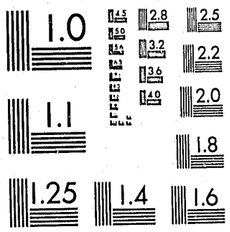


OPERATIONS RESEARCH CENTER

working paper

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ANALYSIS OF A CONTINGENT EXPERIMENTAL DESIGN

A BEFORE-AND-AFTER EXPERIMENT WITH A BASELINE PERIOD OF RANDOM DURATION

by

Thomas R. Willemain

OR 079-78 August 1978

Prepared under Grant Number 78NI-AX-0007 from the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice. Points of view or opinions stated in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice.

Table of Contents

Abstract

This paper introduces a simple "contingent experimental design" and outlines how the contingent design would operate and how it would be evaluated. The familiar "before-and-after" type of experiment is modified so that the duration of the "baseline" period, rather than being fixed before the experiment, is made contingent on the experimenter's prior estimate of the experimental impact and on the baseline data as they appear. At the conclusion of each day of the baseline period, a decision is made as to whether to terminate the baseline at that time, weighing the costs of extending the baseline by one day against the benefits of better estimating the experimental impact. An analytic framework is proposed for making this decision and for comparing the contingent design against an alternative having a baseline period of fixed duration.

Acknowledgements

Professor Richard Larson stimulated interest in this work and offered valuable suggestions. Ms. Pam Ersevim patiently typed the manuscript from a formidable draft.

1. INTRODUCTION

This paper addresses the following question: given that one is planning a prospective experiment of the "before-and-after" type, how long should one extend the "baseline" period? One simple answer is to make the baseline as long as the anticipated trial period. A more intriguing answer, which this paper explores, is to make the duration of the baseline period contingent on the baseline data as they unfold. Thus we imagine that at the conclusion of every day of the baseline period, the managers of the experiment have the option of either beginning the trial period on the next day or extending the baseline period one more day and then reconsidering the question. We develop below an algorithm for making that decision.

There are many possible contexts in which this problem of contingent experimental design might be analyzed. For concreteness we pick one which seems both tractable and realistic, making the following five assumptions. First, we assume that the data are counts of events generated by homogeneous Poisson processes with rates λ_b counts/day during the baseline period and $R\lambda_b$ during the trial period. For instance, the events may be serious crimes, the experimental treatment an increase in police patrol presence, and the goal of the experiment an estimate of the parameter R [1]. Second, we assume that the potential duration of the baseline I_b is unlimited but the anticipated duration of the trial period is fixed at I_t.

Third, we assume that the cost of any day of baseline data collection is C_b and that of any day of trial operation is $C_t > C_b$. This creates a pressure to minimize the duration of the baseline which trades-off against the need to extend the baseline period to smooth out sampling fluctuations and better estimate the true experimental impact. Fourth, we assume that the managers of the experiment have a prior sense of the possible values of R and their relative credibilities. This prior distribution for the experimental impact is essential for the real-time reaction to baseline data. Fifth, we assume that the only other costs of note are the costs of errors in the estimation of R, the experimental impact. In general, these costs will be a function both of the actual value of R and of its estimate $\hat{R}$. The decision itself is treated as costless, or as part of the fixed costs of data analysis.

2. DECISION RULE FOR TERMINATING BASELINE PERIOD

After any day during the baseline period, the decision to be made is to stop the baseline then or to add one more day. If the baseline period is stopped after I_b days, the total cost of the experiment is the cost in errors of estimation plus the implementation costs $C_bI_b + C_tI_t$. If the baseline period is stopped after $I_b + 1$ days, the total cost is the cost in errors of estimation plus $C_b(I_b+1) + C_tI_t$. Let the cost or disutility of estimating impact $\hat{R}$ conditional on the true

impact R_o be $U(\hat{R}|R_o)$. Then the total cost of stopping after I_b days ("now") is

$$C_{now} = U(\hat{R}_{now}|R_o) + C_b I_b + C_t I_t \qquad (1)$$

and the total cost of stopping after $I_b + 1$ days ("next") is

$$C_{next} = U(\hat{R}_{next}|R_o) + C_b(I_b+1) + C_t I_t \ . \qquad (2)$$

In general one should conclude the baseline period whenever

$$C_{now} < C_{next} \qquad (3)$$

which condition corresponds to

$$U(\hat{R}_{now}|R_o) - U(\hat{R}_{next}|R_o) < C_b \ . \qquad (4)$$

Of course, the estimates $\hat{R}_{now}$ and $\hat{R}_{next}$ will be random variables whose values will be unknown at the time of the decision. Suppose that during the first I_b days of the baseline period there have been N_b random events, that during the next baseline day (if it is decided to extend the baseline period) there will be m additional events, and that during the I_t days of the trial period there will be n random events. In that case the estimates of experimental impact would be

$$\hat{R}_{now} = \frac{n/I_t}{n_b/I_b} = \left(\frac{I_b}{I_t}\right)\left(\frac{n}{N_b}\right) \qquad (5)$$

and

$$\hat{R}_{next} = \frac{\cdots\ n/I_t\ \cdots}{(N_b+m)/(I_b+1)} = \left(\frac{I_b+1}{I_t}\right)\left(\frac{n}{N_b+m}\right) \qquad (6)$$

At the time of the decision both m and n will be unknown, but both will presumably be generated by Poisson processes with parameters λ_b and $\lambda_t I_t$ respectively. The parameter λ_b can be estimated by Bayesian updating [2] from the accumulated baseline data (I_b and N_b). The parameter λ_t would then be estimated as $R_o\lambda_b$.

As shown below, this line of analysis produces the probability distributions of m and n , which via (5) and (6) provide the distributions of $U(\hat{R}_{now}|R_o)$ and $U(\hat{R}_{next}|R_o)$ in (4). It is reasonable and customary when faced with a stochastic criterion like (4) to base decisions on the expected values of the stochastic terms [3]. Thus the criterion for stopping the baseline after I_b days becomes

$$E\left[U(\hat{R}_{now}|R_o)\right] - E\left[U(\hat{R}_{next}|R_o)\right] < C_b \ . \qquad (7)$$

Finally, we note that the true experimental impact R_o cannot be known at the time of the decision, but a prior distribution is presumably available. Taking this prior into account leads to the unconditional expected cost decision rule: stop the baseline period after I_b days whenever

$$\int_{R_o=0}^{\infty} \left\{ E\left[U(\hat{R}_{now}|R_o)\right] - E\left[U(\hat{R}_{next}|R_o)\right] \right\} f(R_o)\, dR_o < C_b \quad . \qquad (8)$$

Note that since both $\hat{R}_{now}$ and $\hat{R}_{next}$ are functions of the number of events in the baseline period, N_b, the use of (8) may produce different decisions depending on the particular history of events over the baseline period. Whether a particular history calls for terminating the baseline will depend on the actual current count N_b, on the form of the disutility function for errors in estimation (e.g., are over-estimates more serious than under-estimates?) and on the prior estimate of experimental impact. This general format can accomodate two very different types of evaluator. One, whom we might call the "scientist", would seek to establish "objective" evidence of experimental impact by using a diffuse prior and by reacting equally to over-estimates and under-estimates. The other, whom we might call the "advocate", would seek to "confirm" a rather strong prior and would have different sensitivities to false-negative and false-positive conclusions. Without here arguing the merits of these philosophical approaches to evaluation, we note that either perspective can be embodied in the decision rule (8).

3. PROBABILITY DISTRIBUTIONS OF THE ESTIMATES OF EXPERIMENTAL IMPACT

We noted above that the estimation errors conditional on a

prior estimate of experimental impact R_o have distributions which depend (through (5) and (6)) on m, the number of events in the next baseline day if the baseline is extended, and on n, the number of events in the trial period. This section derives the probability distributions of the random variables m and n and their joint distribution.

The number of events m in any baseline day has, by assumption, a Poisson distribution with parameter λ_b. Thus

$$\text{Prob}\,[m|\lambda_b] = \exp\,[-\lambda_b]\ \lambda_b^{\,m}/m! \quad , \ m \geq 0 \quad . \qquad (9)$$

Now the exact value λ_b cannot be known, but a Bayesian estimate can be made from the baseline data. Assume that before the baseline data collection begins we have a diffuse prior distribution for λ_b (one could instead chose a logarithymically flat prior or a gamma prior with no essential change in the form of the results to be derived). It is well known [2] that updating this prior with the baseline data of N_b Poisson events in I_b days leads to a gamma posterior distribution for λ_b

$$f(\lambda_b) = \frac{I_b^{\,N_b+1}}{N_b!}\ \lambda^{N_b} \exp\left[-I_b\lambda_b\right] \quad , \ \lambda_b \geq 0 \quad . \qquad (10)$$

Combining (9) and (10) we get the unconditional distribution of m, the count in the next baseline day

$$\text{Prob } [m] = \int_{\lambda_b=0}^{\infty} \left[\exp\ [-\lambda_b]\ \lambda_b{}^m/m!\right] \times \left[\frac{I_b{}^{N_b+1}}{N_b!}\ \lambda^{N_b}\ \exp\ [-I_b\lambda_b]\right] d\lambda_b \quad (11)$$

$$= \frac{I_b{}^{N_b+1}}{N_b!\ m!} \int_{\lambda_b=0}^{\infty} \lambda_b{}^{N_b+m}\ \exp\left[-(I_b+1)\lambda_b\right]\ d\lambda_b \quad (12)$$

$$= \frac{I_b{}^{N_b+1}}{N_b!\ m!} \quad \frac{(N_b+m)!}{(I_b+1)^{N_b+m+1}} \quad (13)$$

$$= \binom{N_b+m}{m} \left(\frac{I_b}{I_b+1}\right)^{N_b+1} \left(\frac{1}{I_b+1}\right)^{m} \quad , m \geq 0 \ . \quad (14)$$

A similar analysis holds for n, the number of events in the trial period, conditional on the estimates R_o of experimental impact and λ_b of baseline rate. During a trial period of duration I_t the conditional count will be Poisson

$$\text{Prob } [n|R_o,\ \lambda_b] = \exp\ [-\ R_o\lambda_b I_t]\ (R_o\lambda_b I_t)^n/n! \quad , n \geq 0 \ . \quad (15)$$

We can next use (10) to obtain

$$\text{Prob } [n|R_o] = \int_{\lambda_b=0}^{\infty} \left[\exp\ [-R_o\lambda_b I_t]\ (R_o\lambda_b I_t)^n/n!\right] \times \left[\frac{I_b{}^{N_b+1}}{N_b!}\ \lambda^{N_b}\ \exp\ [-I_b\lambda_b]\right] d\lambda_b \quad (16)$$

$$= \left(\frac{R_o I_t}{n!}\right)^n \quad \frac{I_b{}^{N_b+1}}{N_b!} \int_{\lambda_b=0}^{\infty} \lambda_b{}^{N_b+n}\ \exp\left[-(R_o I_t + I_b)\lambda_b\right]\ d\lambda_b \quad (17)$$

$$= \frac{(R_o I_t)^n}{n!} \quad \frac{I_b{}^{N_b+1}}{N_b!} \quad \frac{(N_b+n)!}{(R_o I_t + I_b)^{N_b+n+1}} \quad (18)$$

$$= \binom{N_b+n}{n} \left(\frac{I_b}{R_o I_t + I_b}\right)^{N_b+1} \left(\frac{R_o I_t}{R_o I_t + I_b}\right)^n \quad , n \geq 0 \ . \quad (19)$$

We note that since the counts m and n both appear in expression (6) for $\hat{R}_{next}$, we require their joint distribution. Conditional on R_o and λ_b, the counts m and n are independent Poisson variates with parameters λ_b and $R_o\lambda_b I_t$, respectively (since they arise in non-overlapping time periods).

$$\text{Prob } [m,n|R_o,\lambda_b] = \text{Prob } [m|\lambda_b] \times \text{Prob } [n|R_o,\lambda_b] \qquad (20)$$

$$= \left[\exp\ [-\lambda_b]\lambda_b{}^m/m!\right] \times \left[\exp\ [-R_o\lambda_b I_t]\ (R_o\lambda_b I_t)^n/n!\right] \quad . \qquad (21)$$

Again using (10) to uncondition with respect to λ_b

$$\text{Prob } [m,n|R_o] = \int_{\lambda_b=0}^{\infty} \left[\exp\ (-\lambda_b)\lambda_b{}^m/m!\right] \times \left[\exp\ (-R_o\lambda_b I_t)\ (R_o\lambda_b I_t)^n/n!\right]$$

$$\times \left[\frac{I_b{}^{N_b+1}}{N_b!}\ \lambda^{N_b}\ \exp\ (-I_b\lambda_b)\right] d\lambda_b \qquad (22)$$

$$= \frac{(R_o I_t)^n\ I_b{}^{N_b+1}}{m!\ n!\ N_b!} \int_{\lambda_b=0}^{\infty} \lambda_b{}^{m+n+N_b} \exp\left[-(1+R_o I_t+I_b)\lambda_b\right] d\lambda_b \qquad (23)$$

$$= \frac{(R_o I_t)^n\ I_b{}^{N_b+1}}{m!\ n!\ N_b!}\ \frac{(m+n+N_b)!}{(1+R_o I_t+I_b)^{m+n+N_b+1}} \qquad (24)$$

$$= \left(\frac{(N_b+m+n)!}{m!\ n!\ N_b!}\right) \times \left(\frac{I_b}{1+R_o I_t+I_b}\right)^{N_b+1}$$

$$\times \left(\frac{R_o I_t}{1+R_o I_t+I_b}\right)^n \times \left(\frac{1}{1+R_o I_t+I_b}\right)^m \quad . \qquad (25)$$

Note that the counts m and n are not independent, since (25) is not (14) times (19). Expressions (19) and (25), together with (5) and (6), would be used in the decision rule (8) to decide, after observing N_b events in I_b days, whether or not to terminate the baseline period. Unfortunately, even for a simple disutility function such as $(R - R_o)^2$, it does not appear to be possible to obtain an analytical expression for the decision rule (8) or even for the conditional decision rule (7). It should be quite easy to obtain numerical results on a digital computer, however.

To summarize, terminating the baseline period at I_b, having observed a total of N_b events and believing the true experimental impact to be R_o, one would expect estimates of impact of the form

$$\hat{R}_{now} = \left(\frac{I_b}{I_t}\right)\left(\frac{n}{N_b}\right) \qquad n \geq 0 \qquad (5)$$

where

$$\text{Prob } [n] = \binom{N_b+n}{n}\left(\frac{I_b}{R_o I_t+ I_b}\right)^{N_b+1} \left(\frac{R_o I_t}{R_o I_t+I_b}\right)^n \quad . \qquad (19$$

If one were to extend the baseline by one day, the possible estimates would be of the form

$$\hat{R}_{next} = \left(\frac{I_b+1}{I_t}\right)\left(\frac{n}{N_b+m}\right) \qquad n\ ,\ m \geq 0 \quad . \qquad (6)$$

where

$$\text{Prob } [n,m] = \left(\frac{(N_b+m+n)!}{N_b!\ m!\ n!}\right)\left(\frac{I_b}{1+R_oI_t+I_b}\right)^{N_b+1}\left(\frac{R_oI_t}{1+R_oI_t+I_b}\right)^{n}\left(\frac{1}{1+R_oI_t+I_b}\right)^{m} \quad (25)$$

If, instead of the diffuse prior, one were to choose a gamma prior for the baseline rate λ_b

$$f(\lambda_b) = \frac{T^{k+1}}{k!}\ \lambda_b^{\,k}\ \exp\ (-T\lambda_b) \quad (26)$$

then one merely replaces I_b with I_b+T and N_b with N_b+k in the probability expressions, i.e.,

$$\text{Prob } [n] = \binom{N_b+k+n}{n}\left(\frac{I_b+T}{R_oI_t+I_b+T}\right)^{N_b+k+1}\left(\frac{R_oI_t}{R_oI_t+I_b+T}\right)^{n} \quad (27)$$

and

$$\text{Prob}[n,m] = \left(\frac{(N_b+k+m+n)!}{(N_6+k)!\,m!\,n!}\right)\left(\frac{I_b+T}{1+R_oI_t+I_b+T}\right)^{N_b+k+1}\left(\frac{R_oI_t}{1+R_oI_t+I_b+T}\right)^{n}\left(\frac{1}{1+R_oI_t+I_b+T}\right)^{m} \quad (28)$$

4. COMPARISON OF CONTINGENT AND FIXED-LENGTH BASELINES BY MONTE CARLO SIMULATION

We have developed above an algorithm for contingent termination of the baseline period in a before-and-after type experiment. Given a prior guess about the experimental impact R_o, a prior for the baseline rate λ_b, a set of disutilities expressed by $U(\hat{R}|R_o)$ and C_b, and the current baseline duration I_b and count of events N_b, one can numerically evaluate the decision rule (7) and decide whether or not to stop the baseline.

In this section we address the issue of comparing this contingent approach with fixed-length baselines. Several fixed-length alternatives come to mind: (1) equal baseline and trial periods; (2) longer baseline, so that equal costs are devoted to baseline and trial periods; (3) longer trial period, so that the (presumably) smaller rate of events during the trial period can be estimated with equal precision. Whatever the choice of fixed-length alternative, we require both a measure of comparative performance and a mechanism for estimating the measure. The measure to be used is based on that used in the decision rule; the mechanism is Monte Carlo simulation.

Earlier we based the decision to terminate or extend the baseline on a measure which combined the cost of error in estimation of the experimental impact with the cost of the baseline data collection. To be consistent, we must use this same measure, only now we compare the contingent baseline against a fixed-length alternative. Let c index the contingent case and f the fixed-length case, and let the true measure of experimental impact be R^*

Then the performance measure

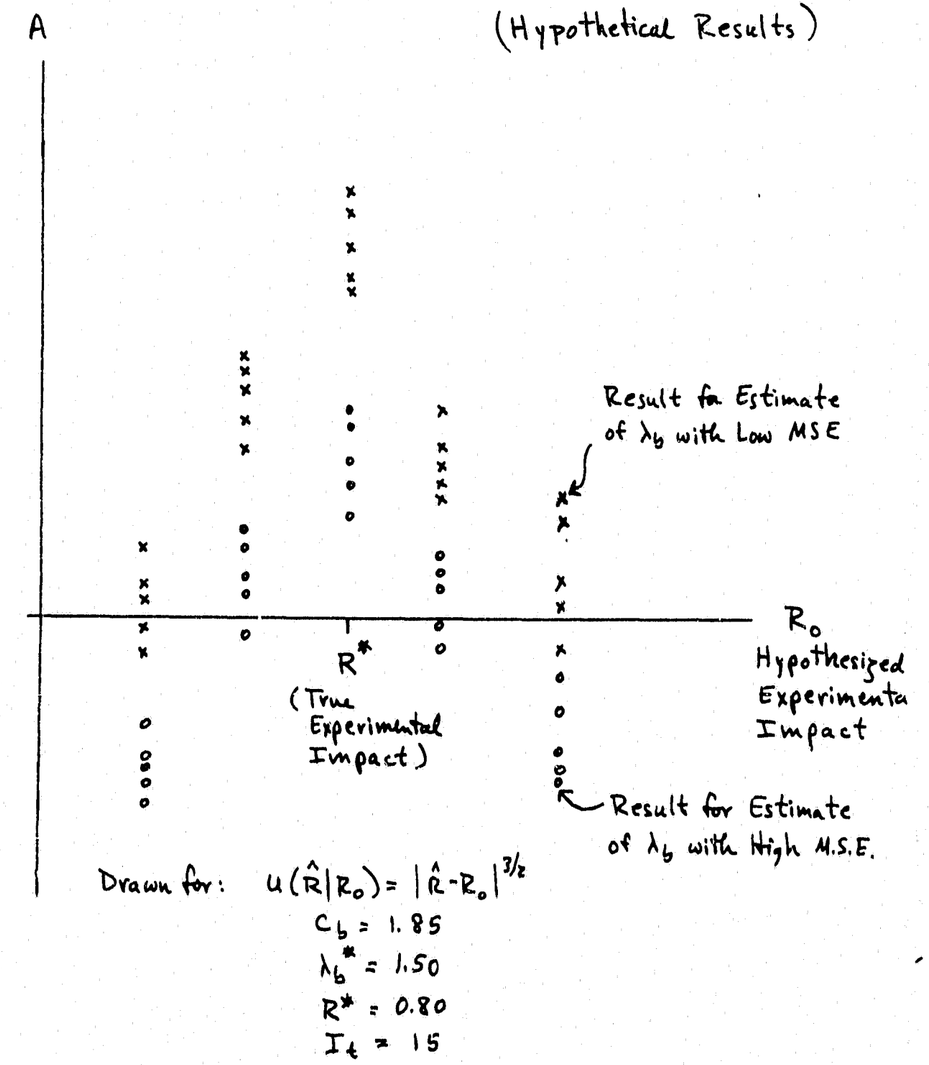
$$A = \left[U_f(R^*)+C_b I_{bf}+C_t I_t\right] - \left[U_c(R^*)+C_b I_{bc}+C_t I_t\right] \quad (29)$$

$$= \left[U_f(R^*)-U_c(R^*)\right] + \left[C_b(I_{bf}-I_{bc})\right] \quad (30)$$

Thus the relative advantage A depends on the difference in estimation errors and the difference in baseline durations (and therefore costs). Any value of $A > 0$ indicates that the contingent approach out-performed the fixed length alternative.

One would expect that the relative advantage of the contingent approach would depend on five factors: the prior distribution of experimental impact, the prior distribution of the baseline rate, the length of the trial period, the form of the disutility function for estimation errors, and the cost of baseline data collection. All five of these factors represent parameters in a Monte Carlo simulation, so a full investigation of the relative merits of the contingent approach promises to be a rather large undertaking. For any given settings of the last three factors, we could systematically explore the dependence of A on the first two factors. A useful format for reporting the simulation results would be as shown in Figure 1. Any prior distribution for λ_b could be summarized by its mean squared error ("M.S.E.") around the true value. For a given M.S.E one would choose a prior estimate of impact R_o and run several simulations

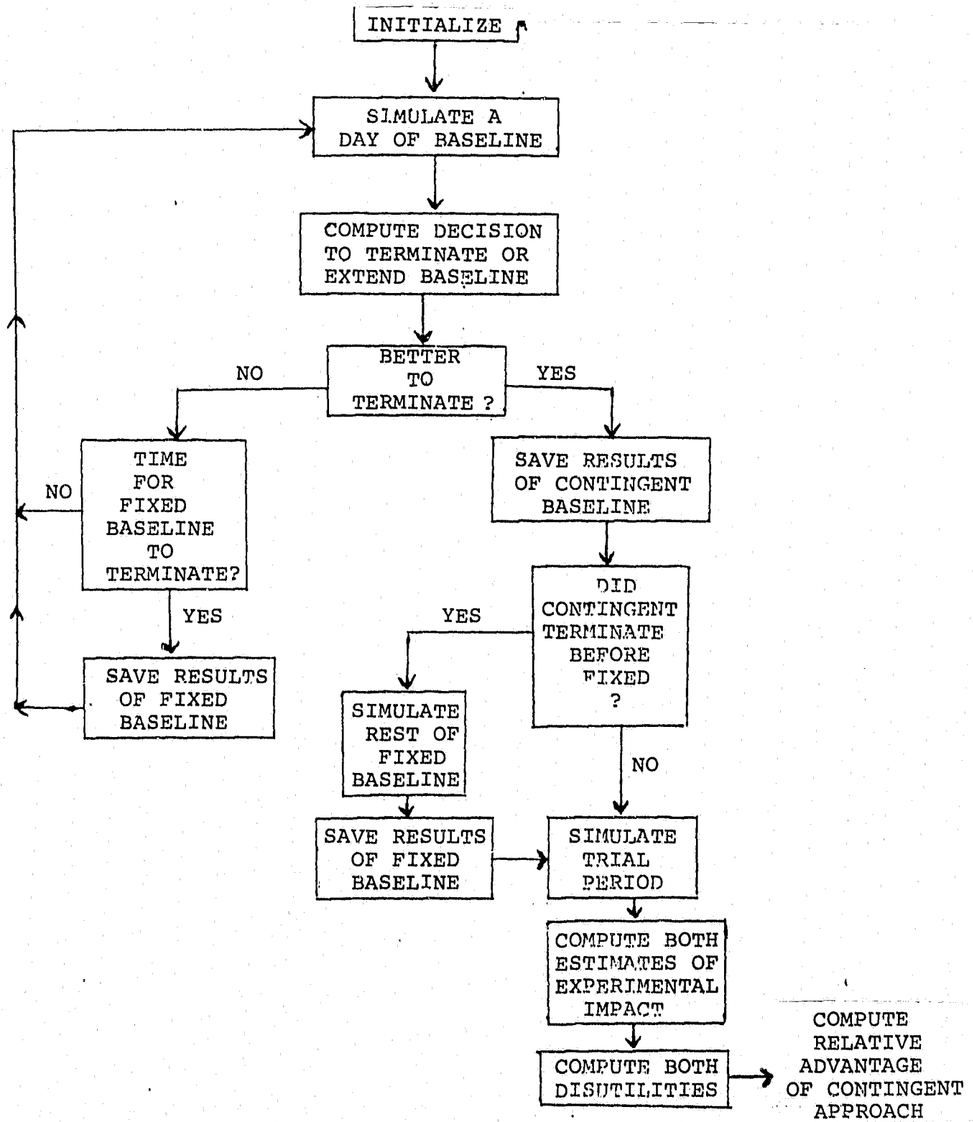
FIGURE 1: FORMAT FOR DISPLAY OF MONTE CARLO COMPARISON OF CONTINGENT vs. FIXED-LENGTH BASELINES



to form an estimate of the relative advantage of the contingent approach.

The simulations themselves would be executed in accordance with the flowchart shown in Figure 2. Initialization involves selection of those parameters which will not be varied during the course of one simulated experiment, such as the prior distributions for baseline rate λ_b and the hypothesized experimental impact R_o, their actual values λ^*_b and R^*, the duration of the fixed-length alternative I_{bf}, the duration of the trial period I_t, and the disutilities $u(\hat{R}|R_o)$ and C_b. Simulation of a day of baseline or trial requires the generation of Poisson counts at a given rate for a given period of time. Numerical solution of (7) constitutes evaluation of the decision rule. The output of any given simulation would be a value of A, the relative advantage of the contingent approach.

FIGURE 2: FLOWCHART OF SIMULATION COMPARING CONTINGENT WITH FIXED-LENGTH BASELINE



References

[1] Thomas R. Willemain, Bayesian Analysis of Crime Rate Changes in Before-And-After Experiments, M.I.T., Working Paper, OR#075-78, June 1978.

[2] S.A. Schmitt, Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics, Addison Wesley, Reading, MA, 1969.

[3] H. Raiffa and R. Schlaifer, Applied Statistical Decision Theory, M.I.T. Press, Cambridge, MA, 1961.

END