STATISTICAL ISSUES IN EMPIRICAL STUDIES TO ASSESS THE IMPACT OF HEALTH POLICY

Pete W. Shaughnessy, Spectrum Research, Inc.

I. INTRODUCTION

This paper focuses on issues entailed in the conduct of studies designed to ascertain the impact of health policy. As such, the paper has several generalizations to assessing the impact of policy variables in areas other than health. Examples of such generalizations are also given.

From a statistical perspective, the intent is to cover certain statistical methods designed to "filter out" policy effects from the effects of other intervening or confounding variables. Operational procedures involved in the conduct of studies to examine policy impacts are also discussed. Although it is assumed throughout that one is dealing with a large number of variables and/or large sample sizes, the techniques and issues addressed in this paper also pertain to smaller studies both in terms of number of variables and sample sizes. In addition, since such studies frequently necessitate reasonably rapid turn-around with distinct constraints on dollars to be spent in their conduct, it is also assumed that one must develop the study design expeditiously, prior to the empirical work.

Since this paper is expository rather than technical in nature, an attempt is made to cover several issues without the use of extensive statistical or mathematical notation. The remainder of the paper is divided into four sections. These sections are integrated, but deliberately written to be relatively independent of one another in order to allow for selective reading of various sections. The second section deals with examples and generalizations of health care policy issues which are the focus of the statistical methods and operational considerations discussed in the remaining three sections. Section III is concerned with an overall model or paradigm of the health care delivery system which aids in the design and conduct of empirical studies to assess the impact of health policy. Section IV violates slightly the objective to be non-technical; it contains a discussion on five different methodological points which can be useful in policy research studies in health. The final section is concerned with various operational considerations which are pertinent to the implementation and conduct of empirical projects to assess policy impact.

II. EXAMPLES AND GENERALIZATIONS OF FOCUS

A. Health Care Cost Containment Policies: An example of cost containment in the hospital sector is prospective reimbursement. There are several types of prospective reimbursement systems, but basically they represent payment systems for hospitals in which a hospital agrees to a certain rate structure or budget prior to the time it is actually reimbursed (or incurs costs). A Social Security Administration funded project to analyze the impact of the Indiana Prospective Reimbursement System for hospitals is discussed in Section IV. Other types of cost containment policies are certificate of need legislation and utilization review programs. The intent behind certificate of need legislation is to contain the growth of hospitals (and other institutional) facilities in specific areas in order to minimize over-utilization of health care facilities. The cost containment purpose of utilization review programs is to monitor utilization of health care facilites in order to ascertain whether consumers, third party carriers, etc., are over paying due to over-utilization of health care services and facilities. In assessing the impact of such cost containment policies it is important to obtain information on the basic performance variable(s) which the policies are intended to effect. In this case, the performance variable (or dependent variable) is cost, the primary independent variable is the cost containment policy itself, and the mitigating or confounding variables are those which influence cost apart from the policy variables.

B. Quality Assurance Programs: The fundamental purpose of quality assurance programs is to monitor and maintain the quality of health care delivered to consumers. An internal quality assurance program in a hospital would be conducted basically by the hospital itself whereas an external review program would be conducted by an organization outside the confines of the hospital. An example of an external review mechanism is the Professional Standards Review Organization (PSRO), a fairly recent institution legislated into existence at the federal level. PSRO's are responsible for monitoring utilization and quality of care in specific geographic regions. Certain medical societies and foundations have recently become PSROs serving the geographic region in which they are located. In examining quality assurance programs, the primary performance variable to examine is quality of care. There are two basic types of quality indices which have been used to date. First, there is the outcome or health status type of index which has been utilized only on a small scale, usually for highly focused empirical studies. Next, there is the process-type measure of quality which is somewhat more common in larger studies. An example of process indices is provided in Section IV. The policy variables of interest are those which characterize the quality assurance programs under consideration. Confounding variables would be those which influence the quality of care apart from the quality assurance program, such as the physician specialty mix in an area.

C. Health Care Financing Policy: There are several types of financing arrangements available to consumers of health care. The two largest government sponsored alternatives are the Medicare and Medicaid programs. Commercial health insurance is regulated primarily at the state level. Thus, not only are government run programs important policy issues in health care financing, but also state health insurance regulatory activities are also relevant policy variables in this area. Also in the financial area, there are various federal and state programs which provide capital for the construction of health care facilities and other resources. For health care financing, the major policy variables are those which describe the type of financing system, legislation, or regulations under consideration. The primary performance or dependent variables would depend on the type of financing alternatives being examined, but they could be variables such as total dollars paid for health care per capita, quality of care measures, measures of the distribution and specialty mix of physicians (if a financing policy is intended to influence such variables), etc. Confounding or mitigating variables would include factors such as the initial health status of the consumer population and the availability of different types of health care facilities and resources.

D. Less Traditional Health Care Institutions Which Have Been or Likely Will be Encouraged by Legislation: Perhaps the best example of such an institution is the Health Maintenance Organization (HMO). This type of prepaid health care is thought by many to provide better incentives for both consumers and providers in terms of the delivery of quality health care at minimal cost. Another example of an innovative type of institution is the outpatient surgi-center, which allows for certain types of surgery to be performed without hospitalization. In evaluating the impact of components of the health care system such as the HMO and the surgi-center, it is important to cover a wide range of performance variables since, although cost may be reduced by such innovations, quality may possibly suffer. Thus, a range of perforamnce measures should be considered; the primary policy variables would characterize the HMO's or surgi-centers being evaluated. The mitigating or confoudning variables are similar to those mentioned in the above three paragraphs.

E. The Practical Orientation of Results of Policy Research: One of the most important components in the delivery of health care is health planning at federal, regional, state, areawide, and local levels. The results of research studies designed to assess the impact of health policy should be directed toward implementation at the appropriate level. Comprehensive health planning deals with a series of issues which have been, are being, and should continue to be explored by policy research aimed toward implementation and improvement in the provision of health care. Problems such as the maldistribution of physicians have several causes and ramifications which policy research studies can address. With respect to health care provided by institutions such as hospitals, extended care facilities, ambulatory care facilities, mental health facilities, drug addiction and alcoholism facilities, there are several pertinent considerations which policy research can address. Such considerations include facility location, optimal utilization, the need for and types of improvements, needed resources (including financing), manpower considerations, supply, and productivity. In addition, there is the practical problem of the integration of public and private health care providers, funders, regulators, and administrators.

F. Generalizations of Focus: The types of statistical issues and quantitative considerations discussed in this paper extend to physiological or biological areas of health as well as to the somewhat more consumer oriented aspects of health care delivery, including preventive and environmental health. Generalizations to other service-oriented programs also exist. One such area is the welfare delivery system. For example, a major issue in welfare policy is the impact of the Aid to Families with Dependent Children (AFDC) income disregard formula for employed parents. This is and has been the subject of policy research for the past several years. There are several confounding variables such as the employment background of welfare clients, socio-demographics, and so on. The policy variable is the income disregard formula itself and the chief dependent variables are employment related outcome measures such as employment status and/or income level adjusted for time employed.

The comments in this paper pertain largely to assessing the impact of existing policy. This can be referred to by the statistician as "autopsy statistics" in the sense that the treatments have been applied to the subjects prior to the study, without much regard for a scientific evaluation at some future date. The opposite approach would be a bona fide social experiment in which different policies are administered to different subject groups in accord with an experimental design approach. Nonetheless, several of the operational considerations and some of the methodological considerations discussed here pertain to the experimental as well as the autopsy type of study.

III. A PARADIGM WHICH LENDS ITSELF TO POLICY RESEARCH

A. Summary: The below model of the health care delivery system was developed as part of a National Science Foundation study to examine the health care delivery system in metropolitan areas. It is a flexible paradigm which can be altered in various ways to meet the needs of the researcher, depending on evaluation objectives, units of analysis, and a variety of other circumstances. An approach utilizing this type of paradigm is probably most functional if the paradigm is employed at the outset of the study in order to develop the conceptual framework. In this process, it often becomes evident that certain types of hypotheses are highly relevant or, in some cases, quite irrelevant to the mainstream objectives of the study. It is also possible that the conceptual stage can point out the need for certain types of modeling prior to the finalization of the study design. Modeling can encompass a wide range of activities from pure conceptual work to fairly extensive simulation or pseudo-simulation work -- depending on cost and time limitations.

The following two subsections, B and C, describe the major components of the paradigm which is diagrammed and further discussed in subsection D. This paradigm is intended to provide a vehicle which assists in both the development and conduct of studies whose objectives deal with assessing the impact of health policy initiatives. (Analogous paradigms can be constructed for other types of policy work.) Since the initiatives availabe to the policymaker are usually implemented through some type of regulatory or financing program, the major "change agents" with which the paradigm is concerned are these two.

B. Major (Non-performance) Characteristics of the Health Care Delivery System: The non-performance variables associated with the health care delivery system as defined here can be categorized into three major groups: regulatory variables, financing variables, and structural variables. If one is examining the performance of the delivery system (as a set of dependent variables), then these three categories can be regarded as classes of independent variables. In some instances, however, the purpose of an evaluation study may be to examine one of these categories as the primary set of dependent variables. For example, it may be appropriate to examine the impact of regulatory and financing variables on structural features of the health care delivery system.

The class of regulatory variables includes characteristics such as the nature of the licensure regulations for health care professionals including physicians, nurses, etc. In addition, it would include different types of capital expenditures regulation, such as certificate of need laws. This variable category further includes the nature of health insurance regulation which, as mentioned earlier, has its locus chiefly at the state level, although national health insurance may change this locus in the near future. Other types of regulatory activities include quality assurance and utilization review programs. Different types of cost and price regulation such as prospective reimbursement programs and the Economic Stabilization Program are also in this class of variables.

The financial characteristics of the health care delivery system include variables such as hospital costs and charges. Physician fee schedules and private health insurance reimbursement schedules are also in the domain of financial variables. The number of subscribers and general types of coverage for third party health insurance carriers; the breadth (in terms of benefits) and depth (in terms of eligible recipients) of the Medicaid programs in different states; and the characteristics of the Medicare Part A and Part B programs (the hospital and physician insurance components, respectively) are also in the class of financial variables.

The structural characteristics of the health care delivery system can be subdivided into provider structure, consumer structure, and situational/environment structure. The provider and consumer structural characteristics can be regarded as supply and demand variables, respectively. On the provider side, structural variables include hospital characteristics such as numbers, types, and distribution (assuming the unit of analysis is a specific geographic entity such as the county, say). Characteristics of nursing homes, once again including number, type, and distribution (among other variables), are also elements of the provider structure category. Physician parameters which describe the specialty mix, numbers, and distribution of physicians are examples of other types of structural variables. In addition, the nature and penetration of HMOs are part of the provider structure of the health care delivery system in a specific location. The availability (in some cases duplication) of tertiary care facilities and certain types of medical technology provides another illustration of a structural variable. The emergency medical service capabilities of the area fall into this category as well.

The consumer structure of the health care delivery system includes variables such as the health status of the consumer population. This, of course, just as with the vast majority of variables mentioned so far, is actually a vector of several component variables including items such as morbidity and mortality rates. In addition, the actual number of individuals served by the (sub)delivery system under consideration is also a consumer structural variable. Demographic characteristics such as age, education, and family composition of the consumer population further define this cateogry of variables. Financial resources of the consumer population (this does not include the financial variables directly associated with the delivery system described above -- i.e., those variables which related more directly to financing and delivering health care) are also components of the consumer structural category.

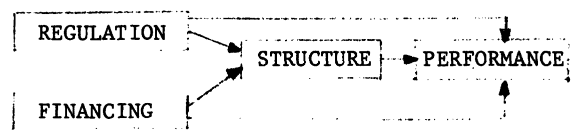
The final category of structural variables includes those which are not necessarily directly related to supply and demand characteristics but nonetheless can mitigate the impact of such characteristics. The situational and environmental variables which could influence the performance of a health care delivery system include the geographic location of the system and the climatological characteristics of that location. The availability of transportation, including private automobiles, public transportation, networks of interstate highways, etc., are important determinants of access to various types of health care. Provider and consumer attitudes, ignorance, and the political environment can also influence the delivery of health care. This class of variables further includes the at times, highly important but hard to measure, characteristics of strong leadership, staff morale, and other similar motivational or psychological characteristics which are largely a function of personality dynamics.

C. Performance Characteristics of the Health Care Delivery System: Under the assumption that the overall objective of any health care delivery system is the maintenance and improvement of the health status of a particular consumer population, one can categorize performance indicators of the health care delivery system into two general classes. First, there is the class of outcome indicators which includes chiefly health status variables and changes in health status as a result of the delivery of health care. The second general class of outcome indicators is used more frequently simply because data are usually more readily available. This category consists of process measures of the performance of the delivery system. As such, it includes measures of quality in terms of how health care is dedelivered from the viewpoints of equity, effectiveness, and efficiency. As stated previously, performance indicators for the outcome category are not usually available for the types of policy research projects addressed by this paper. Hopefully, survey and data gathering activities such as those currently conducted and sponsored by the National Institutes of Health, the National Center for Health Statistics, and the Center for Disease Control will continue to alleviate this problem.

In terms of process indicators of quality, one can examine various types of patient management characteristics by diagnostic and/or procedural cohorts of patients. For example, in the Indiana and the (second phase of the) metropolitan area projects mentioned earlier, we intend to obtain hospital data from the Commission on Professional and Hospital Activities (CPHA) in Ann Arbor which will enable us to construct process measures of quality of care by combining patient treatment variables across various procedural and diagnostic cohorts. We will be examining patient cohorts falling into the general areas of surgery, internal medicine, obstetrics and gynecology, pediatrics, and psychiatry. For each cohort the analyses will be based on specific procedural information which describes the type of care received by patients in that cohort. The actual "goodness" or "badness" of each type of procedure was ascertained by an advisory group of physicians and other health care specialists. As an illustration, in the pediatrics category one of the diagnostic cohorts is pneumonia and acute bronchitis, age 14 and under. For this cohort (for each hospital in the study) we will collect information which describes the cohort in terms of socio-deomographics, the percent of individuals falling into the cohort who undergo complete blood count testing, percent with urinalysis, percent with skin tests, percent with diagnostic x-rays, and approximately 10 more different types of procedural characteristics. In general, 40 to 60 diagnostic or procedural patient cohorts will be utilized in these studies. A general discussion on the types of methods used in constructing process measures of quality from such data is contained in Section IV.E of this paper.

Equity of the delivery system can be measured in various ways. Perhaps one of the most straightforward indices of equity is based on the relative proportion of different consumer subgroups served. For example, if the elderly or the indigent receive a substantially lower amount of health care expressed in terms of dollars expended, say (adjusted for health status, ideally), then the system is said to be inequitable. Even if a delivery system is reasonably equitable, there are varying degrees of effectiveness at which it might operate. Effectiveness can be measured in various ways, perhaps one of the most straightforward being the portion of the demand satisfied. This can be measured in terms of deviations from some preset standard such as a maximum (which might represent over-utilization) obtained by finding the largest portion of consumers or dollars spent on health care in one area and using this maximum or some point below it as a norm which should be met. Efficiency can be measured by several standard cost and productivity indices including different types of services provided per unit cost, needless provision of services, also possibly in accord with a preset norm.

D. Utilizing the Paradigm -- The Policy Perspective and Research Perspective: The basic paradigm described in the above three subsections can be depicted as:



Since the change agents normally available to policymakers include chiefly regulatory actions and financing alternatives, the intent of policy research is thus to assess the impact of these change agents on the performance of the health care delivery system. These impacts can occur through the structural characteristics of the system and in turn influence performance, as is the case with certificate of need legislation which is intended to decrease hospital beds (actually the beds to population ratio) over time, a structural change, which is designed to decrease over-utilization or increase efficiency -- a performance change. In other instances, the impact of change agents on performance can occur directly without necessarily providing a change in the structure of the system, except as a side effect. An example of this would be the availability of Medicare dollars, which provides more money for the care of the elderly and consequently more health care for the elderly, i.e., a change in the direction of increased equity -- a performance change. Thus, the intent of policy research is to assess or predict the impact of past, current, and planned regulatory and financing initiatives on the performance of the delivery system -- in some instances through structure and in other instances on performance directly.

This paradigm is useful in that it can be applied in a variety of ways. Depending on the objectives of the research, it can be used as a conceptual crutch to translate policy initiatives to performance measures; to construct hypotheses directed toward ascertaining the influence of other regulatory, financing, or structural characteristics which could mitigate the intended impact of the policy initiatives; and to design and conduct evaluation efforts to assess the impacts of policy initiatives on the appropriate performance measures, taking these mitigating structural, regulatory, and financing variables into consideration. Such studies begin with the development of a theoretical framework or the specification of testable hypotheses which follow from the problem at hand and the paradigm altered to fit the study under consideration. Assuming that time constraints exist, as is usually the case, two extremes should be avoided in the early stages of any study. The first is underkill or insufficient thought and energy devoted to the conceptual framework which follows from the intent of the study. The most detrimental effect of this extreme is a study which is totally off point. The opposite extreme, namely overkill in terms of undue expenditures of time and energy at this stage of a study, can result in an equally detrimental effect -- namely insufficient time and money remaining to conduct the study.

Flowing from the conceptual framework are various study components such as the units of analysis, the overall time sequencing of the components of the study, the time frame covered by the study in terms of years for which data are to be collected (paying attention to the importance of lagged variables since policy variables usually have a lag in terms of impacting performance or structural variables after their inception), variables on which data are to be collected, data collection methods, and impact assessment methods.

IV. STATISTICAL CONSIDERATIONS IN ASSESSING THE IMPACT OF HEALTH POLICY

A. The Confounded Effects Problem: The general problem of isolating the impact of a particular factor is one of the most frequently encountered problems in empirical research. Although statistical methods do not provide unequivocal answers to questions regarding causality, they are designed to shed light on cause-effect relationships through the analysis of various types of associations and correlations. In policy research, impact assessment is normally the primary objective of the research effort. Thus, the data analyst is normally operating in a multivariate setting where it is desirable to assess the influence of one given variable or set of variables on another variable or set of variables in the presence of a series of related factors, which both interact with variables whose impact is being assessed and partially determine the dependent variables. In regression analysis terminology this confounding of the influences among independent variables as they relate to dependent variables is called (multi)collinearity. The problem of collinearity refers to those interrelationships among independent or predictor variables which tend to vitiate the interpretability of the individual impact of specific (sets of) independent variables apart from their joint impact with other variables.

The confounding of results due to collinearity among independent variables has long been one of the major obstacles to bona fide impact evaluation in non-laboratory research environments. Much methodological research in the area of collinearity and confounding of associations during the recent past has centered on assessment of the severity of the problem. Thus, several measures of the intensity of collinearity have been developed and associated statistical tests can thus provide the researcher with some notion of the presence, overall severity, and possible location of the collinearity (confounding problem). These measures and tests include the determinant of the correlation matrix, statistical procedures to test for the departure of the determinant of the correlation matrix from singularity, multiple correlations among independent variables, and variance inflation factors, among others [1,2,6]. Different estimation methods have been proposed for developing linear models which can be useful in both impact assessment and prediction in the face of collinearity. One of the more practical methods which is gaining widespread popularity in this respect is ridge regression [3,4,5].

Notationally, let Y be a dependent variable and $\underline{X} = (X_1,\ldots,X_p)$ be a vector of independent variables. Further assume that a sample of size N has been drawn, that $\underline{R}_y$ is the p x 1 vector of correlation coefficients between Y and the respective components of $\underline{X}$, and that $\underline{R}$ is the p x p correlation matrix for the independent variables. Denote the determinant of $\underline{R}$ by $\det|R|$ and the $i,j^{\underline{th}}$ element of the inverse of $\underline{R}$ by r^{ij}. Under the hypothesis that $\underline{R}$ is singular (i.e,, "perfect collinearity"), the Bartlett statistic

$$\chi^2(\nu) = c\cdot\ln(1 - \det|\underline{R}|)$$

is approximately chi-square with $\nu = p(p-1)/2$ degrees of freedom, where $c = 1 - N + (2p + 5)/6$. The multiple correlation obtained by regressing the $i^{\underline{th}}$ independent variable on the remaining independent variables is given by

$$R_i^2 = 1 - (1/r^{ii}),$$

and the variable inflation factor for the $i^{\underline{th}}$ independent variable is

$$V_i = 1/(1 - R_i^2).$$

It can be shown that $r^{ii} \to \infty$ as a $\det|\underline{R}| \to 0$, resulting in collinearity expressed by $R_i^2 \to 1$ or $V_i \to \infty$. Assuming that collinearity is not so severe as to prevent inverting $\underline{R}$, or that a generalized inverse method is used, the ordinary regression coefficients (standardized) are given by

$$\underline{b} = \underline{R}^{-1}\,\underline{R}_y$$

and the ridge coefficients are given by

$$\underline{b}'(\lambda) = (\underline{R} + \lambda \underline{I})^{-1} \underline{R}_y,$$

where $0 \leq \lambda \leq 1$ and the optimal ridge coefficients are selected in accord with trade-offs between bias and variance of the coefficients.

Although the above results are not new, they can be useful in policy research studies. First, the chi-square statistic can be used to assess the overall severity of the collinearity problem. If it is not severe, ordinary regression techniques can be used in deriving coefficient estimators which in turn can be interpreted as the (correlative) impacts of the respective independent variables. If the chi-square test indicates that collinearity is severe, then the location of the collinearity problem can be ascertained by examining the multiple correlations and variance inflation factors for the independent variables. If the variables whose impacts are to be assessed are not involved in the collinearity problem then their coefficients can be interpreted in the fashion just described in spite of the fact that other variables may be collinear with one another (assuming that the collinearity is not so severe that it influences the computational accuracy of the matrix inversion method). If the policy variables whose impacts are to be ascertained are involved in the collinearity problem, ridge estimators can be used to provide more stable estimates of the coefficients for these variables than can be obtained through ordinary regression methods. In general, although the ridge estimators are biased, their variances are smaller. Hence, predictive equations derived using the ridge method are generally more stable. Nonetheless, there can be instances where collinearity is sufficiently severe so as to cause significant problems in interpretating ridge coefficients as well as ordinary coefficients, even when generalized inverse techniques are applied in order to guarantee invertibility. If this is thought to be the case prior to sample selection, then the technique described in the next subsection can be of value in selecting observations with the specific intention of minimizing collinearity between a policy variable and other potential confounding variables.

For convenience, this disucssion has assumed that the policy researcher is using a regression method to conduct his analyses. The general problem of collinearity pertains to a variety of techniques, including other more generalized regression methods such as two stage least squares, weighted or generalized least squares, simultaneous equation methods, etc. Nonetheless, it is often wise from a practical point of view to use standard regression methods since although policymakers might have difficulty in understanding them as techniques, the results which follow from the application of such techniques are readily understood -- partly because regression methods are perhaps the most widely used statistical methods outside of experimental research environments.

B. A Matching Procedure to Separate Tangled Variables: The procedure described below was used as part of the currently on-going SSA sponsored study to assess the impact of the Indiana system of prospective reimbursement for hospitals. Although a large number of statistical techniques are being used in the conduct of the analyses for this study, it is once again convenient to discuss the matching procedure from the perspective of a regression framework. The primary policy issue addressed by this study is the question of whether the Indiana system has contained the rate of increase of hospital costs. Since costs cannot be considered apart from case mix and quality, the most important set of dependent or performance variables for this study are cost variables adjusted for case mix and quality. These types of measures and a procedure used to derive adjusted cost variables are outlined in subsection E below. For purposes of this present discussion, it is assumed that the dependent variables have been defined and data are available on them.

The Indiana Prospective Reimbursement system is a controlled charges system which is operated by Blue Cross and the Indiana Hospital Association. Because the system has been in existence since 1959, the paucity of data prior to the program's inception led to the necessity of conducting a comparative study in which the performance of Indiana hospitals is contrasted with the performance of hospitals from states which do not have prospective reimbursement systems. A variety of considerations went into selecting the states from which the control hospitals were chosen. In many ways, the considerations which went into this selection process represent a conceptual analogue of the more quantitative matching procedure described below. In essence, the control states were selected on the basis of statewide health care delivery system characteristics, socio-demographic characteristics, and other mitigating variables which could influence hospital performance. Without going into detail on the various characteristics used to select the control states, four states were selected: Illinois, Iowa, Minnesota and Michigan. Due to budgetary considerations as well as analytical considerations, it was decided that a control group of size 100 would be selected from the 600 potential control hospitals, this represents a control group of the same size as the group of Indiana hospitals participating in the study.

Assume the policy variable Z is dichotomous (as is often the case when one is concerned with the presence or absence of a particular type of policy). That is, define the prospective reimbursement variable as: Z = 0, if the hospital is an Indiana (prospective reimbursement) hospital and Z = 1, if the hospital is a control hospital. Theoretically, there are several types of variables which could influence hospital cost apart from prospective reimbursement. These variables include characteristics such as hospital ownership (for-profit, non-profit, government run, etc.), hospital size, metropolitan versus rural location, population served, and so on. Clearly hospital case mix and quality could also influence hospital costs. However, one of the reasons for reducing the number of control hospitals from 600 to 100 was the additional per hospital cost associated with obtaining the case mix and quality data. Thus, rather than incorporate the case mix and quality measures into the matching procedure, it became necessary to compensate for them analytically after the matching procedure was conducted. Basic data on variables such as those mentioned above were collected for all 600 potential control hospitals. Profile analyses indicated that there were substantial discrepancies between the 600 potential control hospitals taken as a group and the 100 Indiana hospitals taken as another group. Consequently, it was evident that there was substantial collinearity between Z and certain of the confounding variables. The purpose of the matching procedure was to select the sample of 100 hospitals from the potential control group so as to minimize this collinearity and therefore reduce the confounding between Z and the independent variables used in the matching procedure.

Given that the unit of analysis for this study is the hospital, denote the variables which confound the effects of Z by $\underline{U} = (U_1,\ldots,U_p)$. Intuitively, the idea is to select a set of control hospitals which match the Indiana hospitals as closely as possible on the components of $\underline{U}$. Let the Indiana hospital observations on $\underline{U}$ be denoted by $X = (X_1,\ldots,X_p)$ and the potential control hospital observations on $\underline{U}$ be denoted by $\underline{Y} = (Y_1,\ldots,Y_p)$. Thus we have two data arrays

X ARRAY	Y ARRAY
INDIANA HOSPITALS	POTENTIAL CONTROL HOSPITALS
$X_{11},X_{12},\ldots,X_{1P}$	$Y_{11},Y_{12},\ldots,Y_{1P}$
$X_{21},X_{22},\ldots,X_{2P}$	$Y_{21},Y_{22},\ldots,Y_{2P}$
.	.
.	.
.	.
$X_{N1},X_{N2},\ldots,X_{NP}$	$Y_{M1},Y_{M2},\ldots,Y_{MP}$ $(M > N)$,

where X_{ik} = the value of the $k^{\underline{th}}$ variable (U_k) for the $i^{\underline{th}}$ Indiana hospital, and Y_{jk} = is defined similarily for the $j^{\underline{th}}$ potential control hospital. As mentioned earlier, N = 100 and M = 600 for this application.

The problem is to select N observations from the potential control group (the Y array) in order to match them as closely as possible with the N Indiana hospitals contained in the X array. In other words, given that we have a study sample for which Z = 1, the control sample for which Z = 0 is to be chosen such that the multivariate empirical distribution function of the control sample is approximately the same as the multivariate empirical distribution function for the X array.

The procedure begins by selecting for the first $\underline{X}$ observation that $\underline{Y}$ observation which is "closest" to $\underline{X}$ in terms of a distance measure. The distance between the $i^{\underline{th}}$ $\underline{X}$ observation and the $j^{\underline{th}}$ $\underline{Y}$ observation is given by

$$D(\underline{X}_i,\underline{Y}_j) = \sum_{k=1}^{p} (X_{ik} - Y_{jk})^2.$$

Proceeding sequentially in this fashion, a control hospital is chosen for each Indiana hospital. The computer program used to conduct this matching procedure provides a variety of options including the ability to:

1. Force or exclude certain matches on a priori basis.
2. Allow a potential control hospital to qualify as a matching hospital for a specific Indiana hospital only if it passes tests which the user can specify, on a variable by variable basis, necessitating that the $k^{\underline{th}}$ variable for the potential control hospital deviate no more than a certain amount from the corresponding value for the Indiana hospital under consideration, either in absolute magnitude or percentage terms.
3. Weight the separate summands of the distance measure by user-selected weights chosen in accord with the importance of the respective variables in the matching profile.
4. Establish a cut-off value v for the overall distance function so that no matches can occur between study and control pairs which are more than v units apart.
5. Iterate through the potential control data several times, progressively loosening various criteria in order to obtain matches for hard-to-match Indiana hospitals.
6. Standardize the data prior to the conduct of the procedure in order to put all variables on the same scale.

Upon selection of the matched sample, discriminant function procedures or similar profile comparison techniques can be used to assess the overall strenght of the final set of matches. Results could indicate that it may be appropriate to exclude certain matches from the analyses or iterate further in order to rematch certain study cases with new degress of tightness on specific profile variables.

Assuming that the control sample is finalized, denote the performance measure (cost, say) by C. As mentioned earlier, several analytic techniques are being utilized on this project beyond the straightforward dummy variable approach discussed here. Nonetheless, the coefficient of Z in the regression equation

$$C = b_o+b_1U_1 + \ldots + b_pU_p + bZ,$$

is reasonably unconfounded by the remaining terms in the equation, and can therefore be interpreted as the correlative impact of the Z variable. In this instance, the Z variable clearly connotes more than the presence or absence of prospective reimbursement. It symbolizes other potential state level differences between Indiana and the control states. One reason for paying careful attention to the selection of the control states was to minimize the number of other factors which could be involved in Z. Several of the other analyses planned as part of this study and the NSF metropolitan area study described earlier are directed toward separating out the effects of prospective reimbursement from other potential factors which could complicate the interpretation of Z in this respect.

The matching procedure discussed here can be generalized to polychotomous and continuous policy variables by subdividing the range of such variables into several intervals and then applying the matching procedure as if each of the intervals were a dichotomous variable.

C. Sample Selection Based on Minimizing Collinearity Among Relevant Policy Variables: The preceding dealt with a sampling procedure to minimize confounding between a single policy variable Z and a set of mitigating variables denoted by $\underline{U}$. Frequently more than one policy variable must be analyzed in a health policy research study. For example, in the NSF study to analyze the delivery of health care in metopolitan areas, it is necessary to analyze the impact of HMO penetration, different types of Medicaid programs, several different versions of prospective reimbursement, different types of quality assurance programs, etc. Suppose that there are three policy variables to be analyzed, Z_1,Z_2, and Z_3, and that the other independent or mitigating variables are once again denoted by $\underline{U}$. Further suppose that the purpose of the sampling procedure is to provide a sufficient degree of disparity among Z_1,Z_2, and Z_3 so as to allow for the untangling of their differential effects on peformance measures. Further, it is desirable not only to minimize the confounding among the components of $\underline{Z}$, but also between $\underline{Z}$ and $\underline{U}$. Thus, the purpose of the sampling procedure is to minimize the collinearity among the components of $\underline{Z}$ and the overall collinearity between $\underline{Z}$ and $\underline{U}$, taken as total vectors rather than individual sets of independent variables. Phrased in this fashion, the problem sounds something like a new twist to the old cannonical correaltion problem with the difference being the minimization of, rather than the maximization of, correlations. Although cannonical correlation techniques can be helpful in this instance, they do entail transformations which are difficult to interpret from a practical point of view and therefore often are of little value in assessing the impact of policy variables.

In the NSF study, which is also on-going, our sampling purpose was to select a group of 90 Standard Metropolitan Statistical Areas (SMSAs) from the 201 SMSAs which qualified as potential units of analysis for this study. Once again, costs associated with collecting information did not permit us to include all 201 SMSAs in our more intensive analyses. We did collect information on $\underline{U}$ and $\underline{Z}$ in order to select the sample in accord with the criteria just given. Two other considerations entered into selecting the sample. First, performance variables were not allowed to enter into the selection procedure and second, it was necessary to stratify on certain variables in order to be certain that representative ranges for the components of $\underline{U}$ as well as the components of $\underline{Z}$ were encompassed by the sample observations.

Observations from the group of 201 SMSAs were selectively eliminated in order to (1) maximize the determinant of the correlation matrix for (Z_1,Z_2,Z_3), thereby minimizing the degree of confounding among the policy variables and (2) maximize the determinant of the correlation matrix for C_1,C_2,C_3,Z_1,Z_2,Z_3 where the three C_i's are the first three principal components of $\underline{U}$. In general, the number of principal components to be used depends on the extent of the overall variation explained by the first few principal components. The actual procedure used in selecting the sample was heuristic in spite of the fact that the two objectives of the sampling procedure could be rigorously stated. We are currently working on an algorithm to make the selection procedure itself more rigorous. In any event, upon selection of the sample, the confounding of the effects of Z_1,Z_2,Z_3 and the vector $\underline{U}$ have been minimized and the analyses therefore strengthened.

D. The Target Population Adjustment Problem: In studies designed to analyze the impact of specific types of health policies, it is sometimes necessary to specify the consumer population serviced by the health care delivery system in a particular area. Depending on the type of health care services one is concerned with this may be a complex problem since people travel from one geographic area to another for different types of health care. In those instances where the number of consumers is of primary concern (as opposed to specific characteristics of the consumers), it is appropriate to adjust the population in the geographic areas in which the delivery systems being examined are located. The adjustment should compensate for patients who live in such geographic areas yet do not receive their health care within the confines of the delivery system under consideration, as well as for those who travel into the geographic area under consideration to receive care.

Such adjustments are necessary especially in those instances where one is examining per capita measures such as physicians per capita, hospital beds per capita, tertiary care facilities per capita, etc. Once again, using the NSF study as an illustration, it was necessary to adjust the SMSA population for the inflow and outflow of hospital patients. Since the study focuses on short-term general hospitals in metropolitan areas, exlusive of military and Veterans Administration hospitals, it also became necessary to adjust the SMSA population for SMSA residents receiving hospital care at Veterans and military hospitals. The acutal number of SMSA residents who receive hospital care at VA and military hospitals is not easily obtainable, hence the SMSA population was deflated by an estimated factor. The Veterans Administration provided patient origin information which enabled us to compute the percent, P_v, of total bed days attributable to SMSA residents for each VA hospital located in an SMSA. Next, since VA hospitals tend to have longer lengths of stay and may have different occupancy rates than other short-term general hospitals, it was also necessary to adjust by a factor which took this into consideration. For this purpose the ratio, R_v, of VA admissions per bed to other short-term general admissions per bed in the SMSA was used. Total VA beds in the study, B_v, was utilized as the variable to be multiplied by these two factors.

Similar information was obtained from the Army, Navy and Air Force for the military hospitals in each SMSA. Let the corresponding military factors and variables be given by P_m, R_m, and B_m. Finally, denoting the number of non-VA and non-military short-term general beds in the metropolitan area by B_o, the factor used for the VA/military adjustment is:

$$Q(v,m) = B_o/(B_o + P_vR_vB_v + P_mR_mB_m).$$

Patient origin studies were also avaiable for roughly half of the SMSAs. For these SMSAs two percentages were computed: the percentage of hospital admissions attributable to non-SMSA residents, P(in), and, for all the residents of the SMSA, the percentage who received hospital care outside the SMSA, P(out). Let the SMSA population be given by T. Then the adjusted population which reflects the number of consumers in the market served by the short-term general hospitals of a particular metropolitan area is given by:

$$T(adj) = T \times Q(v,m) \times \{1+P(in)\} \times \{1-P(out)\}.$$

In those instances where the inflow and outflow percentages were not available from patient origin studies, we substituted expected values based on predicting percentages in terms of demographic variables for the SMSA and surrounding counties.

The intent of this subsection has been to demonstrate that certain key variables in policy research must be carefully considered, particularly when they deal with a service oriented delivery system whose consumers are not necessarily located in the same area as the delivery system. The example given does not necessarily exhibit the best possible type of adjustment factor (several factors which were considered more appropriate were not available due to constraints on obtaining data), rather it is intended to exemplify a typical problem (and solution) encountered in policy research.

E. The Case Mix/Quality Adjustment: In Section III it was indicated that for both the SSA sponsored and NSF sponsored projects to analyze the Indiana Prospective Reimbursement system and the delivery of health care in metropolitan areas respectively, data are being collected on a series of specific diagnostic and procedural patient cohorts, on an individual hospital basis. Suppose that one wishes to construct a process measure of quality, Q, on the basis of cohort specific patient management indices (variables characterizing how patients are treated -- such as those discussed in Section III.C in the pneumonia and acute bronchitis pediatrics example). Considering just one diagnostic category, suppose there are r patient management indices, $P_1, P_2, \ldots, P_r$, where higher values for each P_i are regarded as appropriate treatments. As an illustration, P_1 could be the percent of children receiving a diagnostic x-ray if the category is the pneumonia and acute bronchitis classification just mentioned. One overall measure of quality for a specific diagnostic category might be the sum

$$Q = W_1P_1 + W_2P_2 + \ldots + W_rP_r,$$

where the weights are chosen in accord with medical considerations. Certain multivariate techniques, including variants of principal component analysis, can also be used to compute the weights in accord with statistical considerations. Assuming that such quality indices are constructed for each diagnostic category, then a weighted quality index can be computed for an entire hospital by taking a weighted sum of the form specified above where the P_i's are replaced by Q_i's and the index i pertains to the diagnostic categories. The W_i's would be functions of both the frequency of the particular diagnostic category and the intensity of the category as measured by both medical and cost considerations. Assuming such an overall index of quality has been constructed (the actual process of doing so is both costly and time consuming) for hospitals, let C* = a standardized cost variable and Q* = the standardized quality variable. In this instance, cost and quality are on the same scale -- i.e., zero mean and unit standard deviation. Define the quality-cost trade-off variable as

$$Y = Q^* - C^*.$$

High positive values of this variable represent a "good buy" in terms of quality care for the consumer dollar. Lower values, especially negative values connote poor quality care relative to cost. Interpreting Y as a cost variable adjusted for quality, it can then be used as a dependent variable in order to study the impact of a policy variable on adjusted cost. As indicated earlier this measure is being used in the Indiana Prospective Reimbursement study in order to assess the impact of the Indiana system on cost adjusted for quality and case mix (although the actual case mix adjustment has not been discussed here, it is analogous).

V. OPERATIONAL CONSIDERATIONS IN THE CONDUCT OF HEALTH POLICY ASSESSMENT STUDIES

A. The Theoretical Model: The topic of development of a conceptual framework during the design state of a policy research project is a controversial one. The importance of establishing a sound conceptual base as a starting point for an empirical project cannot be overlooked. Yet, if a project is to proceed on schedule it is important that this phase not be overemphasized needlessly. The policy researcher can usually complain that policy questions are not asked well in advance of when the answers are needed. Such is the nature of our political system, however, and the researcher must cope with the fact that a large amount of time and energy cannot be spent designing long term studies needed to address short run policy questions. This is not to say that there have not been instances in which largely policy studies have been appropriate. Thus, although the standard axioms of good research, such as the need to develop an overall model and operational hypotheses, must be heeded, they must also not preclude the possiblity of providing at least some basic information to policymakers at the time policies must be made. This is perhaps one of the most frustrating aspects of policy research to the individual who is schooled in classic research methodology. Basically, it means that even the more "creative" aspects of such work must be completed in accord with a tight time schedule. This is rarely easy and at times not possible; nonetheless, it is a crucial ingredient to a solid approach to most types of policy research.

B. Computer Considerations: The importance of data management and proper utilization of computer technology cannot be overemphasized in empirical studies to assess policy impacts, especially empirical studies with large data bases. The past decade has seen a remarkable development in computer technology and there is every reason to believe that the next decade will witness a greater degree of advancement. The role of quantitative researchers and methodologists is increasing due in no small way to the "computer revolution". One of the strongest temptations which researchers and statisticians must overcome is the drive to get to the heart of the analyses as quickly as possible without paying sufficient attention to data collection and processing activities prior to the analysis. Data analysts should consider the collection, editing, and management of data as part of the actual analysis process. This means an active involvement or at least a keen awareness of all data processing activities which took place prior to the derivation of results using statistical or mathematical methodologies. Many economies can be realized in conducting empirical projects under a tight schedule by careful planning and management of data collection and processing.

On the analysis side, since multivariate methods are frequently necessary in policy research, one should take advantage of the utility of the correlation matrix approach to conducting multivariate statistical analyses. In particular, the correlation matrix (covariance matrix, cross product matrix), the mean vector, and the standard deviations (variances) can be used as input to a wide variety of multivariate procedures without any need to access the raw data, except to compute the statistics. Such multivariate procedures include regression analyses (ordinary, forward and backward stepwise, two stage, etc.), discriminant function analyses (k-group and stepwise procedures), principal component analyses, canonical correlation analyses, factor analyses, and various types of multivariate T^2 procedures, among others. Not only can this amount to considerable savings in terms of time, but for studies where a large number of multivariate analyses must be run, it can also amount to substantial savings in computer costs.

C. Application of Several Statistical Methods: For a particular hypothesis testing problem there are frequently two or three strong contenders for the position of the 'best test'. Rather than spend a great deal of time deciding which method should be used or arguing in a somewhat academic fashion that a particular procedure should not be used, it is sometimes simplest and safest to utilize all of the top two or three contenders. This not only has the virtue of heading several of your critics off at the pass, but it also can serve to lend credibility to research results. In this regard, it is wise to apply both parametric and nonparametric procedures in testing the same hypothesis. If the results achieved using the two different methods agree, then the overall results have even greater credibility. If they do not agree, the statistician is not much worse off than he was when someone (perhaps himself) was telling him to use one procedure and someone else (perhaps his alter ego) was telling him to use the second procedure.

Concluding Remarks

In summary, policy research studies, be they in health care or other areas, are beset with problems which many other types of studies do not have. Yet, although this can result in curbing certain methodological considerations, it represents a different type of challenge to the methodologist. In short, there is a need to deal pragmatically with some degree of imperfection and yet derive objective conclusions which will constructively influence the direction of public policy. Further, the integrity and objectivity of the researcher who conducts such studies is not tainted by his actively advocating policy changes based on results which he has derived or been close to as a researcher.

References

[1] Farrar, D.E., and Glauber, R.R.(1967): Multicollinearity in Regression Analysis: The Problem Revisted, *The Review of Economics and Statistics*, 49, 92-107.

[2] Haitovsky, Y.(1969): Multicollinearity in Regression Analysis: Comment, *The Review of Economics and Statistics*, 51,486-498.

[3] Hoerl, A.E., and Kennard, R.W.(1970);Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics*, 12,55-67.

[4] Hoerl, A.E., and Kennard, R.W.(1970); Ridge Regression: Applications to Nonorthogonal Problems, *Technometrics*, 12-69-82.

[5] Marquardt, D.W.m and Snee, R.D.(1975): Ridge Regression in Practice, *The American Statistician*, 29,3-20.

[6] Snee, R.D.(1973): Some Aspects of Nonorthogonal Data Analysis. Part I. Developing Prediction Equations, *J. Qua. Technol.*, 5,67-69.