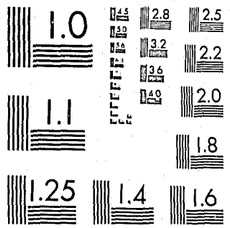
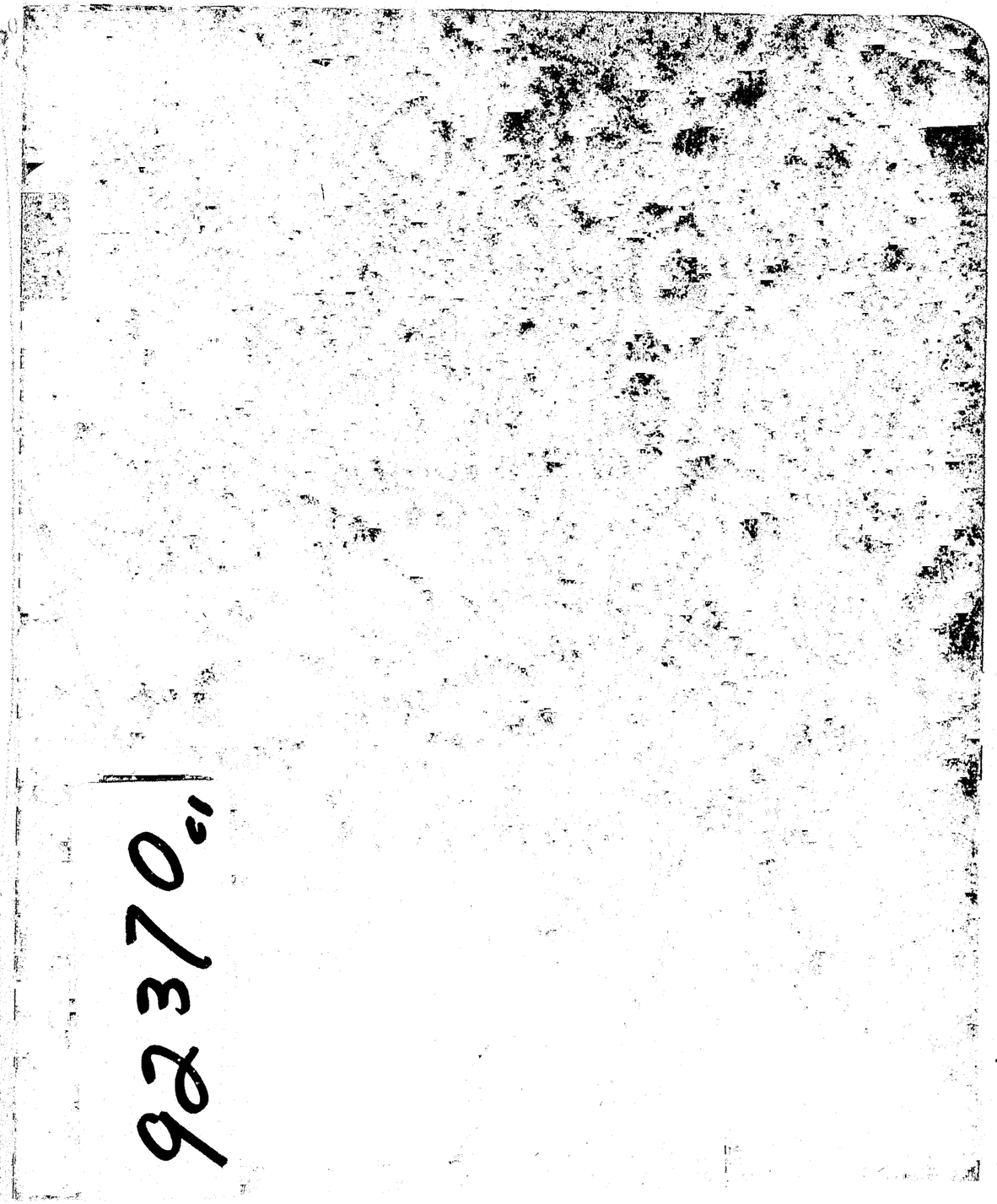




Rensselaer Polytechnic Institute

TROY, NEW YORK 12181

FINAL REPORT

On Developing Evaluation Designs: A Summary Report

by

James M. Tien, Ph.D.
Rensselaer Polytechnic Institute
Troy, New York 12181

November 1983

This study was supported by Grant Number 81-IJ-CX-K026, awarded to Rensselaer Polytechnic Institute by the National Institute of Justice (formerly, the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration), U.S. Department of Justice, under the Omnibus Crime Control and Safe Streets Act of 1968, as amended. Points of view or opinions stated in this document do not necessarily represent the official position or policies of the U.S. Department of Justice.

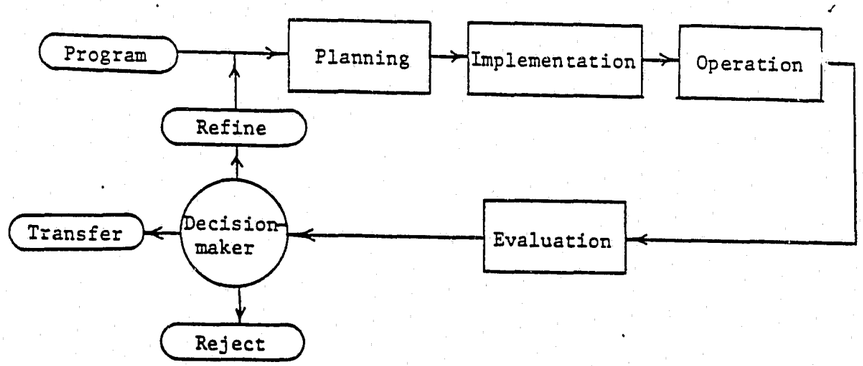
1. Introduction

Does the program work? Is it worth the cost? Can and should it be implemented elsewnere? It is the purported purpose of evaluation to provide answers to these and related questions. Unfortunately, program evaluation has not lived up to its expectations. The field of evaluation is littered with efforts which do not adequately address the important issues or objectives; which do not employ valid controls for comparison purposes; which rely on inadequate measures or include expensive collection of data on measures that are in fact never used in the evaluation; which rely on inappropriate measurement methods; or which employ inadequate analytic techniques. Most, if not all, of the above cited problems could be mitigated by developing -- at the beginning of an evaluation effort -- a valid and comprehensive evaluation design. Although there is no stock design that can be taken off-the-shelf and implemented without revision, there should be an approach or process by which such designs can be developed.

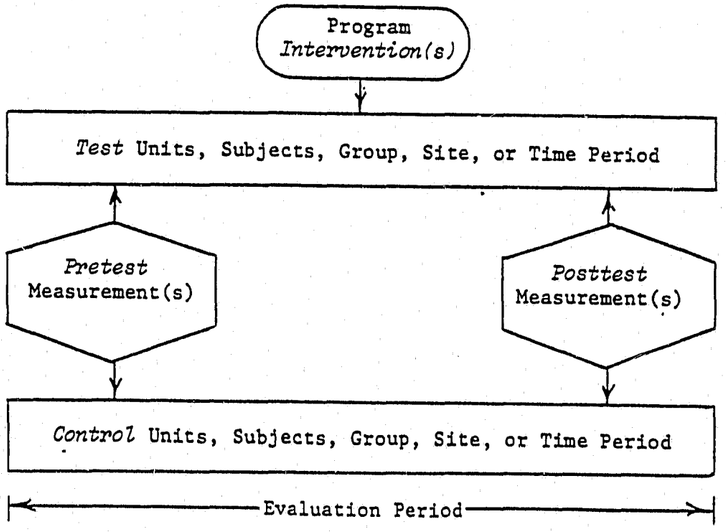
Indeed, Tien (1979) developed such an approach. The objectives of this study were, first, to extend Tien's earlier work, and, second, to demonstrate the viability of the extended approach by applying it to at least two on-going evaluation efforts. The first objective has been met primarily through the development of statistical models which extend Tien's approach by detailing certain critical links in the approach. (As an example, an enlightening linear statistical model is presented herein.) The second objective has been achieved through the application of the extended approach to three actual evaluation efforts (Colton et al. 1982, Cahn and Tien 1982, Cahn and Tien 1983). In regard to the

study's products, it should be noted that in addition to this summary report and evaluation-oriented write-ups for the three applications, an invited paper -- entitled "A Systems and Model-Based Approach to Evaluation Design" -- has been submitted to the Encyclopedia of Systems and Control (Edited by M. Singh and A.P. Sage, which should be published in mid 1984 by Pergamon Press. Furthermore, it should be noted that the results of this study has contributed to a graduate level course -- entitled "Evaluation Methods for Decision Making -- which the author teaches every spring semester at Rensselaer Polytechnic Institute.

Before summarizing the results of the above stated study, it is helpful to provide some background information and terminology. Figure 1 details a program's conduct in terms of its development and evaluation steps. Figure 1(a) assumes, first, that the program design and its evaluation design are developed concurrently (so that the program is indeed amenable to evaluation), and, second, that the traditional paradigm of evaluation is in effect (i.e., evaluation provides feedback to the program administrator or decisionmaker, who decides whether the program should be refined, rejected, and/or transferred). In terms of the evaluation process, Fig. 1(b) notes that, in general, every unit (i.e., subject, group, site or time period) can be designated as being either test or control. During the period of evaluation, pretest measurements are first made of both sets of units, followed by the administration of the program intervention on each test unit, and concluding with appropriate posttest measurements. There may, of course, be several test units, control units, program interventions, pretest measurements, and posttest measurements.



(a) Development



(b) Evaluation

Figure 1 Program development and evaluation

It should also be recognized that Fig. 1 best depicts a social program. This depiction is not by coincidence but reflects the fact that evaluation as an area of research interest has its roots in the social sciences, especially in the discipline of psychology. Obviously, the conduct of evaluation has extended beyond social programs, and includes, for example, technology assessments or evaluations (Porter et al., 1980) and evaluation of computer aids (Sage 1981).

In many respects, this report can be regarded as a guide to program evaluation design. The report identifies a design framework which links program characteristics to design elements; defines five related design components which contain the essential design elements; and develops a linear statistical model which highlights some of the key underlying issues in program evaluation. The report does not, however, address issues relating to the conduct and management of an evalution effort; nor does it give advise about how to communicate the evaluation findings. In these respects, it is different than other evaluation-related guides and manuals -- see, for example, Rossi and Freeman (1982).

The remainder of this report is divided into four sections which address, respectively, the design framework, the design components, a linear statistical model, and some concluding remarks.

2. Design Framework

A general evaluation design framework is depicted in Fig. 2; it assumes that the program and its evaluation design are developed concurrently. The framework is based on a dynamic roll-back approach which consists of three steps leading up to a valid and comprehensive evaluation design.

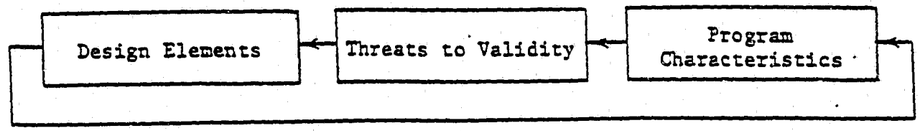


Figure 2 **Design framework: a dynamic roll-back approach**

The "roll-back" aspect of the approach is reflected in the ordered sequence of steps which are identified in Fig. 2: the sequence rolls back in time from i) a projected look at the range of program characteristics (i.e., from its rationale through its operation and anticipated findings); to ii) a prospective consideration of the threats (i.e., problems and pitfalls) to the validity of the final evaluation; and iii) a more immediate identification of the evaluation design elements. The logic of this sequence of steps should be noted; that is, the anticipated program characteristics identify the possible threats to validity, which in turn point to the design elements that are necessary to mitigate, if not to eliminate, these threats. The sequence of steps can be stated in terms of two sets of links which relate, respectively, an anticipated set of program characteristics to an intermediate set of threats to validity to a final set of design elements.

The "dynamic" aspect of the approach refers to its nonstationary character; that is, the components of the framework must constantly be updated, throughout the entire development and implementation phases of the evaluation design. In this manner, the design elements can be refined, if necessary, to account for any new threats to validity which may be caused by previously unidentified program characteristics. In

sum, the dynamic roll-back approach is systems-oriented; it represents a purposeful and systematic process by which valid and comprehensive evaluation designs can be developed.

The first two steps of the design fremework are elaborated on in the next two subsections, while the third step is considered in Sect. 3.

Program Characteristics

In general, the characteristics of a program can be determined by seeking responses to the following questions: What is the program rationale? Who has program responsibility? What is the nature of program funding? What is the content of the program plan? What are the program constraints? What is the nature of program implementation? What is the nature of program operation? Are there any other concurrent programs? What are the anticipated evaluation findings?

Again, according to Fig. 2, it should be noted that the purpose for understanding the program characteristics is to identify the resultant problems or pitfalls that may arise to threaten the validity of the final evaluation.

Threats to Validity

After almost two decades, the classic monograph by Campbell and Stanley (1966) is still the basis for much of the on-going discussion on threats to validity. However, as listed in Table 1, the original 12 threats by Campbell and Stanley (1966) have been expanded to include 7 additional threats. The 20 threats to validity can be grouped into the following five categories.

- Internal validity refers to the extent that the statistical association of an intervention and measured impact can reasonably be considered a causal relationship.

Threats to Internal Validity

1. *Extraneous events* (i.e., history) may occur during the period of evaluation, inasmuch as total test or experimental isolation cannot be achieved in social experimentation.
2. *Temporal maturation* of subjects or processes (e.g., growing older, growing more tired, becoming wiser, etc.) -- including cyclical maturation -- may influence observed impacts.
3. *Design instability* (i.e., unreliability of measures, fluctuations in sampling units or subjects, and autonomous instability of repeated or equivalent measures) may introduce biases.
4. *Pretest experience*, gained from a response to a pretest measurement (e.g., questionnaire test, observation, etc.), may impact the nature and level of response to a subsequent posttest measurement.
5. *Instrumentation changes* (e.g., changes in the calibration of a measurement instrument, changes in the observers or evaluators used, etc.) may produce changes in the obtained measurements.
6. *Regression artifacts* may occur due to the identification of test or control subjects (or time periods) whose dependent or outcome measures have extreme values -- these extreme values are artificial and will tend to regress toward the mean of the population from which the subjects are selected.
7. *Differential selection* -- as opposed to random selection -- of subjects for the test and control groups may introduce biases.
8. *Differential loss* (i.e., experimental mortality) of subjects from the test and control groups may introduce biases.
9. *Selection-related interaction* (with extraneous events, temporal maturation, etc.) may be confounded with the impact of the intervention, as, for example, in the case of a self-selected test group or in the case of test and control groups which are maturing at different rates.

Threats to External Validity

10. *Pretest-intervention interaction* (including "halo" effect) may cause a pretest measurement to increase or decrease a subject's sensitivity or responsiveness to the intervention; thus rendering the results obtained for a pretested population unrepresentative of the impacts of the intervention for the unpretested universe from which the test subjects are selected.
11. *Selection-intervention interaction* may introduce biases which render the test and/or control groups unrepresentative of the universe from which the test subjects are selected.
12. *Test-setting sensitivity* (including "Hawthorne" and "placebo" effects) may preclude generalization about the impact of the intervention upon subjects being exposed to it under non-test or non-experimental settings.
13. *Multiple-intervention interference* may occur whenever multiple interventions are applied to the same subjects, inasmuch as the impacts of prior interventions are usually not erasable.

Threats to Construct Validity

14. *Intervention sensitivity* may preclude generalization of observed impacts to different or related interventions -- complex interventions may include other than those components responsible for the observed impacts.
15. *Measures sensitivity* may preclude generalization of observed impacts to different or related impact measures -- complex measures may include irrelevant components that may produce apparent impacts.

Threats to Statistical Conclusion Validity

16. *Extraneous sources* of error (including "post hoc" error) may minimize the statistical power of analyses.
17. *Intervention integrity* or lack thereof may invalidate all statistical conclusions.

Threats to Conduct Conclusion Validity

18. *Design complexity* (including technological and methodological constraints) may preclude the complete and successful conduct of the evaluation.
19. *Political infeasibility* (including institutional, environmental and legal constraints) may preclude the complete and successful conduct of the evaluation.
20. *Economic infeasibility* (including hidden and unanticipated costs) may preclude the complete and successful conduct of the evaluation.

Source: Tien 1979, p. 498.

Table 1 Design considerations: threats to validity

- External validity refers to the extent that the causal relationship can be generalized to different populations, settings, and times.
- Construct validity refers to the extent that the causal relationship can be generalized to different interventions, impact measures, and measurements.
- Statistical conclusion validity refers to the extent that an intervention and a measured impact can be statistically associated -- error could be either a false association (i.e., Type I error) or a false nonassociation (i.e., Type II error).
- Conduct conclusion validity refers to the extent that an intervention and its associated evaluation can be completely and successfully conducted.

In evaluation terms, the threats to validity can be regarded as plausible rival hypotheses or explanations of the observed impacts of a program. That is, the assumed causal relationships (i.e., test hypotheses) may be threatened by these rival explanations. Sometimes the threats may detract from the program's observed impacts. For example, the model in Sect. 4 shows how a regression artifact may result in such a detraction. It is therefore the purpose of an evaluation design to minimize the threats to validity, while at the same time to suggest the causal relationships.

In conclusion, it should be stated that the evaluation design framework presented in this section is very much dependent on the threats to validity. It is through these threats that program characteristics are linked to design elements.

3. Design Components

This section provides both a summary of and an update to the earlier work by Tien (1979), who found it systematically convenient to describe a program evaluation design in terms of five components, including test hypotheses, selection scheme, measures framework, measurement methods, and analytic techniques.

Test Hypotheses

The test hypotheses component is meant to include the range of issues leading up to the establishment of test hypotheses. In practice and as illustrated in the dynamic roll-back approach in Fig. 2., the test hypotheses should be identified only after the program characteristics and threats to validity have been ascertained.

The test hypotheses are related to the rationale or objectives of the program and are defined by statements that hypothesize the causal relationships between dependent and independent measures, and it is a purpose of program evaluation to assess or test the validity of these statements. In order to be tested, a hypothesis should i) be expressed in terms of quantifiable measures, ii) reflect a specific relationship that is discernible from all other relations, and iii) be amenable to the application of an available and pertinent analytic technique.

Finally, it should be stated that while the test hypotheses themselves cannot mitigate or control for threats to validity, poor definition of the test hypotheses can threaten statistical conclusion validity, since threats to validity represent plausible rival hypotheses.

Selection Scheme

The purpose of this component is to develop a scheme for the selection and identification of test groups and, if applicable, control groups, using appropriate sampling and randomization techniques. The selection process involves several related tasks, including the identification of a general sample of units from a well-designated universe; the assignment of these (perhaps matched) units to at least two groups; the identification of at least one of these groups to be the test group; and the determination of the time(s) that the intervention and, if applicable, the placebo are to be applied to the test and control groups, respectively. A more valid evaluation design can be achieved if random assignment is employed in carrying out each task. For example, random assignment of units to test and control groups increases the comparability or equivalency of the two groups, at least prior to the program intervention. The statistical model in Sect. 4 shows how the equivalency of the two groups can affect the net observed impact of the program intervention.

Tien (1979) identifies numerous selection schemes or research designs, including experimental designs (e.g., pretest-posttest equivalent design, Solomon four-group equivalent design, posttest-only equivalent design, factorial designs), quasi-experimental designs (e.g., pretest-posttest nonequivalent design, posttest-only nonequivalent design, interrupted time-series nonequivalent design, regression-discontinuity design, ex-post facto designs) and non-experimental designs (e.g., case study, survey study, cohort study). In general, it can be stated that non-experimental designs do not have a control group

or time period, while experimental and quasi-experimental designs do have such controls -- even if it is just a before-after control. The difference between experimental and quasi-experimental designs is that the former set of designs have comparable or equivalent test and control groups (i.e., through randomization) while the latter set of designs do not.

Although it is always recommended that an experimental design be employed, there are a host of reasons which may prevent or confound the establishment -- through random assignment -- of equivalent test and control groups. One key reason is that randomization creates a focused inequity because some persons receive the (presumably, desirable) program while others do not. Whatever the reasons, the inability to establish equivalent test and control groups should not preclude the conduct of an evaluation. Despite their inherent limitations, some quasi-experimental designs are adequate. In fact, some designs (e.g., regression-discontinuity designs) are explicitly nonrandom in their establishment of test and control groups. On the other hand, other quasi-experimental designs should only be employed if absolutely necessary and if great care is taken in their employment. Ex-post facto designs belong in this category. Likewise, non-experimental designs should only be employed if it is not possible to employ an experimental or quasi-experimental design.

In terms of selection scheme factors which could mitigate or control for the various threats to validity, it can be stated that randomization is the key factor. In particular, most, if not all, of the internal and external threats to validity can be mitigated by the experimental designs which can be achieved through randomization.

Measures Framework

There are two parts to the measures framework component. First, it is necessary to specify the set of evaluation measures which is to be the focus of the particular evaluation. Second, a model reflecting the linkages among these measures must be constructed.

In terms of evaluation measures, Tien (1979) has identified four sets of measures -- input, process, outcome and systemic measures. In general, the input and process measures serve to "explain" the resultant outcome measures. Input measures alone are of limited usefulness since they only indicate a program's potential -- not actual -- performance. On the other hand, the process measures do identify the program's performance but do not consider the impact of that performance. Finally, the outcome measures are the most meaningful observations since they reflect the ultimate results of the program. In practice, as might be expected, most of the available evaluations are fairly explicit about the input measures, less explicit about the process measures, and somewhat fragmentary about the outcome measures.

An increasingly popular outcome measure of social programs is the multiattribute utility measure (Keeney and Raiffa 1976). In a multiattribute utility framework, each outcome measure is considered an attribute which can be combined with other measures or attributes by means of an aggregation rule, most often simply a judgmentally weighted linear combination or an additive aggregation. The resultant combination or aggregation is a value of the utility which may be used to compare the outcomes of different programs or alternative versions of the same program. The attraction of being able to compare the outcomes

of one program with those of another will undoubtedly hasten the introduction of utility theory into evaluation research.

The fourth set of evaluation measures -- the systemic measures -- can also be regarded as impact measures but have been overlooked to a large extent in the evaluation literature. The systemic measures allow the program's impact to be viewed from a total systems (i.e., organizational, longitudinal, programmatic and policy-oriented) perspective.

The second part of the measures framework concerns the linkages among the various evaluation measures. A model of these linkages should contain the hypothesized relationships -- including cause-and-effect relationships -- among the measures. Thus, the model should help in identifying plausible test and rival hypotheses, as well as in identifying critical points of measurement and analysis. In practice, the model could simply reflect a systematic thought process undertaken by the evaluator, or it could be explicitly expressed in terms of a table, a block diagram, a flow diagram, or a matrix.

In conclusion, concise and measurable measures can mitigate the measures-related threats to validity. Additionally, the linkage model can help to avert some of the other threats to validity.

Measurement Methods

The list of issues and elements which constitute the measurement methods component include measurement time frame (i.e., evaluation period, measurement points, and measurement durations), measurement scales (i.e., nominal, ordinal, interval, and ratio), measurement instruments (i.e., questionnaires, data collection forms, data collection algorithms, and electromechanical devices), measurement procedures

(i.e., administered questionnaires, implemented data collection instruments, telephone interviews, face-to-face interviews, and observations), measurement samples (i.e., target population, sample sizes, sampling technique, and sample representativeness), measurement quality (i.e., reliability, validity, accuracy, and precision) and measurement steps (i.e., data collection, data privacy, data codification, and data verification).

Measurement methods which could mitigate or control for threats to validity include a multi-measurement focus, a long evaluation period (which, while controlling for regression artifacts, might aggravate the other threats to internal validity), large sample sizes, random sampling, and pretest measurements.

Analytic Techniques

Analytic techniques are employed in evaluation for a number of reasons: to conduct statistical tests of significance; to combine, relate or derive measures; to assist in the evaluation conduct (e.g., sample size analysis, Bayesian decision models); to provide data adjustments for nonequivalent test and control groups; to model test and/or control situations.

Next to randomization (which is usually not implementable), perhaps the single most important evaluation design element (i.e., the one which can best mitigate or control for the various threats to validity) is, as alluded to above, modeling. Unfortunately, most evaluation efforts to date have made minimal use of this powerful tool. However, more recent efforts have begun to recognize the importance of adopting a model-based approach to evaluation. Larson (1975), for example, developed

some simple structural models to show that the integrity of the Kansas City Preventive Patrol Experiment was not upheld during the course of the experiment -- thus casting doubt on the validity of the resultant findings. As another example, Willemain (1978) developed a Bayesian model to assist in the implementation of a contingent experimental design.

Finally, in order to hightlight some of the critical evaluation-related issues, a linear statistical model is developed in the next section. A variation of the model was recently employed by Cahn and Tien (1983) to characterize a retrospective "split-area" research design or selection scheme which was used to evaluate the impact of security surveys on commercial burglary.

4. A Linear Statistical Model

To begin with and for simplicity, it is assumed that:

i) There is a single selection measure X (e.g., "before" crime rate)

ii) There is a single impact measure Y (e.g., "after" crime rate)

iii) There are two groups: $j = t$ (test), c (control)

iv) There is a single intervention Z_j, where $Z_j = \begin{cases} 0, & j = c \\ 1, & j = t \end{cases}$

v) There is a disturbance or error term e, which is uncorrelated with other measures and which possesses an expected value of zero.

vi) There is a linear causal relationship between Y_{ij} and X_{ij}; that is,

$$Y_{ij} = a + bZj + dj(X_{ij} - \overline{X}_{..}) + e_{ij} \qquad (1)$$

where Y_{ij} = value of impact measure for unit i in group j

X_{ij} = value of selection measure for unit i in group j

e_{ij} = value of error associated with unit i in group j

Z_j = value (i.e., presence) of intervention in group j

$\overline{X}_{..}$ = X_{ij} averaged over both i and j (i.e., the "grand mean" of the selection measure)

In the above equation, it should be noted that i) b reflects the (net) impact of the intervention; ii) $\overline{X}_{.t} \neq \overline{X}_{.c}$ reflects the presence of a regression artifact threat to validity; iii) $d_j \neq 0$ reflects the presence of a selection-regression artifact interaction threat to validity, and iv) $d_t \neq d_c$ reflects the presence of a selection-intervention interaction threat to validity.

In deriving the impact b, it is first helpful to determine

$$\overline{Y}_{.c} = E[Y_{ij}|j=c] = a + bE[Z_c] + d_c(\overline{X}_{.c} - \overline{X}_{..}) + E[e_{ic}]$$

$$= a + d_c(\overline{X}_{.c} - \overline{X}_{..}) \qquad (2)$$

Similarly,

$$\overline{Y}_{.t} = E[Y_{ij}|j=t] = a + bE[Z_t] + d_t(\overline{X}_{.t} - \overline{X}_{..}) + E[e_{it}]$$

$$= a + b + d_t(\overline{X}_{.t} - \overline{X}_{..}) \qquad (3)$$

Subtracting Eqn. (2) from Eqn. (3) and solving for b, one can show that

$$b = \overline{Y}^*_{.t} - \overline{Y}^*_{.c} \qquad (4)$$

where

$$\overline{Y}^*_{.t} = \overline{Y}_{.t} - d_t(\overline{X}_{.t} - \overline{X}_{..}) \qquad (5)$$

and

$$\overline{Y}^*_{.c} = \overline{Y}_{.c} - d_c(\overline{X}_{.c} - \overline{X}_{..}) \qquad (6)$$

The above expressions can perhaps be better understood by a graphical presentation, as contained in Fig. 3. As indicated in Fig. 3, b is actually the net impact of the intervention on a unit with $\overline{X}_{..}$ as its selection measure. In general, for a unit with a different selection measure -- say X_a -- the net impact would be

$$b|X_a = \overline{Y}^*_{.t}|X_a - \overline{Y}^*_{.c}|X_a \qquad (7)$$

where

$$\overline{Y}^*_{.t}|X_a = \overline{Y}_{.t} - d_t(\overline{X}_{.t} - X_a) \qquad (8)$$

and

$$\overline{Y}^*_{.c}|X_a = \overline{Y}_{.c} - d_c(\overline{X}_{.c} - X_a) \qquad (9)$$

It can also be seen from Fig. 3 that if $d_t = d_c$, then $b|X_a = b|\overline{X}_{..} = b$; that is, the impact of the intevention is the same for all units, even if they possess different selection measure values -- thus, in such a situation, there is no selection-intervention interaction threat to validity. Moreover, if $d_t = d_c = 0$, then not only is $b|X_a = b|\overline{X}_{..} = b$ (i.e., no selection-intervention interaction threat to validity), but, in combining Eqns. (4), (5) and (6), b is also equal to

$$b = \overline{Y}_{.t} - \overline{Y}_{.c} \qquad (10)$$

which implies that there is no selection-regression artifact interaction threat to validity. Further, b can likewise be defined by Eqn. (10) if the test and control groups are equivalent (i.e., $\overline{X}_{.t} = \overline{X}_{.c} = \overline{X}_{..}$), in which case there is no regression artifact threat to validity.

In general, Eqn. (4) can be used to determine the net impact b; however, such a determination would first require calculating the other measures identified in Eqns. (5) and (6). Alternatively, one could determine b in terms of several covariance measures. Further, one could use a t-test of the difference between two sample means to determine if the net impact b is statistically significant.

In sum, it should be noted that the model developed in this section, although relatively simple, is able to adjust for three critical threats to validty. Further, the resultant adjustment is a consequence of the assumed linear relationship between the impact

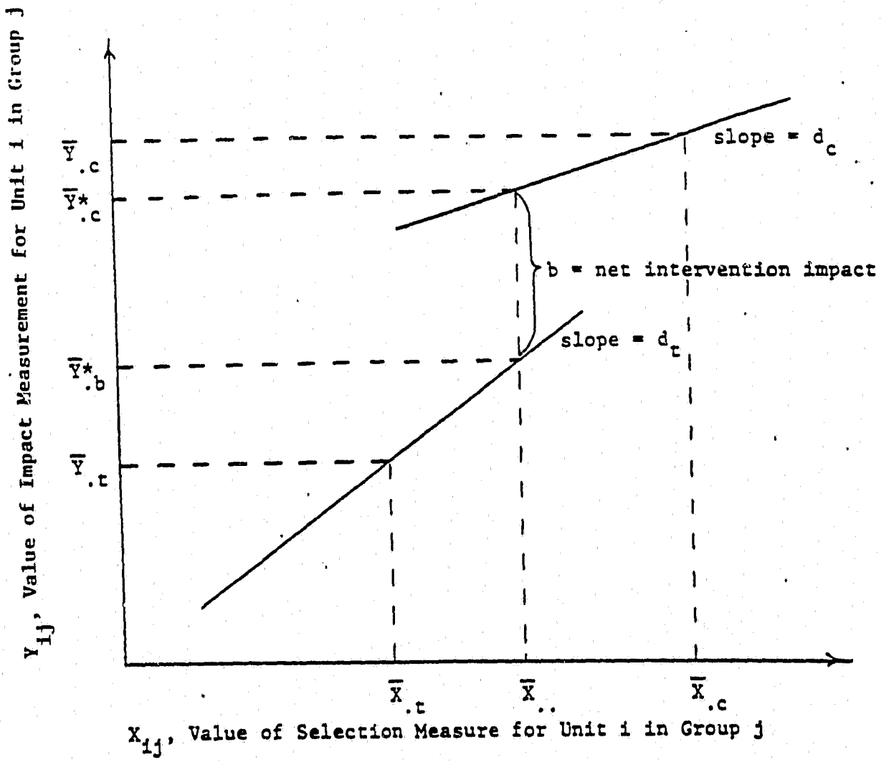


Figure 3 Analytic techniques: a linear statistical model

measure and the selection measure. Finally, in a situation where there are H selection measures (i.e., $X_1, X_2 \ldots X_H$), then Eqns. (5) and (6) can be extended and replaced by

$$\bar{Y}^*_{.j} = \bar{Y}_{.j} - \sum_{h=1}^{H} d_{hj}(\bar{X}_{h.j} - \bar{X}_{h..}) \quad \text{for } j = t, c \tag{11}$$

where d_{hj} is the slope of the regression of the Y_{ij}'s on the X_{hij}'s.

5. Concluding Remarks

In conclusion, three remarks should be made. First, although the focus of this report has been on program evaluation, the systems and model-based approach considered herein is, for the most part, applicable to the design of any analysis effort. According to Webster's Dictionary, to evaluate means "to examine and judge"; thus, evaluation includes the step of analysis (i.e., examination) and can be thought of as a more judgment-oriented form of analysis.

Second, while this report provides a purposeful and systematic approach or guide to the development of an evaluation design, it does not constitute a "cookbook" or handbook. The author feels that an adequate handbook will not be forthcoming in the near future; it will require many more years of evaluation experience and careful analysis of that experience.

Third, the need for evaluation is growing in the U.S., and it will continue to grow in the foreseeable future. Government at every level is being increasingly required to justify the value of its programs. On the other hand, increased federal deregulation, increased domestic and foreign competition, and high interest rates have resulted in similar pressures on the leaders of private industry. Given a growing

need for evaluation, it is critical and necessary that proper procedures exist for the development of evaluation designs which are valid and comprehensive. Certainly, the systems and model-based approach summarized in this report attempts to provide such procedures.

Bibliography

Cahn, M.F. and J.M. Tien, An Evaluation of the Wilmington Management of Demand Program. Cambridge, MA: Public Systems Evaluation, Inc., 1982.

Cahn, M.F. and J.M. Tien, An Evaluation of the Commerical Security Field Test. Washington, DC: National Institute of Justice, 1983.

Campbell, D.T. and J.C. Stanley, Experimental and Quasi-Experimental Designs for Research. Chicago, IL: Rand McNally, 1966.

Colton, K.W., M.L. Brandeau, and J.M. Tien, A National Assessment of Police Command, Control, and Communications Systems. Washington, DC: National Institute of Justice, 1982.

Keeney, R.L. and H. Raiffa, Decisions With Multiple Objectives, New York, NY: Wiley, 1976.

Larson, R.C., "What Happened to Patrol Operations in Kansas City? A Review of the Kansas City Preventive Patrol Experiment", Journal of Criminal Justice, 3, pp. 267-297, 1975.

Porter, A.L., F.A. Rossini, S.R. Carpenter, and A.T. Roper, A Guidebook for Technology Assessment and Impact Analysis. New York, NY: North Holland: 1980.

Rossi, P.H. and H.E. Freeman, Evaluation: A Systematic Approach. Beverly Hills, CA: Sage, 1982.

Sage, A.P., "A Methodological Framework for Systemic Design and Evaluation of Computer Aids for Planning and Decision Support", Computers and Electrical Engineering, 8, pp. 87-101, 1981.

St. Pierre, R.G., "Congressional Input to Program Evaluation: Scope and Effects", Evaluation Review, 7, pp. 411-436, 1983.

Tien, J.M., "Toward a Systematic Approach to Program Evaluation Design", IEEE Transactions on Systems, Man and Cybernetics, SMC-9(9), pp. 494-515, 1979.

Willemain, T.R., Analysis of A Contingent Experimental Design. Cambridge, MA: Massachusetts Institute of Technology, 1978.

END