In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Health Care & Epidemiology

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date Sept. 19, 1988
ABSTRACT

In recent years, concern about the rising costs of health care has prompted the development of programs aimed at reducing utilization of hospital services and facilities while maintaining an acceptable standard of care. One of the major strategies that has emerged in the effort to accomplish these dual objectives, is utilization management. Although there are a number of different approaches, the primary aim of all utilization management programs is to identify and eliminate unnecessary and inappropriate hospital use.

To date, most of the utilization research and program development has taken place in the United States. To a great extent, this effort has focussed on the development and use of norms for utilization based on a breakdown of length of stay data by diagnostic-related groups (DRG's). Canadian interest in this type of approach is reflected in the recent development of data bases defined by case-mix groups (CMG's). However, while continued efforts are being made to refine these schemes, they have been vulnerable to the criticism that they do not provide adequately objective criteria for establishing what constitutes appropriate patterns of hospital use. In addition, because they are based on statistically derived norms, they have been criticized as lacking sufficient clinical relevance to encourage physician support. Since hospital utilization is largely determined by the medical staff, utilization management programs that fail to obtain physician support are unlikely to succeed.

An alternative approach, which appears to be gaining in popularity, involves the formulation of criteria which can be used to determine what constitutes appropriate and necessary hospital use. Essentially, it is argued
that by directly identifying the source and nature of misutilization, it should be possible to develop more effective strategies for the resolution of identified problems. The American Appropriateness Evaluation Protocol designed by Gertman & Restuccia (1981) is one of the earliest and most highly tested examples of a criterion-based system. In Canada, interest in this type of approach is more recent and, consequently, little attention has as yet been focussed on the development and use of clinical criteria in utilization review and management.

One exception, however, is the SWITCH Index System. This system, which was developed and implemented in 1984 by the Peace Arch District Hospital (White Rock, B.C.), makes a direct attempt to identify and eliminate days of hospital stay during which no appropriate acute care services are being provided. The criteria used in this system are classified under the headings Signs, Wind, Intramuscular Therapy, Tubes, Consultant, and Hospice. Patients are considered to be appropriately placed in the hospital if, on any given day, at least one of the specified criteria are met. Otherwise they are classified as Off-Index and action is taken to identify the source of the problem and to initiate corrective action.

Since a major objective of the SWITCH system is to identify and eliminate inappropriate use, an observable outcome, if the program is successful, should be a reduction in length of stay. The present study investigated this hypothesis by comparing pre- and post-intervention length of stay trends at the Peace Arch District Hospital. In addition, to take into account any general secular trends in length of stay over time, the Peace Arch length of stay was compared to the length of stay observed for a control group of three peer-group member hospitals.
Although data covering the four year period 1982 to 1985, indicated that the length of stay at the Peace Arch District Hospital had been decreasing over time, no component of this general decline could be attributed to the SWITCH Index System. Time series regression analyses failed to detect changes in either the slope or the height of the estimated response curve. However, limitations in the study design do not permit any conclusions regarding the potential effectiveness of this system. Characteristics specific to the Peace Arch District Hospital may have prevented the detection of an effect. In addition, because it is likely that there would be a lag between when the program was implemented and when it might be expected to effect a reduction in length of stay, the follow-up period of eleven months may have been too short for the determination of the program's effectiveness.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>x</td>
</tr>
<tr>
<td>Acknowledgement</td>
<td>xi</td>
</tr>
<tr>
<td>CHAPTER I. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>I.1 Statement of the Problem</td>
<td>1</td>
</tr>
<tr>
<td>I.2 Thesis Scope and Organization</td>
<td>4</td>
</tr>
<tr>
<td>CHAPTER II. American Perspectives on Hospital Utilization and</td>
<td>6</td>
</tr>
<tr>
<td>Utilization Management</td>
<td></td>
</tr>
<tr>
<td>II.1 Overview</td>
<td>6</td>
</tr>
<tr>
<td>II.2 Evidence of Excessive Hospital Use</td>
<td>8</td>
</tr>
<tr>
<td>II.2.1 Comparative Geographic Studies</td>
<td>10</td>
</tr>
<tr>
<td>II.2.2 Comparative Cost Reimbursement and Prepaid Practice Studies</td>
<td>11</td>
</tr>
<tr>
<td>II.3 Perspectives on Hospital Utilization</td>
<td>12</td>
</tr>
<tr>
<td>II.3.1 General Conceptual and Normative Perspectives</td>
<td>12</td>
</tr>
<tr>
<td>II.3.2 Modelling the Hospital Utilization Process</td>
<td>16</td>
</tr>
<tr>
<td>II.4 Early Developments in Utilization Review and Management</td>
<td>21</td>
</tr>
<tr>
<td>II.4.1 The Introduction of Retrospective Utilization Review</td>
<td>21</td>
</tr>
<tr>
<td>II.4.2 A New Approach: Concurrent Utilization Review</td>
<td>26</td>
</tr>
<tr>
<td>II.4.2.1 Objectives of the PSRO-Prototypes</td>
<td>27</td>
</tr>
<tr>
<td>II.4.2.2 Characteristics of the PSRO-Prototypes</td>
<td>28</td>
</tr>
<tr>
<td>II.4.2.3 Effectiveness of the PSRO-Prototypes</td>
<td>31</td>
</tr>
<tr>
<td>II.4.3 The Transition to PSRO's</td>
<td>39</td>
</tr>
<tr>
<td>II.4.4 Major Characteristics of PSRO Review</td>
<td>42</td>
</tr>
<tr>
<td>II.4.4.1 Admission Certification</td>
<td>42</td>
</tr>
<tr>
<td>II.4.4.2 Continued Stay Certification</td>
<td>43</td>
</tr>
<tr>
<td>II.4.5 The DHEW Evaluation of PSRO Review</td>
<td>44</td>
</tr>
<tr>
<td>II.4.5.1 Status of PSRO Program Development</td>
<td>44</td>
</tr>
<tr>
<td>II.4.5.2 Effectiveness of PSRO Review</td>
<td>45</td>
</tr>
</tbody>
</table>

(cont...)
### Table of Contents (cont.)

II.4.5.3 Impact of PSRO Review on Health Care Expenditures 49
II.4.6 Other PSRO-Related Research 52
II.4.7 The Development of DRG's 63
  II.4.7.1 The Construction of DRG's 63
  II.4.7.2 Implications for Utilization Review 65
  II.4.7.3 Initial Reactions to DRG's 66
  II.4.7.4 The Reformulation of DRG's 69
  II.4.7.5 Suggested Applications for DRG's 70
II.4.8 An Alternative Approach: Assessing Appropriate Utilization 71
  II.4.8.1 Measuring Appropriate Use 74
  II.4.8.2 Assessing the Reasons for Inappropriate Use 84
  II.4.8.3 Empirical Findings on the Amount and Nature of Inappropriate Use 86
  II.4.8.4 Implications of the Empirical Studies 93
II.5 Current Trends in Utilization Management: The Era of Prospective Payment 95
  II.5.1 The Transition to the Federal Prospective Payment System 97
  II.5.2 Major Objectives of the PPS 99
  II.5.3 Major Characteristics of the Medicare PPS 101
  II.5.4 Implications of PPS for the Cost, Use and Quality of Hospital Care 102
    II.5.4.1 Potential Problems Associated with PPS Incentives 102
    II.5.4.2 Concerns Regarding the Adequacy of DRG's 107
II.5.5 The Role of External Review Organizations Under PPS 115
  II.5.5.1 Goals and Objectives of PRO's 116
  II.5.5.2 Characteristics of the PRO Program 116
II.5.6 The Role of Internal Review Committees Under PPS 119
II.5.7 Current Trends in Utilization Review Procedures 120
II.5.8 Early Effects of the PPS on Hospital Utilization, Costs, and Quality 124
  II.5.8.1 National Trends in Utilization, Costs, and Quality 125
  II.5.8.2 The Results of the New Jersey Experiment: Implications for the Medicare PPS 127
II.5.9 Comments 131

(cont...)
Table of Contents (cont).

CHAPTER III. Utilization Management in Canada: The SWITCH Index System and Other Developments

III.1 The Conceptual Context of Utilization Management in Canada 134

III.2 Current Status of Utilization Management in Canada 142

III.2.1 Retrospective Review Programs 143

III.2.1.1 Diagnosis-Specific Profile Analyses 143

III.2.1.2 Service-Specific Profile Analyses 150

III.2.2 Concurrent Review Programs 151

III.2.2.1 The Hospital for Sick Children 151

III.2.2.2 The Vi-Care Program 153

III.3 The SWITCH Index System 154

III.4 Study Purpose and Rationale 160

III.5 Hypotheses 161

CHAPTER IV. Methods

IV.1 Study Design 163

IV.2 Sample 164

IV.3 Data 166

IV.4 Length of Stay Measures 169

IV.5 Data Analyses 171

CHAPTER V. Results

V.1 Graphical Analyses 174

V.1.1 Unadjusted Average Length of Stay 174

V.1.2 Case-Mix Adjusted Average Length of Stay 174

V.2 Regression Analyses 177

V.3 Supplementary Analyses 187

CHAPTER VI. Discussion

VI.1 The SWITCH Index System 191

VI.2 Alternative Explanations for the Failure to Detect a Program Effect 195

VI.3 Methodological Issues and Study Limitations 198

(cont...)
Table of Contents (cont).

VI.4 Implications for Future Research and Program Planning 200  
VI.5 Policy Implications 204  
VI.6 Conclusion 205  

References 206  

Appendices: 

Appendix A. Medicare Levels of Care Criteria for Medical/ Surgical Patients 221  
Appendix B. Pediatric-Modified Appropriateness Evaluation Protocol 223  
Appendix C. List of Potential Barriers to Appropriate Utilization 224  
Appendix D. Data Elements Included on the HMRI Hospital Discharge Abstract 227  
Appendix E. Release of Data Form 231
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table I</td>
<td>Appropriateness Evaluation Protocol: Day-of-Care Criteria</td>
<td>77</td>
</tr>
<tr>
<td>Table II</td>
<td>Appropriateness Evaluation Protocol: Admission Criteria</td>
<td>83</td>
</tr>
<tr>
<td>Table III</td>
<td>Studies of Inappropriate Hospital Utilization</td>
<td>87</td>
</tr>
<tr>
<td>Table IV</td>
<td>Studies Indicating Reasons for Inappropriate &quot;Day-of-Care&quot; Use</td>
<td>92</td>
</tr>
<tr>
<td>Table V</td>
<td>Studies Indicating Reasons for Inappropriate Admissions</td>
<td>94</td>
</tr>
<tr>
<td>Table VI</td>
<td>SWITCH Index System Criteria</td>
<td>156</td>
</tr>
<tr>
<td>Table VII</td>
<td>Characteristics of PADH and Non-SWITCH Hospitals: 1982-1985</td>
<td>165</td>
</tr>
<tr>
<td>Table VIII</td>
<td>Patient Characteristics of PADH and Non-SWITCH Hospitals: 1982-1985</td>
<td>168</td>
</tr>
<tr>
<td>Table IX</td>
<td>Regression Coefficients Using ALSR as the Dependent Variable: PADH and Non-SWITCH Hospitals - Standard Set of CMG's</td>
<td>179</td>
</tr>
<tr>
<td>Table X</td>
<td>Regression Coefficients Using ALSR as the Dependent Variable: PADH and Non-SWITCH Hospitals - OCC Removed - Standard Set of CMG's</td>
<td>185</td>
</tr>
<tr>
<td>Table XI</td>
<td>Regression Coefficients Using ALSR as the Dependent Variable: PADH and Non-SWITCH Hospitals - Patients Aged 18 to 69 - Standard Set of CMG's</td>
<td>190</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page No.</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Figure 1</td>
<td>Subgroup Partitioning of MDC 55 - Urinary Calculus</td>
<td>66</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Kardex Acuity Index</td>
<td>159</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Hypothetical Trends Detectable with Time Series Regression Analysis</td>
<td>172</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Unadjusted Average Length of Stay by Hospital Group</td>
<td>175</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Case-Mix Adjusted Average Length of Stay by Hospital Group</td>
<td>175</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Case-Mix Adjusted Length of Stay Ratio by Hospital Group</td>
<td>176</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Case-Mix Adjusted Adjusted Length of Stay Ratio by Hospital Group - Patients Aged 18 to 69</td>
<td>189</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENT

The research reported in this thesis was supported by a grant from the British Columbia Health Care Research Foundation (Grant # 87 (86-1). I gratefully acknowledge the members of my thesis committee and, in particular, Dr. S.B. Sheps for his patient support and excellent supervision. Thanks are also extended to the Peace Arch District Hospital and the peer group member hospitals which kindly permitted access to their patient data files, and to the HMRI for their timely assistance in providing the requested data.
CHAPTER I
Introduction

I.1 Statement of the Problem

Historically, the financial and structural mechanisms created in the development of Canada's universal health insurance program did little to encourage economy or efficiency in the delivery of health care services. Rather, they favoured the use and expansion of the most expensive forms of health care. The growth and prominence of the hospital sector, which currently accounts for over 40% of total health care expenditures (National Health Expenditures, 1987), has been a prime example (Soderstrom, 1978; Iglehart, 1986a, 1986b).

During the past decade, however, as federal and provincial governments have become increasingly concerned about the continued affordability of Canada's publicly financed health insurance program (Canadian Comprehensive Auditing Foundation, 1987), economic restraint policies enacted at both levels of government have challenged hospitals to find new ways to improve efficiency and contain costs while maintaining an acceptable standard of care. In response, a few hospitals have begun to develop and implement innovative utilization management strategies designed to identify and eliminate inappropriate and unnecessary use of hospital facilities and services.

Although utilization management is a recent innovation in Canadian hospitals, American hospitals have been experimenting with such programs for over twenty years. Utilization review, which involves "the assessment of the appropriateness and efficiency of hospital care through review of the medical record" (Payne, 1987, p. 109), has, since 1966, been a
mandatory requirement for hospitals participating in government-funded health insurance programs. In subsequent years, additional legislation provided the impetus for the development of the more comprehensive cost-containment programs known as utilization management.

In essence, utilization management involves a "deliberate action by payers or hospital administrators to influence providers of hospital services to increase the efficiency and effectiveness with which services are provided" (Payne, 1987, p.709). More specifically, as defined by the American Hospital Association, utilization management is:

The planning, organizing, directing, and controlling of the health care product in a cost-effective manner while maintaining high quality care and contributing to the overall goals of the institution. This is accomplished through the judicious use of resources to control inappropriate inpatient admission, lengths of stay, and use of ancillary services (Connor, Mack, & Handleman, 1983).

Thus, while utilization review typically involves the measurement and evaluation of hospital use against some pre-established criteria (Dodds, 1974), utilization management incorporates this information into the development of a systematic action plan designed to correct identified problems.

As noted by Anderson & Lomas (1988) and Rachlis & Fooks (1988), although most Canadian hospitals have implemented some form of utilization review, in response to accreditation requirements, few of these programs involve more than measurement. An absence of defined standards for use have, until recently, inhibited the development of evaluation strategies and, consequently, remedial action plans have been almost nonexistent.
In recent years, however, a few hospitals have begun to implement more comprehensive utilization management programs. For the most part, these programs have been designed, in concert with their American predecessors, to facilitate reductions in length of hospital stay. In contrast to the United States, however, few attempts have been made to quantify the amount of inappropriate use reflected in Canadian hospitals' length of stay experience. In addition, to date few systematic attempts have been made to determine whether current utilization management programs are effecting reductions in length of hospital stay nor have any studies been undertaken to compare the cost-effectiveness of the different program initiatives.

Research conducted in the United States suggests that ongoing evaluation of new utilization management programs is critical because such programs can become extremely costly and time-consuming without producing sufficient benefit to warrant their costs. Therefore, since it does appear that interest in utilization management is growing in Canada, it is becoming increasingly important for evaluation studies to be initiated. The study reported in this thesis thus represents a preliminary exploration of the effects of one recently implemented utilization management program. This program, called the SWITCH Index System, was implemented in 1984 at the Peace Arch District Hospital in White Rock, British Columbia. As will be detailed in Chapter III, the SWITCH system was designed to facilitate the identification and elimination of inappropriate and unnecessary days of hospitalization. Assuming that such days exist, and that they are controllable by the hospital and/or its medical staff, then if the system is effective an observable outcome ought to be a reduction in the hospital's
average length of stay. The present study was designed to explore this hypothesis.

1.2 Thesis Scope and Organization

As noted in the American Hospital Association's definition of utilization management, resource management and quality assurance are both important components of the utilization management process. Clearly, if hospitals are to begin emphasizing cost-consciousness in the delivery of hospital services, a system for monitoring the quality of care is essential. However, although resource management and quality assurance are in many respects inseparable (Donabedian, 1969), the focus of this thesis is on the development of systems aimed specifically at improving the efficiency and appropriateness of hospital resource utilization. No attempt is made to provide a comprehensive review of the large and growing body of quality assurance literature.

To date, most of the literature on utilization (resource) management has been based on the American experience. Few publications on the nature and status of utilization management in Canada have as yet begun to appear in the academic literature. Therefore, literature relevant to both Canada and the United States is discussed. In describing the American experience, however, attention is limited to developments that have occurred in response to federal and state initiatives. Innovations in utilization management occurring as a result of the actions of private health insurance agencies are not discussed because 1) public sector developments are likely to be more relevant to Canada, and 2) very little information on private sector developments has been found in the academic literature.
This thesis, which provides a history of utilization management and describes an exploratory evaluation of the Canadian-based SWITCH Index System, is organized as follows:

Chapter II: details the evolution of utilization management in the United States from 1966 to the present, describes the factors believed to influence hospital utilization, in particular, hospital length of stay, and discusses the current status of theoretical models of hospital utilization.

Chapter III: examines the factors prompting Canada's awakening interest in the utilization management process, describes the various program initiatives undertaken in hospitals across Canada, and provides a detailed description of the SWITCH Index System. In addition, the purpose and rationale for the evaluative study are delineated.

Chapter IV: describes the methods used in evaluating the SWITCH Index System.

Chapter V: describes the results of the study.

Chapter VI: discusses the results, the limitations of the study, and the policy implications. In addition, recommendations for future research are made.
CHAPTER II
American Perspectives on Hospital Utilization
and Utilization Management

II.1 Overview

The United States, like most western industrial nations, has experienced rapid growth in health care expenditures during the past twenty-five years. In 1965, for example, total health care expenditures accounted for only 5.9% of the nation's Gross National Product (GNP); by 1987, this total had climbed to 11.4%. Moreover, it has been estimated that by 1990, if expenditures continue to grow at current rates, as much as 12% of the GNP will be devoted to the health care sector (Ginzberg, 1987).

Two of the factors identified as major contributors to the observed increases in health care expenditures include: 1) the establishment, in 1965, of publicly financed health insurance for the elderly (Medicare) and the indigent (Medicaid); and 2) the high, and still rising costs associated with treating patients in hospitals (Lave & Leinhardt, 1976a; Bean, 1988).

Although concerns about the costs and efficiency of publicly financed hospital care remain current, they are not new. For over twenty years, government authorities, health care professionals, and health service researchers have been searching for cost-effective ways to contain the cost and improve the efficiency of the publicly financed hospital sector (Brown, 1986; Dans, Weiner, & Otter, 1985; Lantos, 1984; McMahon, 1984). Complicating the situation, however, has been the problem of how to curtail costs while at the same time maintaining the quality of care at an acceptable standard (Boaz, 1979; Huber, Wolfe & Hardwick, 1974; Kelley,
In the ongoing effort to achieve the dual objectives of cost containment and quality assurance, federal and state authorities have launched a variety of regulatory programs aimed at reducing over-utilization of hospital services and facilities (Brown, 1986). For years, it has been assumed that a substantial portion of hospital use is unnecessary or inappropriate\(^1\) (Dans et al., 1985; Goran, 1979). Thus, during the late 1960's and throughout the 1970's, it was thought that hospital costs could be contained, without jeopardizing the quality of care provided, simply by eliminating excessive use. Accordingly, in 1966, legislation was enacted which called for the establishment of hospital utilization review committees and reimbursement controls (Dans et al., 1985). In the mid-1970's, more stringent utilization controls were implemented and final authority for utilization review and management was delegated to external review agencies called Professional Standards Review Organizations (PSRO's).

Throughout this period, the primary aim of utilization review was to reduce the average length of hospital stays. This emphasis was prompted by a widespread belief that most instances of excessive use occurred during the

---

\(^1\) Unnecessary utilization is defined, in this context, as the provision of care that has no demonstrable effect on patients' outcomes. For the most part, this term is used in reference to hospital stays that are longer than necessary. It can also refer, however, to the use of services or procedures that are ineffective. Inappropriate utilization, on the other hand, can refer either to the admission (treatment) of patients to (in) hospitals for conditions not requiring that intense a level of care (Goran, 1979) or to the failure to provide needed services (Payne, 1987). Thus, inappropriate use can refer to situations reflecting either over- or under-utilization. Historically, inappropriate overutilization has attracted the most attention. More recently, as quality of care has become a more salient issue, problems associated with underutilization have also increased in importance.
latter stages of a hospital episode (Goldberg & Holloway, 1975). In addition, because each day of stay is associated with hotel costs, at the minimum, it was thought that reductions in length of stay would be accompanied by proportionate reductions in hospital costs (Lave & Leinhardt, 1976b). By the 1980's, however, it had become clear that, on their own, utilization review procedures were not sufficient to contain costs (Averill & McMahon, 1977; Boaz, 1979; Dobson et al., 1978). Therefore, in 1983, the federal government enacted legislation which called for a massive restructuring of the Medicare hospital reimbursement system. Specifically, the reimbursement method was changed from cost-based, retrospective reimbursement to case-based, prospective reimbursement (Iglehart, 1982a, 1982b; May, 1985). With the advent of the new prospective payment system (PPS), utilization review has evolved into utilization management; i.e., it has developed from a medical review program into a strategy employed by federal officials and hospital administrators to effect changes in provider and physician behaviour (Payne, 1987). In the transition, the utilization review process has taken on a new focus, emphasizing the appropriateness of hospital level care rather than the length of the hospital episode. The basic assumption, however, is the same as before, namely, that a substantial portion of hospital use is unnecessary or inappropriate (Hughes et al., 1984).

II.2 Evidence of Excessive Hospital Use

During the 1950's, private health insurance companies had begun to promulgate the notion that insured hospital services were being "abused" (i.e., overutilized) by patients and physicians (Shain & Roemer, 1959). Aware of such claims, a number of researchers began trying to identify and measure the factors that were contributing to overall patterns of hospital
utilization and misutilization. In one of the most important of these early studies, Shain & Roemer (1959) found that hospital utilization (measured as hospital days per thousand population) was strongly related to bed supply (beds per thousand population). While acknowledging that utilization tended to increase with capacity, these investigators did not concur with the general opinion that such a relationship indicated abuse. Instead, they suggested that in the presence of excess supply, physicians' perceptions of what constituted appropriate hospital use were modified; i.e., that as capacity increased, the range of medical conditions deemed appropriate for hospitalization widened. As noted by Evans (1984) and McLure (1982), subsequent studies have so consistently supported the original findings of Shain & Roemer, that their conclusion ("beds that are built tend to be used") has become known as "Roemer's Law". In fact, to date, Roemer's Law remains one of the most commonly accepted generalizations about hospital use (Evans, 1984). Moreover, because it suggests that there is, to some extent at least, a discretionary component in hospital utilization decisions, it has itself been cited as one of the earliest indications of excessive hospital use (McLure, 1982).

Since the publication of Shain & Roemer's classic study, numerous other investigations have been undertaken in an effort to identify and measure unnecessary and inappropriate hospital use. The studies that have been primarily responsible for promoting the notion that a considerable portion of hospital use is unnecessary have been grouped into two categories: 1) Comparative geographic studies, and 2) Studies comparing fee-for service with prepaid practice groups. Although both of these groups of studies provide only indirect evidence of excessive hospital use, supporting
evidence from a third group, one that attempts to define and measure inappropriate use more directly, has recently begun to appear in the literature. However, since these studies have been done within the context of refining and modifying current utilization review procedures, they are not included here. Instead, they are described in section II.4.8.

II.2.1 Comparative Geographic Studies

During the past twenty years, there has been a proliferation of utilization studies based on geographic comparisons. For the most part, these studies have emphasized length of hospital stay as the chosen measure of hospital utilization. However, other measures that have been employed include patient days per thousand population, rates of surgical procedures, and admission rates. A detailed discussion of the problems associated with the definition and measurement of hospital utilization is provided in a section II.3.2.

As noted by the Congressional Office of Technology Assessment (OTA) in an advisory published in 1983, the results of studies comparing average length of hospital stay have been consistent and persistent over time. The Northeast region of the United States has tended to have the longest average length of hospital stay while the West has had the shortest (OTA, 1983). According to OTA data, the eastern hospital lengths of stay have been approximately forty percent higher than those in the west.

The reported variations in length of stay among geographic regions over time do not, in and of themselves, provide evidence to support the assumption of unnecessary use. They could, for example, reflect differences in regional demographics (e.g., age distributions) or severity of illness. In addition, it could be that regions with shorter lengths of stay have poorer
patient health outcomes. In reviewing the literature on regional lengths of stay, however, neither Evans (1984), Goran (1979), McLure (1982), nor the OTA (1983) was able to find evidence which suggested that the observed variations in length of stay could be adequately accounted for by differences in demographic or severity of illness variables. Similarly, as noted by Goran (1979) and McLure (1982), population studies standardized for age and sex, comparing regional admission rates, and days of hospital care per thousand population consistently reported wide variations in utilization rates. Likewise, Wennberg and Gittlesohn (1982), who found substantial variations in surgical rates across 193 geographic areas in six New England states, were unable to attribute their results to differences in demographic, economic or patient condition variables. Rather, such differences were deemed to be more closely associated with the supply of physicians within a given area and their preferred patterns of practice. Finally, although there is general agreement that the evidence on health status outcomes associated with utilization has been somewhat equivocal, there is a consensus that high levels of hospital use are not a guaranty for quality (Enthoven, 1983; Evans, 1984; McLure, 1982; OTA, 1983; Wennberg & Gittlesohn, 1982).

II.2.2 Comparative Cost Reimbursement and Prepaid Practice Studies

The costs and utilization rates associated with prepaid practice groups, such as Health Maintenance Organizations (HMO's), have frequently been cited to support claims that expenditures and utilization by patients with traditional insurance coverage (i.e., cost-based hospital reimbursement and fee-for-service physician reimbursement) may be unnecessarily high (Enthoven, 1978, 1983; Luft, 1978). To examine the validity of such claims, Luft (1978) conducted a major review of comparative studies performed
between 1958 and 1971. With respect to costs, he found that, in general, the studies he reviewed did support claims that expenditures were lower for HMO's. In particular, he noted that the Kaiser Permanente plan, one of the oldest and most well known of the HMO's, had total per capita costs that ranged from 10 to 40 percent lower than those with traditional insurance coverage. In addition, he observed that most of the difference in costs was attributable to reductions in hospital use. Prepaid practice groups were found to have between 15 and 40 percent fewer admissions and between 25 and 45 percent fewer hospital days per thousand population.

Although questions about the comparability of the populations and the quality of care provided can be raised, little evidence has been provided to suggest that either concern can account totally for the observed results. Many of the studies reviewed by Luft had incorporated population standardization procedures. In addition, as noted by Evans (1984), observed reductions in use apparently have not affected mortality or morbidity, two commonly used measures of quality. Despite these observations, however, the effect of prepaid group practice on quality of care remains a matter of debate (Ginzberg, 1985; Luft, 1978). Furthermore, as Luft (1978) and Enthoven (1983) have indicated, the issue of whether prepaid practice group achieve reductions in costs and use through selection of healthier patients, has not yet been resolved.

II.3 Perspectives on Hospital Utilization

II.3.1 General Conceptual and Normative Perspectives

When the United States Congress enacted the 1965 Social Securities Act it provided a mandate for the establishment of two publicly-financed health insurance programs, Medicare and Medicaid. Medicare, a federal
program, was to pay for hospital and medical services provided to the elderly and the disabled; Medicaid, a joint federal-state program, was to pay for services provided to the indigent (Enthoven, 1983).

The passage of this legislation clearly reflected the conceptual and normative perspectives that dominated the health care scene at the time. According to Brown (1986), the development of the American health care system, prior to the 1970's, was guided by six beliefs or propositions. These he summarized as follows:

- First, curing illness is the primary function of the health care system. An improved health care system will improve health status outcomes.
- Second, one of the main problems with the health care system is that there is too little of it.
- Third, the only other serious problem is limited financial access to mainstream care among disadvantaged citizens, notably the poor and the elderly.
- Fourth, payments for professional and institutional services should be based on the actual incurred costs of hospitals and the usual and customary charges of physicians, so long as these are reasonable.
- Fifth, in health care, markets do not and cannot work.
- Sixth, in health care, regulation is the wrong answer to market failures.

(Adapted from Brown, 1986, p. 572).

As originally formulated, the Medicare and Medicaid legislation embraced the propositions outlined above. Within this framework, the only role envisioned for public policy intervention was in the removal of financial barriers to care (Brown, 1986; Evans, 1984). Accordingly, the primary aim of the legislation was to increase the quality and quantity of health services available, and to ensure that the elderly and the indigent had access to these services. Moreover, reimbursement strategies were to
include actual cost (retrospective) reimbursement for hospitals and fee-for-service payments for physicians. In line with the dominant views on reimbursement and regulation, the only constraint placed on reimbursement was that submitted claims reflect 'reasonable' costs (Grimaldi & Micheletti, 1984).

Within this framework, referred to as the "traditional" (Ro, 1969) or "Medico-Technical" (Evans, 1984) model, utilization controls were not considered necessary because the utilization of hospital and other health care services were assumed to be a direct response by providers to the medical needs of patients, subject only to currently available medical technologies (Evans, 1984; Ro, 1969). Any observed variations in use were explained solely in terms of differing medical philosophies about the "technologically-prescribed method of treatment" (Ro, 1969, p. 298) and, occasionally, 'extra-medical' factors such as the age and sex of the patient.

When the legislation was enacted, however, the experience of private insurers with "overutilization" of insured hospital services had already begun to suggest that utilization decisions were not a "micromanifestation of objective, scientific, professional judgement" (Brown, 1986, p. 574). Instead, it appeared that such decisions were subject to the influence of such non-clinical factors as medical practice pattern preferences and the availability of facilities and services (Shain & Roemer, 1959), insurance coverage (Studnicki & Honemann, 1983), income-increasing objectives (McLure, 1982; Goran, 1979), and hospital organizational procedures (Studnicki & Honemann, 1983; Querido, 1963). Therefore, the inclusion of open-ended reimbursement strategies and the absence of utilization controls in the Medicare/Medicaid legislation were interpreted by some as reflecting
a rather naive acceptance, on the part of the government, of the "traditional" viewpoint (Brian, 1972, 1973). Consequently, critics of the federal program began to raise serious questions about the impact public financing would have on the utilization and, thereby, the costs of insured hospital services (Fulchiero, Miller, Foley, Ballantine, & Amorosino, 1978). In particular, it was feared that the provision of public health insurance would promote overutilization of hospital services and, consequently, inflate hospital costs. In essence, it was argued that with the provision of public insurance, the function of price, as a stabilizer between supply and demand, would be negated (Evans, 1984). Thus, by removing price as a barrier to care, it was anticipated that utilization would increase excessively, subject only to the constraint of available supply. Moreover, because reimbursement was to be provided on an actual cost basis, it was feared that hospitals would be motivated to promote overutilization. Under the federal reimbursement scheme, hospitals could increase their revenues by admitting more patients and by keeping them in hospital longer. In addition, physicians, who were responsible for admitting and discharging patients, received a higher fee for treating patients in hospitals than in less costly settings. Hence, incentives favouring excessive hospitalization were believed to be operating on both the hospital and its physicians (Enthoven, 1983; May, 1985). To counteract this possibility, the federal government amended the original Medicare/Medicaid legislation to include utilization review requirements. The amended legislation was enacted in 1966 and provided a major impetus for the subsequent proliferation of utilization review and management programs and hospital utilization research.
In the following pages, the evolution of utilization review from its initial formulation under the 1966 amendment to its present status will be described. In addition, the literature describing the effectiveness of the various utilization strategies will be reviewed. Before turning to that discussion, however, it is important to examine the literature that has attempted to define, measure, and model the factors that influence hospital utilization decisions. For the most part, the regulatory mechanisms that have been implemented, over the years, in the effort to reduce excessive hospital use, have focussed on length of hospital stay as the primary measure of hospital utilization. Consequently, a considerable body of literature has grown up in an attempt to identify and quantify the factors that are important in determining the length of a hospital episode. A knowledge of such factors, and how they interact, is necessary if predictions about the effects of policy initiatives are to be made, and if evaluations of the effectiveness of utilization review/management programs are to be meaningful. Therefore, the remainder of this section will be devoted to the general hospital utilization literature. In particular, studies that have attempted to model length of stay will be reviewed.

II.3.2 Modelling the Hospital Utilization Process

Hospital utilization is a complex, interactive process involving patient, physician, and hospital characteristics. In addition, it is vulnerable to the influence of policies aimed at effecting changes in hospital and physician behaviour. Thus, for example, it is subject to the influence of changes in reimbursement strategies and insurance coverage. A model of hospital utilization that fails to include a description of the implicit or explicit model of the hospital to which it refers will, therefore, be limited in
its usefulness. As noted by Evans (1984), however, most of the literature that has been accumulated on hospital utilization "has been at best loosely linked to theoretical models of hospital behaviour" (p.185). Generalizations induced from empirical research have been far more common. Moreover, the attempts that have been made, to date, to develop formal models of hospital behaviour have not been very successful (Evans, 1984). The numerous processes, factors, and interrelationships that govern hospital behaviour are extremely complicated and, often, idiosyncratic and culture-bound. Consequently, as Evans states,

"[r]ealistic description and analysis of such complex processes can lead to very useful generalizations about how hospitals are likely to behave, and to respond to changes in their external environments, but they may never be expressed in a formal analytic framework which is either realistic or useful" (p. 173).

Thus, although as yet lacking a formal theoretical framework, the literature on hospital utilization has provided some valuable insights into the factors that influence patterns of utilization and misutilization. For example, on the basis of the research done by Shain and Roemer (1959), it is now widely accepted that physicians' utilization decisions can be modified by changes in bed capacity. Likewise, in contrast to the "traditional" view of utilization, an accumulation of research has contributed to widespread acceptance of the view that utilization decisions, "though predominantly based on clinical considerations, are also influenced by nonmedical factors" (Boaz, 1979, p. 316). Thus, for example, studies demonstrating wide variations in surgical rates across geographical regions (Wennberg & Gittlesohn, 1982), and in admission rates within diagnosis-defined patient groups (Wennberg, McPherson, & Caper, 1984), have contributed to
acceptance of the notion that utilization is affected by factors such as physician supply and practice pattern preferences.

As mentioned previously, the major variable employed in the formulation of utilization policies, the development and evaluation of utilization management programs, and the conceptualization of hospital utilization models, has been length of hospital stay. This emphasis on length of stay has occurred for several reasons. First, when the initial utilization review regulations were enacted in 1966, a measure of utilization was needed that could be obtained, objectively and easily, nation-wide. Length of stay satisfied these requirements (Goldberg & Holloway, 1975). Second, when utilization review procedures were first being developed, "no definitive standard separating necessary from unnecessary use [had] been established" (Goran, 1979, p. 7). Evidence available from comparative studies, however, suggested that most instances of misutilization were occurring during the latter stages of inpatient episodes. Although this belief has since been dispelled, it was a major factor giving rise to an emphasis on hospital lengths of stay (Goldberg & Holloway, 1975; Zimmer, 1974). Third, length of stay has traditionally been accepted as an important measure of hospital performance and resource consumption (Berki, Ashcraft, & Newbrander, 1984; Lave & Leinhardt, 1976b; Ro, 1969). Finally, length of stay has been identified as one of the most important variables contributing to hospital costs, since the hotel costs associated with each bed day are, on average, the single most costly component of each hospital episode (Knapp, Speedie, Yaeger, & Knapp, 1980; Lave & Leinhardt, 1976b).

With the advent of utilization control policies, a major impetus was provided for hospital utilization research. Previous studies, such as those
described above, in the section on unnecessary utilization, had generated questions about the premise that utilization decisions were based solely on medical and technological imperatives. But, little was as yet known about the relative importance of non-clinical variables on outcome measures such as length of hospital stay. As one of the major aims of the newly enacted utilization review requirements was the elimination of unnecessary or inappropriate use, as measured by reductions in length of stay, a better understanding of the factors, other than utilization review, that could influence the length of a hospital episode was becoming increasingly important. In particular, information concerning the variables that needed to be taken into account when attempting to evaluate the effectiveness of utilization review strategies was sought.

In response to the need for more information, several researchers attempted to identify variables that could, potentially, influence length of stay (Becker, Shortell, & Neuhauser, 1980; Berki et al., 1984; Boaz, 1979; Boone, Coulton, & Keller, 1981/82; Lave & Leinhardt, 1976b; McLure, 1982; Ro, 1969). It was hoped that the results of such studies would eventually lead to the development of a formal model of hospital utilization (based on length of hospital stay). To date, however, no comprehensive and realistic model of hospital length of stay has been developed. The major factors inhibiting the development of such a model have been the lack of appropriate data and, accordingly, the inability to study more than a few variables at a time (Berki et al., 1984; Studnicki & Honemann, 1983).

It has become clear, however, that two of the most important determinants of length of stay are the "patient's health status at entry and the medical condition prompting the admission" (Lave & Leinhardt, 1976b,
In addition, a positive relationship between the patient's age on admission and the length of the inpatient episode has consistently been reported in the literature (Berki et al., 1984; Goldfarb, Hornbrook, & Higgins, 1983; Lave & Leinhardt, 1976b; Ro, 1969). More specifically, clinical variables such as diagnosis, admission status, severity of illness, complications, and number of surgical procedures, as well as patient age, have repeatedly been found to be more important determinants of length of stay than non-clinical variables - for example, occupancy rate, income, insurance coverage, distance from hospital, size of the hospital, teaching status of hospital, day of the week of admission, type of accommodation in hospital, physician age, physician specialty - (Berki et al., 1984; Boaz, 1979; Goldfarb, Hornbrook, & Higgins, 1983; Goldfarb, Hornbrook, & Rafferty, 1980; Lave & Leinhardt, 1976b; Ro, 1969). This is not to say, however, that non-clinical variables are unimportant. Significant and, often, complex relationships have been identified between each of the above variables and length of stay. But, in comparison to the importance of the clinical variables, their overall influence has been small. For example, in one of the most comprehensive studies (Lave & Leinhardt, 1976b), variables such as the patient's primary diagnosis, number of diagnoses, number of surgical procedures, admission and discharge status accounted for 40 per cent of the variability in log length of stay. The addition of hospital variables such as occupancy rate and day of week admitted plus patient variables such as age, marital status, race and sex increased \( R^2 \) by only 2 per cent. Thus, while many of these variables were found to have significant beta coefficients, their overall contribution to the explanatory power of the regression model was small. However, because only a few of the potentially important non-clinical variables have as yet
been included in any one study, the simultaneous effects and interactions among these variables have not been adequately tested (Berki et al., 1984; Lave & Leinhardt, 1976b). Nonetheless, while efforts to develop a realistic model of hospital length of stay continue, it has become clear that any studies that attempt to assess the effects of non-clinical variables on length of hospital stay must control, at the minimum, for differences in clinical attributes. Thus, for example, efforts to evaluate the effectiveness of utilization review or management programs on length of hospital stay will be meaningful only if the important effects of patient age and medical condition have first been taken into account. In addition, because complex relationships have been found to exist between length of stay and variables such as hospital size, bed capacity, occupancy rate and teaching status, an attempt should be made to ensure the comparability of hospitals included in the analysis.

II.4 Early Developments in Utilization Review and Management

II.4.1 The Introduction of Retrospective Utilization Review

As mentioned in section II.2, one of the earliest indications of hospital overutilization came from private health insurance companies (Shain & Roemer, 1959). In response, a number of hospitals began to develop and implement utilization review programs. In addition, the insurance companies themselves began to institute retrospective claims review procedures. By 1960, trustees of the American Medical Association had approved in principle the idea of utilization review. In 1963, it was recommended by the Joint Commission on Accreditation of Hospitals and, by 1966, it had become a standard for accreditation (Foster, 1968).
Despite these private sector initiatives, the major impetus for what has since become the proliferation of utilization review and management programs was the 1965 enactment of the Medicare and Medicaid legislation. In response to concerns that public insurance would promote overutilization of hospital services, and lead to an unacceptable cost burden, the federal government decided to follow the lead provided in the private sector and pursue cost containment by invoking utilization controls (Fulchiero et al., 1978; Lantos, 1984). Accordingly, in 1966, utilization review was made a mandatory requirement for hospitals participating in the Medicare and Medicaid programs (Knowlton, 1975; Lantos, 1984; Schmitz & Schoenhard, 1976). In addition, the fiscal intermediaries responsible for administering the plans, were mandated to conduct retrospective claims review and to deny payment for care that was deemed to reflect unnecessary or inappropriate use of hospital facilities or services. In particular, claims were to be denied if it was concluded that the services provided could have been provided in a less costly setting (eg. outpatient clinic) or with fewer days stay in hospital (Flashner, Reed, White, & Norris, 1972; Goldberg & Holloway, 1975; Snider, 1977).

Under the terms of the amended legislation, each participating hospital was required to establish a Utilization Review (UR) committee. This committee was to be comprised of physicians who were to be responsible for ensuring that hospital utilization was both appropriate and medically necessary. In addition, the committee was to review the utilization of professional services, giving particular emphasis to the ascertainment of overuse or underuse of such services (Lantos, 1984). The primary objective of the committee was to be "the maintenance of high
quality patient care, and an increase in the effective utilization of hospital services" (Health Insurance for the Aged, 1966).

As initially enacted, the regulations governing the utilization procedures to be implemented by the hospital UR committees provided few guidelines with respect to when and how the reviews should be done. Accordingly, the utilization review programs that were implemented tended to emphasize retrospective analysis of inpatient episodes (Flashner, Reed, White, & Norris, 1972; Snider, 1977). That is, patient stays were, for the most part reviewed long after the actual patients involved had been discharged from the hospital. An exception to this process, however, occurred with patient stays of extended duration. Under the terms of the utilization review regulations such cases had to be reviewed on the 12th and 18th days of hospitalization and every 30 days thereafter (Schmitz & Schoenhard, 1976). This emphasis on extended stays derived from a general belief that most instances of misutilization would occur during extended hospital stays. Although this assumption would later be challenged, at the time it was widespread and pervasive (Zimmer, 1974).

In addition to utilization review, the newly mandated UR committees were also required to conduct, on an ongoing basis, medical care evaluation studies (MCE's; Knowlton, 1975). These studies, essentially medical audits, were designed to assure that the quality of care provided was of an adequate professional standard. In particular, they were to "emphasize quality improvement through continuing medical education targeted specifically to correct identified problems" (Goran, Roberts, Kellogg, Fielding, & Jessee, 1975, p. 15). At the time, however, a lack of adequate tools for quality of care assessment discouraged the effective development of
MCE's (Goran et al., 1975). Moreover, because the federal regulations included the possibility of retroactive claims denial, more emphasis was placed on utilization review than on quality assessment (Flashner, Reed, White, & Norris, 1972).

Over the next five years many UR committees were established and a variety of utilization review programs were developed and implemented. As noted above, with the exception of extended stay cases, almost all of the patient reviews were conducted retrospectively. Likewise, the reviews conducted by the fiscal intermediaries were retrospective. Moreover, when instances of misutilization were identified, hospital claims were denied retroactively. This procedure put the hospitals at considerable financial risk for situations over which they had little control. Physicians were responsible for admitting and discharging patients from the hospital; the hospital administration could do little to intervene in this process. Consequently, the threat of retroactive denials became a continuing source of tension between hospital administrators and their medical staffs (Flashner, Reed, White, & Norris, 1972; Lave & Leinhardt, 1976a; Tom, 1976).

By 1970, it was generally conceded that the federal effort to control utilization was not working (Dans et al., 1985; Flashner, Reed, Coburn, & Fine, 1973; Flashner, Reed, White, & Norris, 1972; Fulchiero et al., 1978; Tom, 1976). In a review of several studies which attempted to evaluate the effectiveness of retrospective review procedures, Foster (1968) was unable to find any evidence to support claims that utilization review reduced length of stay or admission rates. It should be noted, though, that all of the studies reviewed by Foster suffered from major methodological problems. Thus, in
the absence of any well designed studies, Foster's conclusions may have been premature. In a later review of the literature, however, Clendenning, Wolfe, Shuman, & Huber (1976) found only one significant change that could be attributed to the federal utilization review effort; namely, that "hospitals with high lengths of stay in 1966-1967 had a larger reduction in lengths of stay by 1969-1970 than hospitals which started with lower lengths of stay" (p. 753). The possibility that this reduction may simply reflect a regression to the mean effect, however, cannot be discounted.

Although empiric evidence on the effectiveness, or lack of effectiveness, of retrospective review was limited, the fact that hospital costs were continuing to increase prompted the conclusion that the federal effort had been a failure (Brian, 1972; Dans et al., 1985; Fulchiero et al., 1978). Many reasons were cited for the program's lack of success. One of the most frequent criticisms was the retrospective nature of most of the review procedures. Since these procedures did not identify utilization problems until the patient had already left the hospital, the UR committee could do little to effect change in physician behaviour. Moreover, the threat of retroactive denial of claims, and the tensions this generated, did little to encourage cooperation between the hospital and its medical staff (Flashner, Reed, White, & Norris, 1972; Fulchiero et al., 1978). Consequently, support for the program was difficult to maintain.

Other reasons cited to explain the program's failure include the inability or unwillingness of UR committees to uphold decisions that had a negative impact on their peers, haphazard enforcement of the regulations, poor coordination between the activities of Medicare and Medicaid, and the lack of appropriate norms for evaluating the medical necessity and
appropriateness of care (Flashner, Reed, White, & Norris, 1972; Flashner, Reed, Coburn, & Fine, 1973; Fulcheiro et al., 1978).

II.4.2 A New Approach: Concurrent Utilization Review

By the late 1960's, acknowledgement of the failure of the federal review strategy to control either hospital costs or use had prompted several responses. First, state Medicaid authorities had begun reporting that, at current rates of cost increases, Medicaid trust funds would soon be insufficient to meet demands for medical care (Flashner, Reed, White, & Norris, 1972). Consequently, they began to appeal to regional Foundations for Medical Care (FMC's) for assistance in developing more effective utilization review strategies (Flashner, Reed, White, & Norris, 1972; Frederick, 1975). Second, in 1970, Senator Wallace Bennett (R., Utah) introduced a proposed amendment to the federal Medicare and Medicaid legislation which called for the establishment of Professional Standards Review Organizations (PSRO's). Finally, and also in 1970, the National Center for Health Services Research and Development launched the Experimental Medical Care Review (EMCRO) program (Sanazaro et al., 1972).

The combined impact of these initiatives was the development of new approaches to utilization review which emphasized: 1) the use of admission screening or certification procedures to set initial expected lengths of stay (target dates) for each hospital admission, 2) the establishment of concurrent rather than retrospective review procedures, and 3) the institution of extended stay certification procedures. Until the

---

1 FMC's are voluntary medical review organizations which represent a large number of physicians practicing within a defined geographic region.
mid-1970's, however, most of the new programs were developed as a result of the state Medicaid and EMCRO initiatives. The enactment of the PSRO legislation, introduced by Senator Wallace in 1970, did not take place until 1972, and the regulations delineating PSRO responsibilities were not established until 1974. Nonetheless, by 1975, the PSRO's had gradually begun to replace the state and EMCRO programs and, by 1981, they had been established in 187 of 195 geographically designated areas (Dans et al., 1985).

To a large extent, the major components of the federal PSRO programs were derived directly from the pre-PSRO Medicaid program strategies. Moreover, because the transition to PSRO's took so many years, most of the evaluative studies done during the 1970's concentrated on the Medicaid concurrent review programs. Some of the major pre-PSRO programs, developed between 1970 and 1973, that were later used as prototypes for the development of the PSRO program initiatives included the Certified Hospital Administration Program (CHAP), the Hospital Admission and Surveillance Program (HASP), the Predischarge Utilization Review Program (PDUR), the New Mexico Experimental Medical Care Review Organization (NMEMCRO), and the Commonwealth Health Agencies Monitoring Program (CHAMP). The objectives and major characteristics of these programs are described below.

II.4.2.1 Objectives of the PSRO-prototypes

Although the various prototype programs listed above differed to some extent in administrative details, they all had the same objective. The primary goal was cost containment, which was to be achieved through reductions in inappropriate and unnecessary hospital use. Like the federal retrospective review program, however, this goal was tempered, at least
implicitly, by the caveat that such reductions were to be obtained without jeopardizing the quality of care provided. The EMCRO programs, in fact, were established with quality assurance as their primary objective (Sanazaro et al., 1972). Moreover, important developments in quality assurance did take place as a result of the EMCRO initiatives (Goran et al., 1975). Nonetheless, cost concerns were so widespread at the time that all of the programs, the EMCRO's included, tended to emphasize cost containment objectives over quality assurance activities.

With the advent of concurrent review, the major strategy employed to achieve cost containment involved the development and use of norms for utilization based on length of hospital stay. As with the federal retrospective review program, it was still widely assumed that most cases of misutilization would be found during the latter days of an inpatient episode. Therefore, because each extra day in hospital added hotel costs at the minimum, such days were targeted for elimination. But, whereas the federal program required extended stay reviews of all patients, regardless of diagnosis, at certain specified times, the new concurrent review programs began to use diagnosis-specific length of stay norms (Goldberg & Holloway, 1975).

II.4.2.2 Characteristics of PSRO-prototypes

As noted above, the 1966 federal utilization review regulations placed responsibility for the review process was placed in the hands of hospital utilization review committees. With the new approach to utilization review, this responsibility was transferred to either the state Medicaid authority, a regional FMC, or some other designated external reviewer. This change was initiated in an effort to avoid the criticism that hospital UR
committees were often reluctant to "uphold difficult and unpopular
decisions which could have an adverse effect on their peers" (Flashner,
Reed, White, & Norris, 1972, p.1154). In addition, the previous emphasis on
retrospective review, which had been found to be both ineffective and
contentious, was replaced, in all of the new programs, with concurrent
review procedures. Typically, these procedures included two components:
1) admission screening and the approval of a specified length of hospital
stay, and 2) extended stay certification for cases that could not be discharged
within the time frame approved upon admission. While all the new
programs incorporated extended stay certification, the extent to which
admission screening was carried out varied from one program to another.
Thus, CHAP (Brian, 1972, 1973) and the NMEMCRO (Brook & Williams,
1976; Brook, Williams & Rolph, 1978) required preadmission certification
for all elective hospital admissions and length of stay certification, within 24
to 48 hours after admission, for emergency cases. The HASP (Flashner,
Reed, White, & Norris, 1972) and the CHAMP (Fulcheiro et al., 1978)
required screening of admissions for certain diagnoses and for physicians
with a history of hospital misutilization, as well as length of stay
certification for all cases, emergent and elective, within 24 hours of
admission. The PDUR (Clendenning et al., 1976; Lave & Leinhardt, 1976a),
on the other hand, had no admission screening procedure, only a
prospectively determined length of stay certification. For programs with
admission screening procedures, whether pre-admission or at the time of
admission, the purpose was to identify and eliminate inappropriate
hospitalizations. The length of stay certification process, based on diagnosis-
specific length of stay norms calculated for the region or for each individual
hospital, was intended to eliminate days of hospitalization which could not be substantiated on the basis of medical need.

Within 24 hours before the expiration date for the initial length of stay certification, all of the programs required that the attending physician request extended stay certification for those patients who were not to be discharged. The diagnosis-specific norms used to set the initial target-date, however, differed from one program to the other. The CHAP, HASP, NMEMCRO, and CHAMP all used data compiled by the Professional Activities Studies (PAS) for the state in which the program was located. With the exception of the CHAMP, all of these programs set their initial target dates using the 50th percentile length of stay for the diagnostic groupings recorded by the PAS; the CHAMP used the 75th percentile. In contrast, the database used by the PDUR to establish initial lengths of stay was the Hospital Utilization Project (HUP). The HUP recorded hospital data, by county, for the State of Pennsylvania. The PDUR, therefore, used 50th percentile norms for the county, rather than the state, in which the program had been launched.

Typically, nurse coordinators or physician consultants were hired by the external reviewer to perform the necessary reviews and to grant the necessary approvals. These reviewers were located in each hospital; off-site review was only required in situations where there were conflicting views or physician-initiated appeals. When the reviewer was a registered nurse, rather than a physician, admission or length of stay requests could not be denied without authorization from a designated physician consultant.

As an incentive for hospitals and physicians to cooperate with the new approach, some of the programs (i.e., CHAP, HASP, CHAMP)
abandoned the previously used control mechanism of retroactive claims denial. By requiring that admissions and length of stay be prospectively certified such a punitive mechanism was deemed unnecessary (Flashner, Reed, White, & Norris, 1972; Flashner, Reed, Coburn, & Fine, 1973).

Despite these major changes in the approach to utilization review, few attempts were made, prior to 1972, to determine whether the new concurrent review programs were any more effective in containing costs or reducing utilization than were the federal retrospective review programs. Thus, the subsequent passage of the PSRO legislation was based, for the most part, on anecdotal reports and the intuitive appeal of the review mechanisms that were being employed (Sanazaro, 1977). The ongoing development of PSRO policy formulation, however, depended heavily on the perceived effectiveness of the prototype programs (Sayetta, 1976). Consequently, between 1972 and 1979, a number of evaluative studies began to appear in the literature.

II.4.2.3 Effectiveness of the PSRO-Prototypes

In early evaluations of the effectiveness of two of the pre-PSRO programs, Brian (1972, 1973) and Flashner, Reed, Coburn and Fine (1973) maintained that significant reductions in average length of stay had been obtained with the CHAP and HASP, respectively. Brian further claimed that admissions had declined by 11.4% in the eight month period following implementation of CHAP. In addition, both sets of investigators estimated potential cost savings, as a result of the observed reductions in admission rates and length of stay, that ranged from $9 million to $250 million.

As these were among the earliest attempts to evaluate the effectiveness of concurrent review programs their initial impact on PSRO
activity was considerable. Subsequent critiques of the studies (Davidson, Wacker, & Klein, 1973; Sayetta, 1976; Clendenning et al., 1976), however, cited weaknesses in the evaluation methods that were sufficient to invalidate the purported program successes. Specific concerns included the absence of appropriate comparative data, the use of baseline periods that were too short to establish appropriate utilization trends, and the failure to report the methods used to compute expected utilization rates in the follow-up periods. The major criticism, however, was that in none of the investigations did the researchers attempt to adjust for differences in the case mix among the hospitals studied, or within the hospitals over time. Hospital utilization research had, by this time, already begun to suggest that one of the most important factors contributing to hospital length of stay was the patient's medical condition or diagnosis (Lave & Leinhardt, 1976b; Ro, 1969). It was concluded, therefore, that no evaluations of program effectiveness aimed at reducing length of stay would be valid if, at the minimum, they failed to control for this important variable.

In an attempt to address some of these criticisms, Wylie & Flashner (1979) reevaluated the HASP program using data from the six month period prior to the program's implementation (July to December, 1971) and the comparable six month period following its implementation (July to December, 1972). A total of 46 Illinois hospitals, covered by HASP, were included in the study (for a total of 327,943 discharges). To adjust for possible case mix variations in the two time periods, each Medicaid patient discharge was matched to a comparison group of non-Medicaid patients according to the following characteristics: final diagnosis, age, presence of more than one diagnosis, and whether or not surgery had been involved.
The comparison group was formed using PAS data for all hospitals in the Central Region of the United States during the calendar year 1971. This group was used to compute 'expected lengths of stay' for each Illinois study patient in each of the pre- and post- intervention periods.

In analyzing the data, Medicaid patients in the HASP hospitals were compared to non-Medicaid patients in the HASP hospitals. (The HASP procedure was used with Medicaid patients only). For each group, the observed average length of stay in the before and after periods were computed. In addition, expected values for each group, in each period, were computed using the larger Central Region hospital statistics. The authors reported a reduction in per patient length of stay in the follow-up period that was significantly greater for the Medicaid (0.8 day reduction) than the non-Medicaid (0.2 day reduction) group (p<.01). They concluded, therefore, "that concurrent review, as implemented by HASP, helped shorten [length of stay] in 1972" (Wylie & Flashner, 1979, p. 607). In a change from the study done in 1973, no claims were made regarding the program's cost-reducing effects. The authors did maintain, however, that the program had contributed to a slow down in growth.

Although this study represented a methodological improvement over the earlier studies, it was still problematic. In particular, the use of the calendar year 1971 for the calculation of expected rates, instead of the comparable six month periods from 1971 and 1972, raises questions about the validity of the authors reported conclusions. By not using the 1972 regional data, the possibility of decreasing trends in the comparison group were not taken into account. It was also unfortunate that they did not expand the pre- and post- study periods beyond the reported six months in
each direction. Longer time periods would have helped to establish clearer trends over time and, also, to show whether the observed effects were temporary, enduring or simply regression to the mean. Finally, their comments regarding the program's effects on costs were based solely on impressions derived from the observed reduction in length of stay. Since no attempt was made to analyze the program's effect on admission rates, and given the lack of any analysis of cost data, their contention that the program had slowed growth was unsubstantiated.

Subsequent studies, designed to evaluate the effectiveness of the PDUR (Clendenning et al., 1976; Lave & Leinhardt, 1976a), the NMEMCRO (Brook & Williams, 1976; Brook et al., 1978), and the CHAMP (Fulcheiro et al., 1978), were more thorough. With the exception of the CHAMP study, however, none of these evaluative efforts were able to demonstrate a reduction in length of stay specifically attributable to the introduction of concurrent, target based review. Clendenning et al. in their comparison of 11 PDUR and 11 non-PDUR hospitals (matched for bed size, location, teaching status, and number of services offered) examined length of stay data on Medicaid patients, all of whom were under the age of 65. Thus, in contrast to the HASP study, Medicaid patients were used in both the PDUR and non-PDUR groups. This is an important distinction because, in this study, the Medicaid patients included in the non-PDUR group were also subject to review. The difference between the two groups, however, was that the PDUR group was subject to the new target-date based concurrent review whereas the non-PDUR group was subject to retrospective review and retroactive claims denial.
In analyzing the data, to control for differences in case mix, 14 of the most common Medicaid diagnoses were examined. Data encompassing eleven months prior to the implementation of the PDUR was compared to data encompassing eleven months after implementation (Feb 1, 1972 - Dec. 31, 1972 and Feb 1, 1973 - Dec. 31, 1973, respectively). With the exception of maternity cases (i.e., two of the fourteen diagnoses), no differences in average length of stay could be attributed to the PDUR. The authors concluded, therefore, that, the concurrent review process had not been any more effective than the policy of retroactive denials in reducing length of stay, with the possible exception of maternity cases.

In a second evaluation of the PDUR (Lave & Leinhardt, 1976a), essentially the same conclusion was reached. In this study, Medicaid patients in 9 PDUR hospitals were compared to Medicaid patients in 8 non-PDUR hospitals. The non-PDUR hospitals had all implemented the federally required post-discharge utilization review programs and, in contrast to PDUR hospitals, were subject to retroactive claims denial. Thus, the primary purpose of the evaluation was to determine whether pre-discharge utilization review was any more effective than post-discharge review in reducing length of hospital stay. In addition, comparisons were made between Medicaid patients in the PDUR and non-PDUR hospital and Blue Cross patients in the same hospitals. This was done to determine whether the PDUR or post-discharge review programs differed in effectiveness from the non-regulated review programs used by private insurers.

To analyze the data, regressions were done using institutional characteristics such as occupancy rate, bed capacity, teaching status and
percentage of hospital population covered by Medicaid as control variables. As well, the dependent variable, length of stay, was adjusted to control for case mix. The results indicated that under both the PDUR and non-PDUR programs, significant reductions in length of stay had been incurred over time. Moreover, these reductions were, for the most part, more significant than those observed for Blue Cross patients. In comparing the PDUR to the non-PDUR hospitals, however, no difference in length of stay was observed. Downward trends had started prior to the introduction of the PDUR and, therefore, the authors concluded that the observed reductions in length of stay could not be attributed to the post-discharge review program. Instead, it appeared that both pre-discharge and post-discharge programs were equally effective. Consequently, the authors suggested that, prior to further policy development, more attention should be given to the determination of which method was more cost-effective and, also, to the ascertainment of the impact of different utilization review strategies on the quality of care provided.

In 1978, Brook and his associates published the findings from a major study that had been undertaken to ascertain the impact of the NMEMCRO on Medicaid hospital costs, use, and quality (Brook et al., 1978). These investigators reviewed data accumulated over a four year period (1971-1975) relating to the use of ambulatory services, hospital admissions, hospital days billed per 100 eligible persons, average length of stay, hospital costs, and cost savings. In all cases, the primary method of analysis used was time series regression. Where possible, comparisons to national trends were used in order to account for secular trends. However, no pre-program baseline was established. Instead, it was argued that because the program had been
implemented slowly (between September 1971 and March, 1972) and because there were no observed discontinuities following its implementation, it was appropriate to compare the later months to the earlier months of the study period.

The major findings of this study are summarized as follows:

1) With the exception of medically reviewed injections, ambulatory service use (Prescriptions, office visits, and laboratory visits) increased over time. In contrast, utilization of reimbursable injections decreased by 75% between 1971 and 1975. The largest decreases were observed among antibiotic injections such as short-acting penicillin, lincomycin, ampicillin, and tetracycline. It was concluded that, since all of these injections were used primarily to treat viral infections, the observed "decreases [reflected] clinically significant improvements in the quality of care" (Brook, et al., 1978, p. 257). The authors attributed these reductions directly to NMEMCRO activities.

2) Neither admission certification nor extended stay certification had any demonstrable impact on length of hospital stay (whether adjusted for case mix or not), total hospital days, or the proportion of stays exceeding the 50th PAS percentile. Moreover, there was evidence to suggest that admission rates increased, rather than decreased, during the study period.

3) The percentage increase in Medicaid expenditures per eligible beneficiary per year was significantly greater than the percentage increase in services billed per eligible person. Consequently, Brook et al. suggested that, even if peer review was successful in curtailing utilization, it would be ineffective as a cost containment device because of cost increases associated with price inflation, technological innovations, and changes in case mix.
4) The costs of operating the concurrent medical care review program were between $750,000 and $1,000,000 per year whereas total savings attributed to this process averaged only $700,000 per year. Thus, the savings incurred did not cover the costs of operating the program. Substantial cost savings (approximately $16,000,000) were realized as a result of NMEMCRO administrative activities (eg. downward adjustments of hospital rates, comparisons of claims to fee schedules, and payment denials for administrative or medical reasons). However, none of these activities required professional medical review.

Although these findings are interesting, it should be recalled that they refer to the experience of only one state, the State of New Mexico. Consequently, it is impossible to say whether similar results would be obtained in other jurisdictions. In addition, the lack of both a pre-program baseline and a comparable control group make the results difficult to interpret, and the authors' conclusions open to debate.

In the final evaluative study to be considered here, Fulchiero, et al., (1978) compared the length of stay experience of Medicaid patients to that of non-Medicaid patients, in the same 57 hospitals. Medicaid hospitalizations were subject to CHAMP review whereas non-Medicaid hospitalizations were not reviewed. A total of seven six month periods were studied; two periods before the implementation of CHAMP and five periods after implementation. In each period, the data was standardized for age and case mix.

The results of the CHAMP study indicated a consistently decreasing trend in length of stay for Medicaid patients that began in the pre-program period and continued after the implementation period. In comparing the
average of all pre-CHAMP periods to the average of all post-CHAMP periods, the age and case-mix adjusted length of stay for the Medicaid sample decreased by 11.9%. A similar trend was observed for non-Medicare patients, but the magnitude of the decrease was smaller for this group (6.6%). The authors concluded, therefore, that a 5.3% differential decrease could be attributed to the CHAMP. Moreover, because the two groups studied were from the same hospitals, the authors suggested that the decrease observed in the non-Medicaid group may, at least in part, have been due to an educational effect of the CHAMP that was not limited to the Medicaid patients.

It should be emphasized, however, that this evaluation used non-Medicaid patients for its length of stay comparisons. Since these patients were not subject to systematic review, this study does not address the issue of whether concurrent review is any more effective than post-discharge review. In fact, the observed decrease in length of stay during the pre-CHAMP period, when only the federal retrospective review program was in effect, suggests that the results attributed to the CHAMP may need to be interpreted with caution.

II.4.3 The Transition to PSRO's

It should be recalled at this point that the above mentioned studies were performed in response to, rather than as precursors of, the 1972 PSRO legislation. Thus, the implementation of the PSRO review system was already well underway by the time evaluators first began to suggest that concurrent utilization review, a major component of the PSRO system, was no more effective than the previously mandated retrospective review strategy. Given the magnitude of the PSRO program and the substantial
costs associated with its operation (Burford & Averill, 1979) it is somewhat surprising that a more intensive effort to establish the effectiveness of the program's major components did not precede its implementation. Despite the absence of objective data, however, at the time the legislation was passed, conventional wisdom held that concurrent review programs would be effective in reducing unnecessary hospital use and, thereby, in containing hospital costs (Sanazaro, 1977). Consequently, the implementation of the program was initiated, and only later did questions concerning the effectiveness and cost-effectiveness of concurrent review begin to attract attention.

As initially enacted, the PSRO program, like its prototypes, was intended to have two basic objectives: "1) to assure the quality and 2) to assure the appropriate utilization of health care services" (Goran et al., 1975, p. 2). These objectives were to be met through the implementation and integration of three review mechanisms: 1) concurrent review, which was to include admission and continued stay certification procedures similar to those developed by the prototypes; 2) MCE's (medical care evaluation studies), which were first introduced with the 1966 Medicare/Medicaid amendments and were to continue to be concerned with the assessment of the quality of care provided; and 3) profile analysis, a new procedure which was to be used to identify potential problems associated with either the utilization or quality of services provided. From the outset, however, the PSRO program was regarded, by the health care community, as primarily a cost containment strategy (Berry, 1977; Dans et al., 1985; Goran, 1979). This perception was reinforced when, in 1977, the responsibility for the PSRO's was transferred from the Public Health Service to the Health Care Financing
Administration (Dans et al., 1985). Because of this perception, most PSRO resources were used to implement the labor intensive concurrent review procedures (Burford & Averill, 1979; Goran et al., 1975). MCE's, which had been required under the previous federal review system, continued to be developed and implemented. As with the previous system, however, they still received relatively less attention than the utilization review procedures (Dobson et al., 1978). The profile analysis component of the PSRO program was, for the most part, ignored (Dobson et al., 1978; Goran, 1979). Thus, throughout the 1970's the implementation of the PSRO program was predominantly characterized by the development and instigation of admission and extended stay certification procedures.

Although the utilization review procedures employed by the PSRO's were similar to the procedures used by the earlier state-initiated review systems, the PSRO's themselves had a number of distinctive characteristics. These have been summarized by Dobson et al. (1978) as follows:

1 PSROs are structured organizations created solely for the purpose of implementing the law. Previously, utilization review had been left to individual hospitals, with supervisory responsibility divided between state certification survey agencies and fiscal intermediaries.

2 PSROs are physician-sponsored organizations, in contrast to state agencies or fiscal intermediaries.

3 PSRO memoranda of understanding with state agencies and fiscal intermediaries assure that PSRO review decisions on medical care necessity are binding on the fiscal agent, thereby precluding any retrospective denial of the claim by the fiscal agent on the grounds the stay is not medically necessary. Prior to PSRO, utilization review called for decisions by fiscal intermediaries on all issues of patient eligibility, coverage, and medical necessity.

4 PSRO implementation has been formally guided through a structure of transmittals, support centers, and technical
assistance. This has tended to produce greater uniformity than other utilization strategies.

5 PSRO concurrent review employs uniform norms and standards established on an areawide basis. In contrast, norms for other forms of concurrent review have varied from hospital to hospital. (p. 117).

Thus, the PSRO system represented a major effort on the part of the federal government to develop a "truly national utilization review program" (Dans et al., 1985, p. 1131). Within this national system, however, each PSRO was to develop its own standards and criteria for determining appropriate and necessary hospital use (Berry, 1977). Thus, although it was hoped that the PSRO program would reduce the observed variability in regional length of stay patterns, no specific targets were set for the reduction of excessive hospital use (Goran, 1979).

II.4.4 Major Characteristics of PSRO Review

II.4.4.1 Admission Certification

Under the PSRO mandate, all Medicare and Medicaid admissions were to be reviewed to assure that a hospital level of care was medically necessary. Each admission deemed appropriate after PSRO review was then certified and assigned an initial length of stay. In most instances, the admission certification procedure was performed by a specially trained review coordinator. Typically, the reviewer was a registered nurse who was hired by the PSRO to conduct reviews in a particular hospital. Admissions deemed appropriate could be certified by the review coordinator without PSRO consultation. However, decisions concerning questionable admissions could only be made a PSRO physician consultant.

In general, the criteria used to determine the appropriateness of admissions could be of two types: "1) criteria specific to a particular problem,
diagnosis, or procedure, and 2) criteria which specify the types of services which should be provided at a hospital level of care" (Goran et al., 1975, p. 9). The actual criteria to be used were to be developed by each PSRO and applied, on a uniform basis, to each hospital within the PSRO's jurisdiction.

Once a patient's admission was certified, the assignment of the initial length of stay checkpoint could be made on the basis of 1) regional length of stay norms for patients with similar characteristics such as age and diagnosis, or 2) hospital level of care criteria or standards, as defined by the PSRO (Goran et al., 1975). However, regional norms, such as those provided by PAS, were to be used whenever possible and, in general, the initial checkpoint was not to exceed the 50th percentile of such norms.

II.4.4.2 Continued Stay Certification

Continued stay review was to be performed on all patients still in the hospital on the assigned target date. It was designed to assure the necessity of an extended stay and, like the admission certification procedure, was to be performed by the review coordinator. The certification of an extended stay was to be based on either hospital level of care criteria, PSRO defined discharge criteria, or diagnosis specific criteria (Goran et al., 1975). When continued stay review indicated that an extended stay was necessary the review coordinator was to certify the extension and assign a new length of stay checkpoint. In questionable cases, the decision to certify or deny an additional stay was to be made by the PSRO physician consultant. This process of continued stay review would then be repeated each time the length of stay checkpoint was reached until the patient was ready for discharge.
With the implementation of the two components of the concurrent review process, admission certification and continued stay certification, retroactive denials of claims were eliminated. The certification process was to be binding on the fiscal intermediary and, therefore, was to ensure that, once certified, the claims of eligible beneficiaries for services covered under Medicare and Medicaid would be paid (Dobson et al., 1978).

For the most part, conclusions about the effectiveness of the PSRO strategy in reducing the utilization and costs of publicly financed hospital care were based on the results of the studies which evaluated the PSRO prototype programs. However, in accordance with a requirement of the 1972 PSRO legislation, a formal evaluation of the PSRO program was undertaken, in 1977, by the US Department of Health, Education and Welfare (DHEW). According to Dobson et al. (1978), this study had three major objectives. First, it was to document the status of PSRO implementation. Second, it was to determine whether the concurrent review procedures employed in the PSRO program were more effective than other forms of review (retrospective and prototype) in reducing length of hospital stay. Third, it was to determine what impact the PSRO review system was having on health care expenditures.

II.4.5. The DHEW Evaluation of PSRO Review

II.4.5.1 Status of PSRO Program Development

As noted by Dobson et al. (1978), the planning, development and funding of PSRO programs was based on a three stage process. During the

---

1 The DHEW evaluation was summarized in a paper published by Dobson, et. al. (1978). All of the material contained in this section (II.4.5.1 - II.4.5.3) is therefore based on the Dobson et. al. publication.
initial phase, when review systems and organizational structures were being developed, the PSRO was considered to be in the 'planning stage'. Once the necessary systems and structures were in place, the PSRO was considered to be in the 'conditional stage'. The 'conditional stage' was to last a minimum of 24 months and was designed to ensure that the PSRO was operationally viable. At the end of the 'conditional stage', by which time the PSRO was expected to be performing both short-term and long-term care review, the PSRO was to be designated as 'operational'. Under the federal mandate, a total of 203 geographically defined areas were identified for PSRO development.

As indicated by Dobson et al. (1978), when the DHEW's evaluation of the PSRO system was undertaken in 1977, 108 PSRO's were operating at the operational stage and 64 were in the planning stage. None had as yet been designated fully operational. The implementation of review activities had typically begun with 100% concurrent review in short-term hospitals. MCE studies were implemented next, usually about one year after the concurrent review procedures. Profile analyses, on the other hand, had not yet received much attention.

II.4.5.2 Effectiveness of PSRO Review

According to Dobson et al. (1978), several different strategies were employed in the effort to ascertain the effectiveness of PSRO concurrent review in reducing hospital utilization. One approach involved the assessment of the impact of PSRO review on three measures of utilization: 1) days of care per thousand Medicare patients, 2) admissions per thousand Medicare patients, and 3) average length of stay of Medicare patients. For these analyses, hospital utilization data from 18 active PSRO areas were
compared with data from 26 non-active areas. In order to isolate the effects
of PSRO review, active and non-active PSRO's were matched on 15
demographic and health system variables (the actual variables used for
matching were not delineated). In addition, statistical adjustments were
made to control for: baseline utilization levels, population density, percent
of families with income under $5000 per annum, hospital beds per 100,000
population, weighted PSRO-wide hospital occupancy rates, and prior
Medicaid concurrent review. The calendar year 1974 was used to establish
the baseline utilization rates for both active and non-active areas; data from
the calendar year 1976 were used to establish utilization rates in the post-
PSRO period. Data from the baseline period was used to forecast expected
utilization rates and these expected rates were compared to the actual rates
observed in 1976. The database used to perform the analyses was derived
from Medicare claims files.

On an aggregate level, no significant differences were found between
the active and non-active areas on any of the three utilization variables
(days of care, admissions, length of stay). However, a significant interaction
was observed between bed capacity and PSRO impact. Greater reductions in
days of care per thousand Medicare patients were observed among PSRO's
with a higher bed capacity than among those with fewer beds per thousand
population. This interaction could, perhaps, be explained through
application of Roemer's Law. That is, prior to PSRO review, hospitals with
greater bed capacity may have been providing a larger amount of
discretionary care. If so, then with the advent of PSRO review, these
hospitals would have had more discretionary days of care available for
elimination than the hospitals with more limited bed capacity.
When the data were disaggregated by active PSRO area, 7 of the 18 active PSRO's were found to be more effective in reducing hospital utilization than were their matched comparisons. The other 11 areas showed either no systematic variation in utilization rates or higher utilization rates than their matched comparison. In addition, although Dobson et al. (1978) did not provide any data, they reported that "of the seven PSRO's showing lower actual rates of days of care than expected, six had imputed dollar savings greater than the incremental costs of PSRO review, i.e., six were found to be cost-effective." (p. 119). These results were not given much weight, however, as the overall proportion of programs exhibiting an effect was no different than would be expected by chance.

Although one of the main purposes of the above analyses was to ascertain the effectiveness of the concurrent utilization review component of the PSRO program, this objective was hampered by the fact that even in the non-active PSRO areas some forms of concurrent review were being employed. Since no data were available which would enable ascertainment of the extent to which concurrent review was being used in non-active PSRO areas, Dobson et al. (1978) maintained that the results concerning concurrent review were inconclusive. In the comparison of PSRO versus non-PSRO review, however, it was concluded that no evidence of a differential effect had been found.

The second approach used in the evaluation involved an analysis of 1.19 million hospital records derived primarily from the PAS discharge abstract data service. For this analysis, 92 PSRO hospitals were compared to 42 non-PSRO hospitals. The two groups of hospitals were examined for differences in length of hospital stay and case mix severity for both Medicare
and Medicaid patients. In performing the length of stay analyses, the data were adjusted to control for case mix, geographic region, major hospital and medical resources characteristics. For the case mix analysis, length of stay trends in the PSRO and non-PSRO groups were observed for patients matched on diagnosis, age, number of diagnoses, and whether or not a surgical procedure took place. In addition, the severity of case mix was examined using the SysteMetrics 'Staging Methodology' for 23 specific diagnoses (Dobson et al., 1978, p. 121).

The results of the length of stay comparisons indicated that, for Medicare patients under PSRO review, average length of stay fell by .23 of a day more than for Medicare patients not under PSRO review. Conversely, it was found that for Medicaid patients under review the average length of stay fell, but by .20 of a day less than for Medicaid patients not under review. The combined impact of PSRO review, for both Medicare and Medicaid patients, was a .09 day reduction. Finally, a significant (p < .05) decrease in the percentage of Medicare patients staying beyond the 75th percentile of matched comparisons in the PAS data base was associated with PSRO review. A similar but nonsignificant decrease was observed for Medicaid patients. Since the actual percentage decreases were not reported, however, it is impossible to tell whether the observed reductions represented a meaningful effect.

With respect to the case mix analysis, it was observed that in both PSRO and non-PSRO hospitals there was a trend towards increasing severity of case mix. This observation was based on the finding that expected lengths of stay for all matched patient groups increased over time. However, using a severity of illness staging method developed by Systemetrics, Inc., it was
observed that among Medicaid patients, the trend towards increasing case mix severity was more pronounced in the PSRO than the non-PSRO hospitals. Problems with the reliability of the severity index and the coding of diagnostic data, however, rendered these findings suggestive rather than conclusive.

II.4.5.3 Impact of PSRO Review on Health Care Expenditures

On the basis of the utilization data, Dobson et al., concluded that, at best, PSRO review had produced only minimal reductions in hospital utilization rates. As one of the major implicit goals of the PSRO system was cost containment, however, the next step taken in the evaluation was to determine 1) how the PSRO system could be expected to affect expenditures and 2) what influences it would have to have on utilization in order to be considered an effective cost containment device. In addition, given that one of the primary assumptions prompting the development of the PSRO system was that inappropriate hospital utilization was at least in part responsible for increasing expenditures, an attempt was made to determine what impact PSRO review had on the appropriateness of inpatient care. Finally, an effort was made to estimate the costs associated with PSRO review and to perform a cost-benefit analysis.

In considering the extent to which the PSRO system could be expected to affect health care expenditures, Dobson et al. list six factors over which the system would have to be able to exert some control over in order to have a significant effect on costs. These factors include:

1) Price of materials and wages;
2) Rate increase in complexity of health care services provided due to technological advancement and shifts in medical care practice;
3) Third party coverage patterns and controls on reimbursable costs;
4) Population (size, age, distribution, and health status);
5) Rate of increase in the supply of health care resources caused by increased fixed costs to be amortized and induced by demand;
6) Utilization and case mix.
(Adapted from Dobson et al., 1978, p. 122).

The PSRO system, however, had no legislated mandate to exert control over any of the these factors, except utilization. Consequently, when viewed in the context of all of the factors that influence hospital expenditures, the impact of PSRO review on utilization would have to have been considerable in order for it to perform effectively as a cost containment strategy. Since the observed impact on utilization was negligible, the evaluators suggested that the PSRO program lacked leverage as a cost containment device. To support this contention, they provided a breakdown of the unit costs associated with PSRO review, national program costs, and incremental costs associated with the 'new' concurrent versus the 'old' retrospective review systems. From these data they estimated the amount by which utilization would have to be reduced in order for costs to be contained. In computing these estimates, two sets of assumptions were used: 1) that a system involving 100% concurrent review was employed and 2) that a 20% targeted review system was employed. Under a targeted review system, MCE and profile analysis data would be used to identify potential problem cases, diagnoses, or areas. Thus, rather than reviewing all cases, only those cases most likely to be problematic would be subject to concurrent review. The decision, by Dobson et al., to use estimates based on
an assumption of 20% of cases reviewed was arbitrary and was intended only to provide a basis for comparison.

Using cost data constructed from hospital, PSRO, Medicare, and Medicaid accounting records, the unit costs associated with PSRO review were estimated to be $16.10 (+- $.77) per eligible beneficiary. The estimated national costs for a fully implemented 100% PSRO review system in 1976 were $268.5 million for all eligible beneficiaries. Assuming only 20% review, the estimated costs for 1976 would have been $184 million. In computing the incremental costs associated with 100% PSRO review, the PSRO system was estimated to cost $187.2 million more than the previous retrospective system. Assuming 20% review, the incremental costs associated with PSRO review were estimated to be $76.4 million.

Although one of the major intents of the DHEW evaluation was to compute a cost-benefit ratio for the 18 active PSRO areas, the lack of a significant reduction in utilization rates (i.e., benefit) prevented the pursuit of this objective. Instead, two alternate approaches were taken. First, a break-even analysis was performed on the aggregate data. Second, cost-benefit ratios were computed for each of the seven individual PSRO’s in which reductions were observed.

In the break-even analysis, it was determined that a 100% review program would pay for itself if it could achieve and maintain reductions in utilization rates of between 1.6% and 2.05%. A 20% review program would pay for itself if reductions on the order of .65% to .83% could be maintained. Recall, however, that the aggregate reduction observed from 100% review was only .09%. Therefore, in aggregate, it was concluded that, in 1976, the system was not breaking even.
The cost-benefit analysis of the individual PSRO's that had effectively reduced utilization revealed that for six of these seven PSRO's the cost-benefit ratio was greater than one. In estimating the impact of these achievements on health care expenditures, however, it was determined that if all of the PSRO's had been able to achieve cost-benefit ratios of 2 the reduction in expenditures would have been less than 1%. Since none of the PSRO's had ratios as high as 2, the effect on expenditures would have been considerably less than 1%. Moreover, in order to achieve ratios of 2 it was determined that PSRO's would have to generate utilization rate reductions on the magnitude of 3.2% to 4.1%. None of the PSRO's included in this evaluation were found to achieve such levels of reduction.

The final question addressed in the DHEW evaluation was that of the impact of PSRO review on the appropriateness of hospital care. At the time the evaluation was performed, however, no empirically derived measures of appropriateness were available. In addition, Dobson et al. indicated that attempts to define measures that could serve as proxies for identifying changes in appropriateness of care were frustrated by the nature of the available data. Consequently, it was stated that no conclusions could be drawn with respect to whether the reductions in utilization observed in some PSRO's were the result of decreases in inappropriate use or, alternatively, the result of a failure to provide necessary care. It was suggested, therefore, that much additional work was needed in terms of defining and measuring the appropriateness of hospital utilization.

II.4.6 Other PSRO-Related Research

Although the DHEW study was the most comprehensive evaluation of the PSRO review system, other studies performed in the late 1970's
tended to confirm its findings. Averill and McMahon (1977), for example, developed a theoretical model of the Continued Stay Certification (CSC) process and used it to test the potential cost-effectiveness of the system. In developing the model, a number of different assumptions were made about the distribution of length of stay, the length of delay between the time of CSC review and implementation of appropriate action, the proportion of patients who experience unnecessary days of hospitalization, the number of days of unnecessary hospitalization, and the distribution of unnecessary hospital stays as a function of length of stay. Although the authors conceded that the results of their analysis were vulnerable to criticism on the grounds that they were based on theoretically derived, rather than empirically derived, assumptions, they maintained that a lack of empirical data necessitated the acceptance of theoretical assumptions. To compensate for this, whenever an assumption was required, a sensitivity analysis was performed to test over the range of possible and reasonable values.

In general, the basic steps in the modelling procedure were invoked as follows. First, the number of patients with each possible length of stay was calculated. Second, the number of patients that would be reviewed was calculated. Third, the number of patients not discharged before the results of the review could be implemented was calculated. Fourth, the proportion of patients from step three who would experience unnecessary hospital days was calculated. Fifth, the proportion of patients from step four who would experience a sufficiently large number of unnecessary days of hospital care, relative to their length of stay, to be unnecessarily hospitalized at the time of the review was calculated. Sixth, the ratio of the number of patients in step five to the total number of patients reviewed was calculated. This ratio,
which the authors referred to as the Yield from Review (YR), provided the fraction of patients who could have their length of stay directly affected by CSC review. Then, by estimating the cost of a CSC review and using the YR computed in step six, the financial benefits that would have to accrue in order for the process to be cost-effective could be determined.

Using length of stay data for all patients discharged from the Yale-New Haven hospital between April 1971 and September 1973, the authors determined that, for each case of inappropriate use discovered, savings of between $500 and $10,000 would have to be realized in order for the CSC process to break even. On average, the amount of savings required per case was $2000. The authors concluded, therefore, that as it was unlikely that any reductions in length of stay attributed to CSC would produce such savings, the cost of the review process would most likely outweigh its benefits. They then suggested that the CSC process be de-emphasized and that more attention be paid to the development of less costly review mechanisms such as Profile Analysis and Medical Care Evaluation Studies.

Hirsch, Rimm, and Welsch (1977) expressed concerns that the PSRO strategy for continued stay review did not take into consideration whether extension requests could have been avoided through the provision of better care or better planning (eg. shorter preoperative stays). In addition, they were concerned that the use of diagnoses to establish length of stay norms might not be the most effective and efficient way of relating the reason for hospitalization to the process of care. Consequently, they proposed, as an alternative to diagnoses derived norms, the use of norms based on the Reason for Admission (RFA).
The RFA categories used in the study were developed by Hirsch et al. with the help of medical records librarians from four community hospitals. A total of 2000 consecutive live adult discharges were categorized according to the reason for the admission. As well, the length of stay associated with each discharge was recorded. The reason for admission was to reflect specifically why the physician admitted the patient to the hospital. Originally, 100 categories were defined but these were subsequently reduced to a final set of 25 RFA categories. Example RFA's include Minor Surgery (with subgroups defined), Major Elective Surgery (with subgroups), Major Emergency Surgery (with subgroups), Acute Nonsurgical Illness (with subgroups), Management of Chronic Illness, Psychiatric, Supportive and Terminal, and Workups. The length of stay norms assigned to each RFA category were computed from the length of stay data collected from the 2000 records used to define the categories. For most RFA categories, the modal length of stay was used as the norm. For the rest, the norm was taken as the mean or, in some cases, it was based on special considerations (eg. empirically defined norms found in the medical literature). Unfortunately, no attempt was made to ascertain the reliability or validity of the RFA categories. Furthermore, as noted by the authors, the RFA designation was arbitrary and based on a small and geographically contained sample.

The primary purpose of RFA review was to identify three types of patient charts: 1) charts with a complication or other patient factor (P) indicating that length of stay could not be reduced without adversely affecting the patient's health; 2) charts indicating overutilization (U) of hospital resources; and 3) charts indicating a delay in discharge due to less than optimal quality (Q) of care. The purpose of the study was to compare
the efficiency of the RFA strategy with that of the diagnosis-based strategy in identifying patients whose length of stay could have been reduced without adversely affecting the quality of care.

To accomplish this objective, 500 hospital charts were categorized by RFA and then screened according to the RFA length of stay norms. Of these, 106 charts were found to have lengths of stay that exceeded the established norms. From the remaining 394 charts, 106 that did not exceed the norms were selected to serve as controls. The two groups of charts were then screened according to diagnosis-based norms using both the 50th and 75th percentiles as target dates for review. The charts were also subjected to individual review by the hospital's medical director and the senior author of the research paper. The purpose of these reviews was to categorize the charts according to the P, U, and Q, designations. Each designation decision represented a consensus on the part of the two reviewers.

The results of the study indicated that the use of RFA and diagnosis derived norms were equally effective in identifying hospital episodes that could have been shortened without adversely affecting quality. However, the use of diagnostic norms selected out 76% more charts for individual review than did the use of RFA norms. Therefore, although recognizing the limitations of the study methods, the authors suggested that the use of diagnosis determined norms may not be the most efficient method of continued stay review and that the RFA procedure might provide a viable alternative.

In another study concerned about the relevance of diagnosis as the basis for establishing norms for continued stay review, Burford and Averill (1979) examined the relationship between diagnostic information available
at admission and discharge. These authors were interested in determining to what extent incorrect assignments, made on the basis of information available at admission, would affect the cost-effectiveness of the continued stay certification (CSC) process. In particular, it was hypothesized that length of stay norms applied at time of admission would result in one of the following situations:

1) Review conducted at the correct time,
2) Review conducted earlier than it should have been,
3) Review conducted later than it should have been.

In addition, it was predicted that there would be errors of inclusion and omission. Errors of inclusion would occur in situations where patients who were reviewed too early, would not have been subject to review at all, had the appropriate norms been employed. Errors of omission would occur among the patients who were discharged before being reviewed because the review date established on admission was later than it should have been.

The database used for the study was the PSRO Hospital Discharge Data Set. The study data set consisted of 52,210 patient records collected during the ten month period ending June 1976 for 68 hospitals in one PSRO area. A comparison of admitting and discharge diagnoses revealed that, of the 52,210 records, 41.3% had a change in primary diagnosis between admission and discharge. Furthermore, although the majority of the 52,210 patients were found to have been reviewed on time (58.5%), 27.6% were reviewed earlier than they should have been, on the basis of their discharge diagnosis, and 13.8% were reviewed late. With respect to errors of inclusion and omission, 6.6% of the 52,210 patients were unnecessarily reviewed, and 2.6% were inappropriately excluded from review. Thus, in total, 41.5% of the patient
records examined were found to have problems in the timing and performance of CSC review. The authors concluded, therefore, that since early and unnecessary reviews increase the number of unproductive reviews, and missed and late reviews decrease the potential for length of stay reductions, all such problems would decrease the effectiveness of CSC review. Because they also found that some diagnostic categories were more variable than others, they suggested that, for those categories, alternative methods for initiating the CSC review process should be explored. They indicated that one possibility would be to set a routine time period that would provide time for diagnostic work-up. Once the work-up was completed, the diagnosis could be confirmed and an appropriate length of stay norm assigned. However, as this would result in an extra review for many patients it would likely increase the costs of the CSC process. Burford and Averill (1979), noting that the cost-effectiveness of concurrent review was already questionable, concluded that it might be more productive to limit CSC review to those diagnostic categories identified as problematic through the less costly mechanism of Profile Analysis.

Boaz (1979), on the other hand, found little evidence to support the use of the CSC process at all. In a study designed to explore the implications for hospital utilization of providing medical care in the most appropriate (i.e., least cost) setting, this investigator concluded that pre-admission certification was more likely than CSC to have an overall impact on hospital utilization. This conclusion was based on a sample of 8,936 nonmaternity patients discharged from 60 short stay general hospitals during the year ended on June 30, 1970. Using this data, a regression model incorporating medical, social, and economic variables was developed in an attempt to
measure the differential effects of these variables on pre-PSRO length of hospital stay. Since one of the major objectives of PSRO review was to eliminate hospital utilization attributable to the systematic influences of nonmedical factors, Boaz argued that the impact of PSRO review on providing care in the most appropriate setting could be assessed through the determination of the magnitudes of the effects of social and economic factors on hospital episodes. She also attempted to predict the impact on hospital utilization of a pre-admission certification process that would screen out all patients who could be treated effectively in an ambulatory setting. Finally, she attempted to determine to what extent hospital stays would be shortened if CSC restricted all hospital stays to the "critical minimum required by medical standards" (p. 317).

Three regression models, each using length of stay as the dependent variable and a variety of clinical, economic, and social variables as predictors, were defined: 1) for all patients, 2) for urgent admissions, and 3) for nonurgent admissions. The overall predictive power of each model was small but significant ($R^2 = .22, .21, .22$, respectively; $p < .0001$ in all cases). In all three models, the medical variables (eg. type of admission, number of diagnoses, discharge status, and surgical status) and age were more strongly associated with length of stay than were the social (eg. living arrangements, occupation) and economic (eg. payment source) variables. Where effects of social and economic variables were observed, they tended to be concentrated among patients whose diagnoses would have been unlikely to qualify them for hospitalization under a preadmission screening process. Because of these findings, the author concluded that CSC would be unlikely to have an appreciable effect on length of stay. She suggested instead that future PSRO
activity should focus on the pre-admission screening process as this would be more likely to lead to substantive reductions in hospital utilization. In addition, she suggested that future evaluations of continued stay review should place less emphasis on measuring reductions in length of stay, and more on determining to what extent the review process effects changes in hospital case mix. In particular, she indicated that a more relevant measure of the procedure's success would be the extent to which hospital case mix shifts towards those diagnoses that, by any criteria, can only be treated in an acute care hospital setting.

None of the above studies employed a randomized controlled trial design and, therefore, they cannot be said to have provided a definitive answer to the question of PSRO effectiveness or cost-effectiveness. However, the accumulation of evidence from these as well as the pre-PSRO program studies was sufficient to motivate the federal government to begin questioning the efficiency with which the program was being delivered (Sanazaro, 1977). Observing that some PSRO's had been more effective than others in reducing utilization (Dobson et.al.,1978), the government attempted to impose greater efficiency in the operation of PSRO review by reducing funding levels and by requiring that PSRO's begin to develop measures by which to demonstrate that the program was having a positive and documentable impact (Adler & Milstein, 1983; Brown & Levy, 1980). Thus, by 1979 the average reimbursement rate for PSRO review was reduced from $12 to $8.70 per inpatient admission and considerable pressure was exerted on the PSRO's to move away from the labour intensive and costly mechanism of 100% concurrent review and to begin to find ways of
focussing the review process on the cases and/or hospitals that were most likely to be problematic (Goran, 1979; Mullin, 1983).

One of the most frequently suggested methods of improving PSRO efficiency and cost-effectiveness was through the expanded use of Profile Analysis (Averill & McMahon, 1977; Burford, & Averill, 1979; Dobson et al., 1978; Goran, 1979; Mullin, 1983). Profile Analysis is a form of retrospective review in which aggregated hospital, practitioner, and patient care data are subject to pattern analysis (Goran, 1979). As mentioned earlier, from the inception of the PSRO review system, Profile Analysis had been included as one of the system's three major components. It failed to gain much attention in the early years, however, because of a lack of appropriate data. As noted by Goran (1979), it took "a considerable amount of time and effort for PSRO's to establish data collection systems, procure data processing services and collect information for a long enough period to develop profiles of practice, all of which are prerequisites to profile analysis" (p.17). In addition, a standard method of defining and measuring hospital case mix was required in order to assure that comparisons based on the development of institutional and patient profiles would be appropriate and valid (Dobson et al., 1978; Doremus, 1980; Goran, 1979).

As noted in an earlier section, hospital case mix had long been acknowledged as a major factor contributing to patterns of hospital utilization. Until the late 1970's, however, a case mix measurement system applicable to any hospital, region, or PSRO population had not been developed (Doremus, 1980; McMahon, 1984). Consequently, in order to comply with PSRO regulations many hospitals and PSRO's developed length of stay norms using diagnostic categories as defined by the
International Classification of Diseases (ICD) or PAS. Both of these classification systems were criticized, however, for their failure to categorize patients into homogeneous groups defined in terms of hospital resource consumption (Fetter, Shin, Freeman, Averill, & Thompson, 1980; Jenkins & Cole, 1984; McMahon, 1984). Thus, there was concern that attempts to compare utilization patterns across hospitals and PSRO's would be invalid; i.e., that any observed variability in utilization would be due to differences in case mix rather than patient management. Similar concerns were expressed by researchers who were attempting to study and model hospital utilization as well as by those who were attempting to study relationships between hospital utilization and costs (Lave & Leinhardt, 1976b; Luke, 1979; McMahon, 1984).

In response to such concerns, a group of researchers from Yale University (Fetter et al., 1980) began working on a patient classification system called Diagnosis Related Groups (DRG's). On the basis of research initiated in the early 1970's, this group of investigators sought to develop a management tool that would relate "the demographic, diagnostic and therapeutic characteristics of patients to the output they are provided so that cases are differentiated by only those variables related to the condition of the patient (eg. age, primary diagnosis) and treatment process (eg. operations) that affect his utilization of the hospital's facilities" (Fetter et al., 1980, p. 2). In other words, the aim was to produce a system of categories within which the observed variability in utilization would be primarily attributable to differential management of similar patient types rather than to differences in case mix (McMahon, 1984; Young, Swinkola, & Hutton, 1980). If successful, this system could then be employed as a tool for utilization...
review, utilization research and, ultimately, for hospital planning, budgeting, and cost control (Fetter et al., 1980; McMahon, 1984).

II.4.7 The Development of DRG's

As described by Fetter et al. (1980), the primary purpose of the DRG approach to patient classification was to "identify in the hospital acute-care setting a set of case types, each representing a class of patients with similar processes of care and a predictable package of services (or product) from an institution" (p.3). In addition, the DRG developers wanted to ensure that the case type definitions would be meaningful to both medical and non-medical users and easy to implement in a variety of settings. To achieve these objectives, the DRG system was developed in accordance with the following criteria:

1. Class definitions must be medically meaningful.
2. Class types should be defined on the basis of variables that are relevant to output utilization and readily available on hospital abstracts.
3. The number of classes must be manageable as well as mutually exclusive and exhaustive.
4. The classes should contain patients with similar expected measures of output utilization.

(Adapted from Fetter et al., 1980, p. 5).

II.4.7.1 The Construction of DRG's

The construction of the initial set of DRG's involved two steps. In the first step, a committee of physicians was asked to identify diagnostic groups that were 1) consistent in terms of their anatomic, physiopathologic classification, or in their clinical management, 2) mutually exclusive and exhaustive, and 3) representative of a sufficient number of patients (Fetter et al., 1980). From this effort, a total of 83 Major Diagnostic Categories (MDC's) were defined. The diagnostic coding system chosen as the standard for the
construction of the DRG system was the ICDA8\(^1\). However, since diagnostic data in the database were coded using different classification schemes (eg. PAS, ICDA8, and HICDA2\(^2\)), this step also involved translating the various coding systems into their ICDA8 equivalent. In the second step, an iterative statistical analysis was undertaken to determine whether any of the 83 MDC's could be further subdivided into groups that would add to the homogeneity of the classification scheme.

The database used to construct the initial DRG's consisted of 500,000 hospital records from 118 hospitals in New Jersey, 150,000 records from one Connecticut hospital, and 52,000 records of federally funded patients from 50 hospitals in one PSRO region (South Carolina). Once the 83 MDC's had been defined, an interactive computer system called AUTOGRP was employed to conduct the subgroup partitioning. This system was designed to facilitate rapid analysis of large volumes of complex data and, through the use of the CLASSIFY facility, provided an algorithm for determining partitions on the basis of interactively specified clinical and statistical criteria (Fetter et al., 1980). The utilization output measure (i.e., the dependent variable) used in the analysis was length of stay. This measure was selected because of its historical acceptance as a proxy measure of resource consumption, its easy availability, and its reliability. The final set of input variables (i.e., independent variables) used in defining the class types

\(^1\) International Classification of Diseases, Adapted for Use in the United States, Eighth Revision.
\(^2\) Hospital Adaptation of ICDA, Second Edition.
\(^3\) Other variables such as sex, tertiary diagnosis and tertiary surgical procedure were examined but were not found to be important in explaining output utilization (i.e., length of stay).
included principal diagnosis, presence of secondary diagnosis, principal surgical procedure, secondary surgical procedure, and age.

Thus, within each MDC, the CLASSIFY algorithm was invoked to identify groups of patients that differed with respect to length of stay. The various independent variables were examined and weighed simultaneously in each analysis. The decision to accept, reject, or modify a suggested subgroup was based on both the statistical evidence and the clinical committee's medical knowledge. The iterative process continued until the introduction of additional variables did not significantly reduce the variance in length of stay or the subgroup was too small to warrant further partitioning (i.e., when there were fewer than 100 observations in the subgroup). The result of the analysis was the definition of 383 DRG's which the system's developers claimed were clinically interpretable, reflected similar patterns of utilization, made use of routinely collected data, and were mutually exclusive and exhaustive (Fetter et al., 1980). An example of the partitioning of one MDC is shown in Figure 1.

II.4.7.2 Implications for Utilization Review

With the development of DRG's, Fetter et al. (1980) maintained that they had provided an "effective mechanism [for use] in profile analysis by providing a structure of consistent patient-class definitions within which institutional performance [could] be compared based on similar types of patients" (p. 27). They thus believed that with the accumulation of sufficient data, DRG's would permit the development of a more timely and cost-effective system of retrospective review than the costly PSRO concurrent review process that was in use at the time. Hospitals or PSRO's would be able to monitor patterns of utilization over time and, through the
use of standard statistical techniques, identify the occurrence of deviant cases. Once identified, these "outliers" could then be investigated to determine whether they were the result of errors, unusual events in the utilization patterns of one or more patients, an overall change in treatment procedures, or misutilization. Moreover, by using DRG's to adjust for case mix differences among hospitals and/or PSRO's, inter-organizational comparisons of utilization patterns would be facilitated (Fetter, et al., 1980).

II.4.7.3 Initial Reactions to DRG's

Initial reactions to the DRG classification system were mixed. It was generally agreed that the DRG system was an improvement over PAS and ICD taxonomies (Berki et al., 1984; Doremus, 1980; Iglehart, 1982b; Simborg,
1981). However, it was also recognized that the system suffered from a number of significant limitations. Of particular concern was the fact that the DRG's were developed using data from only New Jersey, Connecticut, and South Carolina (Doremus, 1980; Young et al., 1980). As noted by Doremus (1980), variations in length of stay by geographic region had long been observed and documented. Thus, the geographically limited data set used by the DRG developers led to questions concerning the external validity of the partitioning process. To investigate this issue, Young et al. (1980) applied the DRG grouping process on data from Western Pennsylvania. Starting with approximately 690,000 patient records, a preliminary analysis was undertaken to identify 10 MDC's for inclusion in the study. The patient records in each of these MDC's were then submitted to the AUTOGRP CLASSIFY algorithm for subgroup partitioning. According to Young et al., none of the terminal groups produced as a result of this analysis were the same as the DRG's identified by Fetter et al. (1980). Specifically, they found that the independent variables (eg. surgery, diagnosis, etc.) that produced the optimal grouping with the Pennsylvania data often differed from those used in the DRG system. As well, the number of terminal groups generated within each of the 10 MDC's often differed from the number of DRG's generated from the same MDC's. On the basis of these findings, the authors concluded that the DRG system was valid only for the patient population from which it was developed, and thus any attempt to generalize the results from one patient population to another would be inappropriate.

Another important issue raised with respect to DRG development concerned the accuracy of the data used in their construction. Doremus (1980) noted that in a study conducted by the National Academy of Sciences
to determine the reliability of Medicare patient abstract data, the principal diagnosis was found to be reliable only 57.2 percent of the time, in one sample, and 65.2 percent of the time in a second sample. Furthermore, the presence of a secondary diagnosis was noted correctly in only 74.5 percent of the cases studied. Similarly, the notation of principal surgical procedure was correct in only 78.9 percent of the cases. It was concluded, therefore, that the data discrepancies were of sufficient magnitude to be considered a major problem in the construction of DRG's.

Finally, questions regarding the clinical homogeneity of the DRG defined patient groups were raised. Berki et al. (1984) investigated this issue by examining the extent to which variations in length of stay (LOS) within DRG's were associated with clinical variables not included in the DRG partitioning process. It was argued that "to the extent that DRG's as currently defined capture the LOS-influencing clinical factors of case complexity and severity, observed variations of LOS within DRG's should be attributable to nonclinical factors" (Berki et al., 1984, p. 126); i.e., no systematic association with clinical factors should be observed. To test this proposition, Berki et al. examined the length of stay distributions of 7 DRG's. Using a regression model that included both clinical (eg. total number of diagnoses, intensity of nursing services, timing of radiologic and laboratory services) and nonclinical (age, sex, distance to hospital, method of payment, day admitted, type of admission, length of preoperative stay) variables as predictors of length of stay, within each DRG, these investigators found that in all cases the clinical variables significantly reduced the observed variability in length of stay. For some of the DRG's studied, significant relationships between length of stay and non-clinical
variables were also identified. However, none of these variables reached the significance of the clinical variables. They concluded, therefore, that DRG's had not yet been developed to the point where they represented sufficiently precise categories for purposes of either utilization research, review or hospital reimbursement.

II.4.7.4 The Reformulation of DRG's

When Fetter et al., (1980) described the method used to construct the DRG's, they indicated that the system would require continuous development and revision as coding changes occurred and more data became available. Accordingly, when an updated version of ICD codes (ICD-9-CM) was implemented in 1979, the DRG system had to be reevaluated and revised. Although the reformulation of the DRG's involved the use of the same AUTOGRP algorithm as had been used previously, an attempt was made to address concerns about the external validity and homogeneity of the system (Fetter, 1985). Thus, in 1981, a new database was developed using 1979 data from a sample of 332 PAS member hospitals. This new database was selected to be as representative of the national population as possible, within the PAS membership criteria, and consisted of a total of 1.4 million patient records. From this database, "a sample of 394,814 records was selected according to a stratified random sampling procedure, with MDC's as the strata" (Hornbrook, 1985, p. 301). It should be noted, however, that these were not the same MDC's as had been used in the previous DRG development process. Originally, 83 MDC's had been defined. In the revised DRG system, the number of MDC's was reduced to twenty-three. The partitioning of all principal diagnoses into these 23 MDC's was based on either the organ system affected or the specialist who would typically
provide care (Fetter, 1985). In most cases, each MDC was then disaggregated according to the presence or absence of a procedure performed in the operating room. Subsequent partitions were based on either the principal diagnosis or procedure category. From these categories, additional partitions were identified according to age, the existence of comorbidities and/or complications, and, occasionally, discharge status. The result of this process was the definition of 467 DRG's in contrast to the 383 DRG's identified in the old system (Fetter, 1985; Hornbrook, 1985; Williams, Kominski, Dowd, & Soper, 1984).

II.4.7.5 Suggested Applications for DRG's

As indicated above, DRG's were developed initially in response to a need for a standard patient classification system that could be used to improve the cost-effectiveness of utilization review procedures employed at local, regional, or national levels. Accordingly, in the mid-1970's "when DRG's were first designed, they were intended as a management system for hospitals, as a tool for utilization review, and as a research tool" (Smits, Fetter, & McMahon, 1984, p. 72). However, by 1980, when the article describing the method used to construct the initial DRG's was published, they were also being promoted as a potential mechanism for case mix accounting in hospital budgeting and prospective reimbursement (Fetter, et al., 1980). In fact, a major experiment with DRG-based reimbursement was already underway in the state of New Jersey (Iglehart, 1982b; May, 1985). Moreover, with the reformulation of DRG's in 1981, it was generally acknowledged that it would only be a matter of time before a DRG-based prospective payment system would be introduced by the federal government (Iglehart, 1982a; Simborg, 1981). This prediction was borne out
in 1983 when legislation calling for the implementation of DRG-based prospective payment for Medicare hospital services was passed (Iglehart, 1983). Although the details of the Medicare prospective payment system (PPS) will be discussed later, it is mentioned here for two reasons. First, the federal PPS was implemented so quickly following the reformulation of the DRG's that almost all of the research concerning the adequacy of the DRG system was done after the enactment of the PPS legislation. Second, although most of the concerns regarding the adequacy of the new DRG's have important implications for utilization review and research, critical appraisals of the DRG system have focussed most of their attention on the implications of using DRG's as a mechanism for hospital reimbursement. Therefore, the specific concerns that have been raised, with respect to both applications (utilization review and hospital reimbursement) will be discussed in the PPS section. Before turning to that discussion, however, another utilization review procedure that began to emerge in the early 1980's, as an adjunct or alternative to DRG-based review, will be described.

II.4.8 An Alternative Approach: Assessing Appropriate Utilization

Since the mid-1960's, the major factor motivating the federal government to implement utilization control strategies had been the assumption that a considerable amount of hospital use was inappropriate and unnecessary. Until the mid-1970's, however, few attempts had been made to objectively measure and quantify inappropriate use or to establish a definitive standard of what would constitute inappropriate use (Goran, 1979). Rather, the government's decision to mandate utilization review was based primarily on indirect evidence provided through comparative geographic and prepaid practice studies; wide variations in regional
utilization rates, after standardization for age and sex, were assumed to reflect inappropriate use. Therefore, because there was no established standard for defining appropriate use, and because it was assumed that most instances of misutilization were occurring in the latter portion of inpatient episodes, the primary intent of the early federal utilization control strategies was simply to reduce the observed regional variability in hospital lengths of stay (Goran, 1979; Goldberg & Holloway, 1975; Zimmer, 1974). Thus, concurrent review, profile analysis, and DRG's were all developed with length of stay as the primary variable of interest.

By the mid-1970's, however, when Fetter and his associates were beginning to work on the development of DRG's, another group of investigators was beginning to question the merits of utilization review procedures that relied solely upon statistically derived length of stay norms as the basis for reducing inappropriate utilization (Gertman & Restuccia, 1981; Goldberg & Holloway, 1975; Restuccia & Holloway, 1976; Rothberg, & Gertman, 1981; Sieverts, 1978; Zimmer, 1974). As will be discussed in more detail below, studies done by members of this group had begun to show that a high percentage of misutilization occurred not only among the longer stay cases but among the shorter stay cases as well (Restuccia & Holloway, 1976; Zimmer, 1974). It was suggested, therefore, that target date concurrent review procedures might fail to identify the cases of misutilization that "occurr[ed] during hospital stays that violate[d] no length of stay norms" (Sieverts, 1978, p. 602). If so, the effectiveness of concurrent review would be seriously limited. Moreover, from a quality of care perspective, it was feared that continued emphasis on length of stay target dates and norms would foster "an attitude that if a patient has reached the average length of stay that
patient 'should' or 'must' be ready to leave the hospital, irrespective of individual medical realities" (Goldberg & Holloway, 1975, p. 478).

As noted by Gertman & Restuccia (1981), although total utilization measures, such as length of stay, can be useful for describing general trends in hospital use, they provide no way to determine whether observed increases, or decreases, in utilization are the result of changes in inappropriate use or changes in appropriate use. Moreover, they provide no information as to the reasons for inappropriate use, the nature and amount of inappropriate use, or the extent to which inappropriate use can be controlled (Gertman & Restuccia, 1981; Goldberg & Holloway, 1975; Restuccia & Holloway, 1976; Rothberg & Gertman, 1981; Sieverts, 1978; Zimmer, 1974).

Given such potential shortcomings in the normative length of stay approach to utilization review, it was suggested that PSRO's would be more effective if they began to focus their attention on actual patient needs for hospitalization (Dobson et al., 1978; Goldberg & Holloway, 1975; Sieverts, 1978). The promulgation of this approach was impeded, however, by the lack of a reliable and valid method for evaluating appropriate and inappropriate hospital use. As stated by Rothberg and Gertman (1981), "[m]easuring unnecessary hospital use in an 'objective' fashion is a difficult, expensive and sensitive task. Merely defining what constitutes 'unnecessary' (the precondition for measurement) has been a problem of major proportions" (p. 48). Thus, to date, most of the research on the appropriateness of hospital utilization has focussed on measurement issues. At the same time, however, attempts have been made to begin quantifying inappropriate use and to document the reasons for it. Although
methodological problems render most of these studies exploratory at best, consistency in the reported results suggests that they may have important implications for both utilization management and quality assurance program efforts.

II.4.8.1 Measuring Appropriate Use

Since the mid-1970's, a number of different methods have been developed to measure the appropriateness of hospital utilization. Zimmer (1974), for example, had physician reviewers examine patient records and talk to staff and patients in order to make a decision about the patient's need for hospitalization. The subjective nature of this approach, however, rendered it particularly vulnerable to criticisms concerning its reliability (Gertman & Restuccia, 1981; Goldberg & Holloway, 1975; Restuccia & Holloway, 1976; Rosser, 1976). Thus, Goldberg & Holloway (1975) suggested that, until more explicit criteria for appropriate utilization review could be developed, the broad service criteria for hospital "level of care" as generated by the Medicare program or the McKay-Dee Hospital Center should be used. In a subsequent study, Restuccia and Holloway (1976) adopted this approach and used the Medicare "level of care" criteria (See Appendix A) to identify instances of misutilization in one hospital setting. Although this approach was considered an improvement over the more subjective approach used by Zimmer (1974), it was generally considered only the first step in the development of explicit utilization criteria (Goldberg & Holloway, 1975; Gertman & Restuccia, 1981).

A completely different, albeit interim, approach was suggested by Rothberg & Gertman (1981). These authors maintained that, since defining and measuring inappropriate utilization is a time consuming and difficult
process, attitudinal surveys designed to assess perceptions of inappropriate use could be used pending the development of more objective measures. This method received little attention, however, because in the same year that it was proposed an article was published describing what has been, to date, the most significant advancement in the measurement of inappropriate utilization; namely, the Appropriateness Evaluation Protocol (Gertman & Restuccia, 1981).

The Appropriateness Evaluation Protocol (AEP) was designed in response to the need for a "valid, reliable and simple technique for determining the appropriateness of hospital use, in terms of the necessity for a patient being in a hospital bed on a given day" (Gertman & Restuccia, 1981, p. 867-868). Because the developers also wanted the AEP to be "applicable to as many patients as possible" (p. 857), it was decided that the instrument should be "diagnosis-independent" and applicable to all patients admitted into "adult medicine, surgery and gynecology services" (p. 857). In addition, it was decided that the AEP should consist of "a set of explicit criteria which, if any single one was met, would indicate that a day of care was appropriate" (p. 857). Finally, in order to improve the efficiency and reliability of the AEP review process, an upper limit of 30 criteria items was established.

After several years of development and testing, 27 "day-of-care" criteria were adopted. These criteria items were organized into three categories: "medical services, nursing/life support services, and patient condition factors" (Gertman & Restuccia, 1981, p. 857). The first two categories contained items that assessed whether or not, on a given day of hospitalization, the patient was receiving a service or services that would
typically only be provided in an acute care setting. The third category
assessed factors that would generally be considered to indicate that the
patient's condition was sufficiently unstable to require hospitalization,
whether or not they were receiving medical or nursing services on the day
of review. A list of all the criteria developed by Gertman and Restuccia is
provided in Table I.

One of the major objectives of Gertman & Restuccia (1981), in
developing the AEP, was to obtain higher levels of inter-rater reliability
than had been obtained with previous measurement techniques. Of
particular concern was the level of reviewer agreement that would be
obtained among the cases deemed inappropriate or uncertain (specific
agreement). As indicated by Gertman & Restuccia, relatively high levels of
overall agreement between reviewers (73-89%) had been fairly common in
previous studies, even in the studies that used subjective criteria. In these
same studies, however, the rate of agreement obtained among the cases
deemed inappropriate, by at least one reviewer, was much lower (0-53%).
Given that one of the primary functions of hospital utilization review, or
PSRO review, is the identification of cases that require further study,
Gertman and Restuccia argued that, although high levels of overall
agreement are important, they are less important, for the purposes of
utilization review, than are high levels of specific agreement. That is,
unless the review instrument can identify reliably the cases that are
inappropriate, the review process will be both ineffective and inefficient.

In order to evaluate the reliability (overall and specific) and the
validity of the AEP, Gertman & Restuccia (1981) conducted two trials. In the
first trial, two nurse reviewers and one physician reviewer applied the AEP
TABLE I. **Appropriateness Evaluation Protocol: “Day-of-Care” Criteria**

### A. Medical Services
1. Procedure in operating room that day
2. Scheduled for procedure in operating room the next day, requiring preoperative consultation or evaluation
3. Cardiac catheterization that day
4. Angiography that day
5. Biopsy of internal organ that day
6. Thoracentesis or paracentesis that day
7. Invasive central nervous system diagnostic procedure (e.g., lumbar puncture, cisternal tap, ventricular tap, pneumoencephalography) that day
8. Any test requiring strict dietary control, for the duration of the diet
9. New or experimental treatment requiring frequent dose adjustments under direct medical supervision
10. Close medical monitoring by a doctor at least three times daily (observations must be documented in record)
11. Postoperative day for any procedure covered in number 1 or 3-7 above

### B. Nursing/Life Support Services
1. Respiratory care - intermittent or continuous respirator use and/or inhalation therapy (with chest PT, IPPB) at least thrice daily
2. Parenteral therapy - intermittent or continuous IV fluid with any supplementation (electrolytes, protein, medications)
3. Continuous vital sign monitoring - at least every 30 minutes, for at least 4 hours
4. Intramuscular and/or subcutaneous injections at least twice daily
5. Intake and output measurement
6. Major surgical wound and drainage care (chest tubes, T-tubes, hemovacs, Penrose drains)
7. Close medical monitoring by nurse at least three times daily, under doctor's orders

### C. Patient Condition Factors
**Within 24 hours before day of review**
1. Inability to void or move bowels (past 24 hours) not attributable to neurologic disorder

**Within 48 hours before day of review**
2. Transfusion due to blood loss
3. Ventricular fibrillation or ECG evidence of acute ischemia, as stated in progress note or in ECG report
4. Fever at least 101 rectally (at least 100 orally), if patient was admitted for reasons other than fever
5. Coma - unresponsiveness for at least one hour
6. Acute confusional state, not due to alcohol withdrawal
7. Acute hematologic disorders, significant neutropenia, anemia, thrombocytopenia, leukocytosis, erythrocytosis, or thrombocytosis yielding signs or symptoms
8. Progressive acute neurologic difficulties

**Within 14 days before day of review**
9. Occurrence of a documented, new acute myocardial infarction or cerebrovascular accident (stroke)

(Source: Gertman & Restuccia, 1981)
retrospectively to 100 randomly selected patient records. The sample consisted of approximately 50% of the patients that were in the medical and surgical units of one hospital on the specific day chosen for review. In the second trial, the same reviewers applied the AEP to a new sample of 100 patients. This time, however, various combinations of concurrent and retrospective review were used by the three reviewers.

To assess the inter-rater reliability of the AEP, overall agreement and specific agreement rates were computed. In addition, overall agreement rates between nurse reviewers, and between each nurse reviewer and the physician reviewer were evaluated. The investigators also compared the rates obtained when the AEP was applied retrospectively as opposed to concurrently and from trial one to trial two. Although no data regarding these comparisons was presented, the authors claimed that the rates were very similar. Hence, in the reported analyses, the data from both trials were pooled.

To evaluate the validity of the AEP, several methods were used. First, the agreement between the physician-nurse pairs was used as an estimate of predictive validity. Second, critical appraisal of the AEP criteria by four PSRO committees was used to assess face validity. Finally, the use of 'overrides' was used as an indication of content validity. The use of 'override' options were included in the AEP in order to account for any defects in the established criteria. Essentially, overrides "allowed a reviewer to 1) judge a day appropriate when no objective criteria were met, usually because the list of criteria was not sufficiently comprehensive; or 2) judge a day as inappropriate even when an objective criteria was met, because the reviewer felt the service or condition was not sufficient in the particular
case to justify use of an acute hospital bed" (Gertman & Restuccia, 1981, p. 865).

On the basis of the 200 patients and patient records reviewed in the study, overall agreement rates, each evaluated using a Kappa statistic, between pairs of reviewers (RNA/RNb, RNA/MD, RNb/MD) ranged from a low of 92% to a high of 94% (p<0.0001 for each comparison). Thus, approximately 93% of the time, paired reviewers agreed on whether or not a given day in a patient's hospital stay was appropriate or inappropriate. The agreement rate among all three reviewers was 88.1% (Kappa statistic; p<0.001). Specific agreement rates (i.e., the percentage of paired agreements on cases identified as inappropriate by at least one reviewer) between pairs were lower, ranging from 73% to 79%, but notably higher than the rates obtained in previous studies. The specific agreement rate among all three reviewers was 63.8%.

The "override" options were used by the RNA/RNb pair 21 times (7 by RNA and 14 by RNb) to make a day appropriate and 19 times (7 and 12) to make a day inappropriate. For the MD/RNa and MD/RNb pairs, each had 24 overrides making a day appropriate and 17 overrides making a day inappropriate. An examination of the override cases indicated that, for each pair of reviewers, the use of the override options "improved the level of agreement between reviewers" (Gertman & Restuccia, 1981, p. 865). Moreover, when both reviewers invoked an override option, "in no case did [the] two reviewers make opposing override decisions" (p. 865).

Another aspect of the Gertman & Restuccia study included an assessment of the appropriateness of admission to the hospital. Unlike the criteria developed for the day-of-care assessment, however, no explicit
criteria were used in the admission assessment. Instead, the reviewers were asked to make decisions about the appropriateness of admission on the basis of their professional judgement. This approach was adopted for two reasons. First, in an appraisal of existing admission review procedures, the authors had been unable to identify a set of criteria that were reliable and valid. Second, by having the reviewers use subjective professional judgement in determining the appropriateness of admission, a comparison of agreement rates obtained using this approach could be made to the agreement rates obtained using explicit criteria.

In examining the appropriateness of admission decisions, Gertman & Restuccia found that the overall agreement rates between reviewer pairs ranged from 89.3% to 91.8% (Kappa statistics; p<0.01 in each case) and among all three reviewers was 87.1% (Kappa statistic; p<.001). Thus, in terms of overall agreement on the appropriateness or inappropriateness of admission, the rates were very similar to those obtained in the criteria-based days-of-care assessment. With respect to specific agreement rates, however, the picture was quite different. Among the admissions deemed inappropriate by at least one reviewer, the level of agreement ranged from 30.5% to 40.0%; rates much lower than the rates computed for the days-of-care review. Although the authors acknowledged that the admission and days-of-care reviews were not directly comparable, they nonetheless believed that the observed differences in specific agreement rates were most likely due to the differential use of subjective and objective criteria.

The final measurement issue addressed by Gertman & Restuccia was to what extent higher levels of inter-rater reliability could, with appropriate adjustments to the AEP, be obtained. To investigate this issue, they
examined all of the reasons for disagreements among reviewers. This analysis was limited, however, by the small number of disagreements that had actually occurred and by the variability in the problems cited. Although no firm conclusions could thus be drawn, three sets of factors were found to have contributed most frequently to the occurrence of inter-rater disagreements. First, over half of the disagreements were associated with patients "for whom there was also disagreement over the appropriateness of the admission (p. 867). Since most of these patients were involved in non-routine treatment regimens, the authors concluded that the formulation of explicit criteria for such cases would be difficult and, therefore, they would be "more conducive to being handled on an exception basis rather than by the application of a general set of criteria such as the AEP" (p. 867). Second, a number of the disagreements centered around the mental status of the patients. And, once again, the authors felt that only in a very few cases would more stringent criteria reduce the frequency of disagreements. Finally, it was observed that a number of disagreements related to one of the nursing service criteria (i.e., intake and output measures). On occasions when this was the only criterion met in a review, it could be problematic. Specifically, it was noted that physicians tended to overuse this service - by requesting it when it was not necessary or by failing to discontinue it at the appropriate time. Hence, the authors suggested that this criterion needed further study. Overall, Gertman & Restuccia estimated that interrater reliability on the AEP could possibly be increased "from the 92-94 per cent level to the 95-96 per cent level" (p. 867). Increases beyond that, however, were considered unlikely.
Through subsequent development and testing, Gertman & Restuccia expanded the AEP to include explicit criteria for admission review (Siu et al., 1986; Walter, 1983). A total of 18 admission criteria, organized into severity of illness and intensity of service categories, were added. In a study done by Siu, et al. (1986), overall inter-rater agreement rates for these criteria were reported to be between 84% and 92% (Kappa statistics; $p<0.001$ in all cases). However, no data on specific agreement rates were provided. The complete list of AEP admission criteria is provided in Table II.

In addition to the adult version of the AEP, a modified AEP has recently been developed for application to pediatric inpatients (Kemper, 1987a). Although modified, the structure and content of the pediatric AEP is very similar to the adult AEP. Hence, no further description will be provided herein. A list of the pediatric "day-of-care" criteria, however, are provided in Appendix B. To my knowledge, admission criteria for pediatric patients have not yet been developed.

As with DRG's, reactions to the AEP have been mixed. On the positive side, it has been praised for being quick and easy to apply (Walter, 1983). In addition, in a survey of over 100 hospitals that were using the AEP, it was found that "both UR officials and physicians prefer[red] it to other forms of review" (Walter, 1983, p. 177). However, although the AEP is clearly one of the most reliable and valid instrument that has yet been developed to assess appropriateness of hospital use, there are potential problems associated with its employment as a tool for utilization management. One particular concern is that, in the effort to contain costs, insurers might implement the AEP criteria alone, without the override options (Walter, 1983). Since most of the overrides in the Gertman &
TABLE II. Appropriateness Evaluation Protocol: Admission Criteria

A. Severity of Illness

1. Sudden onset of unconsciousness or disorientation (coma or unresponsiveness)
2. Pulse rate of less than 50 or greater than 140
3. Systolic blood pressure of less than 90 or greater than 200; diastolic of less than 60 or greater than 120
4. Acute loss of sight or hearing
5. Persistent fever equal to or greater than 100 (orally) or greater than 101 (rectally) for more than five days.
6. Acute loss of ability to move a body part
7. Active bleeding
8. Severe electrolyte or blood gas abnormality
9. Acute or progressive sensory, motor, circulatory, or respiratory embarrassment sufficient to incapacitate the patient (One of the seven intensity-of-service criteria listed below must also be met)
10. ECG evidence of acute ischemia; there must be suspicion of a new MI
11. Wound dehiscence or evisceration

B. Intensity of Service

1. Use of IV medications and/or fluid replacement (does not include tube feedings)
2. Surgery or procedure scheduled within 24 hours requiring general or regional anesthesia or equipment and facilities available only in a hospital
3. Vital-sign monitoring every two hours or more often (may include telemetry or bedside cardiac monitoring)
4. Use of chemotherapeutic agents that require continuous observation for a life-threatening toxic reaction
5. Treatment in an ICU
6. Use of intramuscular antibiotics at least every eight hours
7. Intermittent or continuous respirator use at least every eight hours

(Source: Walter, 1983)

Restuccia (1981) study resulted in a designation of the reviewed day as "appropriate", failure to include the override options could lead to inappropriate reimbursement denials (Walter, 1983) A second concern is that strict adherence to appropriateness criteria might raise the proportion of acute hospital days to a point where most are "filled with high anxiety and low satisfaction for the patient, the physician, and other providers of care"
(Dans et al., 1985, p. 1135). Even the developers themselves have recommended caution in the application and interpretation of AEP criteria (Gertman & Restuccia, 1981). Specifically, they have indicated that the AEP does not adjust for situations "where absolute efficiency is not feasible" (p. 869) nor does it take into account whether or not the "causes of inappropriate use are within the control of the hospital and the patient's physician" (p. 869). Accordingly, the identification of inappropriate use, by itself, could be of limited value, unless it is accompanied by an understanding of the reasons for it. In fact, if examined in isolation from its causes, it could have expensive detrimental effects. As Gertman & Restuccia have stated, "optimizing one objective of the medical care system, such as appropriateness of hospital use, may result in suboptimization of other objectives, such as appropriateness of use of all levels of care. Careful consideration of such trade-offs must be made by health care decision makers" (p. 869). In the following section, therefore, the methods that have been used to identify the reasons for inappropriate use are described.

II.4.8.2 Assessing the Reasons for Inappropriate Use

Since most of the studies on inappropriate use have been concerned primarily with identifying and measuring its occurrence, few attempts have been made as yet to develop a systematic method for assessing why inappropriate use occurs. To date, the most comprehensive attempt to study and classify the reasons for inappropriate use has been undertaken by Restuccia & Holloway (1976). In the only previous attempt, Zimmer (1974) simply asked physician reviewers to indicate whether an observed inappropriate day was the result of: 1) an unnecessary admission, 2) an unnecessary delay in discharge, or 3) an unnecessary delay in performing
surgery or other in-hospital procedure. In addition, he had the reviewers indicate what level of care would have been appropriate, assuming that all levels were available; i.e., on the day of the review, would the patient have been more appropriately placed in a "long-term hospital, nursing home, organized home care with multiple services, at home with one or two home care services, or at home with no home care services" (p. 454).

In contrast to Zimmer's (1974) approach, Restuccia & Holloway (1976) interviewed 9 utilization review nurse coordinators, 11 discharge planners, 3 physicians, and 3 administrators, from five different hospitals, in order to develop an *a priori* list of factors that were known, by these individuals, to have caused inappropriate use at their facility. The final list (called "barriers to appropriate utilization") consisted of 63 factors organized into four categories "according to 'areas of responsibility'; i.e., to whose action or inaction a particular barrier could be attributed" (p. 562). The four categories were thus defined as "Physician Responsibility", "Hospital Responsibility", Patient or Family Responsibility" and "Environmental Responsibility". Barriers attributable to the Physician were those relating to the physician's professional knowledge or availability to render services; Hospital barriers related to operational systems and to the organization's employees; Patient barriers related to the patient's (or family's) participation in the treatment process; and Environmental barriers related to factors beyond the immediate control of the hospital, physician, or patient (e.g. insurance coverage, availability of appropriate alternative services, etc). A list of the specific barriers included in each category is provided in Appendix C.

Gertman & Restuccia (1981) employed a shortened version of this list in order to assess the reasons for inappropriate day-of-care use among the
patients included in their study. In addition, for patients designated as inappropriate admissions, Gertman & Restuccia had their reviewers indicate whether the admission was considered unnecessary because it involved: 1) the provision of services that could have been done on an outpatient basis or in a lower level facility, 2) no clear need for medical care of any kind, or 3) some other specified reason.

II.4.8.3 Empirical Findings on the Amount and Nature of Inappropriate Use

For the most part, the researchers that have attempted to quantify and describe patterns of inappropriate use have focused their attention on one or more of the following four questions:

1) to what extent does inappropriate utilization occur?;
2) at what point in an inpatient episode is inappropriate utilization most likely to occur?;
3) what is the relationship between inappropriate use and length of hospital stay?; and
4) what are the major factors that contribute to inappropriate use?

The findings relevant to each of these questions will be discussed in turn.

1. Amount of Inappropriate Use. Almost all of the "appropriateness of utilization" studies have attempted, in the process of methodological development, to quantify the amount of inappropriate use observed in their samples. Some of these studies have assessed only the occurrence of inappropriate day-of-care use. Others have examined both the inappropriateness of day-of-care use and hospital admissions. A summary of the major findings of these studies, as well as some important characteristics relating to their research methods, are provided in Table III.
### TABLE III. Studies of Inappropriate Hospital Utilization

<table>
<thead>
<tr>
<th>Author</th>
<th>Study (Year)</th>
<th>Study Population and Description</th>
<th>Day-of-Care Reviews</th>
<th>Admission Reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total No.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>% Inappropriate</td>
<td>Total No.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>% Inappropriate</td>
<td>% Inappropriate</td>
</tr>
<tr>
<td>Zimmer (1974)</td>
<td>1968-70</td>
<td>One hospital; All Clinical Services; 3,369 patients; Subjective assessment; 2,695 patients reviewed twice on same day; 674 reviewed once.</td>
<td>6,064</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9.1</td>
<td>-</td>
</tr>
<tr>
<td>Restuccia &amp;:</td>
<td>1973-74</td>
<td>One hospital; Med/Surgical Services; 218 patients; Assessment by Medicare Criteria; All patients reviewed on each day of stay.</td>
<td>1,902</td>
<td>-</td>
</tr>
<tr>
<td>Holloway (1976)</td>
<td></td>
<td></td>
<td>10.6</td>
<td>-</td>
</tr>
<tr>
<td>Gertman &amp; Restuccia (1981)</td>
<td>not given</td>
<td>One hospital; Med/Surgical Services; 192 patients; AEP assessment; most patients reviewed three times on same day.</td>
<td>496</td>
<td>496</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>25.0</td>
<td>10</td>
</tr>
<tr>
<td>Siu et. al.,</td>
<td>1974-82</td>
<td>100 hospitals; 6 cities; Med/Surg Services; 1,132 patients; AEP assessment; Each patient reviewed once.</td>
<td>1,132</td>
<td>1,132</td>
</tr>
<tr>
<td>(1986)</td>
<td></td>
<td></td>
<td>35.0</td>
<td>23*</td>
</tr>
</tbody>
</table>

* An additional 17% would have been inappropriate had ambulatory services been available in the hospitals studied.
A total of 40% of the admission reviews were thus potentially inappropriate.
As can be seen from this table, the amount of inappropriate "days-of-care" use identified in these studies ranges from 9% to 35%, with the higher rates obtained in studies that employed explicit criteria (i.e., the AEP). The proportion of inappropriate admissions ranges from 10% to 23%. Again, the higher rates are associated with the use of AEP criteria. Gertman & Restuccia (1981) have suggested that higher rates of inappropriate use should be obtained with the AEP because it was designed to provide an "absolute type of efficiency standard" (p.869). Subjective techniques, on the other hand, permit reviewers to modify their judgements according to whether or not absolute efficiency is feasible. However, the higher rates may also be a function of temporal differences, differences in study methods, or differences in the study populations. For these reasons, the studies shown in Table III are not directly comparable. Moreover, since all of these studies were done within the context of one hospital or within narrowly defined geographic regions, little can be said about expected rates of inappropriate hospital use across the country. Nor has sufficient data been accumulated to indicate trends in misutilization over time or how such trends have been effected, if at all, by the institution of policy interventions such as PSRO's. Nonetheless, it would appear reasonable to suggest that, despite such interventions, inappropriate use of hospital facilities and services has persisted. Certainly, the problem has remained of sufficient magnitude to warrant, as Rothberg & Gertman (1981) suggested, continued vigilance in the monitoring and management of hospital utilization.

2. **Timing of Inappropriate Use.** Only one of the studies mentioned in Table III attempted to determine at what point in an inpatient episode, an inappropriate day was most likely to occur. To address this question,
Restuccia and Holloway (1976) classified the 201 inappropriate days identified in their study according to the "stage of stay"; i.e., according to the day on which the inappropriate day(s) occurred relative to the overall sequence of days defining each inpatient episode. Thus, inappropriate days classified as "initial-stage" were those that occurred prior to the recording of the first appropriate day or when no appropriate day was recorded prior to discharge; "midstage" referred to an inappropriate day or days that occurred between two series of appropriate days; and "end-stage" referred to inappropriate days that immediately preceded discharge but followed a consecutive series of appropriate days. Using this classification scheme, Restuccia & Holloway found that, of all 201 inappropriate days identified, 10% were classified as initial-stage, 25% were midstage, and 65% were end-stage. Thus, for the hospital in which this study was done, the common assumption that most misutilization occurs in the latter stages of inpatient episodes appeared to be confirmed. Whether or not this was a general or hospital-specific phenomenon was not investigated.

3. The Relationship Between Misutilization and Length of Stay. To explore the nature of the relationship between inappropriate use and length of stay Zimmer (1974) examined the percentage of inappropriate days that occurred among patient stays of 1-7 days, 8-14 days, 15-21 days, 22-30 days, 31-60 days, and over 60 days. He found that this percentage increased from 4.9% for patients hospitalized from 1-7 days, to 24.6% for patients with stays exceeding 60 days. In order to examine this trend further, misutilizers were reclassified according to whether their hospital stay was 21 days or less or
greater than 21 days\(^1\). With this dichotomization of inpatient lengths of stay, it was found that 52.6% of the misutilizers had stays of 21 days or less while only 47.4% had stays over 21 days. Thus, while the overall percentage of misutilization was found to increase with length of stay, a higher proportion of the identified misutilizers had lengths of stay that, at the time the study was conducted, would have failed to trigger an extended stay review. Although these findings were important in the sense that they prompted consideration of misutilization in shorter stay cases, a different pattern would likely have been obtained had a different target date been used. In a more detailed study, however, Restuccia & Holloway (1976) reported findings that confirmed Zimmer's contention that the problem of misutilization among shorter stay cases was worthy of note. These investigators used the diagnosis-related length of stay norms developed by PAS to identify the 50th and 75th percentiles for the patients included in their study. These percentiles were chosen because they reflected the most commonly used target dates in PSRO review. Accordingly, of the 201 inappropriate days identified in their study, 21% occurred before the 50th percentile and 37% occurred before the 75th percentile. Moreover, when the Medicare maximum time limit for certification was used (i.e., 12 days), 55% of the inappropriate days were found to occur before this checkpoint.

Although the instrument used by Restuccia & Holloway (1976) to assess inappropriate use (i.e., Medicare level of care criteria) has been criticized as being overly general and unreliable (Goldberg & Holloway, 1975; Gertman & Restuccia, 1981), the implications of such findings were considered to

\(^1\) The use of 21 days to separate short from long stays was based on the fact that, for the patients reviewed, this was the target used by the state fiscal intermediaries to instigate recertification review.
provide reasonable justification for continued work in the area (Sieverts, 1978). In addition, although Restuccia and Holloway's finding that the greatest percentage of misutilization occurred in the "end-stage" of patient episodes, the application of commonly used length of stay criteria would have missed not only a significant proportion of the "initial" and "midstage" misutilization, but a sizeable proportion of the "end-stage" misutilization as well. This tended to support the notion that, for their study population at least, utilization reviews based on length of stay norms would be of limited effectiveness.

4. Reasons for Inappropriate Use. In order to develop a better understanding of the factors that were contributing to the occurrence of inappropriate days, a few of the investigators mentioned in Table III attempted to document the reasons for each of the inappropriate days identified in their studies. The findings of these studies are summarized in Table IV. Although the lack of a standard instrument for assessing the reasons for inappropriate use makes comparisons from one study to another difficult, certain trends do emerge. For example, across all the studies, the most commonly cited reasons for inappropriate days were discharge-related. Moreover, discharge problems were found to be primarily attributable to a lack of appropriate lower level care alternatives or a failure to access lower level alternatives at the appropriate time. The second major factor found to contribute to inappropriate use was physician-related; i.e., there was a tendency on the part of some physicians to adopt overly conservative treatment management practices. The percentage of inappropriate days attributed to problems in hospital operational procedures, on the other hand, was relatively small.
<table>
<thead>
<tr>
<th>Author (Year)</th>
<th>No. of Inappropriate Days Identified</th>
<th>No. of Reasons Identified</th>
<th>Breakdown of Inappropriate Days Reason</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Delay in hospital procedure</td>
<td>42</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Delay in discharge</td>
<td>425</td>
<td>82.2</td>
</tr>
<tr>
<td>Restuccia &amp; Holloway (1976)</td>
<td>201</td>
<td>229**</td>
<td>Physician Responsibility (related to discharge)</td>
<td>96</td>
<td>41.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hospital Responsibility (related to discharge)</td>
<td>17</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Patient Responsibility (related to discharge)</td>
<td>22</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Environmental Factors (related to discharge)</td>
<td>95</td>
<td>41.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>**</td>
<td>Major Reasons **</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Physician too conservative</td>
<td>16.0%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-delays transfer</td>
<td>7.5%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Environment **</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-terminal patients not transferred</td>
<td>6.5%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-nursing facility not available</td>
<td>23.6%</td>
<td></td>
</tr>
<tr>
<td>Gertman &amp; Restuccia (1981)</td>
<td>124</td>
<td>124</td>
<td>Inappropriate level of care</td>
<td>53</td>
<td>43.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Physician too Conservative</td>
<td>30</td>
<td>24.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Delay in hospital procedure</td>
<td>14</td>
<td>11.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Delay in Transfer</td>
<td>9</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reason not indicated</td>
<td>18</td>
<td>14.5</td>
</tr>
</tbody>
</table>

* For descriptive information concerning these studies refer to Table III.
** 28 of the Inappropriate days had 2 reasons listed.
As can be seen in Table V, the reasons cited to account for inappropriate hospital admissions reflect a tendency to admit patients that could have been treated on an outpatient basis and a failure to use lower levels of care appropriately. It should be noted, however, that at least in the case of the patients included in the Siu et al. (1986) study, the failure to treat patients on an outpatient basis was due to the fact that, in all of the hospitals studied, ambulatory services were not available. Thus, although 17% of the cases studied did not require acute level care, the lack of the appropriate outpatient services rendered their admissions necessary.

II.4.8.4 Implications of the Empirical Studies

All of the findings mentioned above demonstrate the importance of developing an understanding of the factors that impede appropriate utilization. While the results of these studies must be considered preliminary, since they are based largely on case studies and, therefore, lack generalizability, they have important implications for the development of policies aimed at reducing inappropriate utilization. As noted by Restuccia & Holloway (1976), although most of the identified inappropriate use was likely to occur during the latter stage of an inpatient episode, it was primarily attributable to the limited availability of appropriate alternatives. Thus, while normative utilization review procedures may facilitate the identification of such misutilizers, the ability of physicians and hospitals to reduce this form of misutilization would be limited because, in the short run at least, it lies beyond the scope of their control. For similar reasons, it would be difficult to effect reductions in inappropriate use resulting from the second most common cause of misutilization, namely, conservative medical practice. With the use of instruments such as the AEP and the
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>No. of Inappropriate Admissions</th>
<th>Breakdown of Inappropriate Admissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gertman &amp; Restuccia</td>
<td>1981</td>
<td>49</td>
<td>Services could have been provided on outpatient basis.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>33 67.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower level of care needed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6 12.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No need for any care</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 8.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Missing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 10.0</td>
</tr>
<tr>
<td>Siu et. al.</td>
<td>1986</td>
<td>452</td>
<td>Lower level of care needed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>174 38.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Services could have been provided on outpatient basis.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>192 42.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>86 19.5</td>
</tr>
</tbody>
</table>

* For descriptive information concerning these studies refer to Table III.
"Barriers to Inappropriate Utilization", however, hospitals could begin to identify inefficiencies or lapses in operational procedures that contribute to misutilization and, within that realm of responsibility at least, begin to take corrective action. Whether the use of such techniques, either alone or as a supplement to length of stay reviews, will facilitate increased reductions in inappropriate hospital utilization remains to be seen. To my knowledge, no studies have as yet attempted to evaluate the effectiveness of the AEP approach in reducing misutilization. Nor have any studies been done which compare the general effectiveness, or cost-effectiveness, of the AEP approach versus the normative approach.

Since the AEP was developed, however, a major reformulation of federal health care policy has resulted in the abandonment of the PSRO concurrent review strategy (Dans et al., 1985). Under federal legislation enacted in 1983, hospitals are no longer retrospectively reimbursed for the actual costs incurred in treating Medicare patients. Instead, they are paid on the basis of a prospectively determined rate calculated for each of the 467 defined DRG's (Iglehart, 1982a). Accordingly, the nature and objectives of utilization review and management have undergone important changes. The remainder of this chapter will, therefore, be devoted to the new prospective payment system, with particular emphasis given to its implications for hospital utilization and utilization management.

II.5 Current Trends in Utilization Management: The Era of Prospective Payment

Although the development of DRG's and the AEP were important refinements in utilization review technology, by the late 1970's policy analysts and health service researchers had begun to question the merits of
relying upon regulatory mechanisms, such as the PSRO's, to contain publicly financed hospital costs (Brook et al., 1978; Dobson et al., 1978; Goran, 1979). Although decreases in hospital length of stay had been observed throughout the 1970's (Pokras & Kubishke, 1985), little evidence had been provided to suggest that the PSRO program had been effective in containing costs (Dans et al., 1985; Ginzberg, 1985; Goran, 1979; Lantos, 1984). Moreover, Goran (1979) argued that since PSRO's had no mandate to affect costs directly, "they should be held accountable for the elimination of unneeded services and for the shifting of services from inappropriately intense levels of care to lesser ones, but not for the cost and reimbursement consequences of such actions" (p.3). Instead, it was suggested that, in order to contain costs, the PSRO system would have to be supplemented with major statutory changes in the method used to reimburse hospitals (Brook et al., 1978; Goran, 1979); that new financial incentives favouring efficiency would have to be developed to replace the inflationary incentives provided under the traditional method of cost-based retrospective reimbursement (Iglehart, 1982a). In fact, during the latter half of the 1970's, a number of states had already begun to experiment with alternative reimbursement methods (Iglehart, 1982a). Although the details of the various methods differed from one state to another, they all shared one characteristic in common; i.e., they were all based on a prospective system of payment. Thus, rather than retrospectively reimbursing hospitals on the basis of average per diem costs incurred, these states had begun to develop systems based either on cost center budgets, global budgets, or prospectively determined rates-per-case (Iglehart, 1982a). Regardless of the method used, the basic intent was the same; namely, to impose front end limits on
hospital reimbursement (Iglehart, 1982a) and, thereby, reverse previous inflationary trends.

II.5.1 The Transition to the Federal Prospective Payment System

In 1982, the annual increase in total hospital expenditures was determined to be almost three times the national inflation rate (Iglehart, 1983) and estimates of future expenditures predicted that Medicare costs would double by 1987, if no changes in trends could be effected (Iglehart, 1982a). Concerns about the continuing viability of the Medicare trust fund (Dans et al. 1985; Iglehart, 1983; Pettingill & Westell, 1984) and the variable performance of the PSRO's (Brook et al., 1978; Dans et al., 1985; Dobson et al., 1978) led to the 1982 enactment of the Tax Equity and Fiscal Responsibility Act (TEFRA) and the Peer Review Improvement Act (PRIA; Dans et al., 1985; Lipp, 1984). With the passage of TEFRA, "two broad measures were approved: an interim program of tighter Medicare controls on hospitals..., and a directive to the Department of Health and Human Services (DHHS) to develop a prospective payment system (PPS) that, when implemented would effectively impose a spending ceiling for Medicare on hospitals" (Iglehart, 1982a, p. 1288). PRIA called for the creation of "one professional review organization (PRO) in each state to replace the PSRO program" (Lipp, 1984, p. 283). Although the federal administration wanted to abolish the PSRO's immediately, the congress elected to delay taking action until the new PPS could be implemented and the new system of external review could be established (Dans et al., 1985).

In contrast to the traditional method of per diem cost reimbursement, the tighter Medicare controls mandated by TEFRA were to be applied on the basis of cost per discharge (Iglehart, 1982a; Pointer & Ross, 1984). Thus, "to
implement the new regulations and to prepare for the eventual use of a prospective payment system, Congress directed the DHHS to develop a case mix index for every hospital" (Iglehart, 1982a, p. 1289). In order to comply with this directive, however, a method of defining and quantifying the multiproduct nature of the hospital was needed. In the State of New Jersey, an experimental prospective rate-per-case system suggested that DRG's could be used for this purpose (May, 1985). Although a number of concerns were expressed about the appropriateness of using DRG's for reimbursement purposes (Doremus, 1980; Iglehart, 1982b; Simborg, 1981), they were quickly adopted by the federal government as the "only applicable and acceptable approach to product-oriented payment" (Johnson & Appel, 1984, p. 128). Thus, in October 1983, the federal government passed the Deficit Reduction Act which provided the mandate for the implementation of a DRG-based prospective payment system (PPS) for the hospital inpatient care of Medicare beneficiaries (Dans et al., 1985; Iglehart, 1983).

One of the major assumptions prompting this initiative was that "differences in efficiency rather than differences in characteristics of the patient population or in the quality of care [were] responsible for interinstitutional differences in the cost of providing care" (Stern & Epstein, 1985, p. 622). It was believed, therefore, that through the implementation of a system of uniform payments, quality of care could be maintained while hospitals would be provided with financial incentives to improve efficiency and to "provide care at the lowest possible cost for each patient admitted" (Stern & Epstein, 1985, p. 622). As will be discussed in more detail below, however, the incentives provided under the new PPS were complex and, consequently, it was anticipated that they might result in a number of
unintended and deleterious effects (Grimaldi & Micheletti, 1984; Wilson, 1984). Therefore, in recognition of the potential for problematic responses, all hospitals participating in Medicare were required to enter into a contract with an approved external review organization, i.e., PRO, by October 1984, or risk the termination of their Medicare payments (Grimaldi & Micheletti, 1984).

II.5.2 Major Objectives of the PPS

Since its inception in 1983, the primary goal of the federal PPS has been to "moderate the growth of government spending for health care and to force hospitals to assume greater financial risk for the care delivered within their institutions" (Iglehart, 1982b, p. 1655). In order to achieve this goal, hospital reimbursement is provided on the basis of a defined "clinical product" (Gertman & Lowenstein, 1984). The basic objectives of the PPS are thus: "1) to relate the prices paid for caring for illnesses to the actual costs on an illness-specific basis; 2) to reward efficient hospitals and penalize inefficient ones; and 3) to encourage the involvement of key medical decision-makers (the physicians) in decisions regarding costs and resource allocation" (May, 1985, p. 17c).

The first objective is to be met by establishing a national price per case that reflects the actual average cost of treating patients within each of the 467 DRG's (Gertman & Lowenstein, 1984; May, 1985). Hospitals are paid the established DRG price, for each patient that falls into a given DRG category, "regardless of what it costs the hospital to care for that patient" (May, 1985, p. 17c). Thus, the hospital's total reimbursement will be determined by their case mix rather than their resource consumption (May, 1985; Gertman & Lowenstein, 1984; Pointer & Ross, 1984).
The second objective, namely, rewarding efficiency and penalizing inefficiency, is expected to be achieved through the payment process (May, 1985). As noted above, hospitals are paid a set rate per case, regardless of the actual resources consumed or the patient's length of stay. Thus, in contrast to the previous reimbursement method, which rewarded hospitals for keeping patients as long as possible and for providing as many services as possible, the new system is designed to reward hospitals financially for reducing the length of hospital stay, the intensity of service, and the costs of supplies and labour (May, 1985). In addition, because hospitals are permitted to keep any difference between their actual costs and the DRG rate, they are rewarded for increasing the efficiency of their operational procedures, and motivated to ensure the completeness and accuracy of their records (Iglehart, 1983; May, 1985). Conversely, since no additional funds will be provided, in the event of deficits, hospitals that fail to provide necessary care within the established rates will be penalized (Iglehart, 1983; May, 1985).

Finally, the third objective, i.e., increased physician involvement, is expected to be achieved as physicians become cognizant of the effects their prescribed inpatient treatment patterns have on the financial solvency of the hospital (Iglehart, 1982a; May, 1985). The necessity for increased physician involvement is based on three factors. First, since physicians have the primary responsibility for admitting, treating and discharging patients, their involvement is crucial if the new reimbursement strategy is to succeed (May, 1985). Second, physicians are directly involved, through their record keeping behaviour, in determining how much their hospital is paid (May, 1985). Thus, the accuracy and timeliness of their medical record keeping will be an important issue (May, 1985). Third, long term reductions
in hospital utilization require a change in medical practice patterns. In particular, as noted by Thurow (1985), treatment practices that conform with the traditional philosophy of "Do no harm" need to be replaced by practices that are guided by the precept "Employ a treatment only when you are sure that it will make a noticeable improvement" (p. 613). Increased involvement in the management of hospital resources may facilitate such a change.

II.5.3 Major Characteristics of the Medicare PPS

Although the PPS was initiated in 1983, it will not be fully implemented until the end of 1988. Thus, while the ultimate aim was to develop a single national price per DRG (McMahon, 1984), an effort was made, during the early years of the PPS, to protect hospitals from undue financial penalties by calculating the basic DRG rate using a blend of hospital, regional, and national costs (Davis, Anderson, & Steinberg, 1984; May, 1985). When implementation is completed, however, the calculation of the price per case will be based on six variables: 1) the basic DRG rate (calculated using 100% national historical costs for persons within each DRG); 2) the urban or rural location of the hospital; 3) regional labor costs; 4) indirect teaching costs (direct educational expenses are reimbursed at cost); 5) the proportion of the hospital's patients that are eligible for Medicaid or Supplemental Security Income; and 6) the patient's outlier status - i.e., whether the patient has an unusually long length of stay or high cost of care (Davis et al., 1984; Jencks & Dobson, 1987).

At the present time, the PPS applies only to the payment of hospital inpatient care provided to Medicare beneficiaries. "Hospital outpatient services are excluded, as are certain rehabilitative, psychiatric, and long-term
care services [provided] in distinct parts of community hospitals" (Davis et al., 1984, p. 141). In addition, the PPS does not apply to the payment of capital costs nor to the "payment of wages and other direct costs of interns and residents" (Jencks & Dobson, 1987, p. 680).

II.5.4 Implications of PPS for the Cost, Use and Quality of Hospital Care

As indicated above, the financial incentives provided under the PPS are complex. Thus, while it is generally agreed that the PPS provides strong incentives to reduce resource consumption within an inpatient episode (Davis et al., 1984; Gertman & Lowenstein, 1984; Jencks & Dobson, 1987; Pointer & Ross, 1984; Stern & Epstein, 1985; Wyszewianski, Thomas, & Friedman, 1987), critics of the PPS have identified a number of potential problems that could reduce the effectiveness of the system, as it is presently structured (Davis et al., 1984; Horn, Sharkey, Chambers & Horn, 1985; Jencks, Dobson, Willis, & Feinstein, 1984; Omenn, & Conrad, 1984; Simborg, 1981; Speigel & Kavaler, 1985; Stern & Epstein, 1985; Wennberg, McPherson, & Caper, 1984). These concerns generally fall into two categories: 1) potential problems associated with the incentive structure of the PPS itself; and 2) potential problems associated with the use of DRG's as the method of defining the "clinical product". In the following section, the specific problems identified on the basis of theoretical analyses of the PPS incentive structure are delineated. Later, in sections II.5.5 and II.5.8, organizational responses and empirical data relevant to the hypothesized problems are discussed.

II.5.4.1 Potential Problems Associated with PPS Incentives

A number of critics have suggested that, in order to increase revenues, hospitals may respond to the PPS by reducing quality of care
(Iglehart, 1983; May, 1985; Stern & Epstein, 1985; Wyszewianski, et al., 1987). Clearly, the financial incentives provided by the PPS favour hospitals that can effect per case reductions in resource use. As noted by Wyszewianski et al., (1987), if such reductions are achieved through elimination of unnecessary services then quality of care will not be affected adversely. However, if hospitals respond to the PPS by encouraging underutilization of needed services then quality of care will be diminished (Iglehart, 1983; Stern & Epstein, 1985; Wyszewianski et al., 1987).

While the potential certainly exists for the PPS to achieve its objectives at the expense of quality, Wyszewianski et al. (1987) argue that, in fact, it could have the opposite effect. Specifically, they suggest that, in response to the potentially negative effects of the PPS, physicians may become more attentive to quality of care issues. In addition, as utilization goes down, competition among hospitals will likely increase. If so, hospitals will be motivated to provide high quality care because this will become an increasingly important factor in the competition for patients. Nonetheless, until it can be demonstrated empirically that the PPS has not reduced the quality of care provided to Medicare beneficiaries, quality will remain a serious issue (May, 1985).

In addition to quality of care concerns, potential problems have been identified in connection with the cost containment objectives of the PPS. For instance, since payment is by the case, hospitals may attempt to increase their total revenues by treating a larger number of cases (Gertman & Lowenstein, 1984; May, 1985; Stern & Epstein, 1985; Wennberg et al., 1984). In particular, they may attempt to increase the admissions most likely to be profitable (i.e., the less severely ill), and those that, in general, require fewer
services and shorter lengths of stay) while reducing those that are not (Gertman & Lowenstein, 1984; Stern & Epstein, 1985). In addition, through a similar selection process, hospitals may begin to restrict, on the basis of profitability, the mix of services they provide (Eastaugh & Eastaugh, 1986; May, 1985; Omenn & Conrad, 1984). Smits et al., (1984) suggest that since doctors, not hospitals, make admitting decisions, the extent to which hospitals can achieve such selectiveness may be limited. However, Wennberg et al., (1984), in a study of admission variability within DRG's, found evidence to suggest that admission decisions are subject to considerable medical discretion and variations in medical practice preferences. They concluded, therefore, that as capacity increases, in response to PPS-induced reductions in length of stay, hospitals will have no difficulty "recruiting physicians whose clinical interests coincide with a profitable mix of services, or 'product line'" (p. 299). While some analysts of the PPS suggest that specialization by hospitals can have positive benefits, in the sense that by treating larger numbers of the same types of patients greater efficiencies may be achieved (Eastaugh & Eastaugh, 1986; May, 1985; Omenn & Conrad, 1984), others fear that it will result in limiting access to care for many beneficiaries or in making certain types of care unavailable to Medicare patients (Omenn & Conrad, 1984; Stern & Epstein, 1985).

The PPS may also motivate hospitals to encourage physicians to split the inpatient episodes of patients with multiple problems into two or more admissions, or to discharge patients early, even if it means a subsequent readmission will be required (Davis, et al., 1984; Gertman & Lowenstein, 1984; May, 1985; Stern & Epstein, 1985). In this manner they could receive payment for each admission, thereby increasing their total revenues. The
possibility that these practices might be encouraged under the PPS has raised serious concern because, even before implementation of the PPS, readmissions were found to have been a major and costly problem for the Medicare program (Anderson & Steinberg, 1984).

Since the PPS applies only to inpatient care, there may also be a tendency to shift a variety of inpatient procedures to an ambulatory setting (Davis et al., 1984; Eastaugh & Eastaugh, 1986; Gertman & Lowenstein, 1984; Nathanson, 1984a). Although this is, in fact, one of the objectives of the PPS, in the short run it will tend to increase costs because the DRG rates for inpatient services are based on historical costs (Davis et al., 1984; Eastaugh & Eastaugh, 1986). Thus, by performing these procedures on an outpatient basis, the hospital would receive payment twice. While such "double dipping" may create problems initially, it has been argued that they will become less of a concern when the DRG rates have been revised using post-PPS data (Nathanson, 1984a).

Because the assignment of a DRG depends on the sequence of primary and secondary diagnosis, hospitals may also seek to increase reimbursement by manipulating the ordering of these diagnoses; a process referred to as "DRG creep" (Simborg, 1981). Although the PPS legislation ensures that outright manipulation of this sort will be considered fraud, in many cases the resequencing of the diagnoses to favour the more profitable DRG may be medically justifiable; i.e., there may be a difference of medical opinion as to which of several diagnoses should be listed first or second (Nathanson, 1984b; Simborg, 1981; Stern & Epstein, 1985).

Another potential problem that could have a significant effect on the cost, use and quality of care relates to the fact that surgical DRG's receive a
higher rate of payment than medical DRG's (May, 1985; Omenn & Conrad, 1984)). Thus, there may be a tendency to increase the number of surgical procedures performed.

There may also be a tendency to delay the discharge of patients who are nearing the "outlier" trim point (Gertman & Lowenstein, 1984). As noted earlier, hospitals can receive additional payments for patients who are determined to have unusually long lengths of stay or high costs of care. Thus, when it is not possible to discharge a patient within the time frame allotted for his/her DRG, the hospital may attempt to delay discharge until the patient has been in long enough to be eligible for an outlier adjustment.

Finally, because the federal PPS applies to the Medicare program alone, hospitals confronted with reduced Medicare payments could begin shifting costs to private payers (Iglehart, 1982a; Wennberg et al., 1984). In so doing, the incentives to improve efficiency through reductions in length of stay and the use of ancillary services, would be considerably weakened (Davis et al., 1984; May & Wasserman, 1984).

All of these potentially negative responses to the PPS could occur regardless of the accuracy of the case mix measurement system upon which it is based. However, If the classification of cases is inaccurate and results in unfair payments to hospitals, then additional problems could be created. As stated by Jencks & Dobson (1987), "case-mix adjustment for payment must help the prospective payment system to promote efficiency and contain costs while protecting beneficiaries and hospitals" (p. 680). To protect beneficiaries, the payment rates must be "accurate enough so that hospitals cannot identify and avoid patients whose care will cost more than the Medicare payment" (Jencks & Dobson, 1987, p. 680). At the same time, to
protect hospitals, the payment rates must be high enough to ensure that
Medicare patients receive medically appropriate care. Otherwise, "hospitals
might face financial hardship or insolvency, and beneficiaries might face
poor or inaccessible care" (Jencks & Dobson, 1987, p. 680). The DRG system,
which is the cornerstone of the Medicare PPS, has been criticized for failing
to provide the necessary level of accuracy (Conklin, Lieberman, Barnes, &
Louis, 1984; Jencks et al., 1984; Omenn & Conrad, 1984; Stern & Epstein,
1985). In particular, the system has been criticized for failing to account
adequately for differences in severity of illness (Berki, 1984; Brewster, Karlin,
et al., 1985; Conklin et al., 1984; Horn, Bulkley, et al., 1985; Horn & Horn,
1984; Horn, Horn & Sharkey, 1984; Horn, Sharkey, Chambers & Horn, 1985;
Mendenhall, 1984; Young, 1984). Thus, although it is generally agreed that
case-based reimbursement for hospitals is necessary (Johnson & Appel, 1984)
the adequacy of the DRG system has become a topic of considerable debate.
While the following discussion will focus on the limitations of DRG's for
the purposes of reimbursement, most of the concerns identified relate
equally to the use of DRG's for other purposes such as utilization review
and utilization research (Conklin et al., 1984; Williams et al., 1984).

II.5.4.2 Concerns Regarding the Adequacy of DRG's

One of the major requirements of any case mix measurement system
is that it identify groups of patients that are homogeneous with respect to
the outcome variable of interest. Although an effort was made by Fetter and
his associates to improve the clinical homogeneity of the revised DRG's,
critics of the system contend that excessive variability within DRG case mix
groups remains a serious problem. Even the developers themselves have
indicated that "if all DRG's from a large data base are plotted and inspected,
some groupings appear more cohesive than others (Smits et al., 1984, p. 72). Furthermore, many critics of the federal DRG-based PPS have pointed out that, as originally constructed, DRG's were intended to be used for the purposes of utilization review, not reimbursement (Doremus, 1980; Hornbrook, 1985; Johnson & Appel, 1984). Thus, they were designed to identify patients that were homogeneous with respect to length of stay, not costs. Although several investigators (Lave & Leinhardt, 1976b; Luke, 1979) have shown that length of stay and hospital costs are positively correlated, Doremus (1980) has indicated that "it would be naive to imagine that length of stay by itself is an accurate indicator of patient treatment costs" (p. 50). He suggested, therefore, that to improve the accuracy of the prospective reimbursement system, a new set of cost-based DRG's should be developed. He acknowledged, however, that it would take some time for such a reformulation of the DRG system because of problems associated with obtaining the necessary cost data.

Many factors have been identified in an attempt to explain the variability observed within DRG's. However, most of the debate surrounding the use of DRG's, for reimbursement as well as for other applications, has centred on the extent to which the DRG system accounts adequately for differences in illness severity. Several critics of DRG-based reimbursement have suggested that, to improve payment equity, the DRG system should either be replaced or supplemented with severity of illness measures (Conklin et al., 1984; Horn & Horn, 1984; Horn, Horn, & Sharkey, 1984; Nathanson, 1985). Accordingly, a number of investigators have proposed alternative or supplementary classification systems that attempt to identify and group patients along dimensions of illness severity. Included
among these alternative systems are the Severity of Illness Index (Horn, Horn, & Sharkey, 1984), Patient Management Categories (Young, 1984), APACHE II (Knaus, Draper, Wagner & Zimmerman, 1985), Disease Staging (Conklin et al., 1984), and MEDISGRPS (Brewster, Karlin, et al., 1985). Furthermore, in studies done by Horn and her associates, it has been found that the inclusion of a severity of illness adjustment of DRG's reduced the variability observed in both costs and length of stay per case by approximately 40% (Horn & Horn, 1984; Horn, Horn & Sharkey, 1984; Horn, Sharkey, Chambers, & Horn, 1985). Likewise, Conklin et al., (1984) found that by using Disease Staging criteria to adjust DRG's, within-DRG variability in cost per discharge was reduced by 8 to 16%. Similar variance reductions have been attributed to the use of MEDISGRPS-adjusted DRG's (Jencks & Dobson, 1987). As indicated by Smits et al. (1984), however, in order to ensure that the DRG system would be "economical and feasible" only data that were available to all hospitals were used in constructing the DRG's. Thus, it is to be expected that the inclusion of additional clinical variables (eg. severity measures) will reduce the variability within DRG's. However, since most of the severity measures referred to above require the abstraction of data from patients' medical records, a costly and time consuming process, it is uncertain whether the benefits of incorporating such data would outweigh their costs. Moreover, as Jencks, & Dobson (1987) point out, little evidence has been provided to suggest that any of the proposed severity of illness measures offers a major improvement over DRG's, in predicting average cost per case across hospitals, particularly when the PPS adjustments for location, teaching and 'disproportionate share' are taken into account. As noted by McMahon (1984), recognition of the fact
that patients may differ in illness severity does not automatically suggest that DRG's will be invalid. Patients with the most severe illnesses do not necessarily require more resources than those who are less severely ill. Nonetheless, because the presence of systematic variations in severity of illness, within certain hospitals, may have implications for their ability to remain solvent under the PPS, severity of illness remains an important issue. Although Medicare officials are continuing to study the problem (Nathanson, 1985), it is unlikely that severity-adjusted DRG's will be adopted for the PPS, until the relevant data is abstracted on a regular basis (Berki, 1984). Nevertheless, it is recognized that DRG's must be updated on a regular basis, and revised and modified as new data becomes available (Berki, 1984).

Although the proponents of severity adjusted DRG's argue that most of the observed within-DRG variability is due to differences in severity of illness, others have suggested that severity of illness is not the only factor that contributes to heterogeneity within DRG's. As noted by Jencks and Dobson (1987), Smits et al. (1984), and Gertman and Lowenstein (1984), variations in the appropriateness of care provided within a given DRG, and differences in physician practice patterns are also likely to account for some of the observed variability. Neither of these factors, however, represent problems for the DRG system per se. Instead they are sources of variability that the DRG-based PPS is designed to identify and reduce. Other factors identified as potential sources of DRG heterogeneity, that may be problematic for the DRG system, include: 1) coding errors, 2) methodological problems, and 3) misclassification problems.
1) **Coding errors.** One concern raised by critics of the revised DRG system relates to the accuracy of the discharge abstract data that are to be used by hospitals in assigning DRG's (Gertman & Lowenstein, 1984; Johnson & Appel, 1984). To study this problem, Johnson & Appel (1984) compared the Medicare case mix of 26 hospitals, as determined from discharge claims data, to the actual case mix, as determined from each patient's medical record. A total of 75,893 Medicare discharges in 1980 and 77,640 discharges in 1981 were evaluated. In each case, a DRG based on the discharge abstract data was assigned as well as a DRG computed from the medical record. From the 1980 data, agreement between the two DRG's was observed in only 49.4% of the cases. In an examination of the results by hospital, the range of agreement ranged from 31% to 74% with the higher agreement rates being most commonly associated with small suburban hospitals. The lowest agreement rates were typically associated with the larger tertiary care hospitals. By 1981, the overall rate of agreement had increased to 53.0% with a range from 36% to 71%. No changes in the relationships between hospital type and agreement rates were observed.

The major reasons for the observed discrepancies were identified as: incorrect assignment of ICD-9-CM code, discharge diagnosis unknown at the time the claims data was submitted, incorrect procedure codes listed on the discharge abstract, poor documentation and coding of secondary diagnostic and procedure information, and frequent omission of, or errors in, the recording of sex and birth date information (Johnson & Appel, 1984, p. 130).

Gertman & Lowenstein (1984), in a review of studies done by the Institute of Management and Research of the University Hospitals of Cleveland, echoed the concerns raised by Johnson & Appel (1984).
According to Gertman & Lowenstein, the Cleveland studies reported error rates ranging from 20 to 40% which were attributed primarily to "sloppiness and to the use of unqualified record abstractors" (p.80).

Although both Johnson & Appel (1984) and Gertman & Lowenstein (1984) concluded that such coding errors posed a significant problem for all applications of the DRG system which rely upon historical discharge data, they also suggested that with time, experience, and the incentives provided under the new reimbursement system, the error rate would likely decline.

2) Methodological problems. Several concerns have been expressed with respect to the methods used to construct DRG's. Williams et al. (1984), for example, argued that the algorithm used to develop the DRG's (i.e., the AUTOGROUP CLASSIFY algorithm) has three important limitations. First, it places greater emphasis on means than on variances in identifying subgroups. Thus, "may fail to combine subgroups with slightly different means but similar variances, and it will fail to separate subgroups with identical means but different variances" (p. 19). Second, it examines only one variable at a time and, therefore, could miss important interactions. Third, it requires clinical judgement in the selection of the independent variables to use in the partitioning process. Thus, subjectivity is built into the system at each level of subgroup formation. These inherent limitations in the CLASSIFY algorithm prompted Williams et al. to conclude that the DRG's would not represent homogeneous patient groups. However, since no attempt was made to evaluate the effect of these limitations on actual hospital data, they were unable to determine to what extent these methodological limitations would translate into actual problems for the DRG system.
A second methodological concern relates to the fact that DRG's were constructed on the basis of the ICD-9-CM coding system. As indicated by Mullin (1985) and Smits et al. (1984), the ICD-9-CM coding system lacks specificity in the assignment of procedure codes. In particular, it "does not distinguish among different surgical approaches to the same problem. This means that surgical procedures of different lengths and magnitude, and with dissimilar effects on patients and ultimately on resource use, may be coded in an identical fashion" (Smits et al., 1984, p. 75). Moreover, the use of unstandardized "coding conventions" by medical records departments could result in the assignment of ICD codes that are misleading (Mullin, 1985).

The final methodological issue concerns the use of the "principal", as opposed to the "primary", diagnosis in the partitioning of the DRG's. The principal diagnosis is the one which, after study, is determined to have been primarily responsible for triggering the admission. The primary diagnosis, on the other hand, is the one which is found to have accounted for the bulk of the patient's hospital episode (Roberts, Reid, and Irwin, 1985; Smits et al., 1984). Since the primary diagnosis is identified in terms of total resource use, Smits et al. (1984) argued that its use in the formation of DRG's would likely improve the classification of some patients. However, because they lacked a North American data base that included a code for primary diagnosis, they were unable to determine how much of an improvement it would make. Roberts et al. (1985), therefore, investigated this issue in Australia, where the primary, rather than the principal, diagnosis is the convention. Using medical records data collected over a three month period in 1984, Roberts et al. found that, of 1064 records studied, 6.4% had a change in the sequence of diagnosis, when the data were recoded using the
definition for the principal rather than the primary diagnosis. However, only 4.04% of the 1064 recoded records had a change in DRG as a result of the resequencing. These findings support the contention that there will be changes in DRG classification for some patients when a different definition of the principal/primary diagnosis is used. However, the magnitude of the changes observed in the Roberts et al. study were small, and no evidence was provided to indicate whether the changes did, in fact, improve the homogeneity of the affected groups.

3) Misclassification problems. A final issue of importance to the use of DRG's for reimbursement or for utilization review, relates to the fact that several diagnoses may have been underrepresented in the database used in the construction of the DRG's. One example is cystic fibrosis. Of the approximately 1.5 million records included in the original database, only 19 cases were found to have a primary diagnosis of cystic fibrosis (Horn, Horn, Sharkey, Beall, et al., 1986). The major reason identified to account for this underrepresentation was the finding that the DRG database did not include data from hospitals with cystic fibrosis treatment centres (Horn, et al., 1986). Consequently, these patients may have been placed inappropriately into DRG's that do not accurately reflect their resource utilization. The possibility that other infrequent cases may be inappropriately allocated has thus raised concerns that there may be a number of "garbage can" DRG's; i.e., DRG's into which "miscellaneous low-volume cases have been dumped" (Mendenhall, 1984, p. 86). To the extent that this actually occurs, it could create problems for utilization review as well as for other applications, such as reimbursement.
In summary, although the development of DRG's represents a major improvement in case mix measurement for the purposes of reimbursement, utilization review and research, it is still not considered to be problem-free. With time and additional research, however, it is likely that most, if not all, of the current problems can be resolved. Nonetheless, because of the potential for problems, with both DRG's and the perverse incentives provided by the PPS itself, hospitals will have to monitored closely so as to ensure that neither the quality of care nor access to care becomes problematic for Medicare beneficiaries. In the next section, the mechanisms that are currently being used to monitor hospital performance are described.

II.5.5 The Role of External Review Organizations Under PPS

In recognition of the potential negative effects of the PPS, a new external review system consisting of statewide PRO's has been implemented (Grimaldi & Micheletti, 1984). Although many of the PRO's are simply transformed PSRO's (Dans et al., 1985), their goals and objectives differ considerably from those of their predecessors. Under the new system of review, much greater emphasis is placed on quality assurance, admission and readmission review, and the review of medical appropriateness. Whereas the PSRO's were primarily concerned with reducing length of stay, such an emphasis is no longer necessary under the PPS; the payment system itself provides the necessary incentives to reduce length of stay. Moreover, since the rate of payment is based on the case, not the length of stay, any failures to achieve reductions in length of stay will result in financial problems for the hospital, not the government (Grimaldi & Micheletti, 1984).
II.5.5.1 Goals and Objectives of PRO's

The primary goal of the PRO program is to reduce inappropriate and unnecessary hospital utilization while maintaining an adequate standard of care. Specific objectives for the PRO program include: 1) shifting care from inpatient to outpatient settings, when it is safe and effective to do so; 2) monitoring patterns of admission and validating DRG assignments so as to prevent "DRG creep"; 3) monitoring patterns of readmissions so as to detect any increases in readmission rates; 4) reducing the number, and adverse consequences of, invasive procedures; 5) ensuring that medically necessary services are provided; and 6) ensuring that hospitals are not reimbursed for inappropriate or unnecessary use (Dans et al., 1985).

II.5.5.2 Characteristics of the PRO Program

The PRO program has five major components: admission review, outlier review, admission pattern review, DRG validation, and quality review.

1. Admission review. Because the PPS pays a flat rate per Medicare discharge, and thereby provides hospitals with an incentive to increase admissions, PRO admission review is conducted in order to ensure that unnecessary or inappropriate admissions are not reimbursed. Approximately 10% of Medicare admissions are subject to preadmission review (Lipp, 1984). For these patients, once the admission has been approved, no further review is required (unless the patient becomes a "day" or "cost" outlier) and reimbursement is guaranteed. The remaining 90% of Medicare admissions are reviewed retrospectively. Consequently, in the majority of cases, hospitals will face the risk of retroactive denials of reimbursement if admissions are found, after the fact, to have been
unnecessary or inappropriate (Lipp, 1984). To prevent one form of "double dipping", PRO's are directed to give special attention to admissions that are transferred from acute care beds to beds which are exempt from the PPS (eg. rehabilitation units, psychiatric units, or swing beds). In addition, to ensure that patients are not being discharged prematurely, special attention is to be paid to patients readmitted to an acute care hospital within seven days of a previous admission (Grimaldi & Micheletti, 1984).

2. **Outlier review.** Since hospitals can receive additional payment for "day" or "cost" outliers, PRO's are responsible for ensuring the validity of the outlier claim (Grimaldi & Micheletti, 1984). For "day" outliers, the medical reviewer is required to evaluate the appropriateness and necessity of all inpatient days prior to the outlier threshold. If a number of these days are found to have been unnecessary, no additional payment will be made; i.e., the hospital will receive the set DRG rate and no more. For "cost" outliers, the medical reviewer is required to review the necessity of the admission and the services rendered over the course of the inpatient episode. If any of these services are found to have been unnecessary, billed twice, not actually performed, or not ordered by a physician, additional payment may be denied (Grimaldi & Micheletti, 1984).

3. **Admission pattern review.** Since hospitals can increase their revenues by increasing their Medicare admissions, PRO's are required to monitor admission patterns. Some of the potential problems that the PRO will be seeking to identify include: 1) a tendency to treat outpatients on an inpatient basis; and 2) an increase in readmission rates resulting from patients being discharged prematurely or from a tendency to "use more than one stay to care for patients with multiple problems" (Grimaldi and
Micheletti, 1984, p. 31). Admission pattern review is conducted on a quarterly basis. A screening mechanism (eg. profile analysis) is used to identify hospitals with unusual changes in length of stay, number of admissions or readmissions, and case mix. Once aberrant hospitals have been identified, then the PRO is required to take "necessary corrective action, which may involve education activities, denial of payments, or expanded review" (Grimaldi and Micheletti, 1984, p. 32).

4. **DRG validation.** Because reimbursement is provided on the basis of DRG assignment, the PRO's are required to review, on a quarterly basis, the accuracy and completeness of the clinical information used by the hospital in assigning patients to DRG's. In particular, the reviewer is to pay attention to "whether hospitals are coding clinical information in a way that improperly maximizes reimbursement rather than accurately reflects a patient's medical condition" (Grimaldi and Micheletti, 1984, p. 33). If inaccuracies in the clinical information are found, or improper patterns of DRG assignment - "DRG creep" - are discovered, financial penalties may be imposed (Grimaldi and Micheletti, 1984).

5. **Quality review.** There is considerable uncertainty about what effect the PPS will have on the quality of inpatient care (Jencks et al., 1984; Omenn & Conrad, 1984; Stern & Epstein, 1985). In particular, there are fears that hospitals, in the effort to economize, will encourage underutilization of necessary services and premature discharges of some patients (Grimaldi & Micheletti, 1984). If so, the PPS could have a negative effect on quality of care. To counteract this possibility, PRO's are required to conduct quality reviews and to participate in national studies of abusive practices. At the present time, however, no specific criteria or guidelines for quality review
have been established (Grimaldi & Micheletti, 1984). Hence, until such
criteria and guidelines can be developed, the PRO's are expected to use, as
indicators of a reduction in quality, unprecedented reductions in a hospital's
average length of stay, and/or an increase in complaints from nursing
homes and home health care agencies that patients are being discharged
from the acute care setting in worse condition than they had been
previously (Grimaldi & Micheletti, 1984). If such changes are observed,
PRO's will be required to investigate further to determine whether patients
are being denied essential services or are being discharged prematurely. If
patterns of abuse are detected, the hospital can be excluded from
participation in Medicare or financial penalties can be imposed on the
physician(s) involved and/or the hospital (Lipp, 1984).

II.5.6 The Role of Internal Review Committees Under PPS

Under the PPS, a hospital utilization review committee is required
only for those hospitals that are not yet under contract with a PRO. Once,
the PRO program has been implemented nationwide, hospital utilization
review activities will cease to be mandatory requirement (Grimaldi &
Micheletti, 1984). It has been argued, however, that with the advent of the
PPS, the importance of the utilization review committee, to the hospital
itself, will increase substantially (Grimaldi & Micheletti, 1984; Owens &
Averill, 1984). In order to remain financially solvent, hospitals will need to
establish procedures for identifying inappropriate and unnecessary
utilization of hospital days and ancillary services. In addition, they will
need to integrate their utilization review and quality assurance activities in
order to avoid expensive duplication (Grimaldi & Micheletti, 1984). Most
importantly, utilization review committees will need to be empowered to
identify the reasons for misutilization and to implement appropriate corrective action (Grimaldi and Micheletti, 1984; Owens & Averill, 1984).

II.5.7 Current Trends in Utilization Review Procedures

With the new emphasis on appropriateness of acute inpatient care, PRO's and hospitals need tools that will enable them to identify, in an efficient and cost-effective manner, inappropriate admissions, days-of-care use, and ancillary service utilization. Historically, concerns about inappropriate utilization have emphasized overuse. However, inappropriate use also refers to underutilization (Payne, 1987). With the advent of the PPS, and its strong incentives to reduce utilization, concerns about underuse, and its close associate, quality of care, are increasing (Dans et al., 1985; Wyszewianski et al., 1987). To date, however, most of the utilization review procedures developed to identify inappropriate use still remain focussed on the problem of overutilization (Payne, 1987). And most of this effort has been directed at the identification of unnecessary or inappropriate admissions and days-of-care use; little effort has as yet been applied to the problem of assessing inappropriate utilization of ancillary services (Hughes et al., 1984). Consequently, only the methods that have been developed to assess unnecessary or inappropriate admissions and days-of-care use are discussed herein.

As has already been stated, one of the most commonly used procedures to identify unnecessary utilization is profile analysis; i.e, the retrospective evaluation of changes in length of stay patterns by DRG's, physicians, and inpatient service units. By establishing length of stay norms by DRG’s, patients with unusually short or unusually long stays can be identified. Very short stays may reflect inappropriate admissions;
exceptionally long stays may reflect inappropriate days-of-care (Goran, 1979). Length of stay profiles are limited in usefulness, however, because they provide no information as to the source or nature of the problem (Payne, 1987). In addition, they are vulnerable to the coding conventions used by specific hospitals. For example, under PPS there is a strong incentive for hospitals to record multiple diagnoses, complications and/or comorbidities in the sequence that maximizes reimbursement (Simborg, 1981). But, "if a patient is placed in a category with a higher severity level due to coding practices designed to maximize reimbursement, he or she will probably have a lower length of stay than others who are properly in the category" (Payne, 1987, p. 739). Consequently, to the extent that such coding practices occur, length of stay profiles could fail to identify patients that, had they been properly classified, would be in need of further review. In addition, since profile analysis is based on the use of DRG's, it is vulnerable to all of the problems discussed earlier in relation to DRG heterogeneity. Finally, as was also mentioned previously, since empirical evidence suggests there is no direct relationship between length of stay and inappropriate use (Restuccia & Holloway, 1976; Zimmer, 1974), length of stay profiles may miss utilization problems that occur during inpatient episodes that do not violate the established norms (Payne, 1987).

A more promising approach to the identification of overutilization, employs explicit criteria to assess the appropriateness of inpatient days and to identify the reasons for misutilization. At present, the AEP, which was described in section II.4.8, supplemented by a method for recording the reasons for inappropriate use, is considered the "state of the art" in appropriateness of care assessment (Payne, 1987). Since the development of
the AEP, however, other criterion-based review instruments have been
developed and are currently being used by hospitals and PRO's to assess the
appropriateness of acute level care. For example, Interqual, Inc. has
developed a review system called the Intensity of Service, Severity of Illness,
and Discharge Screens with an Appropriateness component (the ISD-A;
Lamprey & Jacobs, 1987). Similarly, Systemetrics, Inc. has developed a tool
called the Standardized Medreview Instrument (SMI; Payne, 1987). Both of
these instruments, like the AEP, can be applied in both admission and day-
of-care review. There is considerable variability, however, among the three
tools in the number of criteria employed, the complexity of their decision
rules, and the extent to which they have been tested for reliability and
validity.

As indicated earlier, the AEP consists of 27 day-of-care criteria and 18
admission criteria. In contrast, the ISD-A consists of 69 generic criteria plus
an additional 20-60 criteria which are applied on the basis of the body system
affected. Specific criteria have been developed for a total of 12 body systems
as well as for psychiatric patients (Payne, 1987). The SMI, on the other hand,
consists of 117 admission criteria (ordered from the most to the least
objective), 30 level of care criteria, and 26 continued stay criteria (Payne,
1987).

The decision rule for the AEP is relatively straightforward. For an
admission or day of care to be considered appropriate, only one of the
relevant criteria needs to be met (Gertman & Restuccia, 1981). With the ISD-
A, "any patient admitted to the hospital must [meet] either one severity of
illness (SI) or one intensity of service (IS) criterion on admission and must
have met both an SI and an IS criterion by the first review following the
completion of 24 hours in the hospital” (Lamprey & Jacobs, 1987, p. 6).
Finally, for the SMI, if one of the 117 admission criteria are met, the
admission is considered appropriate; if one of the level of care criteria and
one of the continued stay criteria are met, the day of care is considered
appropriate (Payne, 1987).

With respect to the reliability and validity of the three instruments,
results have been reported in the literature for the AEP and the SMI, only.
The ISD-A is proprietary and, to date, no results from the systematic testing
of the instrument, if it has even been conducted, have been published
(Payne, 1987). In contrast, the AEP has been subjected to extensive testing
and has been found to be reliable and valid by both the instrument's
developers and others (Gertman and Restuccia, 1981; Siu et al., 1986;
Stumwasser, Paranjpe, et al., as cited in Payne, 1987). The SMI, on the other
hand, has been found to have low reliability and, as a result, it has been
suggested that it no longer be used for the purposes of utilization review
(Stumwasser, Paranjpe, et al., as cited in Payne, 1987).

Appropriateness measures have two major limitations. Firstly, as
with the length of stay profiles, appropriateness measures are intended to be
screening devices. Thus, any problems identified through these procedures
must be considered tentative until a more indepth review of the particular
case can be conducted. In particular, since the initial reviewers are usually
nurses, questions concerning the appropriateness of an admission,
continued stay or reimbursement must be referred to a designated physician
advisor before any decisions are made (Payne, 1987). Secondly, all three
measures include some component of subjective judgement. Thus, there is
a potential for reviewer bias. To minimize this potential, reviewers must by
thoroughly trained, supervised, and monitored so as to ensure that the criteria are applied appropriately and effectively (Payne, 1987).

The application of tools such as the AEP and the ISD-A in utilization review and utilization management is a recent development. Thus, at the present time, it is unknown whether such procedures are effective in reducing inappropriate use. In addition, it is unknown whether they are more or less effective or cost-effective than profile analyses. Intuitively, it would seem reasonable to assume that by targeting all inappropriate use, rather than just unnecessarily long stays, greater reductions in inappropriate use could be achieved. Whether these reductions will be sufficiently large to warrant the further development and implementation of these techniques will depend on the extent to which the inappropriate use they identify can be controlled by the hospital and its medical staff. As noted earlier, preliminary research suggests that the major reasons for inappropriate use are related to physician practice patterns and the lack of appropriate alternatives to acute care. In both cases, time will be required before the necessary corrective action can be taken; neither is immediately under the control of the hospital or its medical staff. Therefore, the development, testing, and evaluation of appropriateness measures will have to be conducted over a sufficiently long period of time to permit the development and implementation of the appropriate long term strategies for change.

II.5.8 Early Effects of the PPS on Hospital Utilization, Costs, and Quality

Given that the PPS is still in the process of being phased in, it is too early to determine whether or not it has been effective in reducing the rate of increase in hospital expenditures. In fact, it will be a number of years before sufficient stability is achieved within the system to permit researchers
to draw any firm conclusions about its effects on hospital costs, utilization or quality (May, 1985; May & Wasserman, 1984). In the interim, a few preliminary studies of national trends in hospital costs, use, and quality have begun to appear in the literature. For the most part, however, the studies that have been undertaken, to date, to evaluate the effectiveness of DRG-based prospective payment have focused on the experience of the experimental program that was implemented in the State of New Jersey in 1980. Since this system was used as the model for the Medicare PPS, the findings from the New Jersey experiment are expected to have important implications for the Medicare system (Hsiao & Dunn, 1987; May, 1985; May & Wasserman, 1984). However, there is one important difference between the two systems that will limit the generalizability of the New Jersey studies. In contrast to the federal PPS, which applies only to Medicare beneficiaries, the New Jersey system applies to all payers (May, 1985; May & Wasserman, 1984; Iglehart, 1982b). Consequently, there is a greater potential for cost shifting under the federal system (Iglehart, 1982b; Wennberg et al., 1984). Therefore, the findings concerning the effectiveness of the New Jersey system are likely to provide an overestimate of what can be achieved under the federal system (May & Wasserman, 1984).

II.5.8.1 National Trends in Utilization, Costs and Quality

For the most part, the findings from the available national trend studies (Des Harnais et al., 1987; Kelly & Bankhead, 1985; Shortell & Hughes, 1988) are in accordance with the objectives of the PPS. Although the data included in these studies relate only to the first year following implementation of the PPS, comparisons to pre-PPS trends indicate that the average length of stay had decreased significantly (Des Harnais et al., 1987;
Kelly and Bankhead, 1985), and that total hospital costs were approximately "$2 billion dollars less than Medicare officials had expected to pay for hospital charges [in 1984]" (Kelly & Bankhead, 1985). Moreover, despite fears that the PPS would generate increases in admission and readmission rates, no significant increases in either rate were observed in 1984 (Des Harnais et al., 1987; Kelly & Bankhead, 1985). In fact, Des Harnais et al. (1987) reported that, for the 729 hospitals included in their data base, Medicare admissions declined by 5.4% in 1984. Similarly, but in a more specific study, a recent examination of readmission rates among Medicare patients treated for congestive heart failure revealed no significant increases in readmissions before and after implementation of the PPS (Rich & Freedland, 1988). These findings are of limited usefulness, however, because they are based on the experience of only one hospital, and are relevant to only one DRG.

Although the above mentioned findings may be suggestive of PPS effects, none of the studies referred to above have been able to attribute their results directly to the influence of the PPS. Most of the observed trends are continuations of trends that had begun prior to the implementation of the PPS (Des Harnais et al., 1987; Kelly & Bankhead, 1985). In addition, no effort has as yet been made to take into account the effects of other factors such as PRO's, or the growth of prepaid practice organizations, on hospital utilization and total health care expenditures (Kelly & Bankhead, 1985). Finally, no studies have as yet begun to examine the cost-effectiveness of the PPS.

With respect to the effect of the PPS on the quality of care, the situation is even more uncertain. Des Harnais et al. (1987) examined trends in mortality rates, comorbidity/complication rates, and severity of illness
from 1980 to 1984 and found no significant increases in any of these measures following implementation of the PPS. However, in a more detailed study of the effects of regulatory constraints, payment constraints, and competition on inpatient mortality rates, Shortell and Hughes (1988) found significant associations between high mortality rates and the stringency of prospective rate-review programs. In addition, hospitals in highly competitive markets (as measured by enrollment in HMO's) were found to have higher mortality rates than hospitals in less competitive markets. While the authors acknowledged that their findings were exploratory, and limited to 16 medical conditions, the fact that they were based on data from 981 hospitals across 45 states led them to conclude that there is a "need for improved monitoring of the issue of the quality of care and patients' outcomes as regulatory and competitive approaches to hospital cost containment continue to become more stringent" (p. 1106).

II.5.8.2 The Results of the New Jersey Experiment: Implications for the Medicare PPS

In one of the earliest studies to examine the effects of the New Jersey DRG-based prospective payment system, May & Wasserman (1984) compared a matched sample of DRG and non-DRG hospitals. All eight of the DRG hospitals included in the study had entered the prospective payment system during the first year of its operation in 1980. Data from 1979 to 1983 was examined to determine whether the DRG system had had a measurable effect on organizational and economic variables. The relevant findings of this study are summarized as follows:

1. The amount and quality of data collected by the DRG hospitals improved following implementation of prospective payment. The number
of incomplete discharge abstracts dropped from 22.8% to 15.8%. However, the time taken by medical records departments to complete the data abstraction process increased from 4.5 days to 5.3 days.

2. Physicians were more involved in managerial decision-making in the DRG hospitals than in the non-DRG hospitals.

3. On average, it cost each of the eight DRG hospitals studied an additional $91,092 per annum to prepare for and implement the new prospective payment system.

4. In 1980, each of the hospitals that entered the new system received an average of $2.3 million more in reimbursement than they would have under the state's previous reimbursement plan.

5. In large part due to the influx of additional revenues, which were provided to hospitals in order to offset start-up costs, there was no measurable reduction in the rate of cost increases that could be attributed to the new reimbursement system.

On the basis of these findings, May & Wasserman (1984) concluded that, in terms of data quality and physician involvement, the New Jersey system was having the anticipated effects. With respect to cost containment, however, they were more cautious. Although none of the data they examined suggested that the system had been effective in containing costs, they noted that it would be some years before it would be possible to determine whether or not the expected gains in efficiency would be sufficient to offset the substantial start-up costs that were required for the system's implementation.

In a more recent study, Hsiao & Dunn (1987) performed time series analyses on data from 80 of the 97 hospitals that had entered the New Jersey
system by the end of 1982. The dependent variables used in their analyses included total per capita expenditures, total costs per case, average length of stay and inpatient admissions. The time period covered in the study extended from 1971 to 1984. It should be noted that, during this period, and prior to the introduction of the DRG-based prospective payment system, a prospective per diem system of reimbursement had been implemented by the State of New Jersey. This system was implemented in 1976 and remained in effect until its replacement by the DRG system in 1980 (Hsaio & Dunn, 1987). Thus, the time trend analysis conducted by Hsaio & Dunn covered three different reimbursement periods: the pre-prospective payment period, 1971-1975; the per diem prospective payment period, 1976-1979; and the DRG-based prospective payment period, 1980-1984. For the purposes of the analyses of expenditures, all cost data was converted to 1969 constant dollars. However, no control group was included in the analysis nor was any attempt made to include additional factors, other than time and intervention, that might affect the study's dependent variables. Although these limitations in the study's design make it impossible to attribute any of the observed trends directly to the new payment system, the implications of the findings for the Medicare PPS merit their discussion. The major findings are summarized as follows:

1. A significant reduction in the rate of growth of per capita expenditures was associated with the introduction of the prospective per diem system. Prior to the implementation of this system, per capita costs had been increasing at a rate of 8% per year; following its implementation the rate was 2%-3% per year. This trend continued following implementation of the DRG system.
2. Total costs per admission increased at an average annual rate of 3%-4% prior to 1976. Following implementation of the per diem system, the rate declined to 2% per annum. With the implementation of the DRG system, the rate declined to less than .5% per annum. For both of these systems, the decrease was significant.

3. Average length of stay changed little between 1971 and 1975. With the introduction of the per diem system, however, a significant increase in length of stay was observed. In contrast, following implementation of the DRG system, length of stay dropped significantly.

4. Prior to the implementation of the per diem system, admissions were increasing at a rate of 4% per year. After the per diem system was implemented this rate slowed to .7% per year. This change was significant. With the advent of the DRG system, admissions started to increase slightly, at a rate of 1% per year, but this increase was not statistically significant.

As noted by Hsiao and Dunn (1987), these results suggest that although the DRG system appears to have been effective in reducing length of stay and, consequently, the costs per admission, it has been no more effective than the per diem prospective system in controlling the rate of growth of overall per capita expenditures. In attempting to explain this result, the authors suggested that the increase in the number of admissions observed, following implementation of the DRG system, may have been sufficient, even though not statistically significant, to offset the cost savings generated by the reductions in length of stay and costs per case. However, it should be emphasized that both prospective payment systems were more effective than the retrospective payment system in containing costs and reducing utilization. Although Hsiao & Dunn did not attempt to measure
the effect of the New Jersey system on the quality of care provided, they concluded that because of the system's effect on length of stay, there is a need to apply vigilance in monitoring discharge practices. In addition, because of the slight increase in admissions observed under DRG-based prospective payment, they suggested that there is "a need to regulate and monitor the appropriateness of clinical services" (p. 219).

II.5.9 Comments

The evolution of utilization review and utilization management in the United States has been a dynamic process spanning more than two decades of study, debate, and, more often than not, expensive trial and error. And, even though knowledge has been gained, much is still unknown about the factors that affect hospital utilization. Clearly, utilization is a complex process that is not easily addressed by simplistic assumptions and policy initiatives. A multiplicity of factors contribute to the pattern of utilization observed in any given hospital, and, consequently, attempts to effect changes in utilization have been hampered by a lack of data and an inability to study more than a few of the important variables at a time. This basic lack of understanding has contributed to the numerous unanticipated effects of the government's attempts to regulate hospital utilization and to impose limits on expenditures for hospital-based medical care. The most salient example of unanticipated effects is the PSRO program that dominated the publicly financed hospital sector throughout the 1970's. Although at the time it appeared to have merit intuitively, the expenses incurred in implementing, maintaining, and monitoring the program far outweighed any of the observable benefits. Yet, it took almost a decade for the limitations of the PSRO program to be recognized and addressed. And
even though, by the early 1980's, it was widely acknowledged to have been a failure, the subsequent development of the PPS, the implementation of DRG's, and the development of the PRO's, has proceeded once again without adequate pilot testing and evaluation. In particular, one of the last topics to be addressed in any of the policy initiatives that have been enacted in the United States, has been their cost-effectiveness. With the advent of the DRG-based PPS, it is somewhat surprising that more time was not taken to study its effects before engaging in nationwide implementation. Although, clearly, this was a move born of necessity, given the state of the Medicare trust fund, one might have expected a little more diligence in view of the expensive and time-consuming failure of the PSRO program effort.

For Canadian policy analysts, hospital administrators, health service researchers, and government officials, thoughtful consideration of the American experience is important for a number of reasons:

1. It reveals the enormous amount of time and money that can be spent on unproductive strategies when program development and policy formation is based on trial and error rather than on empirically derived knowledge.

2. It underscores the importance of pilot projects, complete with formal evaluation plans, in program development.

3. It highlights the need for patience before implementing program strategies that involve major public expenditures.

4. It provides a major rationale for the promotion of health services research, an area of research that, to this day, remains highly undeveloped in both the United States and Canada.
Although the Canadian and American health care systems differ in many important ways, an understanding of the American experience could help Canadian program developers avoid some of the more costly errors incurred in the United States. Thus, for example, it is hoped that through this rather long description of the history of utilization review and management, readers of this document will have gained an increased appreciation for the importance of administrative and organizational research in the hospital environment. As well, since utilization review and utilization management are topics that are only just beginning to attract the attention of Canadian hospital administrators and other health service professionals, it is hoped that a sincere effort will be made to incorporate program evaluation into the design, implementation and maintenance of any utilization management program efforts. In so doing it may be possible to avoid some of the more deleterious effects of such programs and maintain the Canadian hospital-based care at the high standard that it has historically enjoyed.
CHAPTER III

Utilization Management in Canada:
The SWITCH Index System and Other Developments

III.1 The Conceptual Context of Utilization Management in Canada

As indicated previously, hospital utilization management has only recently begun to attract attention in Canada. Consequently, little information on the nature and status of utilization management in the Canadian hospital sector has as yet been published. Nonetheless, there are indications that government officials, health service researchers, hospital administrators and other health care professionals are beginning to recognize and show an interest in techniques designed to monitor and improve the management of hospital resource use (Anderson & Lomas, 1988). For example, at a national level, interest in promoting improvements in resource management methodologies has been reflected in the ongoing development and implementation of the Management Information Systems (MIS) Project (Bolley, 1987; Lowry, 1983). The MIS Project, which "is a co-operative effort of the Canadian Hospital Association, the federal and provincial governments and the provincial hospital/health associations" (MIS Project, 1986, p. 27), is intended "to make data management and collection comparable" (p.27) and to provide a "standard and potentially powerful tool for monitoring resource use" (Anderson & Lomas, 1988, p.17). Likewise, at the hospital level, utilization management program descriptions and case studies have begun to appear in the literature (Coombs & Richter, 1986; Eliasoph & Hassen, 1986; 1987; McDonough & Vaz, 1987; McGeorge, de Mora, & Legros, 1985; Nusbaum, 1988; Winchell, 1985). In addition, health service professionals have begun to identify
appropriate utilization as one of the major challenges facing hospitals in the 1990's (Lazare, 1987); at least one health association has recently held a workshop on utilization management (British Columbia Health Association, 1987); and papers calling attention to the need for improved hospital resource management have become common (Canadian Hospital Association, 1984; Clemenhagen, Champagne, Contandriopoulos, & Pineault, 1985; Klicius & Sawyer, 1986; Komes, 1988; Macies, 1984; MacKenzie, Markle, & Croke, 1987; Schultz, 1986). Finally, at a conference held by the Centre for Health Economics and Policy Analysis (CHEPA) in May, 1988, four papers were presented, each of which dealt exclusively with the topic of utilization analysis in Canada (Anderson & Lomas, 1988; Rachlis & Fooks, 1988; Roos & Roos, 1988; Suttie, 1988).

Despite these indications of interest, utilization management has been slower to emerge as a priority for Canadian hospitals than for their American counterparts. It will be recalled that, in the United States, utilization review and management was mandated by the federal government in an attempt to contain rapidly escalating costs in the publicly financed hospital sector. Although it was subsequently recognized that, in the absence of direct financial controls, utilization management was not producing the desired effect, the mixture of public and private funding sources and delivery systems that comprise the American health care scene prohibited the implementation of the kinds of controls, such as prospective global budgeting, deemed necessary to achieve cost containment (Evans, 1983; Iglehart, 1982a). With the advent of DRG's, however, a tool was provided for the introduction of prospectively applied financial constraints and, consequently, the United States is currently attempting to contain
hospital expenditures through a combination of financial and utilization controls. In contrast, Canada's federal government has not responded to cost concerns by mandating utilization controls. Although it has been recognized that inappropriate and unnecessary utilization of hospital facilities and services contribute to overall hospital expenditures (Bennett & Krasny, 1977; Canadian Hospital Association, 1984; Evans, 1983; 1984; Soderstrom, 1978), and thereby to total health care expenditures, constitutional arrangements that place responsibility for all health and hospital services, health education, and regulation of the health professionals with the provincial governments, prohibit the Canadian federal government from enacting legislation aimed directly at controlling hospital utilization (Evans, 1983; LeClair, 1975; Soderstrom, 1978). However, because Canada has a national health insurance program which is financed, in part, by the federal government, it can attempt to influence the way in which provincial governments structure and administer provincial health services by changing the financial incentives provided under the terms of Federal-Provincial cost-sharing agreements (Evans, 1983). Consequently, when concerns about rising health care costs began to surface in the mid-1970's, the federal government responded by passing the Federal-Provincial Fiscal Arrangements and Established Programs Financing Act (EPFA). With the enactment of the EPFA, the previously open-ended cost-sharing formula, which essentially worked out to a 50-50 federal and provincial contribution, was replaced with block grants and tax transfer payments designed to link "the annual increase in the federal contribution...to the growth of the GNP" (Iglehart, 1986a, p. 207). In addition, annual grants of $20 per capita (indexed to the GNP) were introduced in an attempt to
promote the development of less costly alternatives to hospital care (Iglehart, 1986a). In essence, the aim of this legislation was to transfer responsibility for cost containment to the provinces by placing them at greater financial risk "when aggregate outlays for health care [grew] at a faster rate than the economy as a whole" (Iglehart, 1986a, p. 207). In so doing, provincial governments were provided with greater incentives than previously to find ways to improve the efficiency and reduce the costs of hospital (and other) services.

At the provincial level, governmental responses to the changes in the federal funding formula have similarly been aimed primarily at changing traditional financial incentives. But in this case, the change in incentives has been directed mainly at the hospital sector. Thus, in most provinces, hospitals are now funded on the basis of prospectively determined global budgets (Evans, 1983). This contrasts with the earlier system of line-item budgets and actual cost reimbursement which provided hospitals with little flexibility in the way in which funds could be used and few incentives to contain costs, since budget deficits were typically absorbed by the government (Soderstrom, 1978). With the introduction of global budgets, however, most provincial governments have introduced a policy whereby hospitals that overrun their budget do not automatically receive deficit funding. Moreover, in some instances, hospitals that manage to end the year with a surplus are permitted to retain at least a part of that surplus (Soderstrom, 1978; British Columbia Ministry of Health, 1985; McDonough & Vaz, 1987). Thus, as a result of these changes, hospitals have been provided with greater incentives to contain the costs of their operations
since any deficits they incur may have to be made up through sources outside the provincial insurance program.

A second response of provincial governments has been to curtail hospital expansion (Bennett & Krasny, 1977). Such a response is easier to implement, in Canada than in the United States because of the separation, in Canada, of operating expenses from capital spending (Evans, 1983; Iglehart, 1986b). Hospital's must apply to their provincial ministry of health in order to obtain approval and/or funding for capital expansion. Consequently, although as Evan's (1983) states "the process is not perfect, and is often rather ad hoc, ... on balance it has been substantially more successful in constraining hospital investment than the U.S. free-for-all" (p. 13).

Until recently, at the hospital level, the most typical response of administrators to provincial economic restraint policies has been to close hospital beds and lay off staff (Barrable, 1987; Mercer, 1986;). As noted by Mercer (1986), however, such responses are essentially "short-term solutions to what was viewed as a short-term problem" (p. 39). However, since there are no signs that economic restraint policies are likely to be abandoned in the near future, new management strategies are required to assist hospitals in developing effective long-term responses to conditions of cost constraint. Furthermore, the need for such strategies is becoming more acute since most hospitals, after more than a decade of economic restraint, are currently facing one or more of the following problems: "long waiting lists, bed shortages, forced reduction of beds, need for new and expanded services, discharge problems and the lack of community services, and high occupancy rates" (McDonough & Vaz, 1987, p. 42). In addition, many of Canada's
hospitals are getting old and are in need of restoration and updating (Canadian Hospital Association, 1984; Iglehart, 1986a).

As a result of the various cost containment initiatives described above, Canada has apparently been more successful than the United States in containing total health care expenditures (Evans, 1983; Iglehart, 1986a). Thus, while the United States is currently spending more than 11% of its GNP on health care (Ginzberg, 1987), Canada has managed to hold expenditures relatively stable at approximately 8.6% of its GNP (National Health Expenditures, 1987). Moreover, "Canada's control over health care expenditures has been achieved with comprehensive coverage both of people and of services" (Evans, 1983, p. 4). In contrast, over 30 million American citizens currently have no insurance coverage at all (Iglehart, 1986a). However, the problems that are arising in response to continued economic restraint and, in particular, the growing concern about its eventual impact on the quality and accessibility to care, have led some analysts to suggest that a new spate of increases, within the hospital sector at least, may well be just around the corner (Evans, 1983;1984; Iglehart, 1986a).

At the same time, however, it continues to be assumed by the federal government, provincial governments and a number of health service professionals that a considerable portion of current hospital utilization is inappropriate and/or unnecessary (Anderson & Lomas, 1988; Canadian Hospital Association, 1984; Iglehart, 1986a; Roos & Roos, 1988; Suttie, 1988). Although few studies have as yet been undertaken to quantify the extent of inappropriate or unnecessary use in Canadian hospitals, research on variations in the use of surgical procedures such as cesarean sections, hysterectomies, tonsillectomies and adenoidectomies, and studies
comparing provincial and international variations in length of hospital stay suggest that there is room for improvement in the utilization of hospital services and facilities (Anderson & Lomas, 1984, 1985; Bennett & Krasny, 1977; Dyck, et al., 1977; Evans, 1983, 1984; Roos, 1983; Roos, Henteleff, & Roos, 1977; Roos, 1979; Roos & Roos, 1988). As indicated by Roos & Roos (1988), "[r]esearch using utilization data has challenged the assumption that the knowledge base underlying conventional medical practice produces a rational allocation of resources among clinical priorities and limits demand to those services known to be effective" (p.3). Consequently, there is growing acceptance of the belief that "at least some of the things that happen to patients and some of the time patients spend in hospital are controllable by the hospital and/or its medical staff" (Halperin & Neuhauser, 1976, p 63). To deal with this situation, a number of remedial programs have been suggested. These include the development of a regional referral and admission scheduling process (Canadian Hospital Association, 1984), the development of external utilization review committees (Suttie, 1988), expanded emphasis on prevention and the development of alternatives to hospital care (Bennett & Krasny, 1977; Roos & Roos, 1988; Soderstrom, 1978), and intensification of government intervention in health care delivery, through utilization analysis and management (Evans, 1983; Suttie, 1988). Thus, as one analyst suggests, an appropriate government intervention would be the "requirement that, before requests from hospitals for capital and deficit funding be considered, active utilization analysis be demonstrated in the subject area of the request" (Suttie, 1988, p. 18).
While there is at present no mandate for hospitals to develop comprehensive utilization management programs, a few hospitals have begun to experiment with utilization management procedures. From the perspective of these organizations, such procedures are necessary for the purposes of both cost-containment and quality assurance (Clemenhagen et al., 1985; Coombs & Richter, 1986; Eliasoph & Hassen, 1986; 1987; McDonough & Vaz, 1987; McGeorge et al., 1985; Nusbaum, 1988; Winchell, 1985). Although most hospitals have implemented some form of utilization review, since utilization review is a requirement for hospital accreditation (Canadian Comprehensive Auditing Foundation, 1987; Rachlis & Fooks, 1988), the usual emphasis has been on reviewing utilization for the purpose of medical audit rather than for efficiency or cost-effectiveness (Bennett & Krasny, 1977; Rachlis & Fooks, 1988). This is perhaps understandable given that utilization review falls under the jurisdiction of physicians who, traditionally and appropriately, have been more concerned with clinical issues than with issues of cost (Bennett & Krasny, 1977; Clemenhagen et al., 1985). Given the current era of cost restraint, however, a few hospital boards and hospital administrators are beginning to play a more dynamic role in managing the utilization of the hospital’s resources. Thus, while utilization review still remains the responsibility of the hospital’s medical staff, administrators are becoming more involved in the process by working together with physicians to find ways to eliminate unnecessary and inappropriate use. Clearly, the elimination of misutilization is a desirable goal. From a quality of care perspective, beneficial effects would likely include a reduction in the potential psychosocial and iatrogenic effects of unnecessary services. From a cost perspective,
beneficial effects would likely include the elimination of the expense of unnecessary services and, therefore, the freeing up of resources which could be used to service unmet needs (Restuccia, 1982). In order to work, however, a level of cooperation and communication is needed between administrators and their medical staffs that has been uncommon in previous years Bennett & Krasny, 1977; Evans, 1983, 1984; Iglehart, 1986a; 1986b; Suttie, 1988). With the pressures of reduced bed capacity, high occupancy rates, and long waiting lists, all of which make it more difficult for physicians to admit their patients into the hospital, the opportunity to develop this needed level of cooperation has presented itself to a number of receptive administrators. In response there has been a growth in the development of new and innovative strategies aimed, at least in part, at effecting changes in the way in which providers utilize the hospital's services and facilities. In the following section, the major developments that have occurred, other than the SWITCH Index System, will be described. Unfortunately, few studies have as yet been done to determine whether the programs that have been implemented have achieved their objectives. However, to the extent that any data is available it will be discussed. Following this review, a detailed description of the SWITCH Index System will be provided and the purpose of the present study will be described.

III.2 Current Status of Utilization Management in Canada

On the basis of the available literature, it would appear that the utilization review/management programs implemented in Canadian hospitals tend to be retrospective and diagnosis- or service-specific. Only three hospitals appear to have as yet begun to experiment with concurrent, diagnosis-independent programs comparable to the American AEP.
However, as noted by Rachlis & Fooks (1988), to obtain a more comprehensive picture of the state of utilization management in Canada, a survey of individual hospitals would have to be conducted. Such a survey was beyond the scope of this thesis and, therefore, it is possible that some equally innovative programs, other than those described below, may currently be underway.

III.2.1 Retrospective Review Programs

III.2.1.1 Diagnosis-Specific Profile Analyses

As in the United States, interest in utilization management in Canada has been given impetus, in part, by the development of a patient classification system analogous to DRG's. This system, called Case Mix Groups (CMG's) was developed in the early 1980's by the Hospital Medical Records Institute (HMRI). The HMRI is a not-for-profit, federally chartered company that specializes in information processing for health care institutions (HMRI, 1983; Hodgson & Ormerod, 1985). Canadian hospitals can submit discharge abstract data to the HMRI (or to the American PAS service) for processing and analysis. Information extracted from the medical record and submitted to the HMRI includes province and institution, patient sex, date of birth, length of stay, admission date, discharge date, admission status, discharge status, most responsible physician code, diagnoses, and procedures (HMRI, 1983). A complete list of the data elements included on the HMRI Discharge Abstract is provided in Appendix D.

In developing CMG's, the HMRI used essentially the same procedure as was used in developing the DRG's, with two exceptions: 1) in concert with Canadian practice, the diagnostic and procedure codes used in
constructing CMG's were based on the ICD-9 rather than the ICD-9-CM coding conventions; and 2) CMG's were constructed using the primary diagnosis (i.e., the one most responsible for the patient's pattern of utilization) rather than the principle diagnosis (i.e., the one most responsible for prompting the decision to admit) (MacKenzie et al., 1987). Like DRG's, the dependent variable used to reflect patterns of utilization was length of hospital stay. In addition, age and sex were included as predictors of length of stay when these variables were found to affect the observed pattern of utilization. The result of the classification process was the development of 465 CMG's which, like their American counterparts, were supposed to categorize patients according to their resource utilization (as measured by length of stay) into relatively homogeneous groups (Botz & Singh, 1985; Hodgson & Ormerod, 1985; MacKenzie et al., 1987).

With the development of CMG's, a number of analysts began to promote their use in applications such as cost-accounting, budgeting, resource management, and quality assurance (Botz & Singh, 1985; Hodgson & Ormerod, 1985; Macies, 1984; May, 1983; McGeorge et al., 1985; MacKenzie et al., 1987; Stoughton, 1983; Zuckerman, 1983). To facilitate their use in utilization review and management, the HMRI routinely provides to each member hospital quarterly and annual reports on the hospital's individual performance and the performance of a peer group\footnote{HMRI peer groups are defined according to bed size and type. Thus, hospitals are placed into one of the following groups: 1-99 bed hospitals; 100-199 bed hospitals; 200-399 bed hospitals; 400+ bed hospitals; Teaching hospitals; Paediatric hospitals. Hospitals can be in more than one group; for example, a 300 bed teaching hospital (HMRI, 1984).} of hospitals. A variety of reports are provided including length of stay profiles by CMG, by physician, and by patient service (HMRI, 1983). Despite their availability, profile
analyses based on CMG's have not yet become, to any great extent, a major component of most hospital's utilization review or management activities (Suttie, 1988). One notable exception, is the Kingston General Hospital in Kingston, Ontario. Originally a test site for the CMG system, this hospital has devoted considerable effort to the task of developing and implementing product-line costing and utilization review systems based on CMG's (McGeorge et al., 1985).

Two reasons have been identified for the apparent reticence, on the part of hospital administrators and utilization review committees, to incorporate CMG-based length of stay review. First, according to the results of a survey conducted by Suttie (1988), many hospital utilization review committees lack the necessary technical expertise to understand and apply profile analysis derived from HMRI (CMG-based) length of stay reports. Second, a number of writers have suggested that there is a reluctance to use CMG's because of the criticisms, discussed in Chapter II, that have been raised in connection with the DRG system. Of particular concern are the criticisms related to the failure of DRG's, and presumably CMG's, to account adequately for variations in severity of illness (Hodgson & Ormerod, 1985; McGeorge et al., 1985; Mackenzie et al., 1987; Stoughton, 1983). As noted by McGeorge et. al. (1985), similar concerns about the variability within DRG/CMG categories were raised during the development of the CMG-based management system's at Kingston General Hospital. In response, the system was modified to incorporate the Severity of Illness Index developed by Horn and her associates (Horn, Horn, & Sharkey, 1984). Likewise, the HMRI continues to refine CMG's (MacKenzie et al., 1987). Consequently, some analysts expect that implementation of CMG-based systems will
expand when CMG's have been reformulated to include a severity of illness measure. The inclusion of such a measure is anticipated to occur with the next generation of CMG's. (MacKenzie et al., 1987). It is to be hoped that when and if this expanded use occurs, evaluations of the effectiveness and cost-effectiveness of emerging CMG-based resource management strategies will begin to appear in the literature.

While CMG-based length of stay profiles are, at present, the most commonly available screening mechanisms for use in utilization management programs, six hospitals have begun to develop and implement an alternative form of diagnosis-specific profile analysis called the Value Improvement Process (VIP). These hospitals include: the Foothills Hospital in Calgary, Alberta, the Vancouver General Hospital in Vancouver, British Columbia, the Toronto General Hospital in Toronto, Ontario, the Montreal General Hospital in Montreal, Quebec, the Royal Victoria Hospital in Montreal, Quebec, and the Victoria General Hospital in Halifax, Nova Scotia.

The VIP was developed by the Baxter Corporation (previously known as Travenol, Inc.), an American-based consulting firm. It is a proprietary system and, as such, there appears to be no background information in the academic literature that details the history of its development or provides any data with respect to its reliability, validity, or effectiveness. However, a few articles have been published which describe, and provide preliminary data on, the VIP as it has been implemented in Canadian hospital settings.

Like other utilization management programs, the primary objective of the VIP is to reduce costs while maintaining, or improving, quality (Winchell, 1985). However, the approach taken to achieve this objective is
quite unique. In essence, the program is conducted in two phases. Phase I
involves a detailed review of patient's medical records to identify
"homogeneous groups of patients with similar conditions, operations, and
physical attributes" (Eliasoph & Hassen, 1986). The end result of this review
is the selection of a sample of prototypical patients, for a given therapy, (eg.
myocardial infarction) with no residual complications that would alter the
treatment regimen (Eliasoph & Hassen, 1986). Once this sample has been
identified, "Cost Profiles" are developed on the basis of six cost-component
categories, namely: blood and therapy, lab and tests, pharmacy, supply, O.R.,
and Room & Care (Winchell, 1985). The process is repeated at each of the six
participating hospitals in order to compile "Cost Profiles" that are
comparable from one hospital to the other. Once all six Cost Profiles have
been developed the information is used to construct a "Best Demonstrated
Cost Profile" (BDC). This BDC reflects the "lowest cost performance of each
category from all participating hospitals. In other words, the lowest Room-
and-Care amongst all six hospitals combined with the lowest Blood-and-
Therapy cost, lowest Pharmacy, etc." (Winchell, 1985, p. 85). This
information is then provided to a special hospital Task Force composed of
physicians, nurses, support staff, and administrators (Eliasoph & Hassen,
1987). By comparing their hospital's cost performance to the other hospitals'
performance and to the BDC, the Task Force can begin to target areas, within
specific medical therapies, where opportunities for improvements in the
cost and quality of patient care can be made (Eliasoph & Hassen, 1986). Phase
II then involves the development, within each hospital, of specific
recommendations for improvement, and the formulation of a detailed
implementation plan (Winchell, 1985).
To date, the Foothills Hospital has progressed the furthest of the six hospitals in implementing the VIP. One of the factors that has been suggested to account for the faster response at the Foothills Hospital is the fact that, since 1981, it has also been a participant in an Alberta per-case funding experiment (Coombs & Richter, 1986). Thus, similar to the DRG-based reimbursement system applied in the United States, the Foothills Hospital has a greater incentive to reduce utilization than hospitals that are reimbursed on the basis of global budgets (Winchell, 1985). Moreover, since the VIP process is expensive and time-consuming, some of the original hospitals have had to delay further implementation until additional resources can be made available (Winchell, 1985).

According to Eliasoph & Hassen (1986; 1987), a total of ten VIP studies have been completed at the Foothills Hospital on procedures such as myocardial infarction, total hip replacement, cholecystectomies, hysterectomy, and cataract surgery. As would be expected, a common theme emerging from all of the studies was the importance of length of stay in determining each therapy's total costs. Consequently, length of stay became a major target for reduction. Thus, for example, on the basis of Phase I investigations triggered by the inter-hospital comparison of "Cost-Profiles" for total hip replacements, it was determined that two factors were contributing to unnecessarily long hospital stays: 1) an overly conservative physiotherapy protocol, and 2) the lack of physiotherapy on weekends. Through implementation of a new protocol and by initiation of weekend physiotherapy, the length of stay for total hip replacement (simple cases) was observed to decrease from 21.3 days in the month prior to implementation of Task Force Recommendations to 14.8 days, 6 months later (Eliasoph &
Hassen, 1986). Similarly, in a study of the impact of the VIP on cataract surgery, a comparison of the "Cost Profile" from the Foothills Hospital to that of another major Canadian teaching hospital and five American hospitals, indicated that the Foothills Hospital was spending approximately $300 more per patient than were the comparison hospitals (Eliasoph & Hassen, 1987). Once again, the major factor accounting for the cost difference was found to be length of stay and, in particular, the pre-operative length of stay. Another major factor found to be contributing to the cost difference was the lack of availability of day care beds, which thereby limited the number of patients that could be treated on an outpatient basis. Therefore, the recommendations formulated and implemented were aimed primarily at increasing the treatment of patients on an outpatient basis (by lifting restrictions on the use of day care beds for cataract surgery) and at decreasing the pre-operative length of stay (by making better use of coordinated pre-op work-ups and the hostel facility). Five months after implementation of the recommendations, it was reported that the proportion of patients treated on an out-patient basis had risen from 45% to 74% and that the per patient length of stay had been reduced from 3.1 to 2.2 days. Moreover, these results, combined with reduced lens and supply costs, are expected to generate annual savings of $350,000 (Eliasoph & Hassen, 1987).

While these results appear promising, they must be interpreted with caution because they are based on a before-after study design with no control group. In addition, the time-period over which the data have been tabulated is relatively short. Consequently, there is no way to determine to what extent the observed reductions reflect simply a regression to the mean
effect, secular trends in length of stay, or an actual reduction. In addition, no attempt has as yet been made to determine whether the benefits, in terms of reduced costs and/or increased quality, outweigh the reportedly substantial costs of implementing and maintaining the system (Winchell, 1985) nor has an attempt been made to compare the VIP to other programs for its cost-effectiveness. Finally, as noted by Rachlis & Fooks (1988), VIP comparisons are based on community norms and "it is not clear how the appropriateness of the norm[s] are evaluated" (p.24).

On a more positive note, however, the VIP's team approach to utilization management, reflected in the work of the multi-disciplinary Task Force, has been well received by the hospital's medical staff (Eliasoph & Hassen, 1987). Moreover, the emphasis on improving the quality of patient care first, with emphasis on cost-reduction coming second, has been an important factor in promoting communication and cooperation among the various personnel involved in the process (Eliasoph & Hassen, 1986; Winchell, 1985). Since these are both critical features to the success of any utilization management process, they ought not to be ignored in considering the merits of the system.

III.2.1.2 Service-Specific Profile Analyses

Hospitals that have begun to implement utilization management programs that assess and respond to utilization problems identified on a service or department level include Saint Michael's Hospital, in Toronto (McDonough & Vaz, 1987), and the Chedoke-McMaster Hospital, in Hamilton (Canadian Comprehensive Auditing Foundation, 1987). Although the two programs differ in detail, with the Saint Michael's program being more centralized and the Chedoke-McMaster program being
highly decentralized, the basic approach is similar; utilization profiles are developed for each department, using information such as length of stay and occupancy rate by service. While the Chedoke-McMaster program relies upon internally developed criteria for problem identification, the Saint Michael's Hospital uses both internal and peer group hospital comparative data in its analyses. Once recommendations for improvement have been made, support staff at Chedoke-McMaster, and special Task Forces at Saint Michael's, are provided to assist the clinical services in implementing corrective action plans. To date, no studies have been found in the literature assessing the extent to which either of these programs have been effective in reducing inappropriate, inefficient, or unnecessary utilization.

III.2.2 Concurrent Review Programs

As noted above, information has been located on only three examples of concurrent review programs implemented in Canadian hospitals. These include the program implemented at the Hospital for Sick Children, in Toronto; the Vi-Care program that is currently being implemented by the Victoria Hospital Society, in Victoria; and the SWITCH Index System which has been implemented at the Peace Arch District Hospital, in White Rock. All of these programs make use of diagnosis-independent utilization criteria. However, only the programs from the Hospital for Sick Children and the Victoria Hospital Society will be described in this section. The SWITCH Index System, which is the subject of the evaluation reported in this thesis, will be described separately.

III.2.2.1 The Hospital for Sick Children

Although the focus of this thesis is on utilization review/management in the general acute care hospital setting, a brief
description of the program implemented at the Hospital for Sick Children is included for completeness. According to information compiled by the Canadian Comprehensive Auditing Foundation (CCAF; 1987), Toronto's Hospital for Sick Children has introduced "a detailed concurrent audit of all patient charts" (p.58). The review is conducted by a team of non-medical "quality assurance screeners" selected by the hospital administration and the medical staff. These "screeners" review each patient's record every three days, assessing the care provided for incidents, utilization problems, and documentation problems. To assist in problem identification, the care is assessed against 22 generic criteria. [Unfortunately, no information on the nature of these criteria nor the utilization problems they address was provided in the CCAF report] When problems are identified that require immediate attention, they are related either to the attending physician, the nursing staff, or the "physician advisor" for the particular clinical service. In addition, summaries of all identified problems are compiled by the screeners for monthly reviews with the physician advisors and for follow-up action. Problems are discussed by the physician advisors at divisional rounds and remedial action strategies are developed. Summarized reports of the problems and the actions taken are then submitted to the hospital's multi-disciplinary quality assurance committee, and, finally, to the Quality Assurance Committee of the board. Although no data have been provided to indicate whether the program has been effective, the "approach is believed to have greater educational impact (resulting in behavioural change) than a retrospective process in which the information may be several weeks or months old by the time it is brought forward for review" (CCAF, 1987, p. 59). In light of the American experience with similarly
detailed and labour-intensive concurrent review programs, however, it
would be interesting to see if this expectation would be met if it could be
compared, for its cost-effectiveness, with other retrospective or, less labour-
intensive concurrent review strategies.

III.2.2.2 The VI-CARE Program

The Greater Victoria Hospital Society in Victoria, British Columbia
has recently begun implementation of a utilization management/quality
assurance program called VI-CARE (Victoria Integrated Care Alternatives
Review & Evaluation). This program, which at present applies only to
patients classified as medical emergencies, is initiated upon a patient's
admission, and continues with periodic concurrent monitoring throughout
each inpatient episode (Nusbaum, 1988). When fully implemented, the VI-
CARE program will be comprised of two components: 1) application of the
ISD-A appropriateness criteria developed by Lamprey and Jacob (1987); and
2) application of the MEDISGRPS severity of illness criteria (Brewster, Jacobs,
& Bradbury, 1984; Brewster, Karlin, et al., 1985). Since the ISD-A review
criteria have already been discussed (Chapter II section II.5.7) no further
elaboration on them will be provided here. The MEDISGRPS criteria were
also mentioned in Chapter II, but they were not described. In brief, the
MEDISGRPS severity of illness index classifies patients into one of five
groups based on the following criteria:

   Group 0 - patients with no key clinical findings;

   Group 1 - patients with minimal findings, indicating a low
            potential for organ failure;

   Group 2 - patients with either acute findings connoting a short
            time course with an unclear potential for organ
failure, or severe findings with high potential for future organ failure;

Group 3 - patients with both acute and severe findings indicating a high potential for imminent organ failure;

Group 4 - patients with critical findings indicating the presence of organ failure.

(Source: Brewster, Jacobs & Bradbury, 1984; Brewster, Karlin, et al., 1985)

To date, only the concurrent ISD-A review component of the VI-CARE program has been implemented. Implementation of the MEDISGRPS system is targeted for the fall of 1988. In the meantime, a retrospective review of patients' charts has been initiated in order to develop norms, within CMG's, for use with this system (Sheps, personal communication, August 1988). Later, when the MEDISGRPS severity of illness index has been implemented, and relevant norms have been established, patients will be assigned a MEDISGRPS code upon admission and again upon discharge. The main objective of this process will be to ensure that patients leave the hospital at a lower level of severity than the level at which they were admitted. Thus, by applying the ISD-A appropriateness criteria and by monitoring changes in the MEDISGRP classifications, it should be possible to identify and correct problems related to both utilization and quality (Nusbaum, 1988). However, since the program is still being phased-in, it will be some time before information on its effectiveness and, ultimately, its cost-effectiveness will be available.

III.3 The SWITCH Index System

In 1984, Peace Arch District Hospital (PADH), a community hospital located in White Rock, British Columbia, developed and implemented one
of the first dynamic utilization management programs to appear in a Canadian acute care hospital setting. Several hospital-specific factors motivated this action: 1) consistently high occupancy rates over a period of years; 2) a concomitant shortage of acute care beds; and 3) the infeasibility of routine patient transfers due to the relative isolation of the hospital from other acute care facilities. These factors coupled with province-wide economic restraint policies and the high costs associated with modern medical technologies, prompted the hospital's administration to begin seeking ways to eliminate inappropriate and unnecessary utilization of the hospital's acute care beds (Frobb, personal communication, October, 1985). The result was the development of a comprehensive utilization management program involving admission and discharge planning and inpatient utilization review. In addition, a close working relationship was established between the Utilization Management Committee and the Medical Audit Committee to ensure that the quality of service provided at the hospital was maintained at an acceptable professional standard.

The cornerstone of the PADH utilization management program, and the component that sets it apart from other utilization management efforts, is a utilization review process called the SWITCH Index System (SWITCH). Similar in intent to the AEP (Gertman & Restuccia, 1981) and the ISD-A (Lamprey & Jacob, 1987), the SWITCH uses explicit criteria to identify hospital days during which no appropriate treatment or diagnostic activity is taking place. In contrast to the numerous criteria employed in the ISD-A system, the SWITCH system employs only six categories of criteria aggregated under the headings Signs, Wind, Intramuscular Injections,
Tubes, Consultant (Active), and Hospice. A description of the specific criteria included under each heading is provided in Table VI.

TABLE VI. The SWITCH Index System Criteria

SIGNS - (Q6H) (Q4H):

- Vital Signs
- Neuro Signs

- This index factor to be marked when a physician specifically orders that vital signs be recorded at 6 hourly intervals or less.
- Implies that nursing observation is required in anticipation of a possible deterioration in patient status.
- Dependent on patterns of practice, the clinical care team may opt to use a Q4H description of this factor.

WIND (Respiratory Therapy):

- Continuous oxygen
- Regularly administered sidestream medications
- Chest physiotherapy

- This index factor describes only regularly administered therapies and does not refer to prn orders.
- If the patient has stabilized at their best pre-hospitalization status, direction should be attained from the head nurse and Convalescence Response Team as to continuing the care of the patient at home with outpatient therapy and homecare nursing.

INTRAMUSCULAR INJECTIONS:

- Antibiotics
- Analgesics

- Administered more frequently then Q8H.
- If therapy is expected to continue for a period longer than 5 days

(Table VI cont...)
Table VI (continued).

and patient status is otherwise stable; consideration should be directed to utilizing Outpatient Services or Home Care Nursing for administration of medications.
- Does not describe administration of subcutaneous insulin.

TUBES:

- I.V. Therapy (medications and fluids)
- Nasogastric Tubes
- Foley catheters
- Epidurals
- Chest Tubes/Endotracheal Tubes
- Drains to continuous or intermittent suction

- In most cases, the use of invasive tubes and catheters is self-limited. If use is anticipated for greater than 10 days, the charge nurse should be informed.
- Does not include TKVO I. V. Lines.

CONSULTANT (Active):

- Implies daily visits by a consultant with changing orders, indicating a change in the pattern of care.
- Implies that the changing status of patient condition necessitates daily visits by consultants for further investigations and/or treatment.

HOSPICE:

- Indicates the palliative care patient.
- Implies that the patient is unlikely to recover to a status permitting discharge home or to a Long Term Care Facility.
- Denotes involvement of Palliative Care Team.

(Source: Peace Arch District Hospital, 1985)
Potential cases of inappropriate or unnecessary bed use are identified, on a daily basis, by means of an On-Index/Off-Index classification system. Patients classified as Off-Index are those who fail to meet at least one of the SWITCH criteria. The process of On/Off-Index classification is initiated with the completion of the Kardex Acuity Index (See Figure 2) at the time of the patient's admission. Thereafter, during each nursing shift, the Acuity Index is updated to reflect any changes in the patient's status. This function is usually performed by the supervising R.N. although it may be delegated to the Ward Clerk under the direction of the supervising R.N. During each day shift, the Head Nurse produces a Nominal Roll SWITCH which is simply a listing of all the patients on the ward indicating their On/Off Index status. This Nominal Roll is reviewed by the evening Charge Nurse, for accuracy and updating, and later is finalized by the Night Supervisor. The Night supervisor is also responsible for compiling an OFF-Index census list. This list consists of all patients designated as Off-Index at the time of census collection. In addition, at the time the census is taken, each Off-Index day is assigned to the "most responsible" service; i.e., the service most responsible for the patient's Off-Index status. This information is taken from the Acuity Index (see Figure 2) and indicates whether, at the time it was made, the off-index assignment was due primarily to a delay in physician communication, delays in in-house procedures, or delays in outside services. At 0800 every day, the Head Nurse, the Team Leader, the Director of Social Services and the physician utilization review coordinator review the Off-Index patients for accuracy and initiate the distribution of the Off-Index Census list to hospital service departments for their attention and, if appropriate, corrective action. As well, the Team Leader initiates discussion with
Figure 2. Kardex Acuity Index

<table>
<thead>
<tr>
<th>ACUITY INDEX</th>
<th>#</th>
<th>SIGNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>M.D. COMMUNICATION</td>
<td></td>
<td>Vital Signs Q6H</td>
</tr>
<tr>
<td>Waiting Direction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IN-HOUSE SERVICES:</td>
<td></td>
<td>Nursing Observation</td>
</tr>
<tr>
<td>Radiology</td>
<td></td>
<td>(Indicative of possible deterioration)</td>
</tr>
<tr>
<td>Laboratory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Out-Patients</td>
<td></td>
<td>WIND THERAPY</td>
</tr>
<tr>
<td>O.R.</td>
<td></td>
<td>O2 Sidestream Meds</td>
</tr>
<tr>
<td>Social Services</td>
<td></td>
<td>Chest Physiotherapy</td>
</tr>
<tr>
<td>Convalescence</td>
<td></td>
<td>I.M. MEDICATIONS</td>
</tr>
<tr>
<td>Ortho</td>
<td></td>
<td>TUBES</td>
</tr>
<tr>
<td>Neuro</td>
<td></td>
<td>I.V. or I.V. Meds</td>
</tr>
<tr>
<td>Obstetrical</td>
<td></td>
<td>N.G./Foley/Epidural</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OUTSIDE SERVICES:</td>
<td></td>
<td>CONSULT (Active)</td>
</tr>
<tr>
<td>Diagnostic</td>
<td></td>
<td>Daily Visits</td>
</tr>
<tr>
<td>B.C.C.I</td>
<td></td>
<td>Changing Orders</td>
</tr>
<tr>
<td>Holy Family</td>
<td></td>
<td>HOSPICE</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL # OF DAYS OFF INDEX:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Source: Peace Arch District Hospital, 1985)

physicians regarding discharge plans. An exception occurs with patients designated as Off-Index due to convalescence. These patients are referred, on the first day of being so designated, to a Convalescence Response Team (CRT) for initiation of discharge planning. Finally, when a patient is discharged, the Health Records Department abstracts information concerning the number of days off-index, the specific reasons for each off-index day, and the CRT assessment are entered into the hospital's computer system. This information is used to compile regular historical summaries so that trends in Off-Index use can be identified.
As with the American AEP, one of the most interesting characteristics of the SWITCH is that it focuses the attention of the medical, nursing, administrative, and other professional staff on the problem of identifying what is important about acute care hospital beds. In addition, it provides the hospital with a mechanism for increasing the involvement of the medical staff in the management of the hospital's resources. As with any utilization management program, the success of the SWITCH is dependent on the involvement and cooperation of the medical staff. A previous attempt to implement a utilization review procedure that employed length of stay target dates had met with failure because it was viewed by the medical staff as punitive and arbitrary (Frobb, personal communication, October, 1985). The SWITCH, on the other hand, provides information that is more meaningful and acceptable to clinical staff than the statistical and normative approach of length of stay based utilization review. Finally, unlike some of the other methods of concurrent or in-depth review (e.g. the VIP), the SWITCH is relatively easy to implement and maintain.

III.4 Study Purpose and Rationale

The general purpose of the present study was to provide preliminary data on the effectiveness of the PADH utilization management program. Specifically, the study was designed to determine whether the PADH system of utilization management had been effective, over time and in comparison to a control group of community hospitals, in reducing the length of adult, acute-care hospital stays. The decision to use length of stay as the outcome variable of interest was based on two factors:

1) since the PADH program was designed to eliminate inappropriate and unnecessary days of hospitalization, an observable outcome, assuming
that changes in case mix have been controlled, should be a reduction in length of stay, and

2) since comparison hospitals were to be included in the evaluation, a measure of effectiveness was needed that would be obtainable across all participating hospitals.

Trying to determine whether the SWITCH had an effect on length of stay is, however, a complicated problem. As noted by Lave & Leinhardt (1976a), "over a period of time, exogenous events unrelated to the review program can influence length of stay. Changes in medical technique, therapy, and hospital staffing may each play a part" (p.961), as may variations in case-mix. The PADH was thus compared to a group of peer hospitals on the assumption that secular events would be likely to affect the length of stay patterns of both the PADH and the peer group in a similar manner. The changing case-mix problem was dealt with in two ways, as will be described in Chapter IV. Thus, although the present study is exploratory, it is important because it represents the first attempt to incorporate both a "control" group of hospitals and an adjustment for case-mix changes over time, in evaluating the effects of a utilization management program implemented in a Canadian hospital setting.

III.5 Hypotheses

1) Assuming that secular trends have affected both the PADH and non-SWITCH hospitals in similar ways, then it is expected that during the period prior to the implementation of the SWITCH, average length of stay trends for the PADH will not differ significantly from those of the Non-SWITCH hospitals.
2) In the period following the introduction of SWITCH, it is expected that the average length of stay for the PADH will fall relative to the Non-SWITCH hospitals.
CHAPTER IV

Methods

IV.1 Study Design

A quasi-experimental time-series design with a control group (Campbell & Stanley, 1969) was used to examine trends in the average length of hospital stay at the PADH. Twelve periods of four months each were studied: eight periods before the implementation of the SWITCH (February, 1982 through September, 1984), one period during which the SWITCH was implemented (October, 1984 through January, 1985) and three periods during which the SWITCH was fully operational (February, 1985 through December, 1985). Patients under the age of 18 and patients who died while in the hospital were excluded from the study.

Although the inclusion of a control group provides a baseline for comparing secular changes in average length of stay, secular trends represent only one potential explanation of any observed changes in average length of stay over time. The clinical characteristics of patients and hospital characteristics such as teaching status, occupancy rate, and bed capacity have also been found to affect length of stay (Lave & Leinhardt, 1976a; 1976b). Hence, in designing the study, an effort was made to select hospitals comparable to the PADH in terms of bed size and teaching status (i.e., non-teaching). Potential effects due to hospital occupancy were taken into account in the analytic phase of the investigation. Finally, since the basic unit of observation for this evaluation is the hospital, not a patient case, differences in patient characteristics were dealt with by using the CMG

1 An exception is period 12 which is based on the three month period October, 1985 to December, 1985.
patient classification system to ascertain, and to adjust for, differences in, hospital case mix.

IV.2 Sample

a) The SWITCH Hospital. Because the SWITCH has only been implemented in one hospital, namely the PADH, this one hospital alone constitutes what is usually referred to as the intervention group. The PADH is a public, general, community hospital located in White Rock, B.C. Information concerning the hospital’s bed size and occupancy is provided in Table VII.

b) The Comparison (Non-SWITCH) Group. In order to provide comparability to the PADH, at least in terms of bed size and non-teaching status, the hospitals included in the Non-SWITCH group were initially identified according to the following criteria:

i) A member of the B.C. Ministry of Health's defined Peer Group for the PADH (MOH Peer Group #15)\(^1\).

ii) A member of the B.C. Health Association's (BCHA) defined Peer Group for the PADH (BCHA Peer Group #12).\(^2\)

iii) A member of HMRI since at least January 1, 1982.

---

1 MOH Peer Group 15 includes the following hospitals: Cowichan District Hospital, Duncan; Maple Ridge Hospital, Maple Ridge; Mount Saint Joseph Hospital, Vancouver; Peace Arch District Hospital, White Rock; Saint Joseph's General Hospital, Comox; Saint Mary's Hospital, New Westminster; Saint Vincent's Hospital, Vancouver; Trail Regional Hospital, Trail.

2 BCHA Peer Group 12 includes the following hospitals: Chilliwack General Hospital, Chilliwack; Cowichan District Hospital, Duncan; Langley Memorial Hospital, Langley; Maple Ridge Hospital, Maple Ridge; MSA General Hospital Abbotsford; Mount Saint Joseph Hospital, Vancouver; Peace Arch District Hospital, White Rock; Penticton Regional Hospital, Penticton; Richmond General Hospital, Richmond; Saint Mary's Hospital, New Westminster; Saint Vincent's Hospital, Vancouver; Trail Regional Hospital, Trail; Vernon Jubilee Hospital, Vernon.
On the basis of these criteria, a total of five hospitals were determined to be eligible for inclusion in the study. In order to obtain the data for the evaluation, however, written permission had to be obtained from each of the hospitals. Four of the five eligible hospitals agreed to participate and provided the necessary "Release of Data Forms" (See Appendix E). Thus, when the study began, four hospitals comprised the Non-SWITCH group. Subsequently, however, one of these four hospitals failed to provide

---

Table VII. Characteristics of PADH and Non-SWITCH Hospitals: 1982-1985

<table>
<thead>
<tr>
<th></th>
<th>PADH</th>
<th>Non-SWITCH* Hospitals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Hospitals</strong></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td><strong>Number of Acute Care Beds per quarter</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>per Hospital</td>
<td>Mean</td>
<td>114.9</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>112.0</td>
</tr>
<tr>
<td></td>
<td>St. Dev.</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>112.0</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>129.0</td>
</tr>
<tr>
<td><strong>Quarterly Occupancy Rates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>per Hospital</td>
<td>Mean</td>
<td>89.6 %</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>91.0</td>
</tr>
<tr>
<td></td>
<td>St. Dev.</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>78.0</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>98.0</td>
</tr>
</tbody>
</table>

* Statistics reflect the average across hospitals.
** Occupancy rates were computed on the basis of beds available for acute care use (i.e., set up and staffed) rather than on the basis of each hospital's rated bed complement.
necessary hospital data and was dropped from the study. The final Non-SWITCH group, therefore, consisted of three PADH peer-group-member hospitals. Summary statistics on the bed size, and occupancy rates of the Non-SWITCH group of hospitals, are provided in Table VII.

It is unknown whether any systematically varying characteristic(s) of the excluded hospitals contributed to their non-participation. If so, then it is possible that the length of stay results may have been affected by a self-selection bias. On the basis of statistics reported in the Canadian Hospital Directory (Canadian Hospital Association, 1982; 1983; 1984; 1985) it was estimated that, on average, the excluded hospitals had a higher number of acute care beds (Mean per Quarter = 180) than did the three Non-SWITCH hospitals (Mean per Quarter = 121.3). However, in the absence of length of stay and occupancy data, it is impossible to determine whether the two excluded hospitals differed sufficiently from the others to affect the average length of stay patterns observed in the participating comparison group of hospitals.

IV.3 Data

Data for the study were obtained from two sources: HMRI and each participating hospital. To obtain patient data, requests were made to each hospital for release of HMRI records pertaining to all of the hospital's discharges occurring between January, 1982 and December, 1985. A total of 101,858 patient records were received from HMRI. From this database, 21,512 records were removed because they pertained to the hospital that was dropped from the study. In addition, because of the decision to group the patient data into quarterly intervals, beginning February, 1982, the data from January, 1982 were removed from the database. These two deletions left a
total of 78,674 patient records. Next, since the data-base also included patients under the age of 18 and patients who had died while in the hospital, both of these groups of records had to be removed. This left a final data base of 59,259 patient records; 14,762 from PADH and 44,497 from the Non-SWITCH group of hospitals. The patient data obtained from the HMRI records included: age, sex, length of stay, admission status (emergent, urgent, elective), discharge status (dead or still born, discharged, signed out), and CMG designation. Descriptive statistics on each of these variables (except discharge status, since all deaths were excluded) for the PADH and the group of Non-SWITCH hospitals are provided in Table VIII.

Hospital-specific data (eg., monthly occupancy rates, and bed size) were obtained from each of the participating hospitals (See Table VII). In addition, each hospital was asked to provide an indication of their beds per 1000 population. In 1982, the PADH had a bed per 1000 population ratio of 2.6 which, by 1985 had been reduced to 1.6 beds per 1000 population. Unfortunately, comparative ratios for the Non-SWITCH group of hospitals could not be computed as two of the three hospitals in this group were unable to provide the information.

Finally, each of the Non-SWITCH hospitals was asked to provide a description of its utilization review activities during the period of the study. The purpose of this request was to determine: 1) whether any of the comparison hospitals had utilization review/management programs that were similar in nature and intensity to the SWITCH program; and 2) whether any changes in the review procedures followed by the comparison hospitals had occurred at or around the time the SWITCH was implemented at the PADH. The information obtained indicated that, while
### TABLE VIII. Patient Characteristics of PADH and Non-SWITCH Hospitals: 1982-1985

<table>
<thead>
<tr>
<th></th>
<th>PADH</th>
<th>Non-SWITCH Hospitals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Hospitals</strong></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total Number of Adult Discharges</strong></td>
<td>14,762</td>
<td>44,497</td>
</tr>
<tr>
<td><strong>Number of Discharges per quarter</strong></td>
<td>Mean: 1,230.2</td>
<td>1,236.0**</td>
</tr>
<tr>
<td></td>
<td>Median: 1,238.0</td>
<td>1,276.0</td>
</tr>
<tr>
<td></td>
<td>St. Dev.: 130.4</td>
<td>103.0</td>
</tr>
<tr>
<td></td>
<td>Minimum: 855.0</td>
<td>976.0</td>
</tr>
<tr>
<td></td>
<td>Maximum: 1359.0</td>
<td>1331.0</td>
</tr>
<tr>
<td><strong>Number (%) of Patients by Age group</strong></td>
<td>18-29: 2,680 (18.2%)</td>
<td>10,632 (23.9%)</td>
</tr>
<tr>
<td></td>
<td>30-49: 2,920 (19.8%)</td>
<td>11,193 (25.2%)</td>
</tr>
<tr>
<td></td>
<td>50-69: 3,542 (24.0%)</td>
<td>11,759 (26.4%)</td>
</tr>
<tr>
<td></td>
<td>70+: 5,620 (38.1%)</td>
<td>10,913 (24.5%)</td>
</tr>
<tr>
<td><strong>Number (%) of Patients by Sex</strong></td>
<td>Male: 5,767 (39.1%)</td>
<td>17,654 (39.7%)</td>
</tr>
<tr>
<td></td>
<td>Female: 8,995 (60.9%)</td>
<td>26,838 (60.3%)</td>
</tr>
<tr>
<td><strong>Number (%) of Patients by Admission Status</strong></td>
<td>Emergent: 6,302 (42.7%)</td>
<td>16,707 (37.5%)</td>
</tr>
<tr>
<td></td>
<td>Urgent: 3,224 (21.8%)</td>
<td>10,586 (23.8%)</td>
</tr>
<tr>
<td></td>
<td>Elective: 5,236 (35.5%)</td>
<td>17,201 (38.7%)</td>
</tr>
<tr>
<td></td>
<td>Unknown: 0</td>
<td>3 ( 0.0%)</td>
</tr>
</tbody>
</table>

* Excluding Deaths
** Average across hospitals.
each of the comparison hospitals had engaged in some form of utilization review, during the study period, the SWITCH program was the most systematic and the only one initiated on admission, and activated on a daily basis, throughout each inpatient episode. Moreover, the SWITCH program was the only one to incorporate a documented set of "appropriate day-of-care use" criteria. In contrast, the procedures followed by the Non-SWITCH hospitals were typically retrospective and tended to focus primarily on assessing the need for continued hospitalization of long-stay cases. One of the three hospitals had also incorporated a system of concurrent discharge planning, using length of stay target dates defined by diagnosis, into its utilization review process. Thus, although there was some variation in the intensity and rigour with which utilization review procedures were applied in the comparison hospitals, none of these procedures approached the stringency and frequency of application typical of the SWITCH program. Finally, with the exception of one hospital which, in June 1984, changed its definition of a long-stay case from 10 to 7 days and changed its meeting schedule from weekly to monthly, no other changes in the utilization procedures employed by the Non-SWITCH hospitals during the study period were reported.

IV.4 Length of Stay Measures

Three measures were constructed, using a model developed by Lave & Leinhardt (1976a), to describe the length of stay behaviour of the two hospital groups over the four-year study period. The first, the unadjusted average length of stay (UALS), was computed simply as the mean length of stay for each hospital in each time period. However, because the UALS does not control for differences in the average case mix either over time or across
hospitals, comparisons based on this measure could be misleading. Therefore, to control for inter-hospital variations in case-mix, a standard set of CMG's was defined for use in computing the remaining two length of stay measures. This standard set of CMG's consisted of all the CMG's that appeared in the PADH data set during period 1, except CMG 900 (any unrelated OR procedure) and CMG 901 (missing data necessary for CMG assignment). Across all time periods, the standard set of CMG's accounted for 90% of the PADH patient discharges and 88% of the Non-SWITCH hospitals' discharges.

Next, to control for intra-hospital changes in case-mix over time, the second length of stay measure, an adjusted length of stay (ALS), was constructed. To compute the ALS, the frequencies associated with the standard set of CMG's in time period 1 were determined and were used to weight each CMG-specific average length of stay in each subsequent period, for each hospital. Thus, the ALS reflects the average length of stay each hospital would have had in some period j, if it had the same case mix as the first time period but the length of stay of period j (Lave & Leinhardt, 1976a). Once calculated, the ALS was then used to compute the actual dependent variable used in the length of stay analyses. This variable, the adjusted length of stay ratio (ALSR) was computed as the ratio of the ALS in period j divided by the ALS in period 1. The ALSR was computed for each hospital.

1 For example, if a hospital had 50 cases of CMG 14 in period 1 with an average length of stay of 5.0 days and 75 cases of CMG 75 with an average length of stay of 6.2 days, while in period 2 it had 60 cases of CMG 14 with an average stay of 4.5 days and 50 cases of CMG 75 with an average stay of 6.0 days, then the actual average length of stay in period 1 is \((50 \times 5) + (75 \times 6.2))/125 = 5.72\) and in period j is \((60 \times 4.5) + (50 \times 6))/110 = 5.18\). The adjusted length of stay for period j, however, would be \((50 \times 4.5) + (75 \times 6))/125 = 5.40\).
in each time period and reflects the change in length of stay over time, relative to the length of stay in period 1.

IV.4 Data Analyses

To determine whether the SWITCH was effective in reducing the average length of stay of the PADH, time series analyses were performed using graphical and ordinary least squares regression techniques (Gillings, Makuc, & Siegel, 1976; Neter & Wasserman, 1974; Veney & Kaluzny, 1984). The graphical procedures were used to examine general trends in length of stay over time and to provide a preliminary indication of the appropriateness of applying linear regression techniques. The regression analyses were used to assess the comparability of the PADH and the Non-SWITCH group of hospitals in the pre-intervention period, and to determine whether, in the post-intervention period, there was any evidence of a reduction in the PADH length of stay which could be attributed to the introduction of the SWITCH program.

In general terms, the application of regression techniques to time series data (interrupted by a program intervention) are used to detect the following trends, any one of which, if observed relative to the pre-intervention period and to the comparison group, would be suggestive of a program effect:

1) A change in the slope of the regression line following program implementation. (See Figure 3a);

2) No change in slope, but a change in the height of the response curve (i.e., the intercept) following program implementation. (See Figure 3b); or
Figure 3. Hypothetical Trends Detectable with Time Series Regression Analysis

a

b

x = Program Intervention
O1-O4 = Time Series Observations

c
3) A change in both the slope and the height of the response curve following program implementation. (See Figure 3c).

In order to determine whether any of these changes have occurred, regression equations are defined to include all of the major dependent and independent variables under study plus a number of special indicator variables designed to permit testing of hypotheses relevant to changes in slope and intercept. The specific equations used in the present study - to ascertain whether there had been any reduction, attributable to the implementation of the SWITCH, in the length of stay at the PADH - are defined in Chapter V.

All of the analyses reported in this thesis were performed using the Statistical Package for the Social Sciences (SPSSX). In addition, although primary interest was in the detection of a reduction in length of stay attributable to the SWITCH, because of the exploratory nature of the study, all tests of the regression coefficients were performed using two-tailed tests of significance.
CHAPTER V

Results

V.1 Graphical Analyses

V.1.1 Unadjusted Average Length of Stay

Between Period 1 and Period 12, the unadjusted average length of stay of PADH patients fell from 8.74 to 6.45 days, for a total decrease of 26 percent. For patients in the non-SWITCH hospitals the unadjusted average length of stay also declined, falling from 8.34 to 7.97 days, for a total decrease of 4 percent. Thus, as shown in Figure 4, the unadjusted average length of stay for the PADH started at a slightly higher level and decreased more than that of the Non-SWITCH hospitals. However, as these numbers do not control for differences in the average case mix of the hospital groups, they may be misleading.

V.1.2 Case-Mix Adjusted Average Length of Stay

After performing the necessary procedures to identify a standard set of CMG's and to compute the ALS, it was determined that the case-mix adjusted average length of stay for the PADH fell from 8.63 days in Period 1 to 5.22 days in Period 12, a decline of 40 percent. For the Non-SWITCH group of hospitals the case-mix adjusted length of stay, averaged across the hospitals, dropped from 8.06 days in Period 1 to 7.36 days in Period 12, for a total decrease of 9 percent. Thus, as shown in Figure 5, after adjusting for differences in case-mix, the average length of stay of PADH patients still started somewhat higher and fell further than the average case-mix adjusted length of stay of the Non-SWITCH group of hospitals. To examine the change in average length of stay over time, relative to the length of stay in Period 1, the ALSR was computed and plotted, as shown in Figure 6. From
Figure 4. **Unadjusted Average Length of Stay by Hospital Group**

![Graph 1](image1)

Figure 5. **Case-Mix Adjusted Length of Stay by Hospital Group**

![Graph 2](image2)
this figure it can be seen that, while there appears to have been no systematic change in the ALSR over time for the Non-SWITCH group of hospitals, the ALSR of the PADH has clearly declined, in what could appropriately be described as a linear pattern, over the time period studied. Because of the variability in the ALSR, however, it is difficult to tell whether the observed decrease was more, or less, pronounced in the post-intervention period. Moreover, the graphs do not take into account the possible effect of changing occupancy rates on the observed length of stay trends. Consequently, in the next phase of the analysis, regression equations
were constructed to control for the effects of occupancy and to test for changes in ALSR trends following implementation of the SWITCH.

V.2 Regression Analyses

To determine general linear trends, and to ascertain whether there had been a change in the slope of the regression line following implementation of the SWITCH program, the following equation was estimated for each of the PADH and Non-SWITCH hospitals. Although a significant regression relation was not expected in the latter case, the analysis was included for comparative purposes.

\[
(1) \quad \text{ALSR}_{it} = b_0 + b_1 \text{PERIOD} + b_2 \text{TS} + b_3 \text{OCC}_{it} + e_{it}
\]

In equation (1), \( \text{ALSR}_{it} \) is the case-mix adjusted length of stay ratio for hospital \( i \) in period \( t \). \( \text{PERIOD} \) is a linear time variable for the quarterly time intervals and assumes the values of 1 through 12. The coefficient of \( \text{PERIOD} \) will indicate what the average quarterly rate of decline was over the period of time studied. \( \text{TS} \) is a linear spline (Lave & Leinhardt, 1976a), or indicator variable (Neter & Wasserman, 1974), which takes on the value of 0 for Periods 1 through 9 and the values 1 to 3 for Periods 10, 11, and 12, respectively. Essentially, \( \text{TS} \) provides the information needed to construct two regression lines: one in the period prior to full implementation of the SWITCH, and one in the period after implementation. By including \( \text{TS} \) in the regression equation, it becomes possible to test whether the slope of the regression line changed following implementation of the SWITCH. More specifically, the coefficient of \( \text{TS} \) will indicate whether there was a change in the rate of decline after the introduction of the SWITCH. A positive coefficient will suggest that the ALSR began to increase following implementation of SWITCH; a non-significant coefficient will suggest that
there was no change in slope attributable to SWITCH; and a negative
coefficient will suggest that, following SWITCH, the ALSR started to decline
at a more rapid rate than previously. The final independent variable
included in equation (1), \( \text{OCC}_{it} \), is the occupancy rate of hospital \( i \) in period \( t \)
and is included in order to control for length of stay effects associated with
differences in hospital occupancy.

The results of computing equation (1) are presented in Table IX. As
expected, a significant regression relation was observed for the PADH data
\( (F=15.818, p < .001) \), with the model accounting for 80% of the variability in
ALSR. In contrast, but also as expected, no significant regression relation
was observed for the Non-SWITCH hospitals \( (F=2.000, p > .05) \). Moreover,
for this latter group, the defined regression model accounted for only 8% of
the variability in ALSR. While this lack of fit suggests that, for this group,
there was no linear decline (or increase) in ALSR over the time period
studied, it does not preclude the possibility that some other type of
relationship might exist. An examination of Figure 6, however, suggests
that the most likely explanation for the observed lack of fit is that there has
been no systematic change in ALSR over time for the Non-SWITCH group
of hospitals.

Similarly, a comparison of the coefficients of PERIOD for the PADH
and Non-SWITCH hospitals, reveals that while the ALSR decreased by .05
per period in the PADH, no significant rate of decline was observed for the
Non-SWITCH hospitals. [In fact, the only significant relationship identified
for the Non-SWITCH group, was that of a .005 increase in ALSR with each
unit increase in OCC rate.] Next, to determine whether there had been a
Table IX. Regression Coefficients\(^1\) using ALSR as the Dependent Variable: PADH and Non-SWITCH Hospitals—Standard Set of CMG's

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>PADH Equations</th>
<th></th>
<th>Non-SWITCH Equations</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td></td>
<td>ALSR</td>
<td>ALSR</td>
<td>ALSR</td>
<td>ALSR</td>
</tr>
<tr>
<td>Constant</td>
<td>-.263</td>
<td>-.208</td>
<td>.527</td>
<td>.529</td>
</tr>
<tr>
<td>Period</td>
<td>-.049</td>
<td>-.050</td>
<td>-.005</td>
<td>-.005</td>
</tr>
<tr>
<td></td>
<td>(-4.650)**</td>
<td>(-5.316)***</td>
<td>(-1.168)</td>
<td>(-1.124)</td>
</tr>
<tr>
<td>TS</td>
<td>.033</td>
<td>.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.527)</td>
<td>(.532)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DTV</td>
<td></td>
<td>.089</td>
<td>.019</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.019)</td>
<td>(.495)</td>
<td></td>
</tr>
<tr>
<td>OCC</td>
<td>.015</td>
<td>.014</td>
<td>.005</td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td>(2.230)</td>
<td>(2.367)*</td>
<td>(2.063)*</td>
<td>(2.043)*</td>
</tr>
<tr>
<td>E-Ratio</td>
<td>15.818***</td>
<td>18.935***</td>
<td>2.000</td>
<td>1.981</td>
</tr>
<tr>
<td>R(^2)-Adjusted</td>
<td>.80</td>
<td>.83</td>
<td>.08</td>
<td>.08</td>
</tr>
</tbody>
</table>

\(^1\) Ordinary least-squares estimation: t-statistics appear in parentheses (* indicates p < .05; ** indicates p < .01; *** indicates p < .001). The number of observations for PADH is 12; 1 hospital and 12 time periods. The number of observations for the Non-SWITCH group is 36; 3 hospitals and 12 time periods.
change in linear trends, (i.e., a change in slope), following implementation of the SWITCH, the coefficients of TS were examined. In accordance with expectations, the coefficient of TS for the Non-SWITCH group was not significant. Contrary to expectations, however, the coefficient of TS computed for the PADH was likewise non-significant. Thus, it would appear that the general rate of decline in ALSR observed in the pre-intervention period continued into the post-intervention period. More specifically, on the basis of the observed coefficient of TS, there is insufficient evidence to suggest that, following implementation of the SWITCH, the ALSR declined more rapidly. The absence of a change in the slope of the regression line, however, does not preclude the possibility of a change in the height of the response curve. Therefore, to test for this possibility, the following equation was computed for each hospital group:

\[(2) \text{ALSR}_{it} = a_0 + a_1 \text{PERIOD} + a_2 \text{DTV} + a_3 \text{OCC}_{it} + e_{it}\]

Equation (2) differs from (1) only in the way that the time variable is entered. TS has been dropped from the equation, since no evidence was obtained to suggest a post-intervention change in slope, and DTV has been added. DTV is a dummy variable that takes on a value of 0 for Periods 1 through 9, and a value of 1 otherwise. In contrast to the change in slope measured by the TS variable in equation (1), however, equation (2) is designed to determine only if there has been a discontinuity (i.e., a change in height) in the regression line following program implementation. Thus, a positive DTV coefficient will indicate whether, on average, the ALSR was higher in the post-intervention period than in the pre-intervention period; a non-significant coefficient will indicate that there was no change in the height of the response curve following program implementation; and a
negative coefficient will indicate that, on average, the ALSR was lower in the post-intervention period.

The results of computing equation (2) are also shown in Table IX. Once again, the model provided a good fit to the PADH data ($F=18.93$, $p <.001$), accounting for 83% of the variability in ALSR. Furthermore, as expected, no evidence of a significant regression relation was observed for the Non-SWITCH group. Similarly, as indicated by the coefficient of PERIOD, the same trends showed up with respect to the slope of the regression line. Finally, an examination of the coefficient of DTV indicated that there had been no significant change in the height of the response curve, following implementation of the SWITCH, for either the PADH or the Non-SWITCH group of hospitals. Thus, on the basis of equations (1) and (2), there is little evidence to suggest that, for the PADH, there has been a reduction in length of stay attributable to the introduction of the SWITCH program. Before such a conclusion can be drawn, however, certain factors relevant to the aptness of the regression models must be taken into account.

To determine the aptness of the regression model, for each hospital group, residual analyses were performed in order to detect problems pertaining to non-normality, autocorrelation, and heterogeneity among the error terms. As in any application of regression analysis, certain assumptions must be met in order for the application of time series regression to be appropriate. Specifically, it is assumed that the error terms are independent, normally distributed random variables with equal
variance across all levels of the independent variable\textsuperscript{1}. To determine whether these assumptions were reasonable in the present case, residual plots were constructed and examined for patterns reflecting a violation of the assumptions of normality and variance homogeneity. In addition, because non-independence of the error terms can be a problem with time series data, the Durbin-Watson test for autocorrelation (Neter & Wasserman, 1974) was performed. The results of these analyses indicated that, for the PADH, none of these potential problems were important issues. The residual plots provided no indication that the assumptions of normality and homogeneity had been violated, and the computed Durbin-Watson statistics were 2.55 and 2.61, for the two equations respectively. Both of these values exceeded the critical value of 1.75 and, therefore, reflected no autocorrelation among the error terms. Similarly, for the Non-SWITCH group, there was no evidence of autocorrelation (Durbin-Watson = 2.20 and 2.55, respectively; exceeding critical value of 1.65) among the error terms in either equation. However, the residual plots provided some evidence to suggest that variance heterogeneity may have been a problem. This observation was not unexpected since it was clear from Figure 6 that, for this group, the ALSR had been more variable in the earlier periods of the study than in the later periods. However, no attempt was made to transform the data, in order to correct this problem, because such action would have rendered the models constructed for this group non-comparable to those constructed for the PADH. In addition, given the apparent lack of a

\textsuperscript{1} In the case of a multiple regression model, the distributional properties attributed to the error terms for a given level of the independent variable assume that all other variables in the model are held constant.
systematic relation between ALSR and time, for the Non-SWITCH group, such a refinement was considered unwarranted.

The next potential problem considered was the nature of the relationship among the independent variables, in particular between PERIOD and OCC. In the application of multiple regression techniques, it is desirable to include, as independent variables, only those variables that are highly correlated with the dependent variable and uncorrelated with each other. Although in most research endeavors it is unlikely that all of the independent variables will be completely unrelated to each other, the inclusion of highly inter-correlated independent variables can cause a problem referred to as multicollinearity (Neter & Wasserman, 1974). In essence, when there is a high degree of association among the independent variables, computed regression coefficients become unstable and provide only imprecise information regarding the relative importance of each independent variable. Thus, on the basis of the computed regression model, the "estimated regression coefficients individually may not be statistically significant even though a definite statistical relation exists between the dependent variable and the set of independent variables" (Neter & Wasserman, 1974, p. 339). Accordingly, an examination of the correlation coefficients computed among the independent variables included in the present study, suggested that, for the PADH, the regression coefficients presented in Table IX may have been affected by multicollinearity among the independent variables. Specifically, a high positive correlation observed between OCC and PERIOD (r = +.881; p < .001; two-tailed) may have caused problems. A significant negative correlation between OCC and ALSR (r = -.64; p = .024; two-tailed) suggested that OCC was an important explanatory
variable in the observed length of stay trends. In the computation of equation (1), however, the coefficient of OCC failed to reach significance. Given these apparently contradictory findings, consideration was given to the possibility that the multicollinearity between OCC and Period was masking an OCC effect. Thus, equations (1) and (2) were re-estimated without OCC and then equation (1) was re-calculated, dropping TS and using OCC as a proxy for PERIOD. For comparative purposes, these same equations were computed for the Non-SWITCH hospitals, although multicollinearity between OCC and PERIOD was not an issue for this group. Furthermore, for this group of hospitals, OCC was the only variable found to be correlated with the ALSR. However, in contrast to the PADH, but in accordance with the regression results, this relationship was positive ($r = +.340; p = .042; \text{two-tailed}$).

The results presented in Table X confirmed the previous findings that, for the PADH, the ALSR decreased with time (.03 per period, with OCC excluded), and further that, when no time variables were included in the model, the ALSR decreased by .014 per each unit increase in OCC. Moreover, an examination of the goodness of fit measures (i.e., F-Ratio and $R^2$-adjusted) indicated that, when only OCC was used in the equation, the regression equation retained its significance and the model accounted for 36% of the variability in ALSR. Since, at the PADH, OCC tended to increase over the study period, these findings provide support for the assumption that at least some portion of the overall decline in ALSR observed for the PADH was associated with the increase in the hospital's occupancy rate over time. For the Non-SWITCH hospitals, the results presented in Table X similarly concur with those presented in Table IX; i.e., there was no
Table X. **Regression Coefficients**\(^1\) using ALSR as **Dependent Variable:**
**PADH and Non-SWITCH Hospitals - OCC Removed - Standard Set of CMG's**

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>PADH Equations</th>
<th>Non-SWITCH Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(No OCC)</td>
<td>(OCC Only)</td>
</tr>
<tr>
<td></td>
<td>(No OCC)</td>
<td>(OCC Only)</td>
</tr>
<tr>
<td></td>
<td>(OCC Only)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>.920</td>
<td>.961</td>
</tr>
<tr>
<td></td>
<td>.946</td>
<td>.961</td>
</tr>
<tr>
<td></td>
<td>2.046</td>
<td>.493</td>
</tr>
<tr>
<td>Period</td>
<td>-.029</td>
<td>-.006</td>
</tr>
<tr>
<td></td>
<td>(-4.241)(***)</td>
<td>(-1.234)</td>
</tr>
<tr>
<td>TS</td>
<td>.019</td>
<td>.013</td>
</tr>
<tr>
<td></td>
<td>(.790)</td>
<td>(.747)</td>
</tr>
<tr>
<td>DTV</td>
<td>.070</td>
<td>.031</td>
</tr>
<tr>
<td></td>
<td>(1.314)</td>
<td>(.771)</td>
</tr>
<tr>
<td>OCC</td>
<td>-.014</td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td>(-2.658)(**)</td>
<td>(2.110)(*)</td>
</tr>
<tr>
<td>F-Ratio</td>
<td>14.736(**)</td>
<td>16.942</td>
</tr>
<tr>
<td></td>
<td>7.067(*)</td>
<td>.788</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.807</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.452(*)</td>
</tr>
<tr>
<td>(R^2)-Adjusted</td>
<td>.71</td>
<td>.74</td>
</tr>
<tr>
<td></td>
<td>.36</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>.09</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) Ordinary least-squares estimation: \(t\)-statistics appear in parentheses (* indicates \(p < .05\); ** indicates \(p < .01\); *** indicates \(p < .001\)). The number of observations for PADH is 12; 1 hospital and 12 time periods. The number of observations for the Non-SWITCH group is 36; 3 hospitals and 12 time periods.
significant regression relation between ALSR and the time variables but a rise in ALSR was associated with an increase in hospital occupancy rates.

Residual analyses computed for each of these models indicated that, once again, none of the potential problems of non-normality, variance heterogeneity or autocorrelation among the error terms were important issues.

Although it was clear from the preceding analyses, that in the pre-intervention period the ALSR computed for the PADH had already begun to decline, at a rate of approximately .05 per quarter, a further analysis was performed to determine whether the PADH pre-intervention ALSR was lower, on average, than that of the Non-SWITCH group of hospitals. As hypothesized, it was expected that if the two hospital groups were comparable, then prior to the introduction of SWITCH there should have been no significant differences in trend or average ALSR during the pre-intervention period. It has already been established that there was a difference in trend. To determine whether the two hospital groups differed with respect to their average pre-intervention ALSR, the data from the first 8 time periods for both the PADH and Non-SWITCH hospitals were merged and the following equation was computed:

\[ (3) \text{ALSR}_{it} = \text{OCC}_{it} + \text{HGRP} + e_{it} \]

In this equation, ALSR_{it} is the ALSR in Period i of hospital group t. OCC_{it} is the occupancy rate in Period i of hospital group t. HGRP is a dummy variable which defines the hospital group. Thus, HGRP takes on the value of 1 for the PADH and the value of 0 for each of the Non-SWITCH hospitals. The coefficient of HGRP will indicate whether the pre-intervention ALSR differed, on average, across hospital groups. The results of this analysis
indicated: 1) there was a significant regression relation between the dependent variable and the independent variables ($F = 8.415; p < .01$); 2) the model accounted for 32% of the variability in the pre-intervention ALSR ($R^2\text{-Adjusted} = .323$); 3) no significant effect of OCC was observed ($b_1 = .001; t = .463; p > .05$); and 4) on average, the pre-intervention ALSR was .14 lower for the PADH than for the Non-SWITCH group of hospitals ($b_2 = -.140; t = -3.917; p < .001$). Thus, it would appear that in addition to the observed difference in trend between the two hospital groups, the PADH had a significantly lower pre-intervention ALSR than the Non-SWITCH group of hospitals.

V.3 Supplementary Analysis

An examination of the age distributions of the PADH and Non-SWITCH data bases (see Table VII) indicated that, while the Non-SWITCH patients were fairly evenly distributed among the four defined age groups, the PADH patients were more heavily weighted in the upper age categories ($\chi^2 = 1,051.3; df = 3; p < .001$). Although Age was not included as a variable in the above regression analyses (since Age is one of the variables that was considered in the development of CMG's), to determine whether this higher proportion of older patients may have masked an effect of the SWITCH, equations (1) and (2) were re-estimated using only those patients in the PADH and Non-SWITCH data bases that were under the age of 70 at the time of their discharge from the hospital.

1 The assumptions made here are that older patients may tend to have longer stays, and that they may have less access to appropriate alternatives to acute care. Hence, even if they are identified as ready for discharge, discharge may have to be delayed until an appropriate alternative can be arranged.
The quarterly ALSR's computed for this group of patients, from both the PADH and Non-SWITCH hospitals, are shown graphically in Figure 7 and the results of the regression procedures are presented in Table XI. As can be seen from the regression results, the exclusion of the older age group of patients had little effect on the general trends in ALSR over time, for either the PADH or the Non-SWITCH hospitals. However, the graph suggests that, by excluding older patients, the post-intervention length of stay trend observed at the PADH has become more consistent, with a greater divergence from the Non-SWITCH group, than has been observed previously. If this trend continues, then it may be that with an extension of the follow-up period a reduction in the ALSR, attributable to the SWITCH, could be detected.
Figure 7. Case-Mix Adjusted Length of Stay Ratio by Hospital Group - Patients Aged 18 - 69
Table XI. Regression Coefficients\(^1\) using ALSR as Dependent Variable: PADH and Non-SWITCH Hospitals - Patients Aged 18 to 69 - Standard Set of CMG's

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>PADH Equations</th>
<th>Non-SWITCH Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) ALSR</td>
<td>(2) ALSR</td>
</tr>
<tr>
<td>Constant</td>
<td>-.449</td>
<td>-.501</td>
</tr>
<tr>
<td>Period</td>
<td>-.040</td>
<td>-.041</td>
</tr>
<tr>
<td></td>
<td>(-3.303)*</td>
<td>(-3.519)**</td>
</tr>
<tr>
<td>TS</td>
<td>-.014</td>
<td>-.029</td>
</tr>
<tr>
<td></td>
<td>(-.564)</td>
<td>(-1.022)</td>
</tr>
<tr>
<td>DTV</td>
<td></td>
<td>-.022</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-.387)</td>
</tr>
<tr>
<td>OCC</td>
<td>.017</td>
<td>.018</td>
</tr>
<tr>
<td></td>
<td>(2.223)</td>
<td>(2.347)*</td>
</tr>
</tbody>
</table>

| F-Ratio               | 11.09** | 10.873** | 9.85** | 1.637  | 1.902  | .628    |

| \(^2\)-Adjusted         | .73     | .73      | .62    | .05    | .07    | .00     |

\(^1\) Ordinary least-squares estimation: \(t\)-statistics appear in parentheses (* indicates \(p < .05\); ** indicates \(p < .01\); *** indicates \(p < .001\)). The number of observations for PADH is 12; 1 hospital and 12 time periods. The number of observations for the Non-SWITCH group is 36; 3 hospitals and 12 time periods.
CHAPTER VI

Discussion

VI.1 The SWITCH Index System

Did the SWITCH Index System effect a reduction in length of stay at the PADH? The evidence gathered in the present study would suggest that it did not. On the basis of regression equations computed for both the PADH and the Non-SWITCH group of hospitals, it was determined that while there was a general decline in the PADH length of stay over the time period studied, there was no change in the rate of decline, following the advent of the SWITCH, nor was there a change in the average pre- and post-intervention lengths of stay. It is suggested, therefore, that although the PADH length of stay did decline, over the time period studied, no component of this decrease could be attributed to the initiation of the SWITCH.

To determine whether the observed decline could be explained by general secular trends, the regression equations computed for both the PADH and the Non-SWITCH hospitals were compared. The results of this comparison indicated that while the PADH length of stay decreased over the study period, the length of stay of the comparison hospitals did not. Thus, it does not seem likely that general secular trends were responsible for the observed PADH length of stay trend. However, a limitation in the study methods prohibits the exclusion of secular trends as a possible explanatory variable. To be reasonably confident that the observed decrease was not due to secular trends, it must be assumed that no biases were introduced in the selection of the comparison group of hospitals. That is, it must be assumed that the length of stay trend observed for the Non-SWITCH group of
hospitals was representative of the trend that would have been observed had all of the eligible PADH peer group member hospitals participated in the study. As noted in Chapter IV, it was not possible to determine whether the hospitals that did not participate in the study differed sufficiently from the participating hospitals to affect the observed length of stay trend. Thus, the possibility that the length of stay results were affected by self-selection biases cannot be discounted.

Assuming the observed PADH length of stay trend was not attributable to secular trends, nor to the SWITCH, what other factors could have contributed to the general decline? While no direct attempt was made in the present study to answer this question, the results suggest that at least one of the potentially important variables is hospital occupancy. For the PADH, hospital occupancy was found to be negatively associated with length of stay, while increasing over time. These findings suggest that, for this hospital, the observed decline in length of stay was at least partially associated with an increase in occupancy rate over time. In contrast, for the Non-SWITCH group of hospitals, there was a positive association between hospital occupancy and length of stay and no linear relationship between occupancy and time. These apparently contradictory findings are difficult to explain but a consideration of the distributions of hospital occupancy (See Table VII in Chapter IV), for the two hospital groups, provides a potential clue. During the time period under study, the occupancy rate of the PADH was, on average, considerably higher than that of the Non-SWITCH group of hospitals (Medians = 91% and 83%, respectively). In fact, the median quarterly occupancy rate observed at the PADH even surpassed the maximum quarterly occupancy rate observed for the Non-SWITCH group.
(i.e., 87%). These observations suggest that, perhaps, there is a different functional relationship between occupancy and length of stay depending on the level of occupancy under consideration. More specifically, it is suggested that until a hospital’s occupancy rate consistently approaches maximum capacity, there may not be sufficient incentive for physicians to ensure that patients are discharged as soon as is medically appropriate. In fact, when occupancy rates consistently fall in the 80-85% range, suggesting moderate rather than intense pressure on bed availability, there may be an incentive for physicians to keep patients in the hospital longer, so as to ensure that they will have access to a bed when it is needed for a subsequent patient. On the other hand, as occupancy approaches capacity, peer pressure from other admitting physicians may be sufficient to motivate more timely discharge.

Of course, such an explanation is purely conjecture at the present time. The relationship between hospital occupancy and length of stay is complex, and not yet adequately understood. Nonetheless, the results of the present study concur with previous research which has identified occupancy as a potential confounder in length of stay analyses. (Lave & Leinhardt, 1976a, 1976b).

Although the general decline in length of stay at the PADH may be at least partially attributable to the effects of hospital occupancy, it is clear from the regression results that other factors must also have been involved. When occupancy alone was used as a predictor variable, it was found to account for only 36% of the variability in the adjusted length of stay ratio. On the other hand, when time was the only variable considered, 72% of the variability was accounted for and, when both time and occupancy were considered, the predictive power of the model increased to 80%. The inference drawn from these results is that, in addition to hospital occupancy,
some other, as yet undefined, variable(s) were exerting an influence on the observed trend in PADH length of stay. Unfortunately, because the present study was designed to provide only a global indication of length of stay trends over time, it is not possible to identify what these additional confounders might have been. However, using a similar rationale to that discussed above, it is possible that because of the high and increasing occupancy rate observed over time at the PADH there may have been an increased awareness, on the part of the hospital's medical staff and prior to the development and implementation of the SWITCH, of the need to monitor and ensure efficient and appropriate utilization of an increasingly scarce hospital resource, namely the hospital bed. As is suggested by the findings of previous investigators (Averill, & McMahon, 1977; Dobson et al., 1978; Lave & Leinhardt, 1976a), there may be reason to believe that one of the most important features of an effective utilization management program is not the specific mechanics that are used but, rather, the increased status and priority accorded the utilization process within the hospital setting. Consequently, if in the pre-SWITCH period both the management and medical staff of the PADH had begun to perceive utilization management as a priority issue for the hospital, then it is possible that the observed reduction in length of stay may, at least in part, have reflected an informal response to this perception (i.e., a Hawthorne effect). In this context, the subsequent initiation of the SWITCH may simply have reflected the formalization of a utilization management process that was already well underway. Certainly, it was clear from conversations with hospital personnel that utilization problems had been under discussion for some
time prior to the introduction of the SWITCH. Thus, the possibility of an informal response would appear to have been quite likely.

VI.2 Alternative Explanations for the Failure to Detect a Program Effect

Despite the evidence gathered in the present study, it is too early to state conclusively that the SWITCH Index System did not effect a reduction in length of hospital stay. While this possibility cannot be discounted, there are several reasons why the present study may have failed to detect a reduction in length of stay that could have been attributable to the SWITCH.

1) Prior to the initiation of the SWITCH, concerns about utilization problems may already have begun to have an effect on the hospital's length of stay.

Because the SWITCH Index System was developed and implemented prior to undertaking the present evaluation, it was not possible to get a first-hand assessment of the nature and intensity of the utilization review procedures employed prior to the advent of the SWITCH. Consequently, as discussed above, it is not possible to discount the possibility that, given the consistently high occupancy rate experienced by the hospital throughout the study period, the effectiveness of the SWITCH may have been attenuated by earlier responses to perceived utilization problems. Clearly, utilization issues had been of sufficient concern to prompt the development and implementation of what was, at the time, a highly unique and innovative utilization management program. It is possible, therefore, that had the program been implemented in a hospital for which occupancy was less of a problem and for which utilization was only beginning to emerge as a priority, a different pattern of results may have been observed.
2) The post-intervention period may have been too short (i.e., too few observations) to detect a reduction in length of stay.

Given the small number of observations upon which the regression analyses were based, particularly in the post-intervention period, there may have been insufficient power to detect anything more than a major change in the slope and/or intercept of the regression line following the implementation of the SWITCH. In addition, since the SWITCH was designed to identify and eliminate inappropriate days of care resulting from inefficiencies in organizational procedures, clinical practice patterns, and external services - all of which could require a considerable investment of time to correct- there may be a lag between the program's implementation and the time when it could be expected to have a discernable effect on length of stay. Specifically, as each Off-Index case is identified, and as patterns of Off-Index use begin to emerge, time is required to ascertain the source of the problem, to formulate and provide relevant feedback, and to devise and implement appropriate corrective action. This would be particulary true in the case of Off-Index days which occurred because of a lack of appropriate alternatives. As noted in Chapter II (section II.4.8.3), studies examining the reasons for inappropriate use have typically identified a lack of appropriate alternatives as one of the primary factors contributing to inappropriate acute care bed utilization (Gertman & Restuccia, 1981; Restuccia & Holloway, 1976; Zimmer, 1974). However, the development of the alternative sevices needed to alleviate this problem requires system-wide changes over which an individual hospital may have little immediate control. Hence, if a lack of appropriate alternatives was a significant problem for the PADH, a longer
follow-up period would likely be needed before a length of stay effect could be expected to occur. At the same time, however, if too much time has to pass before an effect is observed, it will be difficult to assert that the effect was due to the SWITCH and not to other intervening variables.

3) It could be that as the length of stay at the PADH declined, its case mix severity increased. If so, and if this increased severity was found to be related to an increase in length of stay, then it is possible that an effect of the SWITCH may have been missed.

Although CMG's were used in an effort to control for changes in hospital case mix over time, it may be that the failure, documented by previous investigators (Brewster et al., 1985; Horn & Horn, 1984; Horn, Horn, & Sharkey, 1984; Young, 1984), of CMG's to adequately account for changes in severity may have masked a length of stay reduction attributable to the SWITCH.

4) It is difficult to discern a length of stay effect in a hospital that is in the midst of a secular decline (for whatever reason).

In order to have a better chance of detecting a program effect, the SWITCH Index System should have been implemented in a hospital known to be in a length of stay steady state. The fact that the PADH did not have a stable length of stay in the pre-intervention period makes it very difficult to identify any component of the general decline that may have been attributable to the SWITCH. This is a major problem for evaluations
initiated after the development and implementation of the program intervention, and one that is not amenable to post hoc control.

VI.3 Methodological Issues and Study Limitations

Although the study is clearly exploratory, it represents one of the first attempts to evaluate a utilization program, implemented in a Canadian acute care setting, using a multiple time series design with a "control" group. While this design represents an improvement over the more typical single observation, pre- and post-intervention case study designs, the study still suffers from a number of serious limitations and methodological problems.

One of the most serious limitations of the present study, is that the SWITCH Index System was implemented in only one hospital. Furthermore, the selection of this one hospital was not random but was decided on the basis of a recognized concern, on the part of the hospital's administrative and medical staff, about hospital utilization. Although an effort was made to take into account the potential effects of secular trends and certain hospital-specific factors (e.g., occupancy rate, bed size, and teaching status, and case-mix), the potential effects of other unidentified, hospital-specific characteristics on the observed length of stay trends at the PADH are unknown. Consequently, there is no assurance that the pattern of results seen at the PADH would be repeated if the SWITCH was implemented in another hospital, or group of hospitals. This limitation would have been even more problematic had an effect of the SWITCH been found. In that event, it would have been very difficult to determine with any certainty whether the observed effect was due to the SWITCH or to other non-related, hospital-specific factors.
Similarly, the selection of the comparison group of hospitals was not random but was determined on the basis of specified eligibility criteria and each hospital's willingness to participate in the study. Although the criteria used were reasonable and necessary (in order to increase the comparability of the hospitals), reliance upon voluntary participation limits the extent to which the length of stay trends exhibited by the comparison hospitals can be considered representative of the trends that would have been observed had all eligible hospitals (or a random selection of eligible hospitals) participated.

A second limitation relates to the fact that no attempt has as yet been made to establish the reliability and validity of the SWITCH Index criteria. Consequently, there is a possibility that the lack of an observed effect could be due to the fact that the defined criteria may be lacking in reliability, validity, or both. Clearly, an attempt to establish the reliability and validity of the program criteria should have been undertaken prior to the implementation of the program.

A third limitation relates to the fact that no attempt was made to ensure on an ongoing and formal basis, that the system was being applied consistently and in accordance with defined procedures. The operation of the system was not monitored because the evaluation was initiated after the final time period included in the study. Thus, the possibility that the lack of an effect was due to incomplete or inconsistent application of the system cannot be discounted.

Because of these problems, the results of the present study provide little basis for drawing conclusions about the effectiveness or ineffectiveness of the SWITCH, as a mechanism for reducing length of stay. Nonetheless, sufficient evidence has been gathered to merit caution in promoting the
SWITCH Index System as a tool designed to facilitate reductions in hospital length of stay.

VI.4 Implications for Future Research and Program Planning

The results of the present study provide only global and preliminary findings relevant to a determination of the effectiveness of the SWITCH Index System of utilization management. The study has focused on only one of the variables considered important in the utilization management process, namely length of hospital stay. The potential effects of the SWITCH on other important variables such as quality of care, intra-hospital communication, the efficiency and/or effectiveness of the hospital's operational procedures, or the utilization of ancillary services, have not been considered.

All of these issues merit further study. In addition, although the results of present study provide some indirect support for the assumption that inappropriate and unnecessary utilization does occur - since, for a standard set of CMG's, the PADH length of stay decreased - no formal attempt has as yet been made to quantify inappropriate use. The SWITCH Index could be used to facilitate this process. In addition, although the present study was unable to detect a length of stay effect of the SWITCH, the system still has intuitive appeal and greater clinical relevance than other, more statistically oriented, programs. For these reasons, it is likely to be more acceptable to medical personnel than programs based on systems such as CMG profile analyses. Since no program will succeed without the support of physicians, this is an important consideration. Thus, given the limitations of the present study, the potential of the SWITCH to be an effective mechanism for reducing inappropriate utilization should not be
discounted too quickly. Moreover, since the SWITCH Index System provides a mechanism for identifying the reasons for inappropriate days of care, it could also be used to begin developing a better understanding of the factors that contribute to inappropriate bed use. This is important because it would promote a better appreciation of the extent to which inappropriate use is or is not under the control of the hospital and/or its medical staff. Clearly, neither hospital administrators nor physicians should be held accountable for utilization problems over which they have no control. For example, lack of available alternatives to acute care hospitalization is a problem of the health care system, not of individual hospitals. Thus, if a sizeable proportion of inappropriate bed use is attributable to a lack of alternatives, then this is a problem that needs to be addressed at the provincial rather than the hospital level. By using the SWITCH Index, or a similar tool such as the AEP, objective and quantifiable data on the reasons for inappropriate acute care utilization could be gathered and employed in developing new strategies for reducing excessive hospital use.

A more immediate concern, however, is the need for studies that focus on criteria development and testing. Neither the SWITCH criteria, nor any other criterion-based utilization index, should be used as a tool for research or utilization management until it has been demonstrated to be both reliable and valid. For this reason, hospitals and researchers interested in using appropriateness-of-care measures, might be advised to begin with indices that have already been subject to extensive testing. For example, the AEP could be employed. However, since this index was developed in the United States, it would still be desirable to conduct reliability and validity
tests to determine whether the results of the American studies could be replicated in Canadian hospital settings.

Once it has been established that a reliable and valid measure of appropriate use has been developed, then to demonstrate its effectiveness as a utilization management tool, evaluative projects should be undertaken to determine not only its effect on utilization measures such as length of stay, but also on quality of care indicators such as readmission and mortality rates and changes in illness severity from admission to discharge. Clearly, one of the most commonly expressed fears about the implementation of utilization management programs has been that reductions in use may be achieved at the expense of the quality of care provided. Since a primary objective of utilization management is to reduce excessive utilization, *while maintaining an adequate standard of care*, it is important to be able to demonstrate that both components of this objective have been met.

In addition, although the present study has focussed on a criterion-based utilization management program, it is apparent from the discussion of the literature in Chapter III, that a number of Canadian hospitals have begun to experiment with other types of programs. Since these different approaches to utilization management can vary considerably in cost and labour requirements, it is important that studies be initiated to compare the cost-effectiveness of these, and other, newly developed strategies.

Finally, it must be recognized that evaluating the effectiveness of a utilization management program is a complex process. As discussed in Chapter II, many variables have been found to have an effect on utilization measures such as length of stay. Moreover, because of the complexity and multiplicity of factors associated with the hospital utilization process, a
comprehensive model of utilization which could be used as a framework for research or program evaluation, has not yet been developed. Since, to date, most of the research directed at hospital utilization has been performed within the American hospital sector, there is a need for the development of research projects that would both facilitate a better understanding of the utilization process, in the Canadian hospital environment, and permit the development of a clearer methodological approach to the evaluation of utilization management programs. In addition, there is a need for the development of alternatives to length of stay as a measure of the effectiveness of utilization management programs. More informative outcome measures might include the number of inappropriate days of care, the number of inappropriate admissions, reasons for inappropriate use, and changes over time in hospital case mix severity.

In summary, because there has been so little research done on utilization in the Canadian hospital sector, there are, at present, many issues that need to be addressed. In the opinion of the present author, it is essential that this research be initiated immediately. Moreover, given the complexity of the issues involved, and the need for inter-hospital collaboration, the author would concur with a suggestion made recently by Suttie (1988) that a national, or a series of interconnected regional, research centres be established with a mandate to perform hospital utilization research and to oversee program development and evaluation. Through the establishment of such centres, it would be possible to develop the knowledge, expertise and coordination required to improve the state of Canadian hospital utilization research. In addition, by coordinating program development and evaluation, it would be possible to avoid some
of the problems encountered in the present study, by making it more feasible for evaluation plans to be incorporated into the development of program initiatives, right from the start.

VI.5 Policy Implications

While it is too early to make specific recommendations for the development of utilization management policies, based on the SWITCH Index System, certain general recommendations can be made. First, an impetus for the development and refinement of utilization management procedures could be provided by provincial governments if, as was suggested by Suttie (1988), they were to begin requiring that hospitals provide evidence of having performed appropriate utilization analyses prior to their receiving approval for requests for capital expansion or deficit funding. Second, in order to ensure that standards of appropriate utilization are developed, implemented, and properly evaluated, regional committees (along the lines of the American PRO's), comprised of physicians, health care administrators, and other relevant health service professionals, could be established with a specific mandate to monitor hospital utilization and quality of care activities. Third, to ensure that policy development is based on adequate knowledge, continuing support and promotion of research and development, particularly in the areas of hospital utilization, quality of care assessment, and program development, is essential. Finally, since the single, most significant determinant of hospital utilization is the medical condition of the patient (Berki et al., 1984; Boaz, 1979; Clendenning et al., 1976; Goldfarb, Hornbrook, & Rafferty, 1980; Goldfarb, Hornbrook, & Higgins, 1983; Lave & Leinhardt, 1976a, 1976b; Mughrabi, 1976; Ro, 1969), there is a need for the development of policies aimed at encouraging
ongoing development of effective medical technologies, while promoting more intensive evaluation of technologies that are of little known value (Anderson & Lomas, 1988).

VI.6 Conclusion

In conclusion, although this study was unable to confirm the expectation that the SWITCH Index System would be an effective mechanism for reducing hospital length of stay at the PADH, further research is needed before conclusions about its potential effectiveness in other hospital settings can be drawn. However, before any attempt is made to promote the use of the SWITCH, even for pilot testing purposes, it is essential that the SWITCH Index criteria be evaluated for reliability and validity. If possible, once it has been demonstrated that a reliable and valid "appropriateness of care" instrument has been developed, a pilot project involving a new set of hospitals should be initiated with an appropriate evaluation plan included as a condition of its implementation. Furthermore, the proposed evaluation, if undertaken, should be expanded to include a broader spectrum of utilization measures, such as length of stay, number of inappropriate days of care, and reasons for inappropriate use, as well as quality of care indicators such as changes in illness severity from admission to discharge and readmission and mortality rates. In this way, more useful, and less ambiguous, information could be obtained about the effectiveness of the SWITCH approach to utilization management.
References


Burford, R., & Averill, R.F. The Relationship between Diagnostic Information Available at Admission and Discharge for Patients in One PSRO Setting: Implications for Concurrent Review. *Medical Care.* 1979; 17 (4): 369-381.


Mughrabi, M.A. The Effects of Selected Demographic and Clinical Variables on the Length of Hospital Stay. Hospital Administration in Canada. 1976; 18: 82-84, 86, 88.


Nathanson, M. 'Double Dipping' can Hike Profits, but Advantages may be Short-lived. Modern Healthcare. 1984a; 14 (3): 154, 156.


APPENDIX A: Medicare Levels of Care Criteria for Medical/Surgical Patients

Medical-Surgical Flow Diagram

I. Availability of Skilled Nursing Services at All Times

Definition: A skilled service is one which cannot safely and adequately be performed by the average, rational, nonmedical person without direct supervision of trained medical or paramedical personnel (including observation and instruction).

I.A. Observation. Does the unstabilized condition of the patient require the skills of a nurse to detect and evaluate (i.e., observe) the patient's need for possible modification of treatment or institution of medical procedures?

I.B. Direct Service. Does the patient require direct skilled nursing services (excluding observation) every day?

I.C. Patient's Condition. In view of the patient's condition, are the range (number of different skilled services) and intensity (frequency of duration) of all skilled services (e.g., extensive diagnostic tests) furnished such that they cannot be performed outside an institution?

I.D. Is the patient terminal?

(Appendix A cont...)
II. Acute Hospital Medical Services

II.a Having established that the patient requires the availability of skilled nursing services at all times (or requires broad and intense skilled services in an institution), does the patient today require the constant availability of medical services provided by a Skilled Nursing Facility (SNF)?

Source: Cited in Restuccia and Holloway (1976).
APPENDIX B. Pediatric-Modified Appropriateness Evaluation Protocol

Medical Services

1. Procedure in OR
2. Pre-op eval. or post-op evaluation (1 day)
3. Cardiac catheterization
4. Angiography
5. Biopsy of internal organ, NOT bone marrow
6. Thoracentesis/paracentesis/cysternal/ventricular tap
7. Test requiring (a or b)
   a. Strict dietary control
   b. Collection of timed sample; > 8 hours
8. New or experimental treatment requiring frequent dose adjustments under direct MD supervision
9. Documented MD exam >= 3 times within 24 hours

Nursing/Life Support Services

10. Ventilator/IPFB/Chest PT >= TID
11. Mist Tent
12. IV Therapy >= 8/24 hours
13. >= IM/sq injections/24 hours, except insulin
14. Strict I's & O's, ordered & done
15. Major surgical wound & Drainage care (chest tubes, hemovacs, Penrose, traction)
16. CR monitor & 24 hours after monitor DC'd
17. >= 3 Vital signs ordered and done.

Patient Conditions

18. Any two conditions:
   a. RT, PT, OT, or ST
   b. > 5 yo, non-ambulatory, and requires help with transfers
   c. Defined patient education program, documented teaching
   d. Rehabilitation, Psychological, or Social services documented

Within 24 hours before day reviewed:
19. Unable to void, not due to neurologic problem
20. Transfusion: acute bleed
21. MD's suspect suicide attempt
22. MD's suspect child abuse/neglect

Within 48 hours before day reviewed:
23. Temperature >= 101 if patient admitted for other reason
24. Coma
25. Acute confusional state
26. Acute hematologic problem, eg, ANC<500, HCT <20, 000
27. Progressive, acute neurological difficulties

------------------------
Source: Kemper (1987b)
APPENDIX C. List of Potential Barriers to Appropriate Utilization

Barriers: Physician Responsibility

1. Patient is admitted for problem outside the area of competency of attending physician (e.g., should have referred patient to physician in different specialty).
2. Some or all of work-up could have been done on an outpatient rather than inpatient basis; e.g., preadmission work-up prior to surgery is inadequate. (Preoperative day, however, is to be regarded as appropriate unless clearly inappropriate).
3. Physician admits patient to hold bed (for possible future patient).
4. "Political admission," as with VIP (Please comment).
5. Physician admits patient for acute condition, then decides to work-up patient for chronic condition, causing delay.
6. Physician delays scheduling of test or procedure.
7. Unavailability of physician causes delay in performing a procedure, (e.g., surgery which could have been performed on a Wednesday is delayed until Thursday because the physician is unavailable on Wednesday).
8. Test sequencing is inadequate, i.e., ordering test in the wrong sequence causes delay in diagnosis and/or treatment. (Please comment).
10. Failure of attending physician to provide patient (or family) with necessary information for informed consent causes delay.
11. Failure of other physician (e.g., anesthesiologist) to provide patient (or family) with necessary information for informed consent causes delay.
12. Physician is not visiting the patient.
13. UR review physician is not forceful enough in supporting UR function.
14. Physician has legal problem with patient, i.e., malpractice threat or involvement with an attorney.
15. Physician's medical management of patient is conservative.
16. Physician delays decision regarding further treatment of the patient because of the complex medical nature of the case, yet does not request a consultation.
17. Patient is kept in hospital because it is easier for his/her physician to have "all" patients in one facility. (Please comment).
18. Physician refuses alternative facility because "its too far away," for him/her to visit patient there.
19. Physician delays transfer or discharge of patient because, "I want to watch him/her for a few more days myself."
20. Physician has no confidence in quality of services available at SNF.
21. "Interesting case" is kept in hospital for teaching purposes.
22. Patient's transfer is delayed because of late date at which physician writes order.
23. Physician does not want to use SNF on prolonged case because "Then Blue Cross will expect this treatment on all cases like this," (e.g., orthopedic problem which could be treated in several different ways).
24. Other. (Please comment).

(Appendix C cont...)
Barriers: **Hospital Responsibility**

25. Problem in hospital scheduling for test or procedure causes delay.
26. Failure of nursing service to obtain consent for a procedure causes delay.
27. Return of test results is delayed, causing delay in diagnosis or treatment. (Please specify type of test).
28. Scheduling or transfer of patient to another institution to perform a special procedure causes delay.
29. Communication failure, e.g., missed doctor's order. (Please comment).
30. Hospital has legal problem with patient, i.e., malpractice threat or involvement with an attorney.
31. Lack of administrative support contributes to inappropriate utilization, (e.g., inadequate clerical support, excessively heavy workload for coordinator). (Please comment).
32. Patient's transfer or discharge is delayed because of inadequate discharge planning on the part of hospital personnel.
33. Other. (Please comment).

Barriers: **Patient or Family**

34. Patient or family insists on admission to hospital.
35. Patient is admitted because he/she is uncooperative with therapeutic or diagnostic program outside the hospital.
36. Indecisiveness of patient (or family) regarding a procedure (despite provision of adequate information by physician and hospital personnel) causes delay.
37. Patient is uncooperative with therapeutic program in the hospital, causing delay.
38. Patient or family refuses alternative facility because it is too far away.
39. Patient or family insists on patient remaining in hospital.
40. Family member (or friend) is unwilling to care for patient after discharge.
41. Other. (Please comment).

Barriers: **Environmental Responsibility**

42. Insurance coverage for diagnostic procedures is more complete on an inpatient than on an outpatient basis.
43. Insurance coverage for therapeutic procedures is more complete on an inpatient than on an outpatient basis.
44. Medicare requires three-day acute hospitalization prior to SNF coverage.
45. Patient must be admitted to remove from environment adverse to health, e.g., unavailability of family or friends to provide care. (Please comment).
46. Terminal patient might die in transit to alternative facility.
47. Patient from unhealthy environment (e.g., home environment) is kept until environment becomes acceptable or alternative facility is found.
48. Patient is convalescing from an illness and it is anticipated that his/her stay in an alternative facility would be less than 72 hours.

(Appendix C cont...)
Environmental Barriers (cont.)

49. Patient is terminal and stable, yet is kept in the acute hospital for humanitarian reasons. (Please comment).
50. Patient is kept in hospital for short period of time (at the end of hospitalization) because patient or family needs to be taught self care, (as with ileostomy).
51. Patient's insurance coverage is more complete in an acute hospital than in an alternative facility or home health care program.
52. Unavailability of SNF bed.
53. Unavailability of sub-SNF level of care facility (e.g., nursing home, boarding house, etc.).
54. Unavailability of SNF with ability to provide the necessary type of skilled nursing services. (Please specify type of service).
55. Patient is Medicare-MediCal "crossover". No alternative facility available or willing to incur financial risk with patient having both Medicare and MediCal coverage because of MediCal requirement that Medicare be billed first with confirmation of no coverage before assuming coverage for patient itself.
56. Unavailability of SNF with ability to provide the necessary amount of (all types of) nursing services. (Please specify amount of service).
57. Unavailability of SNF with necessary ancillary service(s). (Please specify type of service).
58. Awaiting financial clearance from alternative facility for transfer.
59. Awaiting medical clearance from alternative facility for transfer.
60. Awaiting financial clearance from insurance program for transfer to alternative facility.
61. Family member (or friend) is not available to transport patient from the hospital.
62. Family member (or friend) is not available to care for patient after discharge.
63. Other. (Please comment).

Source: Restuccia and Holloway (1976).
APPENDIX D. Data Elements Included on the HMRI Hospital Discharge Abstract

<table>
<thead>
<tr>
<th>ITEM</th>
<th>NO. COL.</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SEGMENT 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>discharge date:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>day</td>
<td>02</td>
<td></td>
</tr>
<tr>
<td>month</td>
<td>02</td>
<td></td>
</tr>
<tr>
<td>year</td>
<td>02</td>
<td></td>
</tr>
<tr>
<td>institution</td>
<td>04</td>
<td></td>
</tr>
<tr>
<td>batch:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>period</td>
<td>02</td>
<td></td>
</tr>
<tr>
<td>number</td>
<td>02</td>
<td></td>
</tr>
<tr>
<td>abstract</td>
<td>02</td>
<td></td>
</tr>
<tr>
<td>coder</td>
<td>02</td>
<td></td>
</tr>
<tr>
<td>chart</td>
<td></td>
<td>02-02-07</td>
</tr>
<tr>
<td>register number</td>
<td>05</td>
<td></td>
</tr>
<tr>
<td>los</td>
<td>04</td>
<td></td>
</tr>
<tr>
<td><strong>SEGMENT 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>province</td>
<td>01</td>
<td></td>
</tr>
<tr>
<td>institution</td>
<td>04</td>
<td></td>
</tr>
<tr>
<td>batch:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>period</td>
<td>02</td>
<td></td>
</tr>
<tr>
<td>number</td>
<td>02</td>
<td></td>
</tr>
<tr>
<td>abstract</td>
<td>02</td>
<td></td>
</tr>
<tr>
<td>coder</td>
<td>02</td>
<td></td>
</tr>
<tr>
<td>chart</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>register number</td>
<td>05</td>
<td></td>
</tr>
<tr>
<td>los</td>
<td>04</td>
<td></td>
</tr>
<tr>
<td>health care number</td>
<td>08</td>
<td></td>
</tr>
<tr>
<td>postal code</td>
<td>05</td>
<td></td>
</tr>
<tr>
<td>residence code</td>
<td>04</td>
<td></td>
</tr>
<tr>
<td>sex</td>
<td>01</td>
<td>male, female, undetermined</td>
</tr>
<tr>
<td>field name not specified</td>
<td>09</td>
<td></td>
</tr>
<tr>
<td>field name not specified</td>
<td>01</td>
<td></td>
</tr>
</tbody>
</table>

(Appendix D cont...)

1 An original HMRI Discharge Abstract is not provided because the pale blue colour of the form prohibits mechanical reproduction.
APPENDIX D (cont).

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>01</td>
</tr>
<tr>
<td>birth date:</td>
<td></td>
</tr>
<tr>
<td>day</td>
<td>02</td>
</tr>
<tr>
<td>month</td>
<td>02</td>
</tr>
<tr>
<td>year</td>
<td>02</td>
</tr>
<tr>
<td>admission:</td>
<td></td>
</tr>
<tr>
<td>day</td>
<td>02</td>
</tr>
<tr>
<td>month</td>
<td>02</td>
</tr>
<tr>
<td>year</td>
<td>02</td>
</tr>
<tr>
<td>hour</td>
<td>02</td>
</tr>
<tr>
<td>discharge:</td>
<td></td>
</tr>
<tr>
<td>day</td>
<td>02</td>
</tr>
<tr>
<td>month</td>
<td>02</td>
</tr>
<tr>
<td>year</td>
<td>02</td>
</tr>
<tr>
<td>hour</td>
<td>02</td>
</tr>
<tr>
<td>institution from</td>
<td>04</td>
</tr>
<tr>
<td>entry codes</td>
<td>01</td>
</tr>
<tr>
<td>stillbirth</td>
<td></td>
</tr>
<tr>
<td>institution to</td>
<td>04</td>
</tr>
<tr>
<td>date ready for discharge:</td>
<td></td>
</tr>
<tr>
<td>day</td>
<td>02</td>
</tr>
<tr>
<td>month</td>
<td>02</td>
</tr>
<tr>
<td>admission category</td>
<td>01</td>
</tr>
<tr>
<td>admitted by ambulance</td>
<td>01</td>
</tr>
<tr>
<td>main patient service</td>
<td>02</td>
</tr>
<tr>
<td>subservice</td>
<td>01</td>
</tr>
<tr>
<td>exit alive codes</td>
<td>01</td>
</tr>
<tr>
<td>death codes:</td>
<td></td>
</tr>
<tr>
<td>coroner/autopsy</td>
<td>01</td>
</tr>
<tr>
<td>&lt; 48 hr after admission</td>
<td>01</td>
</tr>
<tr>
<td>in or suite</td>
<td>01</td>
</tr>
<tr>
<td>&lt; 48 hr. post op.</td>
<td>01</td>
</tr>
<tr>
<td>&gt; 48 hr. post op.</td>
<td>01</td>
</tr>
<tr>
<td>trauma</td>
<td>01</td>
</tr>
<tr>
<td>pregnancy</td>
<td>01</td>
</tr>
<tr>
<td>other</td>
<td>01</td>
</tr>
<tr>
<td>complex case abstract</td>
<td></td>
</tr>
<tr>
<td>incomplete</td>
<td>01</td>
</tr>
</tbody>
</table>

(Appendix D cont...)
APPENDIX D (cont).

SEGMENT 3 -

service transfers:
  service A:
    patient service 02
    sub service 01
    days 02
  service B:
    patient service 02
    sub service 01
    days 02

SEGMENT 4

most responsible doctor 05
other doctors #1:
  type 01
  other doctor 03
  service 02
other doctors #2 - #7:
coded as per 'other doctor #1'

SEGMENT 5-

most responsible diagnosis 10
other diagnoses (n = 15) 10 per

SEGMENT 6-

principal procedure:
  day 02
  month 02
  procedure 04
  doctor 03
  service 02
  tissue 01
  time 03
  OR 01
  anaesthesia 03
  tech. 01

other procedures: (n=11)
coded as per 'principle procedure'

(Appendix D cont...)
APPENDIX D (cont).

### Basic Options:
- a - h 01 per
- i - l 01 per
- m - n 02 per
- o 04

### Segment 7:
- Therapy 01
- Other 01
- Blood Given 01
- Pre-admission Workup 01
- Discharge Planning 01
- Social Services 01
- Infections 01

**Special Care Units:**
- Unit 1 -
  - Unit No. 02
  - Days 02
- Unit 2 - 3 coded as per 'Unit 1'
- Weight 04

### Segment 8:
- Project Codes:
  - Project #1 -
    - Project Code 03
    - 1-13 01 per
    - 14 - 15 02 per
    - 16 03
    - 17 04
  - Project #2 - #5 coded as per 'Project #1'

---

Source: HMRI (1983)
APPENDIX E. Release of Data Form

RELEASE OF DATA FORM

This is to confirm that Ms. Sandra Wiggins has sought and obtained permission to include the ___________ Hospital in the study entitled "Utilization Management of Acute Care Services: An Evaluation of the SWITCH Index System". Consequently, the Hospital Medical Records Institute (HMRI) is hereby authorized to release to Ms. Wiggins the following information:

For each patient discharged from the ___________ Hospital between January 1, 1982 and December 31, 1985, all available data corresponding to the following list of data elements (and only the specified elements):

- Case Mix Grouping
- Length of Stay
- Sex
- Age
- Postal Code
- Admission Category
- Exit Alive Code
- Death Code
- Admission - Institution From
- Discharge - Institution To
- Most Responsible Doctor Code
- Special Care Units
- Main Patient Service
- Sub Service
- Admission Date
- Discharge Date

It is understood that this data will be used for the sole purpose of conducting the study referred to above.

Signature: __________________________
Title: __________________________
Date: __________________________

HMRI Institution Number: __________________________