REGRESSION AND CASE STUDIES OF PUBLIC PROGRAMS: DISCREPANT FINDINGS AND A SUGGESTED BRIDGE*

David P. McCaffrey

David F. Andersen

Paul McCold

Doa Hoon Kim

Nelson A. Rockefeller College of
Public Affairs and Policy
State University of New York at Albany

* The comments of Fred Thompson, Irene Lurie, and Richard Nunez are gratefully acknowledged. This was a fully collaborative effort.

ABSTRACT

Case studies of regulatory and social programs suggest that policy systems are dynamic. In the systems described, outcomes depend on how numerous variables interact over time, and feedback among variables -- "simultaneous" causation over multiple time periods -- is more a rule than an exception. However, the most influential evaluations of public programs are studies using multiple regression. A recognized limitation of multiple regression is its <u>relative</u> insensitivity to multiperiod strategies, feedback among variables, and other dynamics. Accordingly, we maintain, the findings of regression-based and case studies commonly conflict. Simulation modelling can serve as a methodological bridge between case studies and regression-based studies of policy systems, improving theoretical models of the systems and providing a way to evaluate the robustness of alternative regression models. The results of some early experiments along these lines are presented.

SECTION I: INTRODUCTION

Case studies of regulatory and social programs suggest that policy systems are dynamic. In the systems described, outcomes depend on how numerous variables interact over time, and feedback among variables -- "simultaneous" causation over multiple time periods--is more a rule than an exception. For example, weak regulatory strategies or poor management often prompt opposition leading to strong regulation and management reform; "unlikely" coalitions form routinely; and so forth (McCaffrey, 1982: 402-403; Bacow, 1980: ch. 5; Policy Studies-Urban Systems, 1981: Vol. III).

However, the most influential evaluations of public programs are not case studies, but rather studies using multiple regression. A recognized limitation of multiple regression is its <u>relative</u> insensitivity to multiperiod strategies, feedback among variables, and other dynamics. Statistical and data problems tightly constrain the number of variables, interactions, and feedbacks that can be meaningfully built into regression-based analyses of public programs (Meadows, 1980: 40-46). Also, the research strategies associated with regression techniques emphasize cross-sectional, "thin" description of large numbers of individual cases, rather than longitudinal, "thick" description of fewer cases. Such distant examination of even many cases will detect only the most obvious dynamic relationships, and even then not provide data that are sufficiently rich to explain them.

An assumption of regression analysis is that variable interactions and feedbacks, nonlinearities, and other dynamics are not so important as

to prevent the necessarily simplified regression model from picking up the strongest signals in the policy system. The aspiration of regression analysis is to detect central factors that dominate other relationships, however complex.

Arguably, the method works quite well, for it appears to uncover strong, stable signals in numerous policy systems. Regression-based studies find, fairly consistently, that Occupational Safety and Health Administration inspections do not reduce industrial injury rates or increase investment in safety-related capital equipment (Smith, 1979; Viscusi, 1979; McCaffrey, 1983). They find, fairly consistently, that state certificate of need (CON) regulation of hospital investment does not reduce capital investment (Steinwald and Sloan, 1980; Sloan and Steinwald, 1980; Joskow, 1981; Fuangchen and McCaffrey, 1983); that variability in school expenditures or teaching methods are not associated statistically with variations in student performance (Hanushek, 1981); and that the organizational characteristics of medical care, or pschological or ethnic characteristics of individuals, are associated in only trivial statistical ways with use of physicians' services (Mechanic, 1979). The list of stable findings from regression analyses could be lengthened almost at will.

But what is striking is that these stable statistical findings commonly conflict directly with analyses using case study or other "thick" descriptive techniques. As we will see below, intensive studies of firms suggest that OSHA <u>does</u> improve safety performance and increase safety-related investment. Furthermore, case studies suggest that certificate of need regulation does affect the medical technology market and hospital investment (E.g., see U.S. Office of Technology Assessment, 1981: 32-38; Howell, 1981); that school characterististics do affect student performance (Rutter et al., 1979); and that organizational and psychosocial factors do substantially influence physician utilization (Mechanic, 1979). Analysts sometimes label the case studies as "interesting" anomalies, but maintain that greater weight should be given to regression analysis of carefully selected samples (E.g., see Sloan, 1982: 200). We question this argument because the discrepancies between case studies and regression results are too common and systematic to reflect only the anomalous qualities of the cases.

Instead, this paper maintains that regression-based studies, however stable their results, are not definitive. Practically none of the regression based studies of occupational safety regulation, and very few of the studies in other areas, are grounded in credible theories of how the policy systems work. Although the empirical results are consistent, it is therefore difficult to know what the results mean. The problem exists because the data used to operationalize variables measure the variables only grossly or ambiguously. Or, worse, the variables themselves have no theoretical meaning, serving only as rough proxies for variables or processes that are undefined or unknown, but suspected to be somehow important.[1] The gaps between data and variables, or the ambiguities of the variables themselves, mean that even stable empirical results tell us little.

Section II of this paper illustrates these theoretical and empirical problems in detail, focussing on occupational safety regulation. Other policy areas could have been easily used (For extended discussions of similar problems in the research on hospital regulation, education, and physician utilization, see McCaffrey, Andersen, McCold, and Kim, [1984]).

Section III outlines a research strategy, relying on dynamic simulation modelling, that could ease some of these problems. The first element of this strategy is to construct simulation models of a given policy system using both quantitative and qualitative data. Simulation modelling is a method designed specifically to examine and structure dynamic causal relationships; its distinctive strengths are those qualities in which regression methods are weak. The method is also a way to structure in theoretically coherent ways what otherwise would be disparate, noncomparable case studies. The second element of the strategy uses the simulation model--and variations on it--to generate synthetic data on the simulated "reality." Regression models can then be used to estimate properties of the simulation model, and the relative success of alternative regression specifications in describing the known model can be noted. The two elements of the research strategy are related, but each has a distinctive advantage. The first element would improve causal mapping of the policy system; the second element would provide a way to test the robustness of alternative regression models prior to data collection for an evaluation project.

SECTION II: RESEARCH ON OCCUPATIONAL SAFETY REGULATION AND OTHER POLICY SYSTEMS: THEORETICAL AND EMPIRICAL ISSUES

OCCUPATIONAL SAFETY REGULATION

Since 1971 the Occupational Safety and Health Administration (OSHA) has enforced standards for working conditions using inspections and fines. Most of the empirical studies evaluating the impact of the Occupational Safety and Health Administration present a "theory of work injuries" (Smith, 1979: 147-151); a "conceptual analysis of occupational health and safety regulation" (Viscusi, 1979: 118-122); or a "model of injury rate determination" (Mendeloff, 1979: 98-102). These theoretical sections serve as interpretations of the results, if the results are in the predicted direction and statistically significant. The theories focus on the associations between occupational injury rates (or injury rate changes) and OSHA inspections, industry, the size of firms, firms' employment changes, and firm-specific factors. Some studies have also examined the impact of workers' compensation benefit increases on injury rates, but for a variety of reasons these studies have not been integrated into the study of OSHA's impact.

Empirical findings regarding these variables are quite stable, with a mix of positive, negative, and negligible associations. But either the data used to measure the variables are quite incomplete, or the variables themselves have no clear theoretical relationship to injury rates.

Therefore, when the studies try to estimate how a policy variable--usually the OSHA inspection variable--is related to injury rates, the relationship is very ambiguous. Or, when analysts only try to "control" for variables so that the independent effects of the policy variable can be assessed, they cannot in fact assume that control variables do not interact in unmeasured ways with the policy variable because they have only the vaguest idea as to the dynamics of either the control or policy variables (See Footnote 1 above). Consider the results for each variable.

Occupational Safety and Health Administration Inspections

Analysts are interested in the penalties that firms expect for violating OSHA's standards for working conditions. OSHA could induce firms to invest more heavily in safety and health capital by increasing the probability of inspections and/or the penalties associated with standards violations (Mendeloff, 1979: 94, 108; Viscusi, 1979: 123-125; Smith, 1979: 149; McCaffrey, 1983: 132-133). The studies examine how the rates of inspections in industries, the occurrence of inspections in firms, or penalty levels affect injury rates.

The obvious ways to assess the impacts of OSHA inspections, since reliable pre-OSHA and post-OSHA data are unavailable, are to compare the injury rates of industries with relatively high and low rates of inspections, or the injury rates of inspected firms with those of uninspected firms. Several studies have found that high inspection rates in industries are not associated with relative injury rate reductions. Several analysts are skeptical of the value of results for an aggregate category like "industry." Robert Smith, for example, suggested that it is relatively futile to estimate the effects of OSHA from industrial data because the potential effect of OSHA is relatively small and "likely confined" to inspected firms (1979: 147, 169); thus, analysts prefer to use firm-specific data rather than industrial data.

One cannot meaningfully compare inspected and uninspected firms. Comparing them would require that inspected and uninspected firms behave in similar ways, but because inspections can be triggered by employees' complaints, catastrophes, growing hazards, labor-management problems, and so forth, inspected firms may be qualitatively different from uninspected firms. Thus, analysts have opted to compare rates of firms that have been inspected at different times of a given year--usually March-April inspectees with November-December inspectees. (March-April inspectees are used rather than January-February inspectees so that the inspection would not have been prompted by temporary increases in injuries reflected in the prior year's injury rate). Although the late inspectees would presumably be similar in other respects to the early inspectees, their annual injury rate would not be affected significantly by the November or December inspection. Thus, the impact of the inspection on the annual injury rate can be estimated. Quite consistently, the inspection variable is not associated with relative declines in injury rates.

The inspection variable for firms comes from the Bureau of Labor Statistics survey of injuries and illnesses, and represents the month of the first inspection that a firm had in the year (Zero if there was no inspection). This inspection variable does not articulate possibly

important variations in enforcement. The variable does not indicate if the inspection was a routine "general schedule" inspection, or whether it was prompted by a major accident, an employee complaint, or was a follow-up visit to check for abatement of violations cited in previous inspections. The variable does not record multiple inspections, nor does it indicate if the inspection resulted in any penalties.

Finally, even though the inspection data are reported by month, variables which one wants to correlate with the inspection data are reported only by year, and so one cannot measure temporal sequences. Studies test for associations between changes in injury rates and the occurrence of inspections. However, the resulting associations do not indicate whether (1) the inspection altered injury rates (e.g., by reducing the injury rate, or by increasing the injury rate by increasing the reporting of injuries); (2) whether the inspection was induced by a prior increase in injuries; or (3) whether both (1) and (2) occurred in the firm. Indeed, it is not at all unreasonable to suspect that all of these tendencies are real. Interpreting non-significant statistical effects is difficult when, for example, inspections might improve the level of safety in the firm but simultaneously sensitize workers and those who record injuries to safety problems. (In a related observation, W. Kip Viscusi suggested that increases in safety-health investment may not reduce injury rates as workers will respond to safer conditions by "diminish[ing] their level of safety-enhancing actions...that affect either the probability of an accident or the size of the loss...For example, workers may get more careless if the company adds guards or safety cables to the machine" [p. 118]. The point, however, does not figure in his statistical estimates of the net impact of OSHA).

The ambiguities of the inspection data become troublesome when one contrasts studies of OSHA's impacts which use multiple regression with those using other methods. The conclusion commonly drawn from the null regression results for the inspection variable is that the expected penalties for violating OSHA's standards are too low to justify substantial additional safety investment. W. Kip Viscusi, in fact, analyzed a regression of industries' safety-related investments and reported that a greater frequency of inspections was not associated with higher investment in safety-health capital expenditures from 1972 to 1975 (1979). Another conclusion commonly drawn is that OSHA's standards may not address true safety problems of firms.

However, studies which directly ask management or workers about OSHA's effects on safety conditions or operations, or look at disaggregated investment data, imply that OSHA does affect firms' behavior and their level of safety. After studying plants unionized by the International Association of Machinists, Kochan, Dyer, and Lipsky wrote that since OSHA's establishment "management has assigned a higher priority to plant safety, the ability of the union to influence management decision making on safety issues has increased, and the role of union-management safety committees has been bolstered." They reported that 69 percent of the workers, and 48 percent of the managers, said that OSHA had a "strong" or "very strong" impact on safety consciousness and responsiveness in the plants (1977: 5, 36, 76-77). A survey of chemical

workers by Cambridge Research Reports for the Shell Oil Company reported that 64 percent of the workers said that safety conditions had improved in their plants in the 1972-1978 period (1978: 110), and that OSHA was a contributing factor. (See also Freedman, 1981).

Also, an Arthur Andersen and Company survey of 48 large firms, which is the most rigorous study of regulatory costs to date, found that the 48 firms spent $68 million in capital outlays directly related to OSHA, and $184 million in OSHA-related expenditures overall, in 1977 (1979: Section 8, p. 5). The report added that the firms had higher OSH-related capital costs in earlier years, which included the period studied by Viscusi (1972-1975). The Arthur Andersen study, which involved a detailed analysis of 48 firms' investment decisions in 1977, and less detailed analysis of earlier decisions, directly conflicts with Professor Viscusi's, which involved pooled cross-sectional regressions of aggregate industry data, with inspection frequency being the OSHA policy variable.

Thus, multivariate studies examining aggregate industry investment, injury rate, and inspection data, or data on firms' injury rates and the occurrence of inspections, find that OSHA's activities are not associated with injury rate changes. Those studies examining particular groups of firms in great detail, focussing on a variety of practices and decisions within firms, find just as consistently that OSHA's activities do affect firms' behavior and the level of safety in firms. (The "level of safety" is not the same as "injury rates" because the reporting of injuries may vary for reasons other than the level of safety. But this would mean that focussing solely on injury rates would be an inappropriate way to evaluate OSHA's impact on occupational safety).

One could argue that these effects may be limited to specific classes of firms--e.g., the firms unionized by the International Association of Machinists, chemical workers, or the large firms studied by Arthur Andersen and Company. But this would mean that multivariate studies are mis-specified, leaving out important variables that do magnify OSHA's effects. The argument would also--for several other reasons--be weak. (For example, the 48 firms studied by Arthur Andersen accounted for nineteen percent of all capital investments made in 1977 in the United States; this would be a very large "atypical" class of firms [Arthur Andersen Report Executive Summary, p. 7]).

Another possibility is that regulation affects firms' operations and behavior in more complex ways than can be realistically described by multivariate regression studies of national data, or of rather skeletal firm-specific data. Regardless of whether one regards OSHA's possible effects as desirable or perverse, it appears that one will not understand their dynamics except through alternative methods.

Examination of the other variables included in studies of OSHA underscores this point. These are typically entered as control variables; however, as alluded to earlier (See footnote 1), their connections to injury rates are so ill-specified that we cannot assume that they are independent of OSHA's effects. These other variables are now discussed.

Industry Variables

Occupational safety hazards vary by industry. Some technologies are more hazardous than others. The marginal costs and benefits of safety controls also vary by industry. Time series data which indicate the

dominant type of technology, or safety controls' costs and benefits, are often not available at the industry level and hardly ever available at the level of the firm.

Also, trade associations, insurance companies, or other organizations, for any number of reasons, attend to safety problems in some industries more than in others. For example, the chemical industry is reportedly criticized far more heavily for--and is more conscious of--toxic waste disposal problems than the primary metals industry, although the industries' waste outputs are similar, because of the chemical industry's attachment to Love Canal and similar cases (Wall Street Journal, June 30, 1983: 58). A study by Lawrence Bacow found that labor-management safety agreements are implemented more vigorously in the auto industry than in the steel industry for several reasons, although both industries have formally strong agreements (Bacow, 1980: ch. 5). Detailed comparative studies of industry attention to safety matters (such as Bacow's) are rare. Other industry effects, in addition to industry technology or varying attention to safety matters, may be consequential as well.

Analysts try to control for industry effects by entering dummy variables for industry (usually the two-digit Standard Industrial Classification level). Note that this procedure does not indicate what "industry effects" are; it only lets us assess whether or not some non-specified industrial factors are associated with injury rates. It does not let us assess whether OSHA contributes to industry effects by--for example--giving unions in certain industries a "club" to enforce safety agreements; or by affecting the cost and effectiveness of safety equipment by stimulating equipment manufacturing for certain industries; or through several other plausible channels, none of which are adequately described by an inspection variable.

Employment Size

Injury rates vary by employment size in a curvilinear way. Small firms (1-99 workers) and large firms (500 or more workers) tend to have lower injury rates than firms of intermediate size. One explanation commonly noted in the quantitative research is that small firms closely monitor safety conditions, and that large firms are likely to have injury-reducing safety programs. No one really knows if this explanation is accurate. Alternatively, true injury rates could be inversely related to firm size because of economies of scale in safety work, but small firms might have distinctively incomplete injury recording. Data on safety record keeping violations in the product safety area are consistent with this view (Linneman, 1981: 474).

The size category variables do not indicate the size-related factors that influence firms' injury rates. Rather, the size variables only indicate whether or not some non- or ill-defined size-related factors appear to influence injury rates. If, for example, OSHA reduced true injury rates in distinctively hazardous small firms, but simultaneously led the firms to tighten up distinctively incomplete injury recording, one could easily find no net change in injury rates. To conclude that OSHA did not affect firms' behavior would be an error.

Increases in Employment

Increases in employment are associated relatively consistently with increases in injury rates. Changes in employment in firms are usually

represented by a ratio of employment in a year t+1 to employment in a year t (EMP_{t+1}/EMP_t); changes in employment in industries are often represented by new hire rates. Again, while these variables appear to "control" for employment increase effects, there is very little research on what these effects are. Employment increases may be associated with an influx of accident-prone or inexperienced workers; with an influx of young workers who are more willing to complain about job hazards, report injuries, and request inspections; with an increase in the ratio of production workers to supervisory/staff workers in firms (increasing rates of injuries per 100 employees relative to other firms); and other factors. While time and time again employment increase effects appear important, there are no studies that would give us a reasonable way to choose among these or other interpretations of the employment increase variable.

"Firm-specific" Factors

Firms in certain industries or size categories, with certain employment level changes, or inspected by OSHA, undoubtedly have other characteristics associated with injury rates. Particularly outmoded equipment, labor-management difficulties, or characteristics of the plant's immediate location are possibilities. These factors should be reflected in each year's injury rate. One way to control for such factors in a year t is to enter the lagged injury rate as a dependent variable (Smith, 1979: 149). Consistently, this variable accounts for 65-90 percent of the variance "explained" in studies of occupational injury rates, with the figure increasing with sample size (Smith, 1979;

McCaffrey, 1983. Data on the proportion of explained variance due to lagged injury rates are available from McCaffrey). Such a variable does not indicate which plant-specific conditions tend to influence injury rates; it simply "controls" for such conditions. Thus, note that the variable which dominates the results of multivariate studies of occupational safety regulation is a proxy variable with no theoretical referent at all. Imagine any plausible interaction between unique firm circumstances and OSHA's existence--such as proximity to an OSHA area office and the information it distributes, aggressive individuals who commonly threaten to "call in OSHA," managers who for reasons of personal biography stress safety, etc.--and these interactions could contribute to this variable in ways not picked up by an OSHA inspection variable, or an inspection-lagged injury rate variable.

Workers' Compensation Benefits

A final factor thought to affect injury rates is the liberalization of states' workers' compensation laws, although the effects are ambiguous. Benefit liberalization may decrease injury rates as employers strive to reduce injuries to avoid higher insurance premiums. Conversely, benefit liberalization may increase injury rates as workers become more willing to report injuries, claiming compensation. Studies of the association between benefit liberalization and injury rates report, fairly consistently, that benefit liberalization is associated with increased injury rates (Worrall and Appel, 1982; Butler and Worrall, 1983; McCaffrey, 1983). None of these studies have looked closely at the behavioral dynamics underlying the statistical associations--whether the

first effect does not exist; whether the first effect exists but is somewhat offset by the second; or whether other factors might explain the benefit liberalization-injury rate association. Nor have the studies simultaneously considered the workers' compensation variable with OSHA, industrial, or employment variables. Thus, we really do not know how workers' compensation liberalization ought to enter into a theory of occupational safety regulation.

Summary

The results of research on occupational safety regulation are quite stable, showing a mix of positive, negative, and negligible associations. However, the results say surprisingly little about the dynamics of occupational injury regulation. The data used approximate only ambiguously the variables of interest. In most cases the operative variables themselves are not specified or are ambiguously specified, so it is extremely difficult to interpret their individual effects or know of their possible interactions with other variables.

Porter, Connolly, Heikes, and Park (1981) outlined four types of demands that might be made of regression. We might ask regression to describe the characteristics of a single data set; to predict the characteristics of future behavior in the policy system; to causally model (or "explain") the patterns of behavior in the policy system; or to causally predict how future manipulated shifts in one or more variables in the system will affect other variables in the system. Where does the research on occupational injury rates leave us in terms of these levels of policy analysis?

The research does not describe the characteristics of single data sets, because it does not describe the specific factors underlying injury rate patterns in single data sets. Consequently, although one can predict how statistical associations will turn out in additional data sets, one is not predicting the dynamics of injury rates at all. One is predicting only that the current ambiguities will persist. A fortiori, the area has no basis on which to develop causal models or causal predictive models of occupational safety regulation, or appropriate occupational safety regulation policy.

The problems of research in occupational safety and health regulation appear in numerous other policy areas.[2] These common problems involve the relative deemphasis of longitudinal dynamics in cross-sectional (or even longitudinal) regression studies; the extent to which regression results are not linked uniquely to theories of how policy systems operate, or do not suggest what adequate theories might be; and aggregation of variables in ways that may level important sources of variation among variables.

Detailed case studies, which distinctively avoid these problems (and, taken alone, have other problems) often directly conflict with the regression-based research in other areas, just as the two types of research on OSHA conflict. The case studies indicate that certain variables are related strongly in ways that are somewhat complex and unfold only over time. Regression analyses of cross-sectional or even of thin longitudinal data detect few of these relationships (McCaffrey, Andersen, McCold, and Kim, 1984). It is reasonable to suspect that the

discrepancies between the two sets of findings reflect regression's relative insensitivity to dynamic or longitudinal relationships. Certainly the evidence for this view is not definitive. However, the evidence is strong enough to lead us to believe that the marginal payoffs of shifting some effort to process-sensitive analytical methods may be quite high. As will be discussed in the next section, the payoffs would be both conceptual and technical. The conceptual payoff would be more satisfactory attention to theoretical modelling. The technical payoff would be a procedure to determine how well particular regression models do describe the characteristics of policy systems.

SECTION III: RESPONSES TO THE PROBLEMS: SIMULATION MODELLING AND SYNTHETIC DATA EXPERIMENTS

This section discusses two ways to cope with the methodological problems discussed in this paper. The first activity involves extending case-study analyses through the creation of dynamic simulation models of policy systems. The purpose of this activity is to describe more adequately how events, constraints, and feedbacks emerge in policy systems like occupational safety or hospital regulation, and to identify underlying structural properties of the systems. The second activity involves an attempt to estimate experimentally the conditions under which dynamic feedback driven systems may present problems for regression analysis.

CONSTRUCTING DYNAMIC SIMULATION MODELS OF POLICY SYSTEMS

Case studies and other types of qualitative research suggest that policy systems are dynamic and ought to be observed longitudinally. However, regression based methods do not handle such complexity well. Because the use of regression analysis requires large numbers of observations, analysts are necessarily constrained in the time period and richness of the data collected for any single case. Analysts tend to restrict exploration of causal behavior in policy systems, and modelling of the systems, to accomodate what can be roughly tested with these quite skeletal data. Even where the theoretical models are more fully developed, the data are not sufficiently rich to allow adequate tests of the models. Thus, the theoretical and empirical work on the policy systems are, at best, loosely coupled.

Dynamic simulation modelling is a method that can link regression-based results and the large amounts of case material in these areas in a theoretically disciplined way. Like case studies, dynamic simulation modelling uses diverse sources of information. The method uses the numerical data available to regression analysis, but the method also draws on interviews, participants' judgements, and other non-numerical data in constructing models of how the policy systems function. The method goes beyond case studies, however, in providing some clear principles for structuring the information into a theoretical model of the system. These principles are derived from feedback loop theory (Forrester, 1961, 1980; Randers, 1980), and force analysts to specify clearly--in both graphic and mathematical form--the multiple

connections of variables and parameters in the system, the time periods in which these effects play out, and how the multiple connections in the system either amplify or dampen certain behaviors. The behavior of the model can be projected into the future through computer simulation. There is a constant interaction between the implications of the model and behavior in the "real" world; discrepancies force the analyst to refine the theoretical structure of the model.

Once the evidence gathered through essentially case study methods has been assembled into a formal simulation model, several features not available through simple case studies are available. First, most of the theoretical richness of the case study can be retained. This is because the relatively flexible mathematical form of simulation models allows for the inclusion of a large number of "hard" and "soft" effects. Second, the simulation model will contain a mathematically explicit causal structure. Whereas case studies can be rich in descriptive detail, simulation models force the analyst to posit explicit causal hypotheses concerning how the policy system operates. These causal hypotheses can then be tested further via regression or other forms of statistical or empirical analysis. A third feature of the simulation model is that all assumptions of the model are available readily for inspection since they have been cast into an unambiguous mathematical form. The model (and hence the causal theory being tested) must be absolutely consistent and logically complete. These features, often neglected in simple case studies, must appear in a formal simulation model or the model simply will not compile and run on the computer.

A fourth feature of the simulation model is that the behavioral consequences over time of the causal hypotheses can be known with certainty. That is, by actually running the model and creating a simulated scenario, the model itself will generate synthetic time series data that arise solely from the causal assumptions built into the model. By running the model and observing if the simulated behavior is reasonable, the analyst can gain some additional insight into the reasonableness and logical consistency of the assumptions built into the model. A fifth feature of a case study recast as a simulation model is that analysts can easily insert alternative causal hypotheses into the model, and then examine the implications of such changes over time.[3]

Thus, the first benefit of simulation modelling is that it can use diverse sources of information, including descriptively rich case studies, to theoretically model policy systems in a rigorous way.

USING SYNTHETIC DATA EXPERIMENTS TO TEST A REGRESSION MODEL'S ROBUSTNESS

The second benefit of simulation modelling is that synthetic data experiments, using the simulation model, can explore in mathematically precise ways the robustness and sensitivity of regression models.

The basic ideas behind synthetic data experiments are, in principle, simple. First, the simulation model--or a plausible, preliminary model--is run a large number of times under stochastically varying conditions to create a large number of observations. These observations are called "synthetic data." Because the structure and stochastic character of the model is known, the characteristics of the data generated by the model are known as well. The experimenter ought to be able to design a

regression model--relying on simultaneous equations, lagged variables, and variable transformations if necessary--to recover the known structure and parameters of the data generating model.

The nature of the synthetic "reality" is easily changed by altering the simulation model; similarly, the specification of the regression model may be changed easily. These changes may include "corrupting" the data generating model by inserting amounts of measurement error, inserting odd correlations within independent variables, planting plausible misspecifications in the regression model, and other problems that are suspected to exist in real data and analysis. Because the structure of the data generating model is known, the experimenter can monitor in a mathematically precise way how well alternative regression models recover the simulation model's structure and parameters under these imposed measurement and specification problems. Execution of one such experiment usually takes from five minutes to one half hour to complete.

Such synthetic data experiments bear a family resemblance to Monte Carlo simulations and often prove interesting because the statistical properties of feedback-driven systems differ from those assumed by ordinary least squares regression and most of its derivatives (Andersen, 1981). For example, Peter Senge (1977) has used this experimental design to investigate how well regression models can recover parameters from a dynamic system when the observed independent variables have been corrupted by relatively small amounts of measurement error. For feedback driven systems, the estimation models performed well for no or minute quantities of measurement error, but the estimation model's performance degenerated rapidly under small to moderate amounts of measurement error (See also Johnson, 1980).

Also, Mass and Senge have used a similar design to investigate several puzzles that arise when tests of statistical significance are used to include or exclude variables when estimating parameters from a dynamic feedback-driven system. They concluded that simple tests of significance may be poor indicators of which variables are important and unimportant in determining the behavior of policy systems. (Mass and Senge, 1980) Similarly, Luecke and McGinn (1975) used a modified Markov chain to investigate how well regression might estimate schooling effectiveness. Luecke and McGinn discovered that even when expenditures for educational inputs such as better teachers and schools are known to have an effect on student achievement in a simulated reality, estimation models used in major evaluations of education still produced null results.

We have begun to extend this method to the study of regulatory problems; the first results, dealing with a very simple model of OSH regulation, are now discussed.

An OSH Synthetic Data Experiment: Early Results

The Basic Model. A simplified simulation model of accident generation within firms was created. According to the model, accident rates within a firm were an additive combination of factors related to a firm's size, industry type, and safety programs, and a firm-specific constant. As discussed below, one version of the model assumed that OSHA

inspections influenced the impact of firm's safety programs on accident rates; another version of the model assumed that OSHA inspections did not influence safety programs and accident rates. We explicitly set the basic model to resemble closely the model specified by most regression studies of OSHA's impact.

The model generated monthly accident rates for 1000 firms, and accumulated these monthly rates into yearly total rates. The generating formula used was:

$$Acc = K + (.5 * S) + Z + (.5 * I) + E$$

where

Acc = monthly accident rate;

K = a firm specific constant (value 0 through 1);

S = the cumulative accident factor determined by a firm's safety program (value 0 - 2) (Henceforth, this particular factor is called the "safety factor);

Z = the accident factor determined by the size of the firm (value 0 - 1)

I = the accident factor determined by the firm's industry type (value 0 - 2);

E = a random error term (value 0 - 1).

Firms fell into one of five categories of increasing size. If size was 1 or 5, Z = 0; if size was 2 or 4, Z = .5; and if size was 3, Z = 1. Thus, the size-related accident factors followed the curvilinear pattern reported in the literature. Two versions of the model generated data sets. In one version, it was specified that OSHA inspections would reduce the actual safety factor to zero for five months; this is called the "OSHA-effective" model. In the second version, OSHA inspections did not affect the actual safety factor; this is called the "OSHA-ineffective" model. Initially, all values were selected randomly for the firms. A firm's safety factor, size, and industry were allowed to vary between months using a Markov process of varying degrees of stability.

In the model, OSHA inspected all firms in the second year (months 13-24), half early and half late in the year (in March-April or in November-December). The data output for each of the 1000 firms included the yearly accident rate for both years, firm size and type at the end of the second year, and the month of inspection. The inspection variable was coded 1 if the firm was inspected early, and 0 if the firm was inspected late.

A generalized multiple linear regression model was used to evaluate the effects of OSHA inspections on yearly accident rates. Following the earlier studies, the regression equation attempted to fit the accident rate for the second year to the previous year's accident rate, firm size and type, and whether the firm was inspected early or late. The regression coefficient and significance level for the inspection variable were recorded for the data sets generated by the OSHA-effective and OSHA-ineffective versions of the model.

If the regression model accurately described the data generated by the simulation model, then the regression coefficients would match the difference between accident rates of the two years. The <u>actual</u> impact of

OSHA inspections on accident rates could be determined by simply subtracting annual accidents within the simulation where OSHA had an effect from annual accidents within the simulation where OSHA had no effect. This is because in all other ways, including random sequences, these two model variations were identical. Furthermore, the F value for the inspection variable would be highly significant for the OSHA-effective simulation and would be insignificant for the OSHA-ineffective simulation. A more detailed description of both the data generating models (programmed in BASIC) and the regression models (using ordinary least squares in SPSS) is reported elsewhere (McCaffrey, Andersen, McCold, and Kim: 1984).

<u>Basic Model Test</u>. The regression model accurately described the simulated reality under a wide variety of stochastic conditions. The relative size of the random error term, E, and the "stability" of the Markov process that changed safety program's effects, firm size, and industry type were varied substantially (A perfectly "stable" Markov process has all unit values on the diagonal of the state transition matrix--a value as low as .5 was tested). Under stochastic processes of varying stability, the regression model detected that OSHA reduced accidents in the OSHA-effective simulation, and that OSHA did not reduce accidents in the OSHA-ineffective simulation. Bear in mind, however, that none of the empirical and theoretical problems discussed in Section II above were yet built into the data generating model. Two of these problems were then introduced into the simulation, one at a time.

Misspecification of the OSHA Inspection Variable. As discussed in Section II above, numerous empirical and interpretive problems surround the OSHA inspection variable. In the OSHA-effective simulation earlier, OSHA inspections always increase the effectiveness of safety programs; a value of 1 for the inspection variable is a perfect and unique predictor of systematic improvement. We relaxed this assumptions slightly by introducing varying degrees of mismeasurement to the OSHA variable. The mismeasurement involved "miscoding" the inspection variable for certain varying percentages of the firms, rendering the variable inaccurate for those firms. The results of this experiment are presented in Figure 1.

The horizontal axis reports the percent of the inspection variables that have been miscoded in the OSHA-effective simulation. The vertical axis presents the F statistic for the OSHA inspection variable (Scale 1) and the predicted effect (B value) of OSHA inspections (Scale 2). The horizontal dashed line reads on scale 2, and shows the "real" effect of OSHA derived not by statistical inference but by directly comparing the OSHA-effective and OSHA-ineffective simulations.

For no measurement error (i.e., no misspecification of the OSHA variable) the regression model over-predicts OSHA's effectiveness by roughly a factor of 2. This result is highly statistically significant. When the inspection variable for approximately 30% of the inspected firms has been misspecified the regression model's prediction coincides with the correct magnitude of effect for OSHA inspections.

Both the predicted effect of OSHA inspections and the associated F statistic decline monotonically as the mismeasurement increases. At

roughly 45% misspecification, the predicted OSHA effect is not statistically significant, although OSHA is known to be effective in the model.

The results suggest that measurement error in the OSHA variable reduced the predicted effect of OSHA on accident rates, and degraded the F statistic. Also, an estimate under considerable measurement error matched the actual simulated effect of OSHA inspections.

OSHA Inspections Interact with Reporting Behavior. In a second experiment, we specified in the simulation that an OSHA inspection increased the reporting of accidents in the firm. We specified that, prior to the inspection, only some fraction of the accidents were recorded; an OSHA inspection led to reporting of all accidents. In the real world such an effect is plausible, because inspections could increase awareness of safety issues, improve recordkeeping lest violations of reporting requirements be cited by inspectors, and so forth.

Figure 2 presents the results of this experiment. The Percent of Accidents Reported Prior to OSHA Inspection is recorded on the horizontal axis. The vertical Scale 1 and Scale 2 are analogous to those scales in Figure 1. Again, the actual simulated effect of OSHA inspections in reducing accidents is computed by comparing the OSHA-effective to the OSHA-ineffective simulation, and is shown by a horizontal dashed line labelled "actual effect." The line labeled "apparent effect" is derived by comparing the OSHA-effective to the OSHA-ineffective simulations. This line represents the pure effect of increases in accident reporting and actual accident reductions due to inspections. Of course, such a curve could never be computed in a real data analysis situation since the true effect of OSHA on safety, and its effect through changes in reporting are deeply intertwined.

If it is assumed that all accidents were reported prior to OSHA inspection, the regression model predicts an OSHA effect that it roughly twice the actual effect. When roughly 96 percent of accidents had been reported prior to OSHA's inspections, the regression model's estimate coincides with the actual effect of OSHA on accidents reduction. At roughly 91% of accidents reported prior to inspection, the regression model predicts no effect of OSHA in the simulated world where OSHA is known to have an effect. The OSHA effect is not statistically significant from roughly 88% to 93% of accidents reported prior to inspection. For lower prior reporting rates, the model predicts that OSHA inspections actually increase accidents (Again, this is in the simulation where OSHA is known *a priori* to make safety programs perfectly effective).

These results map the statistical implications of a plausible situation in which OSHA inspections increase accident reporting, and simultaneously reduce accidents. Of course, the exact ranges of significance reported here are artifacts of the simulation model being used. The important point is *not* that prior accident reporting rates of between 88 and 93% will produce null results nor that a 96% prior reporting rate produces unbiased estimates of OSHA effect. Rather, the point is that these types of experiments can alert the analyst to a

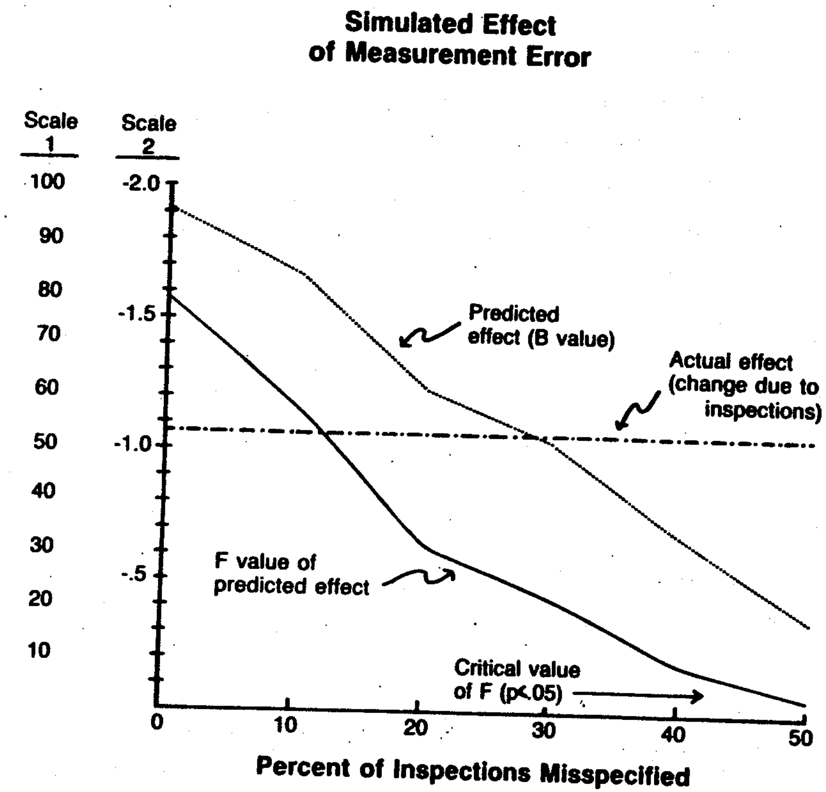
possibly important effect before detailed and expensive data collection begins. Also, such experiments can be used to probe the sensitivity of regression-based results to complications suggested in case studies. Thus, the simulation models provide a bridge between case studies and regression studies via synthetic data experiments.

Next Steps. In the experiments reported above, the simulation model was built so as to conform to the specification most commonly used in regression studies of OSHA effectiveness. The structure of the model was relatively simple, containing only a few lags, non-linearities and feedback effects. A more ambitious task would be to base the simulation model directly on the richer and more detailed insights gained through case studies. Plausible theoretical and empirical complications would be initially built into the model, rather than tacked on as in the simple illustrative experiments described above.

Furthermore, the simulation model should be subjected to more extensive analysis and sensitivity testing before being used in synthetic data experiments. For example, the estimated parameters from the synthetic world should closely resemble in size, magnitude and significance those observed from real data sets. Such a full range of tests were not performed in the results discussed above.

Finally, insights into how such synthetic data experiments should be performed need to be refined by replicating these experiments in several other policy areas (McCaffrey, Andersen, McCold, and Kim, 1984). Such a research program would generate guidelines for using synthetic data experiments to bridge the gap between case studies and regression studies of the impact of government programs. These experiments, and the theoretical simulation modelling that precede them, could begin to remedy methodological problems associated with atheoretical and poorly measured variables, and excessively simplified causal structures.

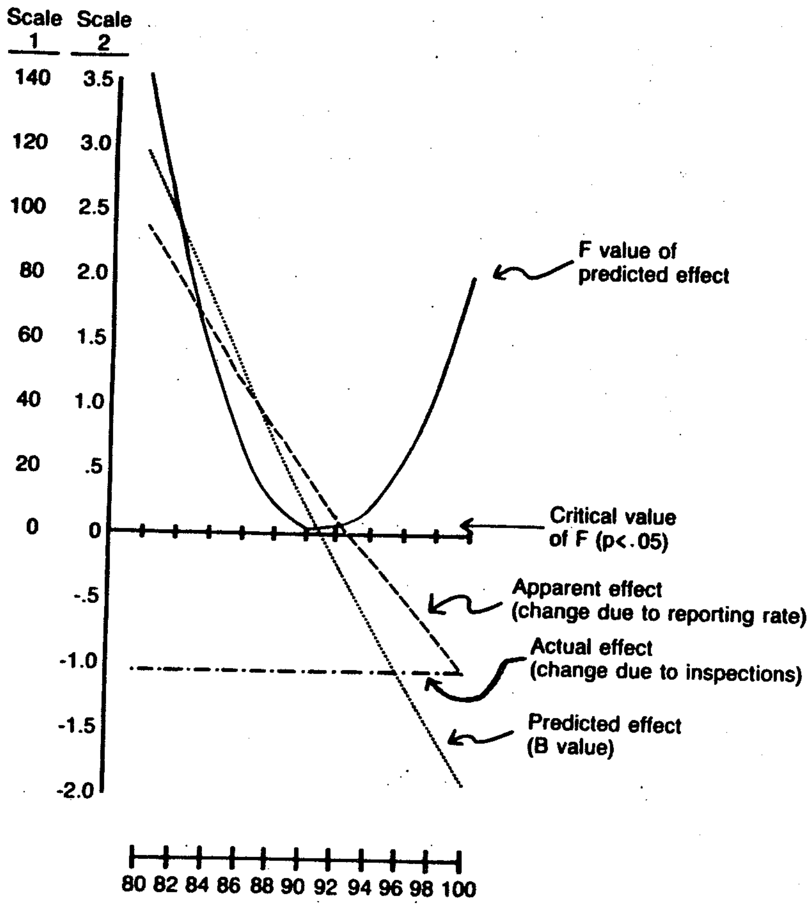
FIGURE 1



NOTE: Scale 1- F value of inspection variable
Scale 2- mean change in accidents due to O.S.H.A. inspections (accidents per year per firm)

FIGURE 2

Simulated Effect of Reporting Error



Percent of Accidents Reported Prior to O.S.H.A. Inspections

NOTE: Scale 1- F value of inspection variable
Scale 2- mean change in accidents due to O.S.H.A. inspections (accidents per year per firm)

FOOTNOTES

1. Often these variables are entered only as "controls" for environmental effects, and so it might seem unreasonable to expect specification of their approximate theoretical relationship to the dependent variable. We do not think it is unreasonable for two reasons. First, analysts commonly do assert a relationship between a control variable and the dependent variable; the problem is that there is usually no basis for the assertion. Second, having only vague (or no) sense of the meaning of control variables can lead us to overlook how they interact with policy variables. Usually the descriptive data on the possible associations among control, policy, and dependent variables are poor. (For example, only sparse descriptive information on how OSHA inspections filter into firms' behavior, or how hospitals make investment decisions [Sloan and Steinwald, 1980: 18] are available). If analysts had a reasonable understanding of the relationship of the control variables to the policy and dependent variables they could assert knowledgeably that control variables are theoretically independent of policy variables. When they have only the slightest, if any, idea of the control variables' dynamics they cannot know when control variables are independent of policy variables. Nor can they just look at correlations between control and policy variables, for if they do not have well founded ideas about theoretical connections of control and

1. (continued).

policy variables they are not able to interpret the correlations. (Nonsignificant correlations could reflect offsetting dynamics or unmeasured interactions as easily as null associations).

2. An extended discussion of how the theoretical and empirical problems discussed in this paper apply to hospital certificate of need regulation, and also hospital rate setting regulation, is found in McCaffrey, Andersen, McCold and Kim (1984).

3. However, the methodological advantages of dynamic simulation modelling are bought at a price. Such models are typically beset by complex questions surrounding parameter estimation (Graham, 1980) as well as overall model validation (Forrester and Senge, 1980). A trade off emerges between the relative strengths of regression and dynamic simulation modelling. Take parameter estimation as a case in point. Simulation modelling uses a diversity of subjective and objective data sources both in formulating a rich causal structure as well as in estimating the parameters associated with the structure. On the other hand, because regression analyses are restricted to available longitudinal or cross-sectional data sets, they usually have less richness of causal structure, but many fewer questions concerning parameter estimation due to the existence of formal tests of statistical significance. The synthetic data procedures reported below represent one of the several possible ways that case studies, dynamic simulation models, and regression models can be combined to capitalize on the relative strengths of all three approaches.

REFERENCES

Andersen, David

1981 "Kalman filter estimates of system states compared to the general regression problem." International Journal of Policy Analysis and Information Systems, 5: 95-109 (No. 2).

Arthur Andersen and Company

1979 Cost of Government Regulation: Study for the Business Roundtable (Three Volumes). New York: The Business Roundtable.

Bacow, Lawrence

1980 Bargaining for Job Safety and Health. Cambridge: MIT Press.

Butler, Richard, and John Worrall

1983 "Workers' compensation: Benefit and injury claims rates in the seventies." Review of Economics and Statistics, 65: 580-589.

Cambridge Research Reports, Inc.

1978 Public and Worker Attitudes Toward Carcinogens and Cancer Risk. Cambridge: Cambridge Reports, Inc.

Forrester, Jay

1961 Industrial Dynamics. Cambridge: MIT Press.

1980 "Information sources for modeling the national economy." Journal of the American Statistical Association, 75: 555-566 (September).

Forrester, J., and P. Senge

1980 "Tests for Confidence in System Dynamics Models." Pp. 209-228 in A.A. Legasto, J.W. Forrester, and J.M. Lynch (Eds.), System Dynamics, TIMS Studies in the Management Sciences, Vol. 14. Amsterdam: North Holland Publishing.

Freedman, Audrey

1981 Industry Response to Health Risk. New York: The Conference Board.

Fuangchen, Somporn, and David McCaffrey

1983 "The impact of certificate of need and rate setting regulation on medical technology." Paper presented to Fifth Annual Research Conference of Association for Public Policy Analysis and Management.

Graham, Alan

1980 "Parameter Estimation in System Dynamics Modeling." Pp. 143-161 in J. Randers (Ed.), Elements of the System Dynamics Method. Cambridge: MIT Press.

Hanushek, Eric

1981 "Throwing money at schools." Journal of Policy Analysis and Management, 1: 19-41 (Fall).

Howell, Julianne

1981 Regulating Hospital Capital Investment: The Experience in Massachusetts. NCHSR Research Summary Series, DDHS Publ. No. (PHS) 81-3298 (March). Washington: U.S. Department of Health and Human Services.

Johnson, Curtis

1980 "Some effects of data errors in econometric models." TIMS Studies in the Management Sciences, 14: 143-159.

Joskow, Paul

1981 Controlling Hospital Costs. Cambridge: MIT Press.

Kochan, T., L. Dyer, and D. Lipsky

1977 The Effectiveness of Union-Management Safety Committees. Kalamazoo: W.E. Upjohn Institute for Employment Research.

Linneman, Peter

1980 "The effects of consumer safety standards: The 1973 mattress flammability standard." Journal of Law and Economics, 23: 461-479 (October).

Luecke, Daniel, and Noel McGinn

1975 "Regression analysis and educational production functions: Can they be trusted?" Harvard Educational Review, 44: 325-350.

Mass, N. and P. Senge

1980 "Alternative Tests for Selecting Model Variables." Pp. 203-223 in J. Randers (Ed.), Elements of the System Dynamics Method. Cambridge: MIT Press.

McCaffrey, David

1982 "Corporate resources and regulatory pressures; Toward explaining a discrepancy." Administrative Science Quarterly, 27: 398-419 (September).

1983 "An assessment of OSHA's recent effects on injury rates." Journal of Human Resources, 18: 131-146 (Winter).

McCaffrey, D., D. Andersen, P. McCold, D.H. Kim

1984 "Dynamic Systems, Ambiguous Variables, and Discrepant Findings." Working Paper, Rockefeller College of Public Affairs and Policy, SUNY-Albany

Meadows, Donella

1980 "The Unavoidable A Priori." Pp. 23-57 in J. Randers (Ed.), Elements of the System Dynamics Method. Cambridge; MIT Press.

Mechanic, David

1979 "Correlates of physician utilization: Why do major multivariate studies of physician utilization find trivial pschosocial and organizational effects?" Journal of Health and Social Behavior, 20: 387-396.

Mendeloff, John

1979 Regulating Safety. Cambridge: MIT Press.

Policy Analysis-Urban Systems, Inc.

1981 Evaluation of the Effects of Certificate of Need Programs Three Volumes). Washington: U.S. Department of Health and Human Services.

Porter, A., T. Connolly, R. Heikes, and C. Park

1981 "Misleading indicators: the limitations of multiple linear regression in formulation of policy recommendations." Policy Sciences, 13: 397-418.

Randers, Jorgen (Ed.)

1980 Elements of the System Dynamics Method. Cambridge: MIT Press.

Rutter, M. B. Maugham, P. Mortimore, and J. Ouston

1979 Fifteen Thousand Hours: Secondary Schools and Their Effects on Children. Cambridge: Harvard University Press.

Senge, Peter

1977 "Statistical estimation of feedback models." Simulation (June), pp. 177-184.

Sloan, Frank

1982 "Government and the regulation of hospital care." American Economic Review Papers and Proceedings, 72: 196-201.

Sloan, Frank, and Bruce Steinwald

1980 Insurance, Regulation, and Hospital Costs. Lexington: Lexington Books.

Smith, Robert

1979 "The impact of OSHA inspections on manufacturing injury rates." The Journal of Human Resources, 14: 145-170 (Spring).

Steinwald, Bruce, and Frank Sloan

1980 "Regulatory approaches to hospital cost containment: A synthesis of the empirical evidence." Pp. 274-308 in M. Olson (ed.), A New Approach to the Economics of Health Care. Washington: American Enterprise Institute.

U.S. Office of Technology Assessment

1981 Policy Implications of the Computed Tomography (CT) Scanner. Washington: U.S. Office of Technology Assessment.

Viscusi, W. Kip

1979 "The impact of OSHA inspections on manufacturing injury rates." Bell Journal of Economics, 10: 117-140 (Spring).

Worrall, John, and David Appel

1982 "The wage replacement rate and benefit utilization in workers' compensation insurance." The Journal of Risk and Insurance, September: 361-371.