Simulating the Large Population Effects of Quality Management Techniques in a Health Care Clinical Process

by

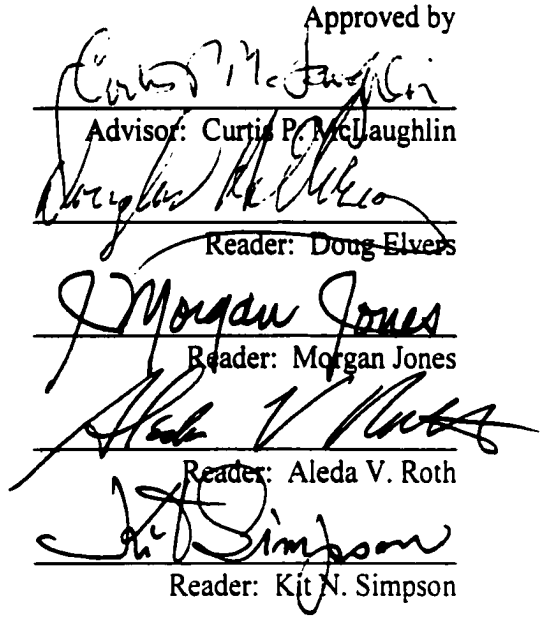
Susan Paul Johnson

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor in Philosophy in the Kenan-Flagler Business School (Operations Management).

Chapel Hill

1997

Approved by



Advisor: Curtis P. McLaughlin

Reader: Doug Elvers

Reader: Morgan Jones

Reader: Aleda V. Roth

Reader: Kit N. Simpson

ABSTRACT
SUSAN PAUL JOHNSON: Simulating the Large Population Effects of Quality Management Techniques in a Health Care Clinical Process
(Under the direction of Curtis P. McLaughlin)

The central area of interest in this research is developing methods by which we can assess the effects of quality management techniques in service operations management. Frequently, service operations management research is undertaken by adapting methods from production operations management efforts. While fruitful, we feel that this research is limited in applicability to service industries. For this reason, we choose to pursue exploratory methodologies uniquely tailored to the service industries.

This exploratory research involves simulation of a service process and investigates modeling specific quality management techniques commonly used in this process. We look at the health care industry and model the process of premature labor and delivery as a context for our investigation. Using Maryland discharge data a model of premature labor and delivery is developed and modified using Monte Carlo simulation techniques to include medical interventions and quality management techniques. Health outcomes, resource utilization and costs are measured.

The success of this methodology is in its many unique attributes. It enables researchers to explore the impact of the use of quality management techniques on a large population rather than looking at specific applications. Additionally, using real data in

conjunction with well-documented effects of medical interventions and expected results from quality management is particularly helpful in linking processes with outcomes. The random selection/random replacement technique adopted from bootstrapping methods allows the models to embrace the inherent variation of the population and measure the results rather than controlling them. Finally, it provides a methodological alternative to randomized controlled trials to the health policy community.

The value of this research is not only in the development of a new research methodology but also in the managerial insights it provides. Simulation modeling can be used both to determine and implement quality improvement goals. It also shows that the goals set can be progressively more difficult to reach and that there are diminishing returns to quality improvement efforts. For the health policy community, the apparent lower bound to the desired cesarean section rate and the evidence of a relationship between respiratory distress syndrome (RDS) and cesarean section are both significant results from this study.

DEDICATION

To Ted

To Maggie and Bridget

May I be one half the inspiration to you that you have been to me.

ACKNOWLEDGMENTS

For a "solo" research effort, there have been innumerable people without whose help I would not have completed this work. First and foremost is Curt McLaughlin, whose insights, guidance, and tireless ability to read mountains of paper in a short amount of time I could not have done without. He has helped me overcome, in his words, the "demons of variation" in more ways than one. Each of the other members of my committee - Aleda Roth, Kit Simpson, Morgan Jones, and Doug Elvers - also guided me skillfully through this investigation. I am thankful for all that I have learned from them. Additionally, I am grateful for the support and teaching of the other faculty members in the OM/QM department and the business school in general. I am especially in debt to the kind leadership of both Ann Marucheck and David Ravenscraft in overseeing the Ph.D. program. I would be remiss also if I did not mention Dean Barrett Hazeltine of Brown University for providing me with the first flickers of passion for studying business. All of these individuals have helped me build a foundation of knowledge which I plan to build on for a lifetime.

There were many people who helped me through some of the more mundane and administrative portions of this work. While I can't name all of them here, I would like to acknowledge the undying support of Liz Griffin, Kim Vaughan and my fellow OM/QM students and colleagues (especially Larry Menor).

While the current wisdom is that it takes a village to raise a child, I have experienced the support of that village in my efforts here. The love and encouragement of my parents and all my family members both got me started and helped me persevere even during the toughest times. Undying thanks go to my close friends in Chapel Hill - Deedee, Linda, and Ina - for always being there with an encouraging word, a moment of relief, or just some extra Diet Coke. These individuals had an uncanny ability to know what I needed and when I needed it. The entire faith community of The Church of the Holy Family helped get me through these endeavors. I praise God for His constant love and know that He will continue to show me the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AHCPR	Agency for Health Care Policy and Research
CDC	Centers for Disease Control
CDRG	Categorical Diagnosis Related Group
CI	Confidence Interval
CQI	Continuous Quality Improvement
DRG	Diagnosis Related Group
HCFA	Health Care Financing Administration
ICD-9	International Classification of Diseases, 9th Revision
LCL	Lower Confidence Limit
LOS	Length of Stay
NHDS	National Hospital Discharge Survey
NICU	Neonatal Intensive Care Unit
NIH	National Institutes of Health
OR	Odds Ratio
PROM	Premature Rupture of Membranes
RDS	Respiratory Distress Syndrome
SUDAAN	Survey Data Analysis
UCL	Upper Confidence Limit

CHAPTER 1

INTRODUCTION

> When we asked Pooh what the opposite of an Introduction was, he said "The what of a what?" which didn't help us as much as we had hoped, but luckily Owl kept his head and told us that the Opposite of an Introduction, my dear Pooh, was a Contradiction; and, as he is very good at long words, I am sure that that's what it is.
>
> A.A. Milne

The central area of interest in this research is developing methods by which we can assess the large scale population effects of quality management techniques that are applied to health care clinical processes. As such, this exploratory research involves simulation of a specific health care clinical process and modeling of specific quality management techniques. The process of premature labor and delivery is the context for our investigation. Using Maryland discharge data we develop a model of premature labor and delivery that can be modified using Monte Carlo simulation techniques to include medical interventions and incorporate the expected results of using these interventions in women experiencing premature labor.

Service Industries and Operations Management

It has been noted that rather than continuing to adapt production operations research to service operations management it is time for service industries to break out on their own and pursue research methods that are tailored to the issues unique to services (Fitzsimmons

and Fitzsimmons, 1994). For this reason, we choose to pursue exploratory methodologies uniquely tailored to the service industries.

Specifically, we have chosen to look at the health care industry, which is growing as rapidly as service industries in general, and to focus on a specific health care clinical process, premature labor and delivery, as a means to focus our research. From the management point of view, we are interested in understanding the potential effects of quality management techniques if they are implemented on a widespread basis. Quality management has developed into a significant research area in both management and health care, so this investigation provides insights uniquely applicable to both arenas. We also use a combination of methods found both in operations management and in health care research. The methodology explored here can best be described as simulation of large population datasets with a random selection and replacement strategy that is a modification of both bootstrapping (used in management) and indirect adjustment for population differences (used in epidemiology). A detailed description of this methodology is documented in this thesis.

Health Care as a Field for Service Operations Management Research

In the United States, the growth of health care expenditures has outpaced that of the rest of the economy for many years. National health expenditures, a combination of health services and supplies (including hospital care, health care provider services, administration and public health activities) and research and construction, have shown almost a tenfold increase between 1970 ($74.4 billion) to 1991 ($738.2 billion) (HCFA, 1993). When measured as a percentage of gross domestic product, these expenditures have risen from 7.2% to 13.1%

during the same time frame. At the same time, the health care system provides a great number of jobs, a number which is also increasing. Since 1950, health manpower has increased five-fold and general hospital beds have doubled (Rice, 1990). The increase in medical care expenditures can be attributed not only to demographic factors (population growth, rising proportion of elderly) and macroeconomic factors (inflation, per capita utilization of health care), but also to changes in the delivery of health care services. This trend of increasing expenditures reveals the need for attention to methods of assessing health care delivery and associated costs.

The combination of economic and regulatory forces are driving health care change and leading to fundamental changes in the way that most Americans think about and pay for their health care. As a result, the management of health care in the United States is undergoing a metamorphosis. The shifting tides in health care financing have resulted in a need for process-focused initiatives. Current quality initiatives in health care are all methods of standardizing health care processes developed in response to the need to provide quality health care at reasonable costs. Changes in health care financing have ranged from increasing private health insurance options (from 1945-1960 due to concerns about access), to an explosion of health care costs (after 1965 with the creation of Medicare and Medicaid), to a current concern with cost containment (Rice, 1990). Cost containment issues are fraught with tradeoffs between access, cost, efficiency, and quality. National priorities are to increase access, moderate cost, and maintain or increase quality (Haggerty, 1990). Cost containment may require producers of health to increase production efficiency - deliver the same amount of services with fewer inputs (Fuchs, 1990). Improvements in efficiency are possible but unlikely to be source of

major cost reductions; there was a reward for inefficiency in the previous health care system that will be eliminated by current initiatives. Thus, new methods are needed that will better assess the tradeoffs between cost, quality and efficiency are needed.

Premature Labor and Delivery as a Context for Investigation

To provide a specific context for this analysis, one health care clinical process has been selected. The health care process investigated is preterm labor and delivery. This process and associated medical interventions are chosen because the end point is common (discharge of the infant from the hospital) and there exists consensus as to what constitutes a positive outcome (a healthy baby and a healthy mother). Labor is considered to be premature when it occurs before 37 completed weeks of gestation (259 days).

We have also selected this process as it enables us to test the ability to model both medical and management interventions. In addition to the processes represented by the Maryland discharge data, three commonly-used medical interventions are included in this discussion of the process. One, cesarean section, is embedded in the data. The other two, the use of corticosteroids to stimulate lung development prior to delivery and the use of surfactant to mimic mature lung development, are incorporated through modeling techniques. Finally, three quality management techniques that have been used with this process and the realized or anticipated effects of each technique are modeled. The results of this modeling effort provide insight as to the utility of simulation techniques used in

conjunction with large empirical datasets in health care management. This in turn provides some inkling of the advantages of this methodology in any service industry.

Overview of This Investigation

This thesis attempts to present a background for these issues. provide a context for investigation, create a simulation model as the basis of the research, conduct simulations and report the results. The subsequent chapters are organized as follows.

Chapter Two presents a foundation of literature on which the investigation is based. It includes literature from both services operations management and health services research. Research questions and propositions are posed at its conclusion. Chapter Three attempts to narrow the focus of the problem by presenting the process of premature labor and delivery as a context for the investigation. Chapter Four provides details of the simulation modeling efforts. The modeling results are found in Chapter Five, and these results are discussed in Chapter Six. Throughout, the focus of this work is that in a unique manner, we utilize simulation models that specifically measure the effects of medical intervention and quality management techniques (process) and outcomes in order to gain policy-related insights.

CHAPTER 2

LITERATURE REVIEW

A man will turn over half a library to make one book.

Samuel Johnson

In considering service operations management research, it is notable that while the service economy is growing at an ever increasing rate, less research is devoted to this arena than to manufacturing. Typically, the results of production operations management research has been "translated" into service operations management. For a long time, this has been sufficient to provide insights into management in the service industries. For instance, quality management research began with Shewart, and later Deming, looking at quality in production processes. After research in quality of production operations management brought gains to manufacturers, service providers began to adopt the methods. As researchers have delved into the service arena more and more diligently, however, it has become apparent that this *modus operandi* has its limitations. In particular, the characteristics that make service industries unique - perishability, intangible, need for customer contact and participation, labor intensive, and variable, nonstandard output - make service quality difficult to measure and therefore make it difficult to achieve the same results as in manufacturing production areas (Fitzsimmons and Fitzsimmons, 1994). Accordingly, research specific to service industries is needed. This need has sparked our interest and thus we take on the task of doing research specifically in the service industries. In particular, we are interested in developing methods that better suit research in service industries. To explore new methods, we use the health care industry as an example.

Why health care? The trend of increasing health care expenditures in the United States reveals the need for attention to methods of assessing health care delivery and associated costs. Additionally, as U.S. health care providers focus on improving the quality of health care delivery and apply quality management techniques to health care clinical processes, there are associated policy questions regarding the anticipated results of these efforts. It would be helpful to have the ability to use large population datasets to investigate the impact of these techniques on a broad level in order to guide policy decisions and development.

The central area of interest in this research is developing methods by which we can assess the large scale population effects of quality management techniques that are applied to health care clinical processes. As such, this exploratory research will involve simulation of a specific health care clinical process and modeling of specific quality management techniques. This chapter provides an introduction to the issues shaping the need for this investigation, combining information from both the perspectives of service operations management and health services research in order to formulate the problem statement and to provide the methodological basis for the investigation.

The background literature comes from three major streams: process management and control, quality management techniques, and research methods. For each of these areas, there is relevant work in both the service operations management and health services arenas. These research streams and their intersections are noted in Figure 2.1. We conclude the literature review with a statement concerning the specific problem that is addressed in this work, along with specific research questions, propositions, and hypotheses. Additionally, to provide a

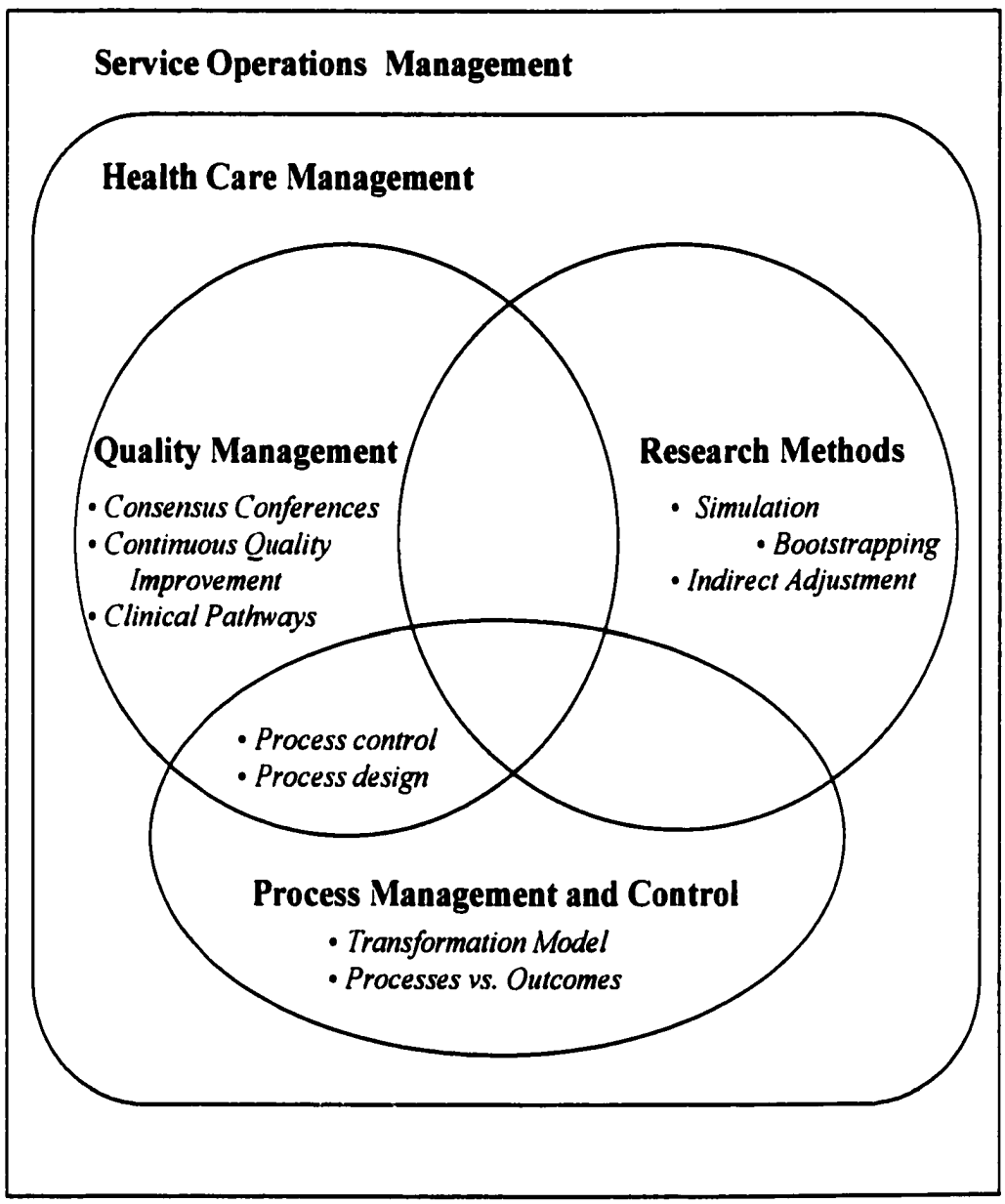
specific context for this research, we review the medical literature on premature labor and delivery.

We use Figure 2.1 to guide the following discussion. We will begin with process management and control in order to set the scene for the research and then begin to focus on health care as a specific application to be addressed in this research. Next, a section on quality management techniques is presented. Broadly characterized, these techniques are focused either on process control or on process design. We consider three techniques in particular that are commonly used in the health care arena, namely consensus conference guidelines, continuous quality improvement (CQI) and clinical pathways. Then, we consider literature regarding research methods, including a discussion of simulation research methodology, its use in the health care arena, and its applicability in this investigation. A complete discussion of the specific physical and clinical process that is the focus of this dissertation, premature labor and delivery, is given in the subsequent chapter. The details about this clinical process and the medical interventions that are commonly used during this process are documented in the literature and form the basis for our modeling efforts.

Process Management and Control - Defining a Model

To begin this investigation , we first needed a model that can help us structure our discussion and provide a basis for process analysis. Specifically, we have identified a transformation model of service operations management that has been adapted to our specific context - delivery of health care services. Thus, we quickly

Figure 2.1. Diagram of Research Streams Contributing to This Investigation



move from the domain of service operations management illustrated in Figure 2.1 to the narrower lens of health care as a field for investigation. We then develop an extension of this model as a useful tool for process analysis in health care.

Models from Operations Management

The model used here starts with a basic transformation model. Transformation models of production are common in operations management and are the simplest representation of a production process. They are equally applicable to the production of goods or services. A transformation model begins with all inputs into a process and through a conversion mechanism transforms the inputs into outputs. The inputs to a system are processed to produce the output desired. Inputs, including customer demand, financial resources, equipment, labor and energy, are independent or exogenous variables. Outputs are dependent or endogenous. Structural and material parameters describe the system as well. Structural, or geometric, parameters include organization and interaction between subsystems. Material, or properties, parameters can include the type of processing. (Reisman, 1979). One of the basic principles of quality management is that productive work is accomplished through processes; with workers having triple roles as customer, processor and supplier. Inputs are received from suppliers (the recipient is a customer), used as processors, and then supplied as services to other customers (Berwick et al., 1990).

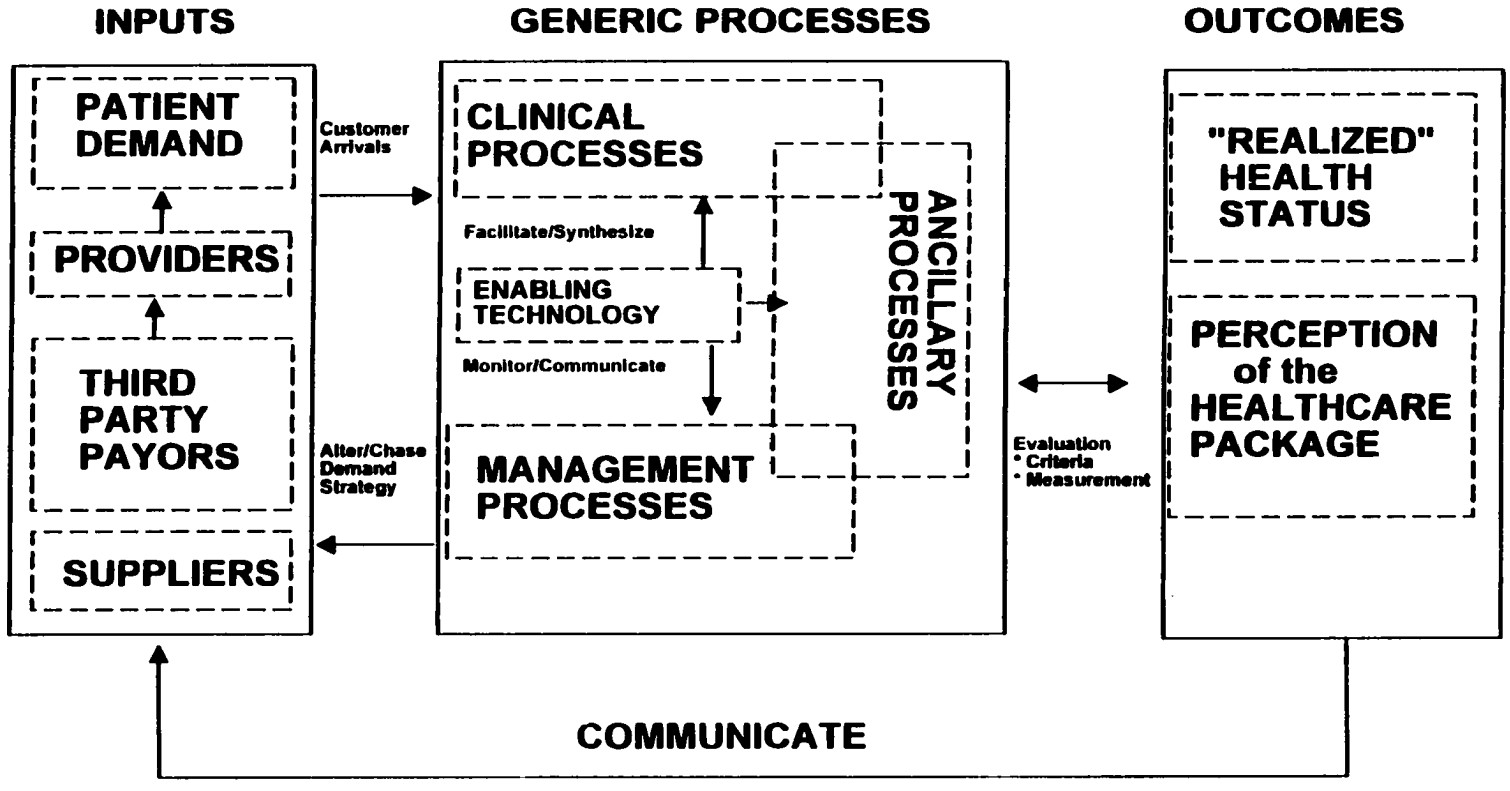
A Transformation Model of Health Care Processes

The production model applies to health care, but is not necessarily readily embraced by health care providers. Berwick et al. noted that seeing health care as a 'production process' by any name would have violated both the self-image of the profession and the desires of the patients (1990, p. 11). This reflects the tension between quality improvement centered on processes to provide health care and the common myth that health care is whatever is provided by doctors. More recent evidence, largely from the quality movement, has shown that health care can indeed be considered a production process and that physicians can embrace this view (Berwick et al., 1990; Johnson et al., 1993). Thus, researchers have adapted this type of model for health care. This model is presented in Figure 2.2 as a Meta-Process Model of a Health Care Delivery System (Roth, 1993; Roth et al., 1996). The Meta-Process Model provides a broad view of health care delivery. There is evidence from the actual provision of medical care that a more detailed model can provide a basis for further analysis. As such, this evidence is incorporated into a phased model of medical care delivery.

Defining the Meta-Process Model

In this model consumer demand (patients) for hospital services is the primary input. Other inputs are providers, suppliers, and payers. All of these impact demand. Three generic types of processes convert inputs to outputs in a hospital delivery system: clinical, management and ancillary. The model illustrates processes organized around outcomes rather than individual tasks, activities, or units. Two types of

Figure 2.2. Meta-Process Model of a Health Care Delivery System (Roth, Johnson and Short, 1996).



outcomes from hospital delivery processes are recognized: 'realized' health status, commonly referred to as clinical outcomes, and external patient perception of care and related services that they received. (Roth et al., 1996). Accompanying this model is the assumption that an effective health care delivery system consists of waste-free, value-added processes that take inputs from suppliers and translate them into desirable outcomes (Roth, 1993).

Other researchers have suggested similar, encompassing process views of health care delivery. A similar four-component model including patient-related risk factors, clinical processes ordered by the attending physician, hospital execution of the physician's plan, and the patient's outcome(s) has been proposed in conjunction with process improvement efforts (Steen, 1994). Another description of medical care provision notes that the process of physician work includes defining a disease, which ends in a diagnosis, selecting a procedure, observing outcomes, and assessing preferences (Eddy in Lee and Estes, 1990).

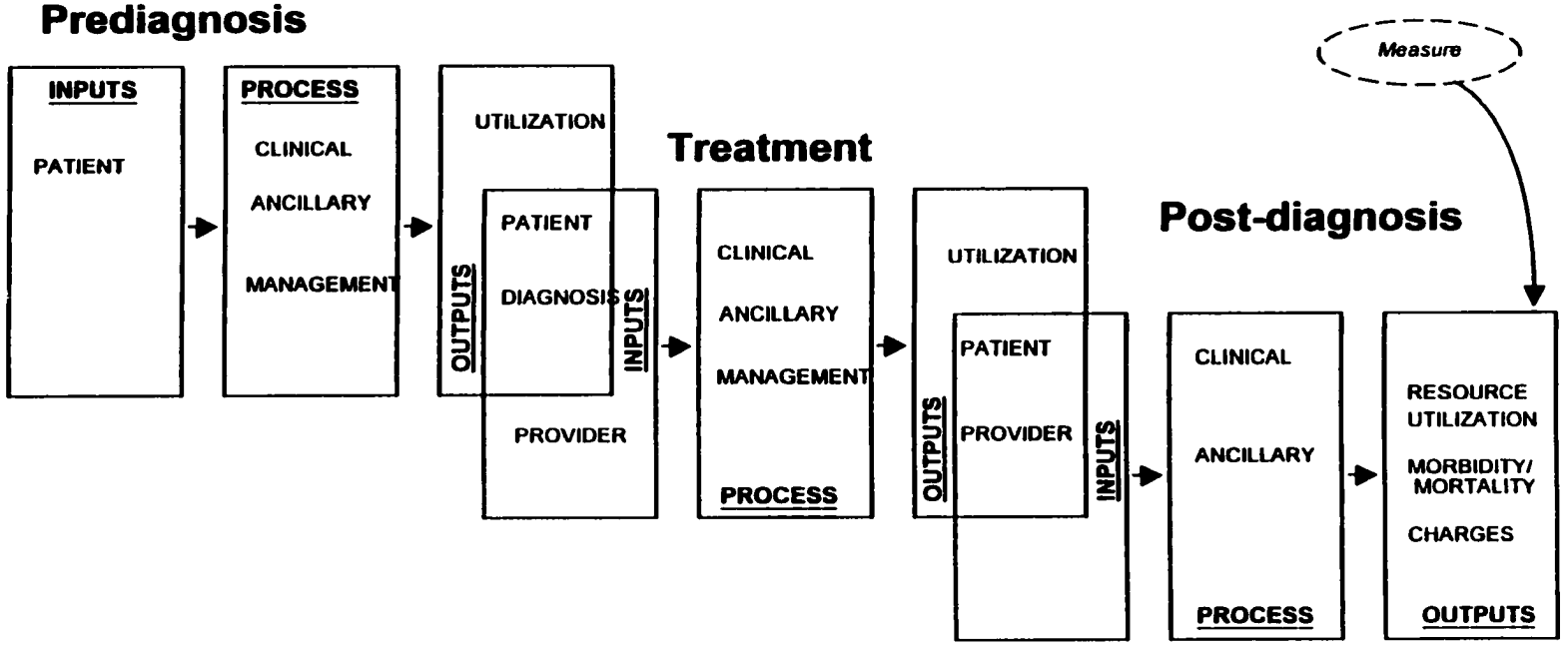
This latter description is intuitively pleasing in that it considers the diagnosis as an endpoint or outcome of one portion of the process. Others consider diagnosis more of a process in itself. Process-oriented definitions of diagnosis exist. For instance, diagnosis can be defined as the act or process of deciding the nature of a diseased condition by examination (Webster's Third New International Dictionary). In epidemiology, diagnosis is the process that labels patients and classifies their illnesses, identifies likely prognoses and propels health care professionals toward specific treatments in the confidence that they will do more good than harm (Sackett et al.,

1991). One definition belies the confusion: diagnosis is described as the decision or opinion resulting from such examination or investigation, with the diagnostic process is how physicians arrive at a diagnosis (Groth, 1985). Gupta (1991) suggests three phases of prediagnosis, diagnosis and/or treatment as appropriate, and post diagnostic care. One additional segmentation of the process is to delineate diagnosis from therapy. The general problem of diagnosis is to classify an object, event or situation on the basis of uncertain information (Solomon, 1985). Therapy does not automatically flow from diagnosis - in a sense, it is a separate process. In this view, diagnostic "success" is when the diagnosis leads to an appropriate therapy that benefits the patient.

Extending the Model

We propose a working model that incorporates these views into a general process of health care delivery decomposed into three sequential processes: prediagnosis (ending in diagnosis), treatment and post diagnostic care. For each of these phases there are inputs, a conversion process, and outputs. These processes overlap in the sense that the outputs for one become the inputs for the next. This view is similar to that of a value chain with each phase having a supplier, a processor and a customer where the customer for one phase is the supplier for the next phase. A diagram of a sequential transformation model is given in Figure 2.3. This model provides the elements of the previously discussed descriptions that are intuitively pleasing while also providing a structure that can be analyzed.

Figure 2.3. Sequential Process Model of Health Care Delivery



The working, phased model should provide several advantages in attempts to assess health care processes. First, it provides a phased structure consistent with descriptions of the health care process while providing analyzable units of the process. This model is expected to be generalizable to both chronic and acute health care processes, as the phases are applicable to both (only the time period may change). Additionally, this model provides a unique structure in which a systems view can prevail and yet specific steps can be investigated. This decomposition into a phased process allows for the holistic analysis of the effects of specific quality management techniques and medical interventions that may be applicable in only one phase. Finally, it also creates an opportunity to use large population datasets for analysis.

Process and Outcomes Management in Health Care

A foundation in health care processes is needed to build a model of outcomes and process management. Medicine can be viewed as playing a role in the production of health as opposed to merely providing health services. This foundation begins with a definition of health care processes that acknowledges the disease-oriented system of medical research and education. The second component of this foundation is a transformation model of production. In this section we detail how this model applies to health care and expand it to build a structure for analysis, a model for assessing outcomes as affected by process management.

Health care processes have long been defined to reflect the disease treatment system that prevails in medical research and education (Senge and Asay, 1988). Such

definitions describe health care processes as encompassing the period within which the patient presents for treatment until the outcome of the processes is realized. It is the means by which an individual or individuals enters the health care system, presenting specific symptoms and an underlying health profile, is provided with an appropriate diagnosis and/or treatment which is followed until general well-being is again attained. A broader definition can be realized by incorporating a production view of health care into our understanding of the health care process. As Berwick et al., note:

> "In health care, as in any complex production system, much of the work can be framed as processes, each with its own set of customers and suppliers: patient flow processes that move people from place to place, information flow processes that create and transport the grist for informed decisions, and material flow processes that move equipment and supplies. Quality management begins with seeing the world of production in process terms." (1990, p. 34).

Expanding the model to represent the health care delivery process as a sequence of conversions provides a structure to examine health care clinical processes more specifically. The sequential model in the sequential order allows for easier identification of the variation in inputs, processes and outputs. In particular, the emphasis on the need for research in health care processes is a result of the historical context of health care delivery. Prior to the introduction of prospective payment systems in the early 1980s, physicians played the role of autonomous gatekeeper in providing health care and were not faced with financial incentives to control costs. With the onset of prospective payment, the environment changed to provide sufficient motivation to improve health care delivery systems. For the first time it was acknowledged that variability in health care delivery played a direct role in health care costs.

Several health services researchers have detailed the need for methods that can assess both processes and outcomes. This working model should provide the basis for exploring new methods that address this need. Additionally, two other comments provide impetus for this work. Lee and Estes note (1990, p. 337):

> "Because of the complexity of emerging treatment modes, physicians face uncertainty about whether the benefits of a particular treatment outweigh the risks. Because our knowledge is incomplete, we see an increased emphasis on outcomes rather than process in order to produce the capacity to better inform physicians and assist them and their patients in decision making and thus ultimately improve the quality of care."

Current research focuses on end-stage illness as treated in the specialty sector, but "upstream interventions" should be studied as well (during early stages of illness progression). A focus on end-stage is appropriate if the time frames in which health status and health care costs are most volatile are of concern, but they may be ameliorated by earlier care (Nutting, 1991). This phased model will specifically consider the possibility of intervention being done at different phases and assess the points at which intervention is appropriate or most effective.

Process or Outcomes - The Debate

In health care research, there has been continual debate as to whether processes or outcomes are more important (Nutting, 1991). Clearly, there is a link between outcomes and processes as noted in the service operations management models. A well-established model in health care research is Donabedian's notion of structure, process and outcome as the three targets of the assessment of care (Berwick et al., 1990). Structure refers to the resources

assembled to deliver care, including the credentials of doctors, characteristics of the building, and standard procedures. Process is the care itself - which medicines are used, how diagnoses are made, and which procedures are performed. The valued results of care, such as lengthening life, relieving pain, and satisfying the consumer are the outcomes. While much research has been done on structures, processes and outcomes, the approaches to measurement concentrate on the notion that outcomes are the 'ideal' targets of measurement and that structures and processes are interesting only to the extent that they are known to be connected to certain outcomes (Berwick et al., 1990).

There are a variety of reports as to what previous research has addressed, though most agree that a new direction is needed. Initial studies of quality focused on the process of care and the structure of services; now there is an emphasis on effectiveness, appropriateness, and outcomes of care (Lee and Estes, 1990). Measurement efforts have largely been concentrated on defining and measuring better quality of care. Research has shown that there are variations in the amounts of care and in the appropriateness and outcomes of care that are too large to be ignored (Brook, 1990). Research in outcomes and medical effectiveness emphasizes patient preferences for outcomes achieved in terms of improvements in health status, functional capacity, and quality of life, rather than purely physiologic measures (Nutting, 1991). Conditions of particular interest include those that are common, that involve substantial patient morbidity or dysfunction, for which the costs and/or risks of the interventions are high, and for which there is uncertainty or variation in the approach used and the outcomes achieved. There is still a need for research that examines multiple determinants of patient outcomes (Nutting, 1991).

Both outcomes and process measures are needed to complete the picture of quality of care. Work is needed to provide evidence of the link between process and outcome measures (Lohr, 1988). The work reported here is designed to do just that. In a unique manner, this investigation specifically addresses both process (medical intervention and management techniques) and outcomes (health status, utilization and costs).

Quality Management Techniques

Research in health care processes will provide a structure to assess the comparative value of different quality management initiatives as well. Current quality initiatives in health care are all methods of standardizing health care processes developed in response to the need to provide quality health care at reasonable costs. However, each may have been implemented in response to different specific objectives and as such may focus only on an isolated problem or portion of the health care delivery process. The lack of process analysis capabilities in medical systems has both contributed to the medical community's willingness to undertake standardization efforts and limited the impact of these efforts (McLaughlin, 1994).

Implications of Variation in Health Care Processes

Historically, under the autonomous physician gatekeeper system, variability was understood as a driver of the system. The fee-for-service system encouraged variability both by its lack of financial incentives for efficiency and by its grounding in individual knowledge and practice style. Even with the introduction of financial incentives to reduce variation, variability exists in the health care delivery process in a number of forms. Patients vary

widely, physician training and practice patterns vary, and measurement systems vary, and they always will. High practice variation can result in inappropriate hospitalizations, medical treatments, and surgical procedures, thus decreasing quality of care. Decreasing variation can be associated with improved patient outcomes. Thus, it is expected that system improvements can lead to measurable improvements in patient outcomes and simultaneously reduce the costs of care. Variation reduction is not the only goal in ensuring quality and cost-effective medical care (McLaughlin, 1996).

Special Cause and Common Cause Variation

Two types of variation can be identified - special cause and common cause. Special (or assignable) cause variation is variation associated with a particular identifiable source, often an error or an aberration in the system. Once isolated, this variation can be relatively easy to remove and the effects are no longer observed. The result is process improvement. Common cause variation, also referred to here as inherent, natural or residual variation, is that variation which is naturally part of the system and is not easily identified or isolated. In fact, some common cause variation will always be present in the delivery of health care. This distinction is consistent with Nelson et al.'s (1994) understanding of variation, which separates error (special cause) from an underlying construct (common cause).

There are existing methods that distinguish common cause from special cause variation. Quality methods grounded in statistics help to distinguish special causes from chronic causes or faults of the system based on the likelihood of the degree of variance (Rosander, 1989; Johnson and McLaughlin, 1994). An example of blood pressure

measurement notes that the goal is to find and eliminate special cause variation so that only natural variability is left in the readings (Nelson et al., 1994). The quality methods that exist are adequate for identifying and eliminating special cause variation. However, dealing with common cause variation is largely left up to the judgement of the service worker. For instance, by isolating natural variability from extra, nonrandom variability due to any other known or unknown effect, a health care provider can point to suspicious instances and study them for cause of variation before deciding whether to believe and accept the result (Nelson et al., 1994). This research is focused on better understanding the management implications of inherent variation with an eye to providing better guidelines and techniques than individual judgement.

Managing Variation

The variation that exists in health care processes has several implications for health care managers. In the service sector, variation in the capability of individuals often has a greater impact on quality than it does in the manufacturing sector (Lefevre, 1989). Second, there is a general expectation that reducing variation will result in significant savings, (Horn and Hopkins, 1994). One estimate is that reduction in physician practice variation could save enough to meet health care needs of Americans currently excluded from system (Wennberg, 1990). These expectations indicate that variation reduction is desirable. There may also be tactical reasons to reduce variability. Stable processes can

be easier to understand and manage, so examining variation may help identify improvements, being a means to an end (Wood, 1994).

Others suggest that reducing variation is not the only way to manage a process that has many variable components In addition, a flexible response can be developed that necessitates a reliance on diagnostic skills within the operation itself and the deliberate cultivation of highly variable processing capability (Morris and Johnston, 1987, p. 19) Variance reduction is analogous to standardization, while flexible response focuses on customization. Other evidence that reducing variation may not be an end in itself comes from analysis of service operations. Distributions in service industries tend to have optimums at one end of a distribution, rather than in the middle as is typical in manufacturing industries (Wood, 1994). Thus, the objective in services is not to reduce variation, but to improve the average value. Note though that in some processes - such as those that are Poisson distributed - reducing variability has the effect of reducing the mean. System-dominated solutions accept only limited variety. Customer-dominated solutions either adapt to individual customers or pre-select several customer groups (rejecting others). Restricting variety leaves the source of variety intact, but may result in dissatisfaction for the customer. (Schary, 1992). Morris and Johnston (1987) note that the goal is to ensure that the final outcome matches specifications. Controlling variability is necessary to meet this goal. Thus, the essence of managing variability is to acknowledge that variability and uncertainty in customer processing operations means that there may be

a need to increase, or at least maintain, variability in process capability in order to cope with the variable input (Morris and Johnston, 1987). Flexibility in processes is likely to achieve this end.

Quality Management Techniques in the Health Care Industry

One way to reduce or incorporate variability in health care delivery systems is to introduce methods of standardization and control into process management. To understand how current initiatives have developed, we first explore the roots of standardization in the history of the health care system. Given the impetus for standardization, we must also consider the associated caveats for taking standardization too far. Finally, we must characterize the standardization and control attributes of the specific methods currently used in health care process management.

Incentives for Standardization and Control

Even as early as the 1970s it was noted that efforts to standardize care such as routines or protocols were useful in enhancing completeness and in providing progress evaluation (Shoemaker, 1974). The environment at that time still supported the cost-independent role of the physician. Two factors that were particularly supportive of processes tolerant of variation were the lack of financial incentives to promote efficient resource utilization, and the resistance on the part of physicians to the apparent loss of autonomy in practice (Coffey et al., 1992). Thus, initial barriers to standardization of

health care practices were great. Four particular trends have helped pave the way for standardization (Coffey et al., 1992, p. 46). In the early 1980s, there were changes in the reimbursement systems, including prospective payment and competitive bidding, that provided financial incentives for standardization. Second, there was mounting evidence that unacceptable variation in clinical care and associated outcomes existed (Wennberg, 1977; Eisenberg, 1986). Third, more and more health care decisions were made by multiple providers, with more frequent requests for consultations with specialists and sub-specialists. Finally, there were added costs of malpractice that infringed upon the physicians ability to contain costs in their practice, costs which might be reduced by clearer definitions of community standards of care. Thus, the environment changed such that there was sufficient motivation for standardization efforts to be beneficial to all parties involved in health care delivery.

Over the past few years, initiatives geared towards process standardization and control have been pursued by nearly every health care providing organization. Government agencies have been investigating protocols for many procedures in an effort to move towards standardized procedures. This change in perspective from the physician as artisan acknowledges that many of the processes that physicians follow are scientifically understood and are repeated and that in these situations standardization is possible.

Health Care Applications

Standardization efforts in health care were initially borrowed from the manufacturing industries. Manufacturing processes, with fewer sources of inherent variation, realized vast improvements from implementing standardization processes that eliminated special cause variation. As standardization efforts progressed, manufacturing has benefited from both increased quality and controlled costs. The subsequent offsetting trend in manufacturing, however, was a demand for increased product variety and specialization. Boynton, Victor and Pine have suggested that industries will move from standardization to dynamic stability (1993). In their view, a dynamic product will be produced by a process that is undergoing continuous improvement, which will eventually move towards a stable process capable of producing high product variability at low cost. Given the current trends towards standardization, inherent variation in the inputs and processes which would require customization efforts can be overlooked. Thus, the opportunity exists to further define where the tradeoffs exist, and identify where there is a way to balance the need for standardization (especially on the aggregate level) with customization (especially on the individual level).

A wide variety of methods exist to attempt to develop and standardize health care processes. Included in these standardization initiatives are outcomes research, total quality management (also known as continuous quality improvement), clinical protocols, case management, critical path management and consensus conferences. The differences in

these methods is often a matter of perspective or degree. Some are focused specifically on eliminating variability and others are more oriented to increasing quality or decreasing costs. In reviewing these methods, we pay particular attention to the focus the particular method has toward variation reduction and the anticipated results of implementing such an initiative.

Process control

While the efforts detailed above focus on the organization and delivery of care, there are other initiatives that take the process of delivering care as a given and instead focus on reducing variation. These initiatives directly acknowledge the relationship between variation and the cost and quality of health care. The most notable of these efforts is Total Quality Management, often adopted as continuous quality improvement in health care settings. Other efforts include quality assurance, clinical practice improvement, outcomes measurement and PORT studies.

The initial movement towards quality in health care began with quality assurance (QA). QA is essentially a comparison of rates of negative occurrences among health providers (Everett, 1993). Thus, the QA system must ascertain the number of occurrences for each provider and the number of occurrences overall over a defined period of time. QA is often implemented through a system of chart review and focused on determining whether a breach of appropriate medical practice by an individual practitioner has occurred (Kritchevsky and Simmons, 1991). These systems assume such situations to involve

special cause variation (get out the "bad apples") and have not been intended to produce system improvements. This reflects the prevailing mentality in health care that problems are the results of individual error rather than a system (Johnson and McLaughlin, 1990). Typically, quality assurance committees are formed to review specific medical cases and assess what may have gone wrong. This post-hoc view was more academic than applied and thus unsuitable for implementation directly in daily procedures (Berwick et al., 1990, p. 11). It focused only on variation in the system that was evident in the final post-diagnostic phase, and thus had limited ability to effectively control variation. QA uses an individual incident view rather than a process view. The result is that quality tends to be assessed after the fact, in a reactive manner (Melan, 1989).

Total Quality Management (TQM) has its roots in the improvement of manufacturing processes, especially in Japan in the 1940s and 1950s. Pioneers of quality improvement efforts included Walter Shewart, W. Edwards Deming, Joseph Juran, and Philip Crosby. In health care settings, these efforts are often referred to as Continuous Quality Improvement (CQI). TQM can be segmented into both a philosophy and tools used to reduce variation. The philosophy behind total quality management is simply that of reducing unnecessary process variation. Often this is done by identifying and isolating special cause variation so that only common cause variation remains, whereby the process is deemed "in control". A number of statistically-based tools are used to assess variation, including brainstorming, Ishikawa (fishbone) diagrams, checksheets, and Pareto charts

(McLaughlin and Kaluzny, 1994). One tool often used in conjunction with TQM or CQI is statistical process control. This allows companies to identify and reduce both special cause and common cause variation. Under the statistical process control driven quality efforts, quality is synonymous with low variation (Birkland, 1991).

In health care, the National Demonstration Project provided evidence that the techniques used in TQM were transferable to the health care process (Berwick et al., 1990). Health care organizations across the nation have implemented total quality management or continuous quality improvement. If properly implemented, the results of these efforts are said to improve efficiency, productiveness and quality (Bluth et al., 1992). Common results reported are reduction of delays, financial savings, and increased patient satisfaction (Berwick et al., 1990; Bluth et al., 1992). It is pervasive in today's health system because it has become a requirement of the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) (McLaughlin, 1994). Thus, TQM/CQI is often used in conjunction with a number of other initiatives.

A modified use of TQM that focuses on process analysis and the use of statistical methods is Clinical Practice Improvement (CPI), developed at Intermountain Health Care, Inc. in Salt Lake City, Utah. This initiative focuses on controlling for patient factors (disease, severity of disease and risk), while using tools to improve and standardize process factors. Multiple regression is used to determine the best outcome. Outcomes are measured at the end of the process in terms of cost/LOS, clinical, satisfaction, access, and

health status. CPI specifically measures process variation, then eliminates it through a combination of consensus and feedback.

Other efforts are centered on reducing the variation in patient outcomes, and are grouped together as outcomes management. On the national level, the AHCPR is involved in outcomes management. In addition to the efforts to develop practice guidelines, the AHCPR is specifically charged with measuring the outcomes of common health care interventions in the belief that this will also help eliminate inappropriate medical interventions and control health care costs (Horn and Hopkins, 1994, p. 128). Clinical outcomes are a widely used measure of quality of care (Raffel and Raffel, 1994).

One specific example of efforts focused on outcomes management is the Patient Outcome Assessment Research Program. Since 1986, through the Omnibus Reconciliation Act of 1986, the program has funded studies for the treatment of a number of health care interventions. Additionally, they have supported methodological research on developing better outcomes measures, meta-analysis, small-area variation, and market-area determination (Salive et al., 1990). Patient Outcomes Research Teams (PORTs), charged with studying medical treatment effectiveness, were formed to study acute myocardial infarction, benign prostatic hyperplasia and locally invasive prostatic carcinoma, coronary artery bypass surgery and low back pain. These teams identify and analyze the outcomes of alternative practice patterns and develop and test methods to reduce inappropriate (or special cause) variation (Salive et al., 1990, p. 698).

Process design

A number of process standardization initiatives are focused on the organization of care giving. As a result, they decrease the variation in the process of giving care. These efforts are often lumped under the blanket term protocol. A protocol is a detailed blueprint of a particular treatment plan. They are often designed and used in conjunction with medical research studies. A protocol details agreed-upon procedures at various points during the care process (AMA, 1995). A more specific use of protocols can be implemented on computers, and incorporates logic that helps health care providers make decisions and alerts them to potential problems in the care plan. From the physicians point of view, using protocols can be a protective measure in the face of an increasingly antagonistic legal climate. A physician adhering to a standard protocol that is well documented and generally accepted is better equipped to refute charges of medical negligence (Raffel and Raffel, 1994). Protocols can include critical path management, case management and practice guidelines.

Critical path management (CPM) is a tool that theoretically gives health care providers a centralized plan for day-to-day care (Healey et al., 1994). It provides an optimal sequencing and timing of interventions by health care providers and encompasses diagnostics, treatments, education and discharge planning for the individual patient, based on the particular diagnosis. Critical paths are best implemented when the course of disease in the patient follows a predictable pattern. In critical path management, variation is

monitored on the individual level and it is essential to identify deviations from the predictable pattern quickly in order to correct the variation. On the aggregate level, critical path management also allows for the assessment of variation by diagnosis. Critical path management is best used for high-volume, fairly predictable but complex procedures. The use of critical path management ideally will result in fewer delays, minimal use of resources, and maximum quality of care (Coffey et al., 1992). In health care, critical paths are also referred to as clinical paths or pathways.

Another similar initiative commonly used is case management, which organizes care specifically around the patient, and assigns resources specifically to meet the individual patients needs (rather than an aggregate pool of health providers meeting the needs of a group of patients). It may also be referred to as patient-centered care. Typically, case management is focused on decreasing the length of stay without compromising the quality of care. A case manager is responsible for tracking patients' stays, but case management also enables patients be involved in directing their own care. The goal of case management is to provide high-quality, cost-effective health care in an efficient manner by providing a coordinating mechanism, the case manager, to integrate the resources available (Horn and Hopkins, 1994, p. 161-2).

All of the above initiatives are developed on an institution-specific basis. Other initiatives are being enacted on a nationwide basis. A variety of agencies in the United States are involved in supporting the development of practice guidelines. There has been a

strong push in the last few years toward the formal creation, dissemination, and enforcement of explicit clinical practice guidelines or parameters relating to the quality and efficacy of various medical interventions, parameters that would guide the decisions and actions of physicians and other health care providers (Kapp, 1993). Medical societies, governmental agencies, and insurers are using a number of approaches to developing these guidelines (informal consensus development, formal consensus development, evidence-based development, and explicit guideline development). The number and variety of practice parameters has burgeoned in response to the wide national variations in medical practice patterns that have been documented by health services researchers, without corresponding differences in clinical outcomes.

The Agency for Health Care Policy and Research (AHCPR) has supported the development of more than fifteen practice guidelines, ranging from acute pain management to post stroke rehabilitation. They define practice guidelines as "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances," (AHCPR, 1993, p.1). This definition is supported by the Institute of Medicine (IOM) (IOM, 1990, Field and Lohr, 1992). While there are acknowledged definitions for practice guidelines, common usage is less unanimous, and can refer to protocols but also encompass practice standards, algorithms, and many other statements regarding appropriate clinical care (Field and Lohr, 1992, p. 26). Three different levels of practice guidelines are acknowledged: standards, recommendations, and

options. While these practice guidelines may be developed in conjunction with national agencies, they are guides for health care provider decisions regarding individual patients, and as such there are innumerable interpretations and uses of these guidelines. At the same time, some practice standards are created without knowledge of their impact on health and economic outcomes and without knowledge of how people would compare the benefit and harm (Eddy and Billings, 1988). Thus, practice guidelines may result in more common physician practices, but are also limited in the ability to assess the entire situation.

Similar nationwide initiatives are encompassed by consensus conferences. Consensus conferences bring experts in a specific medical specialty together with epidemiologists, statisticians and health care researchers in order to develop consensus on the best method to treat specific cases. These efforts are steeped in the concrete knowledge of the disease process, severity, and patient characteristics, and become part synthesis and part group decision-making (Banta and Luce, 1993, p. 72). The logic here is that participation of several experts will both increase the quality of the protocol developed and encourage physician participation that may be lacking with practice guidelines. The goal is to assess both efficacy and safety in the delivery of health care, but they focus on the physician's role and not necessarily on the whole process of care (Banta and Luce, 1993).

In the United States, consensus development is coordinated by the National Institutes of Health (NIH). The NIH has a formal Consensus Development Program. Its stated purpose is "to publicly evaluate scientific information concerning biomedical

technologies and arrive at consensus statements that will be useful to health care providers and the public at large and that will serve as contributions to scientific thinking about the technologies under consideration," (Office of Medical Applications of Research, n.d.). The expectation is that consensus statements will facilitate communications and encourage practice conformity. The Consensus Development Program aims to reach practicing physicians with timely and useful information, to eliminate gaps in clinicians' knowledge, to facilitate informed decision making by physicians and patients regarding complex medical therapies, to foster appropriate changes in medical practices, and ultimately to improve patients' health care outcomes (Kanouse et al., 1989, p. xxviii).

Three styles of consensus statements exist: didactic, scholarly, and discursive. These differ based largely on their audiences, which are clinical, research and more general respectively. (Kanouse et al., 1989). The topics to be studied by consensus conferences involve (1) high priority of the NIH and the sponsoring bureau, institute or division, (2) controversies in clinical care or where there is a disparity between medical practice and research knowledge, (3) medically important issues of interest to practicing physicians, and/or (4) resolvable questions through data evaluation from existing research. Other factors include public health importance, timing, potential for prevention, effect on health care costs, and the public interest (Kanouse et al., 1989). The RAND Corporation was retained to assess the consensus development program and did the following: content analysis of 24 consensus statements, analysis of professional and popular literature,

national physician survey, medical record review/hospital survey, and Washington State physician survey. The aim was to assess the degree of dissemination of information and the degree of compliance.

Research Methods

One major focus in this investigation is to develop methods appropriate for large scale data use in the service industries. As such, this work is exploratory in nature. To introduce the methods that are explored, we briefly introduce some of the problems inherent in service operations research, discuss simulation methods, and then focus on the health care industry and some methods that are commonly used in health services research.

Service Operations Management

Service industries are characterized by the intangible nature of the services offered, perishability, a high degree of customer contact and participation, and labor intensiveness. These characteristics present unique problems to researchers interested in service industries. In particular, while many methods have been adapted from the production operations research arena, there are often problems as these methods do not explicitly take into account the unique nature of services. In more recent years, some methods have been developed which are tailored to the problems found in the service industries. Two examples are data envelopment analysis (Charnes et al., 1994) and site selection analysis (Moutinho and Paton, 1991). These methods are focused on decision making in the service sector. At the same time, there is still a need to develop methods specific to the service industries. One area of need is in process analysis.

Process Analysis

In the operations management arena, a great deal of attention has been paid to process analysis. Most recently, the tools used are considered in conjunction with quality management efforts. Included are sampling, benchmarking, use of control charts, flow process diagrams, and process capability studies. As with many tools, these have been adapted to service industries in general and to health care in particular (Johnson and McLaughlin, 1994). However, due to the uniqueness of service industries, the use of these tools often requires some modification or general assumptions to account for the intangibility, perishability and customer contact inherent in service delivery. Thus, service organizations in general and health care organizations in particular need to improve their ability to analyze processes (McLaughlin, 1994).

Simulation

In a simple sense, simulation is merely the use of a model to mimic, or simulate, reality in order to evaluate ideas. Many varieties and details of models are found. Over the years, simulations models have become very elaborate and complex. One example would be simulators that are used for astronauts to get accustomed to space travel. In the field of operations management, simulation has been shown to have value as a planning tool, particularly in design and development of new systems. In services, however, it has often been used for more limited problems, particularly focused on simple operating systems and queuing problems. It could be used more broadly in this arena, to ensure the productive use of resources and assure quality of service (Smith, 1994).

There are a number of reasons that simulation is used in business situations. Compared to the alternative of real experimentation, simulation can be less costly, less time consuming, safer, and able to be replicated (due to the previous three reasons) (Pidd, 1992). Compared to the alternative of deterministic mathematical models, simulation allows for a dynamic and probabilistic effects. Social scientists turn to simulation for a number of reasons as well. Simulation can represent a diverse array of phenomena, account for a large array of variables, provides efficiency, and can expand the ability to analyze problems even in the absence of empirical data (Whicker and Sigelman, 1991). For many of these reasons, simulation is a reasonable approach to this problem. In health care settings, often randomized controlled trials are used - but they are very expensive and time consuming. Also, while we have a great deal of data, simulation allows us to expand that data in order to develop insights based on a large population of actual patients. Yet, the amount of information encompassed in the data is beyond the limits of mathematical modeling. Additionally, though not mentioned in the reasons above, experimentation can be considered unethical in health care settings and it is useful to turn to simulation as an alternative method.

The type of simulation used in this study can be classified as Monte Carlo simulation. Monte Carlo simulation employs random numbers to generate a range of outcomes (Smith, 1994), and is not based on the passage of time as are continuous and discrete event simulation. The details of the Monte Carlo simulation process are given in Chapter Four, under the section on random selection and replacement. The Monte Carlo method is one of the statistical bases of bootstrapping, a specific simulation technique.

Bootstrapping

The bootstrap is a recently developed computer-based technique of statistical inference that can answer many real statistical questions without formulas by using real data sets to construct confidence intervals and other sample statistics (Efron and Tibshirani, 1993). The origin of this phrase is interesting:

> The use of the term bootstrap derives from the phrase *to pull oneself up by one 's bootstrap*, widely thought to be based on one of the eighteenth century Adventures of Baron Munchausen, by Rudolph Erich Raspe. (The Baron had fallen to the bottom of a deep lake. Just when it looked like all was lost, he thought to pick himself up by his own bootstraps.) (Efron and Tibshirani, 1993, p. 5.)

The bootstrap method involves multiple samples from the same population in order to calculate estimates of population statistics. Thus, this method is a mixture of two techniques: the substitution principle and the numerical approximation (Shao and Tu, 1995). Its application is usually carried out with resampling, though data-resampling is not absolutely necessary. It is intended to mimic sampling behavior, and thus simplify many complex statistical calculations that get more cumbersome with large sample sizes by using many smaller samples as an estimate of the sampling distribution (Shao and Tu, 1995). It has become more useful especially as sampling has become quite simple with the advent of advanced computer technology. Because of its use of random selection, bootstrap sampling is a Monte Carlo integration method (Efron and Tibshirani, 1993). Bootstrap applications are widespread, and include sample surveys, linear models such as regression analysis, nonlinear, nonparametric, multivariate models, and empirical simulation (Shao and Tu, 1995). The

method that we propose in this investigation has some of the same principles as the bootstrap, especially that it makes use of resampling from empirical data in order to calculate statistics.

Applications, especially in Health Care

Simulation, including bootstrapping, is used in many settings. As computers have become more advanced, the application base has widened. Focusing on the health care field, a number of studies have been done that involve simulation. Many of these are applications of the operations research and management research to a health care setting - but not to clinical processes. Specifically, bootstrap simulation has been used to adjust for bias in cost-effectiveness ratios (Stinnett, 1996), in models of carcinogenic risk (Bailer and Smith, 1994), and to validate pharmacodynamic models (Mick and Ratian, 1994). One researcher has combined bootstrap simulation in health care with standardization methods described below (Swift, 1994). In comparing methods, Feng et al. (1996) found that a bootstrap method that resamples clusters was superior to three other estimation procedures if analysts do not want to impose strong distribution and covariance structure assumptions.

Health Care

In clinical medicine, the gold standard of research is considered to be randomized controlled trials. However, it is difficult and often unethical to conduct such research as it would require either unnecessary procedures or the withholding of treatment that is known to be beneficial. The Congressional Office of Technology Assessment has estimated that only 10-20 percent of practices are supported by randomized controlled trials (Eddy and

Billings, 1988). A wide variety of research practices have applied additional methods to the health care arena. One result of this has been an increase in the breadth and depth of research in public health. Two areas of developing health care research - epidemiology and health services research - provide notable contributions relevant to this investigation.

Epidemiology

Epidemiology, the study of the occurrence of illness, is a relatively new field of science, with one author describing its stage of development as an "embryo" (Rothman, 1986). Because of this, the methods used are still in the development stages. This has both its positive and negative aspects. It is helpful as researchers are creative in trying to answer the questions that they have posed; it is difficult as there are often no set standards, definitions or methods to which to turn. In the past twenty-five years, studies have been valuable both for their substantive findings and for their demonstration of the efficacy and efficiency of methods such as the case-control study and multiple logistic regression analysis (Rothman, 1986).

To study the occurrence of illnesses, epidemiologists are concerned with both incidence and prevalence of disease. Incidence describes a frequency of events (or incidents) over a period of time and may be expressed either as a rate or as a cumulative measure of conversion to disease. Prevalence is what exists or prevails; it measures the proportion of individuals having a disease at a specific instant (MacMahon and

Trichopoulos, 1996; Rothman, 1986). When using these measures in investigations, researchers are concerned with how to compare them as the basis from which they are calculated may differ and may be affected by other characteristics of the population that need to be accounted for, such as age, sex, or smoking history. The rate measured in a study is referred to as a crude rate, and it disregards any demographic subgroups. Another way to report these measures is as a specific rate, which limits the rate by the characteristic of the subgroup from which it is calculated. Finally, an alternative to the specific rate is to report standardized rates. Standardization (or adjustment) is used to reduce distortion so that a comparison of rates may be made, especially between populations (Rothman, 1986). For instance, as one text explains, "age is so strongly associated with risk of almost all diseases that two populations that have different age distributions will almost invariably differ in their disease experience, even if their age-specific rates are the same," (MacMahon and Trichopoulos, 1996). Common characteristics for which standardization is used are age and mortality rates; additionally morbidity rates and other characteristics such as race, sex and smoking may also be considered (Kahn and Sempos, 1989). The standardization will vary depending on the method used, but in general it is a method by which to take a weighted average of many categories that make up a sample population and compare it either to another sample or to the entire population.

Standardization methods are referred to as direct or indirect standardization. Either method uses a standard population as a basis of comparison. Direct standardization takes

categorical rates from the sample and applies them to the standard population in order to calculate a composite rate. Indirect standardization derives the rate from the standard population (by category) and then permutes the categorical rates for the sample. While direct standardization is preferable, it is limited by the need for an adequate population to base the standardization on, which is essentially the problem which presents the need for standardization at all (MacMahon and Trichopoulos, 1996). All epidemiology texts reviewed covered discussed the details of direct and indirect standardization (Selvin, 1991; MacMahon and Trichopoulos, 1996; Kahn and Sempos, 1989) and refer to the same specific source articles that propose the measures and detail their use, particularly with indirect standardization (Wolfenden, 1923; Axelson, 1978; Axelson and Steenland, 1988; Steenland et al., 1983). In addition, many researchers have used the indirect standardization methods so that they are becoming a standard part of the epidemiology toolkit (Brackbill et al., 1988; Gail et al., 1988; Steenland and Beaumont, 1989; Merlo et al., 1991; Guyuron and Zarandy, 1994; Merlo et al., 1995). While the potential uses for these measures is great, the studies that have shown them the most useful have to do with adjusting for smoking and adjusting for any number of factors, including exposure to specific substances, in occupational health.

While the indirect adjustment or standardization methods are becoming more commonplace, they are not the entire answer for being able to compare groups. This adjustment cannot completely remove the influence of differences in population

composition (Selvin, 1991). In particular, it only addresses the characteristics by which the adjustment is made (usually one demographic characteristic). Thus, in the same manner that randomized controlled trials can be limiting, so is this method of data comparison. It may be difficult to identify and appropriate standard population. Also, the indirect standardization method is limited in that the only valid comparisons that can be made are between the sample and the standard population. Though some researchers have compared different samples using this method, it has been noted that this is not statistically valid (MacMahon and Trichopoulos, 1996). For example, you would not compare an apple to an orange and a lemon to an orange, and then claim that you have compared the apple to the lemon.

Thus, it would be worthwhile to have methods which would incorporate a wide variety of demographic and health characteristics in order to more accurately reflect the variation underlying an entire population. Yet, synthesizing indirect evidence is extremely complicated. It is neither a surprise nor an insult to say that even when good information is available on all the important factors, putting them together can easily exceed the capacity of the unaided human mind. Mathematical models can be used to structure the existing knowledge and perform the calculations, but the overwhelming majority of problems are analyzed without such models (Eddy and Billings, 1988, p. 25). For example, mortality is often investigated using survival analysis. To use any mathematical model, of course, the model must first be specified - presenting its own set of challenges

(Selvin, 1991). This investigation proposes a experimental methodology which would eliminate the need for this adjustment by controlling for certain categories of variables while allowing the underlying population characteristics to be embraced by the results. A further description of this is found in the research methods chapter under the section regarding random replacement in the simulation.

Health Services Research

Partly in response to the need to address issues regarding increasing health expenditures, health services research has emerged. Health services research is an applied research field encompassing methodological, descriptive, analytical, and experimental research methods. Typically there are two levels of questions addressed, the doctor-patient level and the policy level. The stream of health services research, spurred by concerns with cost containment, originally focused on expenditures but has evolved to focus on the development of analytical tools, quality and outcomes measurement, and process analysis.

Process analysis efforts are not unique to health services research, but they have emerged from systems analysis in a different light. In the 1970s, process analysis in health care settings was grounded in systems science. At that time, systems science in health care encompassed accuracy of decisions, efficiency in resource allocation, policy development, the use of technology in health care, and the flow of information (Sheldon et al., 1970; Flagle, 1977). Queuing analysis and capacity analysis were also commonly done. Much of this work stemmed from the overlap of the domains of health care and operations research. Efforts to decrease costs often resulted in capacity analyses and associated decreases in capacity.

Typically, capacity changes have been a means to deal with issues of variation in health care. Even though these methods are used, however, there is still a lack of process analysis capabilities that address the current needs (McLaughlin, 1994).

More current practices include a hodgepodge of methods and comparisons are difficult when inconsistent data and outcome measures are used (Samuelson, 1995). Analytical tools now commonly used in the health services research arena include assessment of care, disease severity, outcomes of care, statistical adjustment for disease severity, health status measures, decision analysis, and meta-analysis. These methods are designed to develop, use and publicly release information about practice standards. Health status measures can be used to assess whether changes in health policy can affect quality of life and to determine whether clinical interventions work. For instance, for treatment of hypertension one can measure the effect of therapy on lowering blood pressure (increasing life expectancy) and on quality of life (Brook, 1990).

There is also an associated need for a better understanding of how health care delivery processes respond to current and potential quality management techniques. With respect to quality management, the research is in the early stages of developing theories and defining examples. There are many examples of quality improvement techniques being applied to health care settings and a wide variety of research reports directives for how to implement quality programs. There is a gap, however, in understanding how the widespread use of these quality measures will impact the health care system. The research reported here begins to fill that gap by incorporating the impact of quality improvement on a large scale basis by using population data.

Problem Statement

> You would be surprised at the number of years it took me to see clearly what some of the problems were which had to be solved... Looking back, I think it was more difficult to see what the problems were than to solve them.
>
> Charles Darwin

Opportunities Presented in Literature

It has been noted that rather than continuing to adapt production operations research to service operations management it is time for service industries to break out on their own and pursue research methods that are tailored to the issues unique to services. From this review of the literature illustrated in Figure 2.1, we can see that a wide variety of research opportunities exist, particularly if we choose the specific case of health care from within the service industries. First, we considered the use of a process view of health care delivery. Specifically, health care organizations need to have a more process-oriented focus and improve their ability to analyze these processes (McLaughlin, 1994). Our postulated sequential process model provides an enhanced framework to consider health care clinical processes as it provides a broader scope for analysis and yet provides for the modeling of portions of the process. Also, we look at quality management techniques, which have been successfully applied to health care processes in order to identify and reduce special cause variation, particularly in non-clinical processes. However, the field has not addressed as intensely methods for identifying and managing common cause variation. Additionally, the effect of current efforts at quality management in health care can be assessed for large populations.

We combine these areas ripe for research with the potential and desire to have more appropriate methods for their analysis. We use large population data sets, so it is important to develop methods that can handle large amounts of data with a high degree of variation. Having considered methods used in relevant fields, we conclude that it is worthwhile to explore simulation techniques with these larger population data sets. Thus, the methods proposed in this investigation are designed to provide exploratory insights into the impact of quality management techniques over a large scale population. Taking all these streams of research into consideration together, we arrive at the following problem statement:

> PROBLEM STATEMENT: Can we develop new research methodologies that incorporate an integrated view of service processes and performance and allow us to investigate the effects of quality management techniques on large population data sets?

Research Objectives

As a result of the opportunities identified in the literature, this research focuses on the effectiveness of current quality management techniques used in health care. Accordingly, we have identified three research objectives that are addressed by this investigation. Along with each objective, appropriate research questions and propositions are stated.

Objective 1: Explore whether or not we are able to use large population data sets for simulation modeling in service operations management, using health care as a specific case for study.

Research Question 1: How can we design simulation models to adequately investigate processes in service industries using simulation of large populations?

Proposition 1.1: Using a **process management view** leads to better analysis of service outcomes.

Proposition 1.2: Using **random data selection and replacement** allows us to investigate the large-scale population effects of specific interventions on a process.

Objective 2: Model quality management techniques in service operations management in order to demonstrate the impact of these techniques on a large population and provide insight for policy formation, again using health care as the example.

Research Question 2: How does the degree of implementation of the quality improvement techniques affect the amount of improvement in efficiency and the outcomes of the service operations process?

Proposition 2.1: Using **linearly progressive quality improvement targets** enables us to determine whether or not the effects of the quality management techniques are linear.

Proposition 2.2: Using these exploratory methods provides insight as to the **comparable impact of different quality management techniques** (consensus conference, CQI, and clinical pathways) used on the same process.

Objective 3: Demonstrate that the methods developed also lend insight to the health services research arena, both in terms of methods and in terms of the specific results.

Research Question 3: When quality management techniques are used during a health care process what is their impact on specific outcomes, including health outcomes, utilization, and costs?

Proposition 3: Using a process model enables us to link health care processes and outcomes so that greater insights are gained.

The first two sets of objectives, research questions and propositions have been clearly developed. The third set has been developed, but the details regarding the health care process and outcomes measures follows in Chapter Three. In Chapter Six, the discussion of the results is presented in a structure that mirrors these three objectives.

CHAPTER 3

A CONTEXT FOR INVESTIGATION: PREMATURE LABOR AND DELIVERY

> Every baby born into the world is a finer one than the last.
>
> Charles Dickens

To provide a specific context for this analysis, one health care clinical process has been selected - premature labor and delivery. This chapter introduces the clinical process and reviews the issues presented by the literature in terms of this process. This chapter highlights details important to the modeling of the process and the quality management techniques though the specifics of the model design are contained in the subsequent chapter. First, we define premature labor and review the medical literature on the potential impact it has on the health of the infant. Then, the process is described in detail, using the phased process presented in the model from Chapter Two. The description of three commonly-used medical interventions are included in this discussion of the process. Finally, quality management techniques that have been used with this process and the realized or anticipated effects of each technique are reviewed.

The health care process investigated is preterm labor and delivery. This process and associated medical interventions are chosen because the end point is common (discharge of the infant from the hospital) and there exists consensus as to what constitutes a positive outcome (a healthy baby and a healthy mother). The three medical interventions to be considered all concern pregnancies during which early labor is a factor, and each of these

interventions could be used. These interventions are also chosen because there is a great deal of knowledge in the medical field as to the important conditions, risk factors, and results of each of these therapies. Additionally, examples of quality management techniques used in conjunction with each of these medical interventions can be found. Thus, the simulation model is built on information that exists, ideally from meta-analyses, or that can be readily gathered.

Premature Labor and Delivery and Potential Health Consequences

Our working model will be that of premature labor and delivery, encompassing the time from when the mother experiences premature labor until the infant is discharged from the hospital. Labor is considered to be premature when it occurs before 37 completed weeks of gestation (259 days). Disregarding congenital malformations, nearly 85% of neonatal mortality is associated with premature labor. Prematurity is largely recognized as the leading cause of perinatal morbidity and mortality (Ventura et al., 1994b). U.S. Health Statistics taken from birth certificate information indicate that the proportion of infants born at under 37 weeks gestation has risen steadily since 1981 (9.4%), except for a slight decline in 1984 and 1992 (Ventura et al., 1994b). In 1991, the reported national rate was 10.8%; in 1992 it was 10.7%. Yet, some researchers report the incidence of premature labor has remained relatively constant at about 7-9% of all births (Jarvis, 1994; Turnbull, 1989). The rate of mortality associated with premature delivery is reported widely as "up to 25%" (Jarvis, 1994) to "as many as 85%" (Turnbull, 1989), with the difference being the reference group (perinatal

deaths vs. neonatal deaths in structurally normal infants). Additionally, the single largest cause of morbidity and mortality in the neonatal period (the first seven days of life) associated with prematurity is respiratory distress syndrome - RDS (Pachulski, 1990). The incidence of RDS as reported in the literature is about 10% of all premature infants (Pomerance and Richardson, 1993). The incidence rates vary both with gestational age and with birth weight. Hansen and Corbet (1991) reported 29% of infants born at 29 weeks gestation with RDS. Robertson et al.(1992) report an extremely high incidence of 80-90% at 25-27 weeks gestation, lower incidence of 55-65% at 28-30 weeks gestation, and a drop to 30-40% at 31 to 33 weeks gestation. After 34 weeks gestation the incidence rate drops to below 15%. With respect to birth weight, low birth weight infants (under 2500g) have a higher incidence of RDS (14%) (Hansen and Corbet, 1991). Pomerance and Richardson (1993) found the incidence to be highest under 1500g and Jobe reported 70% of infants under 1000g with RDS (1993).

Clearly, if premature labor and delivery procedures can be improved with respect to enhancing lung development there would be improvements in morbidity and mortality. For years researchers have studied the effect of premature labor and delivery on infants and attempted to intervene in order to decrease the associated problems. Since the 1960s, there have been great advances in decreasing neonatal mortality, but not an associated decrease in reducing premature infant births. Thus, more premature infants are surviving, but at a large cost to the medical system of technical advances, intensive care unit facilities and highly skilled labor (McLean et al., 1993). A reduction in infant mortality does not actually save lives, if it merely postpones death (Fries, 1990, p. 38). Thus, there are large difference in

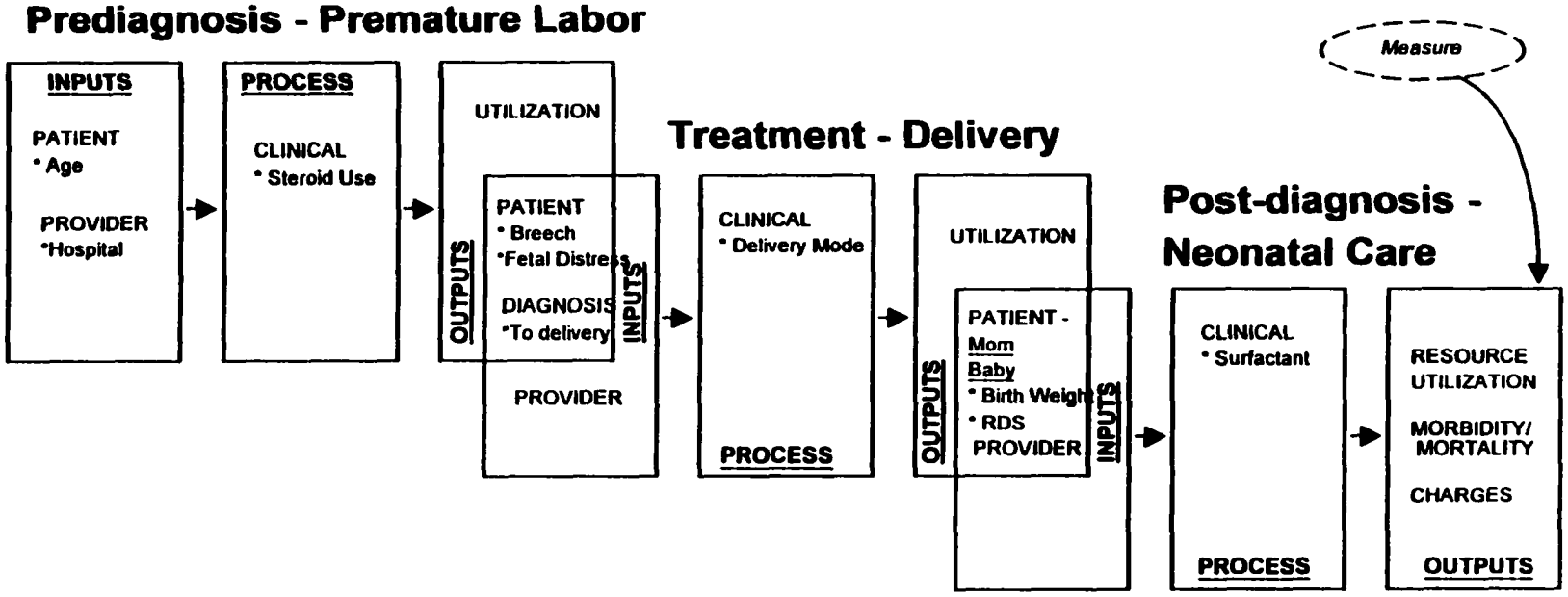
costs associated with the premature births. In one state in 1992, an infant with RDS incurred an estimated $27,901 in medical costs, compared to $7,593 for a healthy infant (Simpson and Lynch, 1994).

Sequential Process Model Applied to Premature Labor and Delivery

The three stage transformation model presented in Chapter Two is adapted to reflect the specifics of this process (Figure 3.1). Prediagnosis is considered the phase when the woman presents with labor or associated symptoms before a decision to deliver the child is made. We will refer to this phase as premature labor. The diagnosis/treatment stage is the delivery of the child. The post-diagnosis stage is postpartum care (for the mother) and the neonatal care and procedures (for the infant).

Three medical interventions that are commonly used during the process of premature labor and delivery are administration of antenatal steroids (with or without tocolitic drug therapy), caesarean section delivery, and surfactant therapy. Antenatal steroids are administered during preterm labor in an effort to speed up lung development. Caesarean section is used as a surgical method of infant delivery. Surfactant therapy is used after delivery to promote lung development in neonates. Thus, these procedures are involved in different phases of the health care delivery process.

Figure 3.1 Premature Labor and Delivery as an Example of the Sequential Process Model of Health Care Delivery



Premature Labor

While the exact process of premature labor is not understood, there are a number of factors that are clearly associated with the phenomenon. Factors strongly associated with premature labor include third-trimester bleeding, twin gestation, chorioamnionitis (infection) Less strongly but other associated factors include prior preterm delivery, vaginal bleeding in first or second trimester, maternal diethylstilbestrol (DES) exposure, uterine anomalies, and urinary tract infection during pregnancy. (Heffner et al., 1993). In general, the associated factors can be grouped as maternal factors, placental factors, fetal factors, and iatrogenic factors (Fuchs et al., 1993). Table 3.1 details the scope of factors known to be associated with preterm labor. Iatrogenic factors are associated with fetal or maternal compromise that makes the continuation of the pregnancy undesirable (McLean et al., 1993). Estimates are that these factors can contribute to anywhere from 30 to 70 percent of preterm births. Additionally, there are a number of comorbidities that often accompany premature labor such as premature rupture of the membranes (PROM), (which accounts for 20 to 50 percent of premature deliveries), preeclampsia and maternal reproductive tract infection, (which mediates preterm labor and PROM and thus increases associated risks of maternal and neonatal infections). While many things are associated with prematurity, the causes of premature labor are not well understood. Thus it is not easy to predict that premature labor will occur and there is little ability to prevent premature labor (McLean et al., 1993). Prevention of preterm labor is a contemporary goal.

The phase of preterm labor includes the time the woman presents herself for medical care through assessment of her condition and a decision as to which course to pursue. At that

decision point, there is an equivalent of a "go/no go" decision: either it is determined that delivery will ensue or that attempts will be made to arrest labor and the woman returns to the starting point. Premature labor accounts for 30 to 50 percent of all premature births (NIH, 1994).

Table 3.1. Possible Factors Attributing to Premature Labor

Factor type	Factor
Maternal	Age (below 20, second or third delivery; over 35, first delivery) Primiparity (first pregnancy) Small stature Maternal prepregnancy weight Maternal weight gain Low maternal birth weight Low socioeconomic category Race Prenatal care Cigarette smoking, cocaine use Small heart volume Congenital cardiac disease Chronic debilitating diseases Anatomic defects of the uterus (i.e. incompetent cervix) Pregnancy complications (i.e. infections, trauma) Previous preterm delivery
Placental	Placentae abruptio Placenta previa Placental insufficiency
Fetal	Multiple gestation Anencephaly Adrenal hyperplasia Anomalies associated with hydramnios
Iatrogenic	Induction of labor for pregnancy complications Preeclampsia Diabetes mellitus Elective induction of labor (i.e. breech, previous cesarean)

Adapted from Fuchs, et al., 1993, p. 85

Delivery

Delivery encompasses the time from which the decision has been made that delivery of the infant is forthcoming (with the woman in the end stages of labor) until the time that the infant is born and an initial assessment of health is made. The most widely used health assessment at this point in the health care process is an Apgar score. An Apgar score is given to a newborn infant at two points: one minute after birth and five minutes after birth. Scores range from 0 (dead) to 10 (thriving) and are based on color, breathing, presence of meconium, and other health factors as assessed by a pediatrician or trained health care provider. Staff constraints are often a factor at this phase of the process, since the number of women in labor and the pace of the labor is difficult to forecast. Premature delivery occurs in seven to ten percent of all pregnancies (NIH, 1994).

Postpartum/Neonatal Care

Postpartum care (for the mother) and neonatal care (for a low birth weight (LBW) infant) follows the delivery and continues until the infant is discharged from the hospital. This period of time can be as short as 24 hours or as long as many months depending on the condition of the infant. Infants born prior to 37 weeks of gestation often have respiratory system difficulties and other associated conditions. They are often assessed based on their weight, with weights under 2500 grams (5 ½ pounds) considered low. These premature infants are 40% more likely to die than infants born at term. They are at

risk for respiratory illness, cerebral palsy, hydrocephalus, blindness, deafness, and complications of NICU care. There is also a greater incidence of neonatal infection with PROM, and infection rates are inversely associated with gestational age (Seo et al., 1992). NICU care facilities are most expensive type of health care, greater than cancer, coronary by-pass surgery, and organ transplant (McLean et al., 1993). Yet, regardless of the expense this care is more cost-effective when measured by life years gained than many other accepted medical interventions (NIH, 1994).

Medical Interventions Used in Premature Labor and Delivery

Clearly, there is a benefit to be gained if the incidence of premature labor and delivery or the associated incidence of RDS can be reduced. To this end, medical researchers have experimented with a number of therapies designed to prolong labor (reduce prematurity) or to promote lung development in an effort to decrease the incidence/severity of RDS. One of these therapies, the use of corticosteroids to promote lung development, is a strategy used during labor but prior to delivery. During the delivery phase of the process, a second intervention, cesarean section, may be used to deliver the infant. A third medical intervention, the use of surfactant in the neonate, which mimics normal lung function until the infant's lungs are sufficiently developed, is used following delivery. These medical interventions will be considered with respect to the effects that they have on the premature labor, delivery and perinatal care phases. Each of these interventions have been in use for different lengths of

time and have been shown to have different effects, as detailed below. These medical interventions and their anticipated effects are discussed below.

Antenatal Steroids

Corticosteroids may be administered to women experiencing premature labor in an effort to promote fetal lung maturity. The benefits of steroids include improved survival, decreased incidence/severity of RDS, decreased incidence of other neonatal morbidities common to prematurity (IVH, PDA, and necrotizing enterocolitis), decreased LOS, and decreased cost (Hansen and Corbet, 1991). This treatment was first introduced in 1972 when Liggins and Howie (1972) demonstrated a statistically significant reduction in the incidence of RDS in infants of less than 32 weeks gestational age. In extensions of this initial research, they found that administration of corticosteroids reduced the incidence of RDS from 15.6% to 10.0% (p=0.02) (Howie and Liggins, 1977).

Since these initial studies, numerous other trials have been undertaken. Corticosteroid use has been studied mainly in gestations of between 24 and 34 weeks. In February, 1994, the NIH convened a Consensus Development Conference on the Effect of Corticosteroids for Fetal Maturation on Perinatal Outcomes, bringing together experts in the field of obstetrics and neonatology in an effort to review all evidence to date and make a recommendation as to the use of corticosteroids during premature labor. As part of the background research presented to the expert panel, Patricia Crowley updated a 1990 meta-analysis of 15 randomized trials (NIH, 1994). She reported an overall reduction of incidence of RDS of

about 50% (typical odds ratio = 0.51; 95% confidence interval 0.42-0.61)[1]. For women delivering 24 hours to 7 days after steroid administration, this odds ratio was improved (0.31; 95% CI 0.23-0.42). With less than 24 hours or more than 7 days, the administration of corticosteroids did not have a statistically significant effect. Other researchers have found that steroids reduce incidence and mortality from RDS if administered on two consecutive days between 24 and 34 weeks of gestation at least 36 hours before birth (Fuchs et al., 1993; NIH, 1994). Accordingly, the recommendations of the NIH Consensus Development Panel were that fetuses between 24 and 34 weeks of gestation at risk of premature delivery, regardless of fetal race or gender, should be considered for corticosteroid treatment, to consist of two (12 mg) doses of steroids 24 hours apart or 4 (6 mg) doses 12 hours apart, in order to reduce the incidence and severity of RDS. Complete consensus conference recommendations are listed in Table 3.2.

Since our data includes exact birth weight rather than exact gestational age, it is also important to consider the weight range in which corticosteroid treatment is recommended. By using standard tables relating birth weight and age, the recommendation of 24 to 34 weeks of gestation would roughly correspond to 820-2500g (Cunningham et al., 1989; Moore, 1983). Some researchers have reported specifically on the relationship between birth weight, steroid use, and incidence of RDS. Ballard et al.(1979), looking at 250g increments, found that the incidence of RDS was significantly reduced for infants between 750 and 1000g and between

---

[1] "A meta-analysis uses quantitative methods to summarize results derived from a systematic review of randomized controlled trials. Results of meta-analyses are customarily reported in terms of odds ratios and 95 percent confidence intervals (CIs). Odds ratios of 1.0 indicate no effect, those below 1.0 imply a protective effect, and those above 1.0 imply an increased risk. Ninety-five percent CIs that exclude 1.0 are considered significant at the p=0.05 level." (NIH, 1994, p. 9)

Table 3.2. NIH Consensus Conference Recommendations

- The benefits of antenatal administration of corticosteroids to fetuses at risk of preterm delivery vastly outweigh the potential risks. These benefits include no only a reduction in the risk of RDS but also a substantial reduction in mortality and IVH.
- All fetuses between 24 and 34 weeks of gestation at risk of preterm delivery should be considered candidates for antenatal treatment with corticosteroids.
- The decision to use antenatal corticosteroids should not be altered by fetal race or gender or by the availability of surfactant replacement therapy.
- Patients eligible for therapy with tocolytics should also be eligible for treatment with antenatal corticosteroids.
- Treatments consist of two doses of 12 mg of betamethasone given intramuscularly 24 hours apart or four doses of 6 mg of dexamethasone given intramuscularly 12 hours apart. Optimal benefit begins 24 hours after initiation of therapy and lasts up to 7 days.
- Because treatment with corticosteroids at less than 24 hours before delivery is still associated with significant reductions in neonatal mortality, RDS, and IVH, antenatal corticosteroids should be given unless immediate delivery is anticipated.
- In PROM at less than 30-32 weeks of gestation in the absence of clinical chorioamnionitis, antenatal corticosteroid use is recommended because of the high risk of IVH at these early gestational ages.
- In complicated pregnancies for which delivery before 34 weeks of gestation is likely, antenatal corticosteroid use is recommended unless there is evidence that corticosteroids will have an adverse effect on the mother or delivery in imminent.

From *Report of the Consensus Development Conference on the Effect of Corticosteroids for Fetal Maturation on Perinatal Outcomes*, NIH, 1994, p. 13.

1000 and 1250g. Looking at very low birth weight infants, Wright et al.(1994) found that for infants weighing 501-1500g receiving a complete course of steroids the odds ratio associated with RDS was 0.62 (95% CI 0.55-0.70); for infants receiving any steroids, regardless of the amount, the odds ratio was 0.70 (95% CI 0.64-0.78). In a similar study, Horbar (1994) found that for infants receiving any steroids, regardless of the amount, the odds ratio associated with RDS was 0.68 (95% CI 0.61-0.75).

Steroids can also significantly affect the mortality rate. The initial studies of Liggins and Howie (1972, 1977) demonstrated a statistically significant reduction of nearly 50% in the early neonatal mortality rate in infants of less than 32 weeks gestational age. Wright et al. (1994) found that for infants weighing 501-1500g receiving a complete course of steroids the odds ratio associated with death within 28 days was 0.34 (95% CI 0.28-0.42); for infants

receiving any steroids, regardless of the amount, the odds ratio was 0.37 (95% CI 0.31-0.43). Horbar (1994) found that for infants receiving any steroids, regardless of the amount, the odds ratio associated with death at 28 days was 0.53 (95% CI 0.45-0.62). These two studies investigated only infants with birth weights of 501-1,500g. Crowley (1994), looking at neonatal deaths, reported an odds ratio of 0.61 (95% CI 0.49-0.78). Crowley notes that there may be a difference in magnitude of effect based on the time that trials are conducted, as the case-fatality rate for RDS was higher prior to 1980. The NIH Consensus Development Conference report indicates that a statistically significant reduction in mortality is not found in all studies; thus, for our purposes we will consider Crowley's odds ratio of .61 (95% CI 0.49-0.78) since it represents the most conservative degree of mortality reduction and can be applied to infants of all birth weights.

Another factor in the use of corticosteroids is the relatively low prevalence of use. Although these numerous trials have shown the efficacy of steroids, they are still not widely used. In one study, reports showed that 42% of practitioners used steroids frequently, 40% only used them sometimes, and 18% never used them (Lewis et al., 1980). One study reported a rate across 20 primarily academic medical centers of 8.4% (Jobe, 1993). More recently, in conjunction with the 1994 NIH Consensus Development Conference, Crowley notes the wide variance of usage rates between hospitals and notes that only in 1993 did the Royal College of Obstetricians and Gynecologists advise members to use corticosteroids whenever there was a risk of RDS occurring. The estimated prevalence of use as of 1994 was that only 12 to 18% of women delivering preterm infants weighing 501-1500g are treated

with corticosteroids (NIH, 1994). They also noted that at some hospitals, 60% of infants were treated with corticosteroids.

The effects of corticosteroids are also seen when given to women experiencing premature rupture of the membranes (PROM). There is still a significant decrease in the incidence of RDS with steroid administration (Hansen and Corbet, 1991), though it is less than when the membranes are intact (NIH, 1994). The typical odds ratio for the incidence of RDS as affected by corticosteroid treatment reported by Crowley was 0.50 (95% CI 0.38-0.66). The associated incidence of perinatal infection is not increased by a statistically significant amount. Thus, in patients experiencing PROM, corticosteroid treatment is still recommended.

Sometimes antenatal steroids are administered in conjunction with tocolitic drugs. When a woman presents with premature labor, the major decision is whether or not to attempt to stop premature labor, based on fetal maturity, judgement of fetal condition, risks for mother and fetus of continued pregnancy, and risks of the treatment to stop labor itself (Fuchs et al., 1993). Tocolitic drugs are used to suppress preterm labor in order to give the fetus more time to develop, particularly for lungs to mature. Tocolysis means pharmacologic inhibition of uterine contractions (Fuchs et al., 1993). Tocolytic therapy as it is currently known can prevent only a small fraction of preterm births (Fuchs et al., 1993); severe perinatal morbidity and perinatal mortality do not decrease even when gestation is prolonged significantly.

Use of tocolytics has previously been used in either an expectant manner, given in an effort to prolong labor for up to 48 hours in order for a full course of steroids to be given, or in an aggressive manner, in an effort to arrest labor (Decavalas et al., 1995). In one study, only in a small fraction of cases was preterm birth prevented (Fuchs et al., 1993). In another, the findings suggested that with aggressive tocolytic therapy the risks of maternal infection and the increased cost were without demonstrated benefit (Decavalas et al., 1995). Current practices rarely subscribe to using tocolytics alone in an effort simply to prolong labor but rather suggest that they be used as short-term therapy (Fuchs et al., 1993). Specific indications for tocolytic use under these conditions include an estimated birth weight less than 2500g (or gestational age less than 35 weeks) with intact membranes (Pachulski, 1990; Fuchs et al., 1993). Although one study suggested that tocolytics could be used even with ruptured membranes (Decavalas et al., 1995), it has been noted that in general patients with PROM do not receive tocolytics due to the risk of infection (Robertson et al., 1992). This practice is supported by other researchers (Lewis et al., 1996; McLean et al., 1993). Thus, the efficacy of tocolytics is equivocal; at best, there is a lack of agreement among specialists and we do not model the use of tocolytic therapy separately .

In summary, our modeling efforts consider the use of corticosteroids in 15% of the population, and that their use decreases the incidence of RDS (OR= 0.51; 95% CI 0.42-0.61).

Cesarean Section

Cesarean section is a mode of delivery used as an alternative to vaginal delivery. It is a surgical mode of childbirth, and as such has potential intraoperative and postoperative

complications. Cesarean sections are commonly done to protect the fetus as they are thought to protect against birth trauma, though this may be true only for very large infants and for small infants presenting breech (Miller, 1988). An analogous reason to perform cesarean delivery is when vaginal delivery is considered unsafe.

Cesarean delivery can affect both the mother and the infant. It is difficult to separate procedure related deaths from underlying disease process (Miller, 1988). Maternal mortality rates are said to be anywhere from two to seven times greater than with vaginal delivery (Miller, 1988). Increased rates of infection occur, particularly in primary section rather than repeat sections (Miller, 1988). Increased endometritis, which may result from prolonged labor, prolonged rupture of the membranes, or maternal anemia (Miller, 1988) and preterm PROM (Seo et al., 1992) may also affect maternal mortality or complications . The surgeon has a large effect, since the best preventative treatment to reduce infection is operative technique. Experienced surgeons have lower complication rates (Miller, 1988) Other factors, including the type of anesthesia, the number of vaginal examinations, and the use of internal fetal monitoring are not associated with postoperative infection (Miller, 1988).

In the infant, prematurity rather than delivery mode is often the cause of adverse outcomes (Miller, 1988). Anesthesia and maternal complications also influence outcomes. RDS is common with cesarean sections, but is a function of gestational age rather than delivery mode (Miller, 1988). One estimate of fetal injury is reported as 0.4 per cent of cesarean deliveries (Miller, 1988). The timing of the delivery (with respect to gestational age) is the crucial issue. Inappropriately timed elective cesarean delivery has been known to result in RDS (Miller, 1988). 3 to 4 percent of infants delivered by elective cesarean section

developed RDS despite reliable prevention methods (Miller, 1988). When cesarean delivery is required due to uterine rupture, stillbirth and asphyxia are common due to the maternal complication. Compared to vaginal delivery, cesarean delivery has shown a lower incidence of low Apgar scores which is attributed to general anesthesia. Under conditions of fetal distress, the fetal mortality rate is associated with the time to cesarean (mortality rate doubled if time increased from 15 to 30 minutes) (Miller, 1988).

With respect to this research, two aspects of cesarean section are relevant - the prevalence of use, and the reasons for choosing this delivery method. Regarding the prevalence of cesarean sections, a 1993 report from the Center for Disease Control and Prevention (CDC, 1993) discussed cesarean section rates in the United States as of 1991. From 1965 when initial data for cesarean section rates were collected, the cesarean section rate has increased from 4.5% to 23.5% as reported by the Center for Disease control. Since 1986 the rate has peaked in 1988 at around 24%. Table 3.3 indicates the cesarean section rate over the past two decades, reported by the CDC (estimated from the number of live births) and from the vital statistics records beginning in 1989 (CDC, 1993; Ventura et al., 1993, 1994a, 1994b).

Table 3.3. Cesarean Section Rate, United States, selected years from 1965-1992.

	1965	1970	1975	1980	1985	1986	1987	1988	1989	1990	1991	1992
NHDS[1]	4.5	5.5	10.4	16.5	22.7	24.1	24.4	24.7	23.8	23.5	23.5	
Vital[2]									22.8	22.7	22.6	22.3
With PROM[2]											28.7	28.0
With Tocolysis[2]											31.5	28.9

[1] Estimated by applying cesarean rates derived from the National Hospital Discharge Survey (NHDS) to the number of live births from national vital registration data (CDC, 1993).

[2] From vital statistics records, compiled from birth certificate information beginning in 1989 (Ventura et al., 1993, 1994a, 1994b).

In recent years, a great deal of attention has also been paid to the reasons for choosing this route of delivery. While the initial rise in cesarean section rates probably reflects improvement in the technique and appropriate intervention, there are those that feel that the increase has probably been too dramatic. Berwick (1994), looking at the same data, comments that in other developed nations, the cesarean section rate has been maintained at less than 10%, and in certain settings have been as low as 1.3%. The U.S. has the third highest cesarean section rate among 21 countries studied (CDC, 1993). Even within the U.S., the rates vary widely from hospital to hospital. In one major city, though the statewide rate of cesarean delivery was commensurate to the national rate, there were hospitals with a rate of 21% and one with a rate of 17%, even with many high-risk deliveries (Walton, 1990). Combined, this information would suggest that there is not the medical need for cesarean section rates to be so high.

More recently improved obstetrical techniques, such as incision methods and fetal monitoring, have potentially decreased the need for cesarean sections. Yet, the continued high rate of cesarean sections reveals that there are factors other than obstetrical advances at work. Many possibilities have been cited, including: (1) higher reimbursement rates for "less"

work (it usually takes less time to deliver an infant by cesarean section than to let labor progress by a natural course) (Walton, 1990); (2) the ability to influence the timing of delivery (the nationwide patterns of delivery by day of the week changed with the advent of cesarean section use so that there were more deliveries during the week than on weekends; previously it had been relatively equal every day of the week) (Ventura et al., 1993); and (3) malpractice concerns (CDC, 1993).

Suggestions have been made as to what the appropriate level of cesarean section should be. A national study by the Public Health Service has set a goal that by the year 2000 the U.S. cesarean section rate should be no more than 15% (Ventura et al., 1994a). Another approach is to set "stretch goals" with respect to the cesarean section rate - reducing the rate to below 10% without compromising maternal or fetal outcomes (Berwick, 1994). This would bring the cesarean section rate back to pre-1980 levels.

Unlike the other medical interventions that we investigate, the use of cesarean section is embedded in our discharge data. In our data, for all mothers, the cesarean section rate is 24.23%; for mothers delivering prematurely it is 24.47%. The cesarean section rate does vary by hospital. For all mothers, the lowest cesarean section rate was 14.48% while the highest rate was 47.96%; for mothers delivering prematurely the lowest rate was 11.67% and the highest rate was 42.86%. The overall cesarean section rates noted in the discharge data are in line with the national average and since the data is from the time where cesarean section rates were at a peak, it would be reasonable to consider this a "worst case" scenario and impute improvements from this point. Thus, we do not initially model the use of cesarean section, but will only model quality improvement efforts that could be employed in conjunction with this

intervention, and will use the suggested reduced cesarean delivery rates as the goals in our quality simulation.

The other consideration in investigating cesarean section rates is the conditions that indicate appropriate use of cesarean section. There are several conditions obstetricians may use as indication to perform a cesarean section, including but not limited to dystocia (failure to progress), breech presentation, fetal distress and previous cesarean section. In 1991, previous cesarean section was the primary reason given for proceeding to cesarean section, accounting for 35% of cesarean sections (CDC, 1993). Table 3.4 below indicates the reasons for cesarean section as noted in different studies and in the data. Due to improvements in the surgical incision techniques, it is no longer expected that previous cesarean section was necessarily a reason for repeat cesarean (Walton, 1990). The atmosphere has changed to

Table 3.4. Reasons for Cesarean Section

Reason	CDC[1]	Walton[2]	EARMOM_D[3]
Dystocia (failure to progress)	30.4	>40.0	unknown
Previous cesarean section	35.0	13.0	18.6
Breech/other presentation	11.7		26.8
Fetal Distress	9.2		25.2
Other	13.7		
Total	100	53	70.6

1 1993; Primary reason only

2 1990; from cause-and-effect diagram indicating all reasons, not just primary. Of repeat cesarean sections, 32% were done for dystocia and 27% at the patient's request. Breech and fetal distress are noted as the second and third most frequent reasons, but percentages are not given.

3 1991-1992; as listed in ICD-9 coding, not noted as reason. Dystocia is not separated from malpresentation.

make nearly all cesarean sections, and the reasons for them, suspect. Thus, depending on the reason that a cesarean section was performed, it might have later been deemed unnecessary. This information will be used in determining which cesarean section deliveries should be replaced during the quality simulation.

Surfactant Therapy

Another option to treat infants who are born prematurely is to use exogenously administered surfactant. Surfactant is naturally produced in mature lungs, secreted by the cells lining the alveoli of the lungs to prevent the alveolar walls from sticking together (Bantam Medical Dictionary, 1981). In its absence, such as in premature births, the lungs often collapse, causing respiratory distress. Surfactant therapy aims to introduce surfactant into the lungs so that the collapse is prevented until the time that the lungs produce surfactant on their own. Surfactant is given by inserting a endotracheal tube in order to distribute 75-100mg/kg of surfactant(Hansen and Corbet, 1991). The effect of using surfactant is to decrease the severity of RDS and to decrease the mortality rate in premature infants.

Typically, one of two strategies of surfactant treatment is employed (Jobe, 1993; Hansen and Corbet, 1991; Fujiwara et al., 1991). The first, a rescue strategy, is to treat infants determined to have RDS after they are stabilized. The second, a therapeutic or prophylactic strategy, is to treat all premature infants nearly immediately, often directly in the delivery room. While this enables infants to be treated more rapidly and thus run less risk for severe complications or morbidity associated with RDS and other respiratory ailments, some infants are treated unnecessarily (30% of infants with birth weights less than 1000g do not develop RDS). Below are detailed the results of studies of the efficacy of surfactant treatment. Under a rescue strategy, Jobe (1993) noted that the mortality rate dropped from 32.1% to 21.2%; Fujiwara et al. (1991) saw a drop of 11%, from 30% to 19% (a 95% CI around the percentage of the drop is 3%-22%, p=.008). Under a therapeutic strategy, Jobe (1993) noted that the mortality rate dropped from 30.3% to 15.7%; Fujiwara et al. (1991) saw

a drop of 17%, from 33% to 16% (a 95% CI around the percentage of the drop is 2%-31%, p=.009). The rates of incidence of complications of RDS such as pneumothorax, pulmonary interstitial emphysema and bronchopulmonary dysplasia also were shown to decrease, as noted in Table 3.5.

Table 3.5. Overview of Literature on Surfactant Administration in Neonates

WITHOUT STEROIDS	Surfactant - Rescue			Surfactant - Therapeutic		
Sample - Jobe	5 studies (n=421), 28.2 wks,1150g, 6 hours			4 studies (n=224), 27.2 wk, 960g, immediate		
	Control % inc.	Surfactant % inc.	% change (95% CI)	Control % inc.	Surfactant % inc.	% change (95% CI)
Mortality						
Jobe	32.1	21.2	10.9	30.3	15.7	14.6
Fujiwara et al.	30	19	11 (3-22), p=.008	33	16	17 (2-31), p=.009
Morbidity				89	66	23
RDS- Complications: Pneumothorax						
Jobe	38.3	14.2	24.1	24.8	12.2	12.6
Fujiwara et al.	27	13	14 (8-21), p<.001	27	10	17 (7-28), p=.003
Pulm. Int. Emphy.						
Jobe	39.1	17.9	21.2	36.7	9.6	27.1
Fujiwara et al.	37	14	23 (14-32), p<.001	38	9	29 (18-40), p<.001
Bronchopulm. Dyspl.						
Jobe	34.9	29.2	5.7	46.8	39.1	7.7
Fujiwara et al.	34	26	8 (0.3-16), p=.056	45	38	NS

Adopted from Jobe, 1993 and Fujiwara et al., 1991.

Quality Management Techniques Used in Premature Labor and Delivery

Of interest in this research are the effects of various management techniques used during the process of premature labor, delivery and perinatal care and their ability to manage variation throughout the system. In order to investigate this, we first provide a brief background on the management techniques that are used in conjunction with this

process, particularly those that are centered on the medical interventions previously discussed.

Consensus Conference - Antenatal Steroids

In the United States, the National Institutes of Health (NIH) began the consensus conference program in 1977 as one mechanism by which to fulfill its responsibilities for appropriate transfer of knowledge developed in biomedical research to the clinical practice of medicine (NIH, 1981, p. iii). This program is national in scope, and aims to provide a recommended treatment plan for specific clinical interventions. Through expert presentation and discussion by an expert panel, evidence for courses of treatment are weighed and a consensus statement is prepared. This statement includes the scientific basis for the treatment and provides a recommendation for treatment use. It also includes remarks regarding research needs that would guide clinical care. Consensus conferences have been conducted on two of the medical interventions commonly used in the process of premature labor and delivery - corticosteroid therapy and cesarean sections (NIH, 1981; NIH, 1994).

Other agencies are involved in developing and disseminating practice protocols on a nationwide basis. The Agency for Health Care Policy and Research develops clinical practice guidelines in an effort to address concerns regarding variability in health care practices and the uncertainty of the effectiveness of health care services (AHCPR, 1995a). None as yet are directed towards premature labor and delivery. Additionally, the American Medical Association (AMA) catalogs practice parameters that are in use, and includes in these national and local efforts. They define practice parameters as strategies for patient management

developed to assist physicians in clinical decision-making. Practice parameters provide an effective mechanism to disseminate the results of outcomes research, technology assessments, and benefit/risk analyses (AMA, 1995, p. vii).

In the early stages of premature labor and delivery corticosteroid therapy can be administered. Corticosteroids given antenatally are intended to reduce overall mortality and the severity and incidence of RDS. The 95th consensus development conference (NIH, 1994) compiled known information about this therapy and provided a recommendation for its use. The recommendation for administration of corticosteroids is based on the gestational age of the fetus. The outcomes of the premature delivery are affected not only by the steroids, but also by comorbidities (particularly premature rupture of the membranes, or PROM) and the time from steroid administration to delivery. Before consensus development only 12-18% of women who delivered prematurely were treated with antenatal corticosteroids (NIH, 1994, p.9). The highest rate of administration observed was 60%. The implementation of this consensus conference recommendation changes the rate of administration of corticosteroids to women experiencing premature labor.

Simulation modeling of this consensus conference recommendation is fashioned as a Monte Carlo simulation. Outcomes are measured as patient outcome, including mortality and presence/severity of RDS. Costs of the program will also be assessed. Previous studies have shown that significant savings are possible due to estimated decrease of length of stay with lower morbidity and mortality (Simpson and Lynch, 1994). Use of consensus conference guidelines will affect the antenatal steroid administration rate (a range from 15% to 60%).

Continuous Quality Improvement - Cesarean Section

CQI efforts in health care have largely been modeled after total quality management (TQM) efforts in the manufacturing arena. CQI uses statistically based tools in order to assess and manage processes. The goal is to bring the process into "control" so the performance of the process is within statistically acceptable limits. Then, process improvements can be made in an effort to improve both the mean performance and to decrease the amount of variance in the process so that the acceptable range is more limited. Common results reported from implementing CQI include reduction of delays, financial savings, and increased patient satisfaction (Berwick et al., 1990).

CQI methods are applicable to any of the phases of premature labor and delivery. We will explore CQI specifically with respect to the use of cesarean section delivery. If CQI is applied, it is expected that the mean and variance of the maternal length of stay and costs associated with that stay will decrease. Also, maternal morbidity (infection rates) could decrease due to the elimination of special cause variation.

One example of CQI involving premature labor and delivery has been documented previously (Walton, 1990). At West Paces' Hospital in Atlanta, they studied cesarean section rates using quality control methods. They identified the reasons for cesarean sections being done and identified that the fourth largest reason for cesarean section was previous cesarean section, accounting for 13%. While other reasons were considered medical and possibly emergent situations, the practice of performing cesarean sections simply because the mother had a previous cesarean section was controversial. They also measured the rate of cesarean delivery on a monthly basis and found considerable variation in the rate.

Clinical Pathways (Local Intervention) - Surfactant

Clinical pathways are plans of care that outline the optimal sequencing and timing of interventions for patients with a particular diagnosis, procedure, or symptom (Ignatavicius and Hausman, 1995). Clinical pathways are typically used to manage a patient visit. They focus on decreasing the amount of time in the system by minimizing delays, minimizing the use of resources, and maximizing the quality of patient care. Because they are locally developed or adapted, the results differ between institutions. The pathways are generally designed around a particular DRG and thus are appropriate after a diagnosis has been made. Continuous review of the pathway is typically done by tracking patients and monitoring cases where deviations from the pathway have occurred. Additionally, retrospective analysis is generally done to compare outcomes of different patients followed on the same clinical pathway (Ignatavicius and Hausman, 1995). Pathways are affected by physician compliance and patient compliance. Over time, it is expected that physician compliance rates would increase, due both to familiarity with the pathway and to iterations of the pathway that increase the physicians' sense of ownership of the process. This predictability of procedural timing and resource utilization has considerable potential for planning, scheduling and controlling direct and indirect patient services (Roth and van Dierdonck, 1995). Clinical pathways are expected to have an impact on both time and resource use, improving the mean and reducing the variance of both of these variables.

In this chapter we have provided detailed information regarding the process of labor and delivery as well as reviewed information regarding the medical interventions and quality

management techniques as they apply to this process. In combination with the literature review done in Chapter Two, this information provides the basis for the simulation modeling, which is detailed in the next chapter.

CHAPTER 4

RESEARCH METHODS

> There is more to doing research than is dreamt of in philosophies of science, and texts in methodology offer answers to only a fraction of the problems one encounters....No matter how carefully one plans in advance, research is designed in the course of its execution.
>
> H. S. Becker

The essence of this dissertation is to investigate the effects of the medical interventions and quality management techniques on the process of premature labor. There is currently no good manner in which to assess the effects of management tools used in health care due to the large degree of inherent variation in the population. Thus, this work is considered exploratory in nature. A simulation model that is designed to encompass a specific health care process has been built in order to assess the impact of specific medical interventions and quality management techniques. These efforts provide useful and meaningful insights into the management of health care processes and particularly the field of outcomes research.

In order to provide a specific context for this analysis, we elected to investigate one health care process, premature labor and delivery. The relevant details of the process have been overviewed in the literature and the three stage transformation model was adapted to reflect this process (see Figure 3.1). Three commonly-used medical interventions are included in the process, each affecting different phases. Finally, quality management techniques that have been used with this process are incorporated and the effects of each are modeled. The

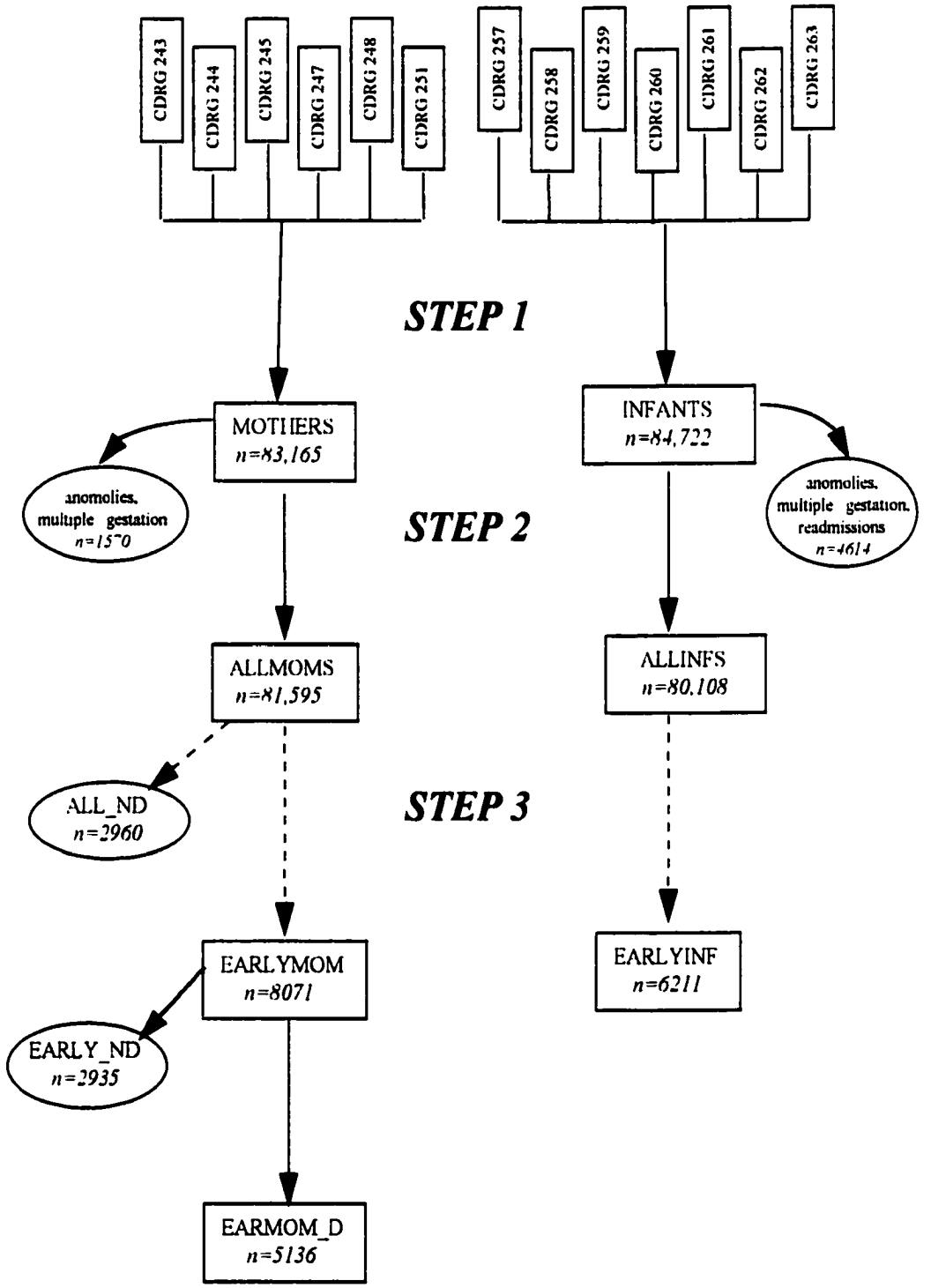
details of modeling the process, the medical interventions and the quality management techniques are given below.

While it is clear by the number of conditions, comorbidities and interactions discussed in the medical literature that a very large number of variables that could be included, it is also desirable to limit the number of variables in the study for two reasons. First, we are not interested in simply modeling known medical knowledge, but in providing an analytic tool for assessing the usefulness of management tools given a large degree of underlying variation. Second, the variables that are included need to either serve as sources of variation that are affected by the medical interventions and management techniques described or to be proxies for the inherent variation in the system.

Initial Data Analysis and Compilation

A data sample of women who experience premature labor (and delivery) is used to investigate the use of simulation methods to assess the effects of the medical interventions and quality management techniques on the process of premature labor. Using 1991 and 1992 discharge data from the state of Maryland, two data sets are compiled, one of mothers experiencing premature labor/delivery and one of infants born prematurely. Figure 4.1 diagrams the process of creating these working data sets. The working data sets are compared to the total population to determine the significant differences associated with prematurity. An attempt is made to match mothers and infants and use a subset of matched data; the validation of this data is reviewed to show that while the matching process was valid, the subset does not accurately reflect the working data sets. Thus, the working data sets are

Figure 4.1. Diagram of Steps Involved to Create Working Data Sets



used as the inputs for the simulation study. The characteristics of these data sets are reviewed.

Maryland Discharge Data Sets

46 hospitals in the state of Maryland participate in a statewide discharge data base. This research makes use of the non-confidential format of case mix discharge data from 1991 and 1992. Each discharge is represented by one record or observation. The 1991 and 1992 discharge data from Maryland discharge data set includes 83,165 discharges of women who were either expecting or had just delivered and 84,722 infants, both normal and premature. The data is coded by Categorical Diagnosis Related Group (CDRG) and by Diagnosis Related Group (DRG) and provided in separate SAS data sets for each CDRG. These codes help identify the population of expectant/delivered women and newborn infants from which our data will be obtained. Table 4.1 below details the number of mothers in each CDRG/DRG category. Premature labor is indicated within specific diagnosis variables of each observation, and the number of women coded with diagnoses indicating prematurity is also noted in this table. Table 4.2 details the number of infants born during the years that the data was collected, by CDRG/DRG. For infants, the indication of prematurity is subsumed in the CDRG/DRG codes. In Figure 4.1, Step 1 reflects the combination of all the separate CDRG data sets into two large data sets, one for mothers (MOTHERS) and one for infants (INFANTS).

For each observation included in MOTHERS and INFANTS, 64 original variables were obtained. These variables contain information on the hospital stay (such as admission day, discharge day, and length of stay), the health care institution where care took place (such

Table 4.1. Women from Maryland discharge database listed by CDRG/DRG coding with incidence of prematurity noted, for both the entire data set (MOTHERS) and the data set less exclusions, particularly multiparity (ALLMOMS).

CDRG	DRG	Description	MOTHERS			ALLMOMS		
			# records	# prem.	% prem.	# records	# prem.	% prem.
243		Cesarean Section	19,788	1503	7.6	19,060	1,257	6.6
	370	with complications		926		7,338	796	10.8
	371	without complications		577		11,722	461	3.9
244	374	Vaginal delivery, with sterilization or D&C	3,027	209	6.9	2,910	189	6.5
245	375	Vaginal delivery with other surgical procedure	35	4	11.4	33	2	6.1
247	372	Vaginal delivery with complicating diagnoses	6,721	746	11.1	6,640	704	10.6
248	373	Vaginal delivery without complicating diagnoses	50,428	3,119	6.2	50,017	2,984	6.0
251	379	Threatened abortion	3,166	3,166	100.0	2,935	2,935	100.0
		TOTALS	83,165	9,184	10.5	81,595	9,184	9.9

as county, bed size, and coding structure), the reason for the admission and the type of care provided (such as diagnoses, procedures and physician identifiers), patient demographics (such as age and race) and charges. Depending on the information provided by the variable, it was used in matching mothers and infants, as an input to the simulation, for measuring the output of the simulation, for selecting data to include or delete, as an experimental variable in the simulation, or simply as information about the characteristics of and care of women delivering prematurely or of premature infants. Some of the information contained in each observation is redundant and thus is discarded from the working data set.

Table 4.2. Infants from Maryland discharge database listed by CDRG/DRG coding with incidence of prematurity noted, for both the entire data set (INFANTS) and the data set less exclusions, particularly multiparity (ALLINFS).

CDRG	DRG	Description	INFANTS		ALLINFS	
			# records	%	# records	%
257	385	Neonates, died or transferred	1618		1344	
258	386	Extreme immaturity or RDS, neonate	1518		1167	
259	387	Premature with major problems	2420		1864	
260	388	Premature without major problems	2269		1877	
		SUBTOTAL, Premature	7825	9.2	6252	7.8
261	389	Full term neonate with major problems	7472		6413	
262	390	Neonate with other significant problems	16,610		16,206	
263	391	Normal newborn	52,815		51,278	
		SUBTOTAL, Full Term	76,897	90.8	73,897	92.2
		TOTALS	84,722	100.0	80,149	100.0

Additional variables needed for all observations are created at this time. Of particular interest are the eight data fields which encapsulate information regarding the applicable diagnoses and procedures used during the episode of care. These data fields are coded using the standard International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) coding practices. Within the codes listed for these diagnoses and procedures is the information regarding premature labor, delivery and relevant neonatal conditions such as respiratory distress syndrome that are applicable for this study. For instance, a principal diagnosis listed as 769 indicates that the patient was primarily treated for respiratory distress syndrome. There may also be up to four secondary diagnoses noted. As the information is presented in the discharge data, it is necessary to review all five of these data fields for each

observation in order to determine whether or not a patient is treated for any specific condition. For instance, the code denoting RDS (769) may occur in any of these five fields. Thus, extensive programming was done to glean the relevant information from these all five diagnoses code fields and to create binomial variables that can be used to analyzed the incidence of the conditions of interest (including RDS, cesarean section, and premature rupture of membranes (PROM)) . A similar procedure is done for the other medical procedures. As with all the variables included in the data, these variables have a variety of purposes. They are primarily used either for matching (where a condition would be noted in both the mother and the infant, such as cesarean section) or for the measurement of the incidence of conditions in the data set. The matching procedure is discussed in further detail later in this section.

Other variables were also created. Each observation in the composite data sets is given an unique identification number (m_id and b_id). For the infant data set, all variables not used for matching for which information needs to be retained are renamed to have an initial letter b. Finally, some variables such as adm_day and payment are used for matching but the information needs to be retained for each individual mother and infant, so the variable is copied to new variables with m and b predecessors to retain the original information when a merge occurs.

Thus, MOTHERS included 83,165 observations with 84 variables and INFANTS included 84,722 observations and 94 variables. These are considered the population data sets for this study. We reviewed the data for the prevalence of prematurity within the population prior to subsetting the premature mothers and infants in order to begin developing a working

data set. Of all the mothers discharged, 8,747, or 10.5%, experienced some form of early labor or gave birth prematurely as identified by principal or secondary procedures coded according to ICD-9-CM standards. For infants, 7825 (CDRG257 through CDRG260), or 9.2% were delivered prematurely. The difference in the percentages for mothers and infants is due to the number of premature labor episodes that did not result in the delivery of an infant (the labor was arrested and the mother discharged, or the pregnancy was terminated). Additionally, for 19,788 mothers, cesarean section was listed as a principal or secondary procedure. The number of cesarean sections represents 24.23% of the 78,635 women who delivered during this episode of care, slightly higher than that of the U.S. population during the same time frame. In the U.S. population during 1991 and 1992 the cesarean section rate was reported to be between 22.6% and 23.5% (CDC, 1993; Ventura et al., 1993, 1994a, 1994b).

The Relevant Subset of Premature Labor Episodes

With all the mothers and infants from the discharge data identified and combined into two data sets for the population, the next task was to identify the relevant observations to use for this investigation. This is noted in Step 2 in Figure 4.1. For mothers, we wanted to include women experiencing premature labor but exclude those that might have medical or other conditions affecting prematurity, such as multiparity, preeclampsia, or history of drug abuse. Infants excluded from the data set included twins, triplets, and other multiples, and those that were readmitted. These exclusions of mothers and infants are consistent with other studies reported in the literature (Robertson et al., 1992; Lewis et al., 1996).

All mothers with diagnoses codes indicating multiparity and all infants who were part of a multiple birth as noted in diagnoses codes (mothers: 659.4, V23.3, V27.2-V27.7 and infants: V31-V37, V61.5) were excluded. Also, infants that were readmissions rather than births (noted by an admission age in days greater than 0) were excluded. The data sets were reviewed to reveal any other anomalies and outliers, odd mothers and infants that should be excluded from the study due to their differences. An example of this is mothers who were in psychiatric bed types - their length of stay, average charges, and other information associated with the episode of care are likely to be different than a mother who is not under psychiatric care. This initial exclusion is represented by Step 2 in Figure 4.1. In all cases, exclusions were retained in separate data files rather than simply eliminated in case they were later deemed necessary to the investigation.

For the mothers, exclusions for multiparity, regardless of prematurity, were represented by 1,570 observations, leaving a master set of mothers used for the study including 81,595 observations (ALLMOMS). While for all mothers the rate of premature labor is 10.5%, the rate of premature labor in the set less exclusions was 9.9%, a reflection of the fact that the exclusions have a higher incidence of prematurity. A breakdown of these exclusions by CDRG/DRG is provided in Table 4.1.

A total of 4,614 of the infant observations had been born as the result of a multiple birth or were observations from a readmission stay rather than the initial stay subsequent to delivery. With these observations removed from the data set, the master set of infants used for the study includes 80,108 observations (ALLINFS). The rate of premature labor in the set less exclusions is 7.8% rather than 9.2%. As with the mothers, this reflects a higher

prematurity rate than the entire set; the difference in the percentages is compounded by the fact that the 9.2% is inclusive of multiple births. A breakdown of these exclusions by CDRG/DRG is provided in Table 4.2.

The next concern was to identify a relevant set of mothers for this study. These mothers should be noted as experiencing premature labor and delivering prematurely. Table 4.3 details the ICD-9-CM codes that indicate premature labor and/or delivery and the number of observations that reflected each code. Observations having these codes as a principal or secondary diagnosis were selected for possible inclusion in the data set. A possible maximum number of observations in the subset is 8,082. Eleven observations (8,082-8,071) were coded with more than one relevant ICD-9-CM as a diagnoses. The separation of 8,071 mothers that experienced premature labor is reflected as Step 3 in Figure 4.1. 8,071 reflects a rate of premature labor of 9.9%. This set is noted EARLYMOM. Of these 8,071 mothers that experienced premature labor, 5,136 (63.64%) delivered during the hospital stay associated with the discharge information. These 5,136 mothers who delivered prematurely are separated from those who did not deliver during that hospital stay in order to eliminate the same mother from possibly being included in the data set more than once, as would happen if they were discharged after one episode of premature labor and then readmitted and delivered prematurely (Figure 4.1, Step 4). These 5,136 mothers comprise the relevant set for the purposes of this investigation (noted EARMOM_D). The rate of premature delivery represented by these 5,136 mothers is 6.29%.

Descriptive statistics and frequencies of the subset of 5,136 observations were reviewed. Of the observations selected, 1,166 (22.7%) were discharged in 1991 and 3,970

(77.3%) were discharged in 1992. This skewness in percentages reflects a reduced data collection period at some hospitals. Table 4.4 reflects the descriptive statistics and relevant frequencies of the subset of premature mothers and compares them to the master set of all mothers in the discharge data.

The subset of mothers delivering prematurely is used to better understand the characteristics of mothers that experience premature labor. The pertinent characteristics are those that provide information regarding the resource use, charges, location, timing, identifying characteristics, diagnoses or procedures of the specific medical episode. These characteristics were also used to create variables relevant to the simulation, to identify attributes associated with cases to be included in the simulation as previously indicated, and to match mothers with infants.

To understand how premature infants as a whole might differ from the population of infants delivered, a similar subset of premature infants was created (Figure 4.1, Step 3). This subset was created by combining all infants noted with CDRG 257-CDRG260 (DRG 385-DRG388) since the premature notation is subsumed in the DRG coding. 6,211 infants were included in this set which is noted EARLYINF. Table 4.5 reflects the descriptive statistics and relevant frequencies of the subset of premature infants and compares them to the master set of all infants in the discharge data.

Table 4.3. Applicable ICD-9 codes indicating women being treated for premature labor and delivery.

ICD-9 codes		Description	PRIN_DX	SEC_DX_1	SEC_DX_2	SEC_DX_3	SEC_DX_4	TOTALS
4th digit	5th digit (a)							
640.0	(0)	Threatened Abortion (before 22 weeks gestation)	4	0	0	0	0	4
	(1)		10	2	1	1	0	14
	(3)		147	0	1	0	0	148
640.8	(0)	Other Specified Hemorrhage in early pregnancy	1	0	0	0	0	1
	(1)		1	1	0	0	0	2
	(3)		19	1	0	0	0	20
640.9	(0)	Unspecified hemorrhage in early pregnancy	0	0	0	0	0	0
	(1)		1	2	0	0	0	3
	(3)		17	2	0	0	0	19
644.0	(0)	Threatened premature labor (after 22 weeks, before 37 weeks)	5	1	0	0	0	6
	(3)		2742	6	0	0	1	2749
644.2	(0)	Early onset of delivery (before 37 weeks gestation)	0	0	0	0	0	0
	(1)		2895	757	671	475	318	5116
		TOTALS	5842	772	673	476	319	**8082**

(a) 5th digit coding: 0-unspecified as to episode of care or not applicable
1-delivered, with or without mention of antepartum condition
2-delivered, with mention of postpartum complication
3-antepartum condition or complication, not delivered
4-postpartum condition or complication, delivery during previous admission or outside hospital

Table 4.4. Comparison of all mothers in database (ALLMOMS) to those mothers experiencing premature labor/delivery (EARLYMOM).

Variable Type/ Variable	ALLMOMS *n* =81595		EARLYMOM *n* =8071	
Descriptive	mean	std dev	mean	std dev
age	27.07	7.53	26.17*	5.94
los	2.30	2.20	3.03*	4.23
dal chg	1716.36	28201.23	2487.39*	35173.35
tot chg	10525.86	282066.89	15688.64*	351698.72

Binomial	percent=1	std dev (a)	percent=1	std dev
deliver	.9637	.187	.6364	.4811
csect (b)	.2423	.429	.2447 (NS)	.4300
prom	.0749	.2633	.1436*	.3507
plprev	.0073	.0851	.0244	.1543
othplac	.0136	.1158	.0368	.1883
preecl	.0276	.1638	.0421	.2009
infect	.0417	.1999	.0706	.2562
breech	.0349	.1836	.0611	.2395
forvac (b)	.1261	.3320	.0750	.2634
prolong (b)	.1056	.3073	.1020	.3027
badcare	.0192	.1373	.0396	.1951
lt 22wks	.0964	.2952	.0250	.1585
wk22 37	.0964	.2952	.9750	.1562

Variable Type/ Variable		ALLMOMS *n* =81595		EARLYMOM *n* =8071	
Categorial	**Values**	freq	percent	freq	percent
race	1 white	52291	64.1	4141	51.3
	2 black	24642	30.2	3628	45.0
	3 asian	1975	2.4	104	1.3
	4 amer ind	85	0.1	4	0.0
	5 other	2593	3.2	192	2.4
	9 unident.	9	0.0	2	0.0
payment	1 medicare	110	0.1	13	0.2
	2 medicaid	22994	28.2	2909	36.0
	3 title V	2	0.0	0	0.0
	4 blue cross	13350	16.4	1085	13.4
	5 insur co	17750	21.8	1382	17.1
	6 other govt	653	0.8	40	0.5
	7 work comp	10	0.0	2	0.0
	8 self-pay	3628	4.4	422	5.2
	9 no charge	11	0.0	1	0.0
	10 other	211	0.3	18	0.2
	11 donor	1	0.0	1	0.0
	12 HMO	20641	25.3	1793	22.2
	13 MSO	541	0.7	103	1.3
	14med/HMO	1679	2.1	299	3.7
	99 unknown	14	0.0	3	0.0
marital status	1 single	26386	32.3	3587	44.4
	2 married	50591	62.0	3954	49.9
	3 separated	1630	2.0	209	2.6
	4 divorced	1373	1.7	159	2.0
	5 widowed	151	0.2	20	0.2
	9 unident.	1464	1.8	142	1.8

(a) calculated as $s = \sqrt{[p(1-p)]}$

(b) calculated as % of those women who delivered during the episode of care

* - significantly different at p=0.01

NS- not significantly different

Table 4.5. Comparison of all infants in database (ALLINFS) to those mothers experiencing premature labor/delivery (EARLYINF).

Variable Type/ Variable	ALLINFS *n 80149*		EARLYINF *n 62112*	
Descriptive	mean	std dev	mean	std dev
bbrth wt	3367.23	714.20	2299.14*	915.70
blos	2.93	7.39	12.53*	22.91
bdal chg	1330.94	14866.06	7488.78*	23319.81
btot chg	3524.09	141429.32	13472.16*	180670.63
days nur (b)	2.63 (n 3518)	8.33	3.49 (n 1942)	9.11
Binomial	percent=1	std dev (a)	percent=1	std dev
bcsect	.2254	.4178	.3048	.4604
bprom	.0018	.0429	.0074	.0857
bplprev	.0002	.0158	.0014	.0380
bpreecl	.0012	.0353	.0042	.0646
bbreech	.0023	.0485	.0052	.0716
RDS	.0191	.1371	.2447*	.4300
othresp	.0615	.2403	.2333*	.4230
vent	.0126	.1115	.1372*	.3441
intub	.0203	.1412	.1697*	.3754
respther	.0149	.1213	.0844*	.2780
lt 28wks	.0076	.0883	.0966	.2954
wk28 37	.0576	.2330	.7389	.4393

Variable Type/ Variable		ALLINFS *n 80149*		EARLYINF *n 6252*	
Categorial	**Values**	freq	percent	freq	percent
sex	1 male	40957	51.1	3195	51.4
	2 female	39188	48.9	3015	48.5
	9 unknown	4	0.0	1	0.0
race	1 white	52129	65.0	3164	50.9
	2 black	23575	29.4	2821	45.4
	3 asian	1954	2.4	97	1.6
	4 amer ind	80	0.1	4	0.1
	5 other	2407	3.0	125	2.0
	9 unident	4	0.0	0	0.0
payment	1 medicare	45	0.1	1	0.0
	2 medicaid	22685	28.3	2468	39.7
	4 blue cross	12953	16.2	767	12.3
	5 insur co	17340	21.6	1005	16.2
	6 other govt	633	0.8	41	0.7
	7 work comp	4	0.0	0	0.0
	8 self-pay	4566	5.7	398	6.4
	9 no charge	10	0.0	0	0.0
	10 other	192	0.2	8	0.1
	12 HMO	19804	24.7	1250	20.1
	13 MSO	177	0.2	22	0.4
	14med/HMO	1726	2.2	250	4.0
	99 unknown	14	0.0	1	0.0

(a) calculated as $s = \sqrt{[p(1-p)]}$

(b) For those infants who spent time in both the NICU and the nursery, the number of days spent in the nursery

* - significantly different at p 0.01

NS- not significantly different

Subsetting and Matched Pair Identification - An Attempt at Wholistic Process Simulation

Ideally a subset of the premature mothers could be used as input for this simulation study and still be large enough to produce reasonable results (allowing us to avoid matching all 5,136 observations). Table 4.6 gives information on all births at each provider number from which discharge data is utilized. Four hospitals were selected from the 46 hospitals (numbers 3,4, 15, and 32). These hospitals were selected on the basis of their underlying cost structure and rates of premature delivery/birth comparable to the statewide population. The data from these hospitals was separated out and used to match mothers with infants in an attempt to provide a basis for a holistic process simulation.

A total of 1,060 mothers who had both experienced premature labor and delivered during the current hospitalization from these 4 hospitals were considered during the matching process (data set MATCHMOM). The matching process used information regarding location, timing, identifying characteristics, diagnosis and procedure from each observation to determine mother-infant pairs. A total of 990 mothers were successfully matched with infants by this procedure, 93.4% of the MATCHMOM data set. When combined with mothers who did not deliver, a data set of 1,680 observations to be used as input into the simulation (DATAUSE). From this data set, further exclusions were made as detailed in Appendix A, and a final set of 1,532 observations was compiled (DATASET).

The paired data set was analyzed by statistical comparisons to the set of all mothers that experience premature labor and all infants that are delivered prematurely. The statistical comparisons revealed significant differences between the subset and the sets of all mothers that experience premature labor and all infants that are delivered prematurely. The results of

these comparisons are given in Appendix A. Thus, although the matching procedure was successful, the subsetting would not be appropriate for generalizing the results. Matching a larger or more adequate subset could be done in later research.

Implications of Using Unpaired Data

Since the matching of mothers and infants on a small scale was not going to be representative of the population, it was determined that the simulation experiments would be based on the EARMOM_D and EARLYINF data sets rather than the paired subset data. Thus, for input into the simulation models we used discharge data from 5,136 mothers who experienced premature labor and delivered during that hospital stay and discharge data from 6,211 infants who were delivered prematurely. This still enables us to model the processes and quality management techniques and measure outcomes as planned. Thus, processes and outcomes are linked in our models. The decision not to pair all the data does limit the scope of this link, however. We are not able to directly assess how interventions during the care of the mother (cesarean section and corticosteroid administration) might affect the infant. For the effects of cesarean section on the mothers and the surfactant models there should be no difference using this approach. A wider implication for the use of quality management techniques can also be taken from the results of this attempted pairing. Since a subset of the hospitals did not adequately reflect the population, we can infer that it may be unwise to take quality improvement results from one hospital and generalize to all hospitals. Thus, the inability to use a subset actually supports our desire to develop methodologies appropriate for investigating the wide-scale effects of quality improvement techniques.

Table 4.6 Hospital Information

PROV NUM	Teach	Beds >400	MOTHERS							INFANTS						
			#	# Prem.	% Prem.	Csect Rate (%)	Delivery Rate (%)	PREMATURE		#	# Prem.	% Prem.	Csect Rate (%)	RDS (%)	PREMATURE	
								Csect Rate	Delivery Rate (%)						Csect Rate (%)	RDS (%)
1			1516	83	5.5	23.8	98.4	11.7	72.3	1549	66	4.3	22.0	1.0	25.8	24.2
2	Y	Y	2359	559	0.2	17.5	93.1	20.7	70.8	2234	409	18.3	15.6	6.5	24.0	34.7
3		Y	3412	366	10.7	23.7	97.3	24.5	74.9	3370	287	8.5	22.9	2.7	35.2	31.7
4		Y	6639	855	12.9	24.2	94.3	27.5	55.7	6230	429	6.9	23.6	0.8	35.7	11.9
5			1925	96	5.0	25.1	98.2	17.7	64.6	1919	92	4.8	24.2	1.2	30.4	23.9
6			445	26	5.8	22.2	98.2	27.8	69.2	539	20	3.7	18.7	1.9	45.0	50.0
7	Y	Y	2088	151	7.2	24.2	97.6	320.0	66.2	2032	108	5.3	23.5	1.5	38.0	27.8
8	Y		1999	305	15.3	25.8	95.4	28.5	70.2	1925	320	16.6	21.6	4.6	25.4	26.9
9	Y	Y	3101	476	15.4	21.3	97.5	18.1	83.8	3046	456	15.0	20.2	2.7	28.1	18.0
10			318	54	17.0	29.3	89.0	36.8	35.2	284	24	8.5	26.4	2.1	37.5	25.0
11	Y	Y	2544	286	11.2	21.2	96.2	18.2	67.1	2490	254	10.2	19.2	3.0	24.0	29.5
12	Y	Y	3789	637	16.8	22.8	92.1	23.3	54.0	3488	295	8.5	21.9	2.0	33.2	23.7
14	Y		3951	346	8.8	19.7	98.2	21.3	78.9	3963	301	7.6	0.2	1.5	23.9	19.9
15			2355	203	0.1	27.7	94.7	22.1	37.9	2255	167	7.4	26.9	1.1	31.7	14.4
16			327	45	13.8	14.5	90.8	26.7	33.3	370	17	4.6	11.4	0.0	29.4	0.0
17			996	50	5.0	32.5	99.0	22.5	80.0	978	43	4.4	31.7	0.5	37.2	11.6
18			1718	160	9.3	32.3	96.4	22.2	61.9	2103	132	6.3	24.2	0.8	26.5	12.9
21			2923	134	4.6	25.2	98.5	28.9	0.7	2884	116	4.0	24.2	0.5	36.2	12.9
22	Y	Y	2295	224	9.8	20.2	95.3	28.7	51.3	1964	124	6.3	19.5	1.7	29.0	25.8

Table 4.6 Hospital Information (Continued)

PROV NUM	Teach	Beds 400	MOTHERS							INFANTS						
			#	# Prem.	% Prem	Csect Rate (%)	Delivery Rate (%)	PREMATURE		#	# Prem.	% Prem.	Csect Rate (%)	RDS (%)	PREMATURE	
								Csect Rate	Delivery Rate (%)						Csect Rate (%)	RDS (%)
23			1004	120	12.0	30.3	92.7	38.8	40.8	1164	83	7.1	23.8	2.0	32.5	27.7
25			1275	52	4.1	21.4	97.3	21.1	36.5	1470	64	4.4	17.8	1.0	23.4	23.4
26			1505	68	4.5	31.4	96.4	42.9	20.6	1466	121	8.3	29.8	1.7	32.2	20.7
27	Y	Y	2158	302	14.0	20.2	96.8	26.2	77.2	2222	381	17.1	17.8	8.8	25.5	51.2
28			385	39	10.1	42.2	94.3	35.3	43.6	366	28	7.7	40.7	1.1	32.1	14.3
30			483	66	13.7	48.0	91.5	34.6	39.4	1058	63	6.0	18.8	0.6	17.5	9.5
31			1212	61	5.0	27.0	98.3	22.5	65.6	1200	49	4.1	25.9	1.0	36.7	24.5
32		Y	4358	326	7.5	23.9	97.9	24.9	71.5	4248	252	5.9	22.4	1.5	28.6	24.6
33			1185	100	8.4	0.3	96.8	25.8	62.0	1157	96	8.3	26.6	1.0	36.5	12.5
35			2036	135	6.6	17.6	97.4	12.4	60.0	1845	135	7.3	17.5	0.8	18.5	10.4
36	Y	Y	2373	303	12.8	18.9	97.9	18.8	84.2	2343	312	13.3	18.3	2.6	25.3	19.6
37			1209	90	7.4	22.8	97.4	23.7	65.6	1289	95	7.4	24.0	1.0	33.7	13.7
42	Y	Y	7951	786	9.9	27.7	95.9	31.9	58.7	7158	416	5.8	25.7	2.4	45.2	39.4
46			2812	151	5.4	21.5	98.4	29.4	72.2	2751	134	4.9	21.2	0.8	34.3	16.4
49			1444	99	6.9	32.1	97.0	0.2	56.6	1407	78	5.5	30.9	1.1	33.3	19.2
50			736	40	5.4	28.9	96.5	14.3	35.0	711	44	6.2	25.6	2.0	22.7	31.8
53			4769	277	5.8	24.1	97.2	30.3	52.4	4671	200	4.3	23.3	0.6	40.0	14.5
TOTALS			81595	8071	9.9	24.2	96.4	24.5	63.6	80149	6211	7.7	22.5	1.9	30.5	24.5

Simulation Design

Two stages of modeling are undertaken in this simulation design. The first stage builds into the data the appropriate medical interventions - use of tocolytics, corticosteroids, and surfactant - that are commonly used in the premature labor and deliver process. The results of this first stage give us an idea of how medical practices can improve the outcomes of the infant population. The second stage of the simulation is to incorporate quality improvement techniques into the process. The results of this second stage provide insight as to how prevalent the improvement of outcomes could be and what the effect of an implementation period would be.

Using the information from the literature given above, we detail precisely how the simulation model is configured. Within the empirical data from the Maryland discharge data set, the only medical intervention represented (embedded in the data) is cesarean section. This provides a baseline indication of the mortality rate and the incidence and severity of RDS.

Subsequently, in the first stage of the simulation, new variables are incorporated into the discharge data set in order to represent the medical interventions. The outcomes as measured by mortality rate and incidence/severity of RDS differ as a result of the use of the additional therapy. This model is then repeated with a likely level of use of each medical intervention based on current (1991) prevalence of the use of each treatment. This model provides the likely results that would have occurred in our population had the medical treatments currently available been used at that time in those health care centers.

Quality interventions can have different effects on the outcomes. An optimal level of use of each medical intervention is determined from the information in the medical literature.

We consider three specific quality interventions, each affecting a different stage of the premature labor/delivery process. First, we consider the use of recommendations from the Consensus Development Conference on Antenatal Corticosteroid Use. Second, Continuous Quality Improvement (CQI) is considered in conjunction with cesarean section. Finally, we model the effects of the use of a protocol recommending a specific surfactant strategy. To mimic the use of a locally developed protocol we also model a combined strategy protocol that more adequately represents what some practitioners are doing.

A synopsis of the simulation models that are used is presented in Table 4.7. The next two sections detail the general simulation process, including the selection and replacement of observations, validation, and measurement. Subsequently, details of each of the simulation models are given, first for the general modeling of medical interventions and then for the modeling of quality interventions.

Simulation Modeling - General Process

The data sets EARMOM_D and EARLYINF on which the simulations are done are SAS data sets and the simulation models are also built in the SAS programming language. SAS allows for variable manipulation via a number of random distributions. For our purposes, the RANUNI function is used to generate uniformly distributed random variables (DiIorio, 1991). In our models, RANUNI is used for both random selection of a treatment group and for replacement by assigning random numbers to each observation and then compared to the expected rate of change to select observations for replacement in accordance with the simulation model design.

Table 4.7 Synopsis of Simulation Models

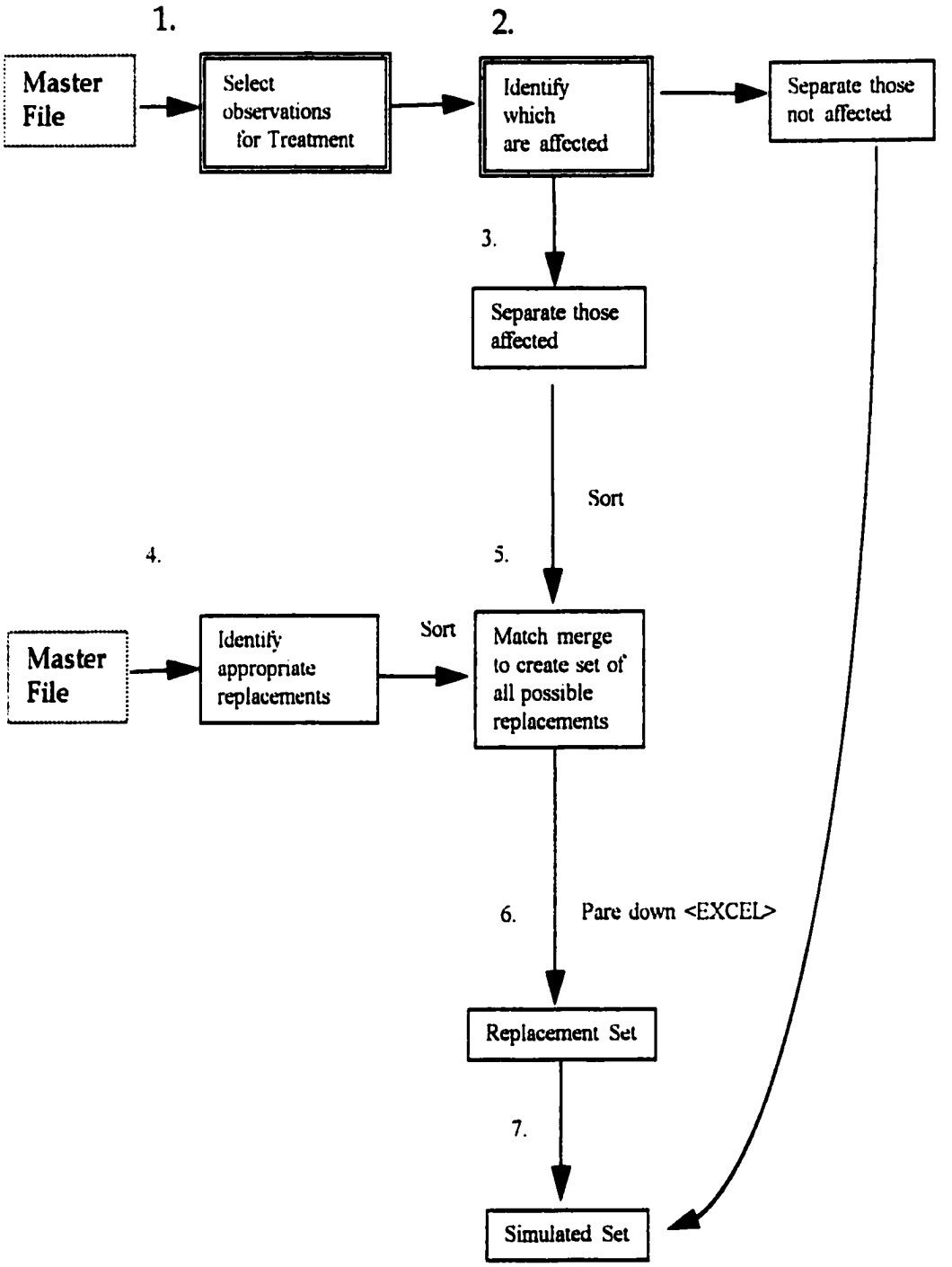
Medical Intervention	Quality Technique	SIMULATION MODEL BASELINE	Selection exclusion/inclusion STEP 1	*choice*	Affected *change* STEP 2	ID random select pop STEP 4	SORT/MERGE *control variables* STEP 5/6	Anticipated Effects
Cesarean Section	CQI							Reduce charges
		MCSECT20	mcsect=1, prom=0	0 20	0 1830	mcsect=0	provider, age, fet dist, breech	
		MCSECT15	mcsect=1, prom=0	0 15	0 3874	mcsect=0	provider, age, fet dist, breech	
		MCSECT10	mcsect=1, prom=0	0 10	0 7235	mcsect=0	provider, age	
		CSECT25	csect=1	0 25	0 1796	csect=0	provider, wt	
		CSECT20	csect=1	0 20	0 3439	csect=0	provider, wt	
		CSECT15	csect=1	0 15	0 5077	csect=0	provider, wt	
		CSECT10	csect=1	0 10	0 6720	csect=0	provider, wt	
Corticosteroid Use	Consensus Conference							Reduce incidence of RDS
		STER15		0 15	0 440132		provider, wt	
		STER30		0 30	0 440132		provider, wt	
		STER45		0 45	0 440132		provider, wt	
		STER60		0 60	0 440132		provider, wt	
Surfactant	Clinical Pathways							Reduce severity of RDS
		RESC	rds=1	n/a		rds=1,sev <3	provider, wt,csect	
		THER	wt<=8	n/a		rds=1,sev <3	provider, wt,csect	
		MIX	rds=1 OR wt<=6	n/a		rds=1,sev.<3	provider, wt,csect	

Random Selection and Replacement Sampling Strategies

The simulation models are done by a combination of random selection for the medical treatment/quality intervention and replacement from an appropriate representative group. A two part random selection process is undergone. First, for those patients selected to "receive" a medical treatment are identified (except in the case of cesarean section, where the selection indicates that cesarean section is not done). Second, patients responding to the effect of that treatment are selected, identifying which observations should then be replaced. A diagram of this simulation process is given in Figure 4.2.

When a quality intervention is introduced, it impacts the selection criteria in the first part of the selection process (the criteria, not the effect). When a change is affected as a result of a quality intervention, two types of changes are possible. In general, these types of changes are analogous to eliminating special cause variation and eliminating common cause variation. The first type of change is the selection criteria for the use of the intervention. For instance, quality improvement methods would call for stricter protocols on which infants receive surfactant or which mothers have cesarean sections. The second is in the rate of use of the intervention or the protocol used. This will not affect the sampling strategy per se; rather, it will affect the number of mothers/infants that are synthetically replaced. These two types of changes are simulated using programming techniques that allow for exclusion or inclusion of certain parameters and for varying rates of change. Details of the specific calculations for each simulation model are given in the sections below on each model and in Appendix B. The second selection, of mothers/infants that will respond to treatment, is based on the medical

Figure 4.2. Overview of Modeling Process



literature or the Maryland discharge data. For instance, for models involving corticosteroids, decreases in morbidity/mortality are designed to mimic the odds ratios from meta analyses. Again, specific details for each simulation model are given below. A mother/infant that is selected in this second selection to respond to treatment is tagged for replacement. At the end of the selection procedure, the data is partitioned into two groups - one of observations to be retained and another of observations to be replaced.

Certain criteria are retained from the data to be replaced so that they may be controlled in the replacement group. To identify which of these should be used as criteria for selection of replacement infants, correlations of each of these variables with the outcome variables (length of stay, costs) were conducted. These analyses indicate that birthweight and mode of delivery are significantly correlated with outcomes and should be used as criteria. Thus, when the selection is made, it is from a pool of similar birthweight infants who have been delivered by the same manner (except in the case where we are deliberately simulating changes in the cesarean section rate). Thus, average birthweight and cesarean section rate should remain relatively constant. For instance, for all simulation models involving EARLYINF, we control for birth weight (none of the medical or quality interventions should affect birth weight) and delivery mode. These controlling variables are supported by the high correlations between the outcome variables. Infant weight is controlled by matching weights by incremental categories. The categories range from 100g increments at the lower end of the weight scale to 250g increments at the top end of the weight scale. This is done because matching the weights exactly would be too limiting in the reselection design. Thus, while the intent of the design is to ensure that average birth weight does not change, there is a

possibility that up to a 499g difference in weight could exist between the original infant and the randomly chosen replacement. However, it is felt that using the weight categories will adequately ensure that infant birth weight is controlled in the simulation models. This assumption will be tested in the assessment of the validation of each model. Similarly, for simulation models involving EARMOM_D, we control for age by replacing mothers only that match in age. Maternal age has been shown to affect birth outcomes, particularly in premature birth, and thus we want to ensure that this is not a factor in our replacement strategy. For each model, the variables controlled during replacement are noted in Table 4.7.

The next part of the simulation design is to identify an appropriate population of mothers/infants for inclusion in the replacement set. The appropriate population varies based on the simulation model being run. As an example, corticosteroids/tocolytics are not typically administered to women experiencing PROM, and thus infants with an indication that PROM was a factor in their care are not eligible as replacements (and they are also eliminated from the selection criteria). Other criteria used to define the replacement set are specific to the quality intervention being modeled. In modeling CQI and cesarean section in mothers, the replacement set for two of the models did not include mothers with breech presentation and/or fetal distress as these are two conditions for which cesarean sections are commonly performed. In all models, the base data set used for the identification of the replacement population is EARMOM_D or EARLYINF as appropriate. After an appropriate replacement population is identified, replacements are selected using randomly generated univariate numbers.

Two final steps are necessary to rebuild the data set. First, the random selection of replacement observations is made. This is done by a matched merge of the original set of variables and the replacement set. The matched merge procedure in SAS is used to pair these two data sets by one or more common variables (DiIorio, 1991). In this case the control variables and appropriate variables for the specific simulation model are used for the matched merge. Both data sets are sorted by these variables so that the merge can take place. In addition, the replacement data set is sorted by the randomly selected number, to enact the random replacement. A matched merge can result in multiplicity of observations as every possible match from the replacement set is used (so an observation tagged for replacement can actually match with as many possible matches as exist in the replacement set). Using EXCEL, the multiple observations are eliminated and only the first, randomly selected match is retained for each of the observations selected to be replaced. Then, these "new" observations are combined with the set of original observations that was *not* selected for replacement. The resulting data set contains the same number of observations as the original EARMOM_D or EARLYINF data set.

For each simulation model, this procedure is coded separately in SAS. Validation of each of the simulation models is undertaken in two manners. First, results are monitored to assure that both the design elements and the control variables are reflected in the outcomes. For instance, if a specific cesarean section rate is specified, the actual resulting cesarean section rate is measured. After this initial validation, each simulation model is subjected to five replications. A second measure of validation is then undertaken by examining the variation on the design and control variables between the simulation runs. For each

replication, means of the outcomes variables are collected. For each set of simulations corresponding to a specific medical intervention/quality intervention model, the means of the means are collected. These are reported in Chapter 5.

Measurement

Three categories of outcome measures are considered - resource utilization, costs/charges, and health conditions/outcomes. Measurement variables are determined by the information available in the Maryland discharge data. Below we indicate the specific variables appropriate for each of these three measurement categories and the anticipated effects of each from the simulation models. In some cases, a variable may be both a measurement variable and a controlling/moderating variable (in different models). When a variable is used to control or moderate a simulation, it is assessed in the model validation process and not considered again in the results. All measurement variables are listed in Table 4.8.

Resource Utilization

Resource utilization measures involve hospital time or specific procedures/resources that are used in the course of stay. Seven variables fall under this category. Two are applicable to simulations involving either mothers or infants. These are length of stay (LOS, BLOS) and the percentage of discharges with a length of stay equal to 0 (LOS_0, BLOS_0). Nationally, in 1991, the average LOS was 2.8 days. For mothers having a cesarean section delivery, the LOS was 4.5 days (prima gravida) and 4.2 days (repeat) (CDC, 1993). Thus, in the simulations where the cesarean section rate is reduced, an associated reduction in average

Table 4.8. List of Measurement Variables

Category Variable	Definition
Resource Utilization	
(B)LOS[1]	Length of Stay
(B)LOS_0	% of observations where LOS=0
NICU_DYS	Average number of days in NICU (only for those infants who spent time in NICU)
DAYS_NUR	Average number of days in nursery (only for those infants who spent time in nursery)
INTUB	% of observations where intubation was used
VENT	% of observations where ventilation was used
RESPTHER	% of observations where respiratory therapy was used
Health Conditions/Outcomes	
DIED	% of infants who died during current hospitalization
RDS	% of infants afflicted with RDS
OTHPREM	% of infants with other conditions of prematurity
OTHRESP	% of infants with other respiratory ailments
PROM	% of (mothers with) or (infants affected by) Premature rupture of membranes
PREECL	% of (mothers with) or (infants affected by) Preeclampsia
PLPREV	% of (mothers with) or (infants affected by) Placenta previa
OTHPLAC	% of (mothers with) or (infants affected by) Other placental conditions
(B)BREECH	% of (mothers with) or (infants affected by) breech presentation
FETDIST	% of mothers with fetal distress during delivery
Costs	
(B)TOT_CHG	Total charges for stay
(B)DAL_CHG	Average daily charge for stay

[1] (B) indicates that the variable is on both mother and infant data sets

LOS is expected. The other five resource utilization measures apply only to simulations models with infants. The first two relate to utilization of specific hospital units - the average number of NICU days and the average number of days spent in the nursery. For both of these measures, only infants that had a length of stay greater than 0 were included in the base set (no immediate discharges, transfers, or deaths are included as needing either of these resources). The last three resource utilization measures refer to specific procedures that are commonly used in infants born prematurely and specifically relate to RDS and respiratory ailments. These are the use of intubation, ventilation and respiratory therapy. The Maryland

discharge data indicates whether any or all of these procedures were used or not, but not the extent to which they were used (i.e. number of days of ventilation). Thus, the indication of use is binomial and the measurement of these variables is noted as a percentage of use.

Health Outcomes

Health outcomes are typically measured in terms of morbidity and mortality. More recently, health outcomes may also include quality of life years (QALYs), measures of quality of care, and patient satisfaction. These types of measures are not encompassed by the Maryland discharge data. Thus, we will consider mortality and major sources of morbidity applicable to premature labor and delivery and premature infants. For mortality, we measure the percentage of deaths (DIED). Our major condition considered in the morbidity measurement is the presence or absence of RDS. Additionally, we consider a number of comorbidities. For infants, we note whether or not they have other common conditions due to prematurity (OTHPREM) or other respiratory ailments (OTHRESP). In the mothers, the common morbidities consider include premature rupture of membranes (PROM), preeclampsia (PREECL), placenta previa (PLPREV), and other placental conditions (OTHPLAC). Two other conditions relevant to the mother's health status include the incidence of breech presentation (BREECH) and fetal distress during delivery (FETDIST). We also consider the incidence rate of infants affected by these conditions in the mother (BPROM, BPREECL, BPLPREV, BOTHPLAC, BBREECH).

Costs

The Maryland discharge data includes a large number of charge variables. We will consider two of these charges categories - total charges for the current stay (TOT_CHG, BTOT_CHG), and average daily charges (DAL_CHG, BDAL_CHG). We use charges as a proxy for costs for these simulation models. It is recognized that the ratio of costs to charges is different for every hospital and thus that it is impossible to make a direct inference about health care system costs from charge figures. However, we can measure the anticipated magnitude of the change from the implementation of management techniques as reflected in the simulation models. For instance, in 1991, the average cost for cesarean section was $7826 and for vaginal delivery $4720, a difference of $3106 (CDC, 1993). We can compare the magnitude of our differences to these figures and consider the impact of different cesarean section rates. For instance, the CDC report notes that if the 1991 cesarean section rate had been 15%, it would have decreased the number of cesarean births by 349,000 nationwide and had a savings of more than $1 Billion in physicians fees and hospital charges (349,000*$3106=$1,083,994,000; 1993). Additionally, it should be noted that the charges encompass only hospital charges and do not reflect physician fees. Of the $3106 difference in costs reported, $611 was from physicians fees and $2495 hospital charges (CDC, 1993).

Details on Simulation Models

This simulation is designed in two stages. The first stage is to build in appropriate medical interventions that are commonly used in the premature labor and deliver process. These medical interventions are designed to decrease the incidence and/or severity of RDS

and other related respiratory ailments common to premature infants. The results of this first stage give us an idea of how medical practices can improve the outcomes of the infant population. The second stage of the simulation is to incorporate quality improvement techniques into the process. The results of this second stage provide insight as to how prevalent the improvement of outcomes could be. Thus, a synopsis of the relevant medical literature is used to direct the simulation design which is discussed in detail.

Medical Interventions

Antenatal Corticosteroids and Tocolysis - STER15

Antenatal medical interventions, the use of corticosteroids and tocolytics, are the first to be included in our model. To model these therapies, we create random variables in the dataset to represent these optional courses of treatment. Once the treatment variables are incorporated into the data, then different levels of treatment can be considered as described in the subsequent section on the details of the simulation models.

We consider the use of tocolytics and corticosteroids taken together. In our baseline model, taken strictly from the data, there is no evidence of use of either tocolytics or corticosteroids. In our model, we introduce a 15% rate of use of corticosteroids/tocolytics, in conjunction with the prevalence of use reported in the NIH Consensus Development Conference Statement (NIH, 1994). In accordance with our review of the literature, this 15% is distributed randomly among infants.

To mimic infants whose mothers had received corticosteroids prior to delivery, we model the associated decrease in incidence of RDS (OR= 0.51; 95% CI 0.42-0.61). In

patients where the random selection of therapy results in a (randomly distributed) reduction in associated morbidity, the infant is synthetically replaced with a suitable infant not experiencing the same morbidity. Thus, the outcomes of these therapies is measured using the resulting incidence of morbidity in the randomly selected replacement infants. After establishing a baseline of 15% rate of corticosteroid use, progressive increases in the rate of corticosteroid administration rate are modeled until the maximum observed rate of 60% is reached. Further simulation details are given below.

Postnatal Use of Surfactant - SURFRESC

Next, we consider postnatal care of premature infants. The therapy that we consider is the use of exogenously administered surfactant. The effect of using surfactant is to decrease the severity of RDS and to decrease the mortality rate in premature infants. The rate of use of surfactant in premature babies has increased dramatically in recent years. However, in our discharge data set, there is no evidence of surfactant therapy among the infants.

Only infants are involved in this medical intervention, so the EARLYINF data set is used for the simulation modeling (n=6211). As with the corticosteroid simulations, the simulation design for these models is based on the known medical information from the literature. Surfactant therapy is often employed using a rescue strategy, where surfactant is administered to all infants determined to have RDS. The results expected from using them have been well documented in the literature. For the rescue strategy, the 1520 infants (24.47%) that were found to develop RDS will be "treated" in the simulation model. This

rescue strategy is then compared to other strategies based on information in the literature. Further details on these models are given below.

Details on Simulation Models - Quality Interventions

Consensus Conference - Antenatal Steroids - STER30, STER45, STER60

The first quality intervention that is considered is the use of recommendations from a consensus conference. The medical intervention that is considered in conjunction with this quality intervention is the use of Antenatal Steroids. Information for the model design for this quality improvement-medical intervention combination was obtained from the Consensus Development Conference on Antenatal Corticosteroid Use. The goal of consensus conferences is to bring together specialists on a specific medical care topic, evaluate all data relevant to the topic, and compile a consensus statement that should guide future medical care practice. The consensus statement thus is a statement of aggregate opinion on the care of patients. Once the consensus statement is issued and health care providers begin to adhere to it, there should be a greater degree of common practice with respect to the medical care topic addressed.

Thus, for the use of antenatal corticosteroids, the consensus conference gave very specific guidelines for the administration of corticosteroids to all fetuses between 24 and 34 weeks of gestation based on the findings that the benefits greatly outweigh the risks (NIH, 1994, p. 13). They also note that tocolytics and corticosteroids should be administered together. The anticipated effect of corticosteroid use is to reduce the incidence of RDS as measured by the odds ratio found in the meta-analysis considered by the consensus conference

participants (OR= 0.51; 95% CI 0.42-0.61). Additionally, we consider the prevalence of use found by the consensus conference participants (estimated at 15% of premature infants nationwide) and the rate of use that is projected based on the highest degree of use reported in the consensus conference statement (60%). We use these widely accepted findings in configuring the simulation model of implementation of consensus conference guidelines in conjunction with antenatal corticosteroid administration.

First, looking strictly at information from the discharge data, there is no evidence of use of either tocolytics or corticosteroids. Thus, we base the simulation model design on the accepted medical literature on corticosteroid use and the results of its use. We configure our baseline model by introducing a 15% rate of use of corticosteroids/tocolytics, in conjunction with the prevalence of use reported in the NIH Consensus Development Conference Statement (NIH, 1994). In accordance with our review of the literature, this 15% is distributed randomly among infants in the data set. For those infants receiving corticosteroids, we model the associated decrease in incidence of RDS based on the odds ratio above. For those infants randomly selected to receive corticosteroids, the incidence of RDS is then changed. Based on the odds ratio, approximately 50% of infants that previously had RDS would not have had RDS had they received antenatal corticosteroids. Thus, infants who received corticosteroids are then again subjected to a random selection process determining whether or not they would have developed RDS. In patients where the random selection of corticosteroid therapy results in a (randomly distributed) reduction in RDS, the infant is synthetically replaced with a suitable infant not experiencing RDS. Then, the outcomes of

these therapies is measured using the resulting incidence of co-morbidity/mortality in the randomly selected replacement infants.

Once the baseline model is complete, further efforts are designed to imitate the implementation of greater adherence to the consensus conference recommendations. The maximum rate anticipated based on rates observed during the consensus conference preparation was 60%. Thus, the simulation models are designed to step up in equal increments toward a 60% rate of corticosteroid use as the rate that would be realized were the consensus conference recommendations followed. Corticosteroid use is modeled at 15% (baseline), 30%, 45% and 60%. The use of progressive, equal increments allows us to consider nonlinearities that might result in comorbidities as an unusual result of the simulation.

The specific simulation manipulations are based on two successive random selections prior to the replacement strategy being employed. First, infants are randomly selected to receive corticosteroid therapy at a rate targeted towards the goal of the simulation (15%, 30%, 45% and 60%). Second, for infants in the "treatment" group (receiving corticosteroids), those with RDS are randomly selected to not have RDS as a result of the corticosteroid administration. The rate of selection of infants to switch from having RDS to not having RDS is determined by the odds ratio relationship documented in the consensus conference proceedings (NIH, 1994). Infants selected in this manner are randomly replaced with infants that do not have RDS. The calculations of the random selection rate for RDS based on the odds ratio is contained in Appendix B. The target numbers for infants treated with corticosteroids and subsequently not afflicted with RDS are detailed in Table 4.9 which lists each of the consensus conference/corticosteroids simulation models.

Only the infants are considered in this simulation as the effects we are concerned with are those on the infants, not the mothers. Thus, the EARLYINF data set is used for this simulation. This data set contains 6211 observations and the observed rate of RDS is 24.47%. As indicated in Table 4.9, the expected rate of RDS subsequent to corticosteroid use will change in accordance with the odds ratio calculations. These expected rates will be compared to the rates seen in the simulation. The costs of corticosteroid administration are negligible (typical charge is $14.32 for a course of steroids plus $38.50 for the IV hookup).

Table 4.9. Simulation Model Design for Consensus Conference/Corticosteroids

Simulation Model	Prevalence of Use of Cortico-steroids	Target # receiving Cortico-steroids	Prior Rate RDS[1]	Expected n. with RDS prior to treatment	Treatment Rate RDS[2]	Target # infants to be replaced	Expected Posterior Rate RDS[3]
Baseline **STER15**	15%	932	24.47%	228	13.7%	100	22.86%
Implementation Phase I **STER30**	30%	1863	24.47%	456	13.7%	201	21.24%
Implementation Phase II **STER45**	45%	2795	24.47%	684	13.7%	301	19.63%
Consensus Conference "Best Practices" **STER60**	60%	3727	24.47%	912	13.70%	401	18.02%

[1] From EARLYINF - Rate of RDS for entire data set

[2] From calculations based on odds ratio; applies only to treatment group (those randomly selected to receive corticosteroids). Rate of RDS for control group would remain the same as the prior rate.

[3] Calculated population rate of RDS after treatment (combined rates, treatment and control group) - calculated as (1520-#replaced)/6211

CQI - Cesarean Section - CSECT25, (M)CSECT20, (M)CSECT15, (M)CSECT10

The second quality intervention that we consider is CQI. CQI efforts are designed to improve the mean performance. The medical intervention that is used in this portion of the

simulation is the delivery mode, measured as the rate of cesarean section. Thus, in modeling CQI and cesarean section, the simulation is based on eliminating both special cause and common cause variation. In eliminating special cause variation, we consider the primary reasons that cesarean sections are done and selectively choose to reduce the number of repeat cesarean sections so that the overall rate of cesarean section does not differ between mothers who have previously had cesarean sections and those that have not. In eliminating common cause variation, we model an improvement (reduction) in the mean cesarean section rate.

First, a baseline measure of cesarean section rates similar to that of the U.S. population (approximately 25%) during the same period as when the data was collected is taken. Then, the rate of cesarean section is reduced to reflect the generally accepted idea that 24% of all live births being delivered by cesarean section are unnecessary. Three "goals" of reduced incidence are considered. The first improvement imposed upon the system reflects an intermediate cesarean rate that would be attained during a CQI implementation. Then, the goal of reducing the overall cesarean rate to no more than 15% as given in the Healthy People 2000 study is modeled (Ventura et al., 1993, 1994a, 1994b). The final improvement level reflects a reduction down to (or below) 10% - the stretch goal advocated by Berwick (1994). Thus, these targets will set the range for the number of cesarean sections that need to be "eliminated". Table 4.10 (infants) and Table 4.11 (mothers) lists some important characteristics for each of these simulation models.

In the data, cesarean section delivery is noted in the discharge records of both mothers and infants. Thus, two sets of simulations for cesarean section rate improvement are undertaken, one for mothers (using EARMOM_D) and one for infants (using EARLYINF),

and the impact on each is measured. Of the 5,136 mothers who experienced premature labor and delivered during the hospital stay for which the discharge data was collected, 1,257 delivered by cesarean section (24.5%). For the 6,211 premature infants, 1,893 were delivered by cesarean section (30.5%). Since multiple births are excluded from these data sets, the difference in these rates is due to the difference in coding for prematurity, not in an overall difference in delivery mode.

Four sets of simulation runs are done for premature infants. The reasons that a cesarean section is done is not noted on an infants discharge record (as it is considered part of the care of the mother, not the infant). Thus, the simulation models for CQI in the infants all reflect a focus on common cause variation leading to reduction of the mean rate of cesarean section. Since the rate of cesarean sections noted in premature infants is significantly greater than that of the mothers, the first simulation model is designed to imitate a cesarean section rate for premature infants that is closer to that of the mothers delivering prematurely. The baseline for the simulation models is the population data set (EARLYINF). The first set of simulation runs is designed to bring the cesarean section rate in the premature infant population down to 25% (close to the population rate). Following this, three more simulation runs are configured by setting goals for the cesarean section rate similar to how a CQI initiative might be implemented. The second simulation model for EARLYINF reflects an intermediate cesarean section rate of 20% due to initial implementation of CQI. This is

followed by the stated quality improvement goals of a rate of 15% (Healthy People 2000) and 10% (Stretch Goal).

Table 4.10 below details the simulation models that are done. For each simulation model, the number of cesarean sections that would be expected to meet the desired rate in a sample of the same size as EARLYINF is noted, along with the imputed number to be changed and the rate of change. These percentage of cesarean sections to be changed is used as an input to the simulation model.

Table 4.10. Simulation Model Design for CQI/Cesarean Section - Infants

Simulation Model	(Desired) Cesarean Section Rate	Number of Cesarean Sections to Meet Desired Rate	Number of Cesarean Sections to be Eliminated	% of Cesarean Sections to be Changed[3]
Baseline[1] **EARLYINF**	30.5%	1893	n/a	n/a
Population[2] **CSECT25**	25.0%	1553	340	17.96
Intermediate **CSECT20**	20.0%	1242	651	34.39
Healthy People 2000 **CSECT15**	15.0%	932	961	50.77
Stretch Goals **CSECT10**	10.0%	621	1272	67.20

[1] For cesarean section, the baseline rates are from the empirical data

[2] Simulation model to bring rate down to population rate

[3] These rates of change are used in the simulation modeling.

Three sets of simulation runs are done for mothers to model the effect of quality improvement. For the mothers, the models are designed to imitate a two-phase CQI implementation, reflecting first an elimination of special cause variation and then a reduction of common cause variation. As in the infant simulation models, the target cesarean section rates are set at 20% (Intermediate), 15% (Healthy People 2000) and 10% (Stretch Goal).

The first simulation model, MCSECT20, focuses on eliminating special cause variation and meeting the 20% goal. We focus on repeat cesarean sections as a special, identifiable cause of variation in the cesarean section rate. This is supported by our data. For women who had not had a previous cesarean section (including primiparous women, who have never had a baby), we observe a cesarean section rate of 21.54%. For women who have had a cesarean section, the rate is 60.47% - clearly much higher than needed. Thus, the first model attempts to have a 20% cesarean section rate in both groups. This is done by setting the rate of change separately for women who have had previous cesarean sections (PCSECT=1) and those who haven't (PCSECT=0). At the same time, women with an indication of breech presentation or fetal distress are not considered eligible to have the delivery mode changed as these are commonly accepted reasons for cesarean delivery. Breech presentation and fetal distress are thus controlled for in this simulation model. The results of this design is that 89.5% of mothers delivering by cesarean section that had a previous cesarean section and do not have a breech presentation or fetal distress are selected for replacement. This 89.5% would be reasonable for CQI implementation focused on eliminating special cause variation. The other 10.5% of women who had previously delivered by cesarean section and not noted to have breech presentation or fetal distress may have other (valid) reasons for a repeat cesarean section. It is noted, however, that having a previous cesarean section is not controlled for during this simulation as there are too few mothers with a previous cesarean section that then do not have a cesarean section.

The second simulation model, MCSECT15, focuses on meeting the Healthy People 2000 goal of a 15% cesarean section rate. In order to meet this rate, the same rate of change

is applied to decrease repeat cesarean sections, assuming that the CQI implementation is now focused on eliminating common cause variation. Again, we control for breech presentation and fetal distress. For women without a previous cesarean section and without breech presentation or fetal distress, 73.0% who had a cesarean section are selected for replacement. This simulation model mimics how the system might implement CQI in an effort to reach the Healthy People 2000 goals.

The third simulation model, MCSECT10, focuses on reducing the cesarean section rate even further - the stretch goal promoted by Berwick. To achieve this goal, more women with breech presentations and fetal distress would have to have vaginal deliveries. If all the women without breech presentation or fetal distress delivered vaginally, the cesarean section rate would be 12.3%. Thus, in this model, breech presentation and fetal distress are not controlled and selection for replacement is made regardless of these conditions. A greater proportion is still selected from women who had previous cesarean sections (83.5%, without regard to breech presentation/fetal distress). 53.6% of women delivering by cesarean without a previous cesarean are selected for replacement.

Complete calculations for the percentages to be changed based on previous cesarean delivery, breech presentation, and fetal distress are given in Appendix B. Table 4.11 below details the simulation models that are done.

Table 4.11. Simulation Model Design for CQI/Cesarean Section - Mothers

Simulation Model	Desired Cesarean Section Rate	# of C. Sections Expected	Number of Cesarean Sections to be Eliminated			% of Cesarean Sections to be Changed[3]		
			Total	PCSECT =0	PCSECT =1	Total	PCSECT =0	PCSECT =1
Baseline[1] **EARMOM_D**	24.50%	1257						
Intermediate **MCSECT20**	20.00%	1027	230	73	157	18.29%	16.19%	89.49%
Healthy People 2000 **MCSECT15**	15.00%	770	487	330	157	38.74%	73.01%	89.49%
Stretch Goals[2] **MCSECT10**	10.00%	514	743	548	195	59.11%	53.58%	83.46%

[1] For cesarean section, the baseline rates are from the empirical data

[2] Stretch Goal Rates of Change are based on changes coming from entire sample, not just those that are not breech or have fetal distress.

[3] Calculations for these rates of change used in the simulation modeling as detailed in Appendix B.

Clinical Pathway (Local Intervention) - Surfactant - SURFTHER, SURFMIX

Finally, we model the effects of the use of a clinical pathway recommending a specific strategy for surfactant use. Clinical pathways are plans of care that outline the optimal sequencing and timing of interventions for patients with a particular diagnosis, procedure, or symptom (Ignatavicius and Hausman, 1995). Clinical pathways are usually locally developed or adopted, causing different results depending on the hospital or system in which it is adopted. We look at three sets of clinical pathways, or strategies, for surfactant administration, each of which could be adopted by the hospitals in Maryland that comprise the database.

As noted earlier, so the EARLYINF data set is used for the simulation modeling (n=6211). The two strategies that have been tested in research are modeled first - a therapeutic strategy (given to all infants under 1500g) and a rescue strategy (surfactant

administered to all infants determined to have RDS). These strategies and the results expected from using them have been well documented in the literature and this literature drives the modeling efforts. For the therapeutic strategy, where surfactant will be "administered" to all infants under 1500g, a total of 1074 infants in our data set, or 17.3%, will be affected. This is noted in the top two rows of Table 4.12. For the rescue strategy, the 1520 infants that were found to develop RDS will be treated (24.47%, first column of Table 4.12).

Table 4.12. Frequency table for Weight versus Incidence of RDS

	RDS	no RDS		
< 1000g	356 5.73%	245 3.95%	60.1968	Therapeutic
1000-1499g	287 4.62%	186 2.99%	473 7.61%	
1500g +	877 14.12%	4260 68.59%	5137 82.71%	
	1520 24.47%	4691 75.53%	6211 100.00%	
	Rescue			

Additionally, to mimic the use of a locally developed protocol we also model a combined strategy protocol that more adequately represents what some practitioners are currently doing. This mixed strategy combines the therapeutic strategy using a lower weight guideline (1000g) and the rescue strategy based on the presence of RDS in higher weight infants. As seen from Table 4.12, noted with double lines, this mixed strategy would apply to 1765 or 28.4% of infants. This strategy is used in an effort to maximize the effect of the

therapy by giving surfactant as soon as possible while minimizing the administration of surfactant to infants that do not need it (those between 1000 and 1499g that did not develop RDS, n=186). Using this strategy also enables us to test the use of simulation modeling for clinical pathways where the outcomes are not yet well documented.

The specific simulation manipulations are based on two successive random selections prior to the replacement strategy being employed. First, infants are selected to receive surfactant on the selection criteria for the appropriate model. Second, for infants in the "treatment" group (receiving surfactant), those with RDS are randomly selected to have a less severe incidence of RDS. Infants selected in this manner are randomly replaced with infants from the group of infants with a lower severity RDS. The target numbers for infants treated with surfactant and subsequently experiencing a less severe case of RDS are detailed in Table 4.13 which lists each of the protocol/surfactant simulation models.

Table 4.13. Simulation Model Design for Protocol/Surfactant

Simulation Model	Selection Criteria	Prevalence	Number Anticipated to be treated
Rescue Strategy **SURFRESC**	evidence of RDS	24.47%	1520
Therapeutic Strategy **SURFTHER**	under 1500g	17.29%	1074
Mixed Strategy **SURFMIX**	under 1000g or evidence of RDS	28.42%	1765

Severity of RDS is modeled as a combination of comorbidities (OTHRESP) and indication of use of procedures related to respiratory ailments (INTUB, VENT, RESPTHER). Several models were tested, including cluster analytic models and algorithms related to the

specific combinations of comorbidities and procedures. The best severity model found quite simply was designed around the number of comorbidities/procedures indicated in the discharge data. Four categories of severity were used, denoted 0,1,2, and 3, to represent the number of comorbidities/procedures. Only 5 observations had indication of all four comorbidities/procedures and they are included in severity category 3.

The four categories were validated by comparing the means on several outcomes measures, including morbidity, birth weight, length of stay, number of days in the NICU, number of days in the nursery. As the severity increased, the means on these outcomes measures were progressively worse. Differences between these categories were tested using ANOVA and found to be significant (Table 4.14). Detailed ANOVA results are contained in Appendix B. Thus, these categories of severity were deemed to be discriminatory and are used in the simulation design as selection criteria for infant replacement.

Table 4.14. RDS Severity Categories

Number of Comorbidities/ Procedures	Number of Observations	BLOS **	NICU_dys[1] **	NICU[2] **	DAYS_ NUR[1] *	BDIED *	BBRTH _WT **
0	431	15.6	24.1	8.3	3.23	2.55	2350
1	380	27	32	19.4	3.05	3.42	2038
2	539	32.9	38.5	23.4	5.58	6.31	1663
3-4	170	48.2	48.4	35.9	4.22	7.06	1489

[1] NICU_dys and DAYS_NUR are calculated as the average number of days only for infants that spent some time in the NICU or nursery (with those that did not eliminated).

[2] NICU is calculated using those that did not spend any time in the NICU as 0.

* indicates significant differences between groups, $p<.05$.

** indicates significant differences between groups, $p<.01$.

Thus, in total, the three combinations of medical interventions/quality management techniques result in four sets of models - STER, CSECT, MCSECT, and SURF. Each has three or four levels of quality improvement implementations, for a total of 14 different models that are detailed in Table 4.7. Five iterations of each model were performed and the results tabulated for each of the outcomes measures previously discussed. The results of these modeling efforts are presented in the next chapter.

Statistical Analysis

For each of the 14 models that are detailed in Table 4.7, statistical comparisons to the base data set (EARLYINF or EARMOMD) are made. Comparisons are made between each individual iteration and the base data set, and then the average of the p values is taken as an indication of the level of significance. Additionally, for variables that are found to change significantly and for variables that are needed to be controlled, additional statistical comparisons are made between models. For these comparisons, the statistical analysis is done for each iteration of a given model with each iteration of the model being compared (25 comparisons in all) and then the average of the p values is taken. The results of the statistical comparisons are detailed in Chapter Five.

The statistical comparisons are conducted using Wilcoxon Rank Sum tests. The Wilcoxon Rank Sum tests is appropriate for independent groups needing nonparametric tests (Schlotzhauer and Littell, 1995). This corresponds to a Kruskal-Wallis test (SAS Institute, 1988). In our case, nonparametric tests are necessary as the observations cannot be assumed to be from a population with normal distributions for any of the variables. In fact, it has been

noted that many health care related data are nonnormally distributed (Morrisey et al., 1995). As for independence, we will assume independence between samples for these tests. It should be noted that given the random selection and replacement techniques, independence cannot be assured. However, given the large sample sizes and the use of empirical data this assumption seems reasonable. While some specific data may be duplicated within a given iterated data set (with a maximum occurrence of any one observation of two times), the desire to incorporate underlying characteristics of the population is greater than the risk of lack of independence.

CHAPTER 5

RESULTS

> Only a little of the first fruits of wisdom, only a few fragments of the boundless heights, breadths, and depths of truth, have I been able to gather.
>
> Martin Luther

This chapter summarizes the major results of the simulation modeling efforts. The simulation modeling process involved modeling three medical intervention/quality management technique combinations for infants and one medical intervention/quality management technique for mothers at three or four different levels of quality improvement each, for a total of 14 different models. A synopsis of these different simulation models and the major performance measures is presented in Table 5.1. We can review this table to determine which of the models performs the best overall with respect to resource utilization as measured by length of stay, health outcomes as measured by mortality, and with respect to costs as measured by daily/total charges. Other measures are not comparable across models due to differences in design that determined the outcome.

For all measures, significance is tested using the Kruskal-Wallis test which corresponds to the Wilcoxon Rank Sum tests for nonparametric, independent data. Using health care data, especially discharge data, poses some unique difficulties in the calculation of confidence intervals. One study noted that the variation in health care cost data is commonly very large, frequently due to bimodal distributions (Morrisey et al., 1995).

Table 5.1 Synopsis of Simulation Results

Medical Intervention	Quality Technique	RESULTS SIMULATION MODEL	Validation BBRTH_WT	Validation CSECT	Utilization LOS	Health Outcome DIED	Costs DAL_CHG	Costs TOT_CHG
INFANTS								
Baseline		EARLYINF	2299.1	30.5%	12.53	4.77%	7,169	10,255
Cesarean Section	CQI	CSECT25	**2299.3**	25.3% ***	12.52	4.75%	7,173	10,260
		CSECT20	**2297.6**	20.5% ***	12.46	4.84%	7,155	10,207
		CSECT15	**2298.4**	15.7% ***	12.36	4.77%	7,087 *	10,117 **
		CSECT10	**2297.5**	11.2% ***	12.32 **	4.81%	7,064 **	10,078 ***
Corticosteroid Use	Consensus Conference	STER15	**2299.0**	30.5%	12.44	4.81%	7,116	10,167
		STER30	**2299.4**	30.4%	12.27	4.88%	7,006	9,991
		STER45	2302.2	30.4%	12.15	4.93%	6,923	9,855
		STER60	**2298.8**	30.1%	12.12	5.02%	6,924	9,849
Surfactant	Clinical Pathways	RESC	**2298.4**	**30.5%**	**12.28**	4.80%	**6,999**	9,937
		THER	**2299.2**	**30.5%**	**12.31**	4.86%	**7,028**	10,021
		MIX	**2298.1**	**30.5%**	12.31	4.81%	**7,025**	9,981
MOTHERS								
Baseline		EARMOM_D		24.5%	3.53		1,419	3,830
Cesarean Section	CQI	MCSECT20		20.8% ***	3.39 ***		1,367 **	3,726 **
		MCSECT15		16.5% ***	3.20 ***		1,297 ***	3,557 ***
		MCSECT10		12.2% ***	3.09 ***		1,259 ***	3,484 ***

* $p <= 0.10$
** $p <= 0.05$
*** $p <= 0.01$
BOLD indicates no significant difference with average p value >= 0.95.

The common measure of resource utilization across all models is length of stay. For infants, there are significant decreases in the length of stay for all steroid models and all surfactant models. For the cesarean section models, the decreases in length of stay are not significant for the higher rates of cesarean sections (CSECT25 and CSECT20). Once the cesarean section rate is decreased to 15%, then there is a significant decrease in length of stay. Of all the models, the steroid models appear to do the best on this measure and get progressively better as the rate of steroid use increases. For mothers, the decrease in length of stay is significant for all decreasing levels of cesarean section. A total decrease of 12% is realized cumulatively through the MCSECT10 model.

The common measure of health outcomes across all infant models is mortality. For mothers, where no deaths were reported in the reference set, mortality is not considered. For the three sets of infant models, mortality either did not change significantly (cesarean section) or increased (steroid and surfactant models). This appears to be a result of the modeling process and possibly an artifact of the simulation's random replacement methodology. More will be said about this in the discussion.

The cost measures, average daily charge and average total charge, are collected for each model and are not affected by the modeling process. For mothers, both daily and total charges are progressively lower with each model, with the greatest improvement seen when decreasing from a 20% cesarean section rate to a 15% cesarean section rate. For infants, the improvements in daily and total charges are not significantly different from the reference set until the cesarean section rate is no higher than 15%. For both the corticosteroid models and the surfactant models, the charges decrease significantly for all models.

The detailed results are discussed in two sections. In the first section, we report the details of the simulation modeling design efforts, both for the two medical interventions that are not embedded in the data and for the quality management technique. This discussion focuses on the control and experimental variables to assess the validation of the models. To assist in this discussion, Table 5.2 details the experimental and control variables for the different models and Table 5.3 reviews the random replacement process in more detail. In the second section of the results we discuss the impact of quality management techniques on outcomes, considering each of the medical intervention/quality management technique combinations as a separate section. Detailed performance measurements for each of the models is presented in Tables 5.4-5.7. There is discussion of each of the three categories of outcomes measures - resource utilization, health outcomes, and charges. In both sections the discussions are ordered in the same order as the sequential transformation process - diagnosis, delivery, neonatal care.

Results of the Modeling Efforts - Model Validation

This section focuses on the simulation modeling design efforts, both for the two medical interventions that are not embedded in the data and for the quality management technique. One of the major modeling challenges was to incorporate medical interventions into the simulation design for the two medical interventions - corticosteroid use and surfactant therapy - that were not encompassed by the data. These two medical intervention simulations are designed using information from the medical literature to direct the simulation. The second major modeling challenge was to reflect the potential use of quality management

techniques in the models at each stage of the process. These simulations were designed to assess the expected impact of the quality management technique and adjust the simulation to reflect this impact. We then assess the results of the simulation in response to these anticipated effects.

We discuss the design efforts in the order that they would occur according to the sequential medical care transformation process. Thus, we first discuss the diagnosis of premature labor, including the use of corticosteroids and the impact of consensus conference implementation. Second, we discuss the modeling of CQI efforts in conjunction with delivery mode. Third, we review the models of surfactant use and the implementation of clinical pathways.

The results are considered from the point of view of model validation, comparing the differences between the medical interventions on the control and experimental variables. For the control variables, we would expect no significant differences. All models were controlled for provider (as a proxy for socioeconomic status and case mix); this variable was modeled to control only for exact matches, so there is no variation between the reference sets (EARLYINF and EARMOM_D) and the simulation groups. For the experimental variables, we would expect that differences occur between models, but that within a given set of models any differences between the expected values of the experimental variables and the realized values of the experimental variables reflects an ability to meet the quality improvement goals. Table 5.2 details the comparisons for the control and experimental variables.

Table 5.2. Results of Experimental and Control Variables for All Models

INFANTS		STEROIDS				CESAREAN SECTION				SURFACTANT		
CATEGORY		MODEL				MODEL				MODEL		
Variable	EARLY INF	STER 15	STER 30	STER 45	STER 60	CSECT-25	CSECT-20	CSECT-15	CSECT-10	RESC	THER	MIX
VALIDATION												
Control												
FREQ	6211	6211	6211	6211	6211	6211	6211	6211	6211	6211	6211	6211
PROVIDER	20.2	20 2	20 2	20 2	20 2	20 2	20 2	20 2	20 2	20 2	20 2	20 2
WT	11.5	11 5	11 5	11 5	11.5	11.5	11 5	11.5	11 5	11 5	11 5	11 5
BBRTH_WT	2299.14	2299.03	2299.38	2302 21	2298.84	2299.25	2297.55	2298.38	2297.5	2298 4	2299 16	2298 12
BCSECT	30.48%	30 42%	30 38%	30 27%	30 14%					30 48%	30 48%	30 48%
RDS	24.47%									24 47%	24 47%	24 47%
Experimental												
TARGET STER		15 00%	30 00%	45 00%	60 00%							
GET_STER		14 82%	30 19%	44 32%	60 13%							
TARGET RDS		22 93%	21 38%	19 84%	18 29%							
NEW_RDS		22 93%	21 43%	20 03%	18 42%							
RDS	24.47%	23 07% *	21 80% ***	20 53% ***	19 34% ***							
TARGET CSECT						25 00%	20 00%	15 00%	10 00%			
NEW_CS						24 99%	19 82%	14 99%	10 06%			
BCSECT	30.48%					25 26% ***	20 47% ***	15 73% ***	11 23% ***			
TARGET SURF										24 47%	17 29%	28 42%
GET_Surf										24 47%	17 29%	28 42%
EXP_SEV										0 38	0 44	0 35
SEVERITY	62.00%									0 51	0 55	0 51

MOTHERS		CESAREAN SECTION - MOTHERS		
CATEGORY		MODEL		
Variable	EARMOM_D	MCSECT20	MCSECT15	MCSECT10
VALIDATION				
Control				
FREQ	5136	5136	5136	5136
PROVIDER	20.2	20 2	20 2	20 2
AGE	26.4	26.4	26.4	26.4
PROM	21.98%	21.98%	21.98%	21.98%
Experimental				
TARGET CSECT		20 00%	15 00%	10 00%
NEW_CS		20 14%	15 07%	10 17%
CSECT	24.47%	20 76% ***	16 54% ***	12 20% ***
BREECH	9.03%	9.03%	9 03%	6 38% ***
FETDIST	20.04%	20.04%	20.04%	19 00%
PCSECT	7.54%	5 19% ***	5 43% ***	4 90% ***

* p <= 0 10
** p <= 0 05
*** p <= 0 01
BOLD indicates no significant difference with average p value >= 0 95

Antenatal Corticosteroids (STER models)

Antenatal corticosteroids can be administered to women experiencing premature labor prior to delivery. Corticosteroids are administered in an effort to promote lung development in the fetus. Information on the effects of the use of corticosteroids has been well documented in the medical literature and thus is used as the basis of the simulation model. In particular, the meta-analysis of previous research on corticosteroid use reported by the NIH as a part of their consensus conference provides the details of the effects of corticosteroids on morbidity and mortality (NIH, 1994). The model is designed to duplicate the effects of corticosteroids on infants experiencing RDS as expressed by the odds ratio of 0.51 (95% CI 0.42-0.61) (Crowley, 1994). Thus, in the models we would expect to see that corticosteroid use decreases the incidence of RDS to 51% of what it is in the group of infants not treated with corticosteroids. An analysis of this effect is part of the validation of the simulation design. The target incidence of RDS for each model is calculated from the odds ratios and the anticipated effect.

To validate the steroid models, four variables are considered. Two control variables, infant weight at birth and the cesarean section rate, and two experimental variables, steroid use and incidence of RDS are measured. The consensus conference report of 15% of corticosteroid use provides the initial level of use of corticosteroids for the model design (STER15); incremental increases of 15% are then considered. Validation of the medical intervention modeling was done with the initial STER15 model. Validation of the quality management tool, use of consensus conferences, was done by focusing on the experimental variables.

Infant weight is controlled by matching weights by incremental categories. It is assumed that using the weight categories will adequately ensure that infant birth weight is controlled in the simulation models. This assumption is tested by measuring the actual infant birth weight. For the baseline data, EARLYINF, the average birth weight is 2299.14 grams. In the STER15 model, the average birth weigh is 2299.03 grams, a difference of 0.11grams (less than .004 ounces), not statistically different from the reference set. Additionally, we consider the cesarean section rate. Cesarean section is not explicitly controlled in the STER15 model by controlling for delivery mode in the simulation. Preference was given during the selection process to selection by matching delivery mode unless there was no match; then an alternative match without the same delivery mode was allowed. As a result, the cesarean section rate in STER15 the cesarean section rate was 30.42% (in EARLYINF it is 30.48%). This slight decrease was significant, though this difference of 0.06% in a population of 6211 translates to less than four fewer cesarean section deliveries. Thus, it seemed that the model adequately represented the cesarean section rate without unduly constraining the random selection process. Thus, for the two control variables we consider the STER15 model is valid and subsequent models were run using the same model with changes only in the experimental variables. We did note that the cesarean section rate did go down progressively (probably due to the relationship between RDS and cesarean section, discussed later); were future modeling efforts to occur, this aspect of modeling would be more specifically addressed.

Next we consider the experimental variables of the rate of steroid use and the realized incidence of RDS. In the STER15 model, the goal was to have 15% steroid use. The actual

amount of steroid use based on randomized selection was 14.82% and ranged over the simulation runs from 14.1% to 15.2% (standard deviation .0041; 95% CI .1446-.1518). Since the target rate of steroid use is contained within the confidence interval we are assured that the model design is valid. For the incremental increases of steroid use (STER30, STER45, and STER60), the confidence interval around the mean of the randomly generated steroid rate also includes the target rate, again validating the models.

The expected incidence of RDS is the rate of RDS that was expected given the random selection process in the modeling effort (prior to replacement). The actual incidence of RDS is that measured after the replacement of infants used to represent the administration of steroids. For the STER15 model, this translates into an expected rate of RDS of 22.93% (see Appendix B for details on the calculations with respect to odds ratios). In STER15 the average expected incidence of RDS (variable NEW_RDS) was 22.93%, exactly as expected. For all other steroid models, the 95% confidence interval of the expected incidence of RDS based on the random selection process always included the target based on the odds ratios, so the selection process was validated. Finally, we consider the actual incidence of RDS reflected in the model results after the random replacement portion of the modeling. For STER15, the incidence of RDS was 23.07%, with a 95% confidence interval of 22.96-23.18%. This confidence interval does not include the target of 22.93%; we see similar results in the subsequent steroid models. As with the cesarean section rate, it is likely that this is an artifact of the constraints on the random replacement process. However, it should be noted that this translates into an average of 9 infants that had RDS in the simulation models that would not have been expected to have RDS according to the consensus conference rates

odds ratios. Thus, more conservative results are presented by the modeling efforts than reflected in the consensus conference reports. For the purpose of constructing and evaluating models, we thus consider the designs valid (they work as we had anticipated) though they possibly underestimate the effects of the results.

Cesarean Section (MCSECT and CSECT models)

The second phase of our sequential transformation model refers to the delivery of the infant. This phase affects both mothers and infants, so two separate modeling efforts are undertaken that emphasize the delivery mode used. Delivery can be either vaginal or by cesarean section. For these models, there is no need to model the medical intervention as the use of cesarean section is already embedded in the discharge data. Thus, given the measured rate of cesarean section use as represented in the reference sets, we consider incrementally lower cesarean section rates and consider the results of each. Again we consider both control variables and experimental variables in assessing the modeling efforts.

Cesarean Section - Mothers (MCSECT Models)

For mothers the control variables considered were age, and incidence of premature rupture of membranes (PROM). All of these were modeled to control only for exact matches, so there is no variation between the reference set (EARMOM_D) and the simulation groups. Additionally, for the first two incremental decreases of the cesarean section rate (from 25% to 20% and from 20% to 15%) the matches were also controlled for breech presentation and for fetal distress; women whose discharge records indicated delivery complicated by one of these

conditions and had delivered by cesarean section were considered ineligible for random selection for vaginal delivery. Thus, there is no difference on the rates of these conditions for the MCSECT20 and MCSECT15 models.

For the experimental variables, two variables were considered. The first is whether or not a mother has had a previous cesarean section (variable PCSECT), used for the model design as a factor in the random selection of a change in delivery mode. Appendix B details the calculations on the selection rates, which vary based on the value of PCSECT. The evaluation of this is obtained by comparing the target cesarean section rate for the model (incrementally 20%, 15% and 10%) to the selection rate (20.14%, 15.07%, and 10.17%). These selection rates represent the average cesarean section rate after the random selection process. In all three of the models, the target rate is included in the 95% CI for the selection rate, so the selection process is deemed valid. The second variable to be considered in how well the models worked is the realized cesarean section rate (variable CSECT), which represents the average cesarean section rate after the replacement phase of the modeling. On average, the cesarean section rate was 20.8%, 16.5%, and 12.2%. In none of the simulation models did the confidence interval for the actual rate include the target rate nor did they overlap with the confidence intervals for the selection rates. As discussed before, this is probably an artifact of the random replacement process but is also a relevant research finding.

Cesarean Section - Infants (CSECT Models)

Control variables for the infants on the cesarean section models included provider and birth weight. As before, the control for the weight variable was done using incremental weight categories (variable WT) and the validity of this control is assessed by measuring

average birth weight (variable BBRTH_WT). For the CSECT25 model, this method of controlling for birth weight was adequate (the 95% CI for bbrth_wt is 2298.2g-2300.3g, including the mean from the reference set of 2299.1g). For the subsequent models with lower cesarean section rates, the average birth weight was slightly low (upper control limits on the 95% CI of 2297.6, 2299.0, and 2298.5, all below the mean of the reference set).

The only experimental variable for the cesarean section models is the cesarean section rate. The target rates for each of the four models were 25%, 20%, 15%, and 10% respectively. First, we compare these target rates to the expected rate (variable NEW_CS). The target rates were included in the 95% CI for the selection rate for CSECT25, CSECT15 and CSECT10, so the selection process is deemed valid even though the selection rate for the CSECT20 models was slightly lower than 20% (UCL 19.9%). The second variable to be considered in how well the models worked is the realized cesarean section rate (variable BCSECT), which represents the average cesarean section rate after the replacement phase of the modeling. On average, the cesarean section rate was 25.26%, 20.47%, 15.73% and 11.23%. In none of the simulation models did the confidence interval for the actual rate include the target rate. For CSECT25, the confidence interval for the actual cesarean section rate did overlap with the confidence intervals for the selection rates, though for the other models there was no overlap. As discussed before, this probably results due to the random replacement process.

Postnatal Use of Surfactant (RESC, THER and MIX models)

The third segment of the sequential transformation process modeling premature labor and delivery is postnatal care of the infant. During this portion of the health care process, surfactant is often used to improve the health of infants born prematurely, particularly those born with respiratory ailments such as RDS. In healthy infants, surfactant is secreted within the lungs and keeps the alveolar walls from sticking together, thus preventing the lungs from collapsing. Surfactant therapy artificially introduces surfactant into the infants' lungs until further pulmonary development is sufficient for surfactant to be produced naturally. The literature documents the effects of surfactant therapy as a reduction in severity of illness rather than a reduction in incidence of illness. Three models based on different clinical pathways are considered, representing a rescue strategy, and therapeutic strategy, and a mixed strategy (models RESC, THER and MIX). These models vary based on the target group that is selected to receive surfactant.

The control variables for these models are weight, delivery mode, and RDS. These controls reflect the fact that delivery happens prior to surfactant therapy (and thus cannot be changed) and that the literature asserts that no difference in incidence of RDS occurs from surfactant administration. In these models, delivery mode and RDS were controlled by using exact matches only for replacement. Thus, for these measures, there is no difference between the models and the reference sets. Weight is controlled in the same manner as for the other models. For THER models, this method of controlling for birth weight was effective (95% CI 2299.10g-2299.22g, includes reference set mean of 2299.14g). For the RESC and MIX

models, we again see confidence intervals that are slightly lower than the reference set mean (UCLs of 2298.6 and 2298.5 respectively).

Two experimental variables are considered, rate of surfactant administration and severity (an index designed to measure the severity of the case of RDS). Surfactant administration is designed based on a inclusion basis rather than a random selection basis. Thus, the rate of surfactant administration is constant within each model. 24.5% of the infants in the RESC model received surfactant, 17.3% of the infants in the THER model received surfactant, and 28.4% of the infants in the MIX model received surfactant. The models assure that these rates are met. The second experimental variable, severity, reflects respiratory comorbidities or procedures that are indicated by the infants' discharge record. Severity is measured on a scale of 0-3; for infants that start with a severity of 1,2, or 3 they are randomly selected to be replaced by infants from a group of infants with a lower severity index. The expected severity index is calculated by proportionately replacements of severity from lower groups. The details of these calculations are given in Appendix B. In actuality, the resulting severity indices on average are much higher than expected. For instance, for the RESC models, the expected severity index average would be 0.376; the actual average was 0.51. However, the 0.51 average does represent an improvement in severity from the original scale average of 0.623 even though it is not as low as expected.

Assessment of the Replacement Strategy

The final discussion regarding the modeling design efforts is focused directly on the replacement strategy. As mentioned previously, often the selection rate of the simulation

model was accurate, but after replacement the actual rate was not within range of the target. This is true for all four of the sets of models presented. Table 5.3 below presents an assessment of the replacement strategy by calculating the percentage of non-replacement for each model. For the surfactant models, the percentages represent only the infants with the highest severity level (severity=3) that could not be matched. It is clear that for the simplest set of simulation models, those for cesarean sections for infants that have the fewest control variables and only one experimental variable, the replacement strategy works better. As the models get more complex and require more controls, the worse they perform compared to the targets set, and that is reflected in the increasing percentages of the number of selected cases that could not be replaced within the constraints of the model.

The question remains, is this non-replacement truly a problem, or does it reflect both the degree of variation in the population and the potential for management techniques to fall short of the goals set (i.e. the difference between goals set and implementation results)? We entertain this question in the discussion and suggest further research to investigate this matter in Chapter Six.

Impact of Quality Improvement Management Techniques

Given valid models, the outcomes variables may be considered and compared to the EARLYINF or EARMOM_D population dataset. We consider outcomes in three categories - resource utilization, health outcomes, and charges. These outcomes are detailed in Tables 5.4-5.7 with significance noted in bold and italics. Outcomes noted in bold indicate that the average for the original dataset (EARMOM_D ro EARLYINF) is not encompassed by a 95%

Table 5.3. Review of Random Replacement Procedure

Medical Intervention	Quality Technique	SIMULATION MODEL	Variable Selected for Change	Average # to Change [1]	Number Matched	Number not Matched	% not Matched
INFANTS							
Baseline		EARLYINF					
Cesarean Section	CQI	CSECT25	CSECT	340.6	322	18.6	5.46%
		CSECT20		662	621.6	40.4	6.10%
		CSECT15		957.4	911.6	45.8	4.78%
		CSECT10		1268	1195.8	72.2	5.69%
Corticosteroid Use	Consensus Conference	STER15	Steroids/RDS	96	87.2	8.8	9.17%
		STER30		188.8	166	22.8	12.08%
		STER45		287.2	256.2	31	10.79%
		STER60		376	318.6	57.4	15.27%
Surfactant	Clinical Pathways	RESC	Surf/Severity	170	145	25	14.71%
		THER		152	121	31	20.39%
		MIX		193	168	25	12.95%
MOTHERS							
Baseline		EARMOM_D					
Cesarean Section	CQI	MCSECT20	CSECT	222.8	190.8	32	14.36%
		MCSECT15		482.8	407.4	75.4	15.62%
		MCSECT10		734.6	630.2	104.4	14.21%

[1] For surfactant, the number selected to be changed represents only the highest severity infants (index=3)

confidence interval around the outcome measure. Outcomes noted in italics indicate that the measure is significantly different from the preceding model, indicating significant incremental improvement within the implementation of the quality management technique. This section is organized in the same order as the sequential transformation model, taking each combination of medical intervention/quality management technique separately and discussing outcomes from the appropriate set of models. The focus of this section is initially on the difference between the reference set and the models and then turns to differences between the different levels of the models.

A wide variety of outcomes measures are considered. For resource utilization, we review the length of stay (BLOS/LOS), the average number of days spent in the NICU (NICU_DYS; only for infants that did spend time there), the average number of days spent in the nursery (DYS_NURS; again, for infants that were in the nursery at all), the patient disposition (BLOS_0/LOS_0 or BTRANS/TRANS) and the procedures utilized during care (INTUB, VENT, RESPTHER). Health outcomes for the infants include mortality rate (BDIED), morbidity rates where it is not dictated by the model (RDS), and the rates of comorbidities (OTHRESP and OTHPREM). For mothers, the health outcomes include mortality (DIED) and many comorbidities including the incidence of placenta previa or other placental conditions (PLPREV and OTHPLAC). Additionally, conditions commonly occurring in conjunction with premature labor or complications of delivery are also measured, including preeclampsia, breech presentation, fetal distress, and previous cesarean section (PREECL, BREECH, FETDIST, and PCSECT). Finally, costs are measured using average total charges and average daily charges as a proxy for costs for both mothers and infants.

Consensus Conference/Antenatal Corticosteroids (STER models)

The first set of models simulate the use of antenatal corticosteroids within the guidelines of the 1994 consensus conference recommendations. The model design for these simulations included incrementally increasing rates of steroid use combined with an associated decreases in the incidence of RDS as noted in the consensus conference report. Looking at differences between the steroid models and the EARLYINF reference set gives us an indication of the level of use of steroids that begins to make a significant difference. Then, looking at differences between the models indicates the value of incremental increases in the use of steroids as recommended. One of the keys that we are looking for here is the presence of nonlinear relationships revealed by the simulation models. These relationships would provide insight into possible recommendations of different levels of implementation of quality management techniques. For each type of difference, we consider measures of resources utilization, then health outcomes, and finally costs. These differences reflect the potential results from reducing RDS in the population. The outcomes measures for these models are tabulated in Table 5.4.

Differences from EARLYINF Reference Set

The main outcome measure considered for resource utilization is length of stay. Use of corticosteroids provides an immediate effect on length of stay in infants; the STER15 model shows a significant decrease in the length of stay from 12.53 days to 12.44 days. This corroborates with infants who have RDS requiring longer hospital stays. With respect to the use of specific use of different units, the NICU and the nursery, the average number of days

Table 5.4. Results from Steroid Models.

CATEGORY / Variable	Description	MODEL: EARLY INF	STER 15	STER 30	STER 45	STER 60
VALIDATION						
Control						
FREQ		6211	6211	6211	6211	6211
PROVIDER	Provider identifier	20.2	20.2	20.2	20.2	20.2
WT	Weight - categorical	11.5	11.5	11.5	11.5	11.5
BBRTH_WT	Infant weight at birth	**2299.14**	**2299.03**	**2299.38**	2302.21	**2298.84**
BCSECT	CSect Rate	30.48%	30.42%	30.38%	30.27%	30.14%
Experimental						
TARGET STER	TARGET Steroid Rate	0.00%	15.00%	30.00%	45.00%	60.00%
GET_STER	Realized/Actual Steroid Rate	0.00%	14.82%	30.19%	44.32%	60.13%
TARGET RDS	TARGET RDS Rate	0.00%	22.93%	21.38%	19.84%	18.29%
NEW_RDS	Expected RDS Rate	0.00%	22.93%	21.43%	20.03%	18.42%
RDS	Realized/Actual RDS Rate	24.47%	23.07% *	21.80% ***	20.53% ***	19.34% ***
RESOURCE UTILIZATION						
BLOS	Length of Stay	12.53	12.44	12.27	12.15	12.12
NICU_DYS	Days spent in NICU	22.01	21.95	21.54	21.40	21.37
DAYS_NUR	Days spent in Nursery	3.49	3.46	3.50	3.43	3.43
N_NURS	Number in Nursery	1942	1933	1918	1886	1899
N_NICU	Number in NICU	2043	2035	2022	1990	2006
BTRANS	% Transferred	17.29%	17.21%	17.23%	17.11%	17.14%
BLOS_0	% with LOS=0	11.24%	11.21%	11.20%	11.15%	11.12%
INTUB	Intubation	16.97%	16.62%	16.33%	15.88%	15.72% *
VENT	Ventilation	13.72%	13.38%	13.16%	12.79%	12.58% *
RESPTHER	Respiratory Therapy	8.44%	8.29%	8.16%	8.06%	7.91%
HEALTH OUTCOMES						
Infant Conditions						
BDIED	% Died	4.77%	4.81%	4.88%	4.93%	5.02%
OTHRESP	Other respiratory cond.	23.33%	23.44%	23.39%	23.38%	23.50%
OTHPREM	Other cond of prematurity	4.67%	4.70%	4.66%	4.59%	4.64%
Related to Maternal Conditions						
BPROM	PROM	0.74%	0.75%	0.74%	0.74%	0.77%
BPLPREV	Placenta previa	0.14%	0.14%	0.14%	0.15%	0.11%
BPREECL	Preeclampsia	0.42%	0.42%	0.43%	0.44%	0.41%
BOTHPLAC	Other placental cond	0.32%	0.33%	0.32%	0.33%	0.33%
BBREECH	Breech	0.52%	0.50%	0.51%	0.52%	0.50%
BFETDIST	Fetal Distress	4.64%	4.67%	4.68%	4.68%	4.76%
SEVERITY	Severity Index	0.62	0.62	0.61	0.60	0.60 *
COSTS						
BDAL_CHG	Average Daily Charge	$7,169	$7,116	$7,006	$6,923	$6,924
BTOT_CHG	Average Total Charge	$10,255	$10,167	$9,991	$9,855	$9,849

* $p <= 0.10$

** $p <= 0.05$

*** $p <= 0.01$

BOLD indicates no significant difference with average p value >= 0.95.

spent in the NICU significantly decreases from 22.01 days to 21.54 days when the steroid use is 30% of all premature infants. Additionally, the number of infants that require NICU care drops from 2043 to 2035 with 15% steroid use. The average number of days spent in the nursery never decreases significantly, though a significantly fewer number of infants spend time in the nursery (1933 with 15% steroid use compared to 1942 without steroids). Additional changes in resource use are reflected in the decreased number of infants transferred (significantly higher with 15% steroid use; 17.21% instead of 17.29%) and the reduced number of infants who do not stay in the hospital overnight once steroids are used at least 45% (length of stay=0; only 11.15% of infants instead of 11.24%). Additionally, we consider the utilization of specific procedures common to infants with respiratory problems. Significant decreases in the use of intubation, ventilation and respiratory therapy are seen immediately with an introduction of steroid use in the population at a 15% rate.

We also consider health outcomes as a result of the use of corticosteroids. There was a significant increase in the mortality rate from 4.77% to 4.81% at 15% steroid use. This is a bit puzzling, and may be an artifact of the replacement strategy. With a decrease in RDS an increase in mortality is not expected; mortality should either be constant or also decrease. Comorbidities associated with other respiratory conditions or other conditions of prematurity also increase significantly with 15% steroid use. However, few of the comorbidities related to maternal conditions increase significantly, and usually only with much higher rates of steroid use. Infants show significantly higher incidence of problems related to PROM or placenta previa in the mother with 60% steroid use; for problems related to preeclampsia the incidence is significantly greater when steroids are used in 45% of the population. Only fetal distress,

which has the highest rate of incidence to begin with, significantly increases with only 15% use of steroids. Complications from breech presentation would be expected to *decrease* significantly with steroid use.

Finally, we consider the impact of corticosteroid therapy on charges. For both average daily charge and average total charge, the reduction of RDS associated with the use of steroids results in a significant reduction of charges.

Incremental Differences from Implementation of Consensus Conference Recommendation

In addition to considering the differences between the original data and the simulation models, we want to consider differences between the different levels of the simulation. For instance, though length of stay improved incrementally with an introduction of steroid use, it does not significantly decrease again until the rate of steroid use is 45%. There is also a significant difference between 30% steroid use and 60% steroid use. Significant differences in the steroid models tend to occur with increments of 30% of use rather than 15%. These differences are noted on Table 5.4 with asterisks. For instance, while the number of infants that went to the NICU or nursery was significantly different with 30% steroid use, there is no further impact on these outcomes measures until steroids are used at a rate of 60% and the associated decrease in RDS is that much greater.

For the commonly used procedures, there continues to be incremental improvements with 30% steroid use and for 45% steroid use for intubation and ventilation. With 60% use, however, no further decreases are seen for intubation and ventilation. The use of respiratory therapy does decrease significantly from 15% to 45% steroid use and from 30% to 60%

steroid use, again reflecting the pattern of significant differences associated with 30% increases in steroid use rather than 15% increases. We see a very similar effect in terms of average daily charges and average total charges. Incremental improvements are not seen for every 15% increase in steroid use, but 30% increases in steroid use does significantly decrease charges. Thus, in terms of resource utilization and costs, recommendations would be to implement steroid use in increments of at least 30% to gain the improvements.

With the exception of the mortality rate, there are not great continuations of differences in the health outcomes associated with the use of steroids. Mortality does continue to increase, and also shows significant differences with 30% increases in steroid use. It is also interesting to note that while there is a continual progression of increasing mortality rate, this progression is not linear. The greatest increases are seen between STER45 and STER60, and then between STER15 and STER30. The only other health outcome affected by progressively increasing rates of steroid administration is the presence of other respiratory conditions, which also increases showing a significant difference when the steroid rate is 60%. The worsening of the health outcomes measures other than RDS is not expected in conjunction with steroid use or in conjunction with decreased incidence of RDS and may be an artifact of the sampling procedure.

CQI/Cesarean Section (CSECT and MCSECT models)

The second set of models considered considers the delivery mode and the improvements made in conjunction with attempts to lower the cesarean section rate using continuous quality improvement methods (CQI). Both infants and mothers are modeled for

these simulations. The infant simulations are the simplest of the models considered, with an incrementally decreasing rate of cesarean section as the only experimental variable. The mother-focused simulations are more complicated in design due to limiting factors that determine when a mother needs a cesarean delivery. Looking at differences between the cesarean section models and the reference sets gives us an indication of the levels at which the decrease of cesarean section rate begins to make significant differences. Then, looking at differences between the models indicates the value of incremental decreases in the cesarean section rate. For each type of difference, we consider measures of resources utilization, then health outcomes, and finally costs. The outcomes measures for these models are tabulated in Table 5.5 for infants and Table 5.6 for mothers.

Differences from EARLYINF reference set and incremental differences due to COI implementation

The most interesting result from the cesarean section simulations comes from the health outcomes measures. As the cesarean section rate drops, the incidence of RDS also drops significantly. Significant differences are found for the first two 5% drops in the cesarean section rate (CSECT25 and CSECT20). Subsequently, the CSECT15 and CSECT10 models show no significant improvement in the rate of RDS compared to the CSECT20 models. These results are interesting for two reasons: first, they indicate that there may be a significant relationship between RDS and cesarean section that has not been investigated extensively, and second, they would suggest that in decreasing the cesarean

Table 5.5. Results from Infant Cesarean Section Models

CATEGORY / Variable	Description	MODEL EARLYINF	CSECT-25	CSECT-20	CSECT-15	CSECT-10
VALIDATION						
Control						
FREQ		6211	6211	6211	6211	6211
PROVIDER	Provider Identifier	20.2	20.2	20.2	20.2	20.2
WT	Weight - categorical	11.5	11.5	11.5	11.5	11.5
BBRTH_WT	Infant weight at birth	2299.14	2299.25	2297.55	2298.38	2297.6
Expenmental						
TARGET CSECT	TARGET CSECT RATE	30.48%	25.00%	20.00%	15.00%	10.00%
NEW_CS	Expected CSect Rate		24.99%	19.82%	14.99%	10.06%
BCSECT	Actual CSect Rate	30.48%	25.26% ***	20.47% ***	15.73% ***	11.23% ***
RESOURCE UTILIZATION						
BLOS	Length of Stay	12.53	12.52	12.46	12.36	12.32 **
NICU_DYS	Days spent in NICU	22.01	22.15	22.39	22.25	22.36
DAYS_NUR	Days spent in Nursery	3.49	3.48	3.42	3.27	3.23
N_NURS	Number in Nursery	1942.0	1935.6	1912.4	1907.8	1897.4
N_NICU	Number in NICU	2043.0	2035.6	2017.6	2008.6	1998.6
BTRANS	% transferred	17.29%	17.20%	16.96%	16.93%	16.68%
BLOS_0	% length of stay = 0	11.24%	11.10%	10.95%	10.82%	10.69%
INTUB	Intubation	16.97%	16.58%	16.33%	15.95%	15.91%
VENT	Ventilation	13.72%	13.48%	13.35%	13.34%	13.36%
RESPTHER	Respiratory Therapy	8.44%	8.32%	8.35%	8.12%	8.08%
HEALTH OUTCOMES						
Infant Conditions						
BDIED	% died	4.77%	4.75%	4.84%	4.77%	4.81%
RDS	RDS	24.47%	24.21%	23.87%	23.78%	23.66%
OTHRESP	Other respiratory cond.	23.33%	23.27%	23.27%	23.01%	22.83%
OTHPREM	Other cond of prematurity	4.67%	4.85%	5.04%	5.01%	5.07%
Related to Maternal Conditions						
BPLPREV	Placenta previa	0.14%	0.13%	0.09%	0.08%	0.07%
BOTHPLAC	Other placental cond	0.32%	0.26%	0.24%	0.24%	0.18%
BPROM	PROM	0.74%	0.77%	0.82%	0.80%	0.83%
BPREECL	Preeclampsia	0.42%	0.35%	0.32%	0.27%	0.24% *
BBREECH	Breech	0.52%	0.49%	0.43%	0.39%	0.39%
BFETDIST	Fetal Distress	4.64%	4.47%	4.34%	4.12%	3.96%
COSTS						
BDAL_CHG	Average Daily Charge	$7,169.08	$7,173.49	$7,154.65	$7,086.81 *	$7,064.33 **
BTOT_CHG	Average Total Charge	$10,255.37	$10,259.69	$10,207.03	$10,116.51 **	$10,078.38 ***

* p <= 0.10
** p <= 0.05
*** p <= 0.01
BOLD indicates no significant difference with average p value >= 0.95.

section rate 20% is sufficient (in premature infants, and with respect to RDS). Considering the other health outcomes, fewer effects are seen. No significant changes in the mortality rate are noted for any of the models. For other respiratory conditions a significant decrease is noted with CSECT15 and for other conditions of prematurity a significant decrease is noted with CSECT20. No subsequent significant changes are noted for either of these comorbidities. For comorbidities in the infants that are related to conditions in the mother, most show a significant reduction with a reduction to a 25% cesarean section rate (preeclampsia, fetal distress, placenta previa, and other placental conditions). Many of these conditions continue to show incremental improvements when the cesarean section rate is decreased by 10% (every other model). The other comorbidities are exceptions to these results. Breech presentation is not significantly reduced until the cesarean section rate is 20% or lower, and PROM actually increases as the cesarean section rate decreases, with the first significant increase noted when the cesarean section rate is 20%. The reductions in these measures do not suggest that there would in fact be such reductions from cesarean sections (since they are conditions that happen independently of and prior to the cesarean section). Rather, they imply the degree of changes in practice that might need to occur in the medical community before the realization of quality improvement goals of reducing the cesarean section rate as low as 20% is possible.

Resource utilization measures for these simulations are consistent with the drop in RDS and the use of resources associated with the care of infants with RDS. Resource utilization in terms of length of stay and demands for nursery or NICU resources are not significantly affected until the cesarean section rate drops to at least 20%, and in some cases

less than that. The length of stay is significantly lower when the cesarean section rate is 15% or less. The need for NICU days, which is a big component of care of infants with RDS, drops when the cesarean section rate is 20%, but then does not drop significantly again (similar to the pattern as for RDS). Resource utilization measures focused on the use of procedures responds more quickly to the drop in cesarean section rate, again a reflection of the drop in incidence of RDS. The use of intubation, ventilation and respiratory therapy are all significantly lower when the cesarean section rate is 25% rather than the 30.5% rate from the reference set. Further improvements in the use of procedures are sparse; less intubation is required when the cesarean section rate drops to 15%; less respiratory therapy when the cesarean section rate is 10%. Use of ventilation is never significantly less than the 13.5% seen with a cesarean section rate of 25%.

Finally, we consider the changes in charges associated with a lower cesarean section rate for infants. No significant difference is seen until the cesarean section rate is 15% and there are no further decreases. Average daily charges decrease from $7,169 in the reference set to $7,087 when the cesarean section rate is 15%. Average total charges decrease from $10,255 to $10,117 over the same associated drop in cesarean sections. Since the cesarean section procedure is more likely to affect the charges of the mother than the infant, this is not a surprising result.

Table 5.6. Results from Mother Cesarean Section Models.

CATEGORY / Variable	Description	MODEL EARMOM_0	MCSECT20	MCSECT15	MCSECT10
VALIDATION					
Control					
FREQ		5136	5136	5136	5136
PROVIDER					
AGE	AGE	26.4	26.4	26.4	26.4
PROM	PROM	21.98%	21.98%	21.98%	21.98%
Expenmental					
TARGET CSECT	TARGET CSECT RATE	24.47%	20.00%	15.00%	10.00%
NEW_CS	Csect Selected Rate	24.47%	20.14%	15.07%	10.17%
CSECT	Actual CSect Rate	24.47%	20.76% ***	16.54% ***	12.20% ***
RESOURCE UTILIZATION					
LOS	Length of Stay	3.53	3.39 **	3.20 ***	3.09 ***
TRANS	% Transferred	0.39%	0.38%	0.37%	0.43%
LOS_0	% with LOS = 0	0.97%	0.99%	1.04%	1.09%
HEALTH OUTCOMES					
DIED	% Deaths	0.00%	0.00%	0.00%	0.00%
PLPREV	Placenta previa	3.41%	3.10%	2.36% ***	2.42% ***
OTHPLAC	Other placental cond	5.43%	5.21%	4.68% *	4.67% *
PREECL	Preeclampsia	6.45%	6.03%	5.01% ***	5.19% ***
BREECH [1]	Breech	9.03%	9.03%	9.03%	6.38% ***
FETDIST [1]	Fetal Distress	20.04%	20.04%	20.04%	19.00%
PCSECT	Previous C Section	7.54%	5.19% ***	5.43% ***	4.90% ***
COSTS					
DAL_CHG	Average Daily Charge	$1,418.60	$1,366.94 **	$1,297.43 ***	$1,259.22 ***
TOT_CHG	Average Total Charge	$3,830.21	$3,725.89 **	$3,557.25 ***	$3,483.63 ***

[1] Breech and Fetal Distress are used as control variables in MCSECT20 and MCSECT15

* p <= 0.10

** p <= 0.05

*** p <= 0.01

BOLD indicates no significant difference with average p value >= 0.95.

Differences from EARMOM_D reference set and incremental differences due to CQI implementation

The differences in resource utilization for mothers as the cesarean section rate is modeled to decrease are quite interesting. Significant decreases in the length of stay are noted with each successive drop in the cesarean section rate. This would support conventional knowledge that women who have cesarean sections have longer hospital stays. However, it is interesting to note that the improvement does not occur in a linear fashion with each successive drop in the cesarean section rate. Length of stay decreased by 4% with a decrease to a 20% cesarean section rate, by 5.4% between MCSECT20 and MCSECT15, and then 3.4% between MCSECT15 and MCSECT10. The percentage of women with a length of stay of 0 days also increases (and contributes to shorter average length of stay), with significant increases noted with a drop to a 20% rate and again when the rate drops to 15%.

Significant differences in the health outcomes are also realized. For placenta previa, preeclampsia, and other placental conditions, significant decreases are noted both for the MCSECT20 model and the MCSECT15 model. No further improvements are made after that; in fact both placenta previa and preeclampsia increase with a yet lower cesarean section rate (though the increase is not statistically significant). The incidence of breech presentation and fetal distress are controlled for in all but the MCSECT10 model; for that model they show significant reductions. This would reflect the model design and indicate that these conditions would have to significantly decrease as factors in the delivery process for the cesarean section rate to approach 10% (or even 12%).

Finally we consider the charges associated with the mother's discharge records and the effects of lowering the cesarean section rate. Both average daily charges and average total charges decrease significantly with each successive decrease in the cesarean section rate as would be expected. Average daily charges decrease from $1,419 in the reference set to $1,297 when the cesarean section rate is 15%. Average total charges decrease from $3,830 to $3,557 over the same associated drop in cesarean sections. Again, the interesting aspect is not the decrease but the non-linearity of the decrease. The drop in cesarean section rate from 20% to 15% results in a much greater percentage change in charges (5.1% less for average daily charge and 4.5% less for average total charge) than either the prior drop (3.6% less and 2.7% less, respectively) or the subsequent drop (3.0% and 2.1% less, respectively).

Clinical Pathways/Postnatal Use of Surfactant (SURF models)

The final set of models consider the use of surfactant in neonates to promote lung function. Here, we model three different strategies for surfactant administration in an effort to mimic the implementation of clinical pathways. Thus, there are really three separate models rather than a progression of models. We will consider the overall differences compared to the EARLYINF reference set and also differences between the models, but not in any particular order of models. Table 5.7 details the outcome measures for these models.

Table 5.7. Results from Surfactant Models

CATEGORY / Variable	Description	MODEL EARLY INF	RESC	THER	MIX
VALIDATION					
Control					
FREQ		6211	6211	6211	6211
PROVIDER	Provider identifier	20.2	20.2	20.2	20.2
WT	Weight - categorical	11.5	11.5	11.5	11.5
BBRTH_WT	Infant weight at birth	2299.14	2298.4	2299.16	2298.12
BCSECT	CSect Rate	30.48%	30.48%	30.48%	30.48%
RDS	RDS rate	24.47%	24.47%	24.47%	24.47%
Expenmental					
TARGET SURF	TARGET % Receiving Therapy	0.0%	24.5%	17.3%	28.4%
GET_Surf	% Receiving Therapy	0.0%	24.5%	17.3%	28.4%
EXP_SEV	Expected Seventy Index	0.62	0.38	0.44	0.35
SEVERITY	Realized/Actual Seventy Index	0.62	0.51 ***	0.55 ***	0.51 ***
RESOURCE UTILIZATION					
BLOS	Length of Stay	12.53	12.28	12.31	12.31
NICU_DYS	Days spent in NICU	22.01	21.08	20.94	21.11
DAYS_NUR	Days spent in Nursery	3.49	3.54	3.50	3.50
BTRANS	% Transferred	17.29%	17.01%	17.16%	16.92%
BLOS_0	% with LOS=0	11.24%	10.97%	11.16%	11.05%
INTUB	Intubation	16.97%	13.01% ***	14.26% ***	12.77% ***
VENT	Ventilation	13.72%	10.60% ***	11.74% ***	10.41% ***
RESPTHER	Respiratory Therapy	8.44%	7.31% **	7.72%	7.28% **
HEALTH OUTCOMES					
Infant Conditions					
BDIED	% Died	4.77%	4.80%	4.86%	4.81%
OTHRESP	Other respiratory cond.	23.33%	20.47% ***	21.15% ***	20.25% ***
OTHPREM	Other cond of prematunty	4.67%	5.09%	4.96%	5.24%
Related to Maternal Conditions					
BPROM	PROM	0.74%	0.72%	0.73%	0.73%
BPLPREV	Placenta previa	0.14%	0.14%	0.14%	0.14%
BPREECL	Preeclampsia	0.42%	0.47%	0.47%	0.48%
BOTHPLAC	Other placental cond	0.32%	0.34%	0.32%	0.32%
BBREECH	Breech	0.52%	0.55%	0.50%	0.52%
BFETDIST	Fetal Distress	4.64%	4.69%	4.59%	4.67%
COSTS					
BDAL_CHG	Average Daily Charge	$7,169.08	$8,999	$7,028	$7,025
BTOT_CHG	Average Total Charge	$10,255.37	$9,937	$10,021	$9,981

* $p <= 0.10$

** $p <= 0.05$

*** $p <= 0.01$

BOLD indicates no significant difference with average p value >= 0.95.

Differences from EARLYINF Reference Set

All three of the surfactant models resulted in significant differences on all measures of resource utilization except the average number of days spent in the nursery. The decrease in the length of stay was nearly 2% for the rescue strategy, and 1.7% for the other two strategies. On average, infants spent less time in the NICU as well. Transfers were down while the percentage of infants with a length of stay equal to zero increased. As for the use of the procedures, the utilization of each went down. This is a reflection of the severity index configuration and the simulation model design. Since those procedures are components of the severity index and severity is randomly decreased, infants are replaced with others with fewer of the procedures. So, it is expected that the use of these procedures would drop.

For health outcomes, again significant differences are found for all three of the models for both mortality and the major comorbidity measures (OTHPREM, OTHRESP). Similar to the procedures utilized, OTHRESP is a component of the severity index, so the decrease in incidence of other respiratory conditions is a reflection of the design. For mortality and other conditions of prematurity, however, there are declines in performance. Mortality increases for each model as does the incidence of other conditions of prematurity. For health outcomes related to maternal conditions, the incidence of PROM decreases for both the rescue and the therapeutic strategies; breech presentation displays a significant increase in incidence for the same two models. Other placental conditions shows a significant increase for the rescue strategy and no change for either of the other models. Both preeclampsia and fetal distress are significantly higher for all three strategies. Since these two conditions occur prior to delivery (and thus prior to the administration of surfactant), this result reflects a difference in

these conditions with respect to severity. Infants with higher severity are replaced with those with lower severity, and these results indicate that the mothers of those infants with lower severity have a higher percentage of preeclampsia and fetal distress. It is most likely, however, that this is a reflection of the data that is used and the limited ability to note comorbidities. In infants with more severe cases of RDS it is likely that the limited data fields are used to note comorbidities and conditions more relevant to the infant's immediate care than preeclampsia and fetal distress.

Differences Between Surfactant Models

For a final assessment of the surfactant models, we consider which surfactant strategy is best for the population as reflected in the model outcomes. Notations of the best strategy are made in Table 5.7 with asterisks; a double asterisk indicates the best strategy as well as significant differences from either of the other strategies. The outcomes measures that are driven by the model design - INTUB, VENT, RESPTHER, and OTHRESP are eliminated from this analysis; the comparative results for these measures is directly proportional to the number of infants considered eligible to receive surfactant based on the selection criteria.

On resource utilization measures, the rescue strategy clearly does the best on the number of infants with a length of stay equal to zero (significantly different) and also performs the best on overall length of stay (12.28 days, 0.03 better than either of the other strategies). The therapeutic strategy performs the best on the average number of days spent in the NICU, and the mixed strategy is best with respect to the percentage transferred. Length of stay is

probably the most important of these measures, which would support the rescue strategy as being the best with respect to resource utilization.

Considering health outcomes is a little more difficult. None of the strategies performed well with respect to mortality. The therapeutic strategy clearly performed the best with respect to other conditions of prematurity and also for the incidence of fetal distress and breech presentation. The rescue strategy did better with respect to PROM. For health outcomes, the mixed strategy did not perform as well as either of the others.

Finally, considering costs, the best model with respect to average daily costs and average total costs was the rescue strategy, though all were significantly better than the EARLYINF average. In general, then, comparing the three models, the mixed model clearly was not advantageous to use. The therapeutic model appeared to be better for some health outcomes, but overall the rescue strategy seemed best when applied in a system-wide fashion. In the discussion chapter we consider the effect of a global application of a clinical pathway, as was modeled, versus the difference of something more locally driven.

CHAPTER 6

DISCUSSION, CONCLUSIONS AND FUTURE RESEARCH

> We shall not cease from exploring, and the end of all our exploring will be to arrive where we started and know the place for the first time.
>
> T.S. Eliot

This chapter discusses the implications of the results detailed in Chapter Five. We focus on some of the major thrusts of the work and some of the more surprising or counter-intuitive results from among the myriad small insights and conclusions that can be gleaned from the numerous output measures. In particular, we note how our results provide insight as to the utility of simulation techniques used in conjunction with large empirical datasets in health care management. This in turn provides some inkling of the advantages of this methodology in any service industry where appropriate data is available.

In Chapter Two, we presented a problem statement focused on the development of new methodologies to investigate the effects of quality management techniques on outcomes in large populations, and followed that with three research objectives and appropriate propositions. In this chapter, we first consider three general categories of conclusions, mirroring the objectives presented in the second chapter: those related to methodology and the modeling process, those related to the implementation of quality management techniques, and those related specifically to the health care environment outcomes. Each category is addressed separately, although there are some overlaps and broader conclusions that are

highlighted later. Subsequently, we return to the original problem statement and highlight the broader implications of this research, focusing on how this research in particular and this type of research in general can lend managerial insights to both health care public policy makers and service operations managers and researchers. The final section of this chapter addresses future research topics that can be addressed as an extension of this current work.

Discussion of the Methodology

Objective 1: Explore whether or not we are able to use large population data sets for simulation modeling in service operations management, using health care as a specific case for study.

Research Question 1: How can we design simulation models to adequately investigate processes in service industries using simulation of large populations?

Proposition 1.1: Using a **process management view** leads to better analysis of service outcomes.

Proposition 1.2: Using **random data selection and replacement** allows us to investigate the large-scale population effects of specific interventions on a process.

This research was highly exploratory in nature, particularly with regard to the methodology used. Thus, a detailed discussion of the simulation modeling process and the results of the models is warranted. We consider our discussion in light of two major facets of the methodology: (1) the use of a random selection/random replacement technique and (2) the use of empirical data as the input to the simulation modeling efforts. Validation issues were discussed in Chapter 4. In this section we consider how well the model results reflect the design and what the implications of these results are with respect to indications for future modeling efforts.

Random Selection and Random Replacement Modeling Technique

The simulation design was unique in that it used a combination of random selection and random replacement to assess the effects of medical and quality interventions in a health care process. Random selection, in this case of infants or mothers to receive a medical treatment, is common to many simulation models and is not unique here. The selection process worked as designed. The use of random replacement is an unusual approach in health care. We consider both the appropriateness of using this technique and the issue of non-replacement in our discussion.

Random Replacement as a Methodological Technique

We developed the random replacement technique in response to some of the shortcomings of classic medical research methods using double-blind randomized controlled trials. One of the difficulties with these classic methods is that they assess the effects of only one variable at a time. After initial studies are done. it is often the interactions of effects and the co-morbidities that are of further interest. Alternatively, the random replacement approach allows a model design focused on the major interventions and demonstrated effects under the assumption that all other related effects are represented adequately in the sample. Thus, random replacement can provide a more realistic reflection of reality when assessing the additional co-morbidities, utilization measures and cost measures. It can, however, lead to some confounding, such as we saw in our cesarean section models where the major effects were to change the incidence of RDS. Many of the other outcomes changed as well, but the changes appear to be an associated result of decreased RDS rather than the direct quality

improvement effort. However, we can at least assess these related effects better than if every comorbidity were also tightly controlled. Two other concerns of double-blind randomized controlled trials are the potential ethical problems associated with experimentation and the high costs of conducting these studies. Both of these problems are minimized with the use of simulation modeling.

It should be noted that the alternative to this random replacement would be to individually model each effect that is noted in the literature and in the data. For instance, the change in the cesarean section rate in mothers could be modeled by separating women with PROM from those who don't and then considering other moderating health care conditions, then randomly selecting for a change in delivery mode, and then considering each possible separate outcome both independently and together. While this may be a more traditional approach, and one frequently used with Bayesian or Markov analysis, it has three specific weaknesses in this case. First, modeling each trait or effect assumes independence. This is an unrealistic assumption in the health care arena. It is the rare patient who has only one medical condition or conditions that are in no way related (some physicians would contend that this would indeed never happen). Secondly, this requires that you have distinct knowledge of the independent effects in each case, many of which are not documented and yet clearly cannot be ignored. Thirdly, the proliferation of possible combinations of co-morbidities translates into too few cases of any given combination of conditions. Thus, the cell sizes of each possible end state are so small that meaningful analysis cannot be undertaken. Therefore, when appropriate data exists, as in this case, random replacement seems appropriate.

Another argument in support of the random replacement technique comes from the results we get when we control more of the selection variables. An example of this is seen in the surfactant simulation models, where the replacements are controlled for delivery mode, incidence of RDS, provider, and weight. In these models, the results are far off of the goal (comparing the expected severity index to the actual severity index). In these cases, given all the constraints on the type of patient, in many of the cases no good replacement could be found. However, by controlling fewer variables we contend that the incidence of the conditions are still reflected in the general population but may not be in the same combinations as before. This seems to indicate that there is a tradeoff between controlling for separate traits and measurement of related effects, and indicates that there may be situations when fewer controls can allow for better measurement of associated traits and conditions.

The Issue of Non-replacement

In Chapter Five, Table 5.3 presented the results of the random replacement and in particular the degree of non-replacement. The non-replacement exists depending on the ability to match based on the criteria variables. There was never complete replacement of all observations selected for change, regardless of the model. This degree of non-replacement is difficult to interpret, however. These results lead us to consider whether or not this non-replacement is a fatal flaw in the simulation modeling efforts or an interesting result that can lead to further insights.

We contend that it is not a fatal flaw; rather, it accurately reflects reality in two senses. First, the observations that are not replaced may be unique enough that no reasonable match

for them can be found. This is encouraging in the sense that it supports our underlying notion that there is inherent variation in health care processes that cannot be removed (McLaughlin, 1996). Essentially, this explicitly identifies those same cases which would have too few observations if they were modeled in an alternative manner, but allows them to be included in the results in the appropriate proportions. Secondly, it can be viewed as a reflection of the difference between goals of quality management techniques and the results often achieved when the techniques are implemented. This view is supported by the increasing gap between the expected rate of the experimental variable and the actual rate of the experimental variable as the quality goals become more ambitious. It is often quite easy to realize initial system improvements and much more difficult to continue to realize the same degree of improvement as more of the patient population must be engaged. As we note in the later section on managerial insights, these results also have implications for the implementation of quality improvement programs.

In the end, the use of the random replacement technique seems well-founded and seems to give us insight into other effects that are seen in the general population. These results would prompt us to continue to use this methodology in the future, with one caveat: you need to have data that can appropriately link the processes to the outcomes. For instance, as discussed in the next section, the surfactant results are not very revealing and it seems that this is largely because it was difficult to mimic the expected results due to the type of data that is available. We also contend that while we have used health care as an illustration, this technique would be valid in any service industry in which appropriate data is available.

Use of Simulation Modeling with Health Care Data

The other highly exploratory portion of this investigation was the use of empirical health care data as an input into the simulation. A great deal of health care related data exists in this country and health care administrators are often bemoaning the fact that it seems to overwhelming in volume and too poorly organized to be useful. This investigation allowed us to explore the usefulness of the specific discharge data base format that we used. This happens to be one frequently used for analysis. We discussed the data from the perspective of model design, quantity of data, and distributions presented in the data.

In general, using this health care data was helpful but limiting. For the corticosteroid models the data was quite useful. In these models, the information embedded in the data was collected on the same parameters as that reported in the consensus conference documents (mortality, incidence of RDS). Thus, the model design could accurately reflect the state of medical knowledge. However, we see the limits of this data with the surfactant models. Given the nature of the data we designed a severity index that best reflected the effects of surfactant use as noted in the literature. The literature was more specific about reduction in the amount of each resource needed rather than whether or not the resource was used, which is all that our data afforded us. Thus, the indexing was quite rough and the model did not seem to accurately reflect the anticipated effects.

One of the positive things about using discharge data bases and other large health care data sets is the ability to have access to large amounts of data. We have already noted in the section about the replacement strategy that the large variation in the population drives a need for very large sample sizes when researching health care processes. Large sample sizes allow

for a high number of variation in the end results due to many different combinations of comorbidities and use of procedures. These databases, while limited in the information they contain, are useful in doing large scale analyses that would be prohibitively expensive, time-consuming, and perhaps unethical using randomized trials or other retrospective techniques such as chart reviews. There are many other service industries - hospitality, transportation, financial services - that have large data bases that may also be untapped and appropriate for this methodology.

A final concern with the use of health care data rests in non-normal distributions of the outcomes. This leads to a number of problems in both modeling and assessment. In terms of modeling, the biggest challenge is deciding which distributions should be used; the random replacement technique eliminates this difficulty. This is comparable to the use of other empirical distributions in simulations rather than assuming normality. While only a preliminary assessment of the resulting distributions was done, standard normality tests clearly indicated that there were non-normal distributions in our results. This also means that the use of significance tests becomes tricky, perhaps in part due to very long tails on many distributions and bimodal distributions in the population as is often found in service industries. In the final analysis, we find that the model design process was very useful and the results pointed us to some interesting conclusions. At the same time, the lack of ability to do statistical tests makes it a difficult methodology. It is possible to continue to explore this methodology using other methods of testing significance, such as SUrvey DAta ANalysis (SUDAAN) as we discuss in the section on future research.

Two other notes about the simulation methodology are of interest here. First, the simulation models yielded many nonlinear results. This suggests that simulation is indeed a good methodology, that we gain insights from using simulation that might not be found using other methodologies. Second, the simulation modeling efforts conducted during this research effort were inherently conservative as the effects of the medical and quality interventions are possibly underestimated (as we see with the incidence of RDS in the steroid models). This is a conservative result in comparison with the literature, and is particularly heartening since a common criticism of simulation models is that they are too optimistic due to broad assumptions. Again, these results could then be viewed as an accurate reflection of the differences expected when the techniques are used in the general population rather than under clinical trial conditions (highly controlled and well-documented; the "best possible" results).

In total, considering the model design, the selection and replacement techniques, and the amount and configuration of data, this health care investigation has enabled us to test a new methodology that appears promising for health care and should be appropriate to use in many service industries where large population datasets exist. This methodology is the primary contribution of this exploratory research.

Discussion of the Implementation of Quality Management Techniques

Objective 2: Model quality management techniques in service operations management in order to demonstrate the impact of these techniques on a large population and provide insight for policy formation, again using health care as the example.

Research Question 2: How does the degree of implementation of the quality improvement techniques affect the amount of improvement in efficiency and the outcomes of the service operations process?

Proposition 2.1: Using **linearly progressive quality improvement targets** enables us to determine whether or not the effects of the quality management techniques are linear.

Proposition 2.2: Using these exploratory methods provides insight as to the **comparable impact of different quality management techniques** (consensus conference, CQI, and clinical pathways) used on the same process.

The results of this investigation also provide insight into the relevance and usefulness of implementation of quality management techniques over a large population. Three major conclusions are reached as a result of the modeling efforts. First, there are differences between the implementation of different types of quality improvement efforts as reflected in the specification of the models. Second, there are important non-linearities associated with the progressive implementation of the quality improvement techniques. These non-linearities have implications that are both general to quality improvement efforts and specific to health care policy. Third, as the improvement targets progressively increase there is an associated inability to meet the goals. Discussion on each of these points follows. Although this investigation is done in the health care arena, we consider the main findings related to these quality techniques to be generalizable to service operations management.

Modeling Quality Improvement Techniques

In this investigation, three quality improvement techniques commonly used in the health care arena were considered: CQI, consensus conferences, and clinical pathways. The modeling efforts for CQI (in conjunction with cesarean section) and consensus conferences (in conjunction with corticosteroid use) worked well and provided insightful results. On the other hand, the modeling efforts for clinical pathways were much less useful. In part, this is

due to the large number of control variables and experimental variables needed to do the clinical pathway modeling.

These results may also be a reflection of the level at which each quality improvement method is applicable. Consensus conferences are aimed at directing practices on a system-wide basis; thus, using data from a whole state to model the effects is appropriate and reveals true differences. While CQI is often implemented on a local level, the modeling efforts seem to indicate that the anticipated effects can be modeled in the same way for every provider. If the technique has gained a level of sophistication and acceptance that ensures that the effects of its implementation are expected to be similar regardless of the environment then the system-wide modeling effort is appropriate and provides useful insights.

Clinical pathways, however, are typically implemented as local efforts at practice standardization and thus may not be accurately represented by a global or system-wide model. The differences in implementation between different providers could be different enough that it could warrant modeling efforts on the level of the individual provider rather than the system. An alternative model that investigates provider differences might produce more valid and generalizable results with respect to clinical pathways but would be extremely costly to implement at that level of disaggregation and data in sufficient volume is unlikely to be available.

These results suggest that the appropriateness of using this new methodology depends upon the match between the scope of the data and the scope of implementation of the quality improvement technique. In the literature review, we discussed the difference between quality efforts focused on process control versus those focused on process design. The methodology

was successful for both process control efforts (CQI) and process design efforts (consensus conferences) and thus could be widely applied to all types of quality improvement efforts that meet the data requirements previously noted.

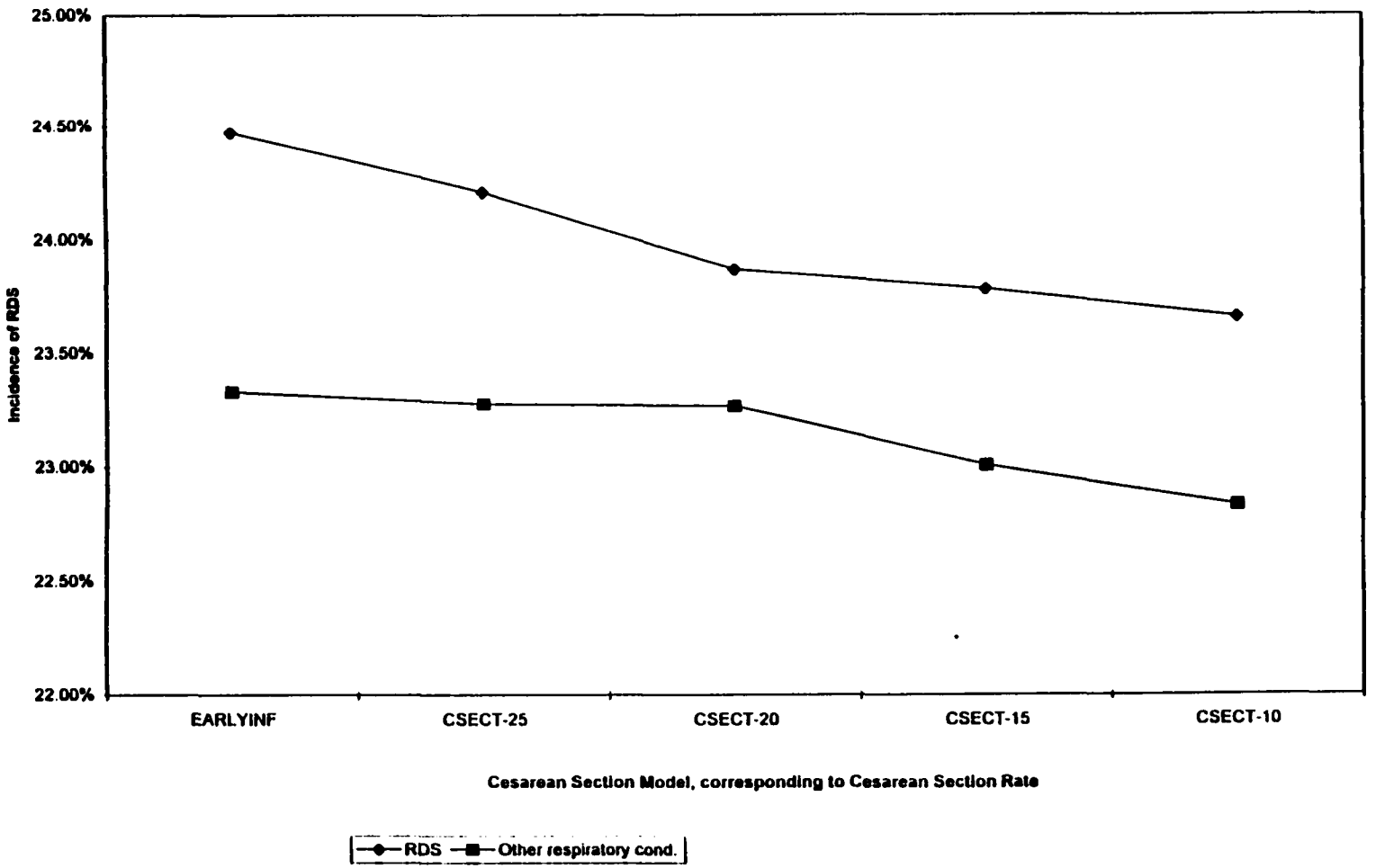
Non-linear Improvements

One of the most important findings of this work is that there are many improvements that exhibit non-linear improvements even in the face of linear design in the quality improvement efforts. Even when the rates are adjusted to reflect the actual rate of the experimental variable rather than the expected rate, there in a difference in the rates at which the changes are realized. One example of this is with the cesarean section rate and the incidence of RDS and of other respiratory conditions (Figure 6.1).

For the first two improvements in the cesarean section rate, to a rate of 20%, there are associated improvements in the incidence of RDS. Subsequently, however, there is little improvement in the incidence of RDS. This would lead you to believe that with respect to significant morbidity most of the gains that can be made have been realized without a huge effort to decrease the caesarean section rate. However, for other respiratory conditions, there is little effect until the cesarean section rate drops below the 20% mark; only after improvements to a rate of 15% are there significant drops in other respiratory conditions.

Other variables and other models show these non-linearities as well. One that is notable is that for the cesarean section rate in mothers. The greatest gains in length of stay and charges (both decreases) are when the rate of cesarean section drops from 20% to 15%. Reductions both prior to that and subsequently did not have as great an effect. In addition to

Figure 6.1 Change in Incidence in RDS with Successive Decreases in the Cesarean Section Rate



a general implication regarding the use of quality improvement programs, there is also a more specific implication for health care policy implied in these results. There is an apparent lower bound to the cesarean section rate that is revealed by these simulations. This lower bound has not been revealed in previous research. Health policy discussions have focused on trying to reduce the cesarean section rate as much as possible, not trying to determine what would be the lowest practical rate. This result is another important contribution of this research.

These non-linearities have important implications for health care managers trying to implement quality programs. These results show that the targets that you set for your quality improvement program (such as cesarean section rate, degree of steroid use) should be tied to various desired outcome levels. Depending on the outcomes, the ideal target level may be different from one's *a priori* calculations.

Progressive Improvement Targets

Finally, we consider the gap between the target rate of the experimental variables and the realized rate as a result of the modeling effort. In all models, the actual rate realized is lower than the target rate due to non-replacement of specific cases, as discussed previously. More interesting, however, is that this gap tends to increase as progressively higher improvement targets are set. As greater improvements are simulated, the less able the model is to meet the goals. The implication of this result is that as quality improvement efforts continue, there is a phenomenon of diminishing returns in the ability to continuously improve quality. We can utilize these results to provide insight to both *setting* quality goals and *implementing* quality programs to meet these goals.

The modeling efforts can be more explicit about how the goals are to be met - and strategies that will fall short of the goal. For instance, with the design for the MCSECT models, the lower improvement goals allowed for the assumption of breech presentation and fetal distress as valid reasons for cesarean section. The modeling efforts showed that for further improvements to occur, these assumptions would have to be allowed to change. Even in the earliest stages of improvement in these models there was a recognition that one of the things that would explicitly have to change is the number of women having repeat cesarean sections. Over time, this would also lead to a reduction in the number of women having previous cesarean sections; the need for this to happen to reach the goals of lower cesarean section rates is reflected in the outcomes of the simulation models. The policy implication of these results is that the criteria used for having a cesarean section, both for the first time and as a repeat cesarean section, will need to change if lower rates of cesarean section are to be realized. It should also be noted that these results are applicable to women experiencing premature labor rather than the general population. Further investigation in the same manner should be done to determine if similar effects are seen in all women.

Finally, the simulations can also point to goals that may be too lofty. If it is very difficult to reach a target and it appears that you will be far off even if you try (like cutting the cesarean section rate to 10%) and simultaneously you find that there is little difference in the outcomes measure for that great an improvement, you have to go back and examine your goals. In this case, the stretch goals for cesarean section rates may be counterproductive. Most of the gains are realized with a reduction to 15% and it is unlikely that you would ever

reach 10% according to the model, at least in the premature birth population which is approximately 10% of the overall universe for this effect.

Using this approach to simulation modeling using empirical health care data has provided insights into the implementation of quality improvement methods in the health care industry. We have been able to better understand some of the differences between alternative quality improvement efforts, analyze non-linearities in the outcomes measures, and consider the resulting inability to meet progressively difficult quality improvement goals. As noted, each of these insights has managerial implications for the use of quality improvement methods.

Discussion of the Health Care Process and Outcomes

Objective 3: Demonstrate that the methods developed also lend insight to the health services research arena, both in terms of methods and in terms of the specific results.
Research Question 3: When quality management techniques are used during a health care process what is their impact on specific outcomes, including health outcomes, utilization, and costs?
Proposition 3: Using a process model enables us to link health care processes and outcomes so that greater insights are gained.

Finally we consider the actual outcomes measures and the significance of the results of these modeling efforts. The first consideration is that while some of the effects look quite small, over the population groups of 6211 infants and 5136 mothers the differences are much more impressive. Additionally, if the results were translated to the general population (premature labor affects only approximately 10% of births), the effects would be magnified even more. In considering the outcomes, we address resource utilization, health outcomes

(morbidity and mortality), and costs. Chapter Five listed the detailed results of each outcome for each model. In this chapter we discuss some of the major effects of each of these categories of outcomes and highlight particularly meaningful, significant or surprising results.

Health Outcomes

In our research, the health outcomes considered included mortality and major sources of morbidity associated with premature labor and delivery. As the use of medical interventions and quality improvement techniques were introduced into the models we expected that these health outcomes would improve, or at the very least not worsen. The health outcomes considered varied depending on the specified model as some of the sources of morbidity were used as controls during the model design. However, there were two results that bear specific mention.

With health outcomes, one of the most interesting revelations from this investigation is the strength of the relationship between RDS and delivery mode. It is clear that decreasing the cesarean section rate (to 20%) has a significant effect on RDS. This is certainly not to say that cesarean section causes RDS but that more infants who develop RDS have been delivered by cesarean section. In our review of the literature, this relationship was not discussed. Yet, all of the results reflect this relationship.

The second point of interest is the increase in mortality in many of the models, particularly those looking at medical interventions of steroid use and surfactant, which presumably would improve health (and accordingly decrease mortality). This appears to be somewhat of an artifact of the random replacement. The suggestion is that mortality needs to

be modeled separately, or that infants that died might not be eligible for use as replacements. This would seem to drive the results in the other direction rather arbitrarily, but the current results would not seem to reflect such expectations. The mortality rate might also interfere with the length of stay measure, again supporting separate models for mortality.

Resource Utilization

Resource utilization measures include the time spent in the hospital or a specific unit and specific procedures/resources that are used during the course of the stay. As the incidence of RDS decreased or the cesarean section rate decreased, we expected to see the resource utilization measures decrease as well. In general, the models gave these anticipated results. However, there are two notables concerns with respect to resource use. Both are concerned with interactions between resource utilization measures.

The first interesting result is the relationship between length of stay and cesarean section in premature infants. When we did the cesarean section modeling, the effect on the length of stay was negligible until significant decreases in the cesarean section rate were made. This is a reflection of our population - we are modeling prematurity and that appears to have a greater effect on length of stay than delivery mode. Yet, when the cesarean section rate was lowered to 10% there was a statistically significant drop in the length of stay from 12.53 days to 12.32 days. The fact that there was eventually a significant difference due to delivery mode indicates that both delivery mode and prematurity might be factors in the length of stay. It would be interesting to further investigate these relationships. In the steroid models, which did not explicitly control for delivery mode, a greater decrease in resource utilization was

observed. This also supports further investigation of the relationship between delivery mode and prematurity and the resulting effect on length of stay.

The other interplay between outcome measures involves the length of stay, particularly in specific units, and the rate of transfer or length of stays that were zero days. The discharge data included information on the average length of stay in the NICU and the nursery as well as overall length of stay. It was not clear from our results how improvements in the number of infants in the NICU and/or nursery and their lengths of stay in those units might effect the percentages of transfers or lengths of stay equal to zero. The expectations could go either way - that lower lengths of stay are a reflection of more lengths of stay of zero, or that transfers and lengths of stay of zero should also go down, reflecting a general improvement in health outcomes. Our outcomes measures show the latter - that all these measures show the same general trends. Further investigation, using a population not limited to premature infants, might help flesh out these relationships.

Costs

Cost information was measured using total and daily charges as embedded in the data. The cost implications for the health care system are quite straightforward. There are significant decreases in costs with the use of steroids, the reduction of cesarean section (in the mothers' charges), and in the use of surfactant therapy. The charges reflect the degree to which each of the therapies should be used. For instance, in the steroid models significant differences in charges are only seen with incremental increases of 30% use. This also indicates

the point at which the implementation of quality improvement measures would be considered to provide a significant return on investment.

These results are limited by the data in the sense that they reflect only hospital charges, a microcosm of the costs involved in the health care system. Consideration should also be given to physician charges separate from the hospital, and the associated improvement in life years (a quality of life consideration). However, all of these considerations would likely only serve to amplify the results found here.

Discussion of the Problem Statement

Problem Statement: Can we develop new research methodologies that incorporate an integrated view of service processes and performance and allow us to investigate the effects of quality management techniques on large population data sets?

Having explored all three of the research objectives, which considered the methodology, the quality management techniques, and the modeling results separately, we now come back to the original problem statement that motivated this research. This problem statement addresses the link of all three of these areas with the use of population data sets. While we have used health care as an example, the problem statement is generalized to service processes. We assert that this generalization is possible for two reasons. First, the quality management techniques used by health care organizations were originally developed in manufacturing and adapted by many service industries, including health care. Second, the ability to do this type of research is based on the availability of appropriate data. There are many other service industries that collect data that can be used to link processes to outcomes

and assess the service delivery process. Thus, this discussion of the research question considers service processes in general and focuses on the exploratory nature of this research with respect to the methodology.

This research has made it quite clear that it is possible to develop new methods of researching services processes and performance. The simulation methods used here seem to be particularly appropriate for investigating these issues for a number of reasons. First, they are well-suited for use with large data sets and for complex problems. The bootstrapping methods translated into the random selection and replacement enhanced this feature of simulation. Also, combining simulation modeling with known effects of medical interventions and quality management techniques documented in the medical and management literature allowed us to develop models that were easy to validate. Finally, this methodology provides an alternative research tool that allows us to investigate the effects of quality management tools over large populations. This enables us to test the generalizability of results from case studies or research with smaller scope and determine the widespread effects of quality improvement. In total, we have successfully addressed all aspects of the problem statement originally posed.

Major Conclusions and Contributions of the Research

We wish to highlight some of the major conclusions from this research and note the resulting contributions to both operations management and health care policy. It was noted from the start that this research was exploratory, particularly in terms of the methods used, so first we note the contributions from the development of a new methodology. Then we focus

on the managerial insights provided from the research and the implications for both the operations management and health care fields.

Methodological Development

A number of particularly important insights have come from the exploratory work on methodological development. We consider the following to be the most important findings from using this methodology:

1. **We have developed an alternative method of modeling health care processes. The ability to use large population data sets and the random selection/random replacement technique developed are the most interesting aspects of this methodological development.** We initiated this investigation with the contention that researchers, particularly in the health care field, are limited in the scope of their investigations by commonly accepted research methods. While we in no way dispute the validity of those methods, we show that the development and use of alternative methods allows for a different scale (large populations) and scope (more inclusion of co-morbidities) of research. While the random selection and random replacement technique developed may have some shortcomings and may not be feasible in all situations, it does appear to mirror reality and thus the new methods are intuitively pleasing and have face validity.

2. **It is possible to use simulation modeling in order to extrapolate information from many sources and from smaller studies and determine the widespread effects of management techniques over a large population.** Often, we look at how

management techniques will affect one organization or one sector. Yet, it is desirable to understand how the results found in a case study or smaller research effort would be distributed if the same techniques were used on a more widespread basis. Although our research focused on quality management techniques, this simulation methodology could be used to provides insight as to the population effects of many management techniques.

3. **We are able to link processes and outcomes using this methodology.** Rather than studying either processes or outcomes, it is advantageous to begin to understand how they are related. This methodology, with its use of large population data sets that contain information regarding both processes and outcomes combined with known process interventions, allows for this connection.

4. **The methodology developed allows for both general and specific insights regarding managerial and policy issues.** For instance, the observation that there may be declining returns to quality improvement efforts was seen in nearly all the models and is very intriguing as a general result. At the same time, a very specific indication of a potential lower bound on the desired rate of cesarean sections for delivery of premature infants was indicated by this research. Both of these insights have managerial and policy implications and can help direct future research efforts.

This is not to say that we think that this methodology is ideal for use in all cases. There are several important considerations for the appropriateness of using the simulation modeling technique developed here. First, the data available must accurately reflect the

process and outcomes of interest. Secondly, the literature on the effects of management (and medical, in the case of health care) interventions must note the effects using the same parameters that are available in the data. For instance, we had difficulty modeling the steroid interventions as the medical literature was represented in terms of severity of RDS while our data included a binomial variable for presence or absence of RDS and many binomial variables for the presence or absence of related conditions or use of procedures. While we could develop these into a proxy for severity, the modeling effort was hampered by this mismatch between the data available and the literature. In another example, had our data included gestational age at delivery we would have been able to make better use of the medical literature on prematurity. Another consideration is the scope of the data, which must match with the management technique being studied. For instance, when we tried to investigate the clinical pathways, the scope of the data was much larger than the scope over which the intervention is used. Alternatively, the data must be adjusted to reflect a smaller scope. Finally, the amount and consistency of data is a concern. The bootstrapping techniques adopted for use here allow for smaller databases, but the number of cases that did not fit the models very well (evidenced by the nonreplacement) could be a concern with smaller cases.

Managerial Insights

The value of this research is not only in the development of a new research methodology but also in the managerial insights that can be gleaned from this method. Our research considered the effect of quality management techniques in a large population and

thus provides general insights to service operations management. Additionally, since our context was a health care process, there are specific health policy implications.

Service Operations Management

The results of this research provide the following important insights for service operations managers and researchers, including but not limited to those in the health care field:

1. **It is possible to use large population data sets in research by means other than regression, especially in order to link processes to outcomes.** As with health care, many service industries that have access to large data bases regarding their operations, yet they do not know how to glean information from this data. Financial institutions, insurance companies, hospitality and tourism, and transportation industries could all find this methodology useful. Often they consider either their processes or the end results, and these methods would allow them to link the two and to test how process changes would be translated to differing results on a large scale basis. This could be helpful not only in looking at quality management techniques, but also in broader efforts such as business process reengineering.

2. **As quality is improved and the bar is raised to increase quality further, the same quality management techniques do not result in the same magnitude of quality improvement.** The models that we developed had progressively more ambitious quality targets. Yet, as the quality target was increased, the ability to meet the target diminished. For instance, with the cesarean section models, when the target

rate was 25%, the realized rate for the models was 25.26%, quite close. By the time the goal was 10%, however, the realized rate was over 1% higher than the goal, at 11.23%. The implication is that increased or alternative efforts might need to be introduced in order to continuously improve processes.

3. **There seem to be diminishing outcomes returns to quality improvement efforts.** As above, with progressively more ambitious quality efforts, the effect on the process outcomes was less dramatic. This was most evident in the cesarean section models. At first, improvements in quality as evidenced by a decreasing rate of cesarean section had dramatic results. This effect diminished after the cesarean section rate dropped to 15%. Thus, if you are improving process quality but there is little or no effect on the outcomes of the service you are providing, then the goals of the quality improvement program could be questioned.

4. **Simulation modeling can be used to assist in both determining and implementing quality improvement goals.** The two previous findings help direct service operations managers in the use of this simulation modeling. First, the model design requires simulation of implementation plans. Secondly, the results of the simulation indicate the anticipated outcomes from different levels of quality. Thus, by using these simulation modeling techniques, operations managers could have some basis of determining what quality goals to set and how to meet them. Conversely, they could determine if quality improvement programs already set in place will be able to achieve the desired results or not.

Health care policy

The results of this research provide the following insights specifically important to health care administrators, researchers and policy makers:

1. **There are valid alternatives to commonly used research methods. The methods developed are particularly interesting as they make use of existing data bases.** The simulation modeling efforts show an alternative to randomized control trials. They are not proposed as an alternative to initial studies, but rather as a method by which to extend research on main causes and effects to study co-morbidities. The proposed methods allow for more measurement of co-morbidities and other related factors. The ability to measure these effects is the result of a tradeoff from not needing to control each variable independently. One advantage to this method is the ability to look at the interaction between effects.

2. **There is an apparent lower bound to the desired cesarean section rate.** Although there is a continual debate as to what the appropriate cesarean section rate should be, there is little experimental justification for the rates proposed. This study provides evidence that there may indeed be a lower bound below which the cesarean section rate (in conjunction with premature labor and delivery) should not go.

3. **There is evidence of a relationship between RDS and cesarean section that is not specifically delineated in the medical literature.** Clearly, cesarean section is not believed to cause RDS, yet our results show that more infants who develop RDS have been delivered by cesarean section. This is an important finding mostly because it points out new, relevant avenues of future research .

Future Research

You are today where your thoughts have brought you; you will be tomorrow where your thoughts take you.
James Allen

Research with this scope and of this experimental nature can seem to reveal more areas to delve into than it does answers. The modeling efforts in themselves presented a number of challenges and decisions, many of which could be approached in a different manner in future research. We list here a number of thoughts on extensions of this research, noted in two categories: topics related to methodology and topics related to specific health care related research.

Methodology

There are many possible avenues for future research that would further refine the methodology that we have begun to develop. These future research topics would be a continuation or extension of this work, and the results would be viewed in conjunction with the results we have already documented. Many of these suggestions would serve to further validate or perfect the methodology as developed thus far.

First we consider topics for further study related to the random selection and random replacement procedures. For instance, we experimented with matching mothers and infants in order to avoid some of the random replacement and in an effort to connect all phases of the process of premature labor and delivery in one model. While we decided that we needed to use all the data rather than our matched subsets, this is still an avenue worth pursuing. In

particular, it would be best to investigate other discharge data bases where the mother and infant discharge records are linked. Then a comparison between the two methods could be made.

There are a number of ways that the issue of non-replacement could be further investigated. First, in order to test the effect of those cases that could not be matched and thus replaced, these observations could be deleted and the effect on the results assessed. While we decided not to pursue this avenue as we were particularly interested in retaining as much inherent variation as possible, it would be worth comparing the results. Alternatively, the criteria for matching could be widened in order to see if a higher percentage of matches could be made. For instance, replacement need not be limited to other discharges from the same hospital, either by not having the hospital as a criteria at all or by matching hospitals based on an alternative proxy for socioeconomic status. This might allow for control for other factors, such as the incidence of complications from breech births, that resulted in counterintuitive results.

It would also be helpful to have other methods by which to test the significance of the results. It appears that SUUDAN would be useful. This technique is applicable when there is no independence between variables as evidenced by t tests. If the above research topics were pursued, SUUDAN might be particularly applicable as a means by which to compare the current experimental methodology with alternative methodologies.

The non-linearities that were observed in the results could also be investigated further. It would be helpful to have a finer discernment of the break points in the non-linearities. This can be investigated by trying smaller intervals on the experimental variables. Alternatively, the

simulation results could be used to guide mathematical modeling efforts for the variables that appear to have these non-linearities.

Finally, from a methodological standpoint, an additional topic for future research would be to look at the distributions of the different outcomes. While the outcomes measures are not normally distributed - many in service industries are not - it would be interesting to better understand how they are distributed. Some preliminary work was done in this regard, using Best Fit software to analyze the distributions. Of the few distributions tested, none were easily identified (thus strengthening the rationale behind the random replacement technique as opposed to any Markovian analysis). If the distributions were better understood, then more specific modeling efforts could be undertaken.

Health Care Policy and Health Services Research

The results of the simulation modeling efforts provided many interesting insights into health care delivery that warrant further research. While some of these future research topics could be investigated with a continuation of the current methodology, other methodologies may be more appropriate than this one for some of these questions. It is the objective of this exploratory methodology to reveal important questions and helping to identify the interactions of interest, particularly given the use of real data.

There are a number of further investigations that the results from this study suggest would be worthwhile with respect to cesarean section. First, it would be helpful to know more about the specific determinants of use of cesarean section in premature deliveries as opposed to full-term births. Larger scale simulations that would compare the effects of

reducing the cesarean section rate in two population groups - premature and full term - would assist in this investigation. Then, it would be easier to delineate the effects of the reduced cesarean section rate from the effects of prematurity. Secondly, the apparent lower bound on cesarean section rate needs to be further investigated and confirmed. This has important policy implications and could provide the basis for sound health policy decisions. Thirdly, further investigation needs to be done to better define the relationship between RDS and cesarean section. The specific characteristics of this relationship should be investigated. Is the cesarean section rate a function of other reasons commonly cited for performing cesarean section, such as fetal distress? Are premature infants who exhibit these other conditions resulting in cesarean section more likely to have RDS? Is there an advantage for the premature or RDS infant to be delivered via cesarean section? The current investigation helps us pose questions such as these.

An important and distressing result was the consistently increasing mortality rate with the different modeling efforts. This result is counterintuitive and seems to be the biggest drawback of the random replacement technique we developed. It is likely that mortality may need to be modeled specifically as a separate category of outcomes from prematurity. Alternatively, cases of premature labor resulting in infant mortality may need to be dropped from the data base in order to eliminate any confounding variables. A better understanding of the drivers of mortality in the cases of premature labor and delivery would be helpful. Additionally, the inability to accurately assess mortality may be a shortcoming of the data since it does not include any means to assess severity of RDS. The relationship between mortality and the severity index used should be investigated as well.

Finally, further investigation into the appropriateness of simulation methods for more local quality improvement methods such as clinical pathways would be useful. Further research would include the means to differentiate between providers or group them and then model local interventions. The difficulty in developing a model that investigates these provider differences is that it would be extremely costly to implement at that level of disaggregation and data in sufficient volume is unlikely to be available. However, such a model might produce more valid and generalizable results with respect to clinical pathways.

This investigation has produced many worthwhile managerial insights, has implications for health care policy, and provides significant contributions to the development of new methodologies that are useful in assessing the effectiveness of service operations by linking processes to outcomes. We began this investigation with a simple notion that the variation underlying service processes is not controllable in the same manner that manufacturing processes can be controlled. The model design was carefully constructed in order to reveal insights regarding the importance of inherent variation in service processes and particularly in health care. At the conclusion, the results show that the inherent and complex variation underlying the process is particularly important in determining the effectiveness of management techniques. This was especially evident in the inability to meet quality improvement targets and the resulting diminishing returns. We conclude this research as convinced as we were at the beginning that the ability to understand and appropriately manage (not necessarily control) this inherent variation is the key to future success in service operations management.

APPENDIX A

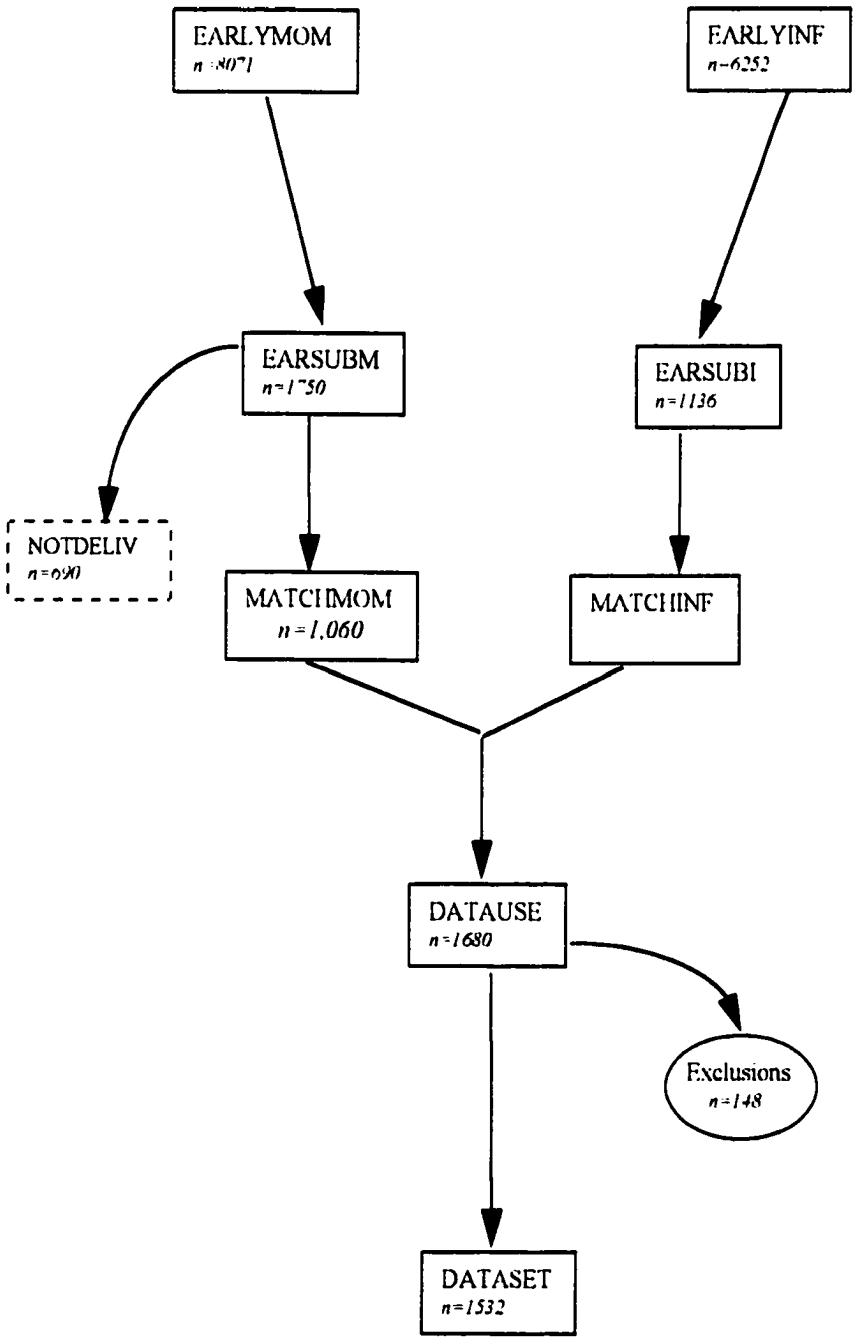
DETAILS OF THE MATCHING PROCESS

In order to investigate the process of premature labor and delivery as a continuum from mother to infant we wanted to identify the pairs of mothers and infants from the discharge data. As the data did not include any specific information linking mothers and infants we used other information contained in the discharge records to match up mothers and infants. This appendix details this matching process, including the discharge information used, major assumptions, and the success of this process. A diagram depicting the matching process is given in Figure A.1. A number of tables are included that give the results of the descriptive statistics for the different subsets. This information is provided for background and for the basis of future research. As described earlier, the methodology used in this research did not employ matched pairs as the matched pairs from a subset of hospitals did not adequately reflect the population of women experiencing premature labor and infants born prematurely.

Overview of the Matching Process

In order to match mothers and infants, we first defined an appropriate subset by hospital, compared discharge information from the mothers and infants, and then paired the data into one dataset. This process is depicted in Figure A.1. As noted in Figure A.1, this process begins with the EARLYMOM and EARLYINF datasets which were defined in Chapter Four.

Figure A.1. Diagram of the Matching Process



Sorting the Premature Mothers by Health Care Provider

Before the record matching was undertaken, it was determined that a subset of the premature mothers could be used as input for this simulation study and still be large enough to produce reasonable results (allowing us to avoid matching all 8,071 observations in EARLYMOM). Considerations for this further sub-sample were that it be representative of all premature mothers/infants, have a consistent cost structure, and provide enough cases of premature labor to get a reliable sample of the variety of outcomes from which to conduct the simulation.

Four health care providers were selected from the 46 hospitals. They were large hospitals (> 400 beds or large number of births), though not teaching hospitals as they tend to have higher costs and lower utilization rates. Provider numbers 3, 4, 15 and 32 were chosen for the initial matching and design.

Table A.1 shows a breakdown of the mothers and infants treated at these hospitals, including the rates of prematurity. In total, this allowed for a subset of 1,750 mothers experiencing premature labor to use for the matching process (EARSUBM). The subset of infants from these providers included 1,136 discharges (EARSUBI). Additionally, it was desired that the hospitals selected have premature labor/delivery rates similar to that of the overall sample. The rate of prematurity represented by this subset of 1,750 mothers is 10.4% compared to 9.9% in the statewide population. The rate of prematurity in the infants is 7.05% compared to 7.8% in the population.

In order to validate the selection of these providers, statistical analyses comparing the subset of four providers to the larger premature subset were conducted for both mothers and

Table A.1. Mothers and Infants Treated at Four Selected Hospitals.

PROV_NUM	3	4	15	32	TOTAL
# Mothers	3412	6639	2355	4358	16764
# Mothers-Premature	366	855	203	326	1750
% Mothers Premature	10.73%	12.88%	8.62%	7.48%	10.44%
# Mothers - Not Delivered	92	379	126	93	690
# Mothers Delivered Premature	274	476	77	233	1060
# Infants	3370	6230	2255	4248	16103
# Infants-Premature	287	429	167	253	1136
% Infants Premature	8.52%	6.89%	7.41%	5.96%	7.05%
Dataset	310	748	189	285	1532
Exclusions	48	58	9	33	148

infants. Table A.2 reflects the descriptive statistics and relevant frequencies of the provider subset for women experiencing premature labor (EARSUBM) and compares them to the subset of all women experiencing premature labor (EARLYMOM). Table A.3 reflects the descriptive statistics and relevant frequencies of the provider subset for premature infants (EARSUBI) and compares them to the subset of premature infants (EARLYINF).

Record Matching

Once the subset of mothers experiencing premature labor/delivery was isolated, then the task was to pair up mothers and infants in a logical fashion so that the observations used in the simulation contained combined information on both mothers and infants. We reviewed the sets of mothers and infants to ensure the appropriate set for matching, determined the

Table A.2. Descriptive Statistics of the Provider Subset for Women Experiencing Premature Labor (EARPROVM) and the Subset of All Women Experiencing Premature Labor (EARLYMOM).

Variable Type/ Variable	EARLYMOM *n=8071*		EARPROVM *n=1750*	
Descriptive	mean	std dev	mean	std dev
age	26.17	5.94	26.82	5.89
los	3.03	4.23	2.66	3.62
dal chg	2487.39	35173.35	1048.66	1319.98
tot chg	15688.64	351698.72	2881.41	2940.03
Binomial	percent=1	std dev (a)	percent=1	std dev
deliver	.6364	.4811	.6057	.4888
csect (b)	.2447	.4300	.2575	.4375
prom	.1436	.3507	.1423	.3494
plprev	.0244	.1543	.0240	.1531
othplac	.0368	.1883	.0297	.1698
preecl	.0421	.2009	.0371	.1892
infect	.0706	.2562	.0503	.2186
breech	.0611	.2395	.0531	.2244
forvac (b)	.0750	.2634	.0840	.2775
prolong (b)	.1020	.3027	.0849	.2789
badcare	.0396	.1951	.0423	.2013
lt 22wks	.0258	.1585	.0343	.1820
wk22 37	.9750	.1562	.9669	.1791

Variable Type/ Variable		EARLYMOM *n=8071*		EARPROVM *n=1750*	
Categorial	**Values**	freq	percent	freq	percent
discharge year	1991	1876	23.2	166	.0949
	1992	6195	76.8	1584	.9051
race	1 white	4141	51.3	718	41.0
	2 black	3628	45.0	863	49.3
	3 asian	104	1.3	38	2.2
	4 amer ind	4	0.0	1	0.1
	5 other	192	2.4	130	7.4
	9 unident.	2	0.0	0	0.0
payment	1 medicare	13	0.2	1	0.1
	2 medicaid	2909	36.0	497	28.4
	3 title V	0	0.0	0	0.0
	4 blue cross	1085	13.4	178	10.2
	5 insur co	1382	17.1	335	19.1
	6 other govt	40	0.5	2	0.1
	7 work comp	2	0.0	1	0.1
	8 self-pay	422	5.2	165	9.4
	9 no charge	1	0.0	0	0.0
	10 other	18	0.2	2	0.1
	11 donor	1	0.0	0	0.0
	12 HMO	1793	22.2	564	32.2
	13 MSO	103	1.3	3	0.2
	14med/HMO	299	3.7	1	0.1
	99 unknown	3	0.0	1	0.1
marital status	1 single	3587	44.4	728	41.6
	2 married	3954	49.9	931	53.2
	3 separated	209	2.6	54	3.1
	4 divorced	159	2.0	33	1.9
	5 widowed	20	0.2	3	0.2
	9 unident.	142	1.8	1	0.1

(a) calculated as $s=\sqrt{[p(1-p)]}$
(b) calculated as a % of those women who delivered during the episode of care
* - significantly different at p= 0.01
NS- not significantly different

Table A.3. Descriptive Statistics and Relevant Frequencies for the Provider Subset for Premature Infants (EARPROVI) and the Subset of Premature Infants (EARLYINF).

Variable Type/ Variable	EARLYINF *n* = 6252		EARSUBI *n* = 1136	
Descriptive	mean	std dev	mean	std dev
bbrth wt	2336.13	1021.03	2363.80	908.16
blos	12.56	23.04	8.41	11.95
bdal chg	7497.92	23295.59	4047.84	6617.72
btot chg	13494.31	180104.90	5680.22	9109.37
days nur (b)	3.51 (n 1952)	9.22	.69 (n 245)	1.57
Binomial	percent=1	std dev (a)	percent=1	std dev
besect	.3045	.4603	.3336	.4717
bprom	.0075	.0864	.0070	.0837
bplprev	.0014	.0379	.0009	.0297
bpreecl	.0042	.0644	.0018	.1316
bbreech	.0051	.0714	.132	.1142
RDS	.2455	.4304	.2016	.4014
othresp	.2342	.4235	.2148	.4109
vent	.1379	.3448	.1171	.3217
intub	.1703	.3760	.1805	.3847
respther	.0845	.2781	.0695	.2545
lt 28wks	.0968	.2957	.0713	.2574
wk28 37	.7396	.4394	.7931	.4052

Variable Type/ Variable		EARLYINF *n* = 6252		EARSUBI *n* = 1136	
Categorial	**Values**	freq	percent	freq	percent
discharge year	1991	1601	.2561	113	.0995
	1992	4651	.7439	1023	.9005
sex	1 male	3216	51.4	568	50.0
	2 female	3035	48.5	568	50.0
	9 unknown	1	0.0	0	0.0
race	1 white	3187	51.0	472	41.5
	2 black	2838	45.4	557	49.0
	3 asian	97	1.6	34	3.0
	4 amer ind	4	0.1	1	0.1
	5 other	126	2.0	72	6.3
	9 unident.	0	0.0	0	0.0
payment	1 medicare	1	0.0	0	0.0
	2 medicaid	2488	39.8	357	31.4
	4 blue cross	770	12.3	106	9.3
	5 insur co	1012	16.2	214	18.8
	6 other govt	41	0.7	1	0.1
	7 work comp	0	0.0	0	0.0
	8 self-pay	405	6.5	135	11.9
	9 no charge	0	0.0	0	0.0
	10 other	8	0.1	0	0.0
	12 HMO	1252	20.0	321	28.3
	13 MSO	23	0.4	1	0.1
	14med/HMO	251	4.0	1	0.1
	99 unknown	1	0.0	0	0.0

(a) calculated as s $\sqrt{[p(1-p)]}$

(b) For those infants who spent time in both the NICU and the nursery, the number of days spent in the nursery

* - significantly different at p 0.01

NS- not significantly

information by which to complete the matches, and completed the matching hospital by hospital for each of the four providers. Provider 3 was used as the test case and the procedure was refined as we went through the matching process for this hospital.

Not all mothers actually deliver infants during the episode of care indicated by the observation. In fact, it is often the goal of clinical intervention to arrest premature labor rather than letting it proceed to delivery. Those observations that did not include delivery are relevant to the current study and thus are retained but not matched with infants. As was noted in Table A.1, for our sample of mothers from providers 3, 4, 15 and 32 who experienced premature labor, 1,060 delivered infants during the current episode of care. Thus, the task was to match these 1,060 mothers with their infants.

The relevant set of infants is less readily evident. Premature infants are noted by DRG codes 385-388 and/or a primary or secondary diagnosis of 765.xx (the fourth and fifth digits refer to the degree of prematurity and the range of birth weight). Ideally, each women experiencing premature delivery would match with a premature infant. The coding of prematurity in infants is not based entirely on gestational age as it is in women, and is largely determined by birth weight rather than gestational age. Infants born under 2500g are considered premature. Thus, it is possible that a woman experiencing a premature delivery with respect to gestational age would give birth to an infant not noted as premature (usually with a higher birth weight). Thus, nearly all infants are potentially able to be born to women delivering prematurely.

To counteract this seemingly contradictory data artifact, a three-pass strategy was employed in the matching process. Preference was given to first matching the infants noted as

premature and then to finding appropriate infants for the mothers that did not match with a premature infant. The first pass consisted of only considering infants that were premature as candidates for matching. The second pass consisted of considering all infants under 2500g (not all were noted as premature, thus confirming the notion that the coding of prematurity in infants is not decisive). On the third pass, for those women who had not already been paired with infants, all infants were considered as possibilities. This results in some women noted as delivering prematurely paired with infants who do not appear to be premature as evidenced by birth weight. There are many scenarios which could in actuality lead to these types of results. In additions to medical conditions such as gestational diabetes which results in larger infants, incorrect calculation of gestational age or a premature delivery very close to term are two possible explanation. Table A.4 details the results of the matching process by pass for providers 4, 15 and 32.

To begin the actual process of matching mothers and infants, we first reviewed the information contained in each observation to determine which pieces of information would contain clues as to which mothers and babies would logically fit together. The information used to pair the mothers and infants included provider number, zip code, discharge year, discharge quarter, an appropriate combination of admission day, discharge day, and length of stay, payment code, and race. Additional information was garnered from the appropriate binomial variables created from the diagnoses and procedure codes. These variables include cesarean section, breech presentation, gestation length, premature rupture of membranes, placenta previa, other placental conditions, fetal distress, use of forceps and/or vacuum extraction, drug use and/or abuse, and infections.

Table A.4. Results of Three Pass Matching Process, by Pass and Provider

	Early Moms	# of Matches Made				Dropped on Review
		First Pass	Second Pass	Third Pass	Total	
Provider 03	274				222 81.02%	
Provider 04	476	278 58.40%	93 19.50%	90 18.90%	461 96.85%	7
Provider 15	77	62 80.50%	13 16.90%	0 0.00%	75 97.40%	2
Provider 32	233	163 70.00%	34 14.60%	35 15.00%	232 99.60%	1
TOTAL	1060				990 93.40%	

To match the mothers and infants we searched for an exact match or equivalent, a near match, or an acceptable match based on the information contained in the discharge data. In an exact match, all criteria would be exactly the same for the mother and the infant. Near matches may not match exactly but reflect a likely combination of information. Acceptable matches are feasible, though not expected to occur frequently. At a minimum, mothers and infants that were matched received medical care at the same hospital (provider number) and recorded their addresses in the same zip code. Most also were discharged during the same year/quarter. It is possible for an infant, usually with a long length of stay, to be discharged in a later quarter than the mother. For the purposes of matching mothers and infants, infants were not allowed to be discharged in a year/quarter prior to the mother.

In order to facilitate the matching process, data were sorted by provider number, zip code, and then sequentially by the other variables of concern. Printouts of the relevant variables for mothers for each provider, in zip code order, were used to facilitate the matching

process. Of particular interest were those zip codes which, for a given provider, only had a handful of observations. These mothers and infants were matched first, and the information from those matches helped in determining what other feasible matches might be. In only a few instances were matches not made because there were no corresponding mothers/infants in the same zip code. For all the other variables, a discussion of near matches and acceptable matches is given below. In all cases, matching was done by hand, and the most infrequent variable combinations were matched first. By process of elimination infants were identified for nearly all of the mothers in the subset of mothers experiencing premature labor.

Arrival/Discharge And Demographic Variables

These variables provide information about the parameters of the stay and demographic information that is essential to the matching process. While these variables may not be exactly the same for the mother and infant, they are likely to be reasonably similar. With the exception of the length of stay (and thus the discharge day), these variables have nothing to do with the reason for the hospitalization; thus, regardless of the course of action during the episode of care, these matches should remain the same.

Admission Day, Discharge Day, and Length of Stay Combination

An exact match , the mother and infant would arrive the same day, stay the same length of time, and be discharged on the same day. There were many valid exceptions to this algorithm, the most common being an infant admission day one or two days subsequent to the mother's admission day with a corresponding length of stay long one or two days shorter than

the mother's. This was considered an equivalent to an exact match (in fact, the difference between one day in admission could actually be a few hours of active labor). The next most common exception was that the admission day was the same or one day different, but the infant's length of stay was greater, and thus the discharge day would not match (and perhaps discharge quarter as well). In general, it was more likely that the discharge day would match than the admission day. It was rare that the mother would be discharged after the infant and in no cases could the infant be admitted before the mother. One of the most difficult scenarios to determine a match was for mothers who had long lengths of stay. In one instance, one mom could match with any of 9 infants. In cases like this, "ties" were broken by considering birth weight (which would be the second round of matching criteria) and preference was given to infants under 2500g as they are considered premature.

Payment Code

For payment codes, there were a number of cases that fell into the near match category. These instances essentially reflected the common occurrence that a mother and infant may not be covered by the same insurance plan. Often, especially for working mothers, the mother may be insured on one plan and the infant insured on its father's insurance policy. This means that any combination of Blue Cross, insurance company, and HMO would be likely. Another near match would be an infant covered by Medicaid, regardless of the mother's insurance, which could simply reflect the fact that the mother's insurance would not have covered the infant. Medicaid inclusion rules are broad, and it would be more likely to be covered than not.

Race

The most common near match in the race category is for the baby to be designated as "other". This baby could be matched with any mother. Any other combination of race would be considered acceptable, reflecting the fact that race is a self-reported characteristic and thus prone to include errors. In other words, a match was preferable, and in some cases even the distinct identifier (and Asian mother and infant, for instance), but a mismatch in race would not necessarily eliminate the match.

Diagnoses And Procedure Variables

These variables, presented as binomial translations of multiple data fields, were useful in providing clues as to matches rather than dictating the results. If there was a positive notation for any one of them in a mother, and a positive notation for an infant that might otherwise be a near or acceptable match, then these notations would add validity to the match. For instance, if it was noted in both a mother and an infant that forceps were used, this would be considered particularly strong evidence that they were a match as long as all the other variables were feasible. At the same time, no notation of these had no effect - positive or negative - on the match. Often, it was more likely for a mother to have a notation of a condition than an infant, or vice versa, and this would not support or negate the match.

This strategy reflects the data coding. While the variables related to diagnoses and procedures that provided clues to matching there unfortunately there was no guarantee that the appropriate diagnosis was noted in the discharge data for both the mother and the infant. There are two reasons for this. First, a mother and infant are not handled the same, and so

while the variables surrounding the parameters of the stay should be the same, the relevant diagnoses may not match. Second, the level of importance of reporting an incident can be different for the mother versus the infant. Placenta previa, for instance, is a critical condition of note for a mother, but it may not significantly affect the infant (and there may be other conditions that are more important to note). Similarly, the use of forceps may be more notable in the care of the infant than the mother, and thus it may be noted only in the discharge data corresponding to the infant.

Cesarean Section

All mothers who have cesarean sections have it noted on their records. Infants, on the other hand, may have no notation made, especially if there are other more important conditions that need attention. Therefore, if a cesarean section is noted in the infant, a match may only be made with a mother who had a cesarean section. Mothers with cesarean sections are more likely to be matched with infants also noted as having a cesarean, but may on occasion be matched with an infant without that notation. A simple two-by-two table indicates the preferred, possible, and impossible combinations (Table A.5a.). Impossible combinations negated a match, regardless of other matching variables. Table A.5b. indicates the cesarean section rates in the different data sets.

Table A.5a. Possible Matches of Mothers and Infants by Delivery Mode.

		Infants	
Mothers		Vaginal	C Sect
	Vaginal	Preferred	**Impossible**
	C Sect	Possible	Preferred

Table A.5b. Cesarean Section Rates by Data Set.

Cesarean Section Rates (%)	**Master**	**Select Providers**	**Early**	**Early/Select Providers**	**Matched Dataset**
Mothers	24.23	24.50	24.47	25.75	23.69
Infants	22.53	23.60	30.48	33.36	22.40

Breech Presentation

The notation of breech presentation is similar to that of cesarean section: it may or may not be noted in both records. The one difference is that a breech presentation can be important in the course of care of the mother or of the infant, or both. Thus, unlike cesarean section where a notation of cesarean section in the infant necessitated a match with a mother that also had a positive notation of cesarean section, breech presentation may be noted only on one record. Thus, from the matching perspective, if a likely match between a mother and an infant also were both indicated to have been breech, that match would be supported by the information. Otherwise, the indication of breech presentation was of little consequence. A simple two-by-two table indicates the preferred and possible combinations (Table A.6a.). Table A.6b. indicates the rate of breech presentation in the different data sets.

Table A.6a. Possible Matches of Mothers and Infants by Breech Presentation.

		Infants	
Mothers		Breech	Not breech
	Breech	Preferred	Possible
	Not breech	Possible	Preferred

Table A.6b. Rate of Breech Presentation by Data Set.

Rate of Breech Presentation (%)	**Master**	**Select Providers**	**Early**	**Early/Select Providers**	**Matched Dataset**
Mothers	3.49	3.45	6.10	5.31	4.96
Infants	0.23	0.55	0.52	1.32	1.52

Gestation Length

Gestation length was used as a check to match appropriate mothers and infants, particularly when there was indication of extreme prematurity (less than 22 weeks for mothers and less than 28 weeks for infants). This was a small proportion of the mothers, but it still helped identify and support some matches. The tricky thing with gestation age is that the cutoff for extreme prematurity is different for infants and mothers, so an infant noted as extremely premature could be matched with a mother noted as premature (but not extremely so). Thus, a one-way association was used, identifying mothers less than 22 weeks and their infants when possible, and the remaining mothers were matched on other criteria as noted in Table A.7.

Table A.7. Possible Matches of Mothers and Infants by Gestation Length

		Infants		
Mothers		Less than 28 weeks	28 to 37 weeks	Unknown
	Less than 22 weeks	Preferred	No match	Unlikely
	22 to 37 weeks	Possible	Preferred	Preferred

Premature Rupture of Membranes (PROM), Placenta Previa, and Other Placental Conditions

PROM, placenta previa and other placental conditions more seriously affect the mother than the infant. The data reflect this difference; for all mothers experiencing premature labor the rate of PROM was 14.36% while for premature infants the rate of PROM was 0.74%. The corresponding rates for PROM in just the four providers used for the matching process was 14.22% for mothers and 0.70% for infants. Similar differences in incidence were found with placenta previa and other placental conditions.

Due to the difference in these rates of incidence, mothers with these conditions often could not be matched with corresponding incidence in the infant. A one-way matching strategy was employed; infants noted as affected by one of these conditions were matched with mothers also noted as having the condition. Once the infants with these conditions were matched, the notation of PROM, placenta previa or other placental conditions was used only as a possible cause of infection in infants. If an infant exhibited infection noted as being associated with conditions in the mother, then one of these placenta-related conditions was suspect. If a likely candidate for a mother that matched the infant also exhibited a placenta-related condition, that was considered more support for the match.

Fetal Distress and the Use of Forceps and/or Vacuum Extraction

Fetal distress and the use of forceps and/or vacuum extraction are more significant as a condition during labor than after delivery, and thus they are more prevalently noted in the mother's discharge record. For mothers delivering prematurely, the rate of incidence of fetal distress was 13.20%. For premature infants, fetal distress was noted 4.64% of the time. For mothers delivering prematurely, the rate of use of forceps or vacuum extraction was 7.5%. For premature infants, forceps were used 0.40% of the time and vacuum extraction 0.16% of the time for a combined rate of 0.56%. Again, a one-way matching strategy was employed; infants noted as affected by one of these conditions were matched with mothers also noted as having the condition.

Drug Use and/or Abuse

Drug use/abuse is commonly noted in both mothers and infants (6.67% and 5.25% respectively for those delivering/delivered prematurely), so it proved helpful as a supporting variable for the matching process. Not all mothers or infants with noted problems related to drug use were matched with corresponding infants or mothers with similar problems, but often this notation helped guide the "tiebreaking" selection of multiple good matches by other criteria.

Infections

Infections were more relevant to the hospital stay of the infant than that of the mother. Premature infants discharge records noted an infection rate of 18.76%; for mothers delivering

prematurely a 7.06% rate of incidence of infection was seen. However, an infection in the infant is an indication that something was up in the mother, such as the conditions of PROM or placenta previa detailed earlier. Thus, this information was used to match in a one-way strategy where mothers with infection were matched with infants with infection and then seen as a support for matching an infant to a mother with PROM, placenta previa, or other placental conditions.

Combining Observations

As the mother-infant pairs were identified, a match_id number assigned to the mother was noted and the infants were given the same match_id number. A further comparison was done by interleaving the datasets, enabling us to recheck the matching assignments and review ties or questionable matches. Then the records were match-merged using the match_id number to create a working data set in which each observation included a mother and an infant. Table A.4 given earlier to note the results of the three pass method also includes information on the number of matches that were made. A total of 990 mothers were successfully matched with infants (93.4% of the subset of mothers experiencing premature labor).

Ideally, all mothers experiencing premature labor would be correctly matched. In reality, based on the information contained in the data sets, there were some observations that could not be isolated to matching with just one infant, some where there was a group of observations of mothers or of infants that were indistinguishable from each other, and some that did not have an appropriate match at all. However, all observations contained in the

working data set are likely combinations if not an exact match. This also means that the actual working data set is a subset of the moms in premature labor diagramed in Figure A.1.

Validity Testing

Finally, observations that could unfairly skew the results (such as odd complications and comorbidities) are deleted from the working data set and both the subsets which contain women experiencing premature labor and premature infants. In accordance with both Robertson et al (1992) and Lewis et al (1996), mothers with preeclampsia and indications of substance abuse were separated out from the data set, along with the infants matched to them. In total, 148 observations were excluded after the matching process. After this process, the final data set to be used included 1532 observations, 857 matched mother-infant pairs and 675 mothers who did not deliver during this episode of care. Each of these observations contain 137 variables.

At this point, we wanted to see whether or not the matched data set was representative of both women experiencing premature labor and infants born prematurely. Thus, we did comparisons of the matched data set to the original sets of premature infants/mothers that formed the basis of the matching. Table A.8 reflects the descriptive statistics and relevant frequencies of the subset of all women experiencing premature labor (EARLYMOM) and compares them to the matched data set (DATAUSE). Table A.9 reflects the descriptive statistics and relevant frequencies of the subset for premature infants (EARLYINF) and compares them to the matched data set (DATAUSE). Unfortunately, as Table A.8 and Table A.9 indicate, the matched data set was not a good representative of the

base data sets for prematurely born infants and women experiencing premature labor. Thus, we abandoned the matched data set for simulation purposes and designed the simulation models using the EARLYMOM and EARLYINF data sets as described in Chapter Four.

While the matched data set was not used, this portion of the research was still fruitful. It was shown that mother/infant records could be successfully matched, thus leading the way to a larger scale process simulation where the mothers and infants could be linked in the same model. Thus, future research efforts could expand upon the results of this matching effort. Possible streams of future research are discussed in Chapter Six.

Table A.8. Descriptive Statistics of the Provider Subset of All Women Experiencing Premature Labor (EARLYMOM) and the Matched Data Set (DATASET).

Variable Type/ Variable	EARLYMOM *n=8071*		DATASET *n=1532*	
Descriptive	mean	std dev	mean	std dev
age	26.17	5.94	26.75	5.9
los	3.03	4.23	2.47	3.5
dal_chg	2487.39	35173.35	979.33	1272.75
tot_chg	15688.64	351698.72	2653.17	2770.19
Binomial	percent=1	std dev (a)	percent=1	std dev
deliver	.6364	.4811	0.5594	0.4966
csect (b)	.2447	.4300	0.2369	0.4254
prom	.1436	.3507	0.1416	0.3488
plprev	.0244	.1543	0.0209	0.1431
othplac	.0368	.1883	0.0235	0.1515
preecl	.0421	.2009	0	0
infect	.0706	.2562	0.0463	0.2103
breech	.0611	.2395	0.0496	0.2172
forvac (b)	.0750	.2634	0.0805	0.2722
prolong (b)	.1020	.3027	0.0875	0.2827
hadcare	.0396	.1951	0.0287	0.1671
lt 22wks	.0258	.1585	0.0352	0.1845
wk22 37	.9750	.1562	0.9647	0.1845

Variable Type/ Variable		EARLYMOM *n=8071*		DATASET *n=1532*	
Categorial	**Values**	freq	percent	freq	percent
discharge year	1991	1876	23.2	179	11.7
	1992	6195	76.8	1353	88.3
race	1 white	4141	51.3	642	41.9
	2 black	3628	45.0	738	48.2
	3 asian	104	1.3	37	2.4
	4 amer ind	4	0.0	1	0.1
	5 other	192	2.4	114	7.4
	9 unident.	2	0.0	0	0.0
payment	1 medicare	13	0.2	1	0.1
	2 medicaid	2909	36.0	414	27.0
	3 title V	0	0.0	0	0.0
	4 blue cross	1085	13.4	159	10.4
	5 insur co	1382	17.1	303	19.8
	6 other govt	40	0.5	2	0.1
	7 work comp	2	0.0	1	0.1
	8 self-pay	422	5.2	142	9.3
	9 no charge	1	0.0	0	0.0
	10 other	18	0.2	1	0.1
	11 donor	1	0.0	0	0.0
	12 HMO	1793	22.2	505	33.0
	13 MSO	103	1.3	2	0.1
	14med/HMO	299	3.7	1	9.1
	99 unknown	3	0.0	1	0.1
marital status	1 single	3587	44.4	616	40.2
	2 married	3954	49.9	844	55.1
	3 separated	209	2.6	42	2.7
	4 divorced	159	2.0	26	1.7
	5 widowed	20	0.2	3	0.2
	9 unident.	142	1.8	1	0.1

(a) calculated as $s=\sqrt{[p(1-p)]}$
(b) calculated as a % of those women who delivered during the episode of care
* - significantly different at p=0.01
NS- not significantly different

Table A.9. Descriptive Statistics and Relevant Frequencies of the Subset for Premature Infants (EARLYINF) and the Matched Dataset (DATASET).

Variable Type/ Variable	EARLYINF *n- 6252*		DATASET *n- 1532*	
Descriptive	mean	std dev	mean	std dev
bbrth wt	2336.13	1021.03	2540.41	837.05
blos	12.56	23.04	5.57*	8.46
bdal chg	7497.92	23295.59	2537.67*	4774.4
btot chg	13494.31	180104.90	3522.22	6354.2
days nur (b)	3.51 (n 1952)	9.22	4.46 (n 781)	7.72
Binomial	percent- 1	std dev (a)	percent- 1	std dev
besect	.3045	.4603	0.224	0.4172
bprom	.0075	.0864	0.0105	0.102
bplprev	.0014	.0379	0.0012	0.0342
bpreecl	.0042	.0644	0.0012	0.0342
bbreech	.0051	.0714	0.0152	0.1223
RDS	.2455	.4304	0.1237*	0.3294
othresp	.2342	.4235	0.1365*	0.3435
vent	.1379	.3448	0.0712*	0.2573
intub	.1703	.3760	0.1179	0.3226
respther	.0845	.2781	0.049	0.216
lt 28wks	.0968	.2957	0.0537	0.2255
wk28 37	.7396	.4394	0.5543	0.4973

Variable Type/ Variable		EARLYINF *n -6252*		DATASET *n- 1532*	
Categorial	**Values**	freq	percent	freq	percent
discharge year	1991	1601	.2561	100	11.7
	1992	4651	7439	757	88.3
sex	1 male	3216	51.4	423	49.4
	2 female	3035	48.5	434	50.6
	9 unknown	1	0.0	0	0.0
race	1 white	3187	51.0	350	40.8
	2 black	2838	45.4	415	48.4
	3 asian	97	1.6	22	2.6
	4 amer ind	4	0.1	0	0.0
	5 other	126	2.0	70	8.2
	9 unident.	0	0.0	0	0.0
payment	1 medicare	1	0.0	0	0.0
	2 medicaid	2488	39.8	284	33.1
	4 blue cross	770	12.3	87	10.2
	5 insur co	1012	16.2	153	17.9
	6 other govt	41	0.7	1	0.1
	7 work comp	0	0.0	0	0.0
	8 self-pay	405	6.5	91	10.6
	9 no charge	0	0.0	0	0.0
	10 other	8	0.1	1	0.1
	12 HMO	1252	20.0	240	28.0
	13 MSO	23	0.4	0	0.0
	14med HMO	251	4.0	0	0.0
	99 unknown	1	0.0	0	0.0

(a) calculated as $s = \sqrt{[p(1-p)]}$

(b) For those infants who spent time in both the NICU and the nursery, the number of days spent in the nursery

* - significantly different at p 0.01

NS- not significantly different

APPENDIX B

CALCULATIONS USED FOR SIMULATION MODELS

For all the models used in the process simulation, there are variables that needed to be calculated based on both the anticipated treatment and the likely effect of that treatment. These variables need to be expressed in terms of percentages so that the simulation modeling can use random number generators to randomly select the observations to be treated or affected by treatment. In this appendix, we briefly overview each of the models and the variables that need to be calculated. We review the information that form the basis for the calculations. Then, as indicated in Chapter Four, we provide the detailed calculations for the variables. These calculations are used in the models and are referenced in many of the tables in Chapter Four. Where applicable, we note these used in the tables.

Steroid Models

The use of steroids in the simulation models is based on the information reported in the NIH Consensus Development Conference Statement (NIH, 1994). These models account for both the random selection of infants whose mothers would "receive" steroids prior to delivery and for the random replacement of infants who are "successfully" treated. The random selection rates are based on the current and maximum rates of steroid use that the consensus conference observed. Thus, we model four levels of steroid use - 15%, 30%, 45%

and 60%. These percentages are used as the rates of steroid use as the models are programmed. In Table 4.9, these percentages are given as the prevalence of use of corticosteroids for each of the models. The random replacement rate is also based on the NIH Statement, but is less readily seen.

Odds Ratios

As reported in the statement (NIH, 1994), the anticipated effect of corticosteroid use is to reduce the incidence of RDS as measured by the odds ratio reported in the meta-analysis. In a meta-analysis, an odds ratio indicates the expected effect of the treatment as measured over multiple studies. Odds ratios of 1.0 indicate no effect while those below 1.0 imply a protective effect and those above 1.0 imply an increased risk. Additionally, confidence intervals are usually given and if the confidence interval does not include 1.0 then the results are considered statistically significant with a p-value of 0.05.

Odds Ratio Calculations- General

Odds ratios are used in epidemiological studies known as case-control studies. These are investigations comparing two groups of individuals, some of whom are given treatment (cases) and some of whom are not (controls). In each group, measurements are taken for the number of individuals exposed to a specific disease or condition. Then, the relative probabilities of being in each group are measured, and the odds ratio is calculated from these

probabilities. A two-by-two table recaps the general information for a case-control study as follows (Kleinbaum et al., 1988):

Table B.1. Two-by-two table used to recap case-control study information.

	Exposed	Not Exposed	Total
Cases			
number	n_{11}	n_{12}	n_1
probability	Π_{11}	Π_{12}	1.0
Controls			
number	n_{21}	n_{22}	n_2
probability	Π_{21}	Π_{22}	1.0

The effect of corticosteroids as the difference between the incidence of RDS observed and the incidence of RDS expected. The odds ratio looks at the ratio between the effect and the variation on the effect (Crowley, 1994). The odds ratio can be calculated using the formula below (Kleinbaum et al., 1988).

$$OR=\frac{\Pi_{11}\Pi_{22}}{\Pi_{12}\Pi_{21}}$$

Our models are designed around the odds ratio desired as given in the literature. The probabilities for the control groups, Π_{21} and Π_{22}, are derived from the population. Thus, we need to calculate Π_{11} and Π_{12} from the formula above and the relationship

$$\Pi_{11} + \Pi_{12} = 1.$$

Solving

$$OR=\frac{\Pi_{11}\Pi_{22}}{(1-\Pi_{11})\Pi_{21}}$$

for Π_{11} gives us the following relationship:

$$\Pi_{11} = \frac{\frac{\Pi_{21}}{\Pi_{22}} OR}{1+\frac{\Pi_{21}}{\Pi_{22}} OR}$$

This relationship allows us to infer the desired probabilities for the cases (those that will be given corticosteroids in our models). In turn, these probabilities are used with the rate of corticosteroid administration to determine the number of infants in the case group that would be expected to develop RDS. Given any odds ratio, we could derive a model that allows for the appropriate results. The specific calculations for our models are given in the next section.

Odds Ratio Calculations - Specific

In our investigation, the cases are those individuals "given" corticosteroids while the controls are those who do not receive corticosteroids (directly from the data set). The exposure is the incidence of RDS. The probabilities for the two-by-two table are derived from the incidence of RDS seen in the population data and the desired odds ratio we want to have as a result. From this information, we calculate the probabilities for the cases given

corticosteroids and then derive the number of individuals that should develop RDS from that probability and the rate of corticosteroid treatment for each model.

First, we look at the two-by two table used for case-control studies. Using only our data, we have only the control group (no evidence of corticosteroid administration exists in the discharge data). Thus, the table for our data would look like:

Table B.2. Control group information from base data set.

	Exposed	Not Exposed	Total
Controls			
number	1520	4691	6211
probability	0.2447271	0.7552729	1.0

The probabilities remain constant in the design of each model. Likewise, as corticosteroids are introduced using the simulation models the probabilities in the case cells of the table will also stay the same. We derive these probabilities from the desired odds ratio and the calculations presented in the previous section.

First we calculate the ratio between Π_{21} and Π_{22} (0.2447271/0.7552729 = .324025). The odds ratio reported for the use of corticosteroids on the effect of RDS was 0.49, with a confidence interval of 0.41 to 0.60 (NIH, 1994, p. 128). This can be interpreted as indicating that the use of corticosteroids results in a 49% reduction in the odds of developing RDS. Using the following relationship,

$$\Pi_{11} = \frac{\frac{\Pi_{21}}{\Pi_{22}} OR}{1 + \frac{\Pi_{21}}{\Pi_{22}} OR},$$

we calculate Π_{11} to be 0.137018. This is the number referred to as the exposure rate for RDS in Chapter Four, Table 4.9. This implies that Π_{12} is 1-0.137018 or 0.862982. As with the percentages for the control group, these percentages for the case groups remain constant with each model.

Each corticosteroid model changes the prevalence of use of corticosteroids and thus shifts the number of individuals in the case cells of the table. For instance, in the STER15 models, 15% of the 6211 infants, or 932 "receive" corticosteroids. In the data, with a 24.47% incidence of RDS, 227 of these infants would have been expected to have RDS without corticosteroid treatment. Using the exposure rate of 0.137018, we determine that after corticosteroid treatment, 127 of the 932 infants will be expected to have RDS. Thus, 100 infants (227-127) need to be synthetically replaced in the modeling efforts. The resulting two-by-two table for the STER15 models would then be as follows:

Table B.3. Case-Control table for STER15 model.

STER15	Exposed	Not Exposed	Total
Cases			
number	127	805	932
probability	0.137018	0.862982	1.0
Controls			
number	1292	3987	5279
probability	0.2447271	0.7552729	1.0

This two-by-two table drives the design of the corticosteroid model with a 15% use of corticosteroids. Similar calculations of the numbers in each cell are done for the STER30, STER45, and STER60 models. Below are given the equivalent tables for these models. Table 4.9 in Chapter 4 uses the information from these tables.

Table B.4. Case-Control table for STER30 model.

STER30	Exposed	Not Exposed	Total
Cases			
number	255	1608	1863
probability	0.137018	0.862982	1.0
Controls			
number	1064	3284	4348
probability	0.2447271	0.7552729	1.0

Table B.5. Case-Control table for STER45 model.

STER45	Exposed	Not Exposed	Total
Cases			
number	383	2412	2795
probability	0.137018	0.862982	1.0
Controls			
number	836	2580	3416
probability	0.2447271	0.7552729	1.0

Table B.6. Case-Control table for STER60 model.

STER60	Exposed	Not Exposed	Total
Cases			
number	511	3216	3727
probability	0.137018	0.862982	1.0
Controls			
number	608	1876	2484
probability	0.2447271	0.7552729	1.0

Cesarean Section Models

Infants

The cesarean section models for the infants are based solely on changing the cesarean section. No intermediate criteria or specific strategy is used in the models. Thus, the calculations are quite simple. The desired rate of cesarean section (for instance, 25%) is multiplied by the number of infants in the dataset (6,211) to give the target number of cesarean sections to realize the desired rate (1,553). The difference between this target number and the number in the data set (1,893) gives the number of infants that need to be "selected" to be delivered vaginally and synthetically replaced in the modeling efforts (for the 25% rate, this number is 340). Then, the ratio between the number needed to be changed and the original number in the dataset is calculated and used in the simulation models (340/1893=17.96). These calculations are completely encompassed by Table 4.10 in Chapter Four.

Mothers

The cesarean section models for the mothers are based on changing the rate of surgical delivery based on specific guidelines and the plans for how to reduce that rate (based on the criteria used for performing a cesarean section). As described in Chapter Four, the conditions considered in the modeling efforts are (1) having had a previous cesarean section, (2) breech presentation, and (3) fetal distress. In the population, there is a great discrepancy between the rates of cesarean section for women who have had a previous cesarean section and those who have not had a previous cesarean section. These models attempt to equalize those rates, while not changing the mode of delivery for women with breech presentation or fetal distress. The two-by-two table below gives the number of women who deliver by cesarean section based on whether or not they have had a previous cesarean section for the population.

The simulation models need to randomly select percentages of women to have the delivery mode changed. Thus, first the total number of cesarean sections to be eliminated are calculated in the same manner as described in the infants' models. The results of these calculations are given in Table 4.11 in Chapter 11. Then these totals need to be broken down into the number of cesarean sections to be eliminated from two different sets of women, those who have had a previous cesarean section and those who have not (PCSECT=1 and PCSECT=0 respectively).

Table B.7. Breakdown of Cesarean Sections by Previous Cesarean Section for EARMOM_D.

EARMOM_D	PCSECT=0	PCSECT=1	Total
CSECT = 0			
number	3728	153	3879
percentage	78.46	39.53	75.53
CSECT=1			
number	1023	234	1257
percentage	21.54	60.47	24.47
Total			
number	4749	387	5136
percentage	92.46	7.54	100.0

To derive the simulation models, we want to adjust the overall cesarean section rate from 24.47% to 20%, 15% and 10% respectively. We look at the calculations for the 20% adjustment first. A similar table is created, with the right hand column percentages fixed to the desired rates. This determines how many women that had cesarean sections (CSECT=1) need to be selected to be replaced by women that did not have cesarean sections (CSECT=0). For the case of a 20% cesarean section rate, the number of cesarean sections would drop from 1257 to 1027. Then, preference is given to changing those women that had previous cesarean sections (PCSECT=1) from having cesarean sections during this delivery, with the exception of those that had breech presentations or fetal distress (a total of 627 cases). The number of previous cesarean sections is modeled using the proportions given in the columns (92.46% without previous cesarean section, 7.54% with previous cesarean section). Applying these percentages to the 1027.2 desired cesarean sections for a 20% rate, there would be 949.8 cesarean section deliveries to women that had not had a previous cesarean section and 77.4

deliveries to women that had had a previous cesarean section. The cell numbers for the women without cesarean sections are imputed from these numbers, with those having had a previous cesarean section and not delivering by cesarean section in the model (153) remaining the same. The revised table for the MCSECT20 model is as follows:

Table B.8a. Breakdown of Cesarean Sections by Previous Cesarean Section for MCSECT20.

MCSECT20	PCSECT=0	PCSECT=1	Total
CSECT = 0			
number	3955.8	153	3879
percentage	80.64	66.41	80.00
CSECT=1			
number	949.8	77.4	1027.2
percentage	19.36	33.59	20.00
Total			
number	4905.6	230.4	5136
percentage	95.51	4.49	100.0

This information allows us to calculate the number of cesarean sections that need to be changed in the simulation model. We want to do the selection from those women that did not experience fetal distress or breech presentation. The figures below show the calculations of those percentages.

Table B.8b. Calculations of Percentages to Change in CSECT20 Model

CSECT20 MODEL	PCSECT=0	PCSECT=1	Total
CSECT=1, Base model	1023	234	1257
CSECT=1, desired	949.8	77.4	1027.2
CSECTS to change	73.2	156.6	229.8
# not breech, fetal distress	452	175	627
% to change	**0.1619**	**0.8949**	

The numbers for the CSECTS to change and the % to change are reported in Chapter Four, Table 4.11. As we see above, this means that nearly 90% of women who have previous cesarean sections and delivering by cesarean section again would need to deliver vaginally for the cesarean section rate to drop to 20%. We feel this is probably the limit for this type of change. Thus, for the CSECT15 model, the reduction in cesarean section rate is made only from women who have not had a previous cesarean section nor experienced fetal distress or breech presentation. The two-by-two table for CSECT15 and the percentages to change are thus given below.

Table B.9a. Breakdown of Cesarean Sections by Previous Cesarean Section for MCSECT15.

MCSECT15	PCSECT=0	PCSECT=1	Total
CSECT = 0			
number	4212.6	153	4365.6
percentage	85.87	66.41	85.00
CSECT=1			
number	693	77.4	770.4
percentage	19.36	33.59	15.00
Total			
number	4905.6	230.4	5136
percentage	95.51	4.49	100.0

Table B.9b. Calculations of Percentages to Change in CSECT15 Model

CSECT15 MODEL	PCSECT=0	PCSECT=1	Total
CSECT=1, Base model	1023	234	1257
CSECT=1, desired	693	77.4	770.4
CSECTS to change	330	156.6	486.6
# not breech, fetal distress	452	175	627
% to change	0.7301	0.8949	

Our final model was designed to reduce the cesarean section rate to 10%. It is not possible to make this reduction without modeling the vaginal delivery of cases of breech presentation and fetal distress. In fact, given that cesarean sections are done in 630 cases of breech presentation/fetal distress, this would put the bottom limit of the cesarean section rate at 12.27% (630/5136) without assuming a difference in delivery for those cases. Thus, to derive the 10% rate we do not control for breech presentation and fetal distress. The two-by-two table and percentage calculations are given below.

Table B.10a. Breakdown of Cesarean Sections by Previous Cesarean Section for MCSECT10.

MCSECT10	PCSECT=0	PCSECT=1	Total
CSECT = 0			
number	4469.4	153	4622.4
percentage	90.40	79.81	90.00
CSECT=1			
number	474.9	38.7	513.6
percentage	9.60	20.19	10.00
Total			
number	4944.3	191.7	5136
percentage	96.27	3.73	100.0

Table B.10b. Calculations of Percentages to Change in CSECT10 Model

CSECT10 MODEL	PCSECT=0	PCSECT=1	Total
CSECT=1, Base model	1023	234	1257
CSECT=1, desired	474.9	38.7	513.6
CSECTS to change	548.1	195.3	743.4
% to change	0.5358	0.8346	

Surfactant Models

The surfactant models are designed to mimic the medical outcomes found when infants are given surfactant using three different strategies. Surfactant is used for infants born prematurely to mimic lung development. Medical findings, as reviewed in Chapter Three, show that surfactant reduces the severity but not the incidence of RDS. First we developed a proxy measure for severity and then we developed models by which that severity could be reduced.

Severity Index

The severity measure was designed around the number of comorbidities and/or procedures used in the course of treatment as indicated by the discharge data. The severity index was validated using ANOVA for several outcomes measures as given in Chapter Four, Table 4.14. The ANOVA results are given below.

RDS Severity Categories - ANOVA Results

Dependent Variable: BLOS LENGTH OF STAY

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	149562.80214311	49854.26738104	49.25	0.0001
Error	1516	1534588.21825162	1012.26135769		
Corrected Total	1519	1684151.02039474			

R-Square	C.V.	Root MSE	BLOS Mean
0.088806	112.6941	31.81605503	28.23223684

Source	DF	Anova SS	Mean Square	F Value	Pr > F
SEVERITY	3	149562.80214311	49854.26738104	49.25	0.0001

Dependent Variable: NICU_DYS

NOTE: Due to missing values, only 833 observations can be used in this analysis.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	46092.53735258	15364.17911753	14.57	0.0001
Error	829	874465.69794154	1054.84402647		
Corrected Total	832	920558.23529412			

R-Square	C.V.	Root MSE	NICU_DYS Mean
0.050070	91.11092	32.47836244	35.64705882

Source	DF	Anova SS	Mean Square	F Value	Pr > F
SEVERITY	3	46092.53735258	15364.17911753	14.57	0.0001

Dependent Variable: NICU

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	107701.31852024	35900.43950675	42.15	0.0001
Error	1516	1291272.76305871	851.76303632		
Corrected Total	1519	1398974.08157895			

R-Square	C.V.	Root MSE	NICU Mean
0.076986	149.3944	29.18497964	19.53552632

Source	DF	Anova SS	Mean Square	F Value	Pr > F
SEVERITY	3	107701.31852024	35900.43950675	42.15	0.0001

Dependent Variable: BTRANS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3.38381169	1.12793723	5.97	0.0005
Error	1516	286.55039883	0.18901741		
Corrected Total	1519	289.93421053			

R-Square	C.V.	Root MSE	BTRANS Mean
0.011671	169.4454	0.43476133	0.25657895

Source	DF	Anova SS	Mean Square	F Value	Pr > F
SEVERITY	3	3.38381169	1.12793723	5.97	0.0005

Dependent Variable: BDIED

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.49356634	0.16452211	3.76	0.0104
Error	1516	66.28274944	0.04372213		
Corrected Total	1519	66.77631579			

R-Square	C.V.	Root MSE	BDIED Mean
0.007391	454.0422	0.20909837	0.04605263

Source	DF	Anova SS	Mean Square	F Value	Pr > F
SEVERITY	3	0.49356634	0.16452211	3.76	0.0104

Dependent Variable: BBRTH_WT BIRTH WEIGHT IN GRAMS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	151916572.64288100	50638857.54762730	43.98	0.0001
Error	1516	1745472063.37751000	1151366.79642316		
Corrected Total	1519	1897388636.02039000			

R-Square	C.V.	Root MSE	BBRTH_WT Mean
0.080066	55.53757	1073.01761235	1932.05723684

Source	DF	Anova SS	Mean Square	F Value	Pr > F
SEVERITY	3	151916572.64288100	50638857.54762710	43.98	0.0001

Dependent Variable: DAYS_NUR

NOTE: Due to missing values, only 814 observations can be used in this analysis.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1046.62738703	348.87579568	2.92	0.0331
Error	810	96672.15885376	119.34834426		
Corrected Total	813	97718.78624079			

R-Square	C.V.	Root MSE	DAYS_NUR Mean
0.010711	256.4210	10.92466678	4.26044226

Source	DF	Anova SS	Mean Square	F Value	Pr > F
SEVERITY	3	1046.62738703	348.87579568	2.92	0.0331

Replacement Strategy

The change in severity for each surfactant model was modeled as matrix transformation. We describe the calculations for the rescue model and provide the results of the calculations for the therapeutic and mixed models.

Rescue Model

The rescue model was designed to mimic surfactant therapy given to all infants who developed RDS, a total of 1520 cases. The severity index for these infants was broken down as in the prior severity table below, with an average of 1.29474. The matrix for change was

used to determine the expected resulting severity table, with all infants moving to a lower severity group. The resulting severity index was 0.28694, which in turn provides the rate of change of 71.64%. This percentage was used in the model definition to determine how many infants were changed to a lower severity.

Table B.11. Matrix Calculations for RESCUE models

		MATRIX For Change						
Prior Severity			After				Expected Results	
		Prior	0	1	2	3		
0	431	0	1	0	0	0	0	1152
1	380	1	1	0	0	0	1	300
2	539	2	0.53	0.47	0	0	2	68
3	170	3	0.32	0.28	0.4	0	3	0
Total	1520						Total	1520
Severity	1.295						Severity	0.29

Ratio for change: 71.64%

Therapeutic and Mixed Models

Table B.12. Matrix Calculations for THERAPEUTIC models

		MATRIX For Change						
Prior Severity			After				Expected Results	
		Prior	0	1	2	3		
0	329	0	1	0	0	0	0	827
1	221	1	1	0	0	0	1	186
2	372	2	0.6	0.4	0	0	2	61
3	152	3	0.36	0.24	0.4	0	3	0
Total	1074						Total	1074
Severity	1.32						Severity	0.203

Ratio for change: 69.37%

Table B.13. Matrix Calculations for MIXED models

		MATRIX For Change						
Prior Severity		**Prior**	**After 0**	**1**	**2**	**3**	**Expected Results**	
0	576	0	1	0	0	0	0	1400
1	417	1	1	0	0	0	1	294
2	579	2	0.58	0.42	0	0	2	71
3	193	3	0.37	0.27	0.37	0	3	0
Total	1765						Total	1765
Severity	2.01						Severity	0.287

Ratio for change: 67.37%

BIBLIOGRAPHY

Agency for Health Care Policy and Research (AHCPR). (1993, August). AHCPR program note - clinical practice guideline development. U.S. Department of Health and Human Services, AHCPR Publication No. 93-0023.

Agency for Health Care Policy and Research (AHCPR). (1995, March). Using clinical practice guidelines to evaluate quality of care, volume 1: Issues. U.S. Department of Health and Human Services, AHCPR Publication Number 95-0045.

Agency for Health Care Policy and Research (AHCPR). (1995, March). Using clinical practice guidelines to evaluate quality of care, volume 2: Methods. U.S. Department of Health and Human Services, AHCPR Publication Number 95-0046.

American Medical Association (AMA). (1995). In Toepp, M. C., Kuznets, N., and Herrera, S., (Eds.), Directory of practice parameters: Titles, sources, and updates. Chicago: American Medical Association.

Axelson, O. (1978). Aspects on confounding in occupational health epidemiology. Scandinavian Journal of Work, Environment and Health, 4, 85-89.

Axelson, O. and Steenland, K. (1988). Indirect methods of assessing the effects of tobacco use in occupational studies. American Journal of Industrial Medicine, 13, 105-118.

Bailer, A. J. and Smith, R. J. (1994). Estimating upper confidence limits for extra risk in quantal multistage models. Risk Analysis, 14, 1001-1010.

Ballard R. A., Ballard P. L., Granberg J. P., and Sniderman S. (1979). Prenatal administration of betamethasone for prevention of respiratory distress syndrome. Journal of Pediatrics 94, 97-101.

Banta, H. D., and Luce, B. R. (1993). Health care technology and its assessment: An international perspective. Oxford: Oxford University Press.

Bantam Medical Dictionary. (1981). New York: Bantam Books.

Berwick, D. M. (1994). Eleven worthy aims for clinical leadership of health system reform. JAMA, 272, 797-802.

Berwick, D. M., Godfrey, A. B., and Roessner, J. (1990). Curing health care: New strategies for quality improvement. San Francisco: Jossey-Bass Publishers.

Birkland, C. (1991). Quality at Ruan. Fleet Equipment, 17, 40-42.

Bluth, E. I., Lambert, D. J., Lohmann, T. P., Franklin, D. N., Bourgeois, M., Kardinal, C. G., Dalovisio, J. R., Williams, M.M., and Becker, A.S. (1992). Improvement in 'stat' laboratory turnaround time. Archives of Internal Medicine, 152, p 837-840.

Boynton, A. C., Victor, B., Pine, B. J. II, (1993). New competitive strategies: Challenges to organizations and information technology. IBM Systems Journal, 32, 40-64.

Brackbill, R., Frazier, T., and Shilling, S. (1988). Smoking characteristics of US workers, 1978-1970. American Journal of Industrial Medicine, 13, 5-41.

Brook, R. H. (1990). Health services research: Is it good for you and me? In Lee, P. R. and Estes, C. L. (Eds.), The nation's health (3rd ed.) (pp 339-346). Boston, MA: Jones and Bartlett Publishers, Inc.

Centers for Disease Control and Prevention (CDC). (1993). Rates of cesarean delivery: United States, 1991. MMWR morbidity and mortality weekly report, 42, 285-289.

Charnes, A., Cooper, W. W., Lewin, A. Y., and Seiford, L. M. (1994). Data envelopment analysis: Theory, methodology and applications. Boston, MA: Kluwer Academic Publishers.

Coffey, R. J., Richards, J. S., Remmert, C. S., LeRoy, S. S., Schoville, R. R., and Baldwin, P. J. (1992). An introduction to critical paths. Quality Management in Health Care, 1, 45-54.

Crowley, P. (1994, November). Update of the antenatal steroid meta-analysis: current knowledge and future research needs. In National Institute of Health Publication No. 95-3784, Report of the consensus development conference on the effect of corticosteroids for fetal maturation of perinatal outcomes.

Cunningham, F. G., MacDonald, P. C., and Gant, N. F. (1989). Williams obstetrics, (18th ed.). Norwalk, CT: Appleton & Lange.

Decavalas G., Mastrogiannis D., Papadopoulos V., and Tzingounis V. (1995). Short-term versus long-term prophylactic tocolysis in patients witih preterm premature rupture of membranes. European Journal of Obstetrics, Gynecology, & Reproductive Biology, 59, 143-147.

DiIorio, F. C. (1991). SAS applications programming: A gentle introduction. Boston,MA: PWS-Kent Publishing Company.

Eddy, D. M. (1990). Variations in physician practice: the role of uncertainty. In Lee, P. R. and Estes, C. L. (Eds.),The nation's health (3rd ed.) (pp 347-355). Boston, MA: Jones and Bartlett Publishers, Inc.

Eddy D. M., and Billings J. (1988). Quality of medical evidence: Implications for quality of care. Health Affairs, 7, 19-32.

Efron, B. and Tibshirani, R. J. (1993). An introduction to the bootstrap. New York, NY: Chapman and Hall.

Eisenberg, J. M. (1986). Doctors' decisions and the cost of medical care. Ann Arbor, MI: Health Administration Press Perspectives.

Everett, W. D. (1993). An epidemiological approach to quality assurance in the hospital. Health Care Management Review, 18, 91-96.

Feng, Z., McLerran, D., and Grizzle, J. (1996). A comparison of statistical methods for clustered data analysis with gaussian error. Statistics in Medicine, 15, 1793-1806.

Field, M. J. and Lohr, K. N. (Eds.). (1992). Guidelines for clinical practice - from development to use. Washington, DC: Institute of Medicine.

Fitzsimmons, J. A. and Fitzsimmons, M. J. (1994). Service management for competitive advantage. New York, NY: McGraw-Hill, Inc.

Flagle, C. D. (1977). An integrating approach to health systems. In Coblentz, A. M. and Walter, J. R. (Eds.) Systems science in health care: Proceedings of an international conference, 1976. London: Taylor & Francis Ltd.

Fries, J. F. (1990). An introduction to the compression of morbidity. In Lee, P. R. and Estes, C. L. (Eds.) The nation's health (3rd ed.) (pp 35-41). Boston, MA: Jones and Bartlett Publishers, Inc.

Fuchs, A. R., Fuchs, F., and Stubblefield, P. G. (1993). Preterm birth: Causes, prevention and management. New York: McGraw-Hill, Inc.

Fuchs, V. R. (1990). Has cost containment gone too far? In Lee, P. R. and Estes, C. L. (Eds.) The nation's health (3rd ed.) (pp 253-260). Boston, MA: Jones and Bartlett Publishers, Inc.

Fujiwara T., Konishi M., and Chida S. (1991). Surfactant therapy in neonatal respiratory distsress syndrome. In Yeh T. F. Neonatal therapeutics, (pp 71-86). St. Louis, MO: Mosby-Year Book.

Gail, M. H., Wacholder, S., and Lubin, J. H. (1988). Indirect corrections for confounding under multiplicative and additive risk models. American Journal of Industrial Medicine, 13, 119-130.

Groth, T. L. (1985). The role of biodynamic models in computer-aided diagnosis. In Carson, E. R. and Cramp, D. G. (Ed). Computers and control in clinical medicine. New York: Plenum Press.

Gupta, Y. P. (1991). Emerging productivity and cost control in the hospital industry. National Productivity Review, 351-367.

Guyuron, B., and Zarandy, S. (1994). Does rhinoplasty make the nose more susceptible to fracture? Plastic and Reconstructive Surgery, 93, 313-317.

Haggerty, R. J. (1990). The boundaries of health care. In Lee, P. R. and Estes, C. L. (Eds.) The nation's health (3rd ed.) (pp 112-123). Boston, MA: Jones and Bartlett Publishers, Inc.

Hansen T. and Corbet A. (1991). Disorders of the transition. In Taeusch, H. W., Ballard, R. A., and Avery, M. E. (Eds.). Scheffer and Avery's diseases of the newborn (6th edition) (pp 498-504). Philadelphia, PA: W.B. Saunders Company.

Healey, K. M., Loukota, S. S., Sears, T. D., Miles, R. R., and Galbraith, T. A. (1994). Innovation and dedication. Hospitals and Health Networks, 68-74.

Health Care Financing Administration (HCFA). (1993). Office of the actuary: data from the office of national health statistics.

Heffner, L. J., Sherman, C. B., Speizer, F. E., and Weiss, S. T. (1993). Clinical and environmental predictors of preterm labor. Obstetrics and Gynecology, 81, 750-757.

Horbar, J. D. (1994, November). Antenatal corticosteroid administration and neonatal outcomes for infants weighing 501-1,500 grams: Vermont-Oxford trials network 1991-1992. In National Institute of Health Publication No. 95-3784. Report of the consensus development conference on the effect of corticosteroids for fetal maturation of perinatal outcomes.

Horn, S. D., and Hopkins, D. S. P. (Eds.). (1994). Clinical practice improvement: A new technology for developing cost-effective quality health care. New York: Faulkner & Gray.

Howie R. N. and Liggins G. C. (1977). Clinical trial of antepartum betamthasone therapy for prevention of respiratory distress in preterm infants. Iin Anderson A. B. M., Beard R. W., Brudenell J. M., and Dunn P. M. (Eds). Preterm labour: Proceedings of the fifth study group of the royal college of obstetricians and gynaecologists (pp.281-289). London: Royal College of Obstetricians and Gynaecologists.

Ignatavicius, D. D. and Hausman, K. A. (1995). Clinical pathways for collaborative practice. Philadelphia: W.B. Saunders and Co.

Institute of Medicine (IOM) Committee on Clinical Practice Guidelines. (1990). Clinical practice guidelines: Directions for a new program. Field, M.J., and Lohr, K.N. (Eds.). Washington, DC: National Academy Press, 1990.

Jarvis, G. J. (1994). Obstetrics and gynaecology: A critical approach to the clinical problems. Oxford, England: Oxford University Press.

Jobe A. H. (1993). Surfactant therapy for respiratory distress syndrome. In Klaus M. H. and Farnaroff A. A. (Eds). Care of the high-risk neonate (pp. 257-266). Philadelphia, PA: W.B. Saunders Company.

Jobe A. H. (1994, November). Beneficial interactions of antenatal corticosteroids and postnatal surfactant. In National Institute of Health Publication No. 95-3784. Report of the consensus development conference on the effect of corticosteroids for fetal maturation of perinatal outcomes.

Jobe A. H., Mitchell B. R., and Gunkel J. H. (1993). Beneficial effects of the combined use of prenatal corticosteroids and postnatal surfactant on preterm infants. American Journal of Obstetrics and Gynecology, 168, 508-513.

Johnson, S. P., Griggs, Thomas R., Williams, Marcus L., and McLaughlin, C. P. (1993). ATQM approach to clinical care in cardiac pre-operative evaluations. In Johnston, R. and Slack, N. D. C. (Eds.), Service superiority: The design and delivery of effective service operations (Proceedings of Operations Management Association - United Kingdom) (pp 239-244). Warwick, United Kingdom: Warwick Printing Company Limited.

Johnson, S. P. and McLaughlin, C. P. (1994). Measurement and statistical analysis in CQI. In Kaluzny, A. D. Kaluzny and McLaughlin, C. P. (Eds.). Continuous quality improvement in health care (pp 70-101). Gaithersburg, MD: Aspen Publishers, Inc.

Kahn, H. A. and Sempos, C. T. (1989). Statistical methods in epidemiology. New York, NY: Oxford University Press.

Kanouse, D. E., Winkler, J. D., Kosecoff, J., Berry, S. H., Carter, G. M., Kahan, J. P., McCloskey, L., Rogers, W. H., Winslow, C. M., Anderson, G. M., Brodsley, F., Fink, A., Meredith, L., and Brook, R. H. (1989). Changing medical practice through technology assessment: An evaluation of the NIH consensus development program. Santa Monica,CA: The RAND Corporation.

Kapp, M. B. (1993). The legal status of clinical practice parameters. American Journal of Medical Quality, 8, 24-27.

Kleinbaum, D.G., Kupper, L. L., and Muller, K. E. (1988). Applied regression analysis and other multivariable methods. Boston: PWS-Kent Publishing Company.

Kritchevsky, S. B. and Simmons, B. P. (1991). Continuous quality improvement: Concepts and applications for physician care. Journal of the American Medical Association, 266, 1817-1823.

Lee, P. R. and Estes, C. L. [Eds.]. (1990). The nation's health, (3rd ed.). Boston, MA: Jones and Bartlett Publishers, Inc.

Lefevre, H. L. (1989). Quality service pays: Six keys to success. Milwaukee: Quality Press.

Lewis, D. F., Futayyeh, S., Towers, C. V., Asrat, T., Edwards, M. S., and Brooks, G. G. (1996). Preterm delivery form 34 to 37 weeks of gestation: Is respiratory distress syndrome a problem? American Journal of Obstetrics and Gynecology, 174, 525-528.

Lewis P.J., de Swiet M., Boylan P., Bulpitt C.J. (1980). How obstetricians in the United Kingdom manage preterm labour. British Journal of Obstetrics and Gynaecology, 87, 574-577.

Liggins G. C. and Howie R. N. (1972). A controlled trial of antepartum glucocorticoid treatment for prevention of the respiratory distress syndrome in premature infants. Pediatrics, 50, 515-525.

Lohr, K. N. (1988). Outcome measurement: Concepts and questions. Inquiry, 25, 37-50.

MacMahon, B. and Trichopoulos, D. (1996). Epidemiology: Principles and methods. Boston, MA: Little, Brown and Company.

McLaughlin, C. P. (1994). Strong medicine for health care. OR/MS Today, 21, 25-33.

McLaughlin, C. P. (1996). Why variation reduction isn't everything: A new paradigm for service operations. International Journal of Service Industy Management, 7, 17-30.

McLaughlin, C. P. and Kaluzny, A. D. (1990). Total quality management in health: Making it work. Health Care Management Review, 15, 7-14.

McLaughlin, C. P., Simpson, K. N., and Johnson, S. P. (in press). A methodology for studying learning rates by organizational form in U.S. health care services.

McLean, M., Walters, W. A. W., and Smith, R. (1993). Prediction and early diagnosis of preterm labor: A critical review. Obstetrical and Gynecological Survey, 48, 209-225.

Melan, E. H. (1989). Process management: A unifying framework for improvement. National Productivity Review, 8, 395-406.

Merlo, F., Costantini, M., Reggiardo, G., Ceppi, M., and Puntoni, R. (1991). Lung cancer risk among refractory brick workers exposed to crystalline silica: A retrospective cohort study. Epidemiology, 2, 299-305.

Merlo, F., Fontana, L., Reggiardo, G., Ceppi, M., Barisione, G., Garrone, E., and Doria, M. (1995). Mortality among silicotics in Genoa, Italy, from 1961 to 1987. Scandinavian Journal of Work, Environment, and Health, 21, Supplement 2, 77-80.

Mick, R. and Ratain, M. J. (1994). Bootstrap validation of pharmacodynamic models defined via stepwise linear regression. Clinical Pharmacology and Therapeutics, 56, 217-222.

Miller, J. M. (1988). Maternal and neonatal morbidity and mortality in cesarean section. In Plauche, W.C., (Ed.) Obstetrics and Gynecology Clinics of North America, Cesarean Section (pp 629-638). Philadelphia: W.B. Saunders Company.

Morris, B. and Johnston, R. (1987). Dealing with inherent variability: The difference between manufacturing and service? International Journal of Operations and Production Management, 7, 13-22.

Moore, K. L. (1983). Before we are born. Philadelphia: W.B. Saunders Company.

Morrissey, J. P., Harris, R. P., Kincaid-Norburn, J., McLaughlin, C. P., Garrett, J. M., Jackman, A. M., Stein, J. S., Lannon, C., Schwartz, R. J., Patrick, D. L. and Koch, G. G. (1995). Medicare reimbursement for preventive care: Changes in performance of services, quality of life, and health care costs. Medical Care, 33, 315-331.

Moutinho, L. and Paton, R. (1991). Site seclection analysis in tourism: The LOCAT model. Service Industries Journal, 11, 1-10.

National Institutes of Health (NIH). (1981, October). Cesarean childbirth: Report of a consensus development conference sponsored by the national institute of child health and human development in conjunction with the national center for health care technology and assisted by the office for medical applications of research. Washington, D.C.: U.S. Department of Health and Human Services, NIH Publication Number 82-2067.

National Institutes of Health (NIH), (1994, November). Report of the consensus development conference on the effect of corticosteroids for fetal maturation on perinatal outcomes. Washington, D.C.: U.S. Department of Health and Human Services, NIH Publication Number 95-3784.

Nelson, F. E., Hart, M. K., Hart, R. F. (1994). Application of control chart statistics to blood pressure measurement variability in the primary care setting. Journal of the American Academy of Nurse Practitioners, 6, 17-28.

Nutting, P. A. (1991). Practice-based research laboratories for outcomes and effectiveness research. In Grady, M. L. (Ed.). Primary care research: Theory and methods (conference proceedings) (pp 277-281). Washington, D.C.: U.S. Department of Health and Human Services (AHCPR)..

Office of Medical Applications of Research. (n.d.). Participants' guide to consensus development conferences. Bethesda, MD: National Institutes of Health.

Pachulski, R. T. (1990). OB/GYN notes. Norwalk, CT: Appleton & Lange.

Pidd, M. (1992). Computer simulation in management science (3rd ed.). New York, NY: John Wiley & Sons.

Pomerance J. J. and Richardson C. J. (Eds.) (1993). Neonatology for the clinician. East Norwalk, CT: Appleton & Lange.

Raffel, M. W., and Raffel, N. K. (1994). The U.S. health system: Origins and functions. Albany, NY. Delmar Publishers.

Reisman, A. (1979). Systems analysis in health-care delivery. Lexington, MA: Lexington Books.

Rice, D. P. (1990). The medical care system: Past trends and future projections. In Lee, P. R. and Estes, C. L. (Eds.) The nation's health (3rd ed.) (pp 72-93). Boston, MA: Jones and Bartlett Publishers, Inc.

Robertson, P. A., Sniderman, S. H., Laros, R. K., Cowan, R., Heilbron, D., Goldenberg, R. L., Iams, J. D., and Creasy, R. K. (1992). Neonatal morbidity according to gestational age and birth weight from five tertiary care centers in the United States, 1983 through 1986. American Journal of Obstetrics and Gynecology, 166, 1629-1645.

Rosander, A. C. (1989). The quest for quality in services. Milwaukee: Quality Press.

Roth, A. V. (1993). World class health care. Quality Management in Health Care, 1, 1-9.

Roth, A. V., Johnson, S. P. and Short, N. M. (1996). Strategic deployment of technology in hospitals: Evidence for reengineering. In Heller, O. and Geisler, E. (Eds.). Managing Technology in Healthcare. Boston, MA: Kluwer Academic Publishers.

Roth, A. V. and vanDierdonck, R. (1995). Hospital resource planning: Concepts, feasibility, and framework. Working Paper.

Rothman, K. J. (1986). Modern eEpidemiology. Boston, MA: Little, Brown and Company.

Sackett, D. L., Haynes, R. B., Guyatt, G. H., and Tugwell, P. (1991). Clinical epidemiology: A basic science for clinical medicine. Boston, MA: Little, Brown and Company.

Salive, M. E., Mayfield, J. A., and Weissman, N. W. (1990). Patient outcomes research teams and the Agency for Health Care Policy and Research. Health Services Research, 25, 697-708.

Samuelson, D. (1995, February). Diagnosing the real health care villain. OR/MS Today, 26-31.

SAS Institute. (1988). SAS/STAT user's guide. Cary, NC: SAS Institute, Inc.

Schary, P. B. (1992). A concept of customer service. Logistics and Transportation Review, 28, 341-352.

Schlotzhauer, S. D. and Littell, R. C. (1995). SAS system for elementary statistical analysis. Cary, NC: SAS Institute, Inc.

Selvin, S. (1991). Statistical analysis of epidemiologic data. New York, NY: Oxford University Press.

Senge, P. and Asay, D. (1999). Rethinking the healthcare system. Healthcare Forum Journal, 31, 32-34 and 44-45.

Seo, K., McGregor, J. A., and French, J. I. (1992). Preterm birth is associated with increased risk of maternal and neonatal infection. Obstetrics and Gynecology, 79, 75-80.

Shao, J. and Tu, D. (1995). The jackknife and bootstrap. New York, NY: Springer-Verlang New York, Inc.

Sheldon, A. P., Baker, F., and and McLaughlin, C. P. (1970). Systems and medical care. Cambridge, MA: MIT Press.

Shoemaker, W. C. (1994). Critical path medicine. Critical Care Medicine, 2, 916-926.

Simpson, K. N. and Lynch, S. R. (1994). Cost savings from the use of antenatal steroids to prevent respiratory distress syndrome and related conditions in premature infants. In National Institutes of Health, Report of the consensus development conference on the effect of corticosteroids for fetal maturation on perinatal outcomes. Washington, D.C.: NIH.

Smith, D. J. (1994). Computer simulation applications in service operations: A case study from the leisure industry. The Service Industries Journal, 14, 395-408.

Solomon, M. (1985). Using computers in the practice of medicine. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Steen, P. M. (1994). Approaches to predictive modeling. Annals of Thoracic Surgery, 58, 1836-1840.

Steenland, K. and Beaumont, J. (1989). Further follow-up and adjustment for smoking in a study of lung cancer and acid mists. *American Journal of Industrial Medicine, 16,* 347-354.

Steenland, K., Beaumont, J. and Halperin, W. (1983). Methods of control for smoking in occupational cohort mortality studies. Scandinavian Journal of Work, Environment and Health, 10, 143-149.

Stinnett, A. A. (1996). Adjusting for bias in C/E ratio estimates. Health Economics, 5, 470-72.

Swift, M. B. (1995). Simple confidence intervals for standardized rates based on the approximate bootstrap method. Statistics in Medicine, 14, 1875-1888.

Turnbull, A. and Chamberlain, G. (Eds.). (1989). Obstetrics. New York: Churchill Livingstone.

Ventura S. J., Martin J. A., Taffel S. M., et al. (1993). Advance report of final natality statistics, 1992. Monthly Vital Statistics Report, Vol 42, No 3(S). Hyattsville, MD: National Center for Health Statistics.

Ventura S. J., Martin J. A., Taffel S. M., et al. (1993). Advance report of final natality statistics, 1992. Monthly Vital Statistics Report, Vol 43, No 11(S). Hyattsville, MD: National Center for Health Statistics.

Ventura S. J., Martin J. A., Taffel S. M., et al. (1994). Advance report of final natality statistics, 1992. Monthly Vital Statistics Report, Vol 43, No 5(S). Hyattsville, MD: National Center for Health Statistics.

Walton, M. (1990). Deming management at work. New York, NY: The Putnam Publishing Group.

Wennberg, J. E. (1990). Sounding board: Outcomes research, cost containment, and the fear of health care rationing. New England Journal of Medicine, 323, 1202-1204.

Wennberg, J. E., et al, (1977). Changes in tonsillectomy rates associated with feedback and review. Pediatrics 59, 821-826.

Whicker, M. L. and Sigelman, L. (1991). Computer simulation applications:An introduction. Newbury Park, CA: Sage Publications, Inc.

Wolfenden, H. H. (1923). On the methods of comparing the mortalities of two or more communities, and the standardization of death rates. Journal of Research Statistics in Sociology, 86, 399-411.

Wood, M. (1994). Statistical methods for monitoring service processes. International Journal of Service Industry Management," 5, p 53-68.

Wright L. L., Verter J. and Younes N. (1994, November) antenatal corticosteroid administration and neonatal outcome in infants weighing 501-1,500 grams. In National Institute of Health Publications No. 95-3784. Report of the consensus development conference on the effect of corticosteroids for fetal maturation of perinatal outcomes (pp 49-52).