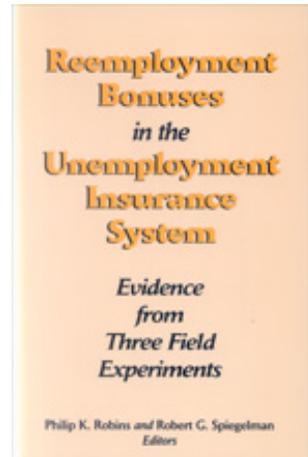

Upjohn Institute Press

Design of Three Field Experiments

Walter Corson
Mathematica Policy Research

Robert G. Spiegelman
W.E. Upjohn Institute



Chapter 2 (pp. 25-75) in:

**Reemployment Bonuses in the Unemployment Insurance System:
Evidence from Three Field Experiments**

Philip K. Robins, and Robert G. Spiegelman, eds.

Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2001

DOI: 10.17848/9780880994217.ch2

2

Design of Three Field Experiments

Walter A. Corson and Robert G. Spiegelman

INTRODUCTION: WHAT MAKES AN EXPERIMENT AN EXPERIMENT?

The modern concept of experimental design is primarily a result of the work of R.A. Fisher, who developed the concept in the 1920s while conducting agricultural field experiments in England (see Fisher 1960; Fisher 1968). Fisher's approach had several statistical features and relied on randomization as the key to providing estimates of response variability. According to Charles Hicks (1982, p. 1): "*A true experiment* may be defined as a study in which certain independent variables are manipulated, . . . and the levels of these independent variables are assigned at random to the experimental units in the study." In essence, an experiment is a process of manipulation and randomization for the purpose of measuring the underlying responses to treatments. Its value over observational studies lies in the increased ability to elucidate cause-and-effect relationships (Fisher 1968, p. 246).

Despite the history of controlled experiments in the laboratory and the agricultural fields, the idea of such an approach did not enter the realm of social science until Heather Ross, a candidate for a Ph.D. in economics at Massachusetts Institute of Technology, presented such a plan to experiment with the negative income tax in a 1966 paper (Ross 1966, cited in Kershaw and Fair 1977). Until that time, the concept of manipulating human beings in the same manner as fertilizer or wheat varieties was not considered. However, the Office of Economic Opportunity (OEO) was eagerly searching for new ways to evaluate alternative negative income tax proposals and the academic community was eager to apply newly developed econometric and data handling tech-

niques. The result was the New Jersey Income Maintenance Experiment, which was soon followed by income maintenance experiments in Seattle, Denver, Gary, and two rural areas. Experiments with housing vouchers, health insurance, and residential electricity pricing followed. New at the time was the idea that it was reasonable to establish a control group that represented a null treatment only in the sense that the particular intervention being tested was withheld. In every other sense, the control group represented the status quo—affected by all the programs, policies, and economic perturbations that existed in the general population. As we will see, experimentation in general and randomization in particular is often not as clean in practice as it is in principle.

Randomization is the process of selecting individuals from a specified population and placing them into one or more groups by a blind process that, in principle, assures that on average the characteristics of the members of the groups are the same.¹ This goal can be accomplished using a random number generator or the last four digits of an individual's social security number to make the assignment. If the process is done correctly and the samples are sufficiently large, then the members of each group will differ, on average, only due to experimental interventions. Basically, there will be two groups: a "treatment" group that receives the experimental intervention and a "control" group that receives no treatment. The number of experimental groups can be expanded, with each group receiving a different treatment. There will invariably be a control group to represent the status quo because a comparison between the treatment and control groups tells the evaluators what can be expected if the treatment is imposed in the then-current environment.

In this chapter, the design of each of the three bonus experiments is presented and the three are compared. The next section examines eligibility requirements for participation in the experiments. The experimental design is then described, comparing the nature of the treatments among the three states and pointing out their differences and their more pervasive similarities. We then discuss the number and allocation of participants in the experiment, followed by a discussion of the decisions as to the location and number of experimental sites. The environment within which each experiment was conducted, including the relevant economic characteristics of the states during the periods in which the experiments were conducted and the characteristics of the

group of claimants selected to serve as a control group, is then described. We next evaluate the randomization process, whose object was to create treatment and control groups that had essentially the same characteristics, and then we describe the operation of the experiments, pointing out the overriding similarities as well as the few differences in operational designs. The final section briefly discusses an unplanned experiment generated by the Federal Supplemental Compensation program in Illinois.

ELIGIBILITY FOR PARTICIPATION IN THE EXPERIMENTS

General Issues

Participation in an experiment should be guided by the expected decision as to who would be eligible to participate in a program modeled on the experiment. Additional considerations that might further limit experimental eligibility are the likelihood of participation in the program and a desire to increase the homogeneity of the sample, thereby reducing the sample size required to achieve a given level of statistical reliability.

Although decisions regarding the characteristics of those eligible to participate in an experiment should reflect the purposes of the experimental program, limiting the sample also limits the information obtained from the experiment. For instance, in an experimental employment program that had the goal of increasing the employability of displaced workers, limiting enrollment to only displaced workers would result in an experiment that provides no information on the relative effects of the program on displaced and nondisplaced workers. Furthermore, a change in the definition of displacement would reduce the usefulness of the results of an experiment that enrolled only displaced workers under the old definition. It is, however, a waste of resources to include persons whose behavior is of no interest in reaching the policy decisions. If the resulting program has the reduction of unemployment insurance (UI) costs as a major goal, then it would seem to be unnecessary to include in the experiment persons not eligi-

ble for UI benefits, unless UI eligibility is affected by the experimental program.

Such an effect can arise because of the potential for an experimental program to generate unwanted increases in program participation (e.g., additional filing for UI benefits). This so-called “entry effect” is usually undetected in an experimental evaluation (see Moffit 1992).² In the UI bonus experiments (henceforth referred to simply as the “bonus experiments”), entry effects could only be determined if both employed and unemployed workers (including those not filing for benefits) were enrolled. This was not the case in any of the bonus experiments, thereby precluding any experimental measurement of entry effects. This issue is discussed more completely in Chapter 6 with regard to the bonus experiments.

Eliminating from a sample some persons who would be eligible to participate in the ensuing program is done primarily to reduce costs. For example, if relatively young and relatively old workers respond to programs differently than those in the age group 25 to 55, concentrating on the middle age group would reduce the sample size needed for any given degree of reliability. Of course, the results fail to inform the policymaker about the effects of the program on the excluded age groups.

Who responds to the treatment is relevant to the determination of net benefits, because of the potential for *windfall gains*, which are gains that accrue to participants in an experiment who receive the treatment benefits (such as a bonus payment) without any change in behavior. An example would be a program to pay moving expenses to unemployed workers who take jobs outside their area of current residence; windfall gain will be the payments to those who would have moved without the reimbursement. Net benefits will be larger as the proportion of payments made to such nonresponders gets smaller. However, windfall gain cannot be reduced to zero for two reasons: 1) equity considerations require a relatively comprehensive definition of eligibility (e.g., if black workers don't respond to the treatment, you cannot design a program that makes payments only to white workers), and 2) behavioral response is measured indirectly by a comparison of group behaviors; individual response is only inferred on a probabilistic basis.

Windfall gain is minimized by excluding from eligibility groups whose members have a low probability of responding to the treatment and whose exclusion does not involve substantial equity issues. For instance, in the bonus experiments, not paying bonuses to claimants who returned to their previous jobs helped reduce windfall gain.

Eligibility to Participate in the Bonus Offer Experiments

All three experiments started from the premise that a major experimental objective was to reduce costs to the UI trust funds by reducing the length of insured unemployment. Offering a bonus to unemployed workers who were not eligible to receive UI benefits would not contribute to achieving this goal because no change in the behavior of ineligible claimants could reduce UI benefit payments. With an exception of a few cases in Washington, all three experiments required that UI eligibility be established.³ The criteria of UI eligibility meant that the claimant must have been monetarily eligible to receive benefits payable out of the host state's trust fund. Thus, veterans receiving benefits under the UCX (unemployment compensation for ex-service members) program, claimants with only interstate claims, and ex-federal employees under the UCFE (unemployment compensation for federal employees) program were all excluded because their benefits would not be paid out of the state trust funds. As shown in Table 2.1, these conditions were uniform across the three experiments; other conditions were not. The eligibility conditions in Table 2.1 are described in the sections that follow.

New benefit year

All three experiments required that the claimant be filing to establish a new benefit year. (See the Glossary on p. 275 for definitions of "benefit year" and other UI terms.) This requirement was imposed for two reasons: 1) it was felt that, in a steady state, the bonus offer would be made at the start of a benefit year and not sometime in its midst and thus the requirement would replicate that expected in a real program; and 2) it created a valuable homogeneity characteristic in that all experimental subjects would start with the same unemployment history (at least with regard to the current spell). In Washington, but not in Illinois or Pennsylvania, claimants filing "transitional" claims were eligi-

Table 2.1 Eligibility Conditions for Participation in the Experiments^a

Eligibility conditions	Illinois	Pennsylvania	Washington
To participate			
New benefit year	Y	Y	Y
Monetarily eligible for UI	Y	Y	Y
Monetarily eligible for UI at time of filing	N	N	Y
No separation issues	Y	Y	N
Not totally interstate, UCFE, or UCX	Y	Y	Y
Not on standby or a referral union member	N	Y	N
Register with job service	Y	N	N
Backdated claims more than two weeks disallowed	N	Y	N
To receive a bonus			
Terminate benefits, become employed for 4 months	Y	Y	Y
Not be recalled, placed by union	N	Y	Y
Nonmonetarily eligible to receive benefits	Y	Y	Y

^a Y: yes, it is a condition; N: no, it is not a condition.

ble to participate. In a real program, it is doubtful that a claimant who received an offer in the initial benefit year would receive a repeat offer in a transitional year.

Monetary eligibility

As mentioned above, the programmatic goal of reducing UI system costs led all three experiments to require monetary eligibility for UI as a condition for participation in the experiments. In Pennsylvania and Illinois, a monetarily valid claim had to be established at the time the claimant filed for the first compensable week and/or the waiting week.

The requirements for monetary eligibility were more stringent in Washington. Only a claimant who had a monetarily valid claim at the time of initial filing was enrolled in the experiment. Thus, certain categories of claimants, particularly laid-off state employees whose wage histories would not be in the file at the time of filing and whose monetary eligibility would not be established until after filing, were excluded from the Washington experiment. This made administration of the experiment easier and permitted fixed dollar offers to be made at the time of enrollment. It is likely that the more equitable approach used in Illinois and Pennsylvania would be followed in a real program.

Nonmonetary eligibility

Claimants who were nonmonetarily ineligible for the duration of their claim were also not eligible to receive a bonus in any of the three experiments. Duration exclusions usually occurred because of “separation” issues; i.e., the claimant had been fired for cause or had voluntarily quit (does not include “good cause” quits, such as for sexual harassment). Nonduration issues, such as not being available for work, did not preclude eligibility for any of the experiments because these conditions were removable by the decision to search actively for work, an event regarded as positive for the experiment.

Nonmonetary ineligibility was handled somewhat differently in each of the three experiments. In all experiments, bonus offers were made to all who were monetarily eligible. However, in Washington and Illinois, an offer letter (officially making the offer) would not be sent until all nonmonetary duration stops had been removed for at least one week. If the stops were removed during the qualification period, then a bonus offer was made and a bonus could be collected. In Pennsylvania, nonmonetary eligibility was checked when the bonus was claimed. Individuals who had been disqualified for the duration of their unemployment were not eligible to receive the bonus, but individuals disqualified for shorter periods could receive the bonus.

Not totally interstate, UCFE or UCX

Interstate claims are filed by claimants whose wage credits were accumulated in another state; that state is then charged with the costs of benefit payments for the interstate claimant. UCFE is the code for federal employees and UCX is the code for veterans who have been

recently discharged from the service. In both of these cases, the benefits are paid under special programs by the federal government. All three experiments eliminated these groups, because UI payments for these claims were not made out of the state UI trust funds, therefore there was no potential saving to the state UI systems for these groups. A universal system would probably include these groups.

Recall status or referral union member

In an effort to reduce windfall gain, groups of individuals whose *ex ante* identification could be used for screening and whose length of unemployment was primarily determined by someone other than themselves were excluded from the Pennsylvania experiment. One such group was UI claimants on standby awaiting recall to a prior job. A second group was members of full referral unions (i.e., unions that operate hiring halls and take full responsibility for placing their members). Claimants on standby with specific recall dates within 60 days and claimants who were members of full referral unions were exempt from UI work-search requirements in all three states.

The three experiments had different rules regarding participation of UI claimants on standby or of members of full referral unions, but the effects were not that different. These claimants were not eligible to participate in the Pennsylvania experiment as stated above, did not generally participate in the Illinois experiment, and were fully eligible to participate in the Washington experiment. In both Pennsylvania and Washington, however, claimants who were in fact recalled to their previous job were ineligible to receive bonuses, even if they had been enrolled in the experiment. In Washington (but not in Pennsylvania), the same was true for claimants who had been placed on their first post-unemployment jobs through union hiring halls. In Illinois, there was no prohibition against paying bonuses to recalled workers or to workers placed on jobs through a union hiring hall. However, workers expecting recall with firm recall dates or expecting to be placed by the union were not required to register with the ES and generally did not do so. If they did not show up the ES office, they would not have been enrolled in the experiment. Thus, most of such UI claimants were de facto excluded from the Illinois experiment.

For the Washington experiment, the rationale ran as follows: even an individual on standby or a member of a full referral union can seek,

and obtain, other employment. Thus, the claimant's behavior can be influenced by a bonus offer, and he or she should have the opportunity to change behavior. However, there is no reason to pay a bonus to someone who is in fact recalled to his or her previous job or placed through the union hiring hall since the bonus offer could not have influenced that decision.⁴ If the claimant could demonstrate that the job with the previous employer was a "new" job, then bonus payment would not be withheld. It was not desirable for the bonus offer to discourage workers who had been laid off from one job in a company from taking an entirely different job in the same company.⁵

In the design stage, several union representatives objected to this provision but went along with it as an experiment, although they stated that they would oppose this provision if the bonus offer were to become a regular UI program. Table 2.2 summarizes the recall requirements in the bonus experiments.

Although these rules were somewhat different among the states, the end results were similar in that recalled workers did not receive bonuses in Pennsylvania and Washington, and most of such workers had not registered with the ES in Illinois and thus were ineligible for the bonus.

Backdated claims

Backdating refers to the process of starting a benefit year at a date prior to the Sunday of the filing week. A one-week backdating to cover the waiting week often occurs if there is good reason for the claimant

Table 2.2 UI and Bonus Eligibility for Workers Expecting Recall^a

Enrollment \ Bonus receipt	Recalled worker can receive bonus	Recalled worker cannot receive bonus
Worker on standby can enroll in experiment	Illinois (only if worker registers with ES)	Washington
Worker on standby cannot enroll in experiment		Pennsylvania Illinois (if worker does not register with ES)

^a Definitions of recall: Washington, return to same job with layoff employer; Pennsylvania, have recall date with layoff employer.

not filing in the first unemployment week. More extensive backdating occasionally occurs for a variety of reasons (e.g., claimant was ill). In the Pennsylvania experiment, a claim backdated more than two weeks made the claimant ineligible to participate in Pennsylvania. There were no such restrictions in Washington or Illinois.⁶

Register with the Job Service

Registration for job search was an explicit requirement for participation only in the Illinois experiment. In Illinois, enrollment took place in the ES office at the time the claimant registered for job search. Such registration was required of all claimants not explicitly exempted from job search.⁷

NATURE OF THE EXPERIMENTAL TREATMENTS

General Issues

A *treatment* is a particular program configuration that participants are offered or required to accept. A treatment may be a single well-defined program element (e.g., the offer of a \$500 reemployment bonus) or it may be a combination of several program elements. For example, a treatment may combine an offer of a reemployment bonus with job-search assistance. If this is the only treatment, then the experimental results will be valid only for that particular combination of program elements. To determine the contribution of any one program component, it is necessary to have more than one treatment that properly nests the program components: for example, one treatment with bonus offer only, a second treatment with job-search assistance and work search requirements only, and a third that combines the two. The third treatment is necessary only if you believe that all the components together will have a different effect than the sum of the effects of the separate components (i.e., a strong interactive effect among the program components). If this is not the case, then either treatment 2 or 3 could be dropped.

In most social program experiments, those eligible members of the population assigned to the experiments are free to accept or reject par-

ticipation in the program.⁸ For instance, UI claimants assigned to receive bonus offers were not required to either change job-search behavior or file for a bonus if they met the conditions for payment. As long as all individuals *assigned* to the treatment group are considered members for evaluative purposes, no harm is done to experimental integrity by having acceptance of the program voluntary.

Variation in the level or quantity of the program elements is necessary if an experimental objective is to evaluate more than one level of a program or, more aggressively, to determine an optimum level for a program. If a program has more than one level (e.g., different support levels for an income maintenance program or different bonus offers for a reemployment bonus program), then interpolation between two treatments in the experiment can be used to estimate the effects of programs levels that have not been tested but that lie between the two tested levels. This, of course, is “modeling” in the strict sense, but the necessary assumptions are sufficiently weak as to not cause undue concern. Of course, if three levels are tested, then the linearity assumption can be more strongly evaluated.

Elements of the Bonus Treatment

The three bonus offer experiments all had the same three basic components: 1) the bonus amount—the dollar value of the bonus offer, 2) the qualification period—the length of time from the date of enrollment into the experiment to the last date on which the claimant must start a qualifying job, and 3) the reemployment period—the length of time after starting a qualifying job that the claimant must remain fully employed in order to collect the bonus.⁹ The first two components varied among the three experiments, but the third did not. The reemployment period was set at four months in all experiments (16 weeks in Pennsylvania). The variations in bonus amount offered and qualification period among the three experiments are shown in Table 2.3.

The value of the bonus offer

Recognizing that acceptance of a bonus offer is intended to generate a change in job-seeking or job-accepting behavior, there is an implicit trade-off facing the claimant between the bonus and the UI benefits that will be foregone by taking a job sooner than would have

Table 2.3 Bonus and Qualification Period Structure of the Three Bonus Experiments

Treatment designation	Bonus offer (\$)	Qualification period (weeks)	Average bonus ^a (\$)	Average qualification period (weeks)
Illinois				
All bonus offers	500	11	500	11
Pennsylvania				
Short-low bonus offers	3 × WBA ^b	6	500	6
Short-high bonus offers	6 × WBA	6	1,003	6
Long-low bonus offers	3 × WBA	12	498	12
Long-high bonus offers	6 × WBA	12	989	12
Washington				
Short-low bonus offers	2 × WBA	(0.2 × UI duration) + 1	302	5.7
Short-medium bonus offers	4 × WBA	(0.2 × UI duration) + 1	610	5.8
Short-high bonus offers	6 × WBA	(0.2 × UI duration) + 1	917	5.7
Long-low bonus offers	2 × WBA	(0.4 × UI duration) + 1	303	11.0
Long-medium bonus offers	4 × WBA	(0.4 × UI duration) + 1	612	11.0
Long-high bonus offers	6 × WBA	(0.4 × UI duration) + 1	924	11.1

^a All dollar values are in nominal dollars. See Chapter 8, note 2, for information on use of the CPI to adjust the values for inflation.

^b WBA: weekly benefit amount.

been the case without the bonus offer. The value of a week's worth of UI benefits is represented by the weekly benefit amount (WBA). Thus, a fixed dollar offer is worth more to a claimant with a low WBA than to a claimant with a high WBA because each week's reduction in benefit payments is more costly to the latter. With this view of the bonus offer, the Pennsylvania and Washington experiments both priced the bonus offer in multiples of the WBA. For example, a bonus offer equal to 3 × WBA implies that all claimants are offered a bonus equal to three weeks of UI benefits.

A second issue facing the designers of the experiment involved the range of bonus amounts—what should be the smallest and largest WBA multiplier and how many different multipliers should be used? The objective was to have a sufficient range and number of bonus offers to enable evaluators to determine how response is affected by the size of the bonus offer, thus permitting policymakers to select the most cost-effective bonus offer. The determination of the range and number of options was partly determined by decisions on a likely policy range, what differences would be large enough to elicit significant differences in response, how many treatments would be necessary to properly map the response surface, and finally, cost.

Two bonus offers, one at the minimum and one at the maximum of the feasible range, would be sufficient to provide a basis for estimating the effects over a full range of options. However, interpolations are constrained to an assumption of linear differences in effects. Thus, if one treatment is at $2 \times$ WBA and another is at $6 \times$ WBA, then it must be assumed that a treatment at $4 \times$ WBA has an effect halfway between that of the two observed treatments. More than two bonus offer levels would permit validation of this linear effect hypothesis or some estimation of a nonlinear response surface. The three experiments differed in regard to ranges of offers. The \$500 bonus offer in the Illinois experiment provided unplanned variations in the bonus/WBA ratio that ranged from $3 \times$ WBA to $10 \times$ WBA.¹⁰ However, the natural differences in this ratio may correlate with other variables that affect behavior and therefore do not provide as a good a basis for decision making as planned variations. Pennsylvania had two and Washington had three planned levels of bonus/WBA offers.¹¹

Although the treatments in Washington and Pennsylvania were specified in terms of multiples of the claimant's WBA, the actual offer was for a fixed number of dollars. Thus, a claimant who had a WBA of \$175 and an assigned treatment of $6 \times$ WBA received a bonus offer of \$1,050. The claimant was not made aware of the algorithm used to calculate the bonus offer. The actual dollar offers varied over a fairly wide range, as follows:

Washington: the minimum bonus offer was \$110 ($2 \times$ the minimum WBA of \$55) and the maximum offer was for \$1,254 ($6 \times$ the

maximum WBA of \$209). At the mean WBA of \$153, a mid-range offer of $4 \times$ WBA would have a value of \$612.

Pennsylvania: the offers ranged from a minimum of \$105 ($3 \times$ the minimum WBA of \$35) to a maximum offer of \$1,596 ($6 \times$ the maximum WBA of \$266, increased from \$252 in 1989).

Qualification period

The second component of the bonus offer is the qualification period, i.e., the maximum duration of the initial spell of insured unemployment that the participant could experience and still qualify for the bonus. The length of the qualification period is important because it may influence the cost-effectiveness of the prospective program. A long qualification period will permit more eligible claimants to be affected by the offer, while at the same time it increases the opportunity for persons who do not alter their job seeking behavior to collect a bonus. These contradictory behavioral outcomes imply that the net effect of differing qualification periods would need to be determined empirically. As a result, both the Washington and Pennsylvania experiments varied the length of the qualification period.

As noted above, earlier acceptance of a job to earn a bonus imposes a cost to the participant equivalent to the amount of UI benefit payments foregone. The qualification period should be considered in this context. A 12-week qualification period is 46 percent of the entitled duration of 26 weeks in Pennsylvania or Illinois.¹² However, in Washington, where the entitled duration ranges from 10 to 30 weeks, a 12-week qualification period may be as low as 40 percent of entitled duration to more than 100 percent. If claimant job-search behavior is related to the length of entitlement, then a qualification period as a fixed proportion of benefit duration rather than a fixed number of weeks represents a more homogeneous treatment in a state with variable benefit durations.

Thus, the Washington experiment design had two qualification periods, one at 20 percent of the time that the individual claimant could draw full benefits (i.e., the compensable duration) plus one week to account for the waiting week, and a second set of treatments at 40 percent of compensable duration plus one week. If the algorithm resulted in a qualification period having a fraction of a week, the qualification

period was rounded to the next highest whole number of weeks. Since the duration of entitlement could run from 10 to 30 weeks, the short-qualification-period treatments ranged from a minimum of three weeks ($0.2 \times 10 + 1$ week) to a maximum of seven weeks ($0.2 \times 30 + 1$ week). The long-qualification period ranged from five weeks ($0.4 \times 10 + 1$ week) to a maximum of 13 weeks ($0.4 \times 30 + 1$ week).

In Pennsylvania, the 6-week qualification period was 22 percent of the 26-week entitlement faced by most claimants (i.e., 6/27ths of the period, including the waiting week), and the 12-week qualification period was 44 percent of the 26-week entitlement. These proportions are similar to those of the short and long qualifications in Washington. In Illinois, the qualification period was fixed at 11 weeks, which was 40 percent of the 26-week fixed entitled duration of benefits in Illinois (plus the waiting week).

The reemployment period

The reemployment period is the length of time that the participating claimant needed to remain fully employed after terminating receipt of benefits in order to qualify for a bonus. Four months was the length of time selected for this parameter of the system in all of the experiments (16 weeks in Pennsylvania). Response was not expected to be sufficiently sensitive to variations in this parameter to warrant varying it experimentally.¹³ The four-month interval was believed to be sufficiently long to avoid encouraging claimants to accept short-term employment simply to qualify for a bonus. It was also sufficiently long to avoid payment for employment on temporary seasonal jobs (e.g., harvesting in the summer, Christmas employment, or canning in the fishing season).

The reemployment period had to be served with the same employer in Illinois, whereas it was only necessary that the claimant remain fully employed in Pennsylvania and Washington, which was defined as taking not more than a week to change jobs and not filing for benefits during the transition period. Eighty to 85 percent of participants who obtained qualifying jobs met the reemployment period conditions. Surprisingly, there was little difference in this regard between Illinois and the other two experiments.

DETERMINING THE NUMBER AND ALLOCATION OF PARTICIPANTS

General Issues

Determining the number of claimants to enroll in the experiment and the number to enroll in each treatment and the control group are critical decisions. Nothing is more frustrating than spending several years and several million dollars on an experiment only to conclude that the number of participants was too small to enable the evaluators to draw statistically reliable conclusions about the results. In the discussion to follow, the term *sample size* is used to describe the number of claimants selected to participate in the experiment. Despite its importance, the subject of sample size is given little attention in books on experimental or evaluation methodology. In a 500-page book by Rossi and Freeman (1989), sampling is mentioned only once and does not merit inclusion in the index. Hausman and Wise (1985) discussed sampling as an issue of randomization and selection, not as an issue of size and cost. Cohen's (1988) important book deals with the sample size implications of statistical reliability, but not with cost and feasibility.

On the other hand, actual experiments must deal with the trade-off between cost and reliability in determining sample size. In the early income maintenance experiments, the question was posed as a mathematical programming problem in which the sample was allocated among income and treatment cells to get the maximum information about labor supply response from a given budget.¹⁴ In this approach, the question was never asked as to whether the response would be sufficient for policy purposes.

The correct approach is to provide answers to the following four sequential questions. 1) What effect must the proposed experimental program have in order to justify its implementation as a full program? 2) What sample size will be needed to generate a result such that, if the true response is the expected response, the result will be statistically significant? 3) How much will this experiment cost? 4) Is the answer sufficiently important to justify the cost? Thus, the experimental designers must be prepared to abort the process if it is determined that

the budget available for the experiment is insufficient to generate a usable response. Each of the four points, especially the first two, are described in more detail below.

Sample Design for the Bonus Experiments

The critical operational question for sample design was the number of claimants that would need to be enrolled into each treatment or control group at each site to meet specified design goals. The principal design goal was to generate experimental samples that would have the ability to detect specified changes in the parameters of interest, namely the length of insured unemployment and the amount of UI benefit payments.

In the first bonus experiment, the Illinois experiment, these questions were not asked explicitly because the experimenters had no guideline as to what changes to expect and no guidance from the state as to what changes were of policy interest. The approach in Illinois was to first establish a budget for payment of bonuses, i.e., \$750,000 for the two experiments based upon the Wagner-Payser funds available to the state of Illinois for these purposes.¹⁵ The next step was to determine how many claimants needed to be enrolled in order to pay 1,500 bonuses of \$500 each.

In Washington, there was no predetermined budget at the point the sample size was being determined. The sample requirements were based on the sample sizes needed in each treatment cell to detect impacts on weeks of insured unemployment as large as those found in the Illinois experiment (1.15 weeks). Based on Washington budget estimates, a budget was established by the U.S. Department of Labor for the Pennsylvania experiment. The sample design task in Pennsylvania was then to allocate the budget to achieve an optimum amount of information and to assure that the experiment could provide reasonable results. In the following discussion, we present the sample design procedures first for Illinois, then for Washington, and then the more elaborate modeling that resulted in the Pennsylvania design.

Sample design in Illinois

In Illinois, there were two treatments and a control group to which assignments were to be made.¹⁶ The decision was made to assign one-

third of the eligible population to each of the three groups. This was to be accomplished by using the last two digits of the claimant's social security number, a number which had been randomly generated, thereby assuring random assignment of the eligible population.

The sample design task was to determine how many bonus offers to make in order to make bonus payments of \$500 each to 1,500 participants in the two treatments (just sufficient to exhaust the \$750,000 bonus budget). The following estimates were made using the data provided by the Illinois Department of Employment Security (DES) on the number of ES registrants filing new valid claims in the last available quarter in the 19 UI offices that had been selected for the experiment: 1) the expected number of filers in Fall 1984 (the anticipated enrollment period), 2) the proportion that would be expected to accept the offer to enroll in the program, 3) the proportion that would be expected to obtain jobs within the qualification period and retain them for at least four months, and 4) an added proportion that might be expected to meet these qualifications as a result of the bonus offer.

The investigators arbitrarily assumed a refusal rate of 10 percent (based generally on experience). The estimated job acquisition rate of 0.317 was equal to the proportion of claimants in the observed period who obtained employment within 10 weeks of filing (the experimental design called for 11 weeks, but the data were provided on a 1–10 and 11–15 week basis). Of those obtaining employment, about 80 percent were estimated to have retained their jobs for at least four months, reducing to 1,212 the number of enrolled claimants expected to earn a bonus without any impact on behavior. We then assumed that two-thirds of those who actually terminated benefits in the 11–15 week period would respond to the bonus and reduce their unemployment spell below 11 weeks. Based on data for the period between July and September 1983, this would add 294 bonus recipients, bringing to 1,506 the expected number of bonus payments.

This rather ad hoc method of estimation was designed to provide a benchmark for enrollment. Continuous monitoring of the enrollment process was used to provide real-time estimates of the number of claimants terminating benefits within the qualification period and filing a notice in that regard. This monitoring process permitted the state to modify enrollment rates and determine the length of the enrollment period based on actual experience.

In the end, however, only 570 bonuses were paid, leaving half of the bonus budget on the table. A major reason for the large overestimate was the failure of the model to consider that many claimants would meet all the conditions for bonus eligibility but not apply (see discussion of the “take-up” issue in Chapter 3). This was particularly true in the employer experiment. While the model predicted that 29 percent of assigned eligible claimants would collect bonuses, the employer experiment paid bonuses to only 3 percent of the employers of eligible claimants. The claimant experiment also fell short of its goal, but by a considerably smaller amount. Fourteen percent of assigned eligible claimants collected bonuses, against a predicted 29 percent. (The shortfall in bonus payments was somewhat mitigated by assigning an additional 2,900 eligible claimants.)

Sample design in Washington

In Washington, sample sizes were selected to enable measurement of changes in key parameters vis-à-vis the control group. In determining the sample, attention was focused on the duration of insured unemployment, and sample sizes in each treatment were set to detect an experimental effect as large as that found in the Illinois experiment, i.e., 1.15 weeks.

Statistical texts such as Cohen (1988) provide the information necessary to determine sample size requirements utilizing four pieces of information: the number of treatment and control groups in the experiment, the estimated (or desired) experimental effect for each treatment group, an estimate of the standard deviation of the population, and the desired statistical properties of the results.

The required sample size for an experimental cell is dictated by two policy criteria: 1) the policymaker wants to have a high degree of confidence that it does not implement an ineffective policy and 2) the policymaker also wants confidence that it does not reject an effective policy. To meet these dual objectives, two statistical tests are administered: 1) a test of statistical significance, which means that, if a program is judged to be effective, the level of statistical significance gives the probability that this judgment is wrong, and 2) the power evaluation, which gives the probability that if a program is effective, our test will reveal it to be so.

A statistical significance level of 5 percent and a power of 80 percent is a typical combination of standards used by many analysts. If a program is judged to have an effect using these standards, there is a 5 percent probability that there is no effect and there is an 80 percent chance of detecting an effect when there is one. A decision must also be made as to whether a one-tailed or two-tailed significance level is to be used. The two-tailed test is generally preferred, unless it is almost certain that policy interest is in only one direction.¹⁷

In Washington, we imposed a more stringent 1 percent statistical significance test but used only the one-tailed criteria.¹⁸ The 0.01 significance test was used with a power of 0.8. A table from Cohen (1977, pp. 54–55), part of which is reproduced as Table 2.4, shows that each cell would need at least 2,000 observations to meet these conditions.

To apply the Cohen table, the appropriate effect size index must be determined. For the Washington experiment, the effect size index was determined to be 0.1, based on a desire to measure a 1.15-week impact on weeks of insured unemployment and the determination that the standard deviation of the distribution of weeks in the population was about 12 weeks. The effect size index is the estimated impact in standard deviation units (i.e., $1.15/12$).¹⁹ The data in the table show that the sample size is very sensitive to the power requirements.²⁰

The sample size requirement relates to that of a single cell. The next step in determining total sample size is to set the number of cells (treatment and control groups) and the degree of variation among treat-

**Table 2.4 Sample Sizes for Treatment Group,
Significance Level (one-tailed) = 0.01**

Power	Effect size index = 0.1 ^a	Effect size index = 0.2
0.5	1,083	272
0.6	1,332	334
0.7	1,627	408
0.8	2,009	503
0.9	2,605	652

SOURCE: Cohen (1977), Table 2.4.1, p. 55.

^a See text for discussion and calculation of effect size index.

ments. In determining sample size, the same considerations hold in designing experiments to distinguish between the impact of two treatments as in measuring the impact of a treatment relative to a control group.²¹ If the basic treatment is an offer of a \$500 bonus (compared with zero), much larger samples would be needed to detect the difference in impact of a \$600 bonus offer. In Washington, each treatment progressed in increments of $2 \times$ WBA. Since the Illinois treatment was equivalent to about $4 \times$ WBA, we should have considered larger sample sizes to detect differences in impact among the treatments.

For a comparison of treatments that represent different quantitative measures of the same variable, imposition of some modeling constraints can reduce the sample size requirement. In the Seattle/Denver Income Maintenance Experiment there were 11 treatments. However, much of the analysis was conducted using a labor supply model in which the differences across treatments in tax rates and support levels were reduced to a model with two variables, thereby substantially reducing the sample size requirements for estimating experimental effects (see Spiegelman, Robins, and West 1983). In the bonus experiments, the bonus values and qualification periods were redefined as continuous variables and estimated in a model with essentially two variables. The cost, of course, is the imposition of constraints on the relationship among the treatment variables.

To estimate the cost of a seven-cell experiment with an average sample size of 2,000 observations per cell, it was necessary to estimate the cost in bonus payments. This required prediction of the average value of the bonus payment and the take-up rate (i.e., the proportion of eligible claimants who collected bonuses). Since the experimental design called for bonus offers to be multiples of the individual WBA, it was no longer a fixed dollar amount as it had been in Illinois. The final sample design called for an unbalanced design underweighting the more expensive treatments: 20 percent of the sample would be controls; 15 percent in each of the four treatments with multipliers of $2 \times$ WBA or $4 \times$ WBA; and 10 percent of the sample in each of the two treatments with a bonus multiplier of $6 \times$ WBA. The average bonus offer was predicted to be \$565 using the 1988 average WBA in the state of Washington of \$148 and an average multiplier of 3.8.

The take-up rate was estimated to be 0.1875, leading to an expected bonus cost per eligible claimant of about \$106. The take-up

rate was loosely based on the Illinois experience of 0.14 and an expectation that improved information to claimants and more extensive follow-up procedures would increase take-up. This did not turn out to be the case.

As a result of these estimates, the total sample of 12,000 treatment-assigned claimants yielded a final budget estimate of about \$1,270,000 (see Spiegelman, O'Leary, and Kline 1992, p. 21). The U.S. Department of Labor allocated \$1.2 million for bonus payments. Despite some differences in actual parameters (the bonuses paid were larger and the take-up rate lower than projected), careful online monitoring led to actual bonus payments within 1 percent of those projected.

Sample design in Pennsylvania

In Pennsylvania, two major design goals were established: 1) the ability to detect a UI cost saving of \$150 (about one week of average UI benefits) from the high bonus offer and 2) the ability to detect a \$150 cost saving of moving from a 12-week to a 6-week qualification period. A third, subsidiary goal was to detect a \$150 cost saving from the removal of the workshop offer.²² Saving of a week's worth of benefit payments was about the average saving in Illinois and in the New Jersey experiment (although the Illinois saving resulted from a bonus offer closer to $4 \times$ WBA, rather than the $6 \times$ WBA in Pennsylvania). The sample sizes per treatment cell were those necessary to meet the measurement goals at a minimum of 80 percent power for a one-tailed test at the 5 percent significant level.

A formal sample allocation model was developed for Pennsylvania, based on the design objectives, on assumptions about the magnitude of the response to the bonus offer, on the cost of individual treatments, and on an overall budgetary constraint for the treatments.²³ The sample allocation obtained from the model provided for 3,000 control and 10,120 treatment group members allocated among the six treatments, with cell sizes ranging from 1,030 to 2,240. In the end, the high-bonus, long-qualification-period treatment, both with and without workshop, had 3,370 observations (see Corson et al. (1991), Table I.2, p. 12).

SELECTION OF EXPERIMENTAL SITES

General Issues

Determining where to conduct the experiment (site selection) is an often trivialized activity that in fact is as important as determining the number of individuals to enroll in the experiment. The number of sites has important experimental ramifications. As the number of sites increase, several things occur: 1) the effects on the results due to the idiosyncratic actions of specific site administrators are reduced; 2) the influence of external shocks that may occur during the experiment in specific sites (e.g., a major plant closing) is mitigated; 3) the experimental sample, and thus the results, become more representative of a larger population group (i.e., a state instead of a city); and 4) the administrative costs increase.²⁴

Although the increased administrative costs of a large number of sites is usually small relative to the total cost of an experiment, control over experimental operations—which is important—is weakened in a multisite experiment. The staff of the organizational unit managing experimental operations is usually small, making it difficult to properly monitor operations occurring simultaneously in many sites.

A case may also be made for a small number of sites if there are effects of scale in the operations. This is of little consequence in a bonus offer program but is important in workfare-type programs, in which a whole office is mobilized to carry out the experimental program. It is also important in training programs, where the breadth of the offering is important, and in counseling programs, where the existence of a dedicated staff is important.

Site selection may not arise in experiments that are conducted statewide. The U.S. Department of Labor had a requirement that experimental populations should be representative of the state in which the experiment was conducted. In the Washington experiment, this was accomplished by conducting the experiment in almost all of the state's Job Service Centers.

When the number of sites is a small proportion of the total available, site selection should be viewed as the first stage in a two-stage sampling process and the principles of randomization should apply to

both stages; that is, there should be random selection of sites and random selection of sample observations within selected sites. Metcalf and Kerachsky (1988) pointed out that failure to recognize that site differences contribute to the variance in the parameter estimates can lead to design errors.

Site Selection in the Bonus Experiments

With a total desired sample size determined, the next step is to determine where within each state to conduct the experiment. This decision is intimately tied to sample size considerations, as well as to determination of the length of the enrollment period. The total sample size, number of sites in which the experiment is conducted, rate of enrollment at each site, and length of the enrollment period are all tied together in a single equation. A decision with regard to any one of these variables affects the parameters of the others.

In Pennsylvania and Washington, UI offices for filing claims served as sites. Employment Service offices for registering claimants (and others) for job search were the sites in Illinois. In Washington, UI and ES offices were coterminous. In Pennsylvania, there are both coterminous and noncoterminous sites, but only the former were included in the experiment. Illinois also had both coterminous and noncoterminous sites, but both were included in the experiment. The decisions as to the number of sites, their location within the state, and the flow rate of enrollees were made on the basis of several complex considerations, including the following.

- 1) Reliance on local office staffs with no previous experience in administering a bonus offer program indicated that a large number of agency personnel should be involved in order to reduce the influence of any individual agents. On the other hand, the need to train agency personnel and monitor their performance put practical ceilings on the number of sites.
- 2) The sensitivity of the results to job-search success indicated the need to minimize the potential impact of specific labor market influences, such as adverse weather or a large plant closing.
- 3) The desire to be able to generalize the results dictated that the experimental sample be representative of the state or other large

populations (coupled with the U.S. Department of Labor requirements for Washington and Pennsylvania that the sample replicate the characteristics of the host state population).

- 4) Treatment consisted only of cash offers in Illinois and Washington, so there was no critical mass needed to conduct the experiment. Pennsylvania, with its job-search assistance component, did have the potential of enrollment rates being too low to maintain a program effort. Even in Illinois and Washington, however, there were minimums because of the need to train agency personnel to conduct enrollment interviews and the desire to keep a sufficient caseload to maintain staff interest and capability. There were no numbers attached to this requirement, but enrolling 16 percent of eligible new claimants in Washington was regarded as a minimum.
- 5) The previous four considerations all deal with establishing minimum sample sizes. There were also some issues that would tend to set maximum sample size. During the four-month enrollment period in Illinois, all eligible claimants were offered the opportunity to enroll. The operational burden on the counseling staff was not a consideration because it was a time that caseloads were falling and counselors were easily able to handle the additional workload. In Washington and Pennsylvania, however, there were practical limits to the enrollment rate, because hiring new personnel was not considered and the ability to add caseloads to the existing staff was limited. The desire to reduce the possibility of displacement of control group members also supported the decision not to enroll all eligible participants at a single site. In Washington, 16 percent of eligibles were enrolled at 20 of the 21 sites, and 32 percent enrolled at the remaining site. In Pennsylvania, 26 percent of eligible UI claimants were enrolled. These percentages are large enough to raise the specter of displacement, an issue discussed in Chapter 6.

Site selection in Illinois

In Illinois, 22 Employment Service offices (also referred to as Job Service Centers) were used as experimental sites. Enrollment into the experiment was carried out by Employment Service counselors.

Although job-search services are often provided by the Employment Service and UI offices under a single roof, only 10 of the 22 Employment Service offices in which the experiment was conducted had coterminous Employment Service and UI offices, and almost all of these were outside of Chicago.²⁵

Offices south of Springfield were eliminated to reduce administrative costs and to make the sample more representative of the industrial base; employment in the south of the state was more heavily in agriculture and mining. The central and northern areas were divided into four regions: Chicago (eight sites), Metro-outlying (four), Central (four), and Northwest (five) areas of the state. The largest offices in each area were selected with a view of acquiring the necessary caseload in four or fewer months and balancing the caseload among the four areas.

Site selection in Washington

Site selection decisions emanated easily from the basic decision to have the experimental sample replicate the characteristics of the state's population. The decision was simply to eliminate as few sites as possible and carry out the experiment in the rest of the Job Service Centers (all joint UI/ES offices.) In the end, 10 of the state's 31 Job Service Centers were eliminated: seven were too small and remote, two were part of other state experimental programs that could have contaminated the results, and one (Vancouver) was part of the Portland, Oregon, metropolitan area. The 21 offices included in the study accounted for approximately 85 percent of the state's UI claims.

The replication of state characteristics in the enrollment sites was increased by doubling the sampling rate at one site, Rainier. Because of the elimination of Pierce County (Tacoma and Lakewood Job Services), a large concentration of the state's black population was eliminated. To compensate, enrollment in Rainier, located in King County adjacent to Pierce County and also containing a large concentration of the black population, was doubled. As a result, the enrolled claimant population had the racial mix of the state as a whole (i.e., about 85 percent non-Hispanic white, 4 percent black, and 11 percent other).

Site selection in Pennsylvania

An early decision was made to conduct the Pennsylvania experiment in 12 sites, both for operational reasons and because analysis showed that 12 sites was a sufficient number to reduce site-specific effects on the variance to an acceptable size. Desiring to enroll participants in only 12 of the state's 87 UC offices and able to eliminate only 24 for cause (mostly for being too small or not having collocation of Unemployment Compensation and Job Service offices), the researchers adopted a process of stratified random sampling to select the 12 out of the 63 eligible sites. The principle was to assure that each eligible claimant in the 63 sites had an equal chance of being selected into the experiment. With only 12 sites to be selected, the researchers wanted to guard against an accidentally skewed selection of sites with regard to administrative region of the state or duration of UI benefits. The latter stood as proxy for employment conditions in the local labor market. Seven regions and four duration categories provided the potential for stratification into 28 cells, which exceeded the number of sites to be selected. As described by Metcalf and Kerachsky (1988), the 28 cells were condensed into 12 clusters with approximately equal size UI caseloads. One site was randomly selected from each cell.

CHARACTERISTICS OF THE ENVIRONMENT

Characteristics of the States of Illinois, Pennsylvania, and Washington

Table 2.5 provides information on the three states in which bonus experiments were conducted, permitting comparison among states and over the two years for each state that were most relevant for the operational phases of the experiments. What do we learn from this table that is of importance to understanding the working of the bonus experiments?

Illinois and Pennsylvania were close in size and population density and experienced little population change over the relevant period. Washington was considerably smaller and less dense, but it experienced some modest population growth in the period. Illinois had the

Table 2.5 Characteristics of the Illinois, Pennsylvania, and Washington States in Selected Years

Characteristic	Illinois		Pennsylvania		Washington	
	1984	1985	1989	1990	1988	1989
Population (000)	11,511	11,535	12,040	11,882	4,648	4,761
Pop/sq. mile	207	207	268	265	70	72
Racial groups (%) ^a						
Black	14.7		9.2		3.1	
Hispanic	5.6		2.0		4.4	
Asian	0.2		1.2		1.3	
Nat. American	0.1		0.1		1.7	
Labor force (000)	5,604	5,673	5,857	5,901	2,295	2,451
Employment (000)	5,093	5,160	5,592	5,583	2,153	2,300
Insured unemployment rate (%)	2.5	2.7	2.4	3.1	3.0	2.9
Total unemployment rate (%)	8.4	8.9	4.8	5.8	5.8	5.5
Labor force participation (%)						
Male	76.8	78.2	73.4	73.2	74.1	76.9
Female	54.3	54.4	52.8	53.5	59.5	61.4
UI claims (000)	835	836	1044	1184	465	457
Exhaust. rate (%)	39.4	38.7	23.3	25.1	27.9	25.7
Avg. weekly wage (\$)	364	377	419	441	393	403
Avg. compensation duration (weeks)	18.6	17.2	14.5	14.8	14.8	15.3
Avg. WBA (\$)	134	136	182	189	151	156
WBA/weekly wage	0.368	0.361	0.434	0.429	0.384	0.387

SOURCE: U.S. Dept. of Labor, Bureau of Labor Statistics (1984, 1985, 1988, 1989, 1990); U.S. Dept. of Commerce (1986, 1987, 1989, 1990, 1991, 1992); U.S. Dept. of Labor, Employment and Training Administration (1984, 1985, 1988, 1989, 1990).

^a Available only for census years: Illinois 1980; Pennsylvania 1990; Washington 1990.

largest non-Caucasian population proportions and Washington the smallest. The proportion of blacks in Washington was considerably below the national average.

Labor force participation was similar, although male labor force participation was somewhat higher in Illinois than in the other two states and female labor force participation was somewhat higher in Washington than in the other two. Employment grew in both Illinois and Washington in the relevant periods but declined slightly in Pennsylvania. The total unemployment rate (TUR) was much higher in Illinois in the period, which reflects the conditions at the time. However, the insured unemployment rate (IUR) did not differ much across the states, despite the differences in TUR.

Relative to the sizes of the labor forces in the three states, it would appear that initial claims were lower in Illinois and were fairly stable across the two-year period. The much higher exhaustion rate in Illinois is consistent with the relatively low ratio of IUR to TUR in the state, as one explanation for the low ratio is a large number of exhaustees who remain unemployed. However, the low initial claims in Illinois may reflect a lower rate of layoffs in the early phases of a recovery that had not yet translated into significant job growth. The longer duration of compensation in Illinois was consistent with greater exhaustion of benefits and with the lingering effects of the recession. (The durations are for regular benefits and, therefore, do not include the effects of the extended benefit program.)

The lower weekly wage rate in Illinois is consistent with the difference in timing of the experiments, fully reflecting the 3 percent rate of inflation that characterized the period. Overall, the Pennsylvania UI program appears to be the more generous as displayed by the higher WBA/weekly wage ratio, known as the “replacement rate.”

What does this information tell us about possible differences in impact of the bonus offer among states? The higher exhaustion rates and longer durations of insured unemployment in Illinois imply either 1) worse job prospects for covered workers in this state, at that time, than in the other two states or 2) a much greater tendency to use the UI system to voluntarily extend unemployment. The much higher TUR suggests the former. Employment growth was very modest, and the increasing labor force participation led to increases in unemployment rates—not suggestive of a situation in which increased job search

would be productive. However, to the extent it is successful, there were potentially greater savings in Illinois, as the bonus responder would be, on average, making a greater reduction in the length of the unemployment spell than his peer in Pennsylvania or Washington. Thus, the economic data from the states are ambiguous as a basis for explaining the larger impact of the bonus offer in Illinois.

The higher wage replacement rate in Pennsylvania implies that a bonus offer representing the same ratio of bonus offer to WBA in the three states is more generous in terms of earning equivalence in Pennsylvania, which leads to an expectation of greater response in Pennsylvania.

Comparison of Control Groups across the Three Experiments

The control group represents the population from which the experimental sample is drawn. Differences and similarities among the control groups of the three experiments can help explain differences in the effects of the experimental treatments. Table 2.6 displays characteristics (including basic demographic and economic characteristics) of the control group members. There are some potentially important similarities and differences worth mentioning. The Illinois sample was slightly more female than the Washington or Pennsylvania samples, which both show a 60/40 male/female split—not unlike the working population. The age distributions are remarkably similar in Pennsylvania and Washington but differ by design in Illinois, which set 55 as the maximum age for enrollment. Nevertheless, the Illinois sample is even younger than indicated by the design, because it differs from the other two in the proportion of those under age 35.

In terms of racial differences, Illinois had a much lower proportion of white non-Hispanics and a much higher proportion of blacks than either Pennsylvania or Washington. The proportion of blacks is particularly low in Washington, consistent with its population mix. The basic manufacturing/nonmanufacturing industrial mixes are essentially the same in the three experiments.

The UI characteristics differ somewhat among the states. The lower WBA and base period earnings in Illinois probably reflects timing, since this experiment preceded the Pennsylvania and Washington experiments by about four years. An annual increase at a compound

**Table 2.6 Characteristics of Control Group Members
in the Three Experiments**

Characteristic	Illinois	Pennsylvania	Washington
Total members	3,952	3,354	3,082
Gender (%)			
Male	45.3	40.5	39.5
Female	54.7	59.5	60.5
Age (%)			
Less than 35	62.2	53.5	52.2
35 to 54	37.8	36.7	39.8
55 and above	0.0	9.7	8.0
Race (%)			
White, non-Hispanic	63.2	83.8	83.3
Black	27.1	12.1	4.3
Hispanic	7.6	3.5	7.0
Other	2.1	0.6	5.4
Industry (%)			
Manufacturing	26.4	25.8	23.1
Nonmanufacturing	73.6	74.2	76.9
Occupation (%)			
White-collar	42.1		34.2
Other occupations	57.9		65.8
Weekly benefit amount (\$)	119.93	164.08	150.51
Entitled duration (weeks)	26.0	25.9	26.9
Base period earnings (\$)	12,753	14,126	15,475
UI Benefits			
Weeks of insured unemployment	20.1	14.9	14.3
Benefits drawn (\$)	2,487.00	2,387.00	2,066.00
Exhaustion rate (%)	47.2	27.7	23.9
Initial UI spell (weeks)	18.3	12.5	11.4

rate of 5.8 percent would bring the Illinois WBA up to the Washington level at the time the Washington experiment was launched. This rate of increase is not inconsistent with the wage growth rates in this period.

Pennsylvania had a somewhat more generous UI program than Washington, as indicated by the higher WBA, despite somewhat lower base period earnings. The average weeks of entitled duration are roughly 26 in all states. In Illinois and Pennsylvania, 26 weeks is the standard (although a few UI recipients get only 16 weeks, making the average for Pennsylvania 25.9 weeks), whereas in Washington it represents the average over a potential range of 10 to 30 weeks.

The poorer economic climate in Illinois at the time is indicated by the considerably longer average weeks of UI benefits, the longer initial spell of weeks on UI, and the much higher exhaustion rate than in either Pennsylvania or Washington. These differences do not carry over to much higher dollars of benefits drawn, however, because of the lower WBA.

EFFECTIVENESS OF THE RANDOMIZATION PROCESS

The objective of random assignment is to create groups that are homogeneous in terms of the characteristics that affect outcomes. If this is accomplished, a simple comparison of mean outcomes among treatment and control groups provides an unbiased estimate of the treatment effects. The characteristics of claimants in the different groups, within each experiment, are compared to test the effectiveness of the assignment process in creating homogeneous groups. Homogeneity is achieved if differences in the distribution of characteristics among the treatment and control groups are small and generally statistically insignificant. Chance alone will create some statistically significant differences. At a 90 percent confidence level, chance alone would be expected to produce a statistically significant difference in one out of ten characteristic comparisons. Tables 2.7, 2.8, and 2.9 show the distribution of characteristics generally found to affect the outcomes of interest—weeks of unemployment or UI benefit payments—in each of the experiments.

For Pennsylvania, the mean values of characteristics across the six treatment groups and the control group are shown in Table 2.7. Sixty-six treatment/control differences were calculated. Seven of the comparisons are statistically different at the 90 percent confidence level, which is what you would expect by chance alone. The three differences at the 95 percent confidence level are what you would expect. In all, the Pennsylvania results are not inconsistent with random assignment.²⁶

For Washington, 14 characteristic means are compared across 6 treatment groups and the control group, providing a total of 84 treatment/control group comparisons (Table 2.8). Chance alone could account for as many as four statistically significant differences at the 95 percent confidence level and eight at the 90 percent confidence level. These are essentially the number of statistically significant differences shown in the table.

For Illinois, the means of 11 characteristics are compared in Table 2.9 for the single treatment and the control groups. These results are somewhat less supportive of the conclusion as to the effectiveness of random assignment, since 4 of the 11 means differed statistically at the 95 percent confidence level, considerably more than the 0 or 1 statistically significant differences that chance alone should account for.

However, validating the random process does not necessarily mean that the samples are sufficiently homogenous to warrant use of unadjusted mean comparisons. It is shown in Chapter 4 that treatment/control differences in UI compensation are affected by the inclusion of WBA (and base earnings) in the Pennsylvania and Washington estimating equations, requiring use of regression-adjusted treatment/control comparisons to provide unbiased estimates of treatment.

Table 2.7 Tests of Randomization in the Pennsylvania Experiment, Control and Six Treatment Group Means^a

Variable	Control	Treatment group					
		Low/ short ^b	Low/ long	High/ short	High/ long	Declin./ long	High/ long
Female (%)	40.5	40.9 (0.22)	39.3 (0.94)	40.2 (0.19)	40.0 (0.37)	40.7 (0.16)	39.8 (0.46)
Age less than 35 (%)	53.5	53.2 (0.23)	54.7 (0.93)	56.4** (2.02)	56.4** (1.99)	53.2 (0.23)	52.7 (0.50)
Age 55 and above (%)	9.7	9.8 (0.05)	9.1 (0.81)	9.8 (0.02)	9.8 (0.11)	10.3 (0.63)	8.6 (1.12)
Black (%)	12.1	10.3* (1.83)	11.5 (0.69)	12.3 (0.16)	12.4 (0.25)	11.5 (0.71)	10.7 (1.31)
Hispanic (%)	3.5	3.9 (0.66)	3.8 (0.45)	3.9 (0.68)	2.7 (1.50)	3.8 (0.55)	3.9 (0.53)
Other non- Whites (%)	0.6	0.5 (0.42)	0.4 (0.58)	0.5 (0.18)	1.0** (2.10)	0.4 (0.67)	0.7 (0.36)
Manufacturing (%)	25.8	25.5 (0.20)	26.0 (0.25)	25.8 (0.00)	25.7 (0.01)	26.1 (0.24)	25.5 (0.15)
Weekly benefit (\$)	164.1	165.5 (0.65)	166.4 (1.27)	167.1 (1.52)	165.1 (0.50)	167.8* (1.82)	166.9 (1.23)
Base earnings (\$)	14,126	14,375 (0.72)	14,650* (1.83)	14,352 (0.73)	14,317 (0.60)	14,695* (1.83)	14,301 (0.49)
Maximum WBA (%)	18.6	19.3 (0.49)	20.7** (1.99)	18.3 (0.27)	20.0 (1.16)	20.5 (1.63)	20.7 (1.57)
Expected recall (%)	10.8	10.1 (0.78)	10.4 (0.47)	10.9 (0.05)	11.4 (0.61)	10.1 (0.85)	11.7 (0.82)

SOURCE: Corson et al. (1991), Table III.7, p. 46; used with permission of Mathematica Policy Research, Inc.

^a The *t*-statistic of the difference from the control group mean is in parentheses.

* = Significantly different from the control group mean at the 90 percent level of confidence; ** = significantly different from the control group mean at the 95 percent level of confidence.

^b Treatment groups are described by the bonus amount (e.g., "Low") and the qualification period (e.g., "Short").

Table 2.8 Tests of Randomization in the Washington Experiment, Control and Six Treatment Group Means^a

Variable	Control	Treatment group					
		Low/ short	Middle/ short	High/ short	Low/ long	Middle/ long	High/ long
Female (%)	39.5	38.9 (0.42)	39.3 (0.17)	38.7 (0.55)	38.5 (0.74)	39.8 (0.22)	38.0 (0.99)
Age less than 35 (%)	52.2	53.3 (0.84)	52.8 (0.43)	51.9 (0.20)	52.2 (0.02)	52.4 (0.14)	53.8 (1.01)
Age 55 and above (%)	8.0	7.7 (0.48)	7.8 (0.36)	8.0 (0.01)	8.2 (0.21)	8.4 (0.49)	6.8 (1.47)
Black (%)	4.3	5.2 (1.48)	4.7 (0.65)	4.0 (0.43)	4.7 (0.67)	4.1 (0.42)	4.2 (0.23)
Hispanic (%)	7.0	6.8 (0.40)	6.4 (0.96)	6.8 (0.28)	7.0 (0.00)	6.8 (0.42)	5.3** (2.18)
Other non-Whites (%)	5.4	4.9 (0.76)	4.6 (1.27)	4.7 (0.92)	4.9 (0.84)	5.1 (0.36)	4.4 (1.47)
Manufacturing (%)	23.1	22.0 (0.92)	22.8 (0.28)	21.9 (0.91)	22.8 (0.24)	22.3 (0.69)	22.5 (0.43)
Weekly benefit (\$)	150.5	150.5 (0.02)	152.1 (1.12)	152.8 (1.40)	153.5** (2.11)	152.8 (1.58)	154.0** (2.12)
Entitlement (weeks)	26.9	26.7 (1.22)	27.0 (0.91)	26.8 (0.75)	26.9 (0.62)	26.8 (0.06)	27.0 (1.16)
Base earnings (\$)	15,475	15,486 (0.03)	15,860 (1.22)	15,537 (0.17)	15,872 (1.27)	16,073* (1.90)	16,148* (1.88)
Maximum WBA (%)	33.0	32.7 (0.18)	33.5 (0.43)	33.5 (0.40)	34.8 (1.40)	34.0 (0.83)	36.0** (2.03)
Search exempt (%)	22.5	21.7 (0.66)	21.8 (0.64)	20.9 (1.23)	22.3 (0.18)	20.7 (1.58)	22.7 (0.14)
White collar (%)	34.2	33.3 (0.70)	35.6 (1.05)	34.8 (0.39)	35.3 (0.83)	36.7* (1.87)	35.6 (0.94)
Years of education	12.3	12.3 (0.34)	12.4* (1.80)	12.3 (0.43)	12.3 (0.34)	12.4 (1.00)	12.4 (1.61)

SOURCE: Spiegelman, O'Leary, and Kline (1992), Table 5-1, p. 85.

^a The *t*-statistic of difference from the control group mean is in parentheses. * = Statistically significant at the 90 percent confidence level for a two-tailed test; ** = statistically significant at the 95 percent confidence level for a two-tailed test.

Table 2.9 Tests of Randomization in the Illinois Experiment, Control and Treatment Group Means^a

Variable	Control mean	Treatment mean
Female (%)	45.3	43.7 (1.45)
Age less than 35 (%)	62.2	61.8 (0.32)
Black (%)	27.1	25.1** (2.10)
Hispanic (%)	7.6	7.4 (0.27)
Other non-Whites (%)	2.1	2.5 (1.01)
Manufacturing (%)	26.4	24.8 (1.62)
Weekly benefit (\$)	119.9	118.8 (1.27)
Base earnings (\$)	12,753	12,888 (0.65)
Maximum WBA(%)	35.3	33.1** (2.15)
White collar (%)	42.1	45.2** (2.79)
Years of education	11.5	11.8** (6.03)

SOURCE: Spiegelman and Woodbury (1987).

^a The *t*-statistic of difference from the control group mean is in parentheses. ** = Significantly different from the control group mean at the 95 percent level of confidence for a two-tailed test.

OPERATIONS

In this section, the operational design and procedures of the three experiments are described and compared. All experiments start from the premise that participation in the experiment is limited to individuals who file claims for UI benefits. Thus, before embarking on a journey through the bonus experiment, we will digress to describe in some detail the process of filing a UI claim.

The UI Filing Process

To start the UI filing process, a prospective UI claimant enters a UI office and files a claim to establish a benefit year. A benefit year is the 12-month period in which the claimant can draw benefits before having to reestablish entitlement. Filing usually occurs the week immediately after the claimant becomes unemployed.²⁷

In all three states, the claimant returns to the UI office to establish the waiting week and to claim the first week of compensation two weeks after filing to establish the benefit year. In all three states, as in most states, an eligible claimant must serve a waiting week before benefits can be paid for subsequent weeks of unemployment.

To receive UI benefits, the claimant must be unemployed at the start of the benefit year and must be both monetarily and nonmonetarily eligible for benefits. The two terms have technical definitions. *Monetarily eligible* means that there are sufficient wage credits in the base year (usually the first four of the five quarters prior to filing for benefits) to establish monetary entitlement. *Nonmonetarily eligible* means that separation from employment did not occur under conditions that made the claimant ineligible to receive benefits and that the claimant is “able and available” to work.

In all three states, a claimant is required to register with the ES some time in the two-week period between filing the initial claim and claiming the first week of benefits. Registration with the ES is tied to the requirement that a UI claimant must be actively seeking work in order to be eligible for UI benefits. The registration requirement is not enforced in Pennsylvania. In Washington, all UI and ES offices were joint, so that ES registration occurred at the same time as filing a claim for benefits. In Illinois, some offices were joint and other were not.

Claimants there were required to register with the ES before returning to the UI office to claim the first week of benefits (plus the waiting week credit).

Assignment and Enrollment

Assignment into the experiment took place in the UI office in the Pennsylvania and Washington experiments, but occurred in the Employment Service office in Illinois. In all three states, however, eligibility was limited to those claimants filing initial claims to establish a benefit year.

We will start with Washington, because the enrollment process there was simplest and most closely linked to the filing for UI benefits. For all UI applicants, an on-line computer file was accessed which provided the status of monetary eligibility and printed an initial statement of entitlement. Software was written exclusively for the experiment. It used the agency data file to randomly select some of the UI-eligible claimants according to last two digits of their social security number and assigned them to treatment or control status. A computer-generated form that was provided to every worker filing a claim would also contain information on the bonus offer if the claimant was assigned to an experimental treatment. A treatment-assigned claimant was tentatively enrolled into the experiment; the enrollment was confirmed by letter when both monetary and nonmonetary UI entitlement were determined. This process was also followed in Illinois.

The process in Pennsylvania differed somewhat, especially with regard to the timing of assignment and enrollment. Upon UI application, local office staff entered information from the UI application form into the state computer. A computer file was then created on the state system to extract information on all claimants in the demonstration offices who were determined to be monetarily eligible for UI benefits during the past week. The file was transferred to the special demonstration data system (the Participant Tracking System), and experiment eligibility screens were applied (described below). Once the screening was completed, a weekly maximum number of individuals was selected randomly from each office from among the pool of eligible claimants. The selected claimants were then assigned randomly to

treatment and control groups (by the last two digits of their social security number).

Enrollment in Pennsylvania differed from that in Washington because it occurred at the time the claimant filed a claim to establish waiting week credit and collect benefits for the first week of compensated unemployment. Thus, enrollment into the experiment occurred two weeks after a claimant filed for benefits and excluded claimants who did not return to claim a waiting week (and a compensable week if also earned).

Illinois differed from the other two in that assignment and enrollment into the experiment occurred in the ES office at the time the claimant, who had previously filed for benefits, registered for job search assistance. Thus, only UI claimants who registered with the ES were eligible to participate, thereby de facto excluding most claimants exempt from job search. At the end of the registration process, the ES counselor would inform the selected claimant that he or she was eligible to participate in an experimental program. The claimant would then be asked to wait for a specially trained counselor who would hand the claimant a printed, one-page instruction sheet and explain the experiment. Unique to Illinois, the claimant would be asked to sign an agreement to participate, but the agreement did not really commit the claimant to any particular actions. This requirement was dropped in Pennsylvania and Washington.

The enrollment process in all three states was one of attempting to communicate the nature of the bonus offer and the necessary requirements to receive the bonus. Either in individual counseling sessions or in group enrollment sessions, the assigned claimant was given a single-page information sheet and a verbal description of the experiment, including an explanation of the participant's actions that would lead to a bonus. In Washington, the UI interviewer would end by asking a series of four questions designed to ensure that the claimant understood the program. The low rate at which apparently eligible participants collected bonuses in Illinois led to increased efforts in Pennsylvania and Washington to improve claimant's understanding of the program. (As described in Chapter 3, there was some improvement in the take-up rate in Washington over that in Illinois, but not in Pennsylvania.)

Differences in timing and location of enrollment among the experiments were not merely administrative details; they affected the charac-

teristics of the sample in each experiment. In Washington, where enrollment took place in the UI office at the time the claimant filed for UI benefits, every claimant eligible for UI benefits was a potential participant. In Pennsylvania, where enrollment took place at the time the claimant submitted his or her first claim for waiting week credit or compensation, claimants who obtained jobs within the first two weeks after filing the initial claim were unlikely to file for the waiting week or the first compensable week and were thus ineligible for a bonus (Pennsylvania data indicated that 6 percent of initial filers didn't return to file a continued claim). The same situation would arise in Illinois for those claimants who did not register for job search assistance at the same time as filing an initial claim—this would differ among ES centers, depending upon whether or not the UI and ES offices were unified.

An external validity problem arises in the Pennsylvania design, because in a real program—unlike the experiment—knowledge of the bonus offer would be universal and claimants could postpone taking jobs until they had served the waiting week and established entitlement to the bonus. The Illinois design also has an external validity issue in that knowledge of the bonus offer could attract some of those exempt from job search to register with the ES. In general, all three experiments have an external validity problem of unknown dimension in that nonfilers who do not expect to be unemployed very long might file anyway in a real program to get the bonus.

Determination of Full Eligibility to Participate

To be fully eligible to participate, the monetarily eligible claimant must not have had a separation issue on the claim that would prevent payment of UI benefits. In Illinois and Washington, but not in Pennsylvania, central office personnel reviewed the claims files of claimants assigned to experimental treatment groups. If there were “issues” on the claim regarding separation from the previous employer, then eligibility for a bonus would be held up until the issue was resolved, and it was denied if the resolution was against the claimant or was not resolved prior to the end of the reemployment period. When the review showed the claimant to be eligible to receive UI benefits, the agent sent an enrollment letter to the assigned claimant, informing that person that he or she was eligible to participate in the experiment.²⁸ A Notice

of Hire form accompanied the letter, with instructions regarding the procedures for submission. In Pennsylvania, the review for nonmonetary eligibility did not occur unless the bonus was claimed.

Submitting the Notice of Hire

The next step in the process occurred when the participating claimant obtained a full-time job within the qualification period. In all three experiments, the claimant informed the central UI office of a job acquisition that would establish eligibility for a bonus by submitting a Notice of Hire (NOH) form. In Illinois, the NOH was actually sent by the employer, thereby providing instant verification of employment. The employer verified that the claimant was hired in a job for 30 hours a week or more, starting on a particular date. The UI claim file was then accessed to verify that the hiring date was within the qualification period and that the claimant stopped receiving UI benefits. In Washington and Pennsylvania, the NOH was submitted directly by the claimant. The interposition of the employer was eliminated to reduce the possible stigmatizing effect of requiring the participating claimant to enlist the support of the employer in his or her quest for a bonus.

In Pennsylvania and Washington, the NOH provided information about the claimant's prior and new jobs, permitting a comparison to assure that the new job was not a recall to a previous job—a non-issue in Illinois. If the employer, occupation title, and wage rates of the prior and new jobs matched, there was a presumption of a recall and the employer would be contacted. Affirmative answers to the query as to whether the claimant was self-employed or placed on the job by his or her union would generate letters requesting clarification. A self-employed individual could receive a bonus but had to provide proof that he or she had a business license and some business income. The NOH process in Washington and Pennsylvania was more complicated than in Illinois because claimants could change jobs and still qualify for a bonus—with the condition that the claimant did not claim benefits and did not take more than a week between jobs. A claimant changing jobs in Washington and Pennsylvania would file an NOH for each job.

Submission of the Bonus Voucher

Sixteen or 17 weeks after reemployment, if the claimant had remained fully employed for the entire period (or changed jobs in Pennsylvania and Washington, with no more than a week interlude and no filing for benefits), the claimant submitted a voucher for the bonus payment. In Illinois, a valid voucher was certified by the employer before submission. In Washington and Pennsylvania, the voucher was submitted without such certification, but office personnel contacted employers to verify continual employment for the reemployment period and checked the UI claim file to ascertain that UI benefits had not been drawn during the reemployment period.

For payment of the bonus to be authorized, verification had to show the following: the claimant had not drawn benefits during the four-month qualification period; he or she had been employed on a full-time job (30 hours in Illinois and Pennsylvania, 34 hours per week in Washington); the first post-unemployment job was not a recall to the job held prior to filing for benefits or a placement through a union hiring hall; and that any self-employment was documented. If any of these issues were not favorable, the claimant would be contacted to resolve the discrepancies. When the agent was assured that the conditions for bonus payment had been met, a check would be issued for the amount of the bonus.

Participation in the Three Experiments

Table 2.10 shows the numbers of claimants offered enrollment into the bonus offer experiments and the proportions who participated at various levels (by submitting NOHs and vouchers for payment of bonuses). The proportion of eligibles submitting valid NOHs was highest in Illinois and lowest in Pennsylvania, as was the case with payment of bonuses. In all three experiments, only a small proportion of eligible enrollees fully participated by receiving a bonus.

The Illinois results pertain only to the claimant experiment. If the employer experiment had been included, the number of assignees would be double the number in the table. The drop-off in number of enrollees was much larger in Illinois than in the other experiments and was due to refusal to participate. Since the assignees included only

Table 2.10 Participation in the Three Experiments^a

Participation	Illinois		Pennsylvania		Washington	
	Number	%	Number	%	Number	%
Assigned	4,186		11,410		14,080	
Enrolled	3,527		10,694		14,080	
Eligible	3,527	100	10,694	100	12,140	100
Submitted NOH	NA ^b		1,604	15.0	2,533	20.9
Valid NOH	765	21.7	1,315	12.3	2,255	18.6
Submitted bonus voucher	NA		1,155	10.8	1,946	16.0
Voucher paid	570	16.2	1,123	10.5	1,816	15.0

SOURCE: Spiegelman and Woodbury (1987), Table 5-1; Corson et al. (1991), Table III.3, p. 40 and Table VI.1, p. 77; Spiegelman, O’Leary, and Kline (1983), Table 3-1, p. 50 and Table 3-2, p. 52.

^a In Illinois, the number assigned includes only those eligible to participate and the number enrolled includes those who agree to participate. In Pennsylvania, some of the eligible assignees did not return to claim a waiting week and therefore were not enrolled. All those enrolled were eligible. In Washington, assignment and enrollment occurred simultaneously at the time of filing for benefits for all those monetarily eligible. Some of those “tentatively” enrolled were found to be ineligible due to nonmonetary issues.

^b NA = not available.

those eligible to participate, there is no difference in the number of enrollees and number of eligibles. In Pennsylvania, the decline from assigned to enrolled represents the nonmonetary ineligibility of some participants. In Washington, all who were monetarily eligible for UI were enrolled, but the ineligible were then eliminated from the sample. Thus, much of the difference shown prior to “eligibility” represents differences in timing of dropping ineligible from the sample and is not substantive.

Despite differences in procedure (discussed above) regarding NOH submissions that should have disadvantaged Illinois, the proportion of valid NOFs submitted was greatest in Illinois. The drop in the number of participants submitting valid NOHs to the number submitting bonus vouchers could be accounted for by failure to remain fully employed for the reemployment period (there might also be some take-up failure,

as discussed in Chapter 3). In the end, the procedures adopted to increase participation in Pennsylvania and Washington were for naught, as Illinois had the highest proportion of eligible enrollees collecting bonuses.

Monitoring for Operational Performance

In all three experiments, the enrollment rates at the various offices were monitored to assure compliance with the experimental design. Monitoring occurred at two levels. In Pennsylvania and Washington, there was an on-line Oracle system developed to manage the paper flow (e.g., tracking claimants and sending out letters and forms) and to track enrollment rates in the offices. Staffs of the state agencies and of the evaluation contractors visited each of the offices, armed with information regarding performance. Real operational problems were uncovered in only one office in Washington—the claimstaker assigned to enroll bonus-eligible claimants into the experiment had been transferred and no one had yet taken her place. This lapse was immediately corrected.

A concern was the possible dwindling of interest and parallel reduction in information flow to the assigned claimant. Monitoring of sites did not uncover this tendency. Overall, it was concluded that the experiment operated according to design and that any lapses were very minor.

THE FEDERAL SUPPLEMENTAL COMPENSATION (FSC) PROGRAM IN ILLINOIS: AN UNPLANNED EXPERIMENT

As will be seen in Chapter 4, the impacts of the Illinois experiment were about double those experienced in similar size treatments in the other two experiments. A possible explanation for the large impacts of the bonus offer in Illinois was the presence in Illinois of a temporary extended benefit program that did not exist in the other states. FSC was a federally funded, national program that existed between 1982 and 1985. Its discontinuation in Illinois in the middle of the experimental enrollment period had not been anticipated by the research

designers and resulted in a “natural experiment.” The FSC program nationally extended benefits for most recipients by 10 to 16 weeks, depending upon the state’s unemployment rate. For Illinois, the extension for those eligible was 12 weeks. Most of those claimants enrolled in the experiment in July and August 1984 had available the 12 weeks of extended UI benefits in addition to their 26 weeks of state regular benefits. After that date, all claimants had only the 26 weeks of regular benefits available.

Because the criteria for receiving extended benefits was somewhat more stringent than that for regular benefits, about 6 percent of the claimants eligible for regular benefits were not eligible for extended benefits. In Chapter 4, we compare the effects of the bonus offer on two groups: the first group is comprised of those claimants enrolled in the experiment (or controls) who were eligible for FSC (“FSC-eligibles”), and the second group is comprised of those experimentals and controls enrolled in the second period who, though not eligible for FSC, would have been eligible if it had been available (“FSC-ineligibles”).

A natural experiment exists because each of these groups has its own control group and differs largely because of the FSC availability. Of course, there are also some other differences between the two groups due to differences in timing of enrollment, but the average difference in date of enrollment of two to three months is probably not substantive.

Table 3.1 in Chapter 3 shows that the FSC-eligible group participated to a greater degree than the FSC-ineligible group. The results in Chapter 4 show that indeed, the impacts were greater on the FSC-eligible group, explaining a large part of the differences in impact among the three experiments.

Notes

1. Obtaining unbiased estimates of program impact by a process that is essentially “model free” is the major motivation for an experimental design. We draw on the discussion by Hausman and Wise (1985) to present the argument. Starting with the premise that any evaluation goal is to estimate an equation of the following simple form

$$(1) \quad Y = XB + e,$$

where \mathbf{B} is a vector of parameters to be estimated, with each element of \mathbf{B} measuring the effect on Y of a unit change on the corresponding element of X , and e are the unmeasured determinants of Y , referred to as the error term. \mathbf{B} will be an unbiased estimate of the effect of X on Y if X is uncorrelated with e . In the operations of a program, so-called “self-selection” is almost invariably going to create this correlation. For example, a government training program (the X variables) offered to those unable to obtain employment attracts those most motivated to get jobs (part of the e term), thus X and e are correlated, and if motivation affects outcomes, then Y and e are correlated, and \mathbf{B} is a biased estimate of the effect of X on Y . At best, according to Hausman and Wise (1985, p. 189), you can never be sure that the estimates are unbiased.

Randomization is a means for assuring that the estimates of \mathbf{B} are unbiased. Suppose there is a treatment, T , and random assignment is used to select persons to receive the treatment. Then in the following equation

$$(2) \quad Y = T\mathbf{B} + u,$$

where T is the treatment and u is the error term, T will be uncorrelated with u , and therefore \mathbf{B} will be an unbiased estimate of the effect of T on Y .

2. Extending the eligibility for an experiment to include potentially eligible as well as actually eligible persons would permit measurement of such an effect. For instance, in the income maintenance experiments, all adults were potentially eligible and families with incomes considerably above the benefit cut-off point were enrolled although they would receive no benefits. Such families, however, could reduce income in order to become eligible for income maintenance payments.
3. The exception in Washington occurred when claimants obtained employment that qualified them for a bonus without having established a waiting week for UI benefits. In this case, issues that might have resulted in the claimant being ineligible for benefits were not resolved, but the bonus was paid anyway; 7.6 percent (138/1,813) of those collecting the bonus did not serve a waiting week.
4. A reason advanced for excluding standbys in Pennsylvania was the desire to avoid antagonizing employers, who would not like standbys to be encouraged to take other jobs. Programmatically, the issue is difficult, because exclusion of standbys from bonus eligibility might simply encourage workers not to report standby status.
5. For a participant to receive a bonus payment while returning to a previous employer, the claimant must complete a form demonstrating that the position was different (e.g., different department, different job title, different salary level).
6. For the first two months of experimental enrollment in Washington (until May 9, 1988), the start of the qualification period for backdated claims was also backdated, reducing the length of the effective qualification period. After that date, the qualification period started with the filing date, the same as for other enrolled claimants. Although the proportion of backdated claims of experimental subjects increased substantially—from 10 percent to over 20 percent—between the two periods, the change could not be attributed to the treatment. This conclusion is

based on the absence of any statistically significant difference between experimental and control group members in the proportions backdating (see Appendix B in Spiegelman, O’Leary, and Kline 1992). At any rate, the proportion of claimants backdating claims more than two weeks was very small; only 1.4 percent of treatments and 1.2 percent of controls in Washington.

7. Of the 12,452 treatment-enrolled claimants in Washington, 1,445 were exempt from job search because of standby status or membership in full referral unions. The impact analysis showed essentially no difference in measured outcome including or excluding the exempt group. Compare Tables 5-4 and 5-5 in Spiegelman, O’Leary, and Kline (1992).
8. The workfare experiments are exceptions. These were a set of state-level experiments in which randomly selected AFDC recipients were placed into an experimental program and required to work or accept training in order to continue receiving welfare benefits. See Gueron and Pauly (1991) for a detailed description of these experiments.
9. The Pennsylvania and Illinois experiments had other components. The Pennsylvania experiment offered a voluntary job-search workshop of four half-days in combination with five of the bonus offer treatments. A sixth treatment was identical to one of the other treatments but did not include the workshop offer (labeled “PT4,” a high-bonus, long-qualification-period treatment). While the workshop was implemented as designed, so few claimants chose to participate (less than 3 percent) that the workshop was discontinued. For analysis purposes, the samples from the two treatments with identical bonus offers were combined and presented herein as a single high-bonus, low-qualification-period treatment.

The Illinois experiment actually comprised two experiments—a claimant experiment in which a \$500 bonus was offered to assigned UI claimants and an employer experiment in which the \$500 bonus was paid to the employer of the participating UI claimant who obtained a job under the same conditions. The employer experiment is not discussed further in this book, but it is described fully in the final report of the Illinois experiment (Spiegelman and Woodbury 1987) and more briefly in Woodbury and Spiegelman (1987).

10. In Illinois, the fixed bonus offer of \$500 was equivalent to an offer 4.2 times the mean WBA (\$120) and ranged in value from 9.8 times the minimum WBA of \$51 to 3.1 times the maximum WBA of \$161.
11. Pennsylvania used a parsimonious two-level experiment because of the need to have treatments that offered job workshops (see note 6 above) and the desire to experiment with a declining bonus offer. In this treatment, the bonus offer was set initially at $6 \times$ WBA, and the offer declined over the qualification period. This treatment is not easily compared with the fixed bonus amount offers and furthermore had no effect on UI outcomes. Therefore, it is not discussed further in this monograph. For details, see Corson et al. (1991).
12. In Pennsylvania, a small proportion of claimants (less than 2 percent) are only eligible for 16 weeks of benefits; the rest are eligible for 26 weeks.

13. Although some claimants could be expected to accept temporary jobs just to qualify for the bonus, it was difficult, given the rules regarding voluntary quits, for a claimant to go on and off jobs at will and still collect UI benefits or the bonus. The evidence from Illinois showed no tendency to shift unemployment from the first spell to later spells in the benefit year.
14. The mathematical programming model used for SIME/DIME is described in Conlisk and Kurz (1972).
15. Administration and research were paid for out of state funds not subject to these budgetary constraints.
16. As discussed previously, the two treatments in Illinois were actually two experiments, the claimant experiment in which bonuses were offered to claimants and the employer experiment in which bonuses were offered to employers to hire claimants.
17. Although logic suggests that a bonus offer would only serve to reduce the length of the unemployment spell, that is not necessarily the case. For those who would normally find jobs within the qualification period, the bonus could create an income effect that would lengthen the unemployment spell. Only if there is a compelling case to consider only reductions in unemployment as of interest should a one-tailed test be adopted.
18. At that time, a one-tailed statistical significance test was considered appropriate because of the belief that the bonus could only serve to reduce the length of the unemployment spell. Subsequently, theoretical considerations led to an interest in possible adverse effects of the bonus offer for at least some groups, which suggests a preference for a two-tailed test.
19. The effect size is a pure number index (called d in Cohen) that is the difference between the treatment and control group means measured in standard deviation units. For a two-tailed test, the effect size index is $d = |M_A - M_B|/s$, where d = effect size index for t tests of means in standard deviation units, M_A and M_B = population means expressed in original units, and s = standard deviation of either population.
20. The power of the test in the table refers to the power of a two-cell test. If there is more than one treatment and it is desired to know the joint power of the tests on all treatments (i.e., what is the probability of rejecting a true effect of any of the treatments), then one must look at the joint probability. Thus, if there are three treatments and each has a power of 0.8, the joint probability of rejecting at least one true effect is 0.51, i.e., $(0.8)^3$.
21. These sample size considerations do not take into account the desire to learn about the effectiveness of the experimental treatment on subgroups of the relevant population, but the costs of having independent treatment groups representing each interesting subgroup is too large. The reasons for needing knowledge about subgroup impacts are either that a policy that strongly affects some population subgroups and not others may not be politically desirable or that it may suggest the need for some restructuring of the treatment to better serve some groups. For instance, a policy ineffective for some racial groups (or one gender) may not be

good policy, even if the overall effect is positive. It must be accepted, however, that estimating subgroup impacts on fully separated samples reduces the reliability of estimates considerably. For instance, dividing the sample into four subgroups (say, two ethnic and two gender groups) implies that, for each of the four groups, only an effect size twice that for the group as a whole will be detectable at the same significance level and power. As discussed in Chapter 4, more parsimonious means of estimating subgroup impacts can be used.

22. See note 10 above for discussion of why the workshop treatment is not being further discussed in this book.
23. A detailed discussion of the sample allocation model is provided in Metcalf and Kerachsky (1988).
24. Another advantage of a larger number of sites, which is not as yet well understood, is related to the displacement effect discussed in Chapter 6. The possibility that displacement would contaminate the control group, thereby compromising the internal validity of the experiment, is small if the sample size relative to the size of the relevant labor market is small. This speaks in favor of smaller samples in many sites rather than larger samples in fewer sites. However, this does not reduce the possibility of displacement in a full-blown program, thereby affecting the generalizability of experimental results.
25. Of the 22 ES offices in which the experiment was conducted, 8 were unified (none of which were in Chicago); 2 others were not unified but were housed in the same building (both in Chicago); and the other 12 were in separate buildings (7 of which were in Chicago).
26. This conclusion is reinforced because not all of the 48 tests are independent. The most important interdependency occurs because each of the six treatment group means are compared with the same control group mean. If one treatment/control test fails, there is an increased likelihood that other treatment/control tests for the same characteristic will fail. As a possible demonstration of this point, among the six statistically significant differences, there are two pairs.
27. It may occur before the claimant becomes unemployed based upon an expectation of imminent unemployment and a desire to establish a benefit year before the event, or it may occur after the week of becoming unemployed. In this case, the benefit year may start in a week prior to the filing week, referred to as a “back-dated claim.”
28. There was an exception to this rule in the Washington experiment. Unlike in Pennsylvania, claimants who filed and were monetarily eligible for benefits but who did not proceed to claim a waiting week were still eligible for a bonus. This created a small group who might have been ineligible for nonmonetary reasons, which would only have been determined if the claimant claimed a waiting week.

References

- Cohen, Jacob. 1977. *Statistical Power Analysis for the Behavioral Sciences*. Revised ed. Orlando, Florida: Academic Press.
- _____. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Second ed. Hillsdale, New Jersey: Erlbaum Associates.
- Conlisk, J., and M. Kurz. 1972. *The Assignment Model of the Seattle and Denver Income Maintenance Experiments*. SRI Research Memorandum no 15, Menlo Park, California, July.
- Corson, Walter, Paul Decker, Shari Dunstan, and Stuart Kerachsky. 1991. *Pennsylvania Bonus Demonstration: Final Report*. Mathematica Policy Research, Inc., Princeton, New Jersey, September.
- Fisher, R.A. 1960. *The Design of Experiments*. Seventh ed. New York: Oliver and Boyd.
- Fisher R.A. 1968. *Encyclopedia of the Social Sciences*, Vol 5. New York: MacMillan and Co.
- Gueron, Judith, and Edward Pauly. 1991. *From Welfare to Work*. New York: Russell Sage Foundation.
- Hausman, Jerry, and David Wise, eds. 1985. *Social Experimentation*. Chicago: University of Chicago and NBER.
- Hicks, Charles R. 1982. *Fundamental Concepts in the Design of Experiments*. New York: Holt, Rinehart and Winston, Inc., p. 1.
- Kershaw, David, and Jerilyn Fair. 1977. *The New Jersey Income Maintenance Experiment*, vol. 8, p. xv, Academic Press, New York.
- Metcalf, C., and S. Kerachsky. 1988. "Sample Allocation and Office Selection." Memorandum dated March 25, 1988, Mathematica Policy Research, Inc., Princeton, New Jersey.
- Moffitt, Robert. 1992. "Evaluation Methods for Program Entry Effects." In *Evaluation Welfare and Training Programs*, C. Manski and I. Garfinkel, eds. Cambridge: Harvard University Press.
- Ross, Heather. 1966. "A Proposal for a Demonstration of New Techniques in Income Maintenance," Memorandum, December, Data Center Archives, Institute for Research on Poverty, University of Wisconsin, Madison.
- Rossi, Peter H., and Howard E. Freeman. 1989. *Evaluation: A Systematic Approach*. Newbury Park, California: Sage Publications.
- Spiegelman, Robert G., Christopher J. O'Leary, and Kenneth J. Kline. 1992. *The Washington Reemployment Bonus Experiment: Final Report*. Unemployment Insurance Occasional Paper 92-6, Washington, D.C.: U.S. Department of Labor, Employment and Training Administration.

- Spiegelman, Robert G., Philip K. Robins, and Richard West. 1983. *Final Report of the Seattle-Denver Income Maintenance Experiment*, SRI International, Menlo Park, California, May.
- Spiegelman, Robert G., and Stephen A. Woodbury. 1987. *The Illinois Unemployment Insurance Incentive Experiments: Final Report*. Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research.
- U.S. Department of Commerce, Bureau of the Census. Various years. *Statistical Abstract of the United States*.
- U.S. Department of Labor, Bureau of Labor Statistics. Various years. *Geographic Profile of Employment and Unemployment*.
- U.S. Department of Labor, Employment and Training Administration, *UI Data Summary*. Various years. Fourth quarter. Unemployment Insurance Service, Division of Actuarial Services.
- U.S. Department of Labor. 1984 and 1985. *Quarterly Unemployment Insurance Compilation and Characteristics*. Fourth quarter.
- Woodbury, Stephen, and Robert G. Spiegelman. 1987. "Bonuses to Workers and Employers to Reduce Unemployment: Randomized Trials in Illinois." *American Economic Review* 77(4): 513–530.

Reemployment Bonuses in the Unemployment Insurance System

Evidence from Three Field Experiments

Philip K. Robins and Robert G. Spiegelman
Editors

2001

W.E. Upjohn Institute for Employment Research
Kalamazoo, Michigan

Library of Congress Cataloging-in-Publication Data

Reemployment bonuses in the unemployment insurance system : evidence from three field experiments / Philip K. Robins and Robert G. Spiegelman, editors.

p. cm.

Includes bibliographical references and index.

ISBN 0-88099-225-5 (pbk. : alk. paper)—ISBN 0-88099-226-3 (cloth : alk. paper)

1. Insurance, Unemployment—United States—States—Case studies. 2. Welfare recipients—United States—States—Case studies. 3. Bonus system—United States—States—Case studies. 4. Insurance, Unemployment—Illinois. 5. Insurance, Unemployment—Pennsylvania. 6. Insurance, Unemployment—Washington (State). I. Robins, Philip K. II. Spiegelman, Robert G.

HD7096.U5 R35 2001

331.13'77—dc21

2001026848

© 2001 by the
W.E. Upjohn Institute for Employment Research
300 S. Westnedge Avenue
Kalamazoo, Michigan 49007-4686
All rights reserved.

The facts presented in this study and the observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of the W.E. Upjohn Institute for Employment Research.

Cover design by J.R. Underhill.
Index prepared by Leoni Z. McVey.
Printed in the United States of America.

01 02 03 04 05 06 8 7 6 5 4 3 2 1