment may be indirectly related to an outcome and mediated by a variety of factors. Third, a series of operational conditions limits the ability of many providers to generate timely and valid quality data. Furthermore, there are many parties with different interests "pulling" quality standards in different directions.

Medical Variation

Medical services present a rich habitat in which to monitor variation. There are many sources of variation (biological, educational, environmental, legal, geographical, organizational, cultural, and economical), making clinical and administrative processes appear chaotic. Researchers have identified substantial variation in medical treatment, including intensity of hospital utilization, delivery of terminal care, elective surgery according to where people live and physician practice styles (Wennberg, 1998). The term "surgical signature" has been coined to describe these idiosyncratic medical differences in surgical rates for various conditions.

The Dartmouth researchers found signatures for many medical treatment conditions, although less variation was apparent for severe "obvious" conditions, such as hip fracture, that could be characterized by: (1) greater consensus on treatment, (2) virtual certainty of correct diagnosis, and (3) virtual certainty for patients seeking treatment.

Greater variation was observed for less-evident conditions, such as forearm fractures, for which little treatment consensus existed, diagnostic certainty was low, and more factors influenced patient service utilization. The Dartmouth Study found a high degree of variation (greater than that observed in forearm fractures) in 37% of medical, and 18.8% of the surgical, diagnostic-related group conditions.

Medical conditions can be difficult to diagnose and treat. There are patient and practitioner characteristics that influence the treatment process. A variety of environmental mediators (e.g. legal/malpractice considerations, insurance status, urban/ rural setting, etc.) influence medical care. The many sources of variation and almost total lack of environmental controls challenge medical quality measurement. Furthermore, in the clinical environment, subjective quality may be indirectly measured as a latent construct that can vary greatly from patient to patient, practitioner to practitioner, time to time, and setting to setting.

Complex Causality

Medical services have been described as "complex adaptive systems", made up of multiple component parts, and micro-systems, each having "freedom to respond to stimuli in many different and fundamentally unpredictable ways" (Institute of Medicine, 2001). Causal relationships in medical care are complex, indirect, more difficult to measure, and, often, conflicting. Medical treatments may be perfectly delivered but result

in unexpected outcomes, complications, or death due to mediating variables such as a virulent disease agent or high-risk patient condition.

Legal considerations can have powerful and complex mediating effects on clinical treatment. Legal constraints can have simultaneously positive and negative influences that pressure service overuse (defensive medicine) or underuse (service denial). Malpractice liability is critically important and directly influences what information is reported, how it is reported, and who has access to patient records. The powerful influence of medical liability places limits on the quality improvement process.

Operational Limits

Operational difficulties also challenge medical quality assessment. Outcome tracking is difficult as patients move through multiple services; records of even obvious outcomes, such as death, often disappear once the patient leaves the immediate service environment. There are cost and reliability issues associated with data collection. Clinical data, frequently obtained through chart abstraction, can be unreliable and expensive to collect. There are liability and patient confidentiality concerns that create additional barriers. Extracting medical quality data from management information systems is frequently difficult; when data is available in an electronic format, it typically is reported for administrative purposes, serving as proxy indicators for quality (e.g. volume, staffing, utilization).

The current status in the typical medical organization is one in which patient records are paper-based, require a lot of time to update, include limited data fields, and have legal considerations regulating their use. Medical encounters are information-heavy, and medical records are information-thick (including patient, practitioner, and billing information), making them cumbersome and expensive to abstract. Time constraints, practitioner characteristics, reporting guidelines, the patient encounter, and legal considerations influence information recording. Even with elaborate coding schemes and transcription experts, the reliable recording of medical and billing information presents an important challenge.

Due to the expense and technical ability of providers to obtain quality-related clinical information, it is likely that existing data streams will be used to monitor quality processes. This means that "secondary" data sources (e.g. administration, billing, and mandated disease reports) may be used to evaluate medical quality. This may be economical, but it challenges the quality measure's internal validity.

Quality assessment may also be challenged by low patient volume that restricts the provider's ability to interpret statistical analyses. Diseases and clinical complications are frequently (and fortunately) rare events. This means that many medical conditions suffer from the statistical disease of low frequency. While non-parametric methods can be used to study small quantities, they are less powerful. Normal probability theory responds to the law of large numbers and, for statistical analysis, small numbers just are not as useful

to generate confident results. For many conditions and in many settings, the medical quality area of interest may be represented by small numbers and reflect slow processes. The primary weakness of medical quality assessment, according to the medical quality pioneer, Avedis Donabedian, is the questionable nature of information about processes and their outcomes (IHI, 2000). It is evident that a medical organization's ability to manage information will determine its ability to improve quality. Good measurement drives reliable and valid statistical conclusions. Measurement issues (e.g. complex causal relationships, sources of variation, information limits, etc.) and statistical concerns (e.g. greater variation, small numbers, etc.) present important challenges.

Level of Analysis

Donabedian, proposed three categorical levels to organize the study of medical quality: Structure – Process – Outcome (Schiff, 2001). Structure refers to how the medical care system is organized reflecting the availability of services; process refers to the actual medical care given; and outcome refers to the consequences of the interaction between the individual and the practitioner. Medical quality measures are typically assigned to each level. According to Campbell, process measures are better indicators of quality of care because they are "common, under the control of health professionals, and may be more rapidly altered. Outcomes are often rare, may lag processes by up to ten years (e.g. management of hypertension), and may be dependent on factors outside the practitioner's control" (Campbell, 2000).

Clinical relationships are frequently considered as an individual dyadic relationship between the practitioner and patient. Even at this micro-level, clinical treatment has two dimensions, technical and interpersonal care, and represents an interaction in which both practitioner and patient characteristics influence the perception of quality. Measuring the treatment experience often includes externalities like front-desk/reception, nursing care, waiting time, out-of-pocket costs, appointment queue, etc. For inpatient care, housekeeping and food service personnel influence the patient experience. Medical quality is challenged by the placement of logical boundaries on processes and environments.

Medical care organizations influence clinical relationships in a variety of ways. For example, physical structure (i.e. building, equipment, location etc.) contributes to the perception of quality. And patient experience may be influenced by a variety of medical system characteristics, such as patient insurance status and reimbursement (by extending or limiting length of stay or not reimbursing for associated pharmacy products). Also, patients may seek and receive different treatment regimens from multiple sources. The construct validity of quality measurement depends on boundary restrictions, necessary because medical and organizational processes are so fluid.

Campbell and Roland developed micro and macro distinctions to describe medical quality (Campbell, et al, 2000). At the micro level, they posit two principal dimensions of medical care quality: access and effectiveness. A significant characteristic is the

placement of clinical and inter-personal care into separate dimensions of clinical practice. This is important, as it emphasizes the duality of clinical quality, having both subjective (patient-centered) perceptions and objective (clinical process) indicators. The clinical relationship is unquestionably influenced by macro-level characteristics. However, it quickly becomes difficult to dis-entangle causal effects exerted at the macro level, making it logically necessary to focus on proximal relationships and restrict the causal field. This issue is especially important when using univariate analytic techniques (i.e. control charts) that are unable to control multiple, sometimes confounding, extraneous variables.

Research Process

Quality control methods, developed in industrial settings, are increasingly being generalized to medical care environments, even though underlying processes are quite different. For example, in the microchip industry, tolerances can be measured to fine distinctions; materials are carefully selected and environments carefully controlled. Medical environments are radically different.

This Study reviews the application of Statistical Process Control (SPC) methods to improving medical quality. Of specific interest is the control chart, a graphic and statistical procedure used to monitor and control variation. The study uses sensitivity analysis to assess the control chart model's validity. The Study's research design

combines descriptive and causal comparative (ex-post facto) methods to address the primary research question: *How is the control chart model related to medical quality*?

Hospital in-patient data for patients who were admitted during a forty one-month period with a diagnosis of Community Acquired Pneumonia (CAP) was used to test control chart sensitivity. During an initial phase, medical quality "events" were pre-specified by hospital staff. The events were assumed to affect the following CAP medical quality indicators (Time-to-Abx., Blood Culture Prior-to-Antibiotic, Influenza and Pneumonia Vaccine Documented, Average Length of Stay, Readmission within 31 days, and Mortality).

An extensive descriptive analysis was undertaken to evaluate fourteen data characteristics (e.g. stationary, autocorrelation, stability, etc.) to assess whether the CAP data would violate statistical assumptions underlying the control chart model. Then, variable and attribute control charts were constructed for each CAP indicator to determine whether signals occurred in association with the CAP-related events. Different types of control charts were constructed using traditional design (Xbar, XmR, and P-charts) as well as charts updated features (e.g. exponential weighting, cumulative sum of deviations). Alternative methods were used during the chart construction process to calibrate charts according to different conditions (e.g. sample size, statistical power, probability rules). Sensitivity was computed as a proportion of event-sensitive signals. Conclusions were drawn from chart sensitivity to the pre-specified CAP events.

Community Acquired Pneumonia was selected because: (1) hospital-based initiatives had been undertaken to improve outcomes and support the use of clinical guidelines; and (2) CAP quality data have been reported for regulatory purposes, according to the Center for Medicaid and Medicare Services 5-7th Scope of Work. CAP represents a good example of a complex medical condition that can be difficult to diagnose and treat, according to patient conditions, and for which treatment standards are evident.

Kaplan wrote that methodology is the "ultimate source and ground of the norms of scientific inquiry" (Kaplan, 1998). Method validity should not be assumed but rather demonstrated by logical and empirical means. Empirical testing of quality improvement methods offers the opportunity to identify how methods could be modified and provides a confidence interval for the interpretation of results. It seems important to evaluate how different operational environments and subject conditions influence quality improvement methodology. This kind of review supports good practices and is consistent with the quality assessment of quality improvement methods. Empirical review of quality-improvement methods is a good fit, as "criticism is the mother of methodology" (Abelson, 1995).

CHAPTER 2. - LITERATURE REVIEW

This section introduces Statistical Process Control (SPC) as a set of quality improvement methods developed in industrial settings and more recently generalized for medical processes. Attention is directed to the control chart, SPC's primary tool used to evaluate process variation. The control chart's logical and statistical conditions are reviewed. Finally, a summary review of community acquired pneumonia is presented to provide a medical context for the Study.

Statistical Process Control

SPC encompasses a set of improvement methods used to "understand, monitor, and improve process performance over time" (Woodall, 2000). It incorporates statistical, analytic, and managerial processes that: (1) organize team support for the improvement process, (2) model the specific production process, (3) specify a quality measurement strategy, (4) monitor process variation using control charts, (5) identify special causes associated with non-random time series, and (6) control production through the elimination of special causes.

An extensive body of literature documents the successful application of SPC to improve different types of industrial processes in different countries, spanning seven decades (Walton, 1986). SPC tools are "user-friendly" and can be managed on the production line. One expert writes, "This is not a theory; this has been proven time after time" (Wheeler, 2001).

However, questions remain about the robust nature, operation, and appropriate analysis of control charts (CC), perhaps the most prominent of SPC's statistical tools. Control charts enable visual and statistical analysis of time series data. As a descriptive tool, the CC enables the practitioner to visually monitor patterns in the production process. As an inferential tool, the control chart establishes baseline, and, as additional data are added during subsequent observations, probability tests are applied to distinguish *special* from *common* cause. When the data series has been shown to be stable or controlled, prediction intervals can be calculated. The control chart uses historical observations to infer to future production outcomes.

A recently published series offers a critical review of control charts (Woodall, 2000). The series frames a debate over "applications versus statistical modeling" and highlights specific disagreements about statistical assumptions (e.g. normality and data independence), sensitizing rules, hybrid methods, analytic interpretation, and computation standards. The debate can be characterized on a continuum on which on one side are claims that SPC is fully adaptable across all kinds of conditions while on the other side are claims that it generates spurious results and false confidence. Some of the disagreement is general, reflecting a division between application and theory, practitioner and professor. These arguments can be traced back to Dr. Walter Shewhart, who wrote that control methods, "cannot be shown to exist by theorizing alone, no matter how well equipped the theorist is in respect to probability theory. We see in this situation the long recognized dividing line between theory and practice... In other words, the fact that the

criterion, which we happen to use, has a fine ancestry of highbrow statistical theorems does not justify its use. Such justification must come from empirical evidence that it works" (Shewhart, 1931).

Historical Development

The two people recognized for having developed and promoted SPC worldwide are Walter Shewhart and W. Edwards Deming (Shewhart, 1931 and 1939, Deming 1950 and 1982). Shewhart's seminal work on SPC was published in 1931, although his work at Bell Telephone Laboratories introduced the control chart concept in 1924 (Montgomery, 2001). Shewhart developed the fundamental logic of SPC, advanced measurement practices, and developed control chart methods. Deming contributed in the areas of sampling theory and research design; however, he is also recognized for incorporating management theory and promoting worldwide dissemination of SPC through his work to re-build industry in post-war Japan. Deming is frequently cited for his "fourteen points for management" which include the elimination of mass inspection and work standards that prescribe quotas and numerical goals (Wheeler, 1995). Deming considered improvement a fundamental personal and corporate value that requires "constancy of purpose" at all times, and in all aspects of production. He assigned responsibility for product quality to corporate management, estimating that 80 - 85% of all product-related quality problems involve production systems (DeVor, 1992). Deming is also recognized for placing emphasis on consumer perspective in the determination of quality. He considered the consumer an essential part of the production line, emphasizing that products should respond to consumer needs, present and future. (Deming, 1982).

SPC has conceptual roots in probability theory. Shewhart and Deming, however, distinguished SPC from traditional statistics in their discussions of analytic and enumerative research (Deming, 1950; please refer to Appendix A.). Analytic studies use practical methods that respond to environmental realities apparent in many industrial processes; they infer to a future production process for the purpose of assessing process capability (Ramberg, 2001). The Analytic Study has no fixed, static inferential population; rather, it infers to a future process. In contrast, Enumerative Studies have a well-defined population that provides an inferential target using traditional sampling methods. Enumerative Studies represent pure scientific research carried out under controlled conditions. The Analytic/Enumerative dichotomy distinguishes SPC from other quantitative methods and provides a framework for understanding its application (i.e. using smaller samples, inferring to future process populations, and preferences for managerial pragmatism over statistical purism).

The Quality Construct

Quality has been defined in the engineering sense as "fitness for use," referring to the product's ability to achieve its intended purpose (Montgomery, 2001). The language derivation is from the Latin word *qualitas:* what kind or how constituted. Quality has also been defined as (1) Any of the features that make something what it is, or (2) The degree of excellence that a thing possesses (New World Dictionary, 2nd Edition). The first meaning shows the objective nature of quality as an attribute of the product or service; the second introduces the evaluative, subjective nature of judgement.

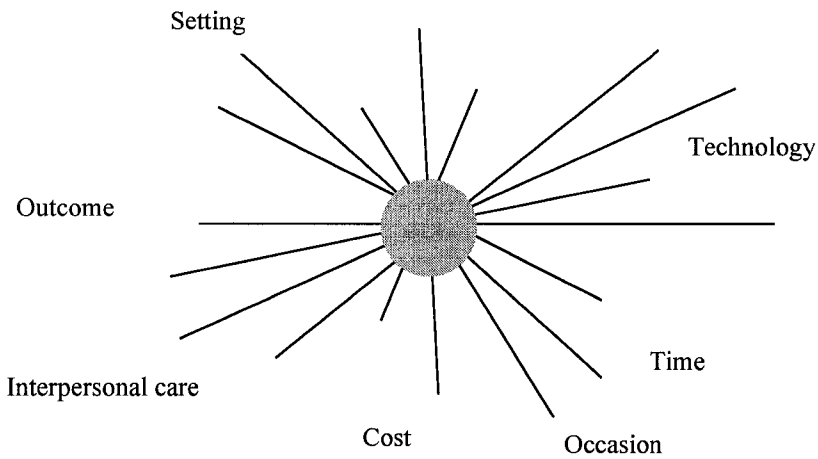
That quality can be evaluated from both objective and subjective perspectives is important. Each perspective requires a different measurement strategy. Objective quality can be reliably measured independent of time and, in the abstract, has a constant value (Shewhart, 1939). Subjective quality is more difficult to measure given differences in perceived value based on individual proclivity. Objective measurement is more concrete and directly observable, while subjective measurement relies on indirect methods of observation. An example of measuring objective quality would include counting the number of scratches on a paint surface or reporting consumer queue time. An example of indirect measurement of subjective quality includes consumer/patient satisfaction. Objective and subjective quality measurement require different approaches, although the same product or service may be evaluated using both.

Subjective quality represents a construct, or abstract variable. Constructs vary in complexity and may have "fuzzy edges," making them difficult to measure. Constructs also vary by the number of observable variables related to their domain and how well defined the indicator variables are (Nunnally, 1978). Subjective constructs tend to have larger domains and be associated with variables having different degrees of association with the underlying domain. This means that subjective quality will likely be measured less reliably and with greater error, challenging construct validity.

Shewhart wrote that quality can be considered in multiple dimensions that exist in "hyperspace" (Shewhart, 1931). These domains may be measured directly or estimated

indirectly, and each may be viewed in magnitude or in relationship to other domains. Establishing quality domains is requisite to better understanding (by sub-dividing) the quality construct. Graphic 1. looks at multiple domains that apply to medical quality.

Graphic 1. Medical Quality Hyperspace



The third-person perspective crosscuts every domain, *who* interprets quality, *who* defines standards, *who* is measuring, and *who* is watching. One would expect differences according to who establishes the criteria for "excellence" and who records, reports, and acts on the data. In manufacturing, producers specify quality targets in deference to consumer perception. In service industries, quality tends to be defined by consumers, while in medicine, traditionally, the medical practitioner, professional organizations, and federal regulatory agencies more likely define quality.

Most products and services have multiple characteristics or elements that describe quality. This requires a measurement strategy that progressively breaks larger concepts into variables that can be measured. Quality is thereby organized by "dimensions," "parameters," "key quality characteristics," (KQC) and finally defined by "quality process variables" (QPV). Montgomery identified eight dimensions or components of quality that address different aspects of the product. These dimensions are summarized in Table 2.

Table 2. Quality Dimensions

Dimension	*Related Question*
Performance	Will the product do the intended job?
Reliability	How often does the product fail?
Durability	How long does the product last?
Serviceability	How easy is it to repair the product?
Aesthetics	What does the product look like?
Features	What does the product do?
Perceived Quality	What is the reputation of the company or its product?
Conformance to Standards	Is the product made exactly as the designer intended?

The quality construct is fraught with different meanings, making it necessary to break these into smaller concepts for measurement purposes and make it operational. Quality must be understood before it can be measured well, and it "must be measured to be controlled or improved" (Blumenthal, 1998). An important step in the measurement process involves identification of the Key Quality Characteristic (KQC) (Carey, 1995). The KQC provides the primary target toward which improvement is directed. Specifying the quality characteristics of a product or service is an essential part of the quality improvement process. Montgomery discusses quality characteristics as "parameters" that can include the Physical (length, weight, voltage, viscosity); the Sensory (taste appearance, color); and Time Orientation (reliability, durability, serviceability) (Montgomery, 2001). For services, quality parameters could be specified in relation to Cost (direct and indirect); Effectiveness (improvement, convenience, sufficiency, accuracy); Perceptual (courtesy, satisfaction, taste, appearance, cleanliness); and Timing (reliability, timeliness).

A subsequent step in the quality measurement process, taken after the KQC is identified, involves specifying the Quality Process Variables (QPV). The QPV is the measured indicator of quality that will be monitored using control charts. It can be considered a process variable, because it is followed over time and is measured as a dependent variable whose value is determined by the production process. Quality assessment involves a series of analytic (deductive) steps taken to refine and direct the measurement process.

KQC and QPV identification is at the base of developing operational definitions that clearly specify evaluation standards in a manner that can be agreed upon, and acted on, by members of the production system to improve the production process. Implementing quality improvement requires production processes be made standard. The flowchart, one of SPC's "magnificent seven" tools, can be used to describe and model the production process as a precursor to making process operations uniform. Once processes are made standard, action can be directed to change the QPV that will, in turn, improve the KQC. Standardizing the medical care production process and establishing operational definitions could present the greatest challenges to the application of SPC methods in the medical environment (Gin, 2001).

Random Variation and Causal Processes

Production creates variation. In the manufacturing environment, variation may be associated with the production process, materials, operators, tooling, quality control, methods (procedures), engineering, management (systems or organizational errors), the environment and/or the measurement process. If products measure outside tolerance limits, the product is rejected or returned for re-production. Product quality is evaluated relative to a specification concerning an acceptable "level of excellence." Quality is measured "proportional to variability" where improvement involves the "reduction of variability in processes and products" (Montgomery, 2001).

SPC emphasizes the distinction between two types of variation: noise and signal. Noise is apparent when time series residuals are randomly distributed, while signals reflect an underlying cause that can be detected in the data structure. The concept of randomness is important, as product quality is measured as a random variable that differs according to underlying causal mechanisms (Wheeler, 2000). In SPC, the control chart is the tool used to distinguish signals from noise.

Sequential order provides an additional analytic dimension for time series data. Time organizes measurement and assists the evaluation of random order and process control. Series values establish the expected level of variation and enable prediction into the future. Current values (together with neighbors, when using zone rules) are used to identify non-random conditions. The sequence of product measurements reflects either random or non-random processes that can be empirically tested using control charts. In writing about a "random order for infinite sequences," Shewhart concluded that, "it appears hopeless to define random order in a useful way for a specific sequence. Instead, the only operationally verifiable way to define random order is in terms of some random operation. A random sequence in this sense is then simply a member of an infinite class of sequences obtainable through repetitions of the chosen random operations" (Shewhart, 1931). Random sequences produce predictable variation.

The SPC practitioner uses statistical monitoring and managerial interventions to improve and control production processes. Control implies that things happen the same way on a

regular basis (even allowing for statistical variation) (Benneyan, 1998). The ability to predict a process demonstrates control of that process; statistical control implies physical control. "A phenomenon will be said to be controlled when, through use of past experience, we can predict at least within limits, how the phenomenon may be expected to vary in the future" (Shewhart, 1939). The inferred population is future-oriented.

Production processes can be viewed as a causal stream; when sequential samples are taken from that stream, and the production process is in control, the sample means will form a distribution hovering around its grand average. Production embodies an activity system where multiple inputs "cause" associated outputs. Shewhart proposed two kinds of causes: *common and special* (or assignable), both of which contribute to product variation. More recently, Wheeler (2000) referred to these causal forms of variation as noise (common cause) and signal (special-cause).

Common-cause variation (noise) is an inherent part of every production process; it is routine, random, and due to natural, regular, or ordinary causes that affect all production outcomes. Common cause processes produce quality values with no discernable pattern. When only common cause is present, the process is in statistical control, and future values can be predicted within limits. That the process is in-control does not mean that it is functioning properly, according to specification. Rather, it implies that large-scale changes to the production system itself are necessary to achieve any reduction in

variation. Incremental attempts to reduce variation may be considered tampering, making the production process more difficult to control. Tampering leads to process instability.

Special-cause variation (signal) is observed as a non-random sequence reflecting the lack of statistical control. When special causes are present, production quality values have been shown to violate probability limits or demonstrate non-random patterns. Special causes reflect irregular or unnatural events not inherent to production; when detected, they should be remedied, because the underlying problem is likely a precursor to even greater variation. Special cause can be assigned directly to the work process, to the employee, or to a machine.

The distinction between common and special cause is "not always sharp" (Quesenbery, 1997). The separation between the two causal forms involves only degree of magnitude. The special cause is a data point that exceeds an arbitrarily established probability limit (e.g. 99.7% certainty), whereas a point close to but located within the probability (control) limit, would be considered common cause variation.

Rational Subgroups and Sample Size

Several study design issues need to be addressed when planning an SPC analytic study. Recommended sampling processes are quite different from those in other quantitative methods. Deming, an expert on sampling methods, proposed that every observation

involves a sampling process, even when considering total populations. He wrote, "Even if we study a complete census, or 100% of a crop in a field in 1960, or all the accounts for a year, or inspect 100% of a month's product, we must interpret the results as one of the samples that the cause system can and will produce, if we hope to reach sensible answers to our problems" (Deming, 1960).

The concept of sampling from an ongoing production stream is important. In SPC, there are four primary sampling concerns: (1) The ability to stratify samples according to causal influence, (2) The role of the CQI practitioner in selecting samples, (3) The practicality of sampling, and (4) Sample size and frequency. Perhaps most important is the ability to stratify subgroups according to causal processes. Fundamental to the logic underlying SPC is the ability to create rational Subgroups (where "rational" means logical and "sub-group" is the set of observations). Production data should be "broken into subsamples upon the basis of human judgement about whether the conditions under which the data were taken are essentially the same or not" (Shewhart, 1939). Rational Subgroups are stratified by causal stream, such as time or setting. For example, in hospital care, emergency department queues may be greater on weekends; work shift differences may be apparent, providing the rationale for sub-grouping.

Traditional methods use a sampling frame that is accessible and is representative of a target, or theoretical population (Gliner, 2001). The SPC sample frame is fundamentally different. In SPC, the ideal is the rational sub-group selected according to underlying

causal processes. Subgroups are selected to "capture process shifts or unnatural variability between, rather than within, subgroups," maximizing differences between subgroups while minimizing differences (due to assignable causes) within subgroups (Benneyan, 1998).

Rational subgroups represent a form of judgement sampling, as expert knowledge about the production process is used to ensure that operational conditions are homogeneous within the subgroup. Production workers can be involved in identifying the rational sub-group, making it a quasi-judgmental, or purposive sample (Carey, 1995). Random sampling will generate probability distributions similar to the population from which they are drawn (when sampling with replacement is used) however, they may not be practical and, looking toward the future, they may not even be relevant. SPC practitioners argue that a rigorous purposive sample is preferable and more economical. When small samples are selected, judgement sampling is more useful in that errors of judgement may be fewer than random errors of probability.

How samples are drawn represents an important SPC Study design consideration. Montgomery identifies two time-oriented methods for selecting samples: (1) taking consecutively produced units, or (2) taking a random sample of all process output over a sampling interval (Montgomery, 2001). The latter method may lead to wider control limits if the process mean drifts between several levels during the interval between

samples. For slower sampling protocols there is the risk associated with process changes between samples, inflating within-group variance estimates.

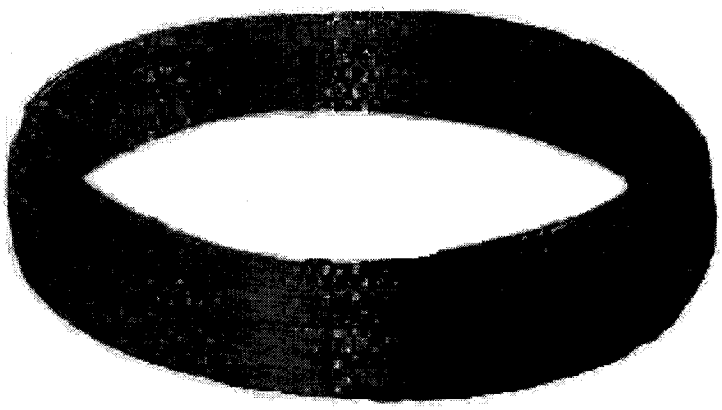
The circumstances in Graphics 2. & 3. (adopted from Shewhart's book) demonstrate practical issues associated with sampling in industrial settings. Graphic 2. shows how selection bias could easily be introduced by selecting easy-to-reach products. In traditional statistical methods, "convenience samples" are discouraged, as they reduce the ability to generalize conclusions (restricting external validity).

Graphic 2. Sampling in the Manufacturing Environment:
Example 1: How should we choose a sample from the poles in this yard?

Item destruction represents another challenge. In Graphic 3., the wire coil would have to be destroyed to sample from its inner core. Practical sampling also presents a challenge in the medical environment. Imagine re-opening the surgical patient to inspect whether the procedure was correctly performed. A related issue concerns sampling with

replacement, which is important to obtaining a true random sample. Sampling with replacement is frequently not possible in medical settings. This can have a large effect on sample size requirements, especially when sampling from small populations.

Graphic 3. Sampling in the Manufacturing Environment
Example 2: How should we choose a random sample of the tensile strength of this coil?



SPC analytic studies rely on small samples. Shewhart considered $n = 4, 5$ for m subgroup samples as economical when sampling from industrial processes, and this remains a SPC "rule-of-thumb for most production processes. In contemporary textbooks, the central limit theorem is used to justify taking small numbers because, over time, the average of the sampling distributions will approach normal, regardless of the shape of the original distribution. Shewhart wrote that larger samples could mask causal processes by

overlapping the effects of different assignable causes (Shewhart, 1939). According to Shewhart, for causal interpretation, the n=1 is the most efficient sub-group size, although unit samples do not produce estimates of within-group variance; with n = 2,3 sizes, there are also variance computation problems, so n = 4,5 became the standard.

Smaller samples may better identify causal change that could otherwise hide in larger data aggregations. Deming wrote, "a small sample taken at frequent time intervals, by showing trends and changes while they are taking place, will furnish much more information than would be furnished at about the same cost by a much larger sample or even a complete count taken at wider intervals of time." Even for a complete census, Deming wrote, "when results are to be generalized to understand the causal systems for future processes, cell frequencies can be unreliable in the sense of having a standard error just as if they had arisen in sampling, as indeed they did" (Deming, 1960).

Sample frequency is another consideration. Both size and frequency impact control chart sensitivity and its ability to identify special causes. A production process can be demonstrated in-control by increasing the span between measurements or by aggregating non-homogeneous data. SPC analytic studies typically draw small samples at frequent intervals. This strategy is considered preferable to a complete census, which, from a traditional perspective, seems contrary to expectation. The sampling strategy is justified by the underlying causal system of what is being counted; looking at the process frequently over time can better identify causal forces.

Control Chart as Quantitative Methodology

Under stable conditions, the control chart can be used to evaluate production process variation. Average and variation control charts are constructed using observed statistics to estimate (μ) and (σ) parameters according to a pre-specified level of confidence (L). Table 3. includes a general model of the control chart (Montgomery, 2001).

Table 3. Control Chart General Model

	Average Chart	*Variation Chart*
	UCL = $\mu_w + L\sigma$	UCL = $L_i\ \hat{\sigma}$
	Center line = μ_w	Centerline = $\hat{\sigma}$
	LCL = $\mu_w - L\sigma$	LCL = $L_i\ \hat{\sigma}$

According to the general model, the process average (i.e. mean, grand mean, proportion, or median) is used to draw the centerline. It represents the baseline process average against which current values are tested to determine if the process remains in-control. The centerline also represents the expected future population average (μ).

Several conditions need to be met for the average statistic to be a good estimator. First, as was previously stated, the process should be stable. Unstable conditions generate unreliable parameters. Stability can be artificially imposed on the process by, for example, "disabling" extreme values from computation of the process average. Dropping

outliers often results in an overly precise estimate that results in unreliable limits. On the other hand, incorporating extreme values that represent special causes can decrease chart sensitivity by creating artificially wide confidence intervals. Either way, if the process average is misrepresented, the chart will likely be erratic.

Estimates of population variation may use the observed range, moving range, or standard deviation taken from each sample as a within-group estimate. Control limits are calculated above and below the centerline based on average within-group variation and the pre-specified confidence level. The estimate of variation used is the within-sample sigma, or, more specifically, the individual sample standard error is computed as: $\sigma_{\bar{x}} = \sigma / \sqrt{n}$.

The sigma parameter requires some clarification. In SPC, sigma refers to all estimates of variation (i.e. standard deviation, standard error, range, etc.) and is adjusted to correct for bias. In practice, the adjustment is made to the L parameter, which is replaced by a constant value before control limits are computed. Because there are different chart types that use different statistics, in application, sigma is adjusted using different constant factors according to chart type, parameter, and sample size. The correction factors (e.g. A, A_2, A_3, C_4, B_3, B_4, B_5, B_6, d_2, d_3, D_1, D_2, D_3, D_4) are identified in SPC textbooks.

The confidence level L is determined according to acceptable risk of making a Type I error. Frequently L is set to encompass 99.72% of all values in the normal probability

distribution, equivalent to +/- 3 standard deviations from the mean. The decision to place control limits at the 3σ level is discretionary, associated with consequences of errors. Some authors suggest using 2σ as warning limits while, in medical care, it has been argued that some medical conditions (e.g. nosocomial infection) justify setting 2σ control limits. Narrow limits will produce more false alarms (i.e. concluding process is unstable when it is under-control) but are justified when the problems associated with signals are serious. Wider limits produce fewer false alarms, but will miss some signals. Setting the *L* parameter involves analysis of cost, benefit, and burden. More recently, the concept of 6σ limits (equivalent to 3.4 defects per million) has been applied as a very rigorous quality standard in manufacturing systems (Elsberry, 2000).

Control charts for process average (i.e. X, $\overline{X}$, p) monitor between-group variation, while charts that monitor variation (i.e. R, mR, S) monitor within-group variation. When a special cause is introduced into the production system, change can be observed in the process average, in the process variation, or in both average and variation. For variable data, both average and variation values are control charted, and special causes can appear on the $\overline{\mathrm{X}} - chart$ and/or the R/S-chart, but not necessarily on both.

The control chart is used to evaluate underlying random conditions. The controlled production process is assumed to reflect random order and, as each observation is plotted on the control chart, the hypothesis of randomness is tested. Each datapoint is significance-tested while the series is tested for non-random behavior using zone rules.

Control of the production process is required to initiate the use of control charts; uncontrolled processes will generate unstable parameter estimates. A preliminary step to constructing the control chart is to create run charts in order to evaluate data stability. Run charts are constructed by plotting quality values across the X-axis (time). There is no minimum number of samples needed to create a run chart. Control (randomness) is evaluated using run tests to determine: (1) whether the series has too few runs (# Runs/m > 33%), (2) whether each run has too many data points, (3) whether statistical trends are apparent, and/or (4) whether zigzag patterns exist.

Deming wrote that for statistics to work, data must "arise from statistical control; until that state is reached, there is no universe, normal or otherwise, and the statistician's calculations by themselves, are an illusion, if not a delusion" (Deming, from Shewhart 1939). Control is apparent when the time series with a sufficient number of data points reflects a random sequence. What is a sufficient number? Shewhart wrote that "it is of far-reaching significance that even after a state of statistical control has been achieved, which is a long process, it is still necessary to have available the results of a thousand or more repetitions of the production process to set valid tolerance limits" (Shewhart, 1931). In practice there is a 25-point rule-of-thumb followed to begin estimating parameters, although for "short run" production processes, as few as 6 data points can be used (Wheeler, 1991). The amount of time necessary to establish control will vary according to characteristics of the process being measured.

The CC has metaphorically been described as a tool that allows the quality practitioner to "listen to the voice of production," implying direct observation of the process. Wheeler suggests that the control chart represents production behavior and recommends it be renamed the "process behavior chart" (Wheeler, 2000). However, the CC allows only an indirect, inductive interpretation of the underlying causal processes. Other SPC tools, sometimes referred to as the "Magnificent Seven" (cause-and-effect diagram, check sheet, control chart, flowchart, histogram, Pareto chart, and scatter diagram), are used to help establish the internal validity of quality improvement processes (Ishikawa, 1972)..

The CC has been mis-represented to have the logical ability to assign cause, which it cannot. The control chart represents observational research and is limited in its ability to detect causes. Causal determination is assisted by the inclusion of other SPC tools, by "ex-post facto" drill-down analysis, and by the prospective use of data to test process changes. The design of experiment (DOE) approach, where "a change is made to a system in a controlled manner for the purpose of learning" is a preferred approach to evaluating causal change in quality processes (Moen, 1999 and Box, 1997).

Aristotle is reported to have said that "an educated man demands no more exactness than allowed by the subject matter being dealt with" (Brodbeck, 1968). Considering the control chart as a model has the advantage of focusing greater attention on activities associated with measurement of the production process. As a model, the control chart tells when to look for unusual circumstances, but it cannot say what happened. The

entire analytic process requires thinking about magnitude of change, cause and effect relations, and other logical conditions. Understanding the production process and having a clear theory of operations will support stronger interpretation of control charts.

Process Modeling

The CC is best viewed as a process model to monitor a "black- box" dependent variable measured as the quality process or outcome. The DV is a "black box" because its value is determined by both observed and unobservable inputs whose effects are not necessarily understood. At each measurement there are two types of independent variable (IV): common and special. Each DV is associated with a common cause IV, and it may also reflect special cause IV. Using the control chart, the analyst looks into the black box to distinguish common from special cause for the purpose of triggering remedial action.

In traditional statistics, multivariate techniques can statistically control for confounding variables and "untangle" complex relationships. Using the univariate control chart, confounding variables are managed (prevented) through design; complex relationships need to be "untangled" before the chart is constructed. This is why design (e.g. homogeneous sub-grouping, sub-group frequency, etc.) is so important. The CC process model represents the one-case, repeated measure time series design. Graphic 4. introduces the process model where common (C) and special (S) cause IVs are observed across time. In this Graphic, the 3rd observation includes both common and special cause.

Graphic 4. SPC Production Process Model

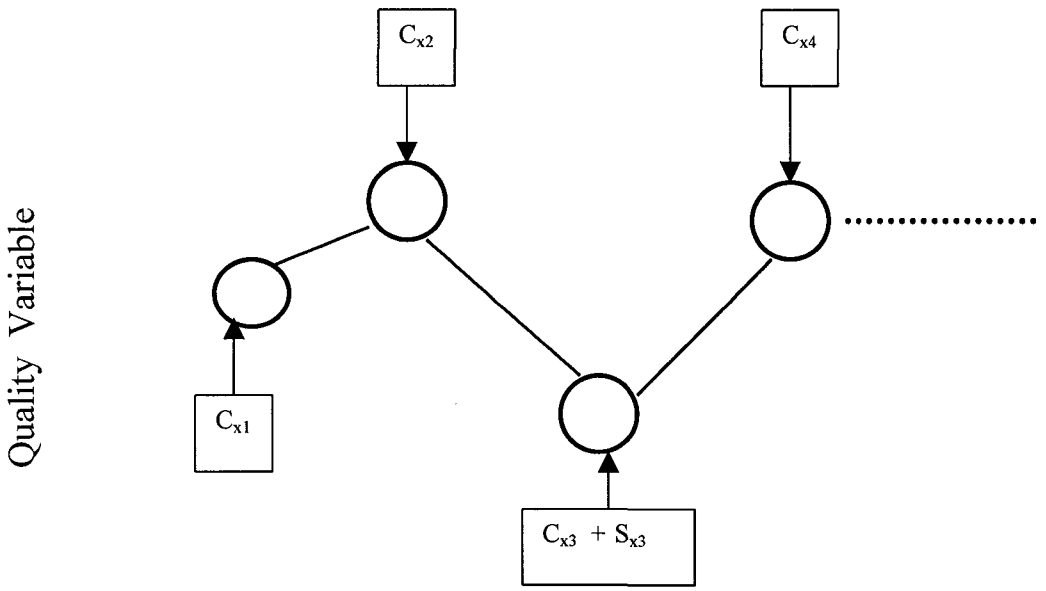


Chart Selection

A series of decisions guide the selection of control charts: (1) quality variable measurement level, (2) sampling, (3) the underlying production process, and (4) probability. Table 4. identifies the most common chart types.

Table 4. Common Chart Types

Name	*Measurement*	*Probability Distribution*	*Parameters*	*Purpose*
X (mR)	Variable & Attribute	Normal	Process mean (Moving Range)	Applied to single observations.
Xbar (S/R)	Variable	Normal	Attribute Std. Dev. Or Range	Traditional Shewhart chart applied to mass production.
P	Attribute	Binomial	Proportion	Fraction Non-conforming.
EWMA	Variable & Attribute		Grand average, Lambda weight	Sensitive to smaller effects. Uses full data series applying weights to recent observations
CUSUM	Variable & Attribute		Reference value. Decision interval	Sensitive to smaller effects. More robust to statistical violations.
Tukey	Variable	Skewed	Median	Applied when outliers common, for underlying skewed distributions
Regression	Variable	Normal	Slope, Std. Dev.	Applied to data series with statistical trend.

Chart selection is guided by whether the quality variable is measured on a continuous (variable) or discrete (attribute) scale. Variable data can take on different values on a continuous scale (e.g. time, height, weight, length of stay, blood sugar levels, # procedures, # discharges, and temperature, etc.). Attribute data are reported in discrete categories according to good or bad (acceptable/unacceptable) specifications. Defects and

defectives are also attribute categories. The defect class is the number of non-conforming events that can happen more than once to the same item (e.g. falls, injuries, number of errors on a dietary tray, etc.) (Amin, 2001). Defects include numerators independent of denominators because the total number of possible occurrences is unknown or unknowable. For example, in reporting the number of patient falls, the number of non-occurrences, (where the patient does not fall), is unknowable. The defective classification is used to report fractions using both the numerator (occurrences) and denominator (non-occurrences) where occurrence count is reported as a percentage or rate (e.g. patient falls per 100 occupied beds).

Counts have "weaker" mathematical properties than do measurements. However, in relation to production processes, discrete (attribute) data have the advantage of creating independent categories whether or not products meet its target specification. Attribute data can also be transformed into variable data; such as for example, plotting time between events for nosocomial infections instead of the actual infection rate.

Two additional conditions that influence the selection and operation of control charts are the sampling strategy and purported effect sizes. Sampling strategies affect control chart operations in several ways. Sampling schemes that result in differing sub-group size challenge interpretation of control limits. Larger samples produce more precise control limits and subgroup size change results in control limit change (Nelson, 1988). Several kinds of adjustments can be made in response to variable sample sizes. Possible

modifications include the use of adjustable control limits, using averaged sample sizes, or creating standardized control charts on which points are plotted in standard deviation units.

Effect size is another factor. Traditional Shewhart control charts are considered sensitive to shifts of 1.5σ or greater. Individual (XmR) charts are less sensitive, as they use smaller sample sizes (n=1) to construct limits. Two chart types recommended for small effects are the EWMA and CUSUM.

Quality outcomes form a probability distribution. According to Montgomery, a distribution is "a mathematical model that relates a variable with the probability of occurrence of that value in the population" (Montgomery, 2001). Frequently, probability distributions used in quality improvement include the normal, the binomial, and the Poisson. Violations to the assumptions underlying these distributions contribute to uncertainty in the conclusions drawn from statistical tests using the distributions.

Chart Analysis

Rules are important when generating standard conclusions. Without rules, different people may interpret visual patterns, like trends and cycles, quite differently. As when reading the Rorschach test, people will interpret patterns according to subjective perception. Here the concern is inter-rater reliability. Visual inspection frequently leads to different conclusions and poor reliability.

The original Shewhart control chart used a single rule (1 point outside the 3σ limit) to identify signals (Davis, 1988). Supplementary rules were developed during the 1950s to increase chart sensitivity to small changes. Seven rules of probability, commonly referred to as the "Western Electric Zone Rules" are frequently used to judge non-random conditions as indicators of special cause. Jaehn proposed a modification of the Zone Rules to simplify the interpretation process (Jaehn, 1987). Jaehn's approach assigns scores to observed data according to its location on the chart; a signal would be identified when the sum of consecutive data values exceeds 8 on either side of the centerline. The Western Electric Rules (#1 - 7), together with the cumulative Zone rules (8 - 11), are identified in Table 5. along with their individual probability of false alarm (Quesenbery, 1997).

Table 5. Chart Sensitizing Decision Rules (Associated False Alarm Probabilities)

Tests to each side of center line:	
1. One point outside the 3-sigma limit	(.0027)
2. 2-of-3 successive points in zone A or beyond	(.00304)
3. 4-of-5 successive points in zone B or beyond.	(.00554)
4. 9 successive points in zone C or beyond on the same side of center line	(.00391)
Tests to entire chart:	
5. (trend) 6-or-more points in a row, steadily increasing or decreasing	(.00277)
6. (saw-tooth) 14 successive up-and-down alternating points.	(.0046)
7. 15 consecutive points in Zone C (+/- 1 sigma)	(.00302)
Zone Control Rules	Score
8. Between target and 1σ limit	1
9. Between 1-2σ limit	2
10. Between 2-3σ limit	4
11. Beyond 3σ limit	8

An important consideration in the use of rules is how rules affect chart sensitivity. Several authors have recommended caution for their routine use (Quesenbery and Montgomery). False alarms are inflated when multiple tests are performed and as the runs are made longer (Quesenbery, 1995). The chart's confidence level and the associated Type I Error (i.e. p =.001, .05, .01) are based on using a single independent (1 point exceeds control limit) test. The application of multiple rules inflates the possibility of error and should trigger more stringent p-levels. False alarm probabilities associated with each rule ranging from .002 - .005. but collectively can increase to .02884 (Quesenbery, 1997). Furthermore, when the run lengths are increased to 100 or more points, the probability becomes 1-test probability.

Control charts are analyzed according to exploratory and confirmatory context, depending on the amount of data available and the existence of statistical control. The use of exploratory and confirmatory analysis expands perspective for the interpretation of chart "evidence." Exploratory conditions exist when charting new processes, when there are fewer start-up observations, or when a process has been unreliable. During the start-up period, charting is carried out to better understand the production process and, as data accumulate, parameters are estimated, assumptions checked, and reliability evaluated. If control is not evident, an incremental approach is taken to identify and remove special causes to better model the production process. Exploratory results provide less confidence; when special causes are identified, they may be considered warning flags for results requiring closer surveillance, not a call to action.

Using confirmatory analysis, a special cause would trigger remedial action when a signal reflects worsening conditions. Confirmatory analysis would also support generalizing observed statistics to predict future outcomes.

Benneynan identified four "more-or-less sequential" charting phases intended to: (1) understand the current process, (2) achieve a consistent level of process performance, (3) monitor process deterioration, and (4) reduce the endemic rate of process variation (Benneynan, 1998). Palm offers another classification framework with three charting phases (Palm, 2001). During the initial chart set-up phase, data are collected or historical data are used to establish preliminary control limits. During the second process improvement phase, real-time charting is applied for the purpose of taking improvement actions, including drilling down to determine special causes. During Palm's third phase, the chart is used to monitor processes, and, if warranted, new charting procedures are applied (i.e. CUSUM or EWMA) according to data and process characteristics. Palm's three charting phases are identified in Table 6.

Table 6. Three Stages of SPC

	Stage	*Timing*	*Statistics*
1.	Chart Setup	Retrospective	Exploratory
2.	Process Improvement	Prospective	Exploratory
3.	Process Monitoring	Prospective	Confirmatory

(Adopted: Palm, 2001)

The CC offers an excellent approach to monitoring process variation. As a graphical procedure, it allows the practitioner to "personally experience" the production process by studying a visual model. The control chart creates a person-chart interaction whereby practitioners can, using confirmatory strategies, directly observe quality improvement processes. This interaction is powerful. "The value of [the] control chart lies not in the novelty of the statistical principles underlying it, but in the ease and reliability with which it converts data into information" (Blumenthal, 1998).

The traditional approach to chart interpretation reflects the SPC paradigm where operators closest to the production process operate, interpret, and act upon the chart. According to SPC theory, this makes feedback more rapid and places responsibility for improvement into the hands of the person directly involved with the process. However, numerous considerations challenge this traditional view, as the time and level of expertise involved in creating and interpreting control charts, pressure to keep production operations moving, and concerns about false-positives all present serious challenges. One author recently discussed SPC in the context of its "de-evolution" from the ideal. In modern environments, Kelly advocates SPC methods be adapted to the socio-managerial environment (Kelly, 2002). The implication is that the management of SPC quality improvement processes will show a degree of variation as methods are applied in different operating environments.

Statistical Assumptions

The control chart is a stochastic procedure used to formally evaluate the probability of an event. Two statistical assumptions justify the application of control charts. The first assumption concerns normally distributed data and the second involves independent observations. Normal distributions are assumed for each sub-group and for time series residuals. The fact that Subgroups are normally distributed is used to justify the amount of data spread within control limits (Bai, 1995). Data series residuals being normally distributed is used to determine process control.

Data independence is important to establishing underlying cause and to computing control limits. Significant autocorrelation implies that the value observed at one time period ($Time_1$) is statistically correlated with the value observed at a subsequent period ($Time_{1-x}$). First lag autocorrelation is diagnosed by correlating an observation with its immediate neighbor; X_1 with X_{1-1}, X_2 with X_{2-2} , and so forth. The essential problem associated with autocorrelation is that it violates the assumption of numeric independence that underlies most statistical procedures (Boyles, 2000). Zhang (2004) writes that autocorrelation affects conditional probabilities; when autocorrelation is high (r=.8), control limits are artificially restricted, which will likely result increase false alarms.

The quality practitioner has been challenged by inconsistencies in the SPC literature in relation to assumption violations. For example, Wheeler states that control charts are fully independent of any distribution assumptions, while other authors have indicated that

control charts (using variances) are quite sensitive to normal departures and can result in a high number of false alarms (Ramsey, 1990). By most accounts, control charting is generally considered robust to violations in normal distribution. The exception is the XmR chart, which does not sample and cannot rely on the Central Limit Theorem. When skewed distributions are expected, alternative charting procedures (e.g. using median values) have been proposed (Carey, 1995).

Control charting is less robust to conditions of non-stationary. Stationary refers to data that vary around a fixed mean. For control charts to be effective, production control is necessary and will produce stationary time series. Non-stationary data will drift in an unstable manner, or demonstrate trend. In either case, the mean average will not provide an adequate measure of the expected value.

In some cases, statistical procedures can be applied to accommodate non-stationary behavior. Quesenbery identifies two kinds of trend. The first reflects special cause and a second represents "necessary trend" where the causal system naturally increases or decreases over time, like monetary inflation. Unnecessary trend would include special causes where, for example a machine not properly calibrated resulting in increasing tolerance values. For production processes that contain necessary trend, regression SPC can be used with slope values used to compute the centerline. Special cause trend is sometimes identified using the 6-datapoint rule (increasing or decreasing values).

The CC, like other statistical techniques, is best interpreted in the context of confidence and risk. It is rooted in probability science and, at some point, obvious violations of statistical assumptions need to be addressed. CC results are conditioned on the statistical assumptions and the magnitude of violation will affect results. Statistical violations will not restrict the QI practitioner from constructing a control chart; rather, he/she will be blinded and (potentially) led to false confidence. If ignored, violations can threaten confidence that provide quantitative estimates and appear to be hard and precise but become soft, imprecise, and perhaps, delusional. It represents good analytic process to review data characteristics before charting.

Continuous Quality Improvement of Medical Processes

In the mid-1980s, a small group of physicians attended Deming's SPC quality improvement seminars. Several of these physicians eventually rose to assume leadership positions in the health care hierarchy, at the Joint Commission for the Accreditation of Healthcare Organizations (JCAHO), the Institute for Healthcare Improvement (IHI), and at major medical centers. Blumenthal documented the process through which SPC was adopted by health care organizations as continuous quality improvement (CQI) (Blumenthal, 1998). The adoption of the CQI "innovation" was rapid and represented a methodological leap across production environments. A survey of U.S. hospitals conducted in 1993 found 69% were considering, or had already implemented, some kind of continuous quality improvement program (Shortell, 1998).

CQI has been described as "the philosophy of continual improvement of the process associated with providing a good or service that meets or exceeds customer expectations. This is accomplished by involving a broad array of organizational members who are trained in basic statistical techniques and tools are empowered to make decisions based on the analysis of data" (Shortell, 1998). Adopting from SPC, CQI is represented as a 4-step cycle, sometimes referred to as the Deming or Shewhart cycle, that involves: Plan – Do – Study – Act. The PDSA cycle parallels the scientific method, with greater emphasis placed on the A...action.

There is a growing body of CQI-related medical literature. Predictably, many early publications involved "how to implement" articles (Chasin, 1998). More recently, authors have described CQI effectiveness in controlling hospital infections and for asthma and diabetic care (Quesenbery, 2000 and Boggs, 1999), but there have been few evaluations using more rigorous evaluation designs (i.e. quasi-experimental conditions) (Blumenthal, 1998). One problem has been the lack of intervention uniformity and program standards, making it difficult to make cross-site comparisons. One study found that, on average, the cost of hospital-based CQI interventions ranged from $148. - $18,590. (Shortell, 1998).

As its name implies, SPC focuses on process, not product (Doty, 1996). However, production processes are not "isomorphic," whereby structural form can be generalized even though products are remarkably different. Engineering methods have been successfully used to improve health care processes (e.g. patient flow analysis, space and information systems planning, inventory control, queuing analysis, forecasting, computer

simulation, facility design, scheduling, and capacity planning) (Benneynan, 1998). The critical issue in generalizing CQI to medical care is not whether to apply, but how to adapt under complex conditions with so much natural variation.

Community Acquired Pneumonia

Community Acquired Pneumonia (CAP) provides an example of medical causal complexity. CAP can be difficult to diagnose, depending on its etiologic agent, patient characteristics, and existing co-morbid conditions. Diagnostic test sensitivity varies. Even chest X-ray, the diagnostic "gold standard," gives time dependent results. Treatment guidelines exist, although the two most-used guidelines, those published by the American Thoracic Society (2001) and by the Infectious Disease Society of America (2002), are not fully consistent.

The diagnosis and treatment of pneumonia is based on medical research that began more than a century ago. Louis Pasteur first isolated the streptococcus pneumoniae (bacterium) in 1881 from the saliva of a patient with rabies. In 1884, the gram stain technique was developed to distinguish pneumonoccal from other causes of pneumonia (Pitluk, 2003). Type-specific antibodies were first developed in 1913 and antibiotics, first developed in the 1930s, were shown to be effective in the 1940s, reducing mortality by 60 - 70% (Metlay, 2003). The first pneumonoccal vaccine was licensed in the USA in 1977.

Definition

Acute pneumonia is defined as an "infection of the parenchyma accompanied by symptoms of acute illness" (Metlay, 2003). The typical presentation of pneumonia involves the inflammation of one or both lungs, caused by bacterial, viral, mycloplasm, fungal, or parasite-related infection that fills air space making it difficult to breathe. Pneumonia-related infection may also invade the bloodstream, the brain membrane, or other organs in which it can infect abscesses. Pneumonia is frequently separated into four categories:

- Community-acquired: during the course of normal life activities,
- Hospital-acquired: during the course of inpatient treatment such as with a medical ventilator or exposure in an intensive care unit (nosocomial infection),
- Aspiration-related: after foreign matter (e.g. vomit) is inhaled, and
- Opportunistic organism-related: striking people with compromised immune systems.

Pneumonia is a special concern for older adults and those with chronic illnesses; however, it strikes young and healthy people as well. Acute pneumonia has been described as the "captain of the men of death" because of its prevalence, and, when occurring in older adults, it is frequently fatal. It also has been referred to as "the old man's friend" because it can lead to a "comfortable" death.

Etiology

There are more than 50 types of pneumonia, including Viral, Pneumococcal, Aspiration, Anaerobic Bacteria, Chlamydia Pneumonie Legionella species and Mycoplasia, Hantavirus Pulmonary Syndrome, and Pneumocystis Carini. Viral pneumonia is the most common and least virulent form; for bacterial pneumonia, the pneumococcus is the most common cause and Streptococcus pneumoniae and Legionella are the most frequent causes of lethal CAP. Pneumococcus can invade different parts of the body. When they invade the bloodstream, they can cause bacteremia; when they invade the brain covering, they can cause meningitis. Pneumococci may also cause otitis media (middle ear infection), sinusitis, and arthritis.

Epidemiology

Pneumonia is a non-reportable disease, which leads to crude approximations of prevalence. Two to four million CAP cases annually result in approximately 10 million physician visits and 500,000 hospitalizations (Bartlett, 1998, Stanton 2003). Overall, prevalence appears to be increasing. Between 1984 and 1995, CAP discharge rates increased by more than 30%. This increase does not appear related to the aging population, aging as age-adjusted mortality also increased (Battleman, 2002). CAP has seasonal and geographic variations with higher rates evident in winter months. The cost of CAP is significant---an estimated $10 billion--- is spent each year, with 92% going to hospital care and $100 million spent for medications.

CAP is the number one infectious cause of death in the USA (and worldwide). An estimated 45,000 people die each year from CAP (Bartlett, 1998 and Stanton, 2003) and from all forms of pneumonia (including CAP), about 90,000 persons die in the USA (Stanton, 2003). CAP-related mortality ranges from 2 - 30% with the crude rate between 12 - 14% (Niederman, 2001). Mortality is influenced by patient characteristics and disease etiology. Younger CAP patients (<65 years) were reported to have a mortality outcome rate of 5.7 - 8% while older patients experienced rates of 24 - 33% (Markowitz, 1996). Eleven patient factors have been identified as having a statistically significant association with mortality, including age, male sex, pleuritic chest pain, hypothermia, diabetes mellitus, hypotension, and neorologic and neoplastic diseases (Fine, 1996).

Co-morbidity challenges the diagnosis of pneumonia. Chronic obstructive pulmonary disease, neoplasms, and congestive heart failure need to be evaluated as potential diagnostic confounders; patients may have simultaneous diseases with both infectious and non-infectious components. One study of 700 pneumonia patients found that more that 60% had co-morbid conditions (Meehan, 2000).

The Pneumonia Patient Outcomes Research Team evaluated outcomes from 2,287 patients in three geographic locations between 1991-1994. (Mortenson, 2002) PORT researchers found pneumonia-related death to be 7.7 times more likely to occur within 30-days of presentation compared to pneumonia-unrelated death and more than 75% of mortality occurred within the first 30 days after the patient was seen. The PORT

researchers also found that mortality unrelated to pneumonia occurred later and with different immediate causes of death.

There are social and demographic characteristics that influence the epidemiology of pneumonia. Ethnic minorities are 3 - 10 times more likely to be diagnosed with pneumonoccus bacteremia, Black Americans have a 3 – 5 time increase in risk, and American Indians have the highest reported rate among American ethnic groups (CDC, 1997). Low-income zip codes had 5.4 times more pneumonia hospitalizations per capita than high-income zip codes (Millman, 1993).

Diagnosis

Symptoms range from mild to life threatening. Among younger populations, pneumonia can be difficult to distinguish from a common cold or flu. Typically, patients may be ill with an upper respiratory tract infection, and, after a few days, suddenly become much sicker with fever, chest pain, shortness of breath, shaking chills, and increased sputum production that can change to rust-colored. The onset of more severe symptoms is usually quite evident and patients can often identify the specific time when "things got worse". Medical examination may identify altered breath sounds or localized rales, fever or hyperthermia, sweats, cough, chest discomfort, patient-reported fatigue, abdominal pain, anorexia, and headache.

According to the American Thoracic Society, the "clinical features of CAP (symptoms, signs, and radiographic findings) cannot be reliably used to establish the etiologic diagnosis of pneumonia with adequate specificity and sensitivity" (Neiderman, 2001). Diagnostic tests are only moderately sensitive. Recently, X-ray diagnostic problems have been identified (Metlay, 2003). One study involving independent and blinded comparisons of 26 confirmed pneumonia cases found that X-ray radiography missed 31% of confirmed cases (Metlay, 2003).

Four laboratory tests have been recommended to confirm the diagnosis of pneumonia. However, each has inherent limits. The agent cause of pneumonia may go unidentified in as many as 50% of CAP patients, even when extensive testing is performed (Niederman, 2001). The four diagnostic tests are:

- Lung biopsy. Needle probes are used to take tissue samples from the patient's lung, produce highly valid results, but are painful. Lung biopsies are recommended only as an alternative diagnostic method for hospitalized patients.
- Chest X-ray. Radiographs are generally considered the "gold standard" to confirm CAP diagnosis. Medical guidelines supported by the Agency for Healthcare Quality and Research and the American Medical Association state that "chest radiography is critical for establishing the diagnosis of pneumonia" (National Guidelines Clearninghouse, 2003). The problem is that X-rays may require specialty reading, and results are time-sensitive (e.g. the appearance of pneumonia may show after the

onset of symptoms). A recent meta-analysis labeled radiography the "imperfect" gold standard and concluded it may not be helpful for all patients (Metlay, 2003).

- Sputum gram stains. Gram stains are a quick method employed to visually confirm the presence of bacteria; however, patients frequently have trouble producing adequate samples or saliva samples are provided that produce false-negatives.
- Sputum and blood culture. Cultures may confirm bacterial pneumonia and identify whether pneumonocci have invaded the blood stream. However, the culture process takes almost twenty-four hours and, at best, identifies causal agents in half the cases.

The Infectious Disease Society of America (2000) recommends the following protocols:

1. Baseline Assessment
    - Chest radiograph for routine evaluation to confirm diagnosis, detect associated diseases, assess severity, and establish baseline to evaluate response.
2. Outpatient Assessment
    - (Optional) Sputum gram stain and culture for conventional bacteria.
3. Inpatient Assessment
    - Complete blood count with differential (before antibiotics).
    - Chemistry panel including glucose and sodium levels.
    - Pretreatment blood cultures (twice).
    - Gram stain and sputum culture.
    - Test for tuberculosis.
    - Alternatives (Bronchoscopy, transtracheal aspiration).

The American Thoracic Society (ATS) disputes the value of sputum gram stains for inpatients, due to the difficulty in generating productive samples, in interpretation, and their non-response to a variety of pathogens (Niederman, 2001). The most recent ATS Guidelines note that "viral cultures are not useful in the initial evaluation ....and should not be performed" (Niederman, 2001).

Clinical Management

Clinical pathways have been established for diagnostic procedures, treatment, patient education and discharge planning (Zynx Health, Inc., 2003). These pathways guide clinical action by specifying performance indicators and recommended treatments.

Disease progression, patient characteristics, treatment location, and medication choice are four factors that guide treatment processes. The disease agent shows an etiologic hierarchy with different forms being extremely aggressive and virulent; these agents require immediate and aggressive treatment. Lower grades of (viral) pneumonia may quickly escalate into higher grades of (bacterial) disease. In addition to agent characteristics, host characteristics are important as older (>64), of male sex, those with co-morbid conditions, or immuno-compromised patients are more vulnerable.

Pneumonia can have serious consequences. For hospitalized patients, the primary clinical objective is successful outcome, while a secondary, but important, treatment concern involves Length of Stay. A second clinical issue concerns the location or site of

treatment. Usually, the primary care or emergency room physician determines the site of care, in consultation with the patient. A general consensus that exists among medical practitioners is to treat high-risk patients in the hospital. However, for younger and lower-risk patients, the benefits associated with alternate treatment locations, inpatient or ambulatory, remain under debate. ATS guidelines state that care location is "perhaps the single most important clinical decision made by physicians during the entire course of illness" (Meehan, 1997).

Research sponsored by the Agency for Health Care Policy and Research found that many low-risk patients could be treated safety on an outpatient basis (Fine, 1997). When low-risk patients are treated at home with appropriate antibiotics, the quality of care can be maintained, patients are more likely to be satisfied, and resources are conserved. It is estimated that more than 75% of pneumonia cases are appropriately treated as outpatients and for younger and healthier patients there is a low risk of mortality (less than 1%). Home-based treatment is also less expensive, as the average cost of outpatient care is $150-$300, compared to $5,700 for inpatient care (Niederman, 2001).

A prediction model for pneumonia was developed to help practitioners evaluate patient risk. Table 7. summarizes the risk prediction model (Fine, 1997). Patients in risk classes I & II do not usually require hospitalization, those in class III may require brief hospitalization, and those in IV & V usually require inpatient care.

Table 7. Prediction Model Scoring System

Patients Characteristics	*Points Assigned*		
Demographic Factors			
Age:			
Male:	Age (in years)		
Female:	Age (in years) - 10		
Nursing Home Resident	+10		
Co-morbid Illnesses		*Laboratory Findings*	
Neoplastic Disease	+30	Ph < 7.35:	+30
Liver Disease	+20	BUN > 10.7 mmol/L:	+20
Congestive Heart Failure	+10	Sodium < 130 mEq/L:	+20
Cerebrovascular Disease	+10	Glucose > 13.9 mmol/L:	+10
Renal Disease	+10	Hemacrit <30 percent:	+10
Physical Examination Findings		PO_2 <60 mmhg (2):	+10
Altered Mental Status	+20	Pleural Effusion:	+10
Respiratory rate ≥ 30/Minute	+20		
Systolic Blood Pressure <90mmHg	+20		
Temperature <35 Degrees C or 40 Degrees C or More	+15		
Pulse 125/Minute or More	+10		

Risk Score

Risk	Risk Class	Based on
Low	I	Algorithm
Low	II	70 or fewer total points
Low	III	71-90 total points
Moderate	IV	91-130 total points
High	V	> 130 total points

The use of prediction models is somewhat controversial. The American College of Emergency Physicians (ACEP) has published guidelines placing the value of prediction models behind clinical opinion and patient social circumstances (ACEP, 2001). One study found that prediction rules can reduce hospitalization, although 30 - 40% of low-risk patients were still hospitalized based on physician judgement in-lieu of prediction rules (Metlay, 2003). This is reflected in a statement made through the ATS that clinical care is still an "art" form, and that appearance of need is sufficient to hospitalize patients at least for an observation period.

After determination is made to admit a patient for inpatient care, he/she needs to be managed through a web of services to ensure appropriate treatment. Antibiotics are the primary treatment for bacterial pneumonia, and recommendations have been established according to disease severity, patient intolerance or side effects, comorbidity and concomitant medications, and treatment location. Clinical decisions around medication are frequently based on three issues: administration, agent/immuno-resistance, and cost.

1. Administration issues concern the time when patients begin and finish antibiotics. Quick administration and completion of the entire antibiotic regimen is essential to successful treatment. Mortality rates for patients infected by penicillin-susceptible strains of S. pneumonia have consistently been reported at ≥ 20%, influenced in part by Time-to-Abx. delays.

2. Antibiotics are considered a miracle treatment and are prescribed for different conditions. The high use of antibiotics has enabled bacterial agents to mutate into antibiotic-resistant forms. Recent estimates suggest that 15% or more of pneumococcal agents are drug-resistant in some locales (VHA, 2000). With repeated use of antibiotics, the patient benefit can also be lowered. While antibiotics have immediate and substantial clinical benefits, their routine prescription has been discouraged.
3. Cost is another challenge. Older antibiotics are equally effective but 30-80% less expensive than many new antibiotics.

Medical Quality

CAP management guidelines have been released by numerous professional associations and healthcare organizations, including the Infectious Disease Society of America (1998 & 2000), the Centers for Disease Control, the American Thorastic Society, CIDS/CTS, the Institute of Clinical Improvement (1999, 2002), the American College of Emergency Physicians (2001), the American Society of American Pharmacists (1999), the Scottish Intercollegiate Guidelines Network (2002), and the Cincinnati Children's Hospital Medical Center (2001). The Infectious Diseases Society of America (1998 and 2000) and the American Thorastic Society guidelines are most frequently cited. Their guidelines include risk grading, diagnostic confirmation, treatment location, antimicrobial therapy, prevention, and performance tracking (Bartlett, 2000).

Two federal agencies, the Agency for Healthcare Research and Quality (AHRQ) and the Centers for Medicaid and Medicare Services (CMS) have had an important influence. AHRQ has developed CAP indicator measure sets to facilitate the choice of guidelines. CMS now requires providers to submit information on selected CAP indicators for those healthcare organizations that seek Medicare reimbursement. The CAP medical quality information being required has been specified in annual Scopes of Work (SOW). Table 8. identifies the CAP quality information required under the most recent CMS 7th SOW, which spans the November 2002 - October 2005 time period.

Table 8. CAP Indicators Included in the 7th Scope of Work Center for Medicare and Medicaid Services (CMS)

1.	Timely antibiotic administration
2.	Initial antibiotic therapy consistent with current guidelines
3.	Collection of blood cultures within 24 hours of hospital arrival
4.	Collection of blood cultures prior to initial antibiotic dose
5.	Screening for influenza and pneumococcal immunization status and vaccination prior to discharge, if indicated.
6.	Smoking cessation counseling during hospitalization.
7.	Arterial oxygenation assessment within 24 hours of hospital arrival.

Process Indicators

Time-to-Abx.

The amount of time that passes before treatment is sought or provided affects many medical conditions and influences medical prognosis. The relationship between time-to-treatment and outcome holds for pneumonia; patients who receive antibiotics quickly are less likely to die and more likely to have a shorter duration of symptoms (as measured through length of hospital stay). Several studies have shown improved survival when antibiotics are administered within 8 hours of admission to a hospital (Battleman, 2002 and Meehan, 1997). The Meehan Study, a multicenter study of 14,000 Medicare patients, found that antibiotic delivery before 8 hours was associated with a 15% reduction in odds of 30-day mortality (95%CI=0.75-0.96) (Meehan, 1997).

Reports of average Time-to-Abx. varies across studies. For emergency department (ED) patients, the average door-to-needle time was 3.5 hours (± 1.4) (Battleman, 2002). A retrospective study of 1700 CAP patients found a median Time-to-Abx. of 4.9 hours, with 76.8% receiving Abx. within 8 hours (Meehan, 2000). Average Time-to-Abx. has been related to prolonged length of stay; patients who experienced longer delays (>9 hours) were almost 2 times more likely to have longer stays (OR=1.75, 95% CI=1.34, 2.29) (Battleman, 2000). Timely Abx. also has been associated with decreased mortality. A retrospective study of 14,000 elderly patients treated for pneumonia at 3,500 hospitals estimated that receiving antibiotics within 8-hours reduced 30-day mortality

(AHRQ, 2003). The Meehan study found that antibiotic delivery of <8 hours was associated with a 15% reduction in mortality.

Time increments are frequently broken into two dichotomies, within 4 or 8 hours. In the original 5th CMS-SOW, the Time-to-Abx. guideline was set to within 8 hours, although in the 7th SOW it has been re-established to within 4 hours. There are a variety of patient and provider factors that influence the time it takes to give antibiotics. Fine, et al. found that timely administration was negatively associated with ethnicity, as nonwhites waited longer to receive the Abx. (African American odds ratio 0.71; 95% CI 0.60 - 0.85, other racial minorities 0.79; 95% CI 0.68 - 0.93) (Fine, 2002). Hospital size may also be a factor, as larger facilities (>250 beds vs. <100 beds) have been shown to take more time to give the Abx. (OR 0.68; 95% CI 0.59 - 0.80). Other conditions that have been investigated include arrival time, staffing ratios, and existing hospital policies.

Antibiotic Selection

Recommendations concerning antibiotic selection depend on whether the causal organism has been identified. When the causal agent is known, the least costly, least toxic and most narrow-scope antibiotic is recommended. When empirical treatment is necessary, recommendations are based on severity of illness, pathogen probabilities, S pneumonie resistance patterns, and comorbid conditions (National Guidelines Clearninghouse, 2003).

The selection of antibiotics has been one area that has reflected considerable variation. The Infectious Disease Society of America has released antibiotic guidelines, although studies have estimated that only 56% of patients were treated with recommended antibiotics. The impact of the appropriate selection of Abx. is less clear. The Battelman Study showed no independent association between Abx. selection and prolonged ALOS (Battleman, 2002).

Blood Culture Taken Prior to Antibiotic

For blood cultures to maintain validity, the culture should be taken prior to the administration of antibiotics. Blood cultures should be performed before antibiotic treatment and sputum gram staining and culture….. "unless these procedures would delay initiation of treatment" (National Guidelines Clearinghouse, 2003). The goal is to proceed from an established diagnosis, because the medical treatment will be more precise and antibiotic use more cost-effective.

Influenza Immunization

Seasonal epidemics of influenza are associated with 20,000 to 40,000 deaths, with about 90% of deaths among those age 65 or older (Bartlett, 2000). The Spanish flu epidemic of 1918 resulted in more than 20 million deaths worldwide. Severe influenza epidemics, when they occur, are estimated to have an annual direct and indirect expense of $12 billion (Federal Register, 2002).

Influenza infects people of all age groups; however, children are the most likely to be infected while older people are the most severely affected, more likely to suffer from severe disease or die. Influenza A and B are the two most common viral types. Both are sub-classified and having immunity to one type or sub-classification does little or nothing to protect people from other types, or sub-classes, of infection.

Immunization is the primary method for preventing influenza and is associated with reductions in influenza-related respiratory illnesses and physician visits among all age groups, hospitalization and death among high-risk groups, otitis media among children, and work absenteeism among adults (CDC, 2002). Immunization can effectively reduce the effect of influenza by 70 - 90% among healthy adults <65 years. A meta-analysis by (Vu et. al., 2002) demonstrated that influenza vaccine reduced hospitalization and mortality due to pneumonia and influenza by 33% (95% CI, 27% - 38%) and 47% (95% CI, 25% - 62%), respectively, for persons >65 years of age. Another study demonstrated that administration of influenza vaccine in persons >65 years of age was effective in preventing respiratory illness, pneumonia, hospitalization, and death. (Gross, 1995).

The Healthy People 2010 goal is to immunize at least 90 percent of all older Americans. Influenza vaccine levels for adults >65 years have increased from 33% in 1989 to 66% in 1999, surpassing the Healthy People 2000 goal of 60% (CDC, 2002).

Pneumonoccal Immunization

About half of the 40,000 annual deaths related to pneumonoccol pneumonia could be prevented through full vaccination coverage. The use of pneumonoccal vaccine has consistently been supported by the American Academy of Pediatrics, the American Academy of Family Physicians, The American College of Physicians, and CDC's Advisory Committee on Immunizations (CDC, 1997). Antibody levels usually last 5 - 10 years following pneumonoccal vaccine. Pneumococcal vaccine is recommended for individuals 65 years and older who have stable immune systems and individuals at increased risk for Pneumococcal disease (VHA, 2000). It has been shown to be 45% effective in preventing hospitalizations in the elderly during peak season. A CMS study conducted in 2000 on 1999 data found that 35% of the population >65 years had received the pneumonoccal vaccine, and 45% of the population had received the flu vaccine (Federal Register, 2002). The same study found that minority populations reported a lower rate of vaccination (Hispanic, 24%, Black, 20%) as compared to Whites (37%).

Provider-based improvement interventions include paper or computer-based reminder systems and practice tracking systems. In tracking systems, providers identify at-risk patients and maintain rosters of those who receive immunization. Physicians who use tracking systems have been shown to administer 30% more vaccines than those not tracking. Reminder systems consist of charts, computers, checklists, and physically posted to remind physicians, nursing staff or other office staff to review patient need for immunization. Evaluations have found that checklists and computer reminder systems

may be associated with increases in pneumonoccal vaccination (from 5-to-50%) (CDC, 1997).

Oxygenation Assessment

The Infectious Diseases Society of America (IDSA, 2000) recommends that all hospitalized patients with CAP should receive blood gas measurement or pulse oximetry. Blood gas measurement or pulse oximetry prior to admission, or within 8 hours of admission, is a recommended performance indicator. Meehan, et al. studied medical charts of 1700 CAP and nursing home patients and found that 94.7% of CAP patients received an oxygenation assessment within 24 hours of hospital arrival (Meehan, 2000).

Medical Stability

The AHRQ has funded research to assess patients' readiness for discharge. Medically unstable patients have been shown to exhibit a 60% increased chance of readmission or death and a 50% higher chance of not returning to their usual activities within 30 days (Stanton, AHRQ). Patient assessment tools have been developed and incorporate six basic measures: temperature, heart rate, blood pressure, respiratory rate oxygen levels, patient mental status, and ability to eat/drink. Using these measures, researchers have found that 1 in 5 discharged patients were considered medically unstable, and that readmission or death was 5 times greater for those patients discharged with 2 or more unstable factors (Halm, 2002). Patients who fulfill one or more of the following instability criteria within 24 hours of discharge have a greater risk-adjusted rate of death

and readmission as compared to those who fulfill none of the criteria: Temperature> 37.8 degrees C, Pulse>100/minute, Respiratory rate > 24/minute, Systolic blood pressure < 90 mm Hg., Oxygen saturation < 90% (Halm, 2002).

Smoking Cessation Advice

The American Thoracic Society (2001) cites smoking as a risk factor for pneumonia and identifies smoking cessation as an important strategy for the prevention of CAP. Documentation of smoking cessation counseling for smokers with CAP is a national performance measure according to the National Quality Forum (2003), a core measure according to the Joint Commission on Accreditation of Healthcare Organizations (2002), and a quality of care measure in the 7th Scope of Work by the Centers for Medicare & Medicaid Services (2002).

Outcome Indicators

Mortality

In a sentinel 1993 study, a hospital system in southern Pennsylvania reported CAP-related mortality reductions of more than 50% and of related expenses by more than 10%. These improvements were related to a reduction in the Time-to-Abx., drawing blood cultures 2 times on each patient, changing laboratory analytic processes, and encouraging pulmonary and infectious disease consultation.

Average Length of Stay (ALOS)

ALOS is routinely used as an outcome measure, even though it is easily confounded by other clinical, demographic, and administrative conditions (e.g. age, comorbidity, payer source, etc.). There is variation in ALOS attributed to patients, physician, and hospital-based factors. Several studies have reported median stays of approximately 7 days (Meehan, 2000 and Batttleman, 2002).

Re-admission

Medically unstable patients have a 60% greater chance of readmission or death if discharged too soon. The determination that the hospitalized patient is medically stable is generally based on vital signs including mental status and ability to eat/drink.

Structural Indicators

Location of Antibiotic

Patients who received Abx. in the Emergency Department (ED) received antibiotics sooner. In one study, CAP patients given Abx. in the ED received it, on average, in 3.5 hours (± 1.4) compared to 9.5 hours (± 3.0) for patients who received the Abx. as an inpatient. Patient characteristics have been studied among those receiving Abx. in the ED or as an inpatient, and co-morbidity was the only significant predictor distinguishing whether patients received it in one site versus the other. Another study showed payer source to be a significant predictor (where Medicaid and self-pay patients would more

likely receive Abx. as inpatient) and presenting clinical symptoms (white blood cell counts and respiratory rates) were also significant clinical predictors (Battleman, 2002).

Standing orders

The Centers for Medicare & Medicaid Services (CMS) made important changes in October 2002 by encouraging the use of "standing orders" for pneumonoccal immunization. Specifically, CMS removed the requirement that physicians sign and date each individual order for the influenza and pneumococcal vaccination. Where allowed by state law, non-physician personnel are allowed to provide vaccinations under a facility-approved, standing-order protocol, in consultation with the physician, after assessment for contraindications. A Journal of Gerontology Study found that standing orders increased immunization rates among higher-risk older adults by 22% (Federal Register, 2002). Standing order programs have also been evaluated by the RAND Corporation and were found to be "very effective" in increasing vaccination rates.

Registered Nurse – Bed Ratios

Fine found that lower nurse/bed ratios to be positively associated both with Time-to-Abx. and timely blood cultures. Having a higher nurse/bed ratio resulted in quicker antibiotic administration (OR 1.23; 95% CI 1.10-1.38) and blood culture collection (OR 1.43; 95% CI 1.26-1.61) (Fine, 2002).

Arrival Time

Shift effects may play an important role in CAP performance indicators. In one Study, patients who arrived during the evening (11pm - 7am) or morning (7am - 3pm) were less likely to have received antibiotics within 8 hours (OR 0.80; 95% CI 0.69 - 0.93) and (OR 0.85; 95% CI 0.77 - 0.94) (Fine, 2002).

CHAPTER 3. - RESEARCH DESIGN

A causal comparative (ex-post facto) design was used to address the principal research question: *How is the control chart model related to medical quality*? Four test questions included:

1. How well do control charts monitor changes in hospital practice?
2. Can charts be made more or less sensitive by changing sampling frequency?
3. Does risk adjustment or stratified analyses improve the sensitivity of conclusions?
4. How do data characteristics influence chart sensitivity?

Control chart performance was evaluated under different conditions using Community Acquired Pneumonia (CAP) process and outcome indicators. CAP was selected due to the sufficient number of cases and because CAP was considered by hospital staff to be a priority area for quality improvement. The hospital had undertaken a series of CAP quality initiatives and had previous experience using control charts to monitor CAP processes and medical outcomes.

Data Management

Patient identifiers were removed from a dataset of people treated for community acquired pneumonia at a mid-sized (393-bed) hospital located in southern Arizona. Data were obtained from Midas + and Perspective On-line with the latter generating severity adjusted estimates using an APR-DRG grouper methodology that adjusts for diagnosis,

procedure, and age. For process variables, the dataset contained N=929 records; outcome variables were reported as single averages for 41- 35- or 30-month periods.

The quality services supervisor queried the hospital's information system, exported, and transferred all data using Microsoft Excel Spreadsheets (MS EXCEL) or downloaded .txt files. Control charts were constructed using MS Excel and verified using Northwest Analytical, Inc. (NWA) Quality Analyst software. Data Desk Software was used for data exploration, and graphics were produced using MS Excel. Descriptive and inferential analyses were performed using SPSS 11.5.

Variable/Indicator Definitions

The initial CAP dataset included thirty-five variables used individually, or paired into rates, to act as direct, or proxy, measures (indicators) of medical quality (see Table 9.). Process and outcome indicators were assigned dependent variable (DV) status. Time-to-Antibiotic (Time-to-Abx.), considered the primary medical process DV, was measured on a continuous scale (in minutes) and then split into dichotomous variables (Abx. lt4 Hours and Abx. lt8 Hours). Other dichotomous DVs include Blood-Culture-Prior to Abx., Influenza Vaccine Documented and Pneumonia Vaccine Documented. Most Outcome DVs were analyzed as rates with the exception of Average Length of Stay (ALOS) which is a continuous variable reported in days. Seven DVs were selected for analysis. Table 9. lists the DV indicators and the variable abbreviations used throughout this Report. For the ALOS and Mortality outcome DVs data were exported using both International Classification of Diseases - 9th Version (ICD-9) and Diagnostic Related Groups (DRG)

coding schemes. Severity adjusted estimates were provided for all outcome DVs, for partial time (N=30 months), the severity adjusted mortality indicator used ICD-9 coding. Indicator definitions, including exclusion and inclusion criteria, are specified in Appendix 2. Variable definitions were adapted from the Joint Commission for the Accreditation of Healthcare Organizations (JAHCO).

Table 9. Indicators Selected for Analysis (Dependent Variable List)

Type	*Indicator*	*Abbreviation*	*N*
Process	Time to Antibiotic	Abx.	929
	♦ Minutes	AbxMin	
	♦ <4 Hours	Abx.<4	
	♦ <8 Hours	Abx.>8	
	Blood Culture prior to Antibiotic	BCP_Abx.	929
	Influenza Vaccine Documented	Influ	929
	Pneumonia Vaccine Documented	Pneum	929
Outcome			
	Average Length of Stay	ALOS	35
	♦ ICD-9		35
	♦ DRG Coded		
	♦ Severity-Adjusted		30
	Readmission within 31 days %	Readmits	35
	♦ Severity-Adjusted		30
	CAP-Mortality Rate	Mortality	
	♦ Adult, ICD-9 Coded	Mort_ICD	35
	♦ Adult, DRG Coded	Mort_DRG	35
	♦ Age >80	Mort_80	35
	♦ Age <79	Mort_79	35
	♦ Severity-Adjusted (ICD-9)		
	Complication Rate, Severity-Adjusted	Complict	30

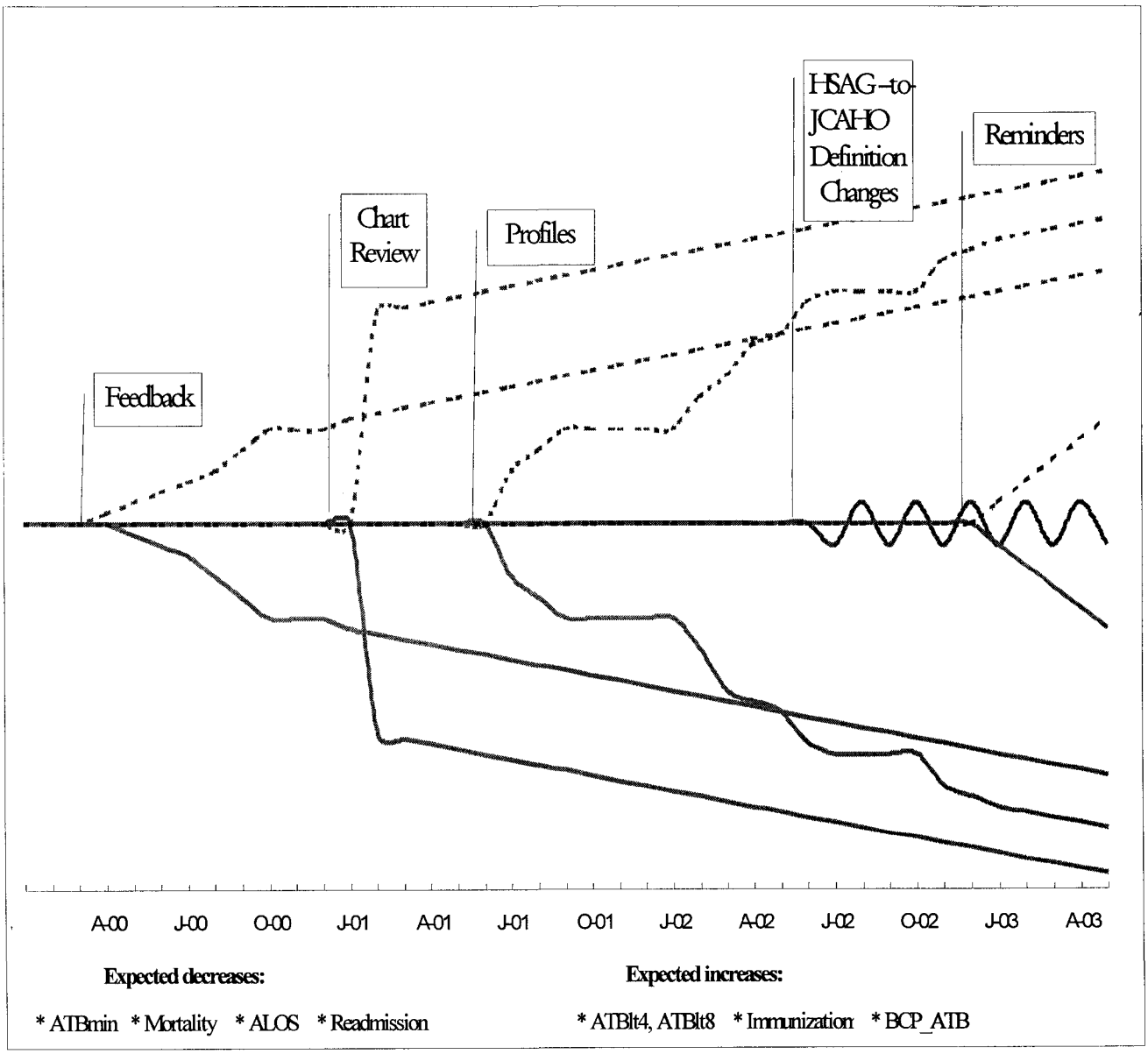
Indicators assigned independent variable (IV) status were causal events identified by hospital staff using the following process. Two members of the Medical Quality Team familiar with the Study were given event timelines with instructions and asked to independently identify major events expected to influence CAP quality indicators. They were also asked to indicate the calendar month. Table 10. identifies the five IV events.

Table 10. CAP Quality Test Events (Independent Variable List)

Event	*Description*
Guideline Presentations	Didactic presentations of varying length and content presented throughout the test period to physicians, nurses, Medical Executive Committee, Medical Quality Council, and the Board of Directors. January 2000: CAP Guidelines drafted, February 2000: presentations initiated, April 2000: ongoing presentations to medical groups.
Chart Review Changes	January 2001: Chart review process as quality audit was changed from retrospective monthly sampling to concurrent review. Retrospective sampling collected 20 cases monthly, with 12 selected for comprehensive review after the patient was discharged. Concurrent chart review (100% if possible) completed while the patient was still in the hospital, N changes month to month.
Indicator Profile by Physician	June 2001: Quality Reports provided to individual practitioner, performance compared to reference group. Information blinded during medical group presentations (e.g. Hospitalists, ED, FP) and for the Guidelines Committee. Un-blinded reports then shared with individual practitioners and group leaders.
HSAG to JCAHO Definitions	[Health Services Advisory Group (HSAG) adopts Joint Commission (JCAHO) reporting standards]. July 2002: New HSAG definitions change hospital reporting requirements (1) requiring admission and discharge diagnosis, and (2) mandating antibiotic time recording even if patient recently received antibiotic at another site. CAP patient reporting sampling scheme was also changed (to 100% mandated).
Clinical Reminders	January 2003: Case managers place reminder sheets on patient medical record highlighting for physicians that (1) patient has CAP diagnosis, (2) identify clinical options, and (3) reminders for clinical treatment.

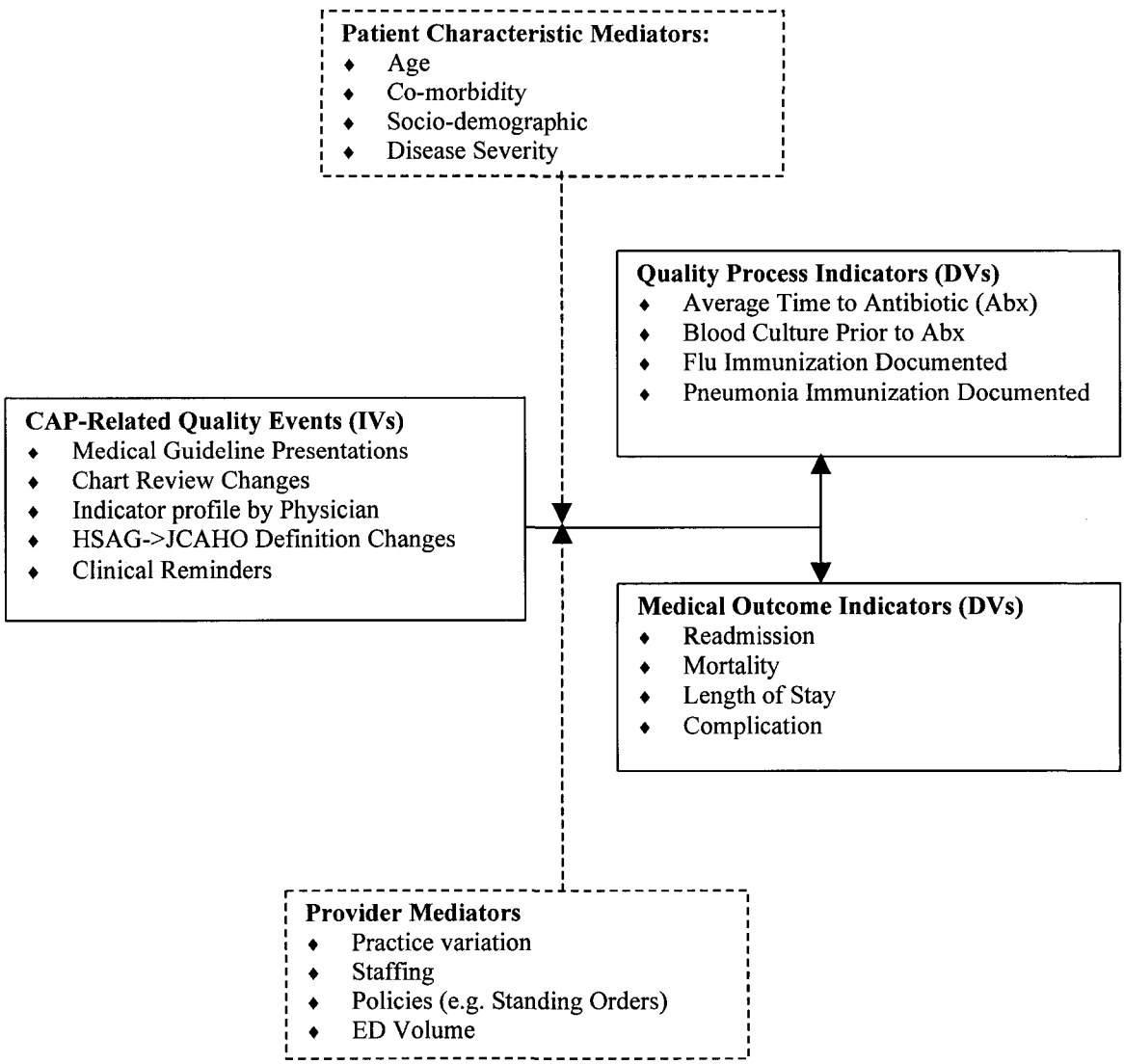
A final CAP Quality Event Timeline was reviewed by the hospital's Medical Director and Clinical Analyst to clarify agreements about the events and time periods to be tested. The following Proposed Effects Model was created to assist estimation of event causal influences and guide control chart sensitivity analysis.

Graphic 5. Proposed Effects Model of CAP Quality-related Events



The following Graphic describes the direct effect univariate relationships to be monitored through CAP quality indicators on the various control charts. Sets of mediating variables (identified in dash boxes) were anticipated but not measured.

Graphic 6. Proposed CAP Univariate IV-DV Relationships with (Unmeasured) Mediators



Analytic Plan

Descriptive Analyses

Descriptive analyses were conducted to: (1) Assess data quality and variable characteristics, (2) Review statistical assumptions associated with control charts, and (3) Determine whether the basic two-parameter control chart model, first developed by Walter Shewhart, is adequate or whether adjustments are warranted. Exploratory graph analysis, descriptive, and inferential statistics were used. Each time series variable was evaluated for thirteen conditions including control (randomness), shape (distribution), trend, seasonality, autocorrelation, sample size, coding, subgroup interval, rate interpretation, association, reliability, reporting bias, and missing data. Then correlation analysis was used to evaluate statistical associations between quality indicators. Factor analysis was used to determine whether common factors could be identified among the quality indicators. Regression analysis was used to determine whether significant trends were evident in any time series and non-parametric ANOVA (Friedman's Test) was used to determine whether seasonal patterns were statistically significant.

Inferential Analyses

Control charts were constructed using Montgomery protocols (Montgomery, 2001). The first step taken for every chart was to evaluate existing control. After the time series was determined to be stable, an initial 25-month baseline period was used to estimate the process average and compute control limits. Then, in a sequential manner, data points were added to the chart and were sequentially evaluated using Western Electric Rules

(Wheeler, 1992) to determine whether the data point indicated a special cause "signal." P-charts were constructed for discrete variables and XmR, Xbar, and EWMA charts for continuous variables (refer to Table 4.).

Most hospital-generated control charts were constructed for monthly time periods. For process DVs, charts were constructed using series reported weekly to determine whether sub-group frequency was differentially associated with causal events (week versus month subgroups). Where information was available, APR-DRG risk adjusted indicators were used to evaluate influences associated with clinical risk adjustment.

Control chart validity is measured by its ability to do what it is intended to do and, for this Study, validity was assessed by sensitivity analysis. Chart sensitivity was defined as a proportion, the number of times the event was correctly identified as a signal divided by the total number of events. The sensitivity analysis assessed the event's effect size relative to common cause variation, rather than validate the individual control chart. Results generated by the P-chart, XmR, Xbar, and EWMA charts were compared to evaluate differences in chart sensitivity.

Event effect sizes were estimated using the standardized mean difference method for stationary indicators and time series regression for indicators that demonstrated significant trend. Effect sizes were computed to help calibrate the control charts and as

an alternate statistical model to confirm the control chart results. Statistical method reliability was evaluated using the Kappa correlation coefficient.

Among the concerns in using control charts to monitor medical quality is their response to violations of statistical assumptions. Simulation methods were used to systematically vary conditions (i.e. effect size, distribution character, and data independence) to evaluate chart performance and calibrate expected events. Simulation was completed to allow testing of control chart behavior under different conditions, addressing its robust character under different conditions of probability.

Anticipated Threats to Validity

The research design was proposed with recognition of various threats to the validity of Study conclusions.

- The lack of experimental design eliminated the Study's ability to rule out alternative explanations.
- History: Univariate control chart analysis eliminated the ability to estimate effects on multiple events or event interactions.
- Reporting bias: The likelihood that events would be correctly specified. Two hospital staff members were asked to independently identify events and time periods.
- Experimenter bias: The possibility that the researcher's experience working with the statistical process control would result in a biased assessment inflating sensitivity estimates. Consequently, all chart construction decisions were recorded, and chart results were compared to previous hospital-constructed charts.

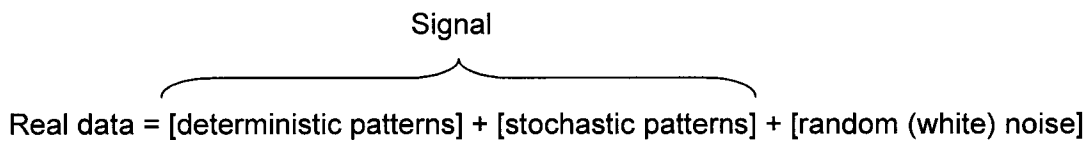
- Rating error: The process of estimating event lag times challenged the ability to designate events as signals. For clinical outcomes, it was assumed that most causal events would be closely associated in time with observed outcomes. Lagged effects proved particularly difficult to estimate, and a 3-month rule was adopted where signals occurring within three months after the event were considered an event-related signal.

CHAPTER 4. – RESULTS

Descriptive Analyses

Descriptive methods were used to evaluate the underlying data structure of each indicator. The following conditions were reviewed: Control (Random Variation), Shape (Distribution), Trend, Seasonal Cycles, Autocorrelation, Coding (DRG & ICD-9), Subgroup Frequency (Week, Month, Quarter), Interpretation of Rates, Association, Reliability, Reporting Bias, Sample Size, and Missing Data.

Different quantitative methods, including time series analysis, can be used to distinguish noise from signal. Whereas the control chart uses straightforward graphic and statistical methods that have the advantage of application ease in difficult environments, the other methods may extract more information. Time series, for example, decomposes signal variance into special event and predictable patterns (Rushe and Gottman, 1993).



Time series methods identify, and then model, stochastic patterns (e.g. autocorrelation), using them for additional information to improve prediction and forecasting. The Shewhart control chart does not distinguish these patterns. If apparent, they would need to be controlled through sampling or with specialized charts. Descriptive analysis helps determine whether statistical assumptions are met and provides information about data patterns and structure.

Control (Stability)

A Runs Test was used to evaluate whether the indicator series was randomly distributed and demonstrated controlled variation. A controlled series would show a stable and consistent pattern reflecting chance cause. Mean/median run charts were created, and the number of runs (crossing average line) was counted. Points on the average line were subtracted and then the number of runs counted; those series <33% were considered not controlled (Carey, 1995). No further Runs Tests were completed. Graphic 7. shows the Run Chart for the Time-to-Abx. indicator.

Graphic 7. Run Chart: Time-to-Abx.

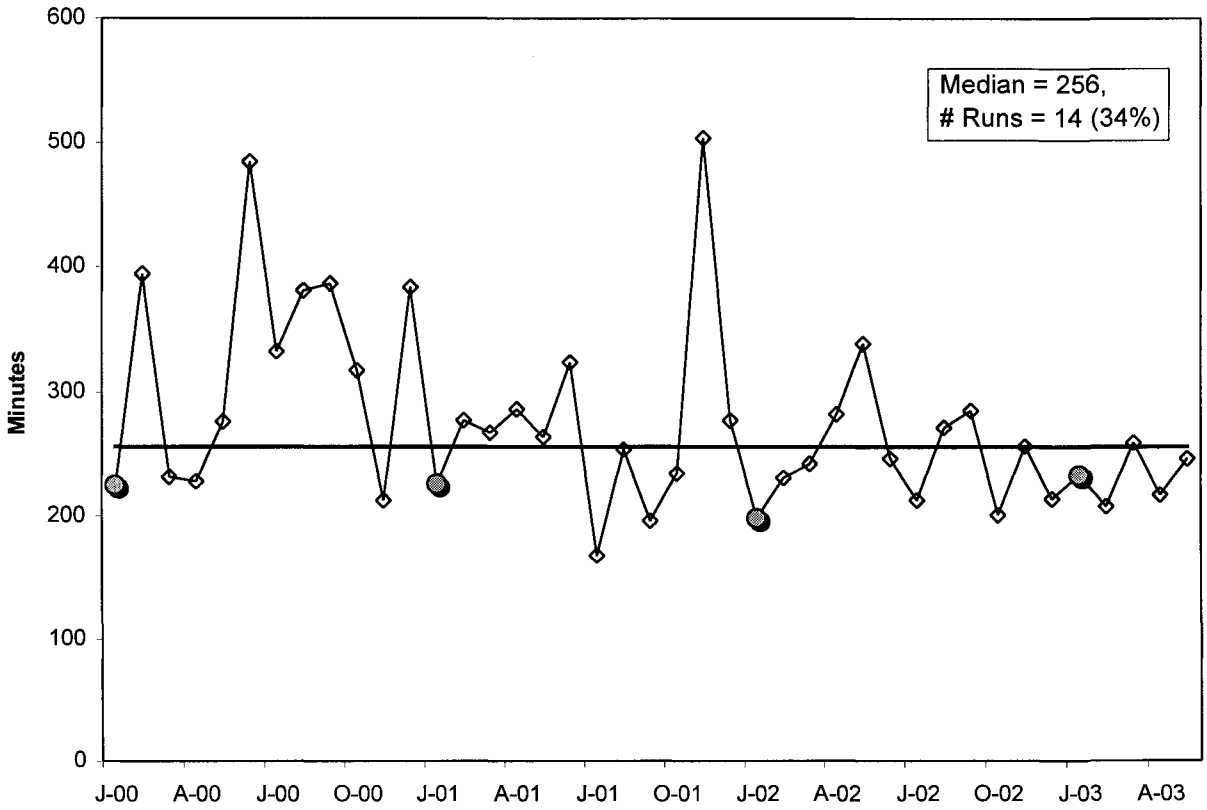


Table 11. summarizes partial results from the descriptive analysis. An insufficient number of runs (<33%) was observed in four indicators: BCP-before-Abx., Influenza Documented, Pneumonia documented, and #Cases. The Influenza and Pneumonia Documentation indicators showed a strong statistical, and non-stationary, trend. Severity adjusted #Cases appears strongly influenced by seasonal effects. The BCP-Abx. indicator reflects changes in average and variation.

Sufficient random variation was observed in: AbxMin, LT4Abx, LT8Abx, LT80 Mortality Rate, GT79 Mortality Rate, Mortality Rate (adjusted and unadjusted), Severity-Adjusted Complication Rate, ALOS (adjusted and unadjusted), and Readmission (adjusted and unadjusted).

Shape (Distribution)

Variable distributions were reviewed using the full dataset (n=944 patients) for CAP process indicators and using the monthly averages, (n=41) for process indicators, (n=35) for outcome indicators, and (n=30) for severity-djusted indicators.

Thirty-five variables were evaluated, including numerators and denominators, severity adjusted and unadjusted. Almost half of all the variables (16/35) had positive skew reflecting time measurement (i.e. Time-to-Abx., ALOS), which has an absolute minimum (0) but no absolute maximum. Outcome rates showed a positive skew having a higher number of small or zero events. Three variables (Abx.<4 and Abx.<8), proportions whose

averages approached an absolute maximum, showed a negative skew. Two indicators had U-shapes, proportions with values at both the absolute maximum and absolute minimum, and one distribution was bi-modal.

Eight variables had extreme values, and fifteen others had outliers (determined by Box-and-Whisker plots). For the monthly reported data, months identified most often as Extremes were June 00, Jun 01, and Apr 02.

The AbxMin Indicator was transformed to review how transformation would affect control charts. One lesson learned concerns the timing of transformation. Transforming the entire dataset before calculating monthly averages created artificially reduced monthly variation. Still, transformation did reduce the amount of skew in the monthly reports. Without transformation, 29/41 monthly averages were positively skewed, which dropped to 4/41 skewed monthly averages after transformation. Transformation significantly reduced the number of outliers in monthly reported averages; this is an important consideration in the interpretation of control charts.

Nine out of thirty five variables, 26%, had normal distributions.

Table 11. CAP Variable Descriptive Summary

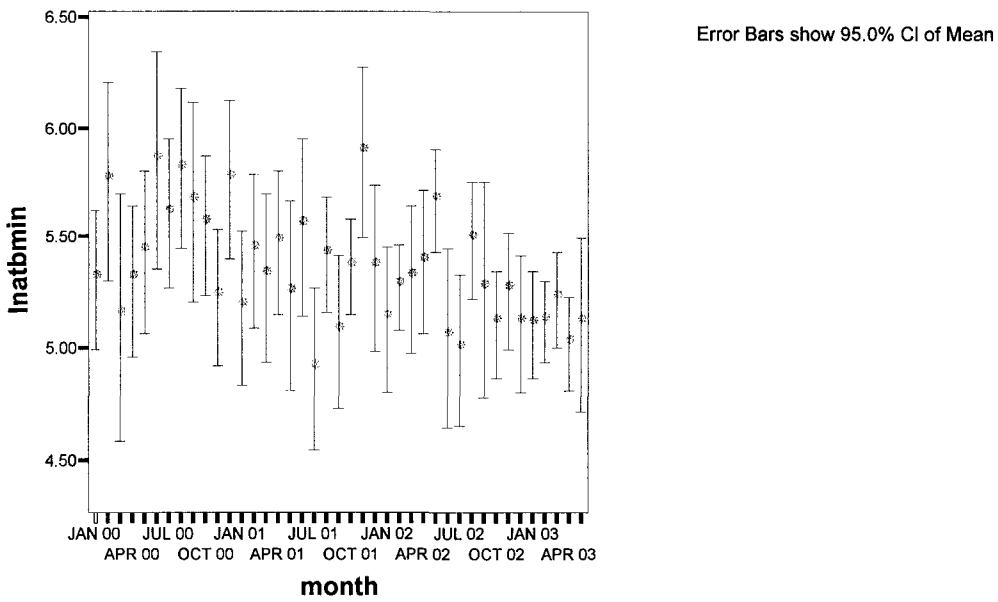
Variable	*N*	*Mean*	*Std. Dev.*	*Median*	*Min - Max*	*Skew*	*Distribution*	*Outliers (Extremes)*
AbxMin	929	264.2	221.8	200	0 - 1350	2.05	Positive Skew	@35 (10)
LnAbxMin	927	5.29	.762	5.30	3.0 - 7.21	-.02	Normal	4
AbxMin- (Month)	41	275	74.5	256.3	169 - 503	1.42	Positive skew	6 (2 June 00, Nov 01)
LnAbxMin	41	5.35	.141	5.31	4.9 - 5.9	.508	Normal	0
ALOS (DRG)	35	4.3	.643	4.24	3.2 - 5.8	.164	Normal	0
Readmtnu	35	3.9	2.65	4.0	0 - 12	1.11	Not-normal	2
Readmtde	35	35.1	15.34	31.0	12 - 73	.83	Positive Skew	0
READMISSION	35	.109	.056	.102	0 - .3	.98	Not-normal	3 (1, June 01)
<80morn	35	1.46	1.36	1.00	0 - 6	1.33	Not-normal	1
<80mord	35	27.2	12.3	24.0	9 - 54	.69	Positive. Skew	0
<80rate	35	.057	.051	.050	0 - .22	1.08	Positive Skew	1
>79morn	35	.71	.860	1.00	0 - 3	1.19	Positive Skew	1
>79mord	35	10.69	4.17	11.0	4 - 21	.44	Normal	0
>79rate	35	.0628	.077	.056	0 - .25	1.14	Positive Skew	1
Peumorn	35	1.17	1.15	1.00	0 - 4	.99	Positive Skew	0
Peumord	35	26.37	12.4	22.00	10 - 58	.79	Bi-modal	0
MORTALITY (DRG)	35	.044	.041	.041	0 - .15	.86	Positive Skew	0
Cases	30	43.0	20.28	38.0	19 - 108	1.29	Positive Skew	Jan 00
APR_read	30	4.4	3.05	4.5	0 - 11	.45	Normal	0
Complica	30	1.20	1.126	1.0	0 - 4	.66	Positive Skew	0
ALOS2	30	4.90	.74	4.9	3.7 - 6.7	.38	Normal	Jul 02
Samrate	30	.053	.030	.050	0 - .14	.87	Normal	Apr 02
Saradmit	30	.102	.068	.102	0 - .32	1.23	Positive Skew	Jun 01
Sacomrat	30	.029	.033	.02	0 - .15	1.94	Positive Skew	1 (1, Jun 00)

Trend

Trend was tested using regression analysis for monthly and full datasets. Necessary trend was apparent when the regression slope coefficient was statistically significant. All of the CAP indicators showed a significant statistical trend with 6 out of 10 being negative. The ALOS indicator model significance was judged marginal (.095) but significant. None of the severity-adjusted indicators showed significant trend. Logistic regression, used to evaluate full datasets for the dichotomous process variables, produced findings consistent with linear regression of the limited month datasets.

Stationarity addresses series stability over time. It is a "statistical property that refers to the condition of relatively constant mean and variance throughout the time series" (Rushe, 1993). Statistical trend, seasonal cycles, and special causal events, all of which are evident in the CAP Indicators, can influence Stationarity. In this descriptive summary, stationarity was first evaluated graphically by examining apparent changes in average value or variation and, for process indicators, by looking at monthly box and whisker plots or error bar charts. Of the process variables checked, none appears to be stationary; all vary in average and variance.

Graphic 8. Transformed Abx. and Abx. <4 Hours, Stationarity Assessment



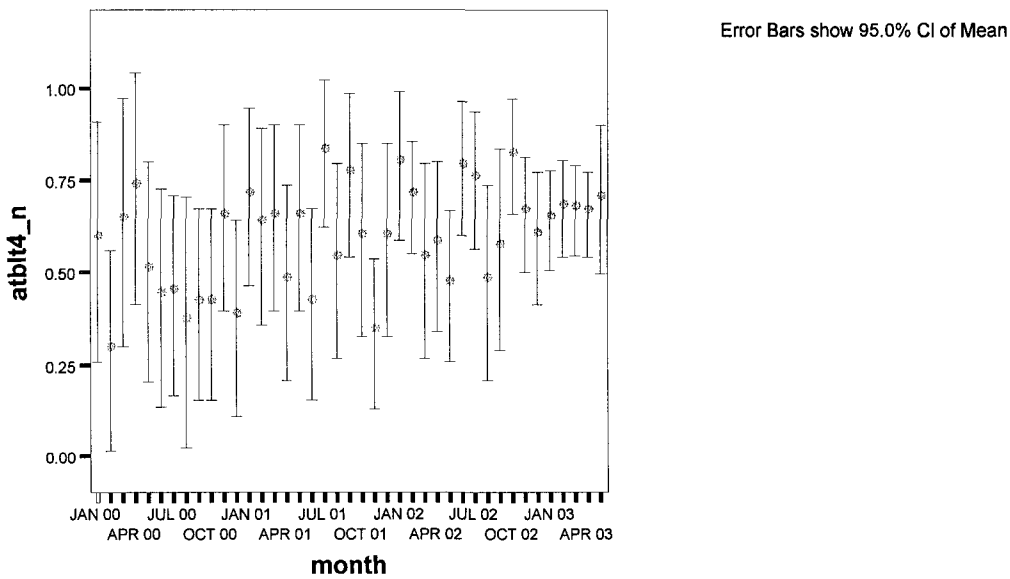


Table 12. Statistical Trend Table

Variable / Indicator	*B*	*sig.*	R^2	*Beta*	*Influential Points*
Process					
Time to Antibiotic					
Minutes	-2.29	.018	.134	-.367	Nov 01
Transformed	-.011	.001	.252	-.502	-
Within 4 hrs	.489	.007	.173	.416	-
Within 8 hrs	.323	.025	.122	.349	-
Blood Culture prior to Antibiotic	-.475	.005	.182	-.426	Sept 00
Influenza Vaccine (documented)	2.967	.000	.857	.926	-
Pneumonia Vaccine (documented)	3.064	.000	.857	.926	-
Outcome					
ALOS (DRG)	-.018	.095	.082	-.287	
Readmission within 31 days %	-.080	.022	.399	-.147	June 01
Age<80 Mortality Rate	.022	.805	.002	.043	Sept 02
Age>79 Mortality Rate	.093	.478	.015	.124	-
Adult Mortality Rate (DRG)	-.237	.000	.320	-.566	-
Severity Adjusted					
Adult Mortality Rate	.001	.289	.040	.200	-
Readmission within 31 days %	.000	.894	.001	-.025	June 01
Complication Rate	0.00	.971	0	.007	June 00
ALOS	-.01	.526	.015	-.121	-

Seasonal Cycles

All variables (numerators, denominators, and rates) were visually and quantitatively analyzed to determine whether significant seasonal cycle patterns were evident. A 3-step process was used to evaluate the influence of seasonal cycles:

1. Review series bar chart to evaluate suggestive series patterns.
2. Rank indicators by month to for each year period to evaluate annual patterns.
3. Use the Friedman nonparametric test, an expansion of the repeated measures ANOVA based on the ranks within each case to test the null hypothesis of "no difference" ranking across months.

The assessment of seasonal effects using quantitative analysis (Friedman's non-parametric test) had limited power due to the small number of observations (n=3 years). To increase power for the outcome measures, the missing month of June 03 was imputed using average values from the previous two months. Due to the small number of observations, graphical analysis was a superior method by which to detect seasonal influences.

There is a clear, increased activity in CAP during the winter months. Three reasons have been offered to remove seasonal effects: when comparing an indicator at different points of the year to understand a purely intra-year phenomenon; to remove seasonal effects from the series to study its other constituents uncontaminated by the seasonal component; or to correct or adjust a current figure to account for seasonal effects.

Seasonal variation was apparent in the unadjusted outcome numerator/denominator variables but not in the rates; creating rates seemed to eliminate seasonal influence. No seasonal effects were observed in any of the process variables, in the outcome rates, or in ALOS. There were fewer seasonal effects observed in the severity-adjusted numerators (only in readmissions indicator). The #Cases denominator also exhibited seasonal effects. Graphic 9. shows how creating rates eliminates seasonal effects.

Graphic 9. Seasonal Effects in CAP Readmission

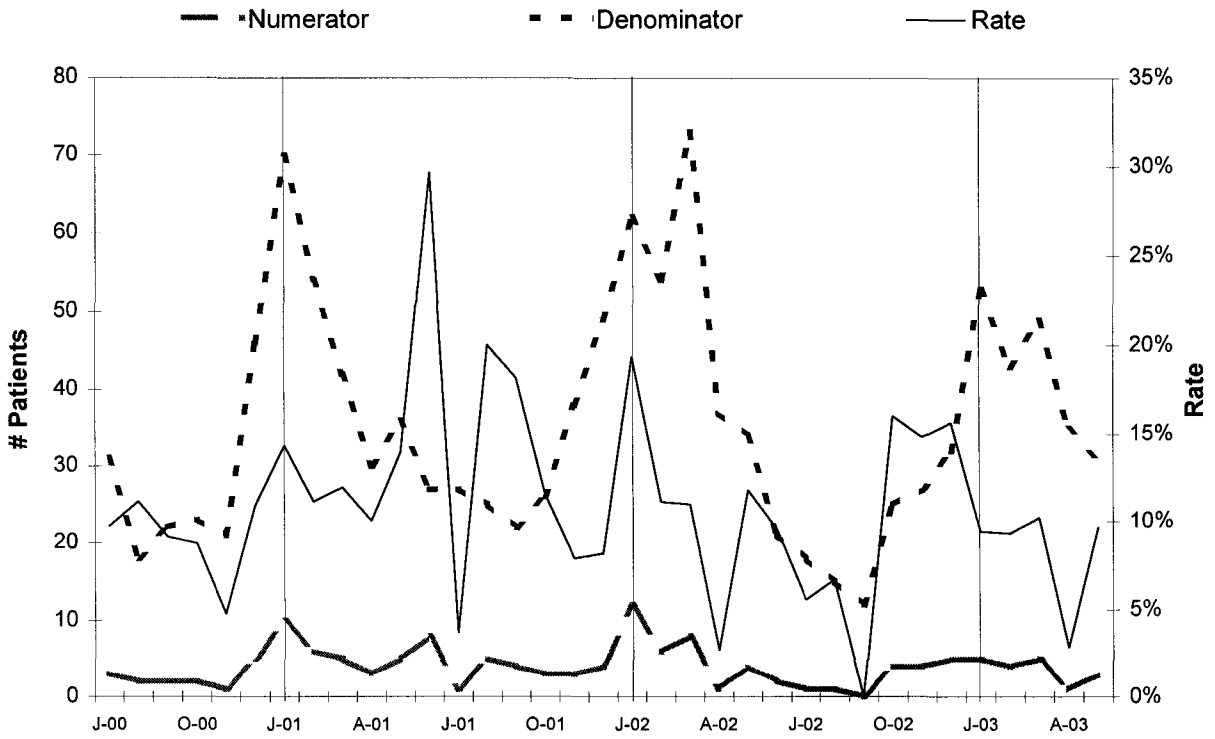


Table 13. Seasonal Assessment: Unadjusted Outcome Indicators Average ranking by Month (n=35)

	Indicators / Rates					*Denominators*			*Numerators*			
Month	ICD-9 ALOS	ICD-9 Adult Mortality	Age < 80 Mortality	Age > 79 Mortality	Readmission within 31 days	Adult Age < 80 w CAP Dx.	Adult Age > 79 w CAP Dx.	# Pts. Discharged W/Pneumonia	All Pneumonia Deaths	Age < 80 Deaths	Age > 79 Deaths	# Pts. Discharged & Readmitted
Jul	8.0	4.7	6.7	5.0	9.7	9.3	5.0	7.7	5.7	6.7	4.7	8.7
Aug	6.0	4.7	4.7	3.7	4.7	9.3	9.7	10.7	5.7	6.0	4.7	7.0
Sep	4.0	6.3	5.7	7.0	8.0	9.7	11.0	10.7	7.7	5.7	7.0	8.3
Oct	4.0	7.0	4.7	7.0	5.7	9.7	7.7	8.7	7.3	5.7	7.0	7.0
Nov	9.3	5.3	6.0	3.7	8.3	8.0	7.0	7.7	4.7	6.0	3.3	8.0
Dec	4.3	6.7	6.3	6.3	6.0	4.3	4.7	4.0	4.7	4.0	5.7	3.3
Jan	7.7	9.0	7.7	5.3	3.3	1.3	2.0	1.3	3.0	3.7	3.3	1.0
Feb	4.3	5.7	7.0	4.7	6.0	2.7	2.3	2.7	2.7	4.0	2.0	3.3
Mar	9.0	4.7	4.0	6.0	5.0	2.0	1.7	2.3	3.7	3.0	4.0	2.3
Apr	7.3	5.3	6.0	2.3	10.0	5.3	6.0	5.7	3.3	4.7	1.7	8.7
May	7.3	9.0	8.0	4.7	4.0	6.0	6.3	6.0	6.7	7.7	3.7	5.3
Jun	4.0	6.0	5.0	7.5	4.5	8.5	11.0	10.0	6.5	4.5	7.5	6.0
Chi sq.	12.1	6.6	5.12	16.72	12.64	26.81	29.55	29.72	12.02	7.03	16.71	20.93
Sign	0.354	0.83	0.925	0.116	0.317	0.005	0.002	0.002	0.36	0.797	0.117	0.034
Visual Appearance	No	No	No	No	No	Yes	Yes	Yes	Uncertain	Uncertain	No	Yes

Autocorrelation

Autocorrelation is related to measurement frequency and concerns the association between subsequent values in the time series. Assessing autocorrelation is important for control charting, as significant positive 1st lag autocorrelation is expected to tighten control limits, while negative 1st lag autocorrelation may result in wider control limits.

More frequent reporting (day, week, month, and quarter comparisons) does not increase 1st lag autocorrelation but does seem to increase the frequency of subsequent lag autocorrelation. More frequent reporting reduces the magnitude of autocorrelation, as there are more values used to calculate the ACF. Quarterly periods seem to reduce the amount of subsequent lag autocorrelation.

1st lag and/or seasonal pattern autocorrelation is apparent for all the outcome indicator denominators and for many numerators. However, the creation of rates seems to eliminate autocorrelation, perhaps due of combining two distinct pieces of information that represent different causal processes. DRG / ICD-9 coding appears to affect autocorrelation patterns more for the Mortality Indicator and less for the ALOS Indicator.

Table 14. Autocorrelation Table

Variable	*Significant 1st Lag Autocorrelation*	*Significant Other Lag Autocorrelation*
AbxMin	No	.310 (6 lag)
AbxMin Transformed	No	.384 (6 lag)
ATB <4 hrs	No	No
ATB <8 hrs	No	.363 (6 lag)
BCP Prior to Abx.	No	No
Influenza Documented	.890	Trending pattern
Pneumonia Documented	.886	Trending pattern
ALOS (DRG)	No	-.291 (15 lag)
Readmission Numerator	No	Seasonal pattern
Readmission Denominator	.659	Seasonal pattern
Readmission Rate	No	-.269 (15 lag)
<80 Mortality Numerator	No	.297 (13 lag)
<80 Mortality Denominator	.656	Seasonal pattern
<80 Mortality Rate	No	No
>79 Mortality Numerator	No	No
>79 Mortality Denominator	.438	Seasonal pattern
>79 Mortality Rate	No	No
Pneumonia Mortality Numerator	No	.289, .338 (11, 14 lags)
Pneumonia Mortality Denominator	.689	Seasonal pattern
Pneumonia Mortality Rate (DRG)	No	.354 (2 lag)
S.A. # Cases	.503	Seasonal pattern
S.A. Mortality Numerator	No	No
S.A. Readmission Numerator	No	No
S.A. Complication Numerator	No	No
S.A. ALOS	No	Seasonal pattern
S.A Mortality rate	-.407	No
S.A. Readmission Rate	No	-.388 (10 lag)
S.A. Complication Rate	No	No

S.A. = Severity Adjusted

Table 15. Autocorrelation Patterns by Reporting Periods (Sig. Tested at 2 SE Limits)

	Month		*Week*		*Quarter*	
Variable	*Significant 1st Lag Autocorrrelation*	*Significant Other Lag Autocorrelation*	*Significant 1st Lag Autocorrrelation*	*Significant Other Lag Autocorrelation*	*Significant 1st Lag Autocorrrelation*	*Significant Other Lag Autocorrelation*
Abx. Minutes	No	.310 6 lag	No	.305, .237 2 & 4	No	No
Abx. Transformed	No	.384 6 lag	No	.245, .242 2 & 4	No	No
Abx. <4 hrs	No	No	No	.182 4 lag	No	.477 2 lag
Abx. <8 hrs	No	.363 6 lag	No	.307 .162 2 & 4	No	No
BCP Prior to Abx.	No	No	No	No	No	No
Influenza Documented	.890	Trending	.824	Uniform pattern	.837	Seasonal
Pneumonia Documented	.886	Trending	.833	Uniform Pattern	.821	Seasonal trend

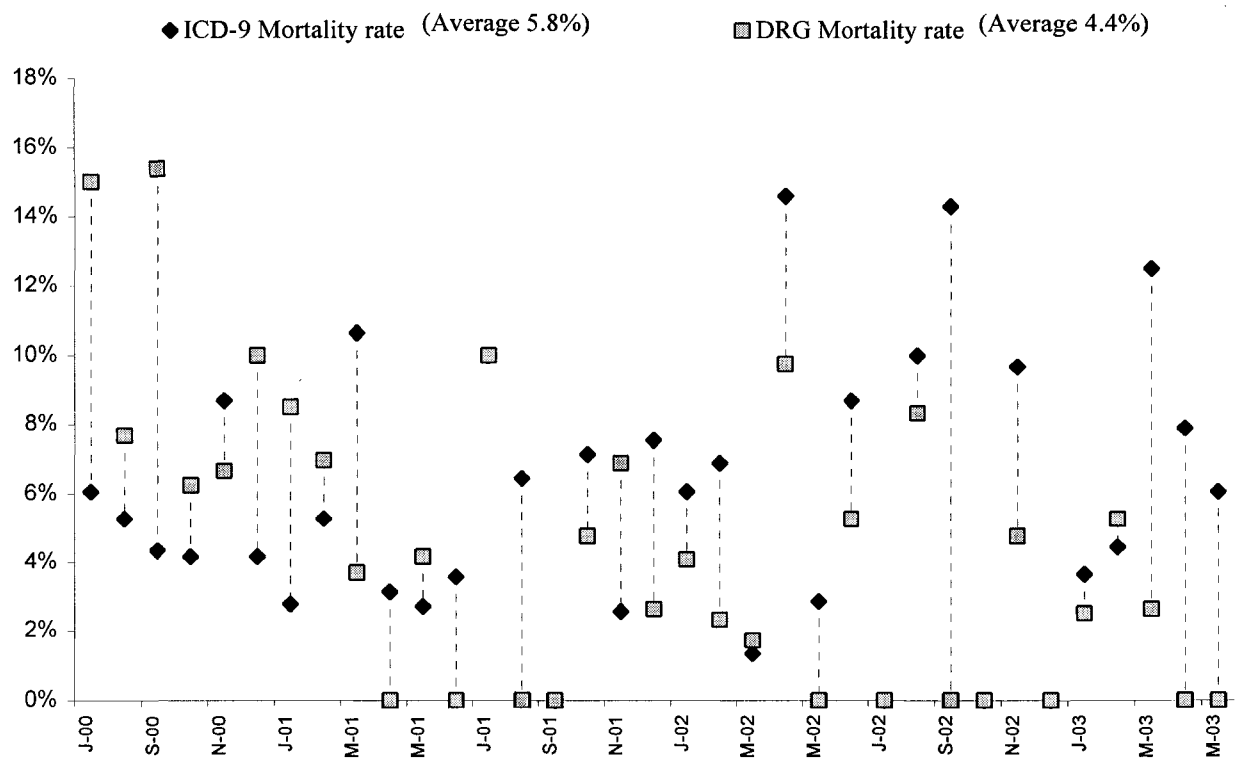
Coding

The ALOS and Mortality Indicators were coded using both ICD-9 and DRG codes, creating differences in nearly every descriptive category. DRG 089 includes simple pneumonia; DRG 079 includes pathogen-identified pneumonia. ICD-9 codes include all pneumonia patients and may have up to 100 different sub-codes. DRG-coded denominators are a subset selected from ICD-9-coded denominators.

There were important differences in the Mortality indicator averages when comparing ICD-9 (average 5.8%) and DRG (average 4.4%) mortality rates. The DRG-ALOS indicator remains consistently below the ICD-9 ALOS; however, the DRG-Mortality Rates fluctuate above and below ICD-Mortality. The differences in mortality are shown in Graphic 9. where DRG rates are higher until February 2001.

Trend patterns were different between the coding schemes. DRG-coded indicators have a significant negative trend while ICD-9 coded indicators have positive non-significant trends. Autocorrelation patterns also differed.

Graphic 10. ICD-9 and DRG Differences in CAP Mortality



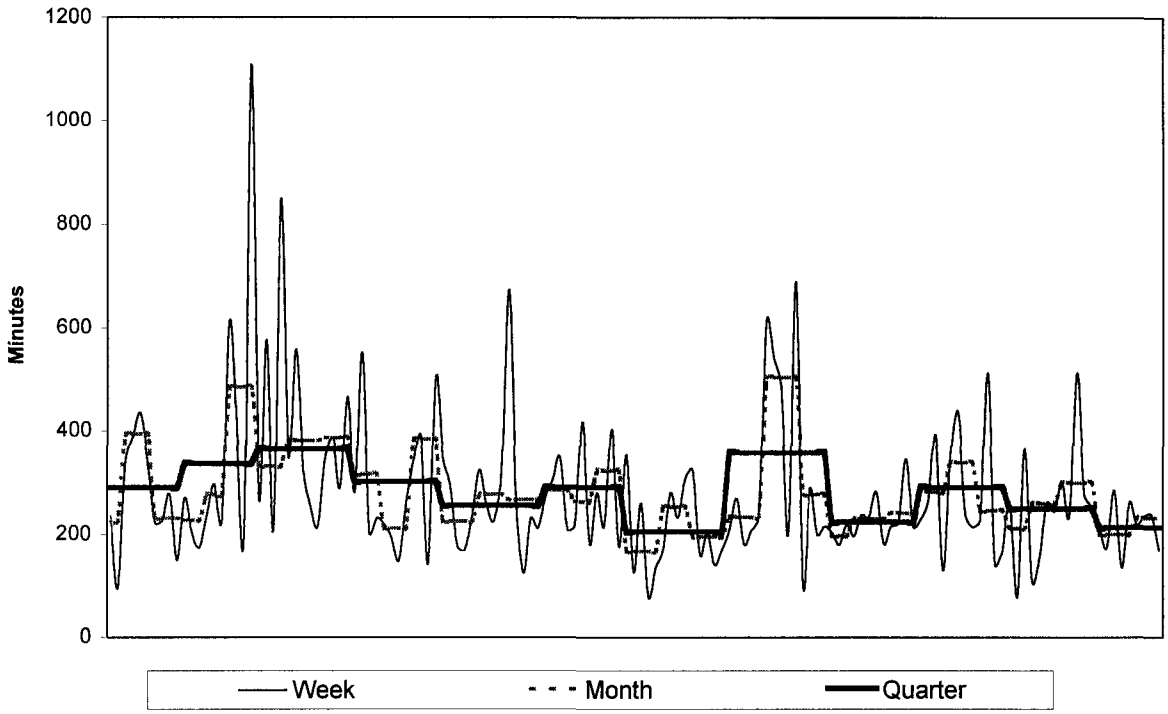
Subgroup Interval

The frequency at which subgroups are selected is one of the most important factors that determine chart sensitivity. As can be observed in the following line graphs, increased reporting does create more opportunities to identify special causes and increases chart sensitivity. As shown in Table 16., summarizing data into longer intervals creates systematic bias in the median, variation, and slope measures. Important differences in identified extreme values were also noted.

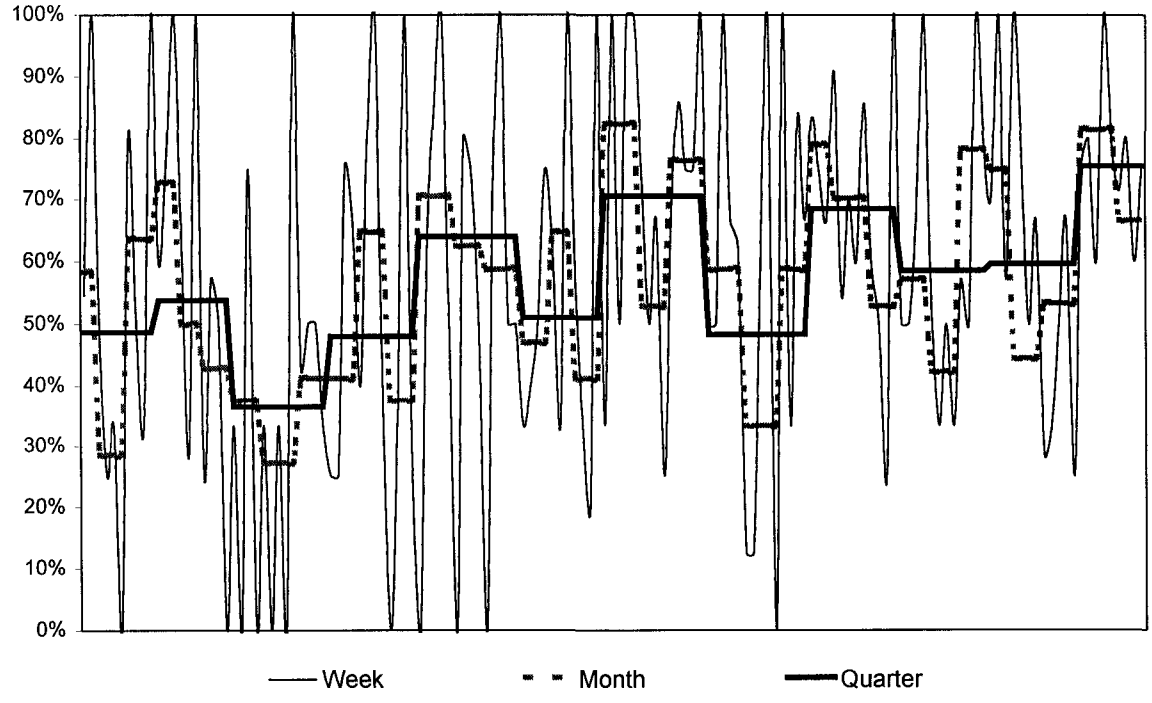
The periodicity of reports affects autocorrelation with shorter time periods showing more complex forms of autocorrelation. The statistical cost of reporting more frequently is increased autocorrelation, although autocorrelation does not appear to be a significant problem for most CAP indicators.

The following graphs show substantial differences in variation depending on subgroup-interval. In Graphic 11., data reported by week shows greater variability than data reported monthly or quarterly. These differences are even more apparent when looking at percentages. In Graphic 12. The proportion of patients who received the antibiotic within 4 hours varied greatly according to subgroup interval, sometimes as much as 100& when using week reports.

Graphic 11. Average Time-to-Abx. by Week, Month and Quarter Report Periods



Graphic 12. Time-to-Abx. Less than 4 Hours by Week, Month and Quarter Periods



Looking at the AbxMin variable reveals that reporting frequency did not influence the mean but did influence the median, all variability estimates, skew, slope, and extreme values. Rapid reporting cycles showed greater series point-to-point variation. Periodicity also influenced the extreme values, as different periods identified (some) different values as extreme.

Table 16. Time-to-Abx. Statistics by Reported Periodicity

	Patient / Date	*Day*	*Week*	*Month*	*Quarter*
Count	929	593	169	41	14
Mean	264.2	266.6	274.4	275.0	275.9
Median	200	215	234	256	274
Standard Deviation	221.7	197.00	138.42	74.99	52.18
Skew	2.051	2.145	2.418	1.419	0.460
Range	1350	1327	1033	336	160
Slope	-0.12	-0.17	-0.58	-2.51	-7.79
Extremes*	@10	@3	@20	@23	@19

* Extreme values change by time-report period

Interpretation of Rates

Rates combine two pieces of information from a numerator (frequency of event over specified period of time) and a denominator (related population size). Rates represent the number of events divided by the number of opportunities. Several challenges are associated with using rates as quality indicators.

In medical care, "controlled quality" may be a function of quality, volume, or both. Rates mix two distinct processes associated with different causal streams; only the numerator is related to quality improvement. For CAP-related mortality, processes associated with dying are mixed with processes associated with the number of people coming into the hospital, being diagnosed with CAP, and being discharged with the CAP as a primary or secondary diagnosis. Both numerator and denominator processes influence the rate; however, quality improvement efforts influence only the event (numerator). The rate is understood as a probability estimate of the event occurring; however, only the numerator directly relates to quality. Rates do not show the occurrence of actual events; rather, they describe the probability (i.e. proportion) of events-to-opportunities. Every rate is an event-to-opportunity measure. Dramatic changes in the indicator may be apparent, even without an increase in events due to increases or decreases in the denominator.

Small numerators present a challenge. A small numeric increase in events can appear larger if the denominator population remains constant. Likewise, a large increase in the numerator can be negated by an increase in the denominator. Many medical events are

fortunately rare or reflect small numbers. Because of the smaller numeric size of the numerator, this part of the rate has a larger impact on the rate overall. Correlation analysis was used to assess the relationship between the Readmission and Mortality numerators, denominators, and rates. The correlations are summarized in Table 17.

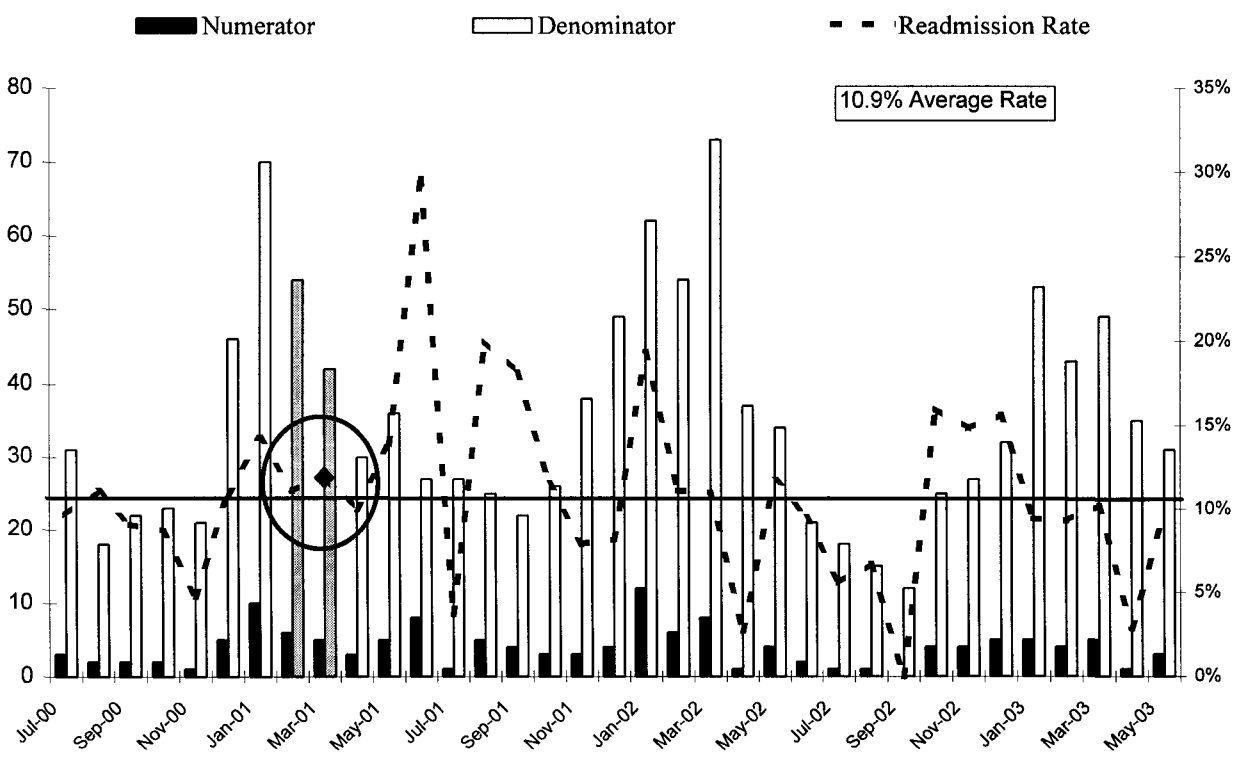
Table 17. Correlation of Numerator and Denominator with Rates

	Numerator	*Denominator*
Mortality		
Denominator	.480	
Rate	.797	.002
Readmission		
Denominator	.764	
Rate	.708	.159

Another consideration concerns measure reliability. When both numerator and denominator exhibit some modest reporting changes, the rate reliability coefficient will change to a greater degree than either the numerator or denominator independently. Rate-change problems are more apparent with small numbers.

Unexpected patterns can emerge when there are disproportionate changes in both numerator and denominator. A modest example of this, as shown in Graphic 13., is during the March 01 period for Readmission Rate where both numerator and denominator fell from the previous month, but did so at different proportion. This results in a modest rate increase.

Graphic 13. Numerator and Denominator Effects on CAP Readmission Rate



Association

Relationships between monthly reported indicators were reviewed using Pearson and Spearman correlation analysis. Overall, there was less association between the process and outcome variables than desired. There are several possible explanations for the lack of association, including the many contributions of variance to each indicator and limited indicator validity. For example, the Readmission indicator may not reflect clinical processes but rather be associated with a natural worsening of the medical condition. ALOS may be affected by insurance coverage or payer status as well as disease severity.

The DRG-coded Mortality Indicator is significantly associated with Time-to-Abx. and vaccine documented (influenza and pneumonia). This relationship disappears when the ICD-9 coded mortality indicator is evaluated.

ALOS and Readmission outcome indicators were not well correlated overall. ALOS was significantly associated with BCP-Prior-to-Abx. but in an unexpected way.

Severity adjustment has a minor effect on the magnitude, and, in some cases, a contradictory effect on the direction of association. Severity-adjusted complication rate is not significantly associated with any other indicator, while the severity-adjusted ALOS is associated with two Time-to-Abx. process indicators.

Table 18. Correlation Matrix for Process, Outcome, and Severity-Adjusted Indicators

	n	*Abx.-Trans*	*Abx.<4*	*Abx.<8*	*BCP_Abx.*	*Influ*	*Pneu*	*ALOS*	*Readmis*	*<80 Mort*	*>79 Mort*	*Mort*	*Cases*	*Severity Adjusted ALOS*	*Mort*	*Read*
Abx.-Trans.	41															
Abx. <4	41	-.911														
Abx. <8	41	-.822	.745													
BCP_Abx	41	.082	-.139	-.006												
Influenza	41	-.511	.469	.422	-.384											
Pneumonia	41	-.501	.455	.408	-.404	.966										
ALOS	35	.006	-.040	.086	.425	-.232	-.211									
Readmission	35	.088	-.013	.018	.115	-.009	-.054	.295								
<80 Mortality	35	.118	-.120	-.149	-.210	-.060	-.054	-.277	-.364							
>79 Mortality	35	-.162	.190	.301	-.025	.220	.222	-.002	-.219	-.095						
Mortality rate	35	.401	-.300	-.391	-.045	-.573	-.544	-.146	-.277	.134	.183					
Cases	31	-.235	.184	.294	.077	.058	.043	-.036	.037	-.138	.135	-.094				
Severity Adj.																
ALOS	31	-.371	.268	.372	.232	-.002	.015	.690	-.024	-.285	.628	.060	.096			
Mortality	31	-.[illegible]	.[illegible]	.[illegible]	-.137	.178	.201	-.113	-.459	.652	.627	.291	-.039	.325		
Readmission	31	.042	-.078	-.011	.223	.041	-.001	.345	.904	-.342	-.136	-.471	.019	-.077	-.413	
Complications	31	.202	-.115	-.177	.150	.019	.025	.300	.208	.247	.369	-.166	-.035	.047	.238	.256

sig. <.5

Reliability

For valid interpretation of control charts, signal-related variation needs to be distinguished from variation associated with the operation of information systems. Different conditions influence an information system's ability to produce consistent reports at different time periods. Information reliability is challenged by human factors, existing technology, indicator measurement, and coding practice, among other conditions. Human factors could include operator error and miscommunication. Technology factors include software and download problems. Indicator measurement factors include changes to definitions, or as indicator series fluctuate after additional clinical information, sometimes abstracted from medical records, becomes available. And there can be a large difference according to coding practice, use of ICD-9 or DRG coding systems.

CAP indicator series values were obtained on different occasions. Information system reliability was evaluated comparing data consistency across the December, April, July, and September time periods. Correlation analysis showed high reliability for the process indicators $r = .9 < x < 1.0$. However, reliability was inconsistent for the outcome indicators, especially for ALOS and Mortality.

Identifying the source of this variation seems important. Human error was involved, as variable definitions were not specified when data were requested. This led to downloading and transferring DRG and ICD-9-coded data at different junctures. The

choice of ICD-9 or DRG coding systems had a noticeable effect on the ALOS and Mortality data series. Series fluctuation in the Readmission indicator seems related to actual rate changes associated with data system updates, as information gleaned from medical records are incorporated into the electronic database (reporting system). Readmission was not affected by the DRG-ICD-9 coding. Lower ALOS and Mortality reliability can be understood as a "Data Handling Error." Readmission indicator reliability was high it though demonstrated a pattern consistent with modified chart reviews. Reliability for both the ALOS ($.6 < r < .99$) and Mortality ($.22 < r < .96$) indicators reflected DRG and ICD-9 coding differences; Mortality reliability reflected small changes in both the numerator and denominator.

Table 19. Test/Retest Reliability Coefficients for CAP Process Indicators

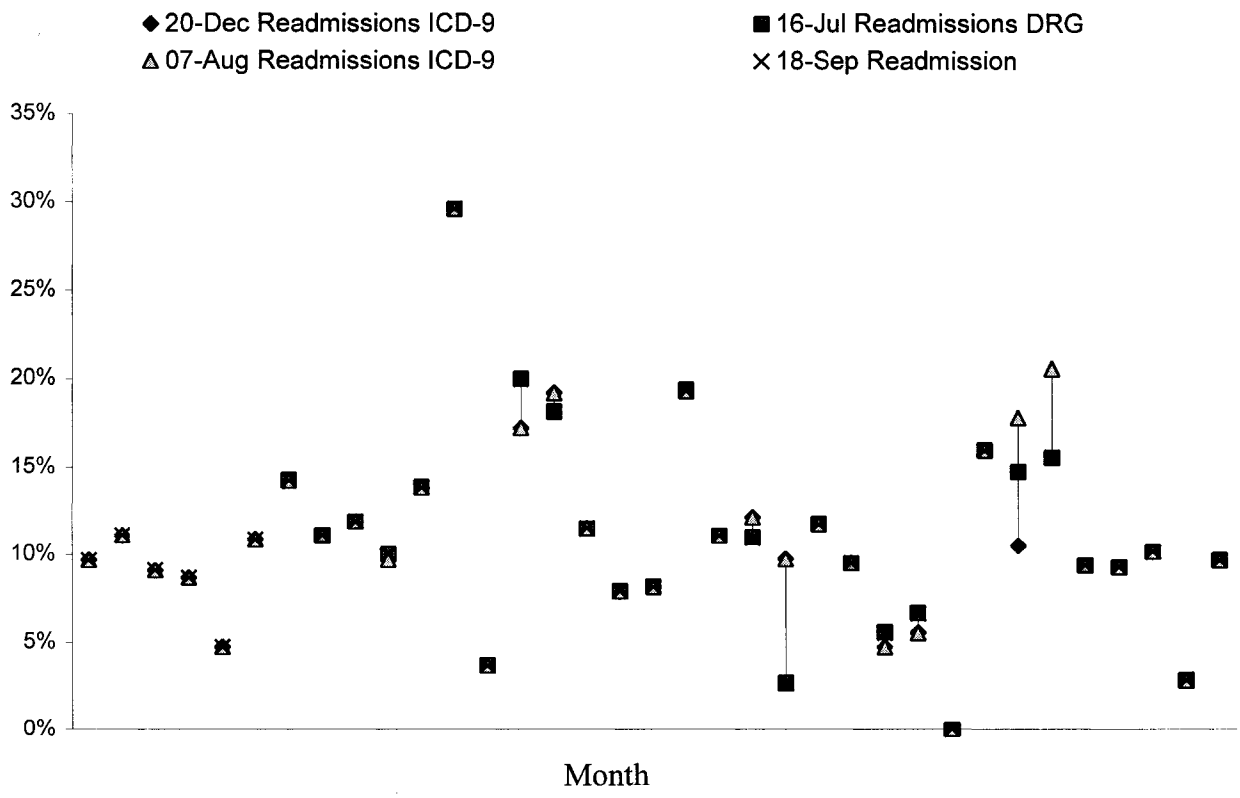
Process Indicator	R
AbxMin	.998
Abx.-Transformed	.948
Abx. <4	.991
Abx. <8	.999
Abx. consistent	.978
BCP-Abx.	.979
Influenza	.999
Pneumonia	.998

Table 20. Test/Retest Reliability for CAP Outcome Indicators

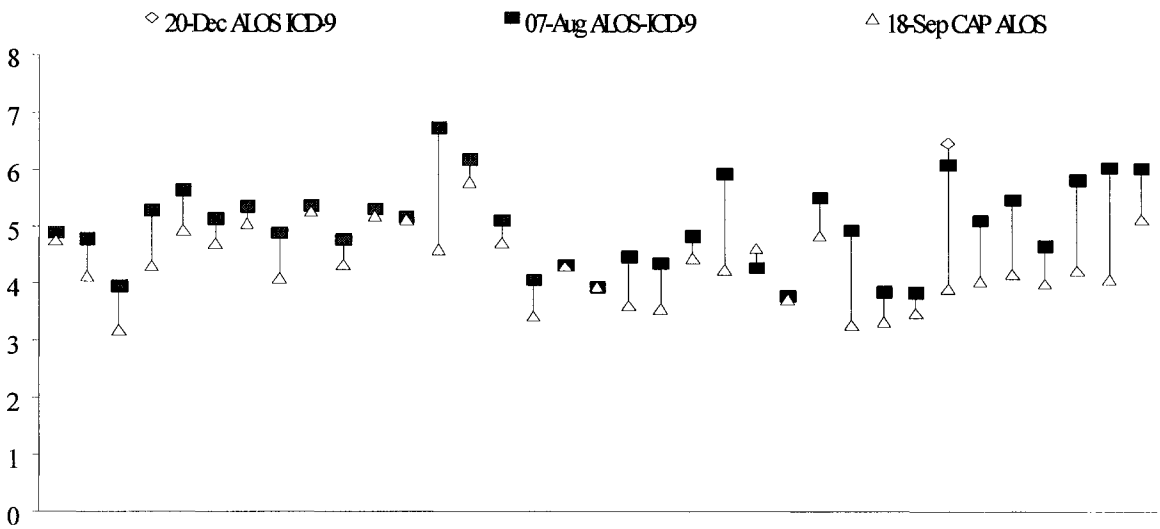
	12/20/02	7/16/03	8/7/03	9/18/03
	ALOS ICD-9		ALOS-ICD-9	CAP ALOS (DRG)
ALOS ICD-9	1			
ALOS-ICD-9	0.996	-	1.000	
ALOS (DRG)	0.613	-	0.600	1
	Readmission ICD-9	Readmission (DRG)	Readmission ICD-9	Readmission (DRG)
Readmission ICD-9	1			
Readmission (DRG)	0.955	1		
Readmission ICD-9	0.972	0.956	1	
Readmission (DRG)	0.958	1.000	0.958	1
	Mortality ICD-9	Mortality (DRG)	Mortality ICD-9	Mortality (DRG)
Mortality ICD-9	1			
Mortality (DRG)	0.435	1		
Mortality ICD	0.967	0.394	1	
Mortality (DRG)	0.262	1.000	0.229	1

Graphics 14-16 show differences in the Readmission, ALOS, and Mortality indicator values as reported at four data downloads periods. Coding differences for the Readmission Indicator shows minimal variation and coding difference across the entire observation period. The pattern observed in the ALOS indicator reflects changes in coding from the DRG->ICD-9 datasets. The patterns for Mortality differ across the reporting periods, with DRG codes being higher during the early months and ICD-9 higher in latter months.

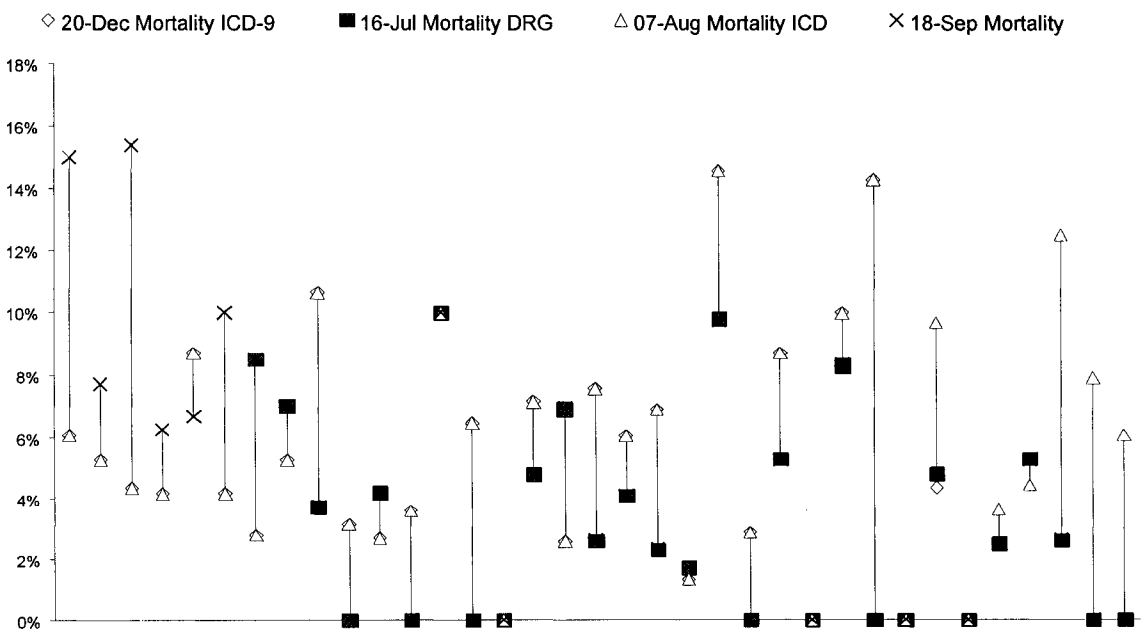
Graphic 14. Reliability Assessment for the Readmission Indicator



Graphic 15. Reliability Assessment for the ALOS Indicator



Graphic 16. Reliability Assessment for the Mortality Indicator



Reporting Bias

Time reporting for AbxMin indicator was reviewed to assess whether any reporting bias might exist as providers "round" their estimates of the time to receiving antibiotics. 60% of the time, time responses were recorded as a "0" or "5," which is greater than if assumed to be by chance.

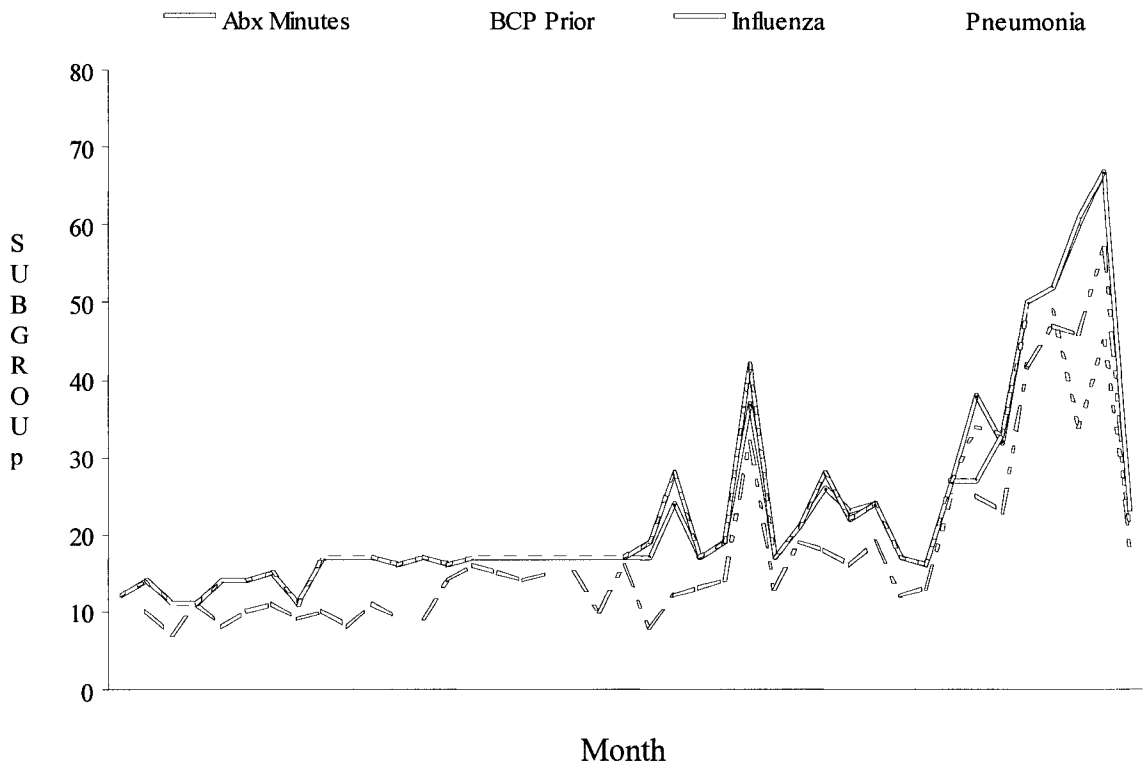
Table 21. Reporting Bias

	Count	*Percent*
1-4, 6-9 Digit Recorded	358	37.9
0, 5 Digit Recorded	571	60.5
Missing	15	1.6
Total	944	100

Subgroup Size

The sample size of each subgroup was found to vary for all of the process indicators. A large variation in subgroup size could reflect a possible confounding relationship, as reduced variation could result from the CAP event, from the sample size increase, or both. Sub-group size line graphs are shown in Graphic 17.

Graphic 17. CAP Process indicators: Subgroup Size



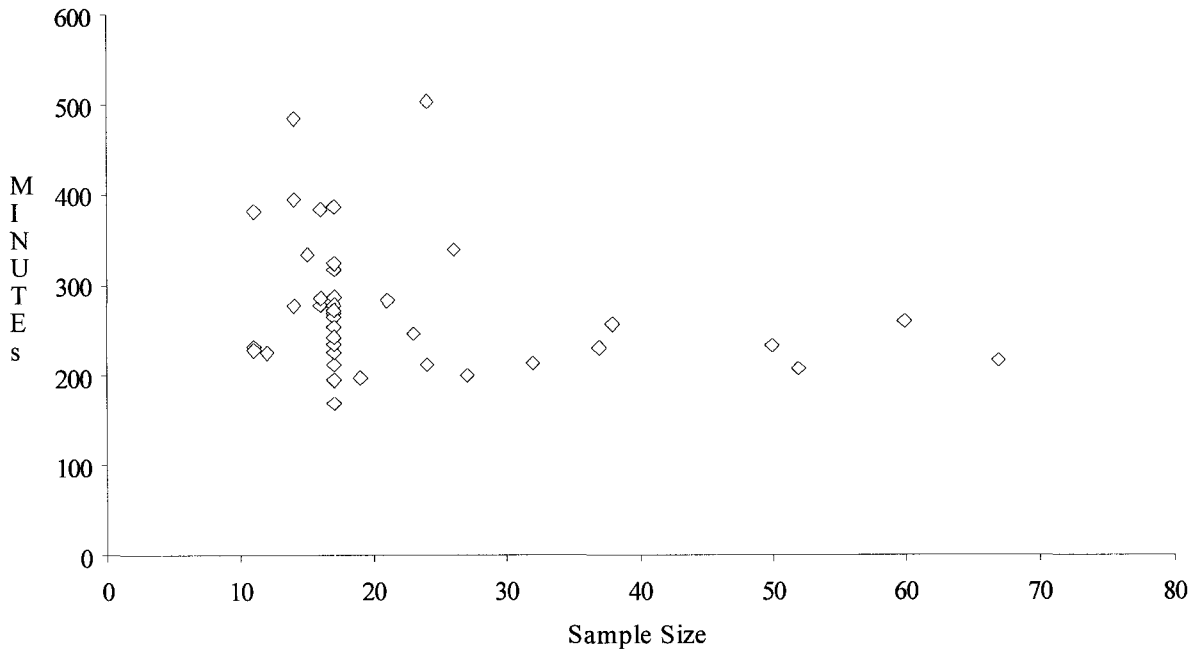
Process variables showed marked changes in sampling strategies, beginning around Oct 01. Table 22. shows the range in subgroup size of about 50 patients. Follow-up analysis was completed to evaluate possible associations between sample size and an increase in average or variation. Significant associations were observed between sample size and average indicator values. Of interest is that, for the Time-to-Abx. and BCP-Prior indicators, significant correlations were negative, indicating sample size increases being associated with decreasing average values, while the associations with variability were positive. As sample size increased, variation also increased. Larger samples would be expected to increase indicator precision and reduce control limits.

Table 22. Subgroup Size Statistics

Indicator	*Median*	*Min.*	*Max.*	*r.* *n : Avg.*	*r.* *n : Std. Dev.*
Abx. (transformed)	17	11	67	-.385 (.013)	.285 (.071)
BCP_Abx.	14	7	57	-.302 (.055)	.375 (.016)
Influenza	17	11	67	.556 (.000)	n.s.
Pneumonia	17	11	67	.618 (.000)	n.s.
Abx._Consistent	16	11	17	n.s.	n.s.

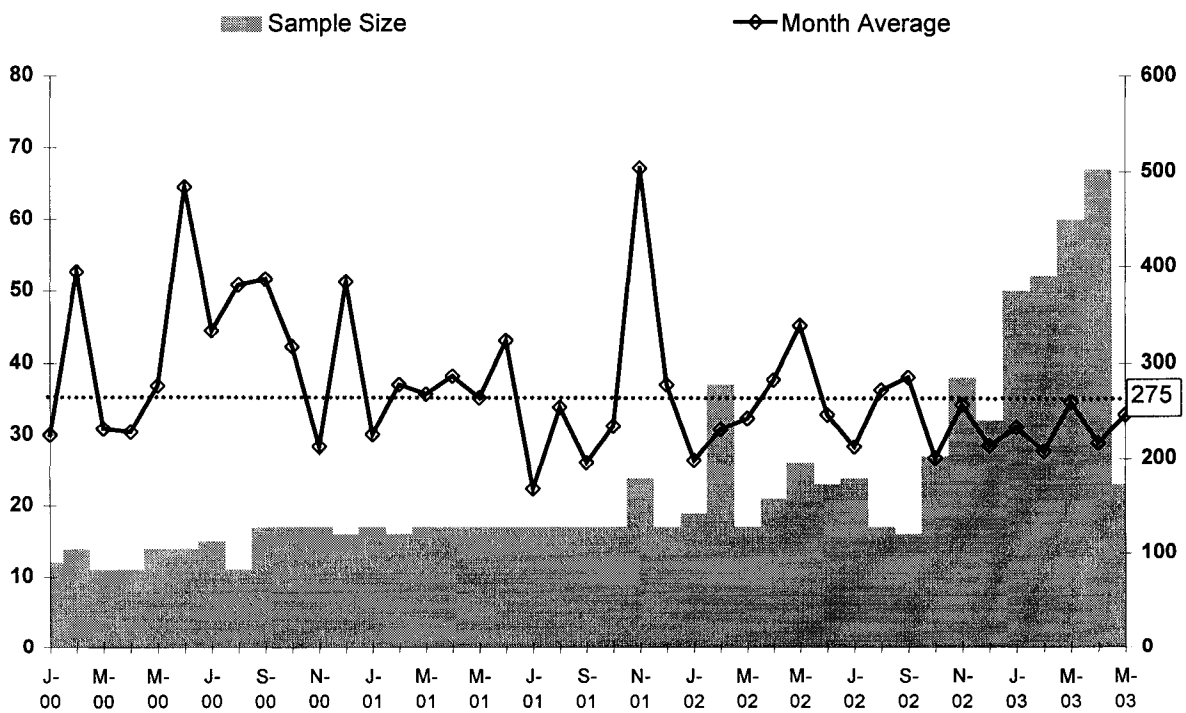
The following scatter plot shows the relationship between larger sample sizes and reduced average (Abx.) values. The critical cutoff appears to be around n=25.

Graphic 18. Scatter Plot of Monthly Time-to-Abx. and Subgroup Size



Significant change in subgroup size challenges the straightforward interpretation of control charts as sample size is used to compute control limits. Normal chart operation requires that control limits be re-calculated as subgroup sizes are changed. The HSAG-to-JCAHO (July 2002) event increased subgroup sample size from an average of 15 to almost 60, the result of reporting all CAP patients instead of only a selected sample.

Graphic 19. Association of Subgroup Size (bars) with Monthly Time-to-Abx. Values (line)



Missing Data

Process indicators were checked for missing data problems. 89% of all patients had no missing data: 110% had one missing value; .5 had two missing values; and .2 had 3 missing values.

Missing data on the Pneumonia Indicator was significantly associated with missing data on the Influenza Indicator. There was an excessive number of missing values in Nov 02 for the BCP-Prior and Influenza indicators. Missing data begins to affect BCP-Prior Indicator in Jun 02 and, beginning Jul 02, nine months have >15% missing, with a maximum 35% missing in Nov 02. Process variables with a large amount of missing data were dropped from most analyses.

Table 23. Indicator Missing Values

Indicator	Interpretation
Abx. Minutes	Minimal missing values
Antibiotic consistent with HSAG guidelines	Minimal missing values. Data collection discontinued after Dec 01. Variable dropped.
Blood Culture Prior to Antibiotic	Missing values escalate beginning Jun 02. After Jul02, nine months with greater than 15% missing, maximum in Nov 02 with 35% missing.
Influenza vaccine (documented) given	Nov 02 excessive amount of missing data.
Pneumonia vaccine (documented) given	Minimal amount of missing data.
Guidelines followed	Partial data collection. Variable dropped.
Blood Culture Prior to leaving ER	Data collection initiated in Jun 02. Excessive amounts of missing data ranging between 18-82%. Variable dropped.

Table 24. CAP Indicator Statistical Quality Rating
(1=Problems Identified, Low Quality 5=No Problems Identified, High Quality)

Indicator	*Seasonal*	*Missing Data*	*Shape*	*Autocorrelation*	*Control*	*Association*	*Subgroup Size*	*Trend*	*Stationary*	*Average*
AbxMinTransformed	5	5	5	3	5	5	2	1	1	3.5
Abx. <4 Hours	5	5	4	5	5	3	2	1	1	3.4
AbxMin	5	5	3	3	5	4	2	1	1	3.2
Abx. <8 Hours	5	5	1	3	4	3	2	1	1	2.8
BCP-Prior-Abx.	5	2	1	5	2	2	2	1	1	2.3
Pneumonia Documented	5	5	1	1	1	3	2	1	1	2.2
Influenza Documented	5	4	1	1	1	3	2	1	1	2.1
ALOS	5	.	5	5	4	2	.	5	.	4.3
Mortality	5	.	5	1	5	5	.	5	.	4.2
Readmission	5	.	3	4	5	1	.	1	.	3.2
Average	5	4.4	2.9	2.9	3.7	2.6	2	1.8	1	

Notes: Lack of stationarity, significant trend, and non-constant subgroup size present the greatest challenges to the control chart model. Mortality and ALOS outcome indicators are least affected by data quality problems.

Chart Calibration

Strategic choices guide the design and operation of control charts. Calibration during the initial start-up involves accommodating data character, signal events and their related effect size, and statistical power. During chart operations, calibration involves making decisions in the context of probability and risk, cost and analytic benefit. Calibrating control charts provides the opportunity to evaluate how control charts function under alternative conditions; it represents the methodological strategy of multiple operationalism.

A 5-step review process was applied to guide calibration of the CAP indicator data: (1) Pattern Analysis, (2) Anticipating Shifts, (3) Intentional Sampling, (4) Process Capability, and (5) Operational Effects.

Pattern Analysis

Random data do not show dependable patterns. The existence of a data pattern implies some causal force that leads to a cluster, trend, or cycles. In psychology, pattern analysis has been advocated to support the analytic process as an incremental approach to establishing construct validity (Trochim, 2003). The analyst attempts to match a hypothetical explanation to an observed dataset.

The control chart provides a formal set of procedures to evaluate patterns in central tendency and variation. Zimmerman identifies nine possible directional change

combinations using average and variation charts in tandem to show no-change, increases, or decreases in either chart (Zimmerman, 1999). Montgomery suggests looking for patterns that identify *cycles* (business cycles, machine fluctuations, etc.), heterogeneous processes (showing extreme changes between opposite control limits), process shifts (represented as a step function), trends, and center-line stratification (reflecting incorrect calculation of control limits) (Montgomery, 2001).

Documenting indicator behavior seems essential to the successful application of control charts to improve medical quality. In addition to the descriptive procedures used in the preceding section, some kind of strategy to classify observed variation would be useful to understanding patterns and anticipating changes. Variation was studied in the CAP-indicator dataset, and results are summarized in Table 25. As expected, most of the changes that occurred were small, within a single standard deviation. There were a number of larger shifts: about 33% of all changes were >1 and only 4% were >2 std. dev.. Table 25. shows there were slightly more negative runs; there were two occasions where change from one month to the next exceeded 3 standard deviations.

Table 25. CAP Indicator Variation Assessment (in Std. Dev. Units)

Indicator	*(-3)-(-2)*	*(-2)-(-1)*	*(-1)-0*	*0-1*	*1-2*	*2-3*	*3-4*
n=41							
AbxMin	-	3	22	10	4	1	1
AbxMin Transformed	-	5	18	10	7	1	-
Abx.<4	1	8	8	17	7	-	-
BCP_ATB	2	5	12	12	10	-	-
n=35							
ALOS		7	11	11	5	1	-
Mortality ICD-9	-	5	13	11	4	2	-
n=30							
Sev.-Adj. ALOS	-	6	9	10	4	1	-
Sev.-Adj. Readmission	-	4	11	12	2	-	1
Sev.-Adj. Mortality	-	3	14	9	3	1	-
Total	3	46	118	102	46	7	2

Effect Size

Common cause variation is expected in every data series, and, in the medical environment, signals are also probable. Stated more directly, "Shift happens." The effect size (ES) can be calculated as the amount of change in a dependent variable related to an independent variable and provides a quantitative estimate of measured effect. ES measures are computed according to the research situation and data characteristics. It may be calculated as (1) absolute measures (i.e. Pearson correlation coefficients (*r*), regression coefficients (*b*), by standardized mean differences (*d*)) or (2) relative measures [odds ratios (*OR*) or Relative Risk (*RR*)]. Cohen (1983) initially suggested standards for evaluating the magnitude of ES (for *r*) and Gliner extended these standards for the standardized mean difference (*d*) where $d = \frac{\overline{X}_1 - \overline{X}_2}{Std.Dev_{pooled}}$.

ES is a signal detection device where the signal is viewed as the difference between observed values, divided by noise, the amount of observed variation. Table 26. provides a gauge for evaluating the magnitude of ES measures using the correlation coefficient and the standard mean difference.

Table 26. Interpreting the Magnitude of Effect Size

Small effect	$d = .2$	$r = .1$
Medium effect	$d = .5$	$r = .3$
Large effect	$d = .8$	$r = .5$

In this Study, two statistical procedures were employed to evaluate the impact of hospital events on each CAP indicator: (1) standardized gain scores for stationary series and (2) time series regression for trended series. The results from both analyses are included in the following Tables. Hospital-related events had modest impact (small effects) on the stationary indicators (range 7-16% for the process indicators).

Table 27. Hospital Event Effect Sizes for Stationary Indicators: Standardized Mean Differences

Event	*n t1(t2)*	*Mortality <80*	*n t1(t2)*	*Abx.*	*BCP_Abx.*
Chart Review Changes	7 (28)	0.12	13 (18)	0.15	0.12
Indicator Profile by Physicians	12 (23)	0.01	19 (12)	0.09	0.07
HSAG-to-JCAHO Definitions	25 (10)	0.07	31 (10)	0.11	0.16

Note: t1 = time 1, number of months before event
t2 = time 2, number of months after event

Table 28. contains ES for trending indicators. A truncated dataset was used for the process indicators to remove the effect of sample size changes. Changes in slope values were mostly small and frequently non-significant. The HSAG Definition Changes and Chart Review events had the strongest effect on the Time-to-Abx. (<4 Hours) and Morality indicators. When standardized coefficients are compared, the Chart Review event appeared to have the strongest effect. Overall, the hospital events appear to have had small effects.

Table 28. Quality Event Effect Sizes Calculated as Regression Coefficients

Indicator	*Event*	*n t1 (t2)*	*Estimated Strength*	*sig.*	*95% CI*		*Relative Strength*
					Low	*High*	
Time-to-Abx. < 4 hours	Chart Review*	13 (18)	0.111	0.040	0.01	0.2	0.37
	Profiles*	19 (14)	0.082	0.139	-0.03	0.2	0.27
	HSAG-to-JCAHO Changes	31 (10)	0.005	0.007	0.00	0.0	0.42
Time-to-Abx. < 8 hours	Chart Review*	13 (18)	0.106	0.017	0.02	0.191	0.424
	Profiles*	19 (14)	0.068	0.141	-0.024	0.161	0.271
	HSAG-to-JCAHO Changes	31 (10)	0.040	0.324	-0.041	0.122	0.158
Readmission	Chart Review	7 (28)	0.013	0.578	-0.04	0.06	0.10
	Profiles	13 (22)	-0.009	0.644	-0.05	0.03	-0.08
	HSAG-to-JCAHO Changes	25 (10)	-0.020	0.357	-0.06	0.02	-0.16
ALOS	Chart Review	7 (28)	1.3	0.578	-3.5	6.2	0.10
	Profiles	13 (22)	-0.9	0.644	-5.0	3.1	-0.08
	HSAG-to-JCAHO Changes	25 (10)	-2.0	0.357	-6.2	2.3	-0.16
Mortality (DRG)	Chart Review	7 (28)	-0.069	0.000	-0.10	-0.04	-0.65
	Profiles	13 (22)	-0.045	0.002	-0.07	0.00	-0.51
	HSAG-to-JCAHO Changes	25 (10)	0.0	0.068	-0.06	0.00	-0.31

* Computed on reduced data series to eliminate effect of process change

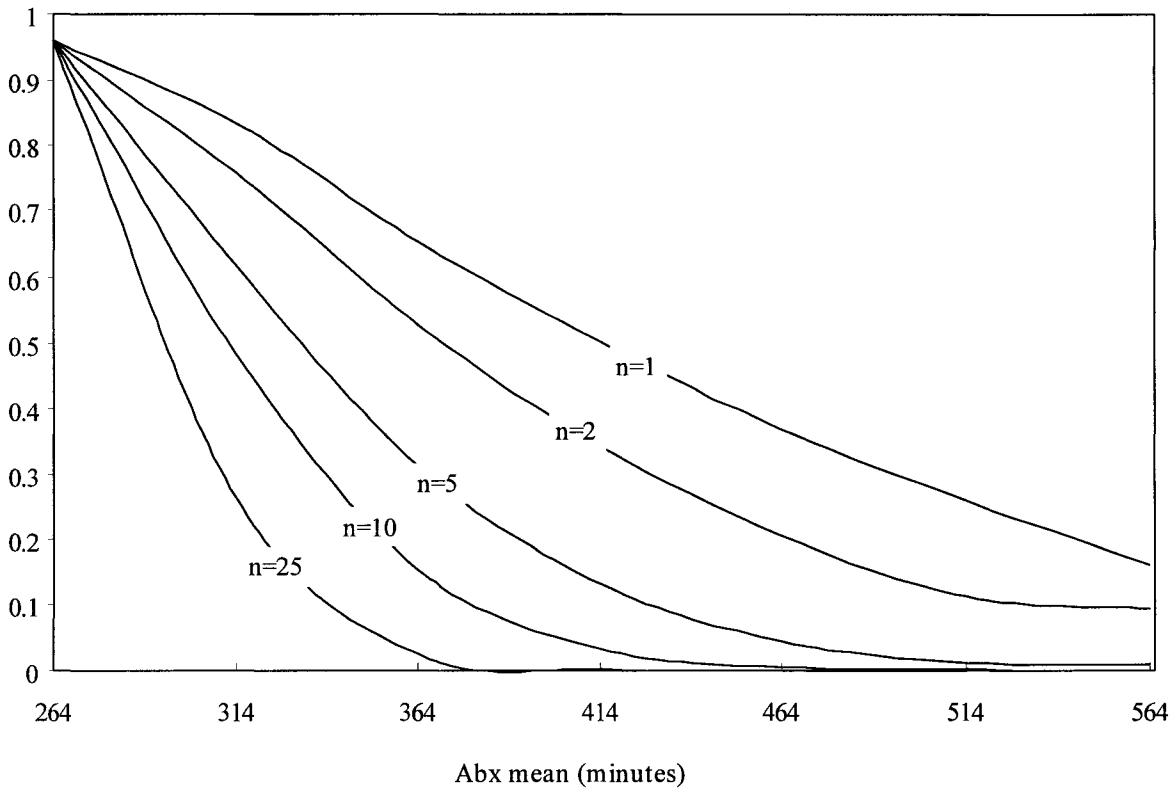
Control charts help determine whether the underlying process is controlled. In making this determination, two kind of errors can occur: Type I (α) and Type II (β). Type I error, referred to as "false alarm," occurs when a datapoint (or series) is identified as special cause, when it really is common cause. It involves the probability of rejecting H_0 when it is true and, perhaps (1) triggering a potentially expensive "drill down" search for the special cause, or (2) tampering with the process (over-control), making it less stable in the future. Type II error occurs when medical processes are assumed to be operating normally when, in fact, special cause variation does exist. Type II error is calculated as the probability of failing to reject H_0 when it is false. It has been defined as "consumer's risk," accepting poor quality or allowing a process to operate in an unsatisfactory manner, resulting in operator over-confidence.

Statistical power ($1 - \beta$) concerns the ability of the chart to correctly identify special causes, to correctly reject H_0 when it is false. Like other statistical methods, power is influenced by (1) sample size, (2) the magnitude of expected difference between sample and population, (3) Type I error, and (4) the observed variability. The further the true mean is from the hypothesized value (i.e. the larger the value of δ), the smaller the probability of Type II error for a given n and α. That is, for a specified sample size and α, the test will detect large differences more easily than small ones. As the sample size increases, the probability of a Type II error decreases for a specified δ and α. The OC Curve presented in Graphic 22., calculated from the AbxMin indicator applied to the XmR chart, shows these relationships.

Graphic 20.

Operating Characteristic Curve - Abx Xbar Chart

(True avg. 264, sd 221, .001 Limits, 2-sided)



Control charts are frequently constructed using 3-sigma control limits, meaning that under conditions of normality, 99.7% of the data will lie inside +/- 3 standard deviations of the mean, and 0.27% of fall outside these limits. To monitor infection control, a more sensitive 2-sigma limit may be used, increasing the amount of data falling outside control limits to 5% (in a normal curve). As is described in the Descriptive Analysis Section, only 9 of 35 variables reviewed exhibited normally distributed conditions. Most variables are either skewed or J-shaped, reflecting time-to-event (non-symmetric) or mortality (infrequent occurrence).

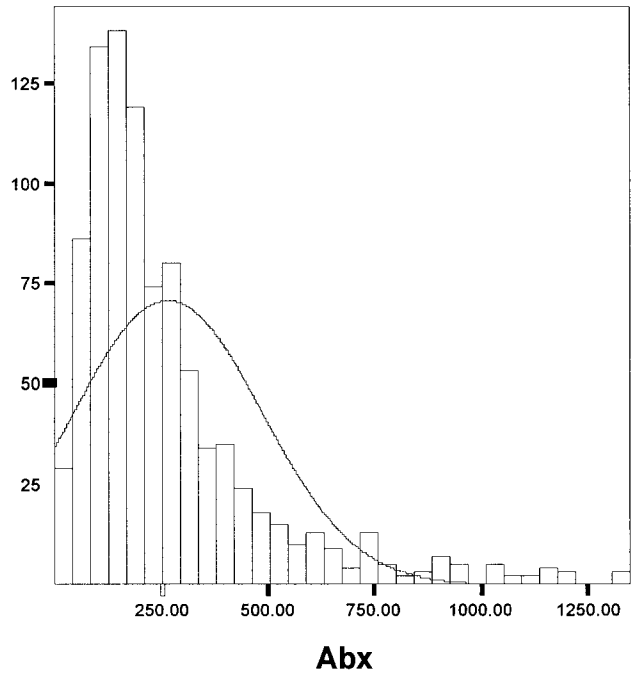
Non-normal distributions affect alpha probabilities. For example, the AbxMin indicator has a positive skew value of 2.1. The Shapiro-Wilk test p-value (p=.000) indicates that AbxMin is a non-normal distribution. Under a normal distribution, 0.27% of the 929 observations (n=28) would be expected to exceed 3 standard deviations. As summarized in Table 29., 61 observations (6.5% of the total) exceeded 3 standard deviations. For the AbxMin variable, the n=28 expected cases corresponds to the 97.2% percentile and a value of 910 minutes. A 99.7 percentile would result in 3 expected observations with a corresponding Abx. value of 1320 minutes.

Table 29. Assumed and Observed Distribution Values for AbxMin Indicator

	Average	*Dispersion*	*Upper Limit*	*# Excess*
Assumed (Normal) Distribution	264 (Mean)	222 (Std. Dev.)	666	61
Observed (Skew) Distribution	200 (Med.)	205 (IQ Range)	910	28

Non-normal distributions can result in misplaced control limits. The following graphic imposes a normal distribution on the AbxMin indicator. This graphic shows that if the upper control limit were placed at 3-standard deviation limit of 790 minutes, it would result in 4% of all values exceeding the control limit.

Graphic 21. Normal Distribution Assumptions Assigned to AbxMin Indicator



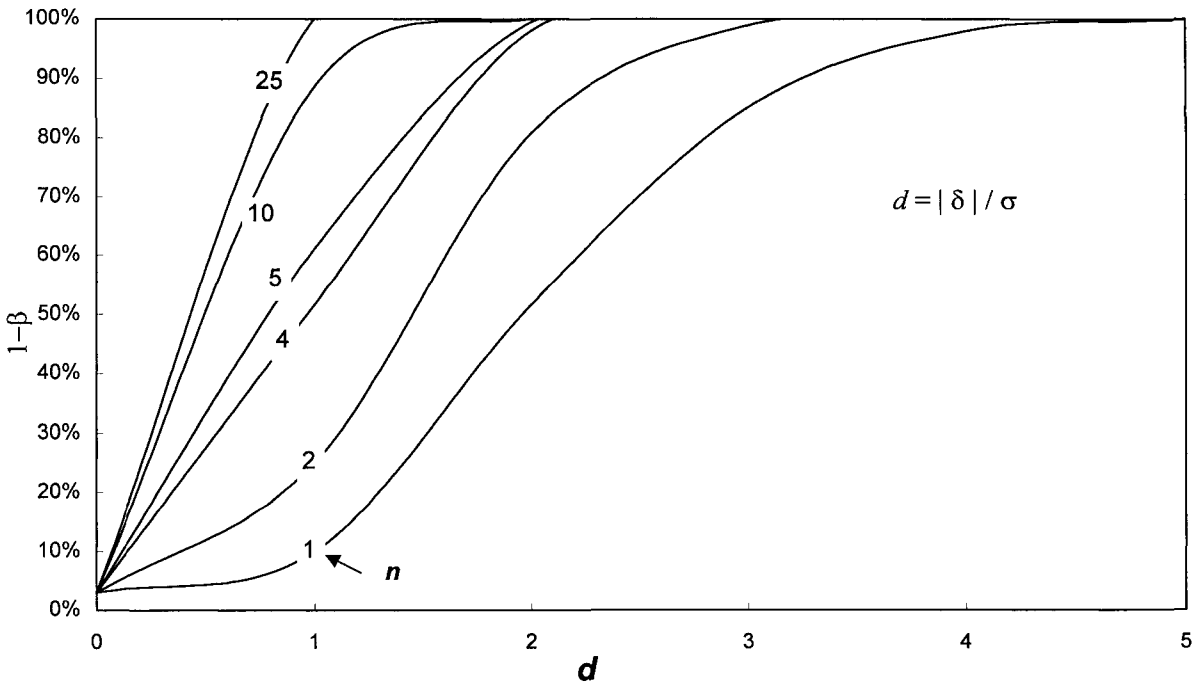
Intentional Sampling

The n = 4,5 standard was designed for high volume, low variation production processes; it was not intended as a general prescription for every condition. Shewhart wrote about sample size, "if we wish to reduce the chance of making an error in estimating the probability associated with chosen tolerance limits, *there is no royal small sample road for doing this*" (italics added) (Shewhart, 1939). Sample size is determined according to how much process variation is evident, the importance of the need to detect magnitudes of change and what are the risks involved, the choice of statistics (with continuous

variables being more powerful), and the frequency of sampling. When adapting control charts to medical quality processes, Lee writes, sample sizes of 10 would be considered small (Lee, 2002). Larger samples will improve statistical precision.

Graphic 22. Power Assessment for AbxMin Indicator

(Operating characteristic Curve for Two-sided Normal Test, Alpha=.05)



The sampling process is perhaps the greatest single issue in calibrating charts, as it is a function of design that can be controlled, and it directly influences chart power. This can be observed in the following graphic that evaluates chart power for the AbxMin variable. Two conclusions can be drawn. First, the further the true mean μ_1 is from the hypothesized value μ_0 (i.e. the larger the value of *d*), the smaller the probability of Type II error for a given n and α. That is, for a specified sample size and α, the test will detect large differences more easily than small ones. Second, as the sample size n increases, the probability of a Type II error decreases for a specified *d* and α. That is, to detect a specified difference, the test can be made more powerful by increasing sample size.

Formulas are available for determining an adequate sample size for different chart types. Montgomery provides a formula for P-charts (Montgomery, 2003).

$$n = \left(\frac{L}{\delta}\right)^2 p(1-p)$$

where
n = Desired sample size
L = Length parameter (σ limits)
δ = Change to detect
p = Proportion

For the Abx.<4 Hours indicator, the observed proportion was (p=.584) during the 41-month observation period. The distance parameter (*L*) is set to 3 sigma above and below the average process value. The overall monthly range in observed Abx.<4 Hours was .3-8, showing a relatively high amount of variation, although most month-to-month changes

were in the .1-.2 range. If the amount of detectable change was established at (δ=.30), the sample size needed would be (n=24).

Montgomery pointed out that this equation can be algebraically manipulated to estimate the effect size that the P-chart could detect $\delta = L\sqrt{p(1-p)/n}$ (Montgomery, 2003). During the 41-month observation period, the average volume of CAP patients was approximately (n=35); the P-chart is able to detect a (δ=25%) change in Abx.<4 Hours. A larger sample size produces greater sensitivity by generating more narrow control limits.

Type I and II error probabilities can be used to calibrate sample size and subgroup frequency decisions. The Average Run Length for in-control conditions (ARL_{IC}) is the average number of samples, or subgroups, that a chart will run before it signals an out-of-control condition. Conversely, the Average Run Length for out-of-control conditions (ARL_{OC}) is the expected number of subgroups before an out-of-control condition will be detected. Finally, the Average time-to-signal (ATS) provides an estimate of the length of time the chart will run before signaling. These methods compute a chart's signal detection capability according to anticipated probabilities and changes observed; Table 30. presents computational strategies.

Table 30. Calibration for Chart Signal Detection Capacity: ARL_{IC} ARL_{OC} & ATS Formulas

Function	Formula
Average Run Length for in-control conditions	ARL = 1/p Where p= probability point exceeds limits
Average Run Length for out-of-control conditions	1/1-β for first subgroup signal
Average Time-to-Signal	ATS = ARL / h Where h = time interval

Using the AbxMin indicator as example, for 3σ limits a single test rule (1 value exceeds limits), there is a (p = .0027) probability that any single point falls outside the limits when the process is in control. That is, ARL = 370 when there is no difference (δ=0) but the ARL rapidly falls as observed differences increase, especially for an Individual X-Chart. Even if no change occurs, an out-of-control signal or false alarm will be generated every 370 samples, on the average. β-risk (β_R) computes the probabilities for the ARL_{OC}. The β_R (probability of not detecting a signal) is related to sample size and shift effect size. Again, using the Abx. example, the probability of not detecting a shift in variance and average using a single test, for an Individual X-chart to detect a 2δ (2*observed standard error) change, is .84. This results in an ARL_{OC} = 6.25. That is, the X-chart can be expected to take up to 6.25 subgroups (months) to detect a change of 2σ (i.e. a change from 264 to 706 minutes).

Table 31. Average Run Length (ARL_{IC}) and Probability of Not Detecting Shifts (β_R) Estimates for Varied Samples and Shifts (1-of-1 test)

		n	*1*	*2*	*5*	*10*	*26*
Function	Size of shift						
β_R	1σ		.9772	.944	.778	.436	.029
	2σ		.8413	6.30	.070	.000	.000
	3σ		.5000	2.00	.000	000	.000
		0					
ARL_{IC}	1σ	370	44	17.7	4.5	1.8	1.0
	2σ	370	6.3	2.3	1.1	1.0	1.0
	3σ	370	1.4	1.1	1.0	1.0	1.0

The Average time to signal (ATS) estimates the number of samples required to detect process shifts. Using the Abx. example, the β_R probability is .84 that a process shift of 2σ will not be detected at the first subgroup after the shift (1-β =.16 that it will be detected). The ARL of detecting this kind of shift is 6.25 subgroups (months). If the frequency of subgroup data were reported as weeks, the ATS would decrease to around 6.5 weeks.

Process Capability

Benchmarks are national, regional, or local best-practice standards used to establish performance goals toward which improvement action is directed. Different benchmarking models have been identified using internal (in-hospital units), cross-industry (other industries e.g. intake at hotels), competitive (other hospital), and strategic (industry standard) comparisons. Medical benchmarks are quality "specifications" that determine whether the product or service meets minimum standards; they are not the observed process parameters. Even controlled processes may not operate according to clinical or administrative needs. The process may not be centered where needed, or it may have more variation than desired.

Table 32. CAP Benchmark Specifications

Indicator	Observed	Benchmark	Source
Average (minutes) Time-to-Abx.	275	169	MIDAS CDB
Abx. < 4hours	58.4%	77.6%	MIDAS CDB
Abx. < 8 Hours	87.2%	95.5%	MIDAS CDB
Blood Culture Prior to Abx.	83.6%	89.7%	MIDAS CDB
Influenza Vaccine	57%	64%	MIDAS CDB
Pneumonia Vaccine	58.9%	64%	MIDAS CDB
Guidelines	14.8%	80%	CHN Goal
ALOS (ICD-9)	4.3	5.1 Days*	Midas CDB
Mortality	4.4%	4.23%	MIDAS CDB

* Depends on payer source

Specifications are assigned rather than being a measured part of the medical process. And while they can represent a reasonable goal to achieve, the existing clinical and/or administrative processes may not support the consistent achievement of any medical specification. Formal engineering methods have been developed to assess process capability (Bothe,1997). Stability and process control are necessary to process parameters and to measure process capability.

Operational Effects

Supplementary run rules, including Western Electric and Jaehn's Zone rules, described earlier (Table 5, pg. 55), are used to increase chart sensitivity. However, they do so at the cost of misclassification and reduced specificity. The routine use of supplementary run rules has been criticized for capitalizing on chance and increasing the probability of false alarm (Montgomery, 2001). The trend rule (6 or 7 continual values increasing/decreasing) has been criticized for not providing unique information (Davis, 1988). For these reasons, Lee recommends limited use of 3 rules (#1, #4, #5) for control chart applications to medical care (Lee, 2002).

The Zone Rules are used to identify three general categories of causal pattern. The first category is the excessive datapoint(s), one subgroup average that exceeds 3σ limit above or below process average or a few excessive values close to the Upper/Lower Control Limit. These patterns are likely associated with a single powerful event or the precursor of a process shift. A second category of signal patterns is positive or negative trend,

identified as seven or more consecutive increasing or decreasing subgroups. This non-random condition reflects gradual process drift or breakdown and, in manufacturing, has been associated with machine degradation. The third pattern category includes alternating values stratified between 1σ-3σ (may be a sign of incorrect calculation of control limits) and 5 points in a row outside zone $0 \pm 1\sigma$ (may show conjoining of two distinct distributions) (Quesenbery, 1997). The following table shows that the Zone Rules were most sensitive to CAP indicators.

Table 33. Average and Variation Charts Patterns Associated with Assignable Causes

Rule	*AbxMin*	*ALOS*	*Mortality*
Probability Rules			
Points Outside 3-Sigma Control Limit	4	1	
2-out-of-3 Consecutive Points in Zone A			
4-out-of-5 Consecutive Points in Zone B or Beyond			
Run Rules			
Less Than 30% Runs Across Centerline			
8-or-more Consecutive points, same side of Center Line			
7-or-More Points Increasing/Decreasing (> 21 data points)			
14 Successive up-and-down Alternative Points (Saw Tooth)			
Zone Rules			
More than 8 Points	12		1

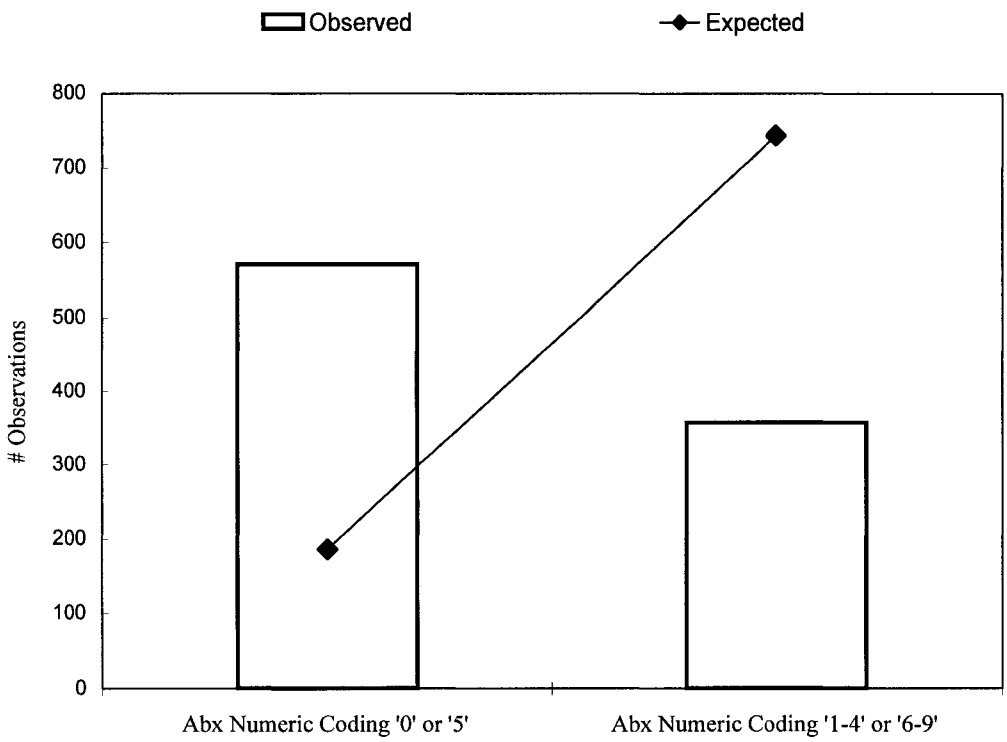
Essential to control chart sensitivity is the estimation of process average (computing the centerline) and process variation (specifying control limits). Different methods of determining these parameters from a minimum number of subgroups, around 25, have been proposed, including using a baseline estimate against which to evaluate ongoing control. This is the method used in computing the CAP-chart sensitivity. However, there are several problems with this approach. The principal problem, using monthly reported subgroups, is the evaluation of current processes using outdated data, more than 2 years old. Under conditions of rapid organizational change, medical quality and administrative conditions that are two years old will not likely reflect current operating processes.

In practice, chart parameters are revised in a number of ways by disabling outliers and non-random data series and by the ongoing re-computation of control limits. Some CC users claim that ongoing revision more easily enables better modeling of current conditions. Some traditionalists warn against ad-hoc, arbitrary methods of estimating process parameters. The SPC literature provides inconsistent guidance on these issues. Montgomery recommends disabling points only if the underlying assignable cause is clearly evident (Montgomery, 2001). The routine elimination of subgroup values will, in the short term, reduce process variation and lead to more false alarms because of smaller control limits. If the underlying special cause has been identified, values were generated by an alternative process and should be disabled. General practice recommendations are that variation charts be checked first, as they give the preliminary justification of points to disable.

There are important process adjustment decisions, such as whether to eliminate extreme values that may represent false alarms and when to re-adjust the process parameters. In these cases, different practitioners can easily make different decisions, or the same practitioner can make different decisions on different occasions. Again, this represents a challenge to reliability and statistical conclusion validity. It has been stated that new control limits should be calculated only after a process change has been introduced, and, when re-calculating new control limits, should only include those datapoints collected subsequent to the change (TQT Manual). Montgomery suggests that control limits and centerlines be periodically revised at regular intervals, while Palm writes that limits should be re-calculated after a chart accumulates more than 100 serial observations. If deliberate authorized changes and improvements to a production process causes control limits to become unuseful, then new limits should be recalculated on an as-needed basis (but not routinely).

Chart calibration implies an ongoing process to check data quality and accuracy. Two principal sources of measurement error were identified, including data reporting, or "rounding," and disease coding. The first involves reporting time in round numbers. At some point in recording the Abx. variable, the number of observed responses recorded as a "0" or "5" exceeded the proportion expected. This "rounding" effect likely reduces measurement precision, although no attempt was made to determine whether there was a constant direction of the bias (i.e. tendency to under-report to the lower digit). Graphic 23. shows the relative size of reporting error.

Graphic 23. Measurement Error from Abx. Numeric "Rounding"



A more serious source of measurement error encountered during this research involved disease-coding procedures: use of Diagnostically Related Group (DRG) verses International Classification of Diseases, Version 9 (ICD-9) coding. The first coding system is used primarily for administrative and billing purposes, while the latter is clinical-based, containing more information about co-morbidity and the disease process. Several important differences between the two were uncovered using the CAP dataset. For the CAP-Mortality and ALOS outcome indicators, there were differences in magnitude, trend, autocorrelation, and signal patterns.

The observed differences between DRG and ICD-9 results appear associated with: (1) the exclusion of infants from the DRG-coded indicators, and (2) that the DRG codes contain more pathogen-specified pneumonia. Because of these differences, the ICD-coded averages were higher for both mortality (ICD-9 5.9% verses 4.4%) and ALOS (ICD-9 5.0 verses 4.3 days). Of particular interest in the mortality indicator were the apparent differences in numerators and denominators. For the ALOS differences, DRG-coded days were uniformly fewer than the ICD-9-coded days. There were differences for mortality coding. The mortality denominators appeared to change consistently with the DRG denominators, consistently below those coded using ICD-9. But the numerators seemed to switch places during the time series; toward the beginning (July 2000–Jan 2001) DRG numerators were more frequently larger, while after February 2001 the ICD-9 numerators were larger, perhaps associated with the Jan 2001 Change in Chart Review event.

Another lesson was learned about the potential effects of coding on the measure validity. During the initial phase of the research, specific indicator definitions had not been specified and, consequently, on several occasions during the handling of data, different coded data were exported. This resulted in what was initially interpreted as poor reliability and confounded interpretation. Data were exported on four occasions, producing a reliability range of (r. coefficient) .26-.97 for mortality and .60-.99 for ALOS. The Readmission indicator was reported using both coding schemes, but its reliability was not affected. Handling error and changes to data definitions can easily affect reliability.

Another source of measurement error, also concerning reliability, happens as indicator values change during subsequent reporting periods because of slow information recovery. This kind of problem is associated with the amount of time needed to abstract data from medical records. Overall, data reliability did not change for the process variables (*r*. .94 - .99) but did affect outcome indicators.

Chart Sensitivity

Sensitivity addresses chart *validity:* does the control chart do what it is supposed to do... correctly discriminate special from common cause? Chart sensitivity was evaluated by testing the chart values against specific events.

The following protocols were used to construct and evaluate charts. For the monthly reported data series XmR, Xbar, and P-charts, an initial dataset of m=25 subgroups (Jan 2000–Jan 2002) was used to evaluate data stability, and, when control was evident, centerline and control limits were plotted on a baseline chart. Thereafter, subgroups were added in a sequential manner to identify signals in a manner consistent with what would have been observed if data were recorded in real time. With each additional datapoint, traditional Western Electric Zone Rules and Jaehn's Zone rules were used to identify out-of-control conditions. All signals were recorded (Refer to Tables 35. and 35.) and patterns associated with chart violations were identified (Refer to Table 33.). Event-associated signals were required to be within 3-subgroups of the IV event time period

pre-specified by hospital staff. "True" signals were defined as those that could be associated with a CAP Event. False Alarms were defined as all signals not associated with a CAP Event. Annotated charts were produced for each indicator. Because of the large number of charts constructed, only final charts are included. Average and variation charts are included for the XmR and Xbar-charts. Variation charts are not constructed with P-chart, EWMA or CUSUM. Signals, false alarms, and disabled values are annotated on all charts.

XmR Charts	
Time-to-Abx.	(Graphics 24. & 25.)
Time-to-Abx. (Transformed)	(Graphics 26. & 26.)
ALOS	(Graphics 28. & 29.)
Severity Adjusted ALOS	(Graphics 30. & 31.)

XmR Chart Background

The X-chart is referred to as the "individual" chart, because each subgroup represents a single summary value taken at regular time periods. Each subgroup represents n = 1. The X-chart can be economical, because it eliminates the sampling and accommodates slower production processes. Because of this functionality, Wheeler has recommended using X-charts for most charting purposes, using it as a special case of other charts (Wheeler, 1992). This requires transforming discrete variables into continuous measures.

There are also problems associated with the X-chart, including low sensitivity and concern with non-normal distributions. Low sensitivity to small or moderate change is related to having wider control limits, based on the small sample size, and the use of the moving range (mR) as a proxy estimate for within-group variation. The mR, used to compute control limits, is a between-group variation measure that stands in for within-group variation. The mR represents the range of a group of n consecutive individual measurements combined artificially to form a subgroup of any size (but usually n=2). If special causes occur within the subgroup, the estimate of common cause variability will be inflated, and the sensitivity of the chart will be eroded. When the time between measures is longer, the opportunity for special causes to contaminate the subgroup and inflate estimates of common-cause variation is greater. The XmR is insensitive to moderate process changes and, according to Crowder, has a long lag time (estimated 44 samples to detect a 1σ shift) (cited in Montgomery, 2001).

Non-normal distribution is also of concern. Skewed distributions prompt some authors to recommend that data be transformed or to use the median to estimate process average. Departure from normality will result in poor estimates of the control limits.

X and mR charts are not independent; successive mR values come from subgroups that are overlapping. Care should be taken when interpreting data patterns on the mR chart, because they are correlated. Consequently, mR patterns do not provide information about shifts, and supplementary decision rules were not used to determine mR chart sensitivity.

Time-to-Abx. (Graphics 24. & 25.)

Visual analysis of the Abx. mR chart indicates a reduction in variation, beginning January 2002, not associated with any signals. The Abx. mR Chart does not indicate any significant shifts in variation, although no supplementary rules were applied. The November 2001 datapoint approached the UCL but is not identified as a signal. (Note: The Nov 2001 subgroup was identified as a signal on the hospital-constructed chart).

One false alarm, a zone rule pattern, was identified on the baseline chart. This signal was not associated with the March 2000 event because it reflected an increase in average value whereas the event was specified to have a negative (downward) effect. The signal values remained part of the centerline computation as no identifiable cause was determined and the four data series was assumed part of the normal operating process.

One event was associated with a (8-points below average) signal. Of minor interest is the observation that the signal occurred 1 month prior to the event. The interpretation assigned is that the June 2002 datapoint was, by chance, below average and became associated with the event-related signal. Visual analysis of the X-chart shows a potential shift reduction beginning as far back as Jan 01 with a single Nov 01 spike that represents an apparent shift in distribution.

Transformed Time-to-Abx. (Graphics 26. & 27.)

A base-e logarithm transformation was used to assess differences associated with the skew in the Time-to-Abx. indicator. The mR chart did not show any signals. When compared to the non-transformed charts, less variation is apparent and the Nov 2001 subgroup has been pulled in. One (8-points below average) signal was identified on the X-chart beginning September 2002, associated with the July 2002 Event. This reflects a 3-month lag (signal delay) compared to the non-transformed dataset. The signal on the Abx.-transformed X-chart shows a proportionally greater reduction in variation compared to the non-transformed dataset.

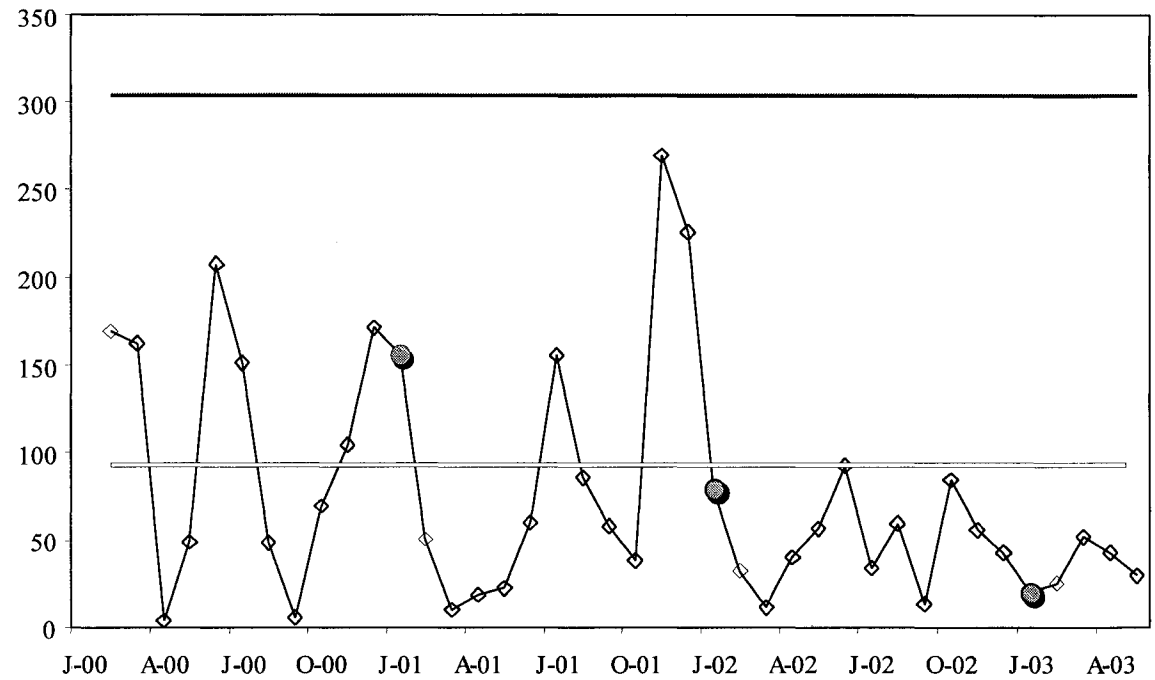
ALOS (Graphics 28. & 29.)

The mR chart does not demonstrate any signal, but a dramatic reduction in moving range variation is evident after the August 2002 period, in association with the HSAG Definition Event. The Baseline X-chart did not demonstrate control; a zone violation was identified beginning October 2001. The Baseline chart was then adjusted, reflecting a process shift at the Oct 2001 time period. The final X-chart identified two signals, the first associated with the Profile event (a 13% reduction in average series values). This signal was associated with the event because it transpired within a 3-month lag period, and it occurred in the negative (downward) direction. The second signal was associated with the Definitions indicator. The ALOS data series have the appearance of multiple trending series that may be better modeled using regression SPC.

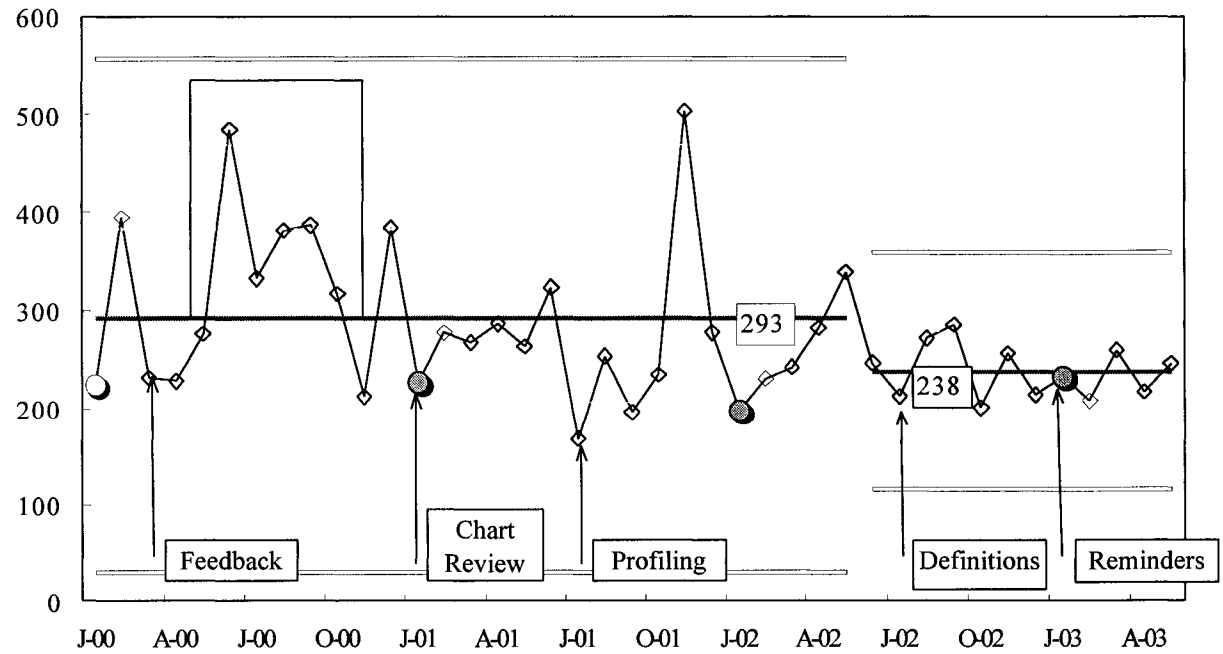
Severity Adjusted ALOS (Graphics 30. & 31.)

Only 30 subgroups are included in the severity-adjusted indicators. The mR-chart does not indicate any signals but does reflect reduced variation during a 7-month period: Nov 00–June 01. One signal was identified on the X-chart associated with the July 01 event (Profiling). This signal, while leading to immediate reduction in ALOS, also appears associated with an increase in series variation. Subsequent to the July 01 subgroup, the data series appeared out of control.

Graphic 24.
Time-to-Antibiotic mR-chart



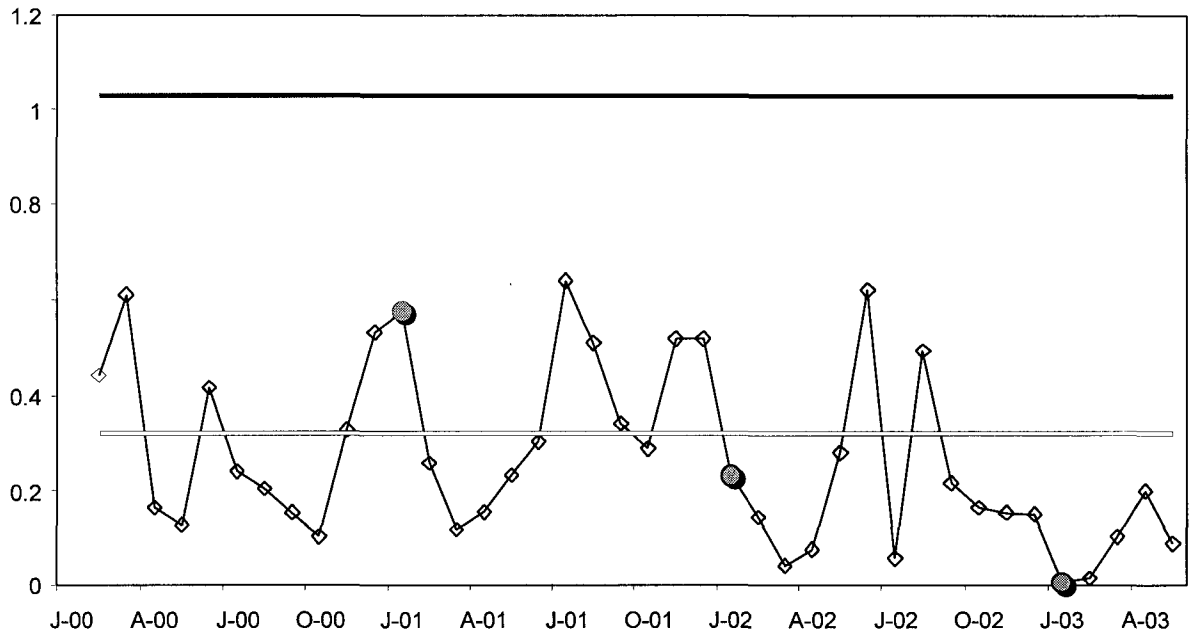
Graphic 25.
Time-to-Antibiotic X-chart
(.001 Limits)



Graphic 26.

Time-to-Antibiotic (Transformed) mR-chart

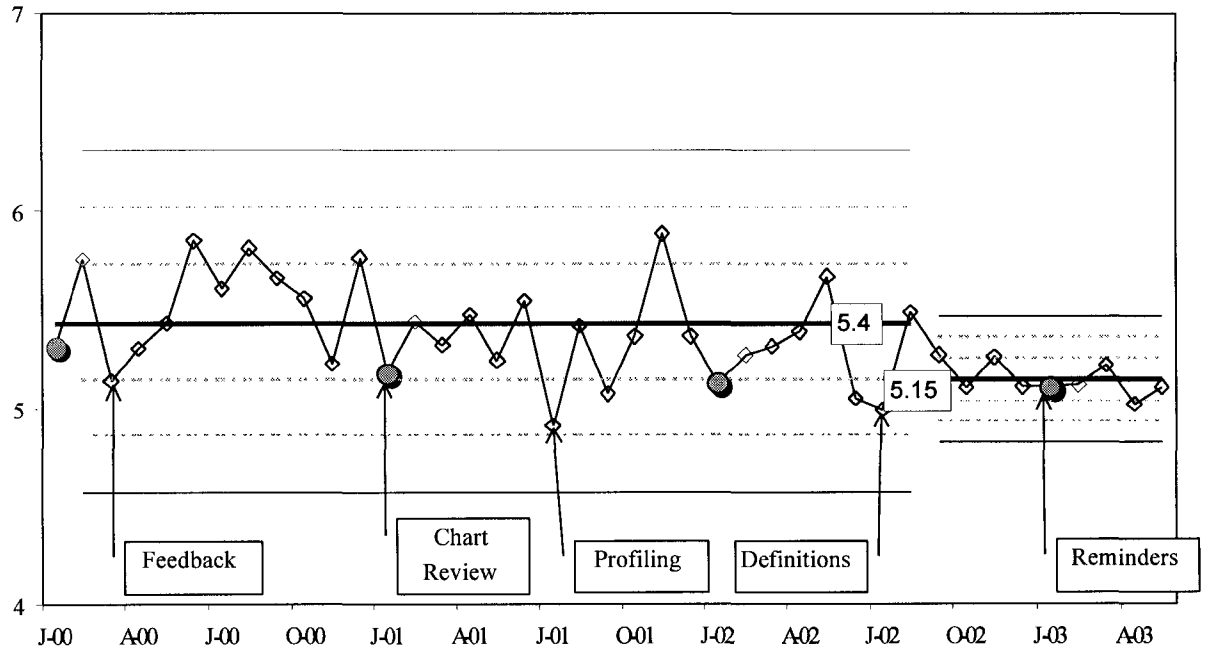
(.001 Limits)



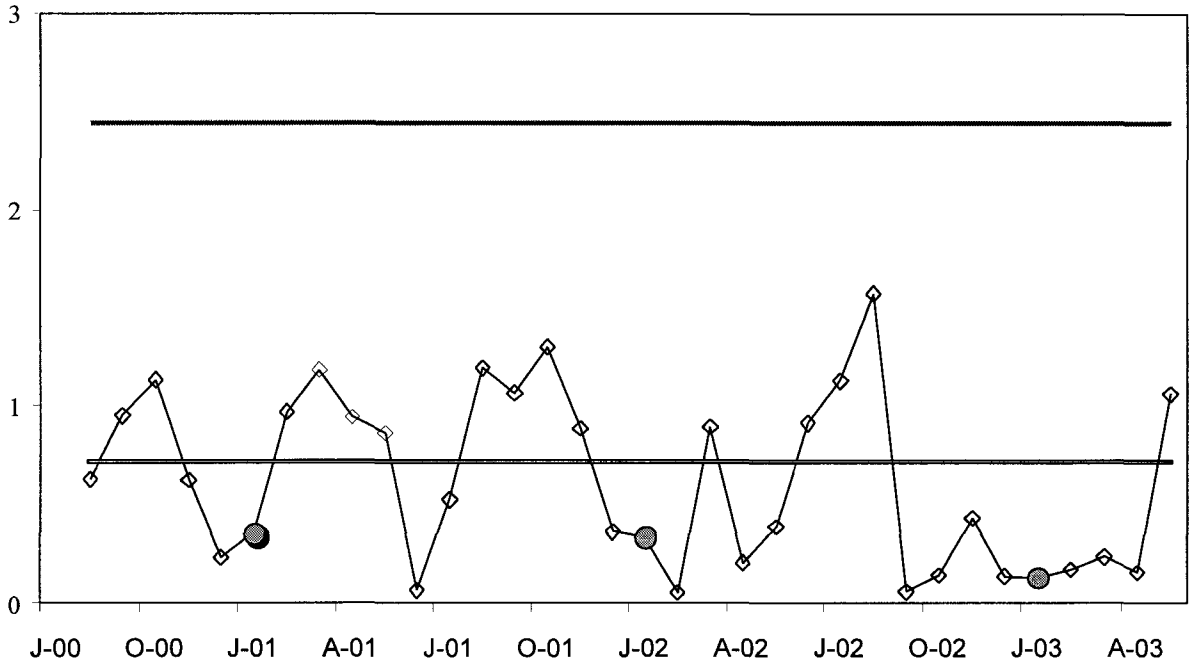
Graphic 27.

Time-to-Antibiotic (Transformed) X-chart

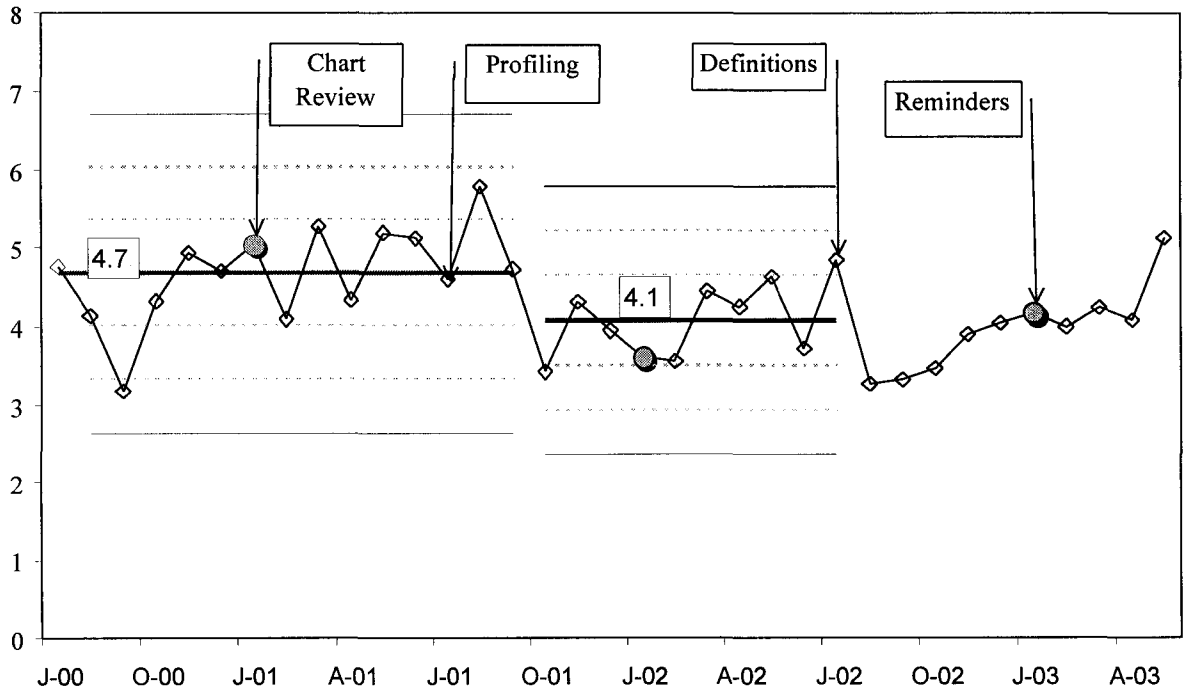
(.001 Limits)



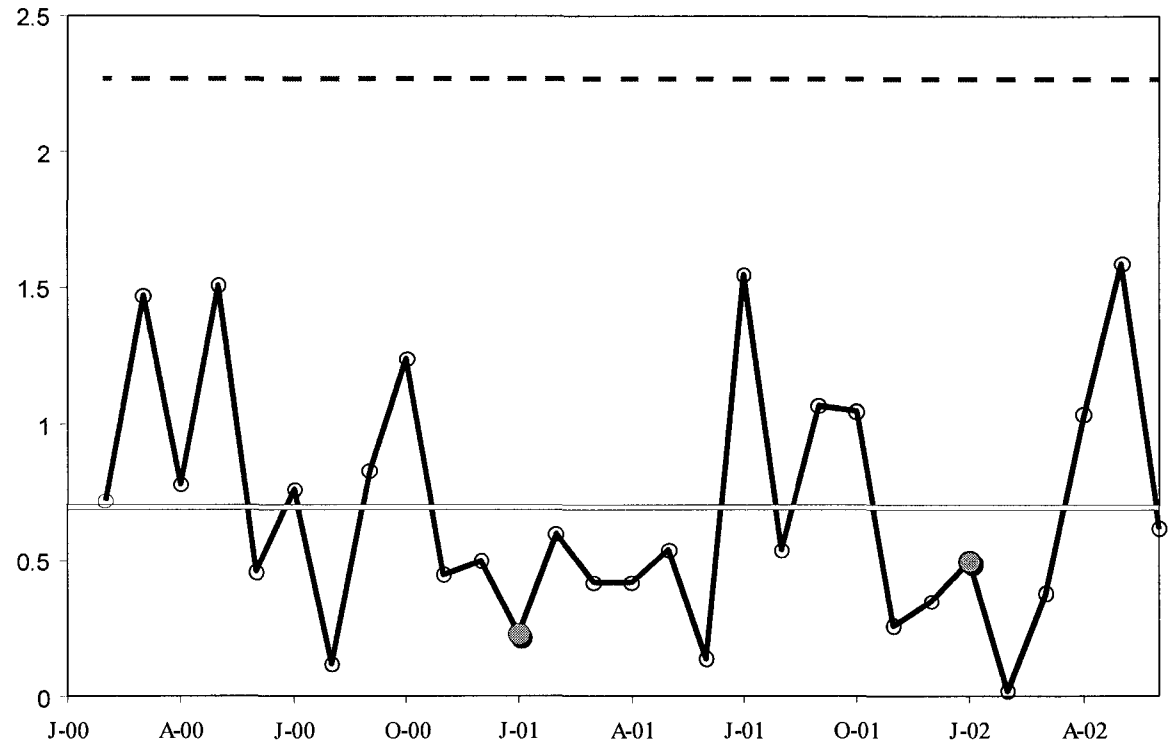
Graphic 28.
ALOS mR-chart
(.001 Limits)



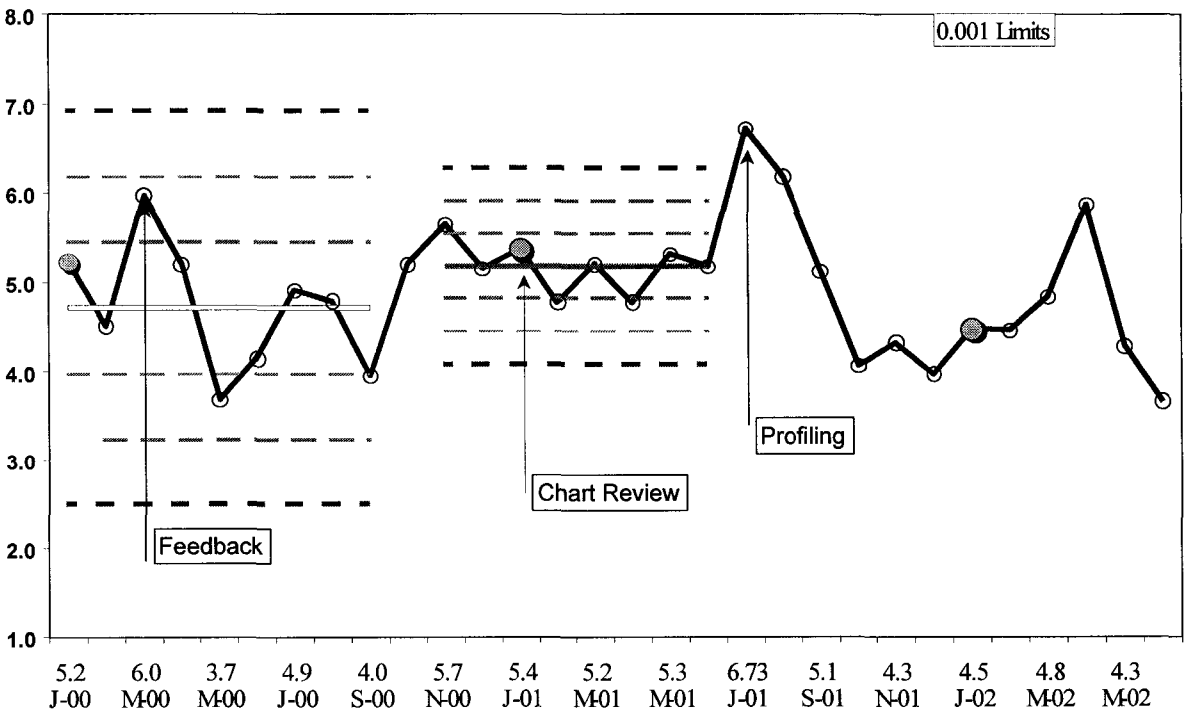
Graphic 29.
ALOS X-chart
(.001 Limits)



Graphic 30.
Severity Adjusted ALOS mR-chart



Graphic 31.
Severity Adjusted ALOS X-chart



$\overline{X}$-S/R Charts	
Time-to-Abx.	
$\overline{X}$-R (n=5)	(Graphics 32. & 33.)
$\overline{X}$-R (n=10)	(Graphics 34. & 35.)
$\overline{X}$-S (n=10)	(Graphics 36. & 37.)

$\overline{X}$-S/R Chart Background

The $\overline{X}-chart$ uses the grand mean from all subgroups $\overline{\overline{X}}$ to represent the process average, centerline, and either the average range $\overline{R}$ or standard deviation $\overline{S}$ for each sample to compute control limits. Sampling is important, because larger samples result in smaller control limits, making the chart more sensitive. Samples of n=4,5 are considered sufficient for evaluating moderate-to-large process shifts of 2σ or larger, although larger samples (n=10-25) are needed to detect smaller process shifts (Montgomery, 2003). Sampling also allows use of within-group variation estimates, which is the true estimate of process variation.

Both R and S-charts measure within-sample variability. S-charts are considered more efficient than R-charts when n >10, because the standard deviation uses every observation while the Range chart uses only two values (maximum and minimum).

Samples (n=5,10) were taken from the Time-to-Abx. data series to evaluate whether sampling affected control chart sensitivity. Range (R) and Standard Deviation (S) Charts were then constructed for sample sizes of 5 and 10. Initially, a random selection process was used; later, systematic sampling was used, taking the first 5 or 10 cases each month. The latter process would likely select homogeneous cases treated closer in time.

Time-to-Abx. $\overline{X}$, R (n=5) (Graphics 32. & 33.)

The baseline R-chart identified the November and December 2001 subgroups as signals. The final R-chart showed a reduction in variation beginning with the January 2002 subgroup. This shift is associated with any CAP event and is considered a false alarm. The September 2002 subgroup's single spike is considered a July 02 event-related signal; however, it could also be associated with the sampling process. The final $\overline{X}$-chart reflects a process shift at the Jan 02 subgroup that is considered a false alarm. A signal was associated with the September 2002 spike.

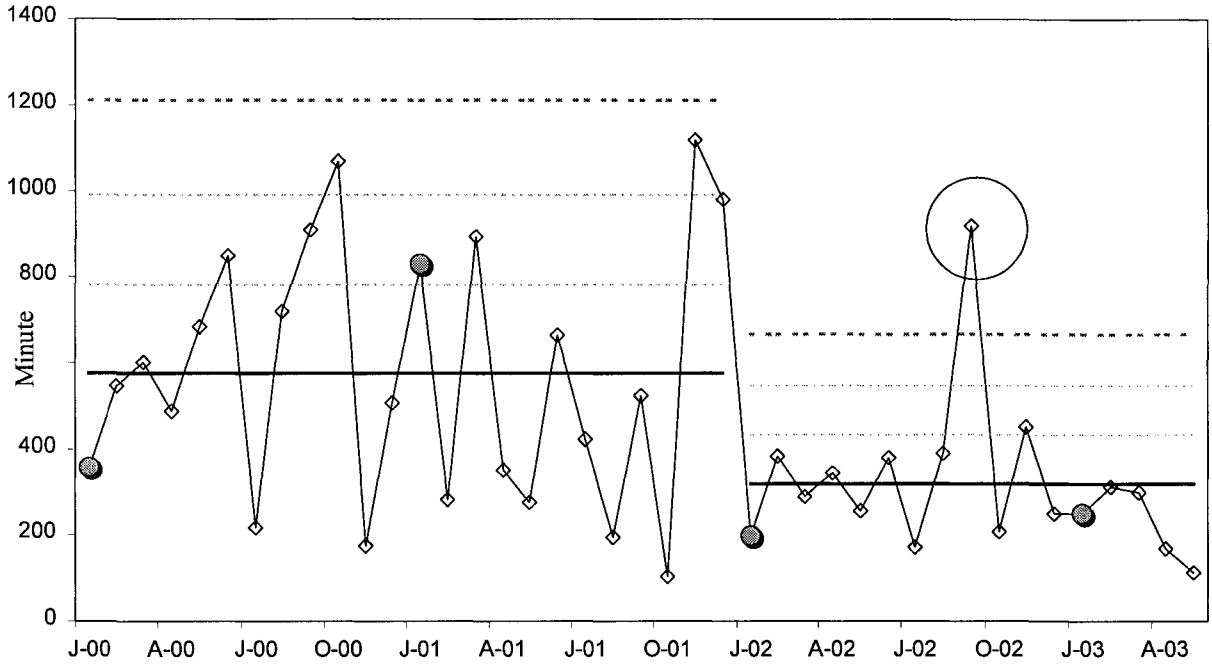
Time-to-Abx. $\overline{X}$, R (n=10) (Graphics 34. & 35.)

Doubling the sample size to 10 increases the number of signals, and false alarms, on the R-chart. Two signals were identified.

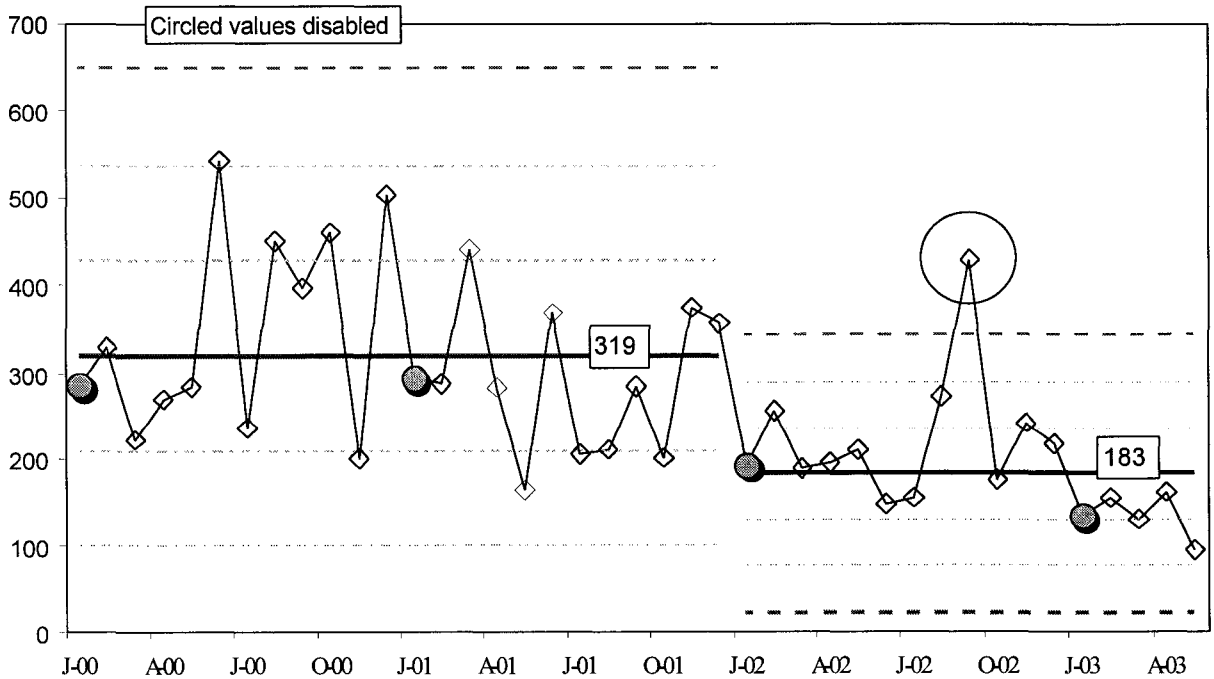
Time-to-Abx. $\overline{X}$, S (n=10) (Graphics 36. & 37.)

Using the std. dev. to estimate process variation resulted in more restricted control limits but not more sensitive charts. The S-chart showed two additional false alarms. The $\overline{X}$-Chart showed a process shift later in the series that is associated with the July 02 event.

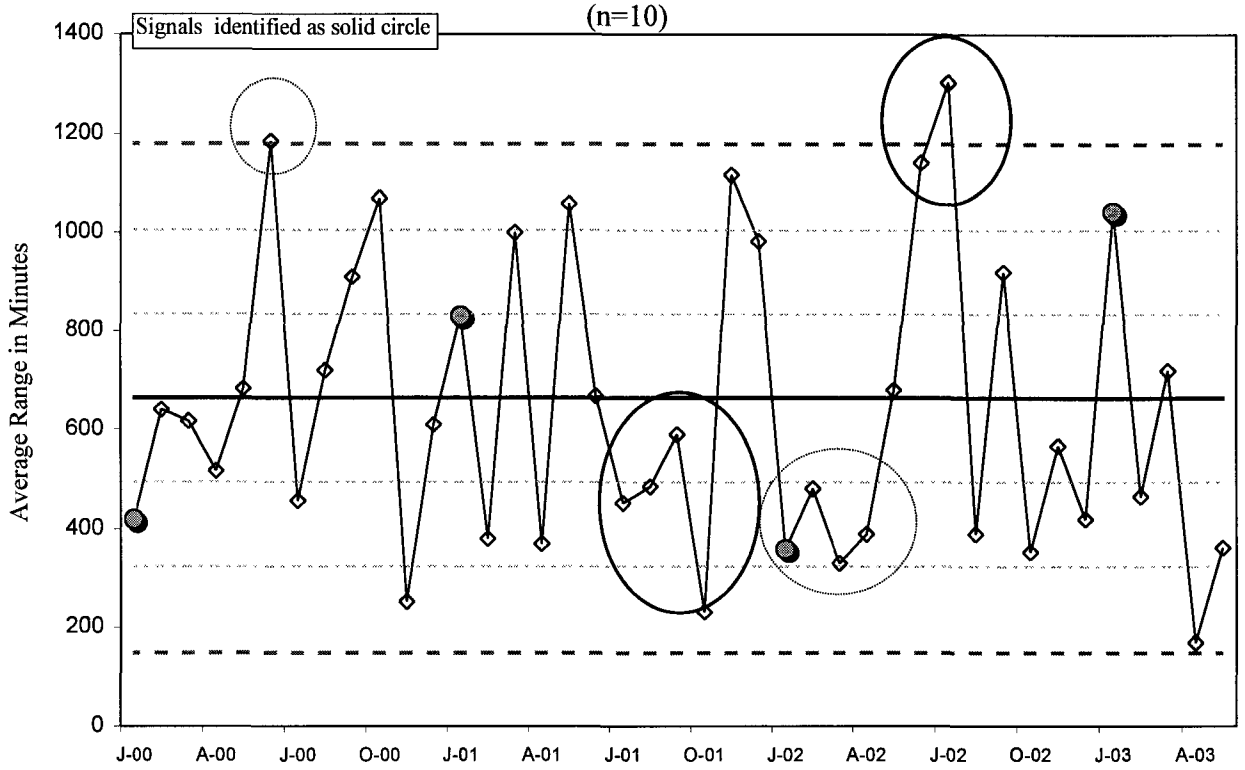
Graphic 32.
Time-to-Antibiotic R-chart
(n=5)



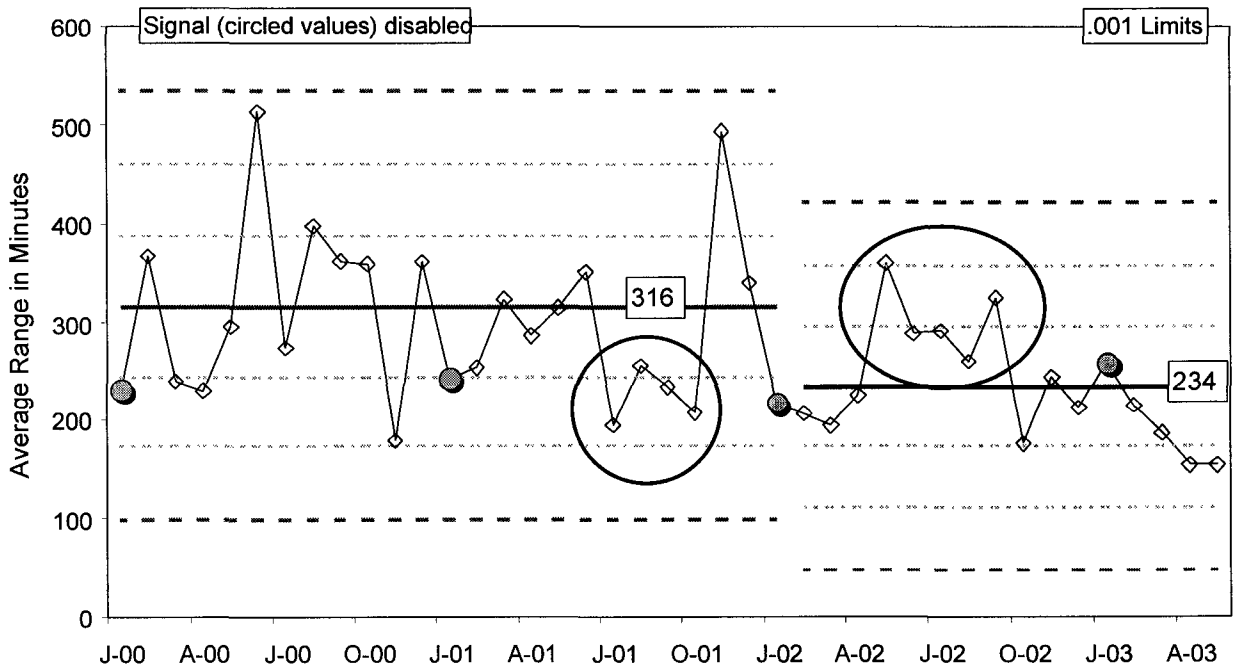
Graphic 33.
Time-to-Abx Xbar-chart (R)
(n=5)



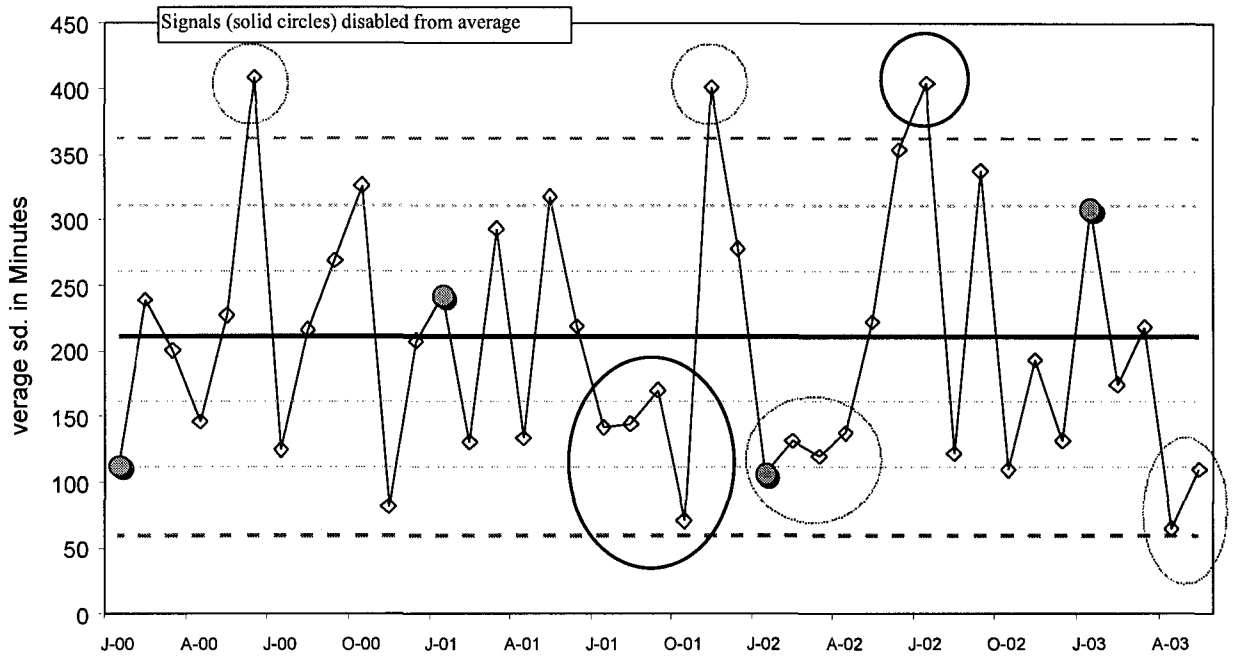
Graphic 34.
Time-to-Antibiotic R-chart
(n=10)



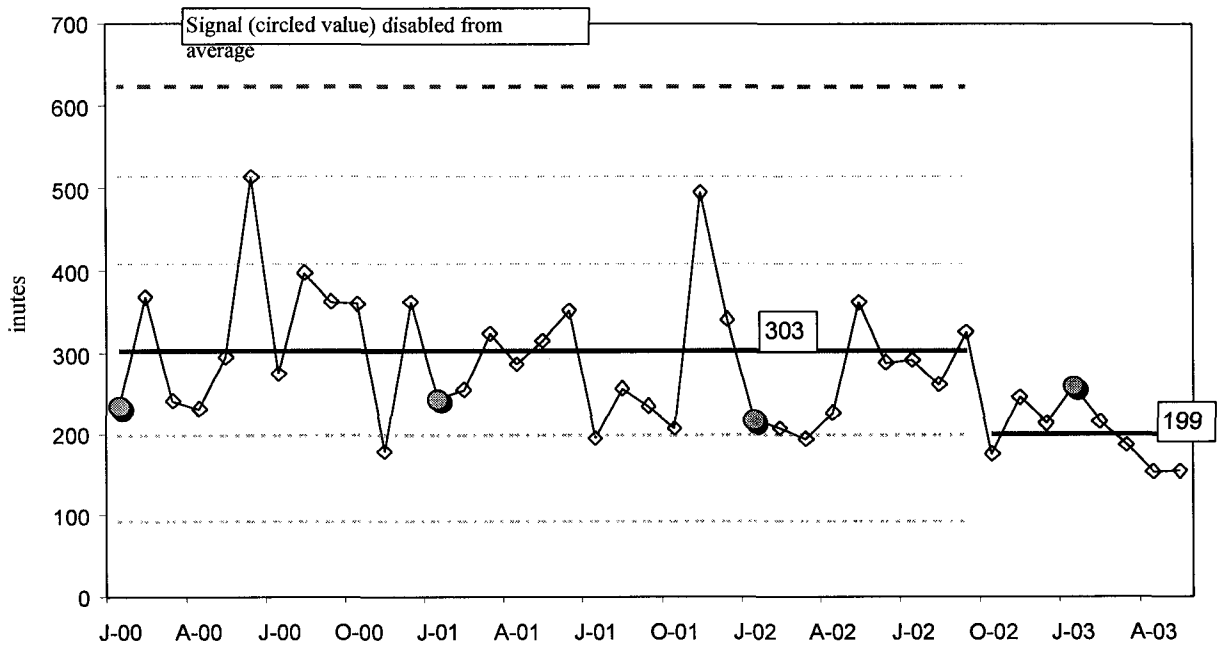
Graphic 35.
Time-to-Antibiotic Xbar (R) Chart
(n=10)



Graphic 36.
Time-to-Antibiotic S-Chart
(n=10)



Graphic 37.
Time-to-Antibiotic Xbar-chart (S)
(n=10)



P-charts	
Time-to-Abx. <4 Hours	(Graphic 38.)
Time-to-Abx. <8 Hours	(Graphic 39.)
Readmission	(Graphic 40.)
Mortality	(Graphic 41.)
Mortality <80 Years	(Graphic 42.)
Mortality >79 Years	(Graphic 43.)
Severity-Adjusted Mortality	(Graphic 44.)
Severity-Adjusted Readmission	(Graphic 45.)
Severity-Adjusted Complications	(Graphic 46.)

P-chart Background

Attribute (discrete) control charts include the P-chart for fraction defective; the c-chart, actual number of defects or non-conforming parts, the u-chart; average number of defectives per unit; and the nP-chart, number of non-conforming items produced. The characteristics that guide selection of these charts is (1) whether data are counts of defects or defectives, and (2) whether there is a constant population at risk. What is the difference between defects and defectives? Both represent discrete variables that do not conform in some way to specification. For defectives, the practitioner has both a numerator and denominator available to count both occurrence and non-occurrence of an event. For example, the percentage of CAP patients who do not receive the Abx. within 4hours would be considered defective. Defect data do not have denominators and are reported as counts (e.g. # patient falls). The constant population-at-risk issue concerns whether there are constant probabilities of an event, or "Areas of Opportunity" (Carey,

1995). Before two counts can be compared, they must have equal Areas of Opportunity and, if they are not equal, or approximately equal, then they must be turned into rates before comparison.

Severity adjustment was performed by Premier Perspective On-line using 3M's APR-DRG grouper. The APR-DRG risk adjustment considers age, procedure, and principal diagnosis. It categorizes patients into four subclasses for disease severity and four additional subclasses for mortality risk. Outliers are excluded from the APR-DRG.

Time-to-Abx. <4 Hours (Graphic 38.)

The Baseline chart identified a (false alarm) signal spanning the period May – Oct 2000. This signal was not associated with the Mar 2000 event, because it reflects a negative trend, contrary to the test model. A second false alarm was identified on the final P-chart at the Jan-Feb 2002 period. The process average was shifted from 55 - 67% at Oct 2002 in association with the HSAG Event. The control limits show overall reductions consistent with the subgroup size increases. Statistical trend is observable in this series.

Time-to-Abx. <8 Hours (Graphic 39.)

The Abx.<8 hours data series is also associated with significant trend. Changes in variation are noted with less variation and with a higher, more consistent average beginning Jan 01. The baseline chart having a process average of 85.4%, identified a false alarm signal spanning June - September 2000. Following visual patterns, the process was shifted beginning January 2001, in association with the Chart Review event,

resulting in a increase in process average from 78.8 - 90.8%. This model results in a single false alarm at the Nov 01 subgroup. The UCL exceeded the 100% value resulting in asymmetric control limits; above average patterns were not evaluated.

Readmission (Graphic 40.)

The June 01 datapoint sits on the UCL, coming close to becoming a signal. It is not assigned signal status, because it is short of the July 01 test period, and it reflects a sharp increase while the test hypothesis is a one-sided value decrease. The June 01 period does reflect change (increase) in series variation.

Mortality (Graphic 41.)

The Baseline Chart shows an initial out-of-control signal for the Jul 00 - Feb 01 period. The final P-chart creates a process shift beginning March 01, reflecting a reduction in CAP mortality from 8.9-2.9%. The shift is recorded as a signal associated with the Jan 2001 Chart Review event.

Mortality <80 Years (Graphic 42.)

Chart control limits reflects the appearance of seasonally higher denominators. The saw-tooth, 14 alternating subgroup signal, begun in May 01, was associated with the Chart Review event (Jan 01).

Mortality >79 Years (Graphic 43.)

There were no signals on the >79 Mortality P-chart. The UCL did seem influenced by seasonal effects, with wider control limits observed during the smaller denominator

summer months. Of interest is the August 02 value that spikes to nearly 25% but reflects a single death, showing the challenge to interpreting rates. The increase is related to the denominator reduction (n=4).

Severity-Adjusted Mortality (Graphic 44.)

No signals identified on the Severity-Adjusted Mortality P-chart, although the April 02 value again sits just below the UCL. The Severity-Adjusted Mortality Chart presents a different visual pattern than the non-adjusted chart. The Mortality Indicator shows a greater influence from the statistical adjustment process compared to Readmission. The non-adjusted Chart is also close to signaling in the April 02 period.

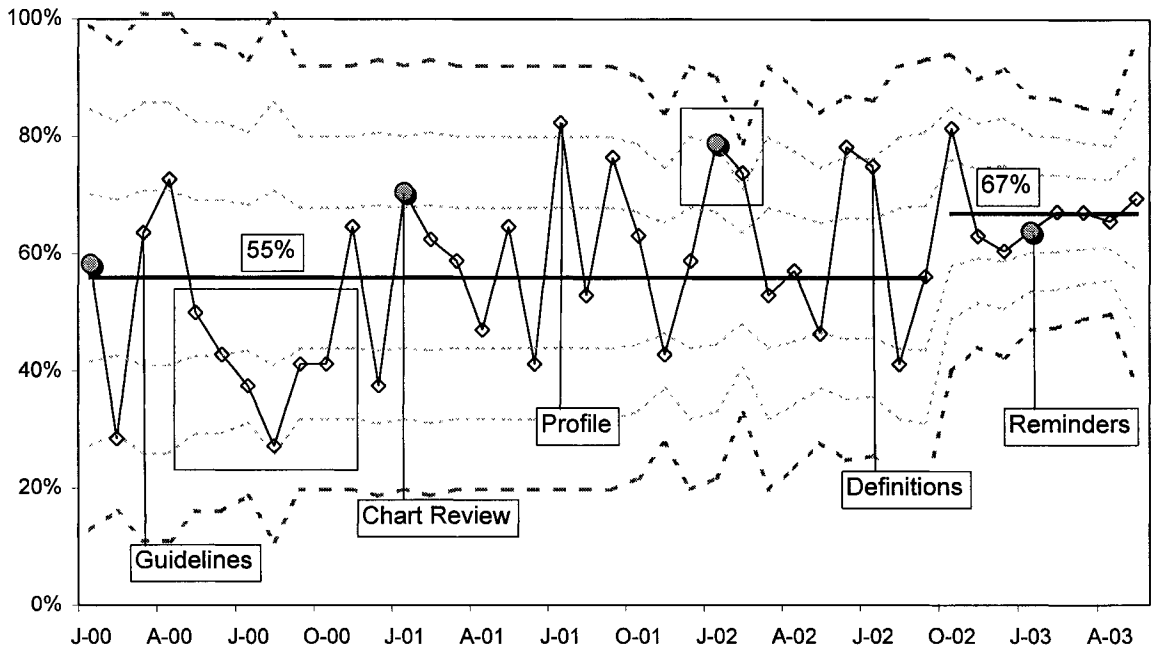
Severity-Adjusted Readmission (Graphic 45.)

The June 01 spike did form a signal on the Severity Adjusted Readmission indicator. It was assigned a false alarm status because it seemed associated with neither a post-event reduction in variation or average value that related to the test hypothesis. The Severity-Adjusted P-chart appears somewhat more volatile than the non-adjusted chart. The process average was very consistent with the non-adjusted Readmission P-chart.

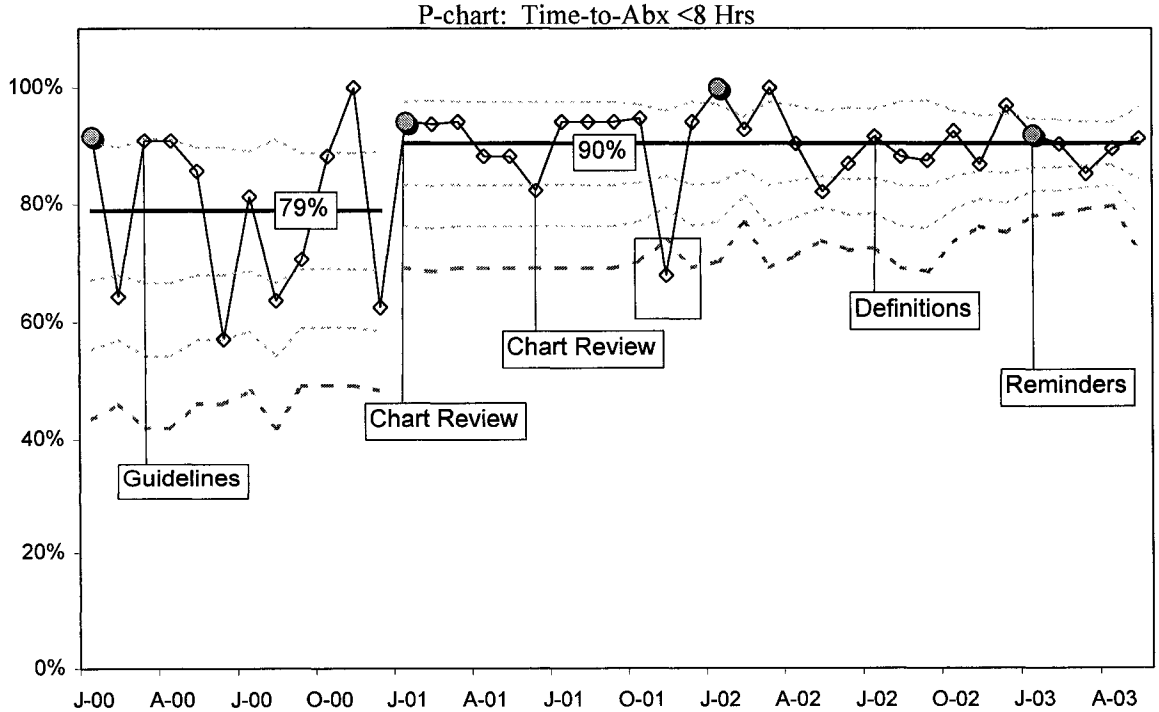
Severity-Adjusted Complication Rate (Graphic 46.)

The June 2000 signal is considered a false alarm because it is inconsistent with the test hypothesis. There are numerous 0-month values (@28%) in the full data series, with the majority in the beginning of the chart.

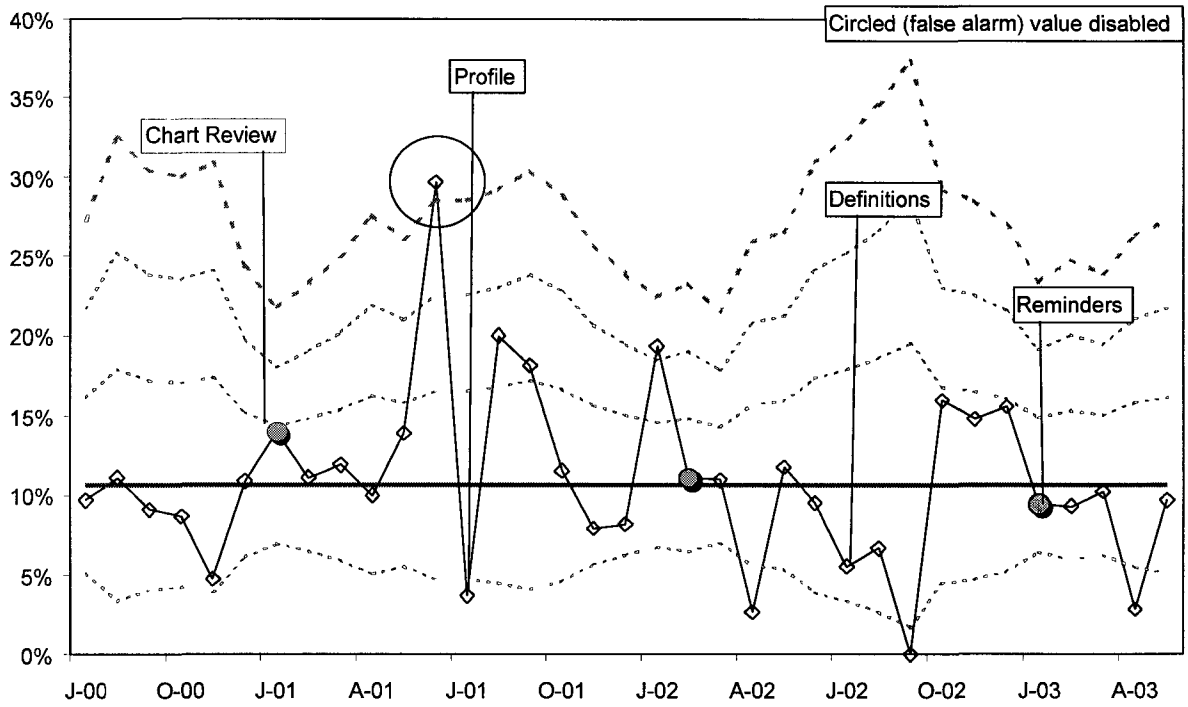
Graphic 38.
Abx <4Hours P-chart



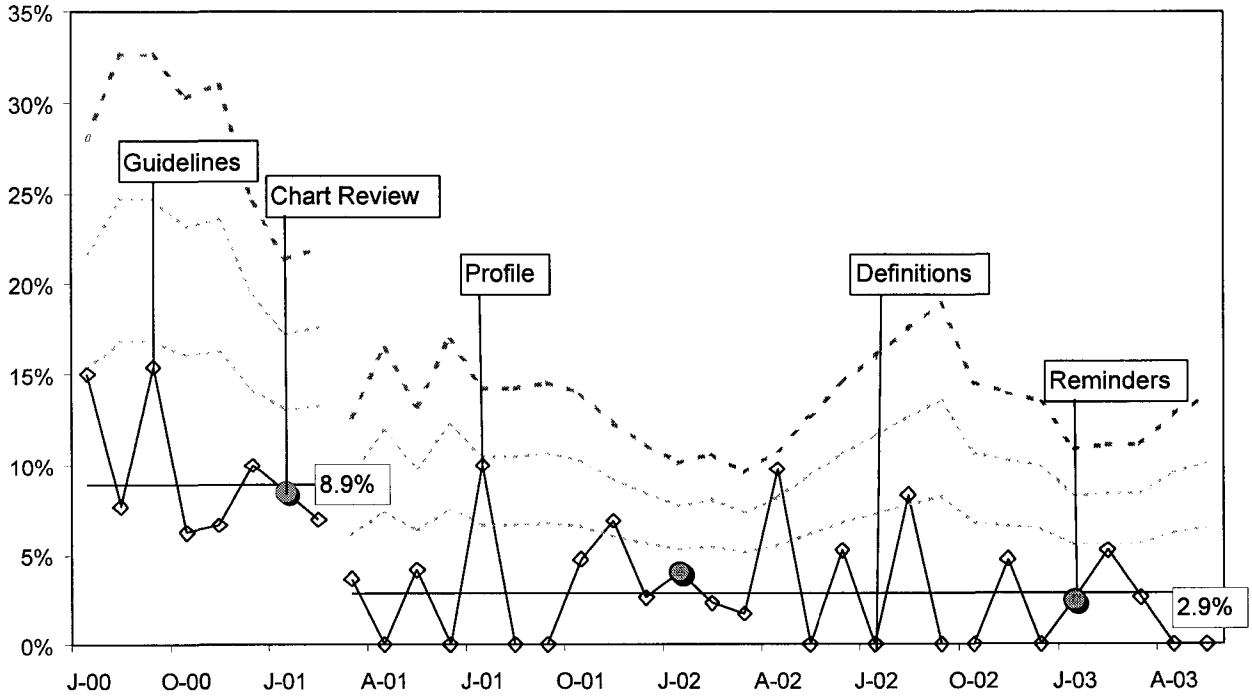
Graphic 39.
P-chart: Time-to-Abx <8 Hrs



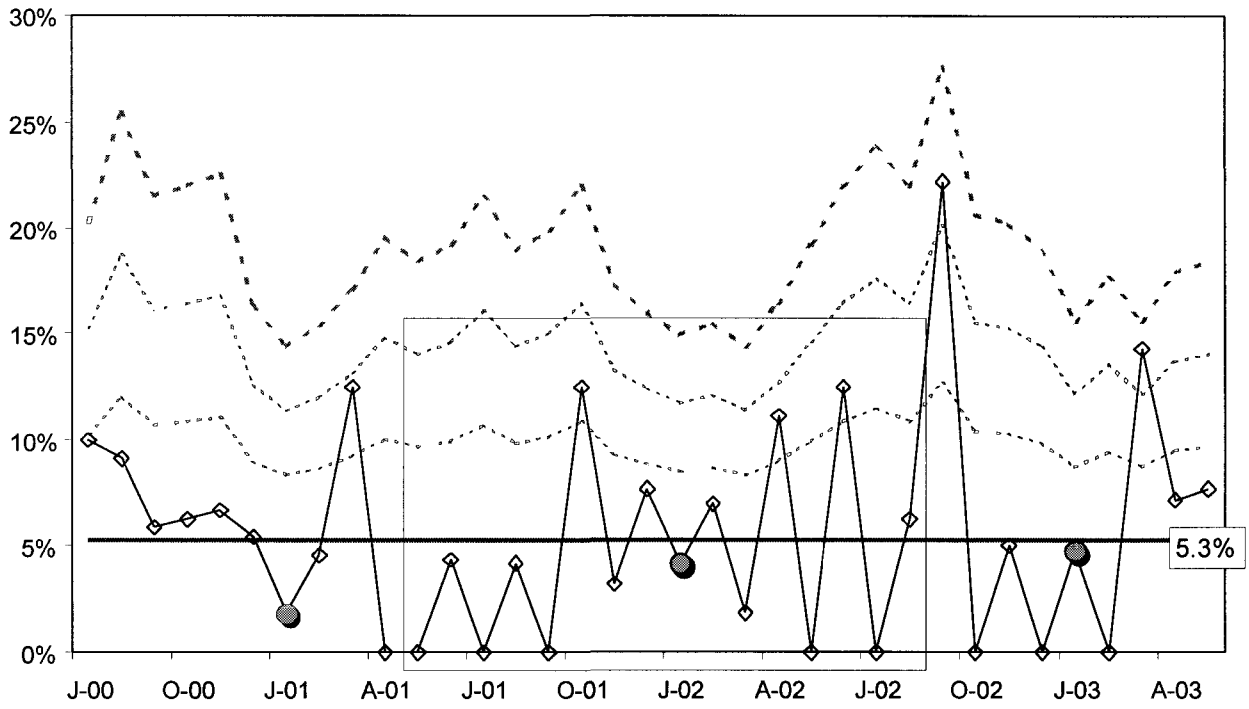
Graphic 40.
CAP-Related Readmission P-chart



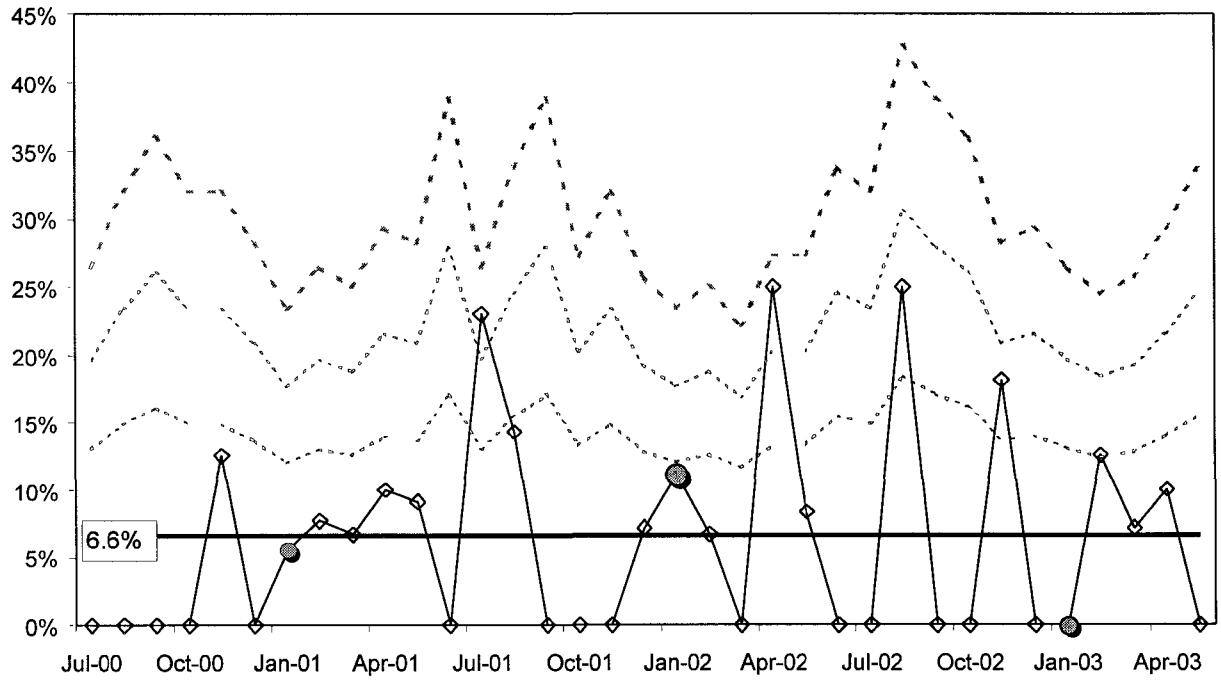
Graphic 41.
CAP-Related Mortality P-chart



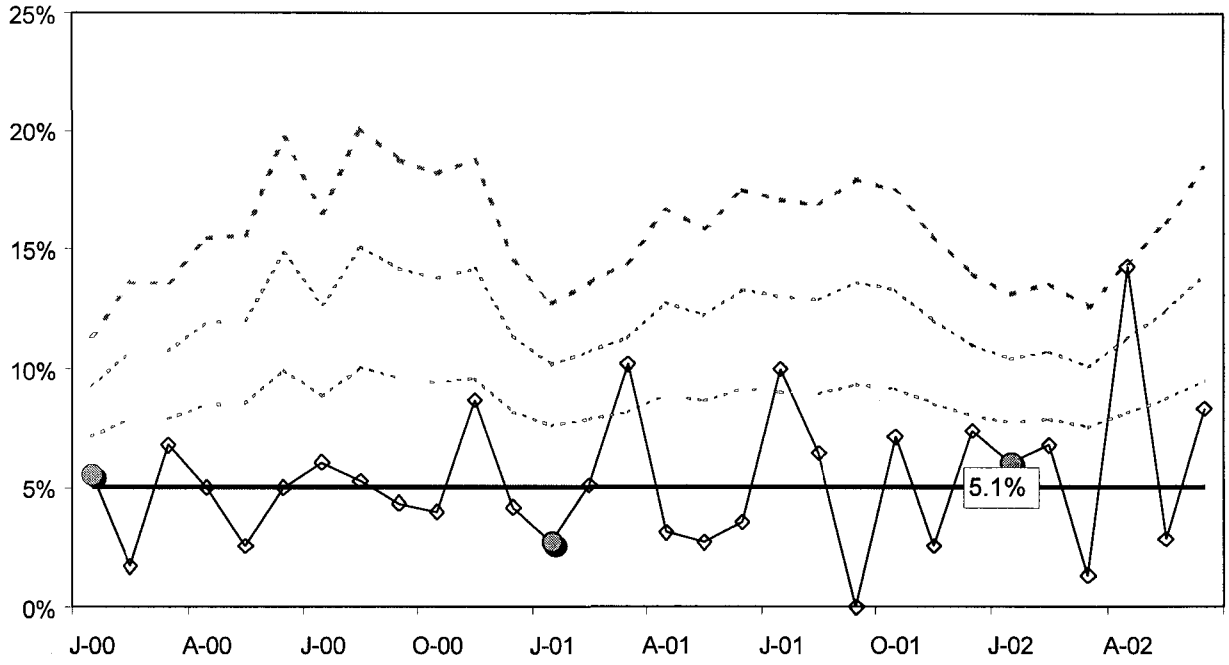
Graphic 42.
<80 Years CAP-Related Mortality P-chart



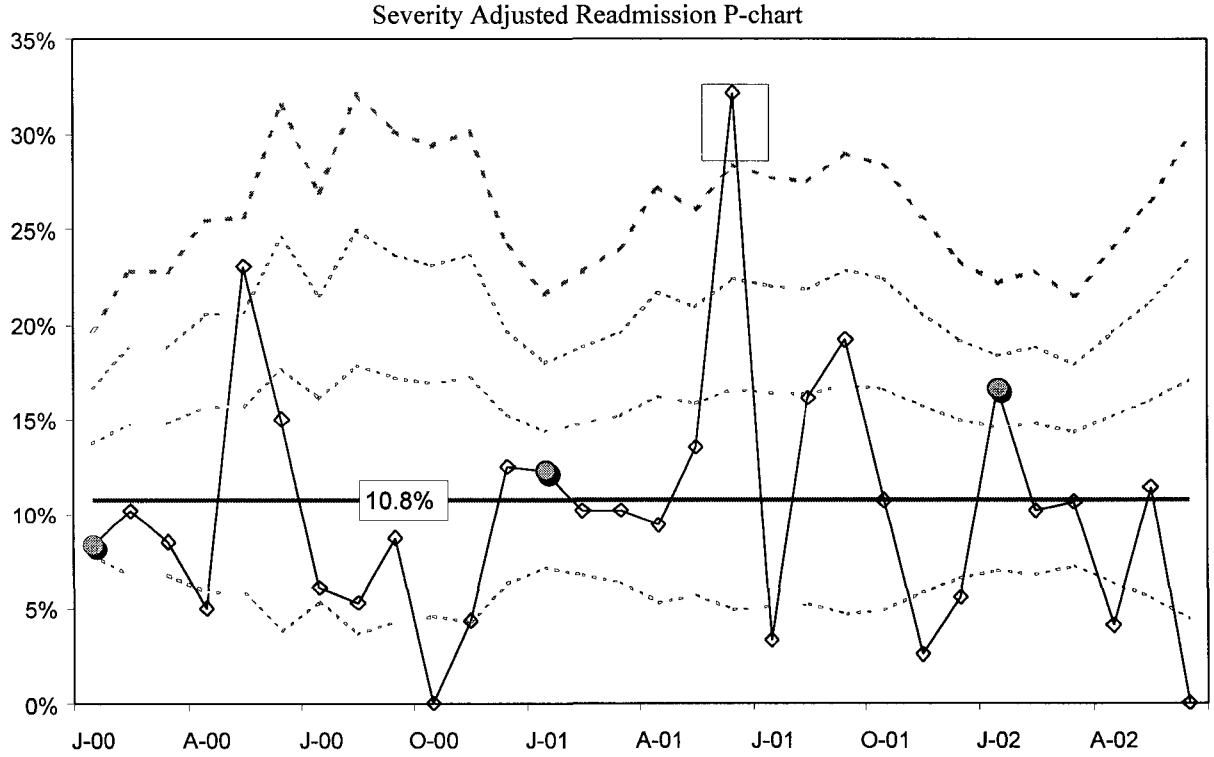
Graphic 43.
>79 Years CAP Mortality P-chart



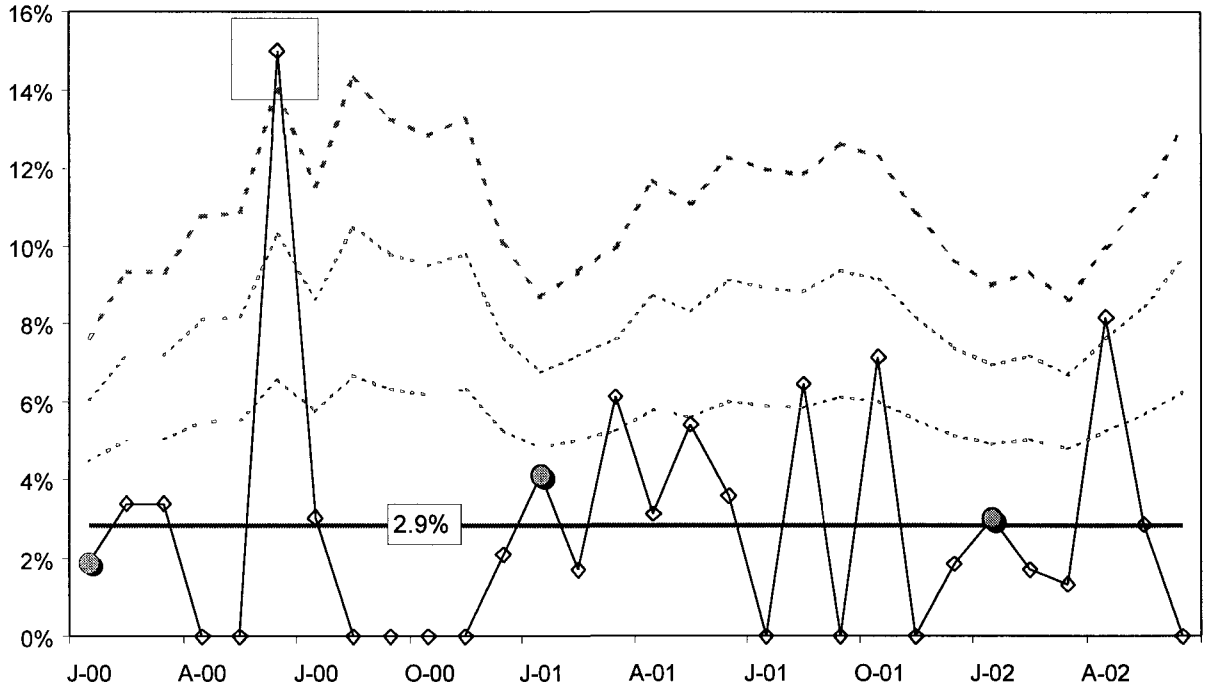
Graphic 44.
Severity Adjusted Mortality P-chart



Graphic 45.
Severity Adjusted Readmission P-chart



Graphic 46.
Severity Adjusted Complications Rate P-chart



Exponentially Weighted Moving Average (EWMA) Charts	
Time-to-Abx.	(Graphic 47.)
ALOS	(Graphic 48.)

EWMA Background

EWMA are special purpose charts used to forecast future values and monitor control. Exponential weighting emphasizes current observations by assigning greater weight to recent data; the weight decreases exponentially across the series. The EWMA is defined as: $z_i = \lceil x_i + (1- \lambda) z_{i-1}$ where:

- $0 < \lambda < 1$ is a constant
- λ is the weight attached to the current value
- i = sample number

- Starting value (first sample at i = 1) is the process target so that $z_0 = \mu_0$
- z_I, the exponentially weighted moving average, is the weighted average that will sum to unity and decrease geometrically with the age of the sample mean

The moving average (M_t) is an average of *kn* observations when $t \geq k$.[(t=time) and k= (span)]. The average age of the EWMA value is inverse the value of λ; with $\lambda = 0.2$, the average age of the values in the EWMA will be five time periods. The choice of a large value for λ will result in an EWMA with less lag; however, using large values will not provide a great deal of information unique from the X-chart. The EWMA is considered more sensitive to small changes and insensitive to the normality assumption (Crowder, 1989). Because the EWMA is a moving average, it lags behind the process itself and, because plotted statistics are dependent, it is difficult to pick up anomalous patterns other than violation of the control limits. For this reason, the EWMA is best used in combination with Shewhart charts.

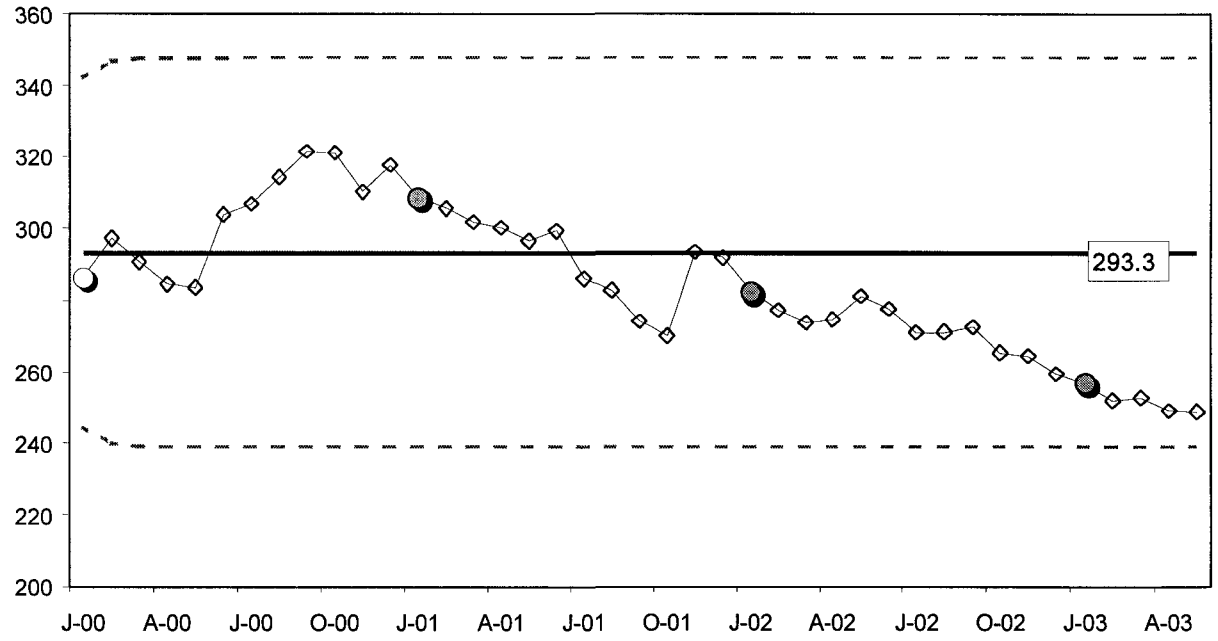
Time-to-Abx. (Graphic 47.)

λ weights (.1-4) were used to assess the sensitivity of EWMA charts. No signals were identified. The EWMA does show a general decline in Time-to-Abx., beginning with the Oct. 00 subgroup. There is brief change in direction observed at the Nov. 01 subgroup.

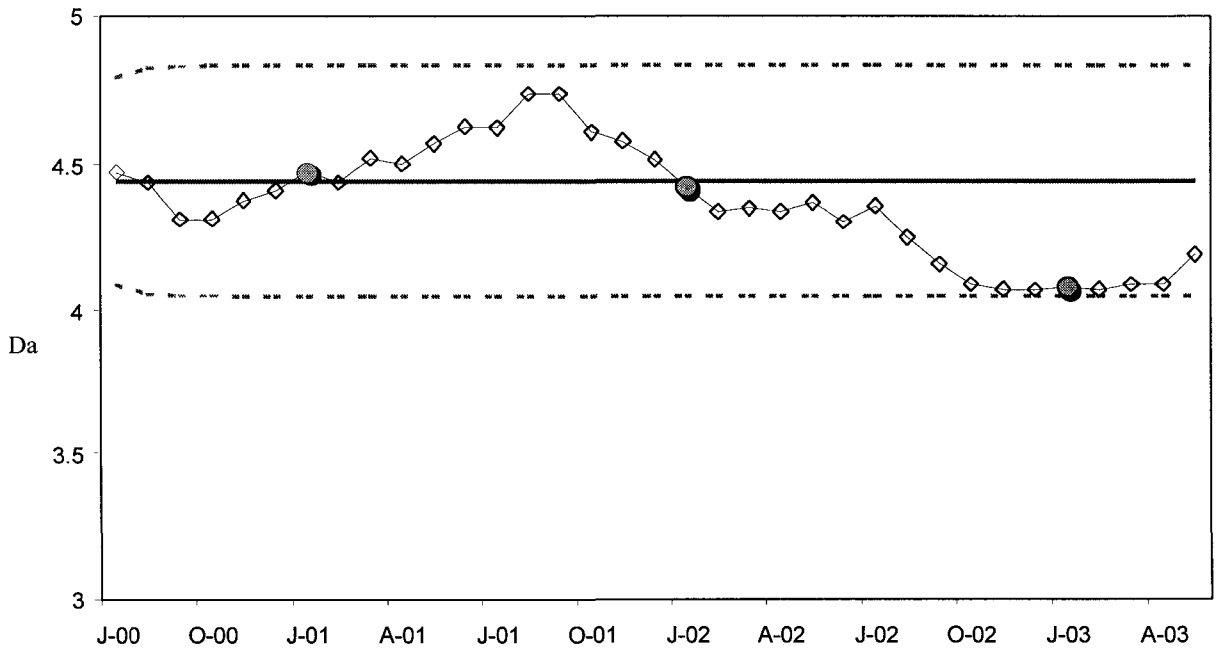
ALOS (Graphic 48.)

No signals were observed using the .1λ, 2.7 *L* estimates. The subgroups increase and then show a decline, beginning Sept 01; they approach the LCL in Oct 02.

Graphic 47.
Time-to-Antibiotic EWMA-chart
(Lambda .1, Length 2.7)



Graphic 48.
ALOS EWMA-chart
(Lambda .1, Length 2.7)



Cumulative Sum (CUSUM) Charts

Time-to-Abx. (Graphics 49. & 50.)

ALOS (Graphics 51. & 52.)

CUSUM Background

CUSUM charts plot the cumulative sum of the deviations of the sample values from the target (reference) value (Lim, 2001). Cumulative sum has a mean that is linear in function; if $\mu = k$, the slope is zero; if $\mu > k$, the slope is positive; and if $\mu < k$, the slope is negative. The charting process involves (1) Selecting target value, (2) Obtaining measure of dispersion, (3) Compute deviations from target, (4) Compute cumulative sums (sum of all deviations from target), and (5) Specifying the Reference value *(K)* (size of shift to determine) and the Upper / Lower limits *(H)*.

The CUSUM detects changes in process mean by accumulating the sum of deviations of the observed sample means $X_1 \ldots . X_t$ from a reference value. It is supposed to have better sensitivity to detect small mean shifts (i.e. 1.5-2 σ) than the traditional Shewhart (using the 1-of-1 rule). However, traditional charts with supplementary rules have the advantage of permitting review of anomalous patterns other than just monotone parameter shifts.

The cumulative sum function is: $S_t = \sum_{t-1}^{t} (\overline{X} - k)$

where k is the reference value and T represents time.

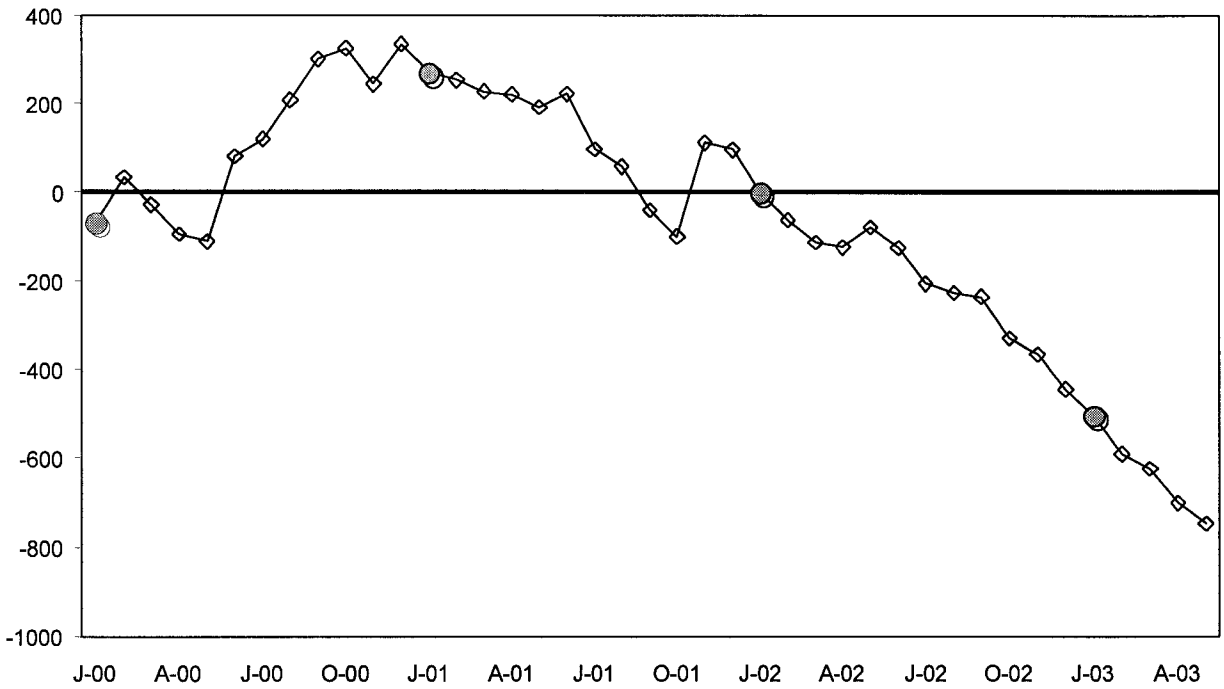
Time-to-Abx. (Graphics 49. & 50.)

The CUSUM Status Chart identified no signals. The target value (293 minutes) was established at the baseline average (m=25). Most cumulative deviations were negative, and a downward trend was observed, beginning at the Dec 00 time period, with a single increase observed at the Oct 01 datapoint. None of the values observed on the CUSUM Status Chart exceeded the Upper/Lower CUSUM that was set at 5σ.

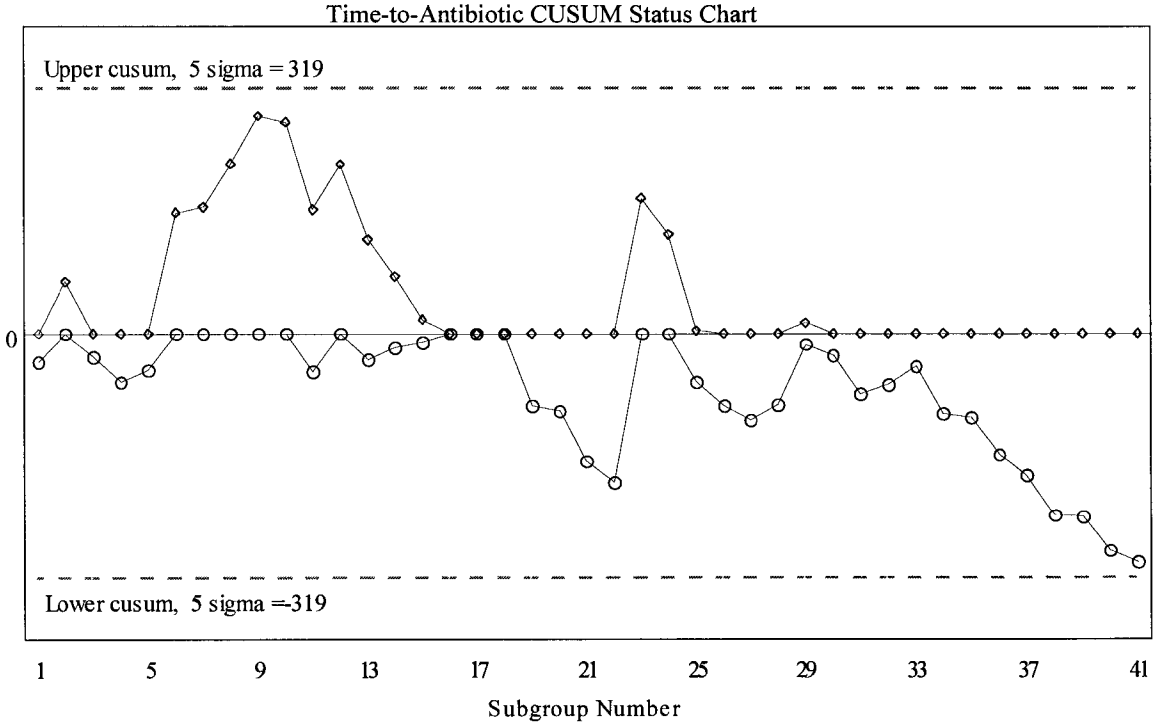
ALOS (Graphics 51. & 52.)

The CUSUM Chart identified no signals. The target value (4.4 days) was established at the baseline average (m=25). Most cumulative deviations were negative, and a downward trend was observed beginning the Sep 01 time period. None of the values observed on the CUSUM Status Chart exceeded the Upper/Lower CUSUM, which was set at 5σ.

Graphic 49.
Time-to-Antibiotic CUSUM-chart

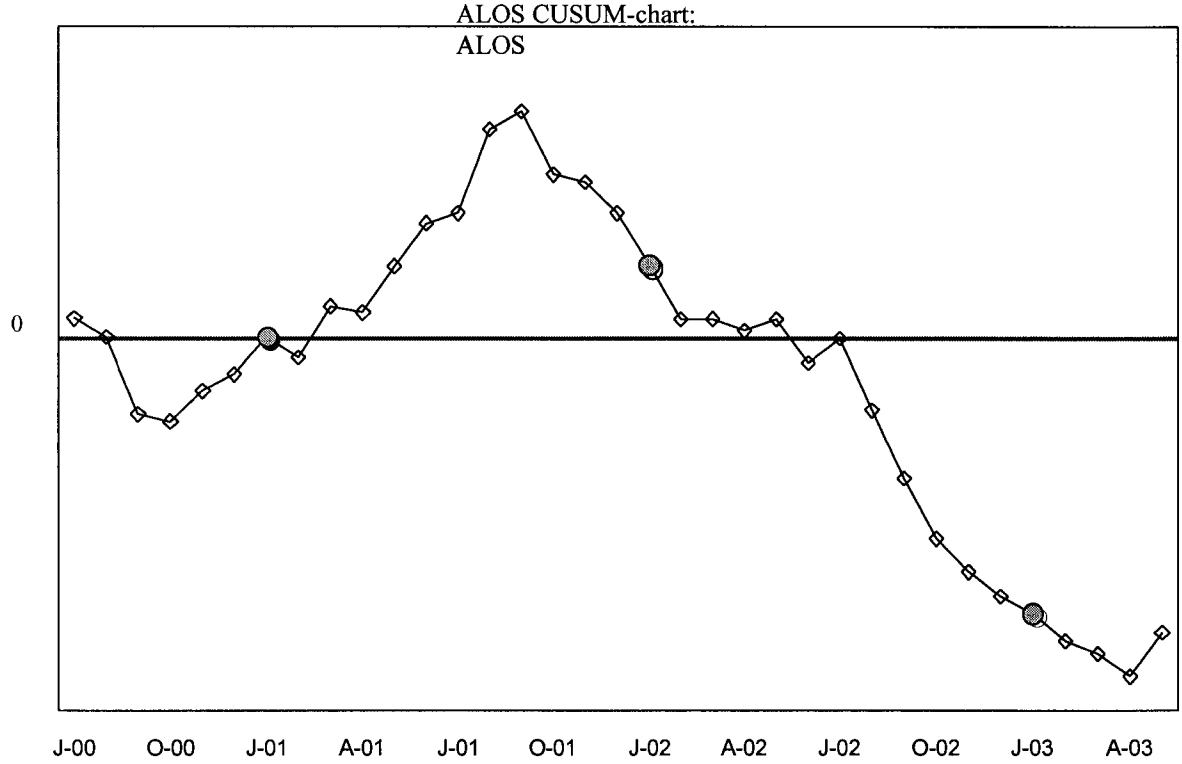


Graphic 50.
Time-to-Antibiotic CUSUM Status Chart



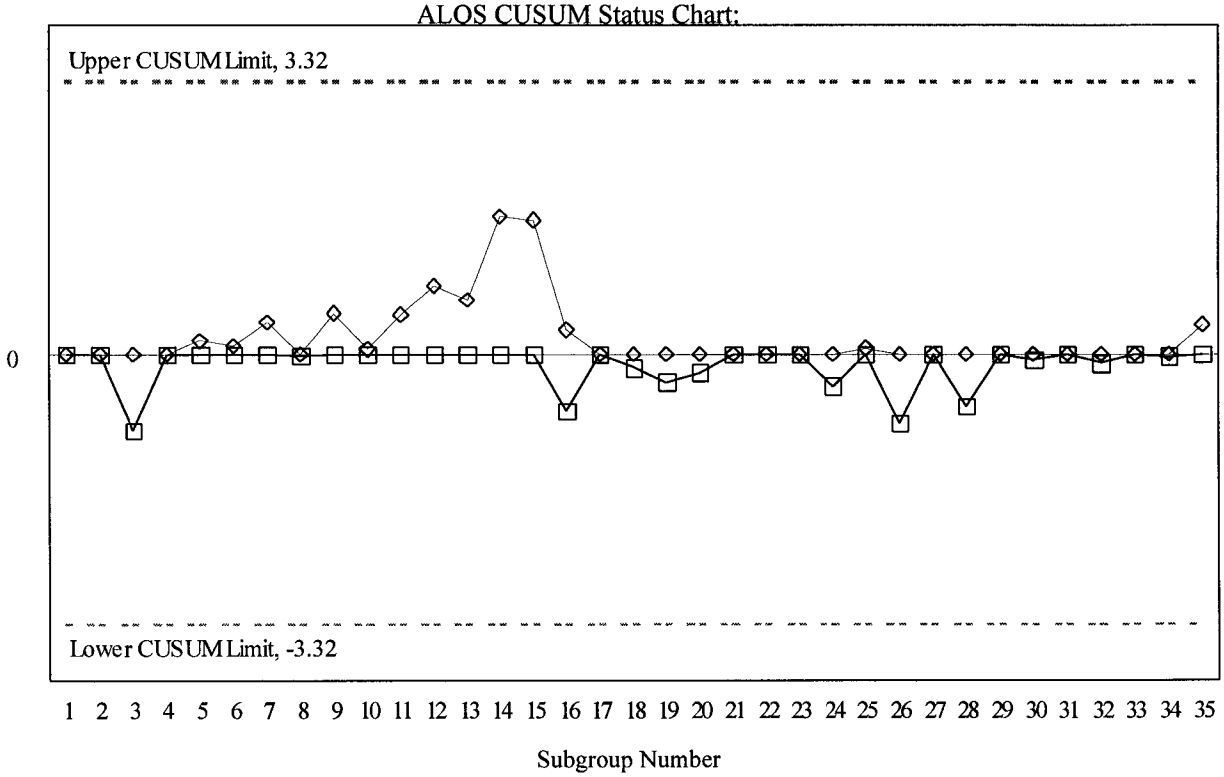
Graphic 51.

ALOS CUSUM-chart:



Graphic 52.

ALOS CUSUM Status Chart:



Sensitivity Comparisons

Table 34. Control Chart Test Results (Codes: 0=No, 1=Yes)

Chart	*Indicator*		*Mar-00*	*01-Jan*	*01-Jul*	*02-Jul*	*03-Jan*	*Totals*
XmR	AbxMin	mR-chart	0	0	0	0	0	0
		X-chart	0	0	0	1	0	1
	Transformed AbxMin	mR-chart	0	0	0	0	0	0
		X-chart	0	0	0	1	0	1
	ALOS	mR-chart	-	0	0	0	0	0
		X-chart	-	0	1	1	0	2
	Severity-Adjusted ALOS	mR-chart	0	0	0	-	-	0
		X-chart	0	0	1	-	-	1
		Sub-total	0	0	2	3	0	5
Xbar	AbxMin (n=5)	R-chart	0	0	0	1	0	1
		Xbar-chart	0	0	0	1	0	1
	AbxMin (n=10)	R-chart	0	0	1	1	0	2
		Xbar-chart	0	0	1	1	0	2
	AbxMin (n=10)	S-chart	0	0	1	1	0	2
		Xbar-chart	0	0	1	1	0	2
		Sub-total	0	0	4	6	0	10
P-chart	AbxMin <4		0	0	0	1	0	1
	AbxMin <8		0	1	0	0	0	1
	DRG-coded Mortality		-	1	0	0	0	1
	<80 Mortality		-	0	0	0	0	0
	>79 Mortality		-	0	0	0	0	0
	Readmission		-	0	0	0	0	0
	Severity-Adj. Readmission		0	0	0	-	-	0
	Severity-Adjusted Mortality		0	0	0	-	-	0
	Severity-Adj. Complications		0	0	0	-	-	0
		Sub-total	0	2	0	1	0	3
		TOTALS	0	2	6	10	0	18

The validity of CAP medical quality indicators concerns their ability to do what they are supposed to do, that is, correctly identify important CAP-related events. Sensitivity analysis of the control-charting analytic process was completed to review chart behavior and evaluate the degree to which important events tested positive and whether non-events tested negative. An exploratory approach was adopted for the testing process to allow multiple comparisons between chart types. Tables 34. and 35. identify whether signals were identified in association with the CAP-specified events.

Table 35. Specialty Chart Test Results

Chart	*Indicator*	*Mar-00*	*01-Jan*	*01-Jul*	*02-Jul*	*03-Jan*
EWMA						
	AbxMin					
	Lambda .1, Length 2.7	0	0	0	0	0
	ALOS					
	Lambda .1, Length 2.7	0	0	0	0	0
CUSUM						
	AbxMin					
	Reference Value (K) = 31.99	0	0	0	0	0
	Decision Interval (H) = 5sigma					
	ALOS					
	Reference Value (K) = .33	0	0	0	0	0
	Decision Interval (H) = 5sigma					

Sensitivity, considered the probability of a chart signal if a CAP event is present, was computed as

$$\frac{\text{\# months when assignable cause is evident and event did occur (a)}}{\text{(a) + \# months where no assignable cause is present but an event did occur}}$$

Specificity, considered the probability of no assignable cause if cause is truly absent, was computed as

$$\frac{\text{\# months where no assignable cause is present and no event occurs (d)}}{\text{(d) + \# months where no assignable cause is present but an event did occur}}$$

The false alarm probability was computed as

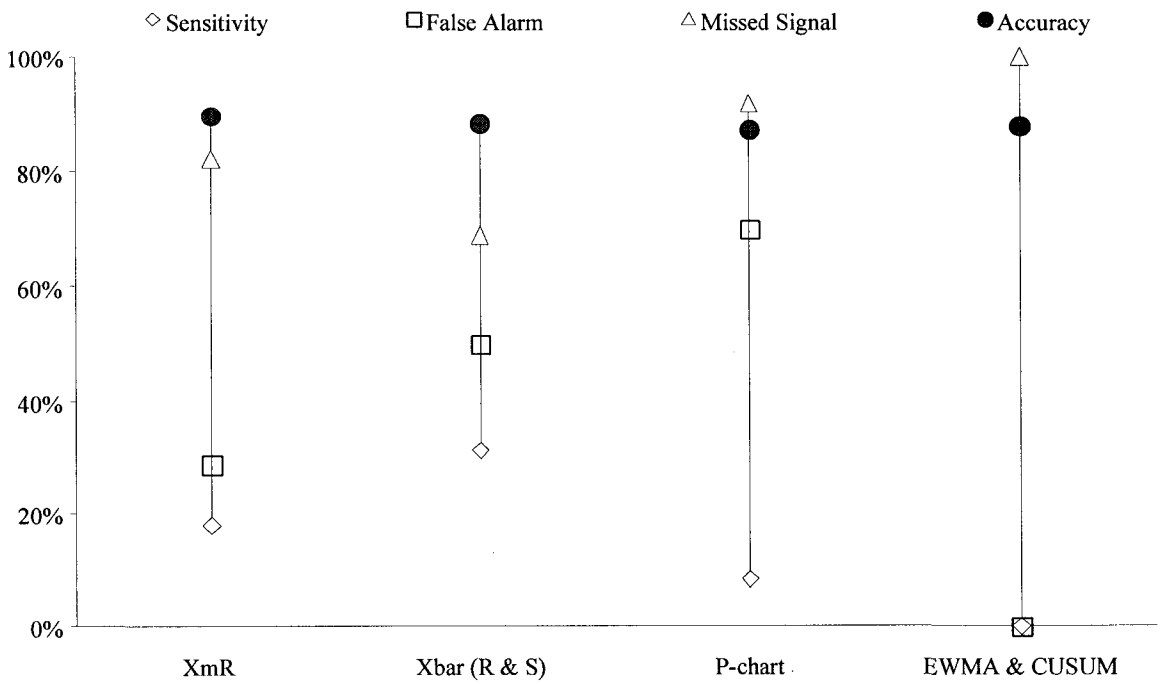
$$1 - \frac{\text{(\# months when assignable cause is evident, and event did occur (a)}}{\text{\# months where signals occurred without an associated event}}$$

Accuracy was computed as the probability that

$$\frac{\text{a signal was correctly associated with an event + non-signals were correctly associated with noise}}{\text{all observations}}$$

Graphic 53. summarizes observed chart sensitivity. These estimates are considered very unreliable with potentially large standard errors.

Graphic 53.
Control Chart Sensitivity to CAP-Events



CHAPTER 5 - DISCUSSION

The control chart is a valuable tool for analyzing medical variation. It is an economical method that combines graphic and statistical analysis of medical quality data reported over time. It is clearly superior to (1) not analyzing available information; (2) single period summaries; (3) multiple, but limited, period comparisons; and (4) the practice of overwhelming practitioners with complex statistics. Huth (1985) offered the following cost-benefit analysis for medical information:

$$Value = \frac{Utility}{Cost}.$$

Under sufficient conditions, the control chart produces good analytic value due to low operational cost and the potential for extracting information about medical operations. Costs rise dramatically when information has poor validity or, alternately, when the purpose of charting is shifted to confirmatory applications that require staff time for intensive "drill-down." By itself, the control chart can distinguish excessive and non-random variation, however it cannot identify the cause of a signal or determine its precise time of occurrence.

Medical care represents a more complex causal field when compared to the manufacturing environment that gave birth to SPC. Medical services are complex adaptive systems with more components of variation. Many medical organizations have unique operating conditions that challenge data collection and reporting. Medical data can also be "messy." These conditions challenge traditional applications and require that the control chart be calibrated to the medical process.

How is the control chart model related to medical quality?

There are logical and statistical considerations that guide response to this question. Results from this Study support an empirical approach to determining acceptable design and optimal operating conditions.

The CAP indicators reviewed in this Study were marginally stable, frequently showing trend, seasonality, skew, changes to definition, sampling variation and autocorrelation. The descriptive analysis of CAP indicators uncovered a somewhat complex data structure. The most important data quality characteristic concerned stability, or existing control. Trend and non-stationary series were observed, although the effects of trend appeared associated with the increase in sample size, beginning in Jan 03, creating less variation and larger, significant slopes. When the time periods with larger samples were not included, or as re-sampling was performed, trend was non-significant. Still, data stability remained an important consideration.

Data stability is influenced by subgroup frequency with more frequently reported subgroups, those by week, being less stable but, more representative, of current events than less frequently reported subgroups, those by month. This issue represents a fundamental tension in control chart design, that of stability versus representativeness. Data stability addresses value consistency during a minimum specified time period. However, stability may be undesirable in the medical environment, given the need for change, and it may be improbable, given the pace of change. Data values collected from

a time period 18 months prior to an event may not represent current periods. To understand rapidly changing conditions, short-run charts are likely preferable, even though they will be less reliable. Data from the local "neighborhood" will be more representative and improve the chart's ability to evaluate proximal quality interventions. Data validity is influenced by rational subgrouping, the ability to segregate information according to medical quality causal streams. Both process and outcome indicators were evaluated in this Study, and, with the exception of a significant association between AbxMin and mortality, the outcome indicators did not appear to have tight affiliation with the process indicators. Caution seems justified in the application of control charts to global outcome measures such as Readmission, Complication, Mortality, and ALOS.

There are several statistical considerations associated with stability. The average value of a stable series creates a baseline against which individual values are determined statistically improbable. The series mean (centerline) is a critical parameter estimate, and on it rests chart sensitivity and validity. Because of the many environmental influences effecting medical care, control charts will likely need frequent adjustment, and moving average charts are being used to continually update process parameters. However, this represents a fundamental challenge to the logic underlying control charts and could be construed as tampering and change the basic tenant of charting: to expose change between subgroups while holding constant change within subgroups. In dynamic environments, the moving average may be more representative of the current process; however, they are a source of additional variation that challenges the logic of stability.

Another issue concerns subgroup size. Changes to subgroup size during the latter observation periods created unstable series and represented an important confounder during the initial analysis. Chart adjustments are needed with subgroup size changes.

Pneumonia is generally recognized as a seasonal medical condition, but there are other medical and organizational conditions that will also show cyclic patterns. Seasonal cycles, created both by the number of events (numerators) and the number of opportunities (denominators), is a source of instability. Creating rates reduces the effect of seasonal pattern but introduces an additional analytic challenge by combining two distinct pieces of information that represent different causal streams. Seasonal patterns affect the computation of control limits, as sample sizes vary.

The measurement of medical quality responds well to its being defined in terms of reduced variation. But there are important structural conditions that require attention. *First,* quality measurement may be more limited by information technology, by the costs associated with abstracting information, and by confidentiality and legal restrictions associated with medical information. *Second,* there are more stakeholders involved in the subjective determination of medical quality. The need to assess quality from multiple stakeholder perspectives will require a greater range of measures and information strategies directed toward consumer groups. *Third,* the medical care workforce is among the best educated in any professional field. The array of practitioners who enable medical care is the essential quality asset contained in the medical care product. Given their

diversity, training, and mobility, it is not surprising that practice variation exists. Also, because of advanced education and high status, medical practitioners may be more reactive to observation. Given this prominent role in determining quality, medical practitioners need to be engaged as full partners for the CQI process to be effective.

1. How well do the XmR, Xbar, EWMA, and P-chart charts monitor changes in hospital practice?

The CAP indicator sensitivity analysis needs to be interpreted within the context that few events were evaluated and that effect sizes associated with events were small.

- Control charts showed limited sensitivity, and for the three principal chart-types, there were more false alarms than event-associated signals.
- As expected, there is support for the conclusion that larger sample sizes increase chart sensitivity, as does the application of Jaehn's Rules to identify chart signals.
- Contrary to expectation, the application of hybrid charting techniques (EWMA and CUSUM) did not increase chart sensitivity.
- The July 02 event that signaled most frequently, adopting JACHO definitions, caused an increase in sample size after Jan 03 subgroups and resulted in a decrease in series variation and average. The effect size associated with the July 02 event is considered small (d.=.11). Two of the events were surrounded by an insufficient number of subgroups to be fully robust for the sensitivity analysis.
- The mR chart did not identify any special cause signals.

- P-charts applied to outcome variables appeared very insensitive. The X-chart applied to the ALOS outcome indicator did signal twice in association with CAP events.

2. Can charts be made more or less sensitive by changing sampling frequency?

Week-reported subgroups were more representative of current events, but less stable. The challenge observed in the CAP sensitivity analysis was that the number of patients (inputs) varied for the shorter time periods (e.g. range 0-20 patients for week reports), resulting in more extreme variation. The process average parameter was not affected greatly by changes in subgroup frequency; however, variation was increased, and the actual outliers changed according to periodicity. Conversely, trend was more likely for the larger aggregations (i.e. quarterly) versus week or daily subgroups.

- Sampling frequency had a big impact on the sensitivity analysis results, with week-reported data being more sensitive than month reports for the AbxMin indicator.
- Increasing subgroup frequency increased the number of more complex significant autocorrelation patterns.

3. How do data characteristics (i.e. distribution and independence) influence chart sensitivity?

CAP indicators showed important deviations from the assumptions of normally distributed data, supporting the conclusion that distribution character should not be assumed for medical data series. Rather, they should be empirically checked. Normally distributed residuals demonstrate series control, which was not evident for some indicators. Subgroup series distributions were mostly non-normal, reflecting their time-

to-event character and for the mortality outcomes, rare events. For mortality, many months had "0" cases, creating negative skew. The assumption that subgroups are normally distributed is not supported by the time-related character (i.e. time-to-event, average length of stay) of many indicators which showed natural positive skew.

- Skew affects the computation of control limits and leads to more false alarms.
- The process of data transformation had mixed effects on the charting process. While log-linear transformation of the AbxMin variable did eliminate the single false alarm observed on the non-transformed X-chart, it also eliminated the observed outliers in most monthly reported datasets.
- Significant autocorrelation of a large magnitude (.8) was identified in two of the process indicators (Influenza and Pneumonia Documented). However, neither indicator was charted, due to strong significant trend.
- Readmission and Mortality denominators had strong autocorrelation (.65). However, the autocorrelation disappeared after creating the rates.
- High autocorrelation (.8) leads to poorly specified control limits and increases false alarms. The magnitude of autocorrelation in the CAP indicators was not of concern.

4. Do risk adjustment or stratified analysis applied to the XmR and P-chart improve the sensitivity of conclusions?

APR-DRG severity adjustment was applied to a limited data series of CAP outcome indicators (ALOS, Mortality, and Readmission) before control charting. The adjustment process had important impact on data series values but did not increase chart sensitivity.

The CAP Mortality indicator was stratified into two age groups (>79 and <80 years); P-charts were constructed for both age groups.

- Contrary to expectation, neither stratified analysis or risk adjustment affected chart sensitivity.

Study Limitations

A variety of limitations should be considered when interpreting results from this Study.

- The CAP events tested had small effect sizes that would challenge control chart sensitivity. The events did not present an accurate test gold standard against which to evaluate chart sensitivity.
- The initial baseline of m=25 subgroups was used to estimate process parameters and construct charts. A decision rule was applied limiting the assignment of signal-related event status to occurring within 3 months of the event time.
- An abundance of data were provided, but few indicators were used for analysis. The AbxMin process indicator was used for most analyses; it was transformed to evaluate the effects of skew, and dichotomized to create P-charts. The outcome indicators were more global and did not appear to have tight affiliation with the process indicators.
- Events were evaluated in isolation from each other.
- The retrospective, causal comparative study design does not allow elimination of rival hypotheses.

Recommendations for Practice

There are design and statistical considerations that affect control chart sensitivity.

- Application to monitor medical quality processes will require calibrating charts to the specific medical process. Determining the control chart's "best fit" will itself require a continuous improvement effort Plan-Do-Study-Act...to the application of methods.
- Creating larger subgroups, increasing subgroup intervals, and applying Jaehn zone rules will increase chart sensitivity. Larger samples will identify smaller shifts.
- The ability to create rational subgroups remains the single biggest challenge to creating a viable control chart strategy. Process indicators may be less complex and more sensitive to monitoring medical quality. Outcome indicators may be important to monitor but will not likely be sensitive to quality interventions. Rational subgrouping is the best strategy available to control confounding.
- Maintain common subgroup sizes or implement necessary adjustments.
- Control charting in the medical environment will need to be adjusted more frequently in response to the multiple sources of variation. Chart operations such as disabling values, process adjustments, identification of cause, etc. should be documented. Disable values only when special cause is known.
- Alternate control charting methods should continually be evaluated for sensitivity.
- Control charts should be strategically placed and intensively operated. The use of Jaehn Decision Rules will increase false alarm, and control chart proliferation could lead to dissatisfaction or process tampering. An intentional strategy would assign multiple control charts to monitor different indicators for single areas of quality.

Appendix A. Analytic and Enumerative Studies

"The real world verses the statistician's world (Deming, 1953)"

Issue	Enumerative Studies	Analytic Studies
Population	Target population exists	Population unlikely to exist now
Frame	Exists at least conceptually, reasonably covers population	Does not exist
Action	Taken on the elements of the population	Action to be taken on the cause system, if it is understood, hope is to affect future processes
Time Frame	How things are now, "snapshot in time"	Future-oriented, about future processes
Prediction	Now or in short term	Projection, extrapolation of what may be if cause system is understood
Estimation	Something of importance about the population	
Traditional statistical methods	Appropriate with assumptions	Of questionable validity
Stability of Process	More or less	Process does not exist now
Statistical sampling	Schemes can be developed to ensure "equal and complete coverage of frame	Likely inappropriate since frame of future may be different; purposeful selection may make sense.
Examples	Census of US to determine congressional representation. Study to determine income characteristics of the residents in Ft. Collins. Determination of the average number of students in school during one week in October for the purpose of obtaining allocation of funds.	Polls used to forecast elections in advance. Reliability studies on laboratory prototype models in order to predict the mean life for products released in the future. Evaluation of the effects of drug A verses drug B on test subjects, and thus the impact of these drugs when released to the public.

Appendix B. Indicator Definitions

Indicator	Description	Numerator	Denominator	Excluded
Time to Antibiotic	Time from hospital arrival to administration of first antibiotic for inpatients with community acquired pneumonia	Received antibiotic between 0-2160 minutes (36 hours) following arrival to hospital and met one of the following condition sets: Pneumonia diagnosis on admission principal ICD-9 codes 480.0-487.0 OR Septicemia codes 038.0-038.9 and secondary pneumonia diagnosis OR discharge diagnosis for respiratory failure and secondary diagnosis of pneumonia	(none)	• Pts less than 29 days of age • Pts received in transfer from another acute care hospital or critical access facility • Pts who have no working diagnosis of pneumonia at time of admission • Patients receiving comfort measures only • Pts who received antibiotics prior to arrival but NOT during hospital stay • Patients who did not receive antibiotics or it cannot be determined from their medical record • Patients whose initial antibiotic was administered later than 2,160 minutes (36 hours) from time of arrival

Initial antibiotic dose within 4 hours	Proportion of pneumonia inpatients who received antibiotics within 36 hours of admission who received their initial dose of antibiotics within 4 hours of hospital arrival	Total number of pneumonia inpatients who received any dose of antibiotics within 4 hours of arrival to the hospital	Inpatients with community acquired pneumonia, including those transferred from long term care facilities, who received antibiotics within 36 hours after arrival to hospital and met one of the following conditions: diagnosis on admission of Pneumonia ICD-9 Codes 480.0-487.0; OR Principal ICD-9 code for Septicemia 038.0-038.9 with a secondary diagnosis of pneumonia OR principal diagnosis of pneumonia and secondary code for respiratory failure ICD-9 518.81 or 518.84	• Pts less than 29 days of age • Pts transferred from another acute care hospital or critical access facility • Pts who did not receive antibiotics within 36 hours after hospital arrival • Pts who have no working diagnosis of pneumonia at time of admission • Patients receiving comfort measures only • Patients who received antibiotics prior to arrival but NOT during hospital stay • Patients who did not receive antibiotics or it cannot be determined from their medical record
Initial antibiotic dose within 8 hours	Proportion of pneumonia inpatients who received antibiotics within 36 hours of admission who received their initial dose of antibiotics within 8 hours of hospital arrival	Total number of pneumonia inpatients who received any dose of antibiotics within 8 hours of arrival to the hospital	Inpatients with community acquired pneumonia, including those transferred from long term care facilities, who received antibiotics within 36 hours after arrival to hospital and met one of the following conditions: diagnosis on admission of Pneumonia ICD-9 Codes 480.0-487.0; OR Principal ICD-9 code for Septicemia 038.0-038.9 with a secondary diagnosis of	• Pts less than 29 days of age • Pts transferred from another acute care hospital or critical access facility • Pts who did not receive antibiotics within 36 hours after hospital arrival • Pts who have no working diagnosis of pneumonia

			pneumonia OR principal diagnosis of pneumonia and secondary code for respiratory failure ICD-9 518.81 or 518.84	at time of admission • Patients receiving comfort measures only • Patients who received antibiotics prior to arrival but NOT during hospital stay • Patients who did not receive antibiotics or it cannot be determined from their medical record
Blood Culture prior to Antibiotic	Collection of blood culture prior to first dose of antibiotic	Pneumonia inpatients whose blood cultures are collected before the first dose of antibiotic is administered in the hospital	Inpatients age 29 days and older with community acquired pneumonia, including those transferred from long-term care facilities who received antibiotics during hospital stay. Pneumonia ICD-9 Codes 480.0-487.0 or Septicemia 038.0-038.9	From denominator: • Pts less than 29 days of age • Pts received in transfer from another acute care hospital or critical access facility • Pts who have no working diagnosis of pneumonia at time of admission • Patients receiving comfort measures only • Pts having no blood cultures obtained • Antibiotics received only prior to arrival and NOT during hospital stay • Antibiotic not received or you are unable to determine medical record documentation
Influenza vaccine (documented)	Pneumonia inpatients 65years and older screened	Number of pneumonia inpatients that were screened for	Inpatients age 65 years and older with community	From denominator: • Patients who have no working diagnosis at time

given	for or given influenza vaccine when needed	vaccination status and were not vaccinated due to contraindications or refusal, or needed vaccine and received it prior to discharge	acquired pneumonia, including those transferred from long term care facilities. Pneumonia ICD-9 Codes 480.0-487.0 or Septicemia 038.0-038.9 or 518.81 or 518.84	of admission • Patients receiving comfort measures only • Patients less than 65 years • Patients who left against medical advice • Patients who expired, or were discharged to inpatient or home hospice program • Patients received in transfer from another acute care hosp. or crit. access facility
Pneumonia Screening and/or Vaccination	Pneumonia inpatients 65 years of age and older screened for and / or given pneumonoccol vaccine when needed	Number of pneumonia inpatients that were screened for vaccination status and were not vaccinated due to contraindications or refusal, or needed vaccine and received it prior to discharge	Inpatients age 65 years and older with community acquired pneumonia, including those transferred from long term care facilities. Pneumonia ICD-9 Codes 480.0-487.0 or Septicemia 038.0-038.9 or 518.81 or 518.84	From denominator: • Patients who have no working diagnosis at time of admission • Patients receiving comfort measures only • Patients less than 65 years • Patients who left against medical advice • Patients who expired, or were discharged to inpatient or home hospice program • Patients received in transfer from another acute care hospital or critical access facility

Average Length of Stay (ICD-9 and DRG)	Total # of midnights pt was counted as part of hospital census .			
Readmission within 31 days % (ICD-9 and DRG)	Rate of non-elective readmission to acute care within 31 days among acute care inpatients with a primary diagnosis of pneumonia	Number of patients discharged from acute care with a diagnosis of pneumonia, who return with non-elective diagnosis within 31 days.	Total number of acute care patients discharged with a diagnosis of Pneumonia from the PREVIOUS month	• Outpatient encounters, or inpatient encounters for rehab, psychiatry, skilled nursing or hospice care • ED encounter • Admission status equivalent to elective
Adult Mortality Rate ICD-9 and DRG	Proportion of pneumonia inpatients who expired or were discharged to hospice	Number of patients discharged from acute care to hospice, or Expired after having been diagnosed with community acquired pneumonia ICD-9 Codes 480.0-487.0 OR from DRG 089 or 079 (Simple Pneumonia and Pleuresy)	Total number of acute care patients discharged with a diagnosis of pneumonia ICD-9 Codes 480.0-487.0 OR From DRG 089 or 079	
Age<80 Mortality Rate	Proportion of pneumonia inpatients <80 years of age who expired or were discharged to hospice	Number of patients age <80 years discharged from acute care to hospice, or Expired after having been diagnosed with community acquired pneumonia ICD-9 Codes 480.0-487.0	Total number of acute care patients age <80 years discharged with a diagnosis of pneumonia ICD-9 Codes 480.0-487.0	Age ≥ 80
Age>79 Mortality Rate	Proportion of pneumonia inpatients >79 years of age who expired or were discharged to hospice	Number of patients >79 years discharged from acute care to hospice, or Expired after having been diagnosed with community acquired pneumonia ICD-9 Codes 480.0-487.0	Total number of acute care patients age >79 years discharged with a diagnosis of pneumonia ICD-9 Codes 480.0-487.0	Age<80
(Severity-Adjusted) Complication Rate	Rate of pneumonia inpatients with complications during hospital stay	# pneumonia patients with complications	Total # pneumonia patients	None

REFERENCES

Abelson, R.P. (1995). Statistics as Principled Argument. Lawrence Erlbaum Associates, Publishers.

Amin, S. Control Charts 101: A Guide to Health Care Applications. Quality Management in Health Care. 2001, 9(3). 2001. Aspen Publishers.

Bartlett J., Breiman R., Mandell, L., & Fine T. Community-Acquired Pneumonia in Adults: Guidelines for Management. Clinical Infectious Diseases. 1998; 26:811-838.

Bartlett J., Dowell S,, Mandell L., File T., Musher D., & Fine M. Practice Guidelines for the Management of Community Acquired Pneumonia in Adults. Clinical Infectious Diseases 2000; 31:347-82.

Barton, R. and Chandra, M. Statistical Methods for the management of Process Quality: A State of the Art Review. International Journal of Reliability and Saftey Engineering. Vol. 3, No. 2. (1996) World Scientific Publishing Company.

Battleman, B., Callahan, M, Thaler, H. Rapid Antibiotic Delivery and Appropriate Antibiotic Selection Reduces Length of Stay of Patients with Community Acquired Pneumonia.. American Medical Association. Archives of Internal Medicine. Vol 162, Mar 25, 2002.

Bai, D. and Choi, I. Xbar and R Control Charts for Skewed Populations. Journal of Quality Technology. Vol. 27, No. 2. April, 1995.

Benneyan, J. Statistical Quality Control Methods in Infection Control and Hospital Epidemiology, Part I: Introduction and Basic Theory. Infection Control and Hospital Epidemiology, Vol. 19, No. 3. March, 1998.

Benneyan, J. Statistical Quality Control Methods in Infection Control and Hospital Epidemiology, Part II:Chart Use, Statistical Properties, and Research Issues. Infection Control and Hospital Epidemiology, Vol. 19, No. 4. April, 1998.

Boggs, P., Hayati, P., Washburne, W. & Wheeler, D. Using Statistical Process Control for the Continual Improvement of Asthma care. Journal on Qualty Improvement. Vol. 25, No. 4. April, 1999.

Bothe, D. (1997) Measuring Process Capability: Techniques and Calculations for Quality and Manufacturing Engineers. McGraw-Hill. New York.

Boyles, R. Phase I Analysis for Autocorrelated Processes. Journal of Quality Technology. Vol. 32, No. 4. October, 2000.

Box, G. and Luceño, A. (1997). Statistical Control by Monitoring and Feedback Adjustment. John Wiley & Sons.

Blumenthal, D. and Kilo, CM A Report Card on Continuous Quality Improvement.. The Milbank Quarterly, Vol 76, No 4. 1998.

Brook, R., McGlynn, E., & Shekelle, PG. Defining and Measuring Quality of Care: A Perspective from U.S. Researchers. International Journal for Quality in Health Care. 2000 Aug; 12(4): 281-95.

Campbell, R, Mo, S..& Buetow, S. Defining Quality of Care. Social Science and Medicine. 51 (2000) 1611-1625.

Campbell, D. and Fiske, D. Convergent and Discriminant Validation by the MultiTrait-Multimethod Matrix.. Psychological Bulletin, Vol 56, No.5. March 1959.

Carey, R. and Lloyd, R. (1995). Measuring Quality Improvement in HealthCare: A Guide to Statistical Process Control Applications. Quality Resources. New York.

Chassin, M. Is Health Care Ready for Six Sigma Quality? The Milbank Quarterly, Vol 76, No 4. 1998.

Clinical Policy for the Management and Risk Stratification of Community-Acquired Pneumonia in Adults in the Emergency Department. American College of Emergency Physicians (ACEP). Annals of Emergency Medicine. July 2001; 38(1):107-13.

Cohen, J., Cohen, P. (1983) Applied Multiple Regression / Correlation Analysis for the Behavioral Sciences. 2nd Edition. Lawrence Erlbaum Associates, Publishers. 1983.

Crowder, S. Design of Exponentially Weighted Moving Averages Schemes. Journal of Quality Technology. Vol. 21, No. 3.July, 1989.

Davis, R. and Woodall, W. Performance of the Control Chart Trend Rule under Linear Shift. Journal of Quality Technology. Vol. 20, No.4, October 1988.

Deleryd, M. The Effect of Skewness on Estimates of Some Process Capability Indices. Journal of Applied Quality Management. Vol. 2, No. 2. 1999.

Deming, WE. (1950). Some Theory of Sampling. John Wiley and Sons.

Deming, WE. (1982). Out of the Crisis. Massachusetts Institute of Technology, Center for Advanced Educational Services.

Deming, WE. (1994). New Economics.for Industry, Government, Education, 2nd Edition. Massachusetts Institute of Technology, Center for Advanced Educational Services.

Deming, WE. On Distinction Between Enumerative and Analytic Surveys. Journal of the American Statistical Association 48, 244-255 1953

DeVor, R., Chang, T. & Sutherland, J. (1992). Statistical Quality Design and Control. Macmillan Publishing Company.

Doty, L. (1996). Statistical Process Control. Industrial Press, 2nd Ed..

Federal Register. Part 482, 483, 484. Vol 67, No. 191. Wednesday, October 2, 2002.

Financial and Clinical Benchmarking: The Strategic Use of Data. Healthcare Financial Management Association. 1997, HCIA, Inc.

Fine, M., Deron, G., Petrillo, M., & Meehan, T. Patient and Hospital Characteristics Associated with Recommended Processes of Care for Elderly Patients Hospitalized with Pneumonia.. Archives of Internal Medicine. Vol 162, April 8, 2002.

Fine, M., Smith, M., Carson, C., Mutha, S., & Steadman, S. Prognosis and Outcomes of Patients with Community –Acquired Pneumonia: A Meta-analysis. JAMA. January 10, 1996-Vol 275, No. 2. 134-141.

Fine M., Auble T., Yealy D., et al. A Prediction Rule to Identify Low-Risk Patients with Community-Acquired Pneumonia. The New England Journal of Medicine 1997 (January 23); 336:243-250.

Gin, J. Class notes, Systems and Industrial Engineering 406/506. Spring 2001 Semester.

Gliner, J. and Morgan, G. (2000) Research Methods in Applied Settings: An Integrated Approach to Design and Analysis. Lawrence Erlbaum Associates, Publishers.

Halm E., Fine M., Kapoor W., et al. Instability on Hospital Discharge and the Risk of Adverse Outcomes in Patients with Pneumonia. Arch Intern Med 2002; 162:178-84.

Huth, E. Needed: An Economics Approach to Systems for Medical Information. Ann Intern Med 1985 Oct; 103(4):617-619.

Institute of Healthcare Improvement (IHI). Quintessential Quality: In Memorium: A Tribute to Avedis Donabedian. (On-line, www.ihi.org). Volume III, Number 22: November 15, 2000

Institute of Medicine. (2001). Crossing the Quality Chasm: A New Health System for the 21st Century. (National Academy Press). Washington, DC.

Institute of Medicine. (1999). To Err is Human: Building a Safer Health System. (National Academy Press). Washington, DC.

Institute of Medicine. (1999). Measuring the Quality of Health Care. (National Academy Press). Washington, DC.

Institute of Medicine. (1999). Statement on Quality of Care: National Round Table on Health Care - The Urgent Need to Improve Health Care Quality. (National Academy Press). Washington, DC.

Institute of Medicine. (2000). America's Health Care Safety Net: Intact but Endangered. (National Academy Press). Washington, DC.

Institute of Medicine. (2001). Coverage Matters: Insurance and Health Care. (National Academy Press). Washington, DC.

Ishikawa, K. (1972). Guide to Quality Control. Asian Productivity Organization.

Jaehn, A. Improving QC Efficiency with Zone Control Charts. ASQC Quality Congress Transactions-Minneapolis. 1987.

Kaplan, A. (1998). The Conduct of Inquiry: Methodology for Behavioral Science. Chandler Publishing Company. San Francisco

Kelly, H. and Drury, C. Sociotechnical Reasons for the De-evolution of Statistical Process Control.. Quality Management Journal. Vol 9, Issue 1. January, 2002.

Kendall, M. and Ord, K. (1990) Time Series. Edward Arnold.

Lee, K. and McGreevey, C. Using Control Charts to Assess Performance Measurement Data. Journal on Quality Improvement. The Joint Commission. Vol. 28, No. 2. February, 2001.

Lim, T., Ding, L., and Morad, Z. Assessing Doctors' Competence Application of the CUSUM Technique in Monitoring Doctors' Performance. International Society for Quality in Health Care. Vol. 14, No. 3. Oxford University Press.

Markowitz, J., Pashko, S., Gutterman E., et al. Death Rates Among Patients Hospitalized with Community-Acquired Pneumonia: A Re-examination with Data from Three States." American Journal of Public Health. Vol. 86, No. 8. August 1996

McCleary, R., and Welsh, W. (1992). Philosophical and Statistical Foundations of Time Series Experiments in Single Case Research Design and Analysis. Kratcochwill, T. and Levin, J. (Eds.) Lawrence Erlbaum Associates, Publishers.

McGarvey, R. and Harper, J. "Pneumonia Mortality Reduction and Quality Improvement in a Community Hospital." QRB. April 1993

Meehan, T., Chua-Reues, J. & Tate, J. Process of Care Performance, Patient Characteristics, and Outcomes in Elderly Patients Hospitalized with Community Acquired or Nursing Home-Acquired Pneumonia. Clinical Investigations. Chest May 2000. 1378 – 1385.

Meehan T., Fine M., Krumholz H., et al. Quality of Care, Process and Outcomes in Elderly Patients with Pneumonia. JAMA. 1997; 278:2080-2084.

Metlay J., Fine M. Testing Strategies in the Initial Management of Patients with Community-Acquired Pneumonia. Annals of Internal Medicine. Volume 138, Number 2. January 21, 2003.

Millman M, (ed.). Access to Health Care in America. Committee on Monitoring Access to Personal Health Care Services. Washington (DC): National Academy Press. 1993.

Moen, R., Nolan, T. &. Provost, L. (1999) Quality Improvement through Planned Experimentation. McGraw Hill Publishers.

Montgomery, D. (2001). Introduction to Statistical Quality Control.. John C. Wiley & Sons. 4th Edition.

Morbidity and Mortality Weekly Report. Prevention and Control of influenza: Recommendations on the Advisory Committee on Immunization Practices. Centers for Disease Control and Prevention. Vol. 51, No. RR-3. April 12, 2002.

Mortensen, E., Coley, C., Singer, D., Marrie, T., Oborsky, S., Wishwa, K., & Fine, M. "Causes of Death for Patients with Community-Acquired Pneumonia." Archives of Internal Medicine. Vol. 162. May 13, 2002.

Murdock, M. Building Improvement into Health Care and Service. ASCQ Quality Congress Transactions. 1992. Nashville, Tenn.

Nelson, L. Calculation of New Limits for Xbar, R Charts When Subgroup Size is Changed. Journal of Quality Technology, Vol. 20, No. 2. April 1988.

Niederman M., Mandell L.,Anzueta A, Bass, J., Fine M. American Thoracic Society Guidelines for the Management of Adults with Community-acquired Pneumonia:

Diagnosis, Assessment of Severity, Antimicrobial Therapy, and Prevention. American Journal of Respiratory and Critical Care Medicine. Volume 163, Number 7. June 2001.

NWA Quality Analyst, Version 5.2. 1997-2002. Northwest Analytical, Inc. Portland, Oregon.

Palm, A.C. Journal of Quality Technology. Vol. 32, No. 4. October 2000.

Practice Guidelines for the Management of Community Acquired Pneumonia in Adults. National Guidelines Clearinghouse. (On-line, www.guideline.gov). Abstracted from the web, May 25, 2003.

Quesenberry, C. (1997). SPC Methods for Quality Improvement. John Wiley & Sons.

Quesenberry, C. The Effect of Sample Size on Estimated Limits for Xbar and X Control Charts. Journal of Quality Technology, Vol 25, No. 4. October 1993.

Quesenberry, C. Statistical Processs Control Geometric G-Chart for Nosocomial Infection Surveillance. American Journal of Infection Control. Vol. 28, No. 4. August, 2000.

Ramberg, J. (personal communication, February, 2002).

Ramsey, P. and Ramsey, P. Simple Tests of Normality in Small Samples. Journal of Quality Technology, Vol 22, No. 4. October 1990.

Rushe, R. and Gottman, J. (1993) Essentials in the Design and Analysis of Time Series Experiments in A Handbook for Data Analysis in the Behavioral Sciences. Ed. Keren, G. and Lewis, C. Lawrence Erlbaum Associates, Publishers. Hillsdale, NJ.

Schiff, G. and Rucker, T. Beyond Structure-Process-Outcome: Donabedian's Seven Pillars and Eleven Butresses of Quality. Journal on Quality Improvement. Vol. 27, No. 3. March 2001.

Schuster, M., McGlynn, E. & Brook, R.. How Good is the Quality of Health Care in the United States? The Milbank Quarterly, Vol 76, No 4. 1998.

Sechrest, L., McKnight, P. & McKnight, K. Calibration of Measures for Psychotherapy Outcome Studies. American Psychologist. 1996.

Sellick, J. The Use of Ststistical Process Control Charts in Hospital Epidemiology. Infection Control and Hospital Epidemiology. Vol 14, No. 11. November, 1993.

Shewhart, W. (1931). Economic Control of Quality of Manufactured Product.. D. Van Nostrand Company, New York.

Shewhart, W. (1939). Statistical Method from the Viewpoint of Quality Control. Department of Agriculture.

Shortell, S., Bennett, J. & Byck, G. Assessing the Impact of Continuous Quality Improvement on Clinical Practice. What Will it Take to Accelerate Progress? The Milbank Quarterly, Vol 76, No 4. 1998.

Stanton, M. Improving Treatment Decisions for Patients with Community Acquired Pneumonia. Agency for Healthcare Research and Quality. Research in Action, Issue 7. May 20, 2003.

The Statistician Who Changed the World: W. Edwards Deming, 1900 – 1993. The American Statistician, August 1994. Vol. 48, No. 3.

TQT Improvement Tools: Total Quality Transformation QIP. Inc./PQ Systems, Inc.

Wennberg, J, Cooper, M, et. al. (1998). The Dartmouth Atlas of Health Care in the United States. American Hospital Publishing.

Wheeler, D. (Online, www.spcpress.com). Webpage accessed July, 2001

Wheeler, D. (1992). Understanding Statistical Process Control. SPC Press, Knoxville, Kentucky.

Wheeler, D. (1995). Advanced Topics in Statistical Process Control. SPC Press, Knoxville, Kentucky.

Wheeler, D. (2000). Understanding Variation: The Key to Managing Chaos.. 2nd Edition. SPC Press.

Willerman, T. and Runger, G. Designing Control Charts Using an Empirical Reference Distribution. Journal of Quality Technology. Vol 28, No. 1. January, 1996.

Woodall, W. Controversies and Contradictions in Statistical Process Control.. Journal of Quality Technology, Vol. 32, No. 4. October 2000.

Khang, N. (personal communication, February, 2004).

Zimmerman, S., and Icenogle,M. (1999). Statistical Quality Control Using Excel. ASQ Quality Press. Milwaukee, WI.

Zynx Health Incorporated. Sample Community-Acquired Pneumonia Pathway. (On-line) Accessed, June 2003.