

SRDC Technical Paper Series 03-01

**How Random Must Random Assignment Be
in Random Assignment Experiments?**

Paul Gustafson

Department of Statistics
University of British Columbia

February 2003

SOCIAL RESEARCH AND DEMONSTRATION CORPORATION

This paper is part of the Social Research and Demonstration Corporation's program of analysis for the Self-Sufficiency Project (SSP) sponsored by Human Resources Development Canada (HRDC). The Self-Sufficiency Project is sponsored by HRDC. This paper was produced for SRDC. The opinions expressed herein are the authors' and do not necessarily reflect those of SRDC or HRDC.

The Social Research and Demonstration Corporation is a non-profit organization and registered charity with offices in Ottawa, Vancouver, and Sydney, Nova Scotia. SRDC was created specifically to develop, field test, and rigorously evaluate social programs. SRDC's two-part mission is to help policy-makers and practitioners identify social policies and programs that improve the well-being of all Canadians, with a special concern for the effects on the disadvantaged, and to raise the standards of evidence that are used in assessing social policies. As an intermediary organization, SRDC attempts to bridge the worlds of academic researchers, government policy-makers, and on-the-ground program operators. Providing a vehicle for the development and management of complex demonstration projects, SRDC seeks to work in close partnership with provinces, the federal government, local programs, and private philanthropies.

Copyright © 2003 by the Social Research and Demonstration Corporation

Table of Contents

Introduction	1
Two Kinds of Randomization	3
Purely Statistical Issues	5
Gaming the System?	7
The Possibility of Blocked Analysis	9
Summary	11
References	13

Introduction

It is well known that the impact an intervention *causes* in a population is best estimated in a study where each participant is *randomized* to either receive the intervention or not. With an observational study comparing participants who choose the intervention with those who do not, an attempt can be made to estimate the causal effect of intervention. In particular, a statistical analysis that adjusts for measured *confounding* variables can be used (i.e. some sort of *regression*). The problem is that seldom, if ever, can the existence of unmeasured confounding variables be ruled out. Hence it is uncertain whether the estimated impact is entirely attributable to the intervention. On the other hand, randomization guarantees that the two groups are very likely to be quite balanced in terms of all potential confounding variables, both measured and unmeasured.

Social policy researchers recognize that randomized studies can yield definitive findings whereas observational studies typically lead to speculative findings. In Canada the Social Research and Demonstration Corporation (SRDC) undertakes large studies of social program impacts using random allocation study designs. An issue that confronts SRDC is that there are different kinds of randomization, and indeed some such schemes appear to be “more random” than others. On the other hand, there are pragmatic reasons why the most random schemes may not be attractive in the specific kinds of studies undertaken by SRDC. The function of this paper is to compare the different kinds of randomization, focussing particularly on whether current SRDC practice is ideal in light of both theoretical and practical circumstances.

Two Kinds of Randomization

Say that an upcoming enrolment period of a random assignment experiment will consist of the next 20 subjects who are eligible and provide informed consent. Consider two different strategies for allocating each enrollee to either the *control* group (C) or the *program* group (P).

Approach A: As each subject enrolls, flip a fair coin (be it real or computer-simulated) to decide the subject's assignment. For instance, the assignment pattern generated might be

C P C C P C C C C P C P C C P C C C P P.

Approach B: At the start of the enrolment period, generate a purely random permutation of ten C's and ten P's, that will then govern assignments as the subjects enter the study. For instance, the computer might generate

P P C C C P C P C C C P C P P P C C P P.

In statistical parlance, approaches A and B might be referred to as *full* and *restricted* randomization respectively. The obvious difference between A and B is that with the former scheme the sizes of the program and control groups are not known in advance. The laws of probability dictate that *on average* 10 subjects will be assigned to each group, and that a severe imbalance between the group sizes is unlikely. But, for instance, the 7–13 split in the example assignment pattern does not constitute a particularly surprising degree of imbalance. On the other hand, Approach B guarantees equal-sized groups.

Potentially, Approach B has both a practical advantage and a statistical advantage. If the number of available spots in the program group during the enrolment period is fixed (at 10 in this case), then obviously Approach B avoids the logistical problem of the program being under or over-subscribed. On the statistical side, Approach B can help in the following way. Say that subjects become eligible to enter the study when they become unemployed. It may be the case that most or all of the enrollees in the upcoming enrolment period are entering the study because their former employer (Company C, say) reduced its workforce. Similarly, during a later enrolment period most or all of the enrollees might have been formerly employed at Company D. If the two companies are quite different in nature, then the enrollees in the earlier enrolment period may differ somewhat systematically from those in the later period in terms of characteristics such as age, education, and previous income. If so, the use of Approach B is more likely to create a better balance of these characteristics across the control and program groups.

On the other hand, a concern about Approach B is an apparent lack of randomness for subjects who enrol late in the enrolment period. Consider an observer who knows that 10 of the 20 enrollees will be assigned to P, and also knows that of the 17 enrollees thus far, 8 have been assigned to P. The observer correctly gauges the probability that the 18th enrollee will get P, *given what he or she knows*, to be $2/3$. Or, more extremely, say that of the 18 enrollees

thus far, 8 have been assigned to the P group. The observer who knows this can infer with certainty that the remaining two enrollees will be assigned to the P group.

Succinctly, both approaches give each enrollee a 50-50 chance of being assigned to the program group. However, under Scenario A each assignment is independent of every other assignment. In Scenario B the assignments are not independent, as is evidenced most clearly by the assignments near the end of the enrolment period. The purpose of this paper is to review the pros and cons of foregoing this independence.

Purely Statistical Issues

It is well known that the only definitive way to estimate the causal effect of an intervention on an outcome is to apply statistical analysis to data obtained by randomly allocating subjects to the intervention and to a control group. In some sense the lack of independence in Approach B might promote the view that this is an “inferior” sort of randomization relative to Approach A. In fact there is no call for concern of this sort, provided the following assumption is valid.

Assumption I: The enrollees will enrol in the same temporal order, regardless of what sequence of assignments is generated in advance.

Under this assumption it is intuitively clear that Approach B precisely corresponds to forming two groups by randomly selecting 10 of 20 subjects. In particular, the assumption lets the enrollees be conceptualized as fixed and the assignments as random.

More generally, think of N enrollees. “About” n of them can be chosen to be the program group by flipping a biased coin (probability n/N for the program group) for each subject. Or exactly n of them can be chosen by selecting at random one of the many subsets of n of the N enrollees. This is most simply understood when n and N are small. Say, for instance, that there are $N = 5$ enrollees, labelled A, B, C, D, and E. And say $n = 2$ of these enrollees are to be assigned to the program group. A quick count reveals there are 10 different subsets of two participants (AB, AC, AD, AE, BC, BD, BE, CD, CE, DE). One of the ten subsets could be selected at random, that is each subset has a one-in-ten chance of being selected. (More generally, the number of subsets is referred to as “N-choose-n,” which can be computed mathematically as $N!/\{n!(N - n)!\}$. Then each subset has a “one in N-choose-n” chance of being selected.) This formalizes the notion of a random permutation in Approach B.

The key point to be made is that *for statistical purposes both kinds of randomization are just as good, and statisticians do not tend to regard Approach B as less random than Approach A*. For instance, in the sample survey literature the fundamental definition of a random sample from a population is based on the restricted randomization of Approach B (see, for instance, Lohr, 1999). Regardless of which randomization scheme is used, the same statistical methods of analysis apply, and the results of the analysis have the same interpretation about the causal effect of the intervention. As a more specific example, say that X represents a numerical outcome measure of interest, and $\bar{X}_P$ and $\bar{X}_C$ are the average values of X among the program group (P) and the control group (C) respectively, so that $\bar{X}_P - \bar{X}_C$ estimates the program’s impact on X . Under either type of randomization is it statistically legitimate to treat $\bar{X}_P$ and $\bar{X}_C$ as independent random variables, so that exactly the same procedures to generate confidence intervals and hypothesis tests can be brought to bear.

Thus, provided Assumption I is valid, Approach B is not inferior to Approach A in any statistical sense, and in fact may have the benefits alluded to in the last section. Given this, it is useful to consider the validity of Assumption I. In the kinds of studies being considered, enrollees often have a strong preference for being in the program group rather than the control group. Given the means to do so, some enrollees would likely attempt to boost their chances

of being assigned to the program group. Or there is a possibility that investigators, perhaps subconsciously, would try to influence the assignment process. Presumably the study design is “double-blinded” in that both enrollees and investigators do not know the future assignments (i.e. only the computer knows). As discussed above, however, Approach B entails some scope for “better than chance” guessing of future assignments having observed past assignments. Thus the next section discusses the extent to which “gaming the system” is possible.

Gaming the System?

The current practice of the Social Research and Demonstration Corporation (SRDC) is to generate a long sequence of assignments by “gluing together” blocks, where each block has the desired ratio of program group (P) and control group (C) assignments. For ease of description, assume a 50-50 allocation is desired so that each block has the same number of P and C assignments. To limit an outside observer’s ability to predict the next assignment given those already made, the sequence is constructed from a random ordering of blocks of different sizes. There is a desire to keep the block sizes small so that as enrolment proceeds the proportion of enrolled subjects assigned to the program group cannot deviate very much from the target proportion.

Even with small blocks, the use of a random ordering of variable sized blocks is an effective guard against an outside observer predicting the next assignment given the previous ones *with certainty*. However, a shrewd outside observer who gets to observe a consecutive string of previous assignments may be able to make a “better than random” guess at the next assignment.

In particular, say the observer is a future enrollee who wishes to time entry into the study in way that increases his or her chances of being assigned to the program group. For the sake of example, say that an assignment sequence of length 24 is constructed from two blocks of size 2, two blocks of size 4, and two blocks of size 6. In addition to restricted randomization within each block, the order in which the blocks appear is also randomized. Furthermore, once the sequence is constructed, the starting point is also chosen at random. The future enrollee might try to game the system by waiting until two consecutive assignments are to C and then enrolling. A simple computer simulation of the process shows that the gamer increases his or her probability of being assigned to P from 50 per cent to 74 per cent. The gamer will not have to wait long to enact his or her strategy, as for a given pair of consecutive enrollees there is a 19 per cent chance they will both be assigned to C. (The latter fact is also a consequence of the short blocks. For a sequence generated by fair coin flips there is a 25 per cent chance that a given consecutive pair of assignments will both be C.)

As a more extreme example, say a more patient gamer waits until three consecutive allocations are to C and then enrolls. The computer simulation shows that this gamer increases his or her probability of being assigned to P from 50 per cent to 82 per cent. This gamer will have to wait longer on average to enact his or her strategy, as for a given trio of consecutive enrollees there is only a 5 per cent chance that all three will be assigned to C (as opposed to a 12.5 per cent chance for a fair coin flip sequence).

Intuitively the effectiveness of the gamer’s strategy can be weakened by using larger blocks. Say the sequence of size 24 is instead created from one block of size 4, one block of size 8, and one block of size 12. Again a random block order and random starting point are used. The gamer who enrolls after two consecutive allocations to C now has a 63 per cent chance of being allocated to P (and the chance that a given pair of consecutive enrollees are both allocated to C is 22 per cent). The more patient gamer who enrolls after three consecutive

allocations to C has a 67 per cent chance of being allocated to P (and for a given trio of consecutive enrollees there is an 8 per cent chance that all three are allocated to C).

Of course this discussion of gaming may not be relevant if outside observers do not have access to sufficient information to enact such gaming strategies. For a given study, it may be difficult for an observer outside of the investigative team to observe all the allocations up to the present time. Moreover, even if a future enrollee could monitor the past assignments, the study protocol may offer very limited ability for that future enrollee to time his or her own enrolment.

As mentioned above, there is also the issue of double-blinding. Presumably the investigators are not privy to the entire computer-generated assignment sequence in advance, but rather are given the assignments one by one as the subjects enrol. Thus an investigator's ability to predict the next assignment are essentially no better than that of an outside observer. If the investigators are able to influence the order in which subjects enrol, then there would be some potential for biasing (either conscious or subconscious) of the assignment process.

The Possibility of Blocked Analysis

Irrespective of how the assignment to the control group (C) and the program group (P) is implemented, there is the possibility that a *blocked* statistical analysis would be more efficient than an unblocked or “regular” analysis. For instance, say a study involves 10 consecutive enrolment periods. If enrolment is influenced by major employers reducing their workforce at particular times, then there may be a tendency for enrollees in the same enrolment period to be more homogeneous than enrollees across different enrolment periods. If this homogeneity is evident in characteristics that might influence the outcome variable (for example, level of education, age, skills), then a blocked analysis might be more efficient.

To be more specific, each enrolment period could be treated as a block. (In fact, this is largely for the sake of argument. In principle, an enrolment period could be subdivided into multiple blocks, or consecutive enrolment periods could be amalgamated into a block.) Roughly put, a blocked analysis estimates the P versus C effect on outcome separately within each block, and then averages across blocks to obtain an overall estimate of the program impact. In practice, with a numerical outcome variable a simple way to accomplish this is by a two-way analysis of variance (ANOVA), as opposed to a standard comparison of two means that can be regarded as a one-way ANOVA. (See, for instance, Montgomery, 2000.)

It should be emphasized that in the present context neither analysis is “wrong” in the sense of providing invalid results. The point is that one method may prove to be more efficient in statistical terms than the other. Put more bluntly, for a given sample size one method of analysis may lead to a narrower confidence interval for the program effect than does the other method. Or, more usefully, for a given desired precision one method may require a smaller sample size than the other.

What governs which method of analysis will work better is the extent to which enrollees are more homogeneous within blocks than across blocks, in terms of characteristics that really do influence the outcome variable. If this phenomenon is substantial, then a blocked analysis can be much more efficient than an unblocked analysis. On the other hand, if the phenomenon is weak or non-existent, then the blocked analysis can be less efficient, though the inefficiency is not likely to be substantial.

Thus in choosing between unblocked and blocked analysis, the following questions must be asked: (1) are enrollees at close enrolment times expected to be more similar, on average, than enrollees at distant enrolment times, and (2) are the subject characteristics that are less/more similar in (1) likely to be associated with the outcome variable. If the answers to both (1) and (2) are “yes,” then blocked analysis should be considered.

Summary

In summary, there are three main points made in this paper:

1. Restricted randomization does have the *appearance* of being less random than full randomization when applied to subjects entering a study in a definite time ordering. However, assuming that subjects will not attempt to manipulate the ordering based on earlier assignments, the two kinds of allocation schemes have the same statistical merit, and the same statistical analysis can be brought to bear on the resultant data. That is, assuming there is no successful “gaming of the system,” the practical advantages of Approach B come essentially without penalty.
2. The practice of gluing together short, randomly permuted blocks of the desired program/control ratio does ensure that the actual group sizes not differ much from the target ratio as the subjects enrol. The use of varying-sized blocks in varying orders does preclude an outside observer guessing the next assignment with certainty, given knowledge of the previous assignments. However, it does permit “better than 50-50” guessing. With short blocks there is no way to avoid this. The relevant question would seem to be whether future enrollees have access to enough information, and enough control over their own enrolment, to really influence the chance of being assigned to the program group.
3. In a study where relevant subject characteristics are suspected to vary systematically over time, there may be merit in statistical analysis that “blocks” together subjects who enrol at similar times. This may be relevant if subjects with a particular former employer tend to enter the study around the same time.

References

- Lohr, S. L. (1999). *Sampling: Design and analysis*. Pacific Grove, CA: Duxbury Press.
- Montgomery, D. C. (2000). *Design and analysis of experiments* (5th ed.). New York: Wiley.

