Research Background Paper

**Maximizing Internal and External Validity:
Recommendations for a Children and Youth
Savings Account Policy Demonstration**

Sondra Beverly and Michael Sherraden

Research Background Paper 01-3

September 2001

Center for Social Development

Maximizing Internal and External Validity: Recommendations for a Children and Youth Savings Account Policy Demonstration

Sondra Beverly
sbeverly@ku.edu

Michael Sherraden
sherrad@gwbmail.wustl.edu

CYSAPD 01-3

September 2001

This paper makes specific recommendations to the Corporation for Enterprise Development (CFED) regarding program and research design for a Children and Youth Savings Account Policy Demonstration (CYSAPD). The first section discusses internal validity, the second discusses external validity, and the third discusses attrition, which threatens both internal and external validity.

1. Maximizing Internal Validity

One primary goal of the CYSAPD evaluation is to attribute observed outcomes to the intervention. In other words, we want to establish, as much as possible, a cause-effect relationship. If observed outcomes can be attributed to the intervention, an evaluation is said to have internal validity. The best way to achieve internal validity is to have two (or more) groups that are equivalent in every way except in the intervention. One or more groups (treatment groups) receive an intervention, and one group (the control or comparison group) does not receive the intervention.

To the extent that groups are comparable on all relevant characteristics (including predisposition toward the program), all of the threats to internal validity described by Campbell and Stanley (1963) (i.e., history, maturation, testing, instrumentation, regression, selection, mortality, and the interaction of selection and maturation) are controlled. Therefore, differences in outcomes should be largely attributable to intervention effects (Johnson & Stromsdorfer, 1990; Rossi & Freeman, 1993).

In this section, we offer two recommendations to increase internal validity.

1A. Use a true experimental research design at one or more large demonstration sites.

A true experiment uses random assignment to create control and treatment groups and thus maximizes the probability that groups will be equivalent.[1] Simple instructions for administrators on a variety of randomization techniques may be found in Fitz-Gibbon and Morris (1978, ch. 8).

1B. If a true experiment is not possible at one or more large demonstration sites, use a quasi-experimental design.

Although randomized experiments are clearly the most powerful, in some cases randomization is infeasible. Quasi-experiments compare outcomes for treatment groups with outcomes for comparison groups that did not receive the intervention, but individuals are not randomly assigned to these groups. If executed properly, quasi-experiments can produce credible results (Posavac & Carey, 1989). However, researchers must make each threat to internal validity explicit and rule each out individually (Cook & Campbell, 1976).

---

[1] Some have voiced ethical concerns about depriving individuals in a control group of the benefits that accompany program participation. However, it is presently impossible for all members of the target population to receive CYSAs, and randomization would give each individual an equal chance of receiving an account (Rossi & Freeman, 1993).

Center for Social Development
Washington University in St. Louis

A quasi-experiment in CYSAPD would likely use constructed controls. This technique involves selecting a comparison group to be as equivalent as possible to the group of individuals who participate in the intervention. Based on a theoretical understanding of relevant social and psychological processes, evaluators choose matching variables that are likely to impact outcome variables. Matching may take place on an individual level, so that each member of the treatment group is matched with a "partner" with similar relevant characteristics, or on an aggregate level, so that the overall distribution of matching variables is similar across both groups. Although individual matching is more precise, it is also expensive, time-consuming, and difficult. It may also result in the loss of a number of experimental cases if appropriate matches cannot be found. In this case, the successful matches may no longer be representative of the sample universe.

The use of constructed controls can be combined with statistical control techniques.

2. Maximizing External Validity

A second primary goal of the evaluation is to be able to generalize findings beyond the research sample. A research design with results that may be generalized beyond the research sample is said to have external validity. External validity in the CYSAPD evaluation will allow us to suggest that a similar CYSA program would have similar outcomes for other individuals and perhaps to suggest that a universal CYSA policy would produce similar outcomes.

In this section, we offer four recommendations to increase external validity.

2A.   When evaluating applicants for the large demonstration sites, CFED should consider how well the resulting research sample would represent the larger low-income population in the U.S.

It would be undesirable, for example, for a large site to serve only a very specific group, such as single-parent families living in a high-poverty inner-city neighborhood or new immigrants from Central America. Along this line, the Request for Proposals should state that the ability to generalize beyond the research sample is an important consideration.

2B.   For the large sites, CFED should choose settings where all participants may be defined as CYSAPD participants.

CFED might choose a school district where all 8th graders (who consent to participate) are CYSAPD participants or a Head Start site where all new entering three-year-olds (whose parents consent to participate) are CYSAPD participants. In these scenarios, there is no need to recruit participants. When participants must be recruited, selection bias becomes a substantial concern. Individuals who volunteer for an intervention are almost certainly different from those who don't volunteer. For example, volunteers are by definition motivated to participate in the program and presumably expect to receive some benefit from the intervention. These differences between study participants and non-participants limit the ability to generalize beyond the research sample.

2C. At the large sites, CFED should aim for account structures that are as consistent as possible with the account structure envisioned under a universal CYSA policy.

For example, if a universal CYSA policy is expected to automatically provide accounts at birth, then CYSAPD participants should automatically receive accounts. Likewise, if a universal CYSA policy is likely to require individuals to take action to open accounts, then CYSAPD participants should have to take action to open account. In this way, empirical findings from the demonstration will be more representative of outcomes under a universal policy.

2D. CFED should encourage sites to make program participation as convenient as possible for participants.

Making participation convenient will increase external validity because individuals who fully participate in CYSAPD are less likely to be atypical (Cook & Campbell, 1976).

3. Minimizing Attrition

Attrition, the loss of sample members, can threaten both internal and external validity. It threatens external validity if the participants who remain in the study no longer represent the population from which they were drawn or the population to which generalization is desired. Attrition threatens internal validity if the treatment and control/comparison groups become less comparable.

In all studies involving human participants, some attrition is inevitable. Because this evaluation of CYSA programs is fairly lengthy and involves a low- to moderate-income population, an above-average attrition rate should be expected. With these obstacles in mind, we recommend a four-pronged strategy to reduce the effects of attrition:

3A. Select a sample which is large enough to handle fairly high attrition rates.

In addition, because individuals who do not receive CYSAs may be more likely to drop out of the study (Cook & Campbell, 1976) and more difficult to track, the control group should be larger than the treatment groups.

3B. Use a variety of techniques to keep participants interested in the study.

These techniques can include making frequent contact with participants through post cards, greeting cards, and reminders, and offering financial incentives for participation in surveys and interviews.

3C. Use a variety of techniques to track participants as their addresses or phone numbers change.

During the initial interview, participants should be asked to give addresses and phone numbers of relatives, friends, employers, social workers, and others who will always know how to find them. We should also ask each individual to give written permission allowing the organization (e.g., school, Head Start site) to release his or her current address and phone number to CYSAPD representatives.

Following the initial interview, holiday cards should be sent regularly to all participants in envelopes marked "Address Correction Requested, Do Not Forward." In this greeting card should be a return card asking participants for any comments or new information they wish to share and any address or phone number updates. When a card comes back with an address change, a new card should be sent to the participant. If a card is returned without a valid address, we should immediately begin tracking procedures via relatives, friends, employers, and so forth. More extensive tracking methods can also use motor vehicle departments, credit bureaus, military branches, and the correctional system. These methods have been successful in research projects involving high-risk adolescents (Gwadz and Rotheram-Borus, 1992).

3D. Because some attrition will inevitably occur, use post-hoc data analysis strategies to compensate for attrition.

One strategy, described by Mark and Cook (1984, pp. 93-4), involves identifying subgroups that have not experienced differential attrition. Other strategies involve statistical controls.

In sum, the attrition problem should be handled through a combination of anticipation, prevention, and statistical compensation.

REFERENCES

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Cook, T. D., & Campbell, D. T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 223-326). New York: John Wiley and Sons.

Fitz-Gibbon, C. T., & Morris, L. L. (1978). *How to design a program evaluation*. Beverly Hills, CA: Sage Publications.

Gwadz, M., & Rotheram-Borus, M. J. (1992). Tracking high-risk adolescents longitudinally. *AIDS Education and Prevention, Supplement,* 69-82.

Johnson, T. R., & Stromsdorfer, E. W. (1990). Evaluating net program impact. In A. B. Blalock (Ed.), *Evaluating social programs at the state and local level: The JTPA evaluation design project* (pp. 43-131). Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.

Mark, M. M., & Cook, T. D. (1984). Design of randomized experiments and quasi-experiments. In L. Rutman (Ed.), *Evaluation research methods: A basic guide* (2nd ed., pp. 65-120). Beverly Hills, CA: Sage Publications.

Posavac, E. J., & Carey, R. G. (1989). *Program evaluation: Methods and case studies* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Rossi, P. H., & Freeman, H. E. (1993). *Evaluation: A systematic approach* (5th ed.). Newbury Park, CA: Sage Publications.

Center for Social Development
Washington University in St. Louis