Approaches to Evaluating Induced Entry Into a New SSDI Program with A $1 Reduction in Benefits for Each $2 in Earnings

By

Nancy B. Tuma

(With the assistance of Donald O. Parsons, Donald Patrick, John Rust, Joseph Sedransk, Judith Tanur, and David Vandergoot)

November 2, 2001

This paper is a working draft prepared by the authors under contract to the Social Security Administration. This paper is for discussion purposes only and not for quotation or attribution without permission of the authors. The views expressed herein represent the views of the authors and do not necessarily represent the position of the Social Security Administration.

SUMMARY

Section 302 of the Ticket to Work and Work Incentives Improvement Act requires the Social Security Administration (SSA) to undertake demonstration projects in order to determine the impacts on induced entry into, and reduced exit from, the Social Security Disability Insurance (SSDI) program if benefits are reduced by $1 for each $2 in earnings. This $1-for-$2 benefit offset is intended to encourage SSDI beneficiaries to work and to earn insofar as they can.

The term "induced entry" refers to a phenomenon in which people who are medically eligible for SSDI benefits but are not current SSDI beneficiaries are induced (i.e., encouraged) by the $1-for-$2 benefit offset to apply for SSDI benefits. The term "reduced exit" refers to a phenomenon in which SSDI beneficiaries remain on the benefit rolls longer than under the existing program because their earnings could exceed the level of Substantial Gainful Activity (the SGA level) without them being terminated from the SSDI rolls. If a $1-for-$2 benefit offset causes sufficient induced entry and reduced exit, SSDI program costs could increase, possibly substantially. But, if beneficiaries receive reduced SSDI benefits from working, and if there is little induced entry and reduced exit, SSDI program costs could decrease. Empirical study of the $1-for-$2 benefit offset is needed to determine if SSDI program costs would increase or decrease.

This paper, prepared by a team of seven consultants, assesses five research approaches to evaluating induced entry. It gives the most attention to demonstration designs, and the several variations of it that have been considered, because the legislation specifically proposed a demonstration project to study induced entry.

Alternative Research Designs

It is extremely difficult and costly to estimate the number and basic characteristics of induced entrants into a new program. Nonparticipants are a heterogeneous group about whom little is known. Consequently, it is hard to predict their response to a new program. Also, the number of induced entrants is likely to be a very small fraction of a very large number of nonparticipants, resulting in research designs that are searching for the proverbial needle in the haystack. The paper considers five approaches to assessing induced entry.

Prediction Based on the Aggregate Response to Previous Program Changes. This approach, which has already been used by the Office of the Chief Actuary, is inexpensive because new data are not collected; however, the quality of such estimates is hard to assess. Extrapolating the aggregate response from prior program changes can be misleading because a benefit offset is a much bigger change than previous changes. A priori, one cannot know if predictions based upon responses to prior changes will over- or underestimate responses to the proposed new program.

Prediction Based on a Dynamic Model of Individual Behavior. Good dynamic models predict average behavior well; very good models also predict the distribution or variability in behavior well. Data from a demonstration involving current beneficiaries could be employed to provide additional insight into the responses and costs of induced entry. The responses of current beneficiaries, however, do not necessarily give good guidance as to who might be induced to apply and how much they will work and earn because the induced entrants are likely to have better prospects for earnings than current beneficiaries.

Prediction Based on Responses to a Survey with Hypothetical Questions. To yield good estimates, a survey with hypothetical questions must explain a new program sufficiently well that

most respondents understand it and how it would affect them. Although this approach offers no silver bullet to the problems of evaluating induced entry, the consultants think that using it in addition to other approaches is worthwhile. The addition of questions to the planned National Study of Health and Activity (NSHA) may furnish valuable information. It can yield a more accurate estimate of the number of nonparticipants who are medically eligible for SSDI, as well as give information on which of them would apply to take advantage of the new SSDI program. A survey with hypothetical questions is much less expensive than a demonstration; the unresolved issue is whether this approach will yield satisfactory estimates of induced entry. A follow-up survey to NSHA permitting the collection of detailed information could improve these estimates.

Evaluation of Responses to a Demonstration. The paper examines several demonstration designs, including randomized assignment of individuals and designs utilizing states or counties as the observational unit. The paper identifies seven key issues that influence the quality and cost of each type of design. In general, the cost of the study increases as the quality of the results improves. Furthermore, a choice that helps to solve one design problem often amplifies another design problem. Indeed, design questions rarely have easy answers even when costs are ignored.

Classical experimental designs have very serious defects leading to unmanageable problems in implementing a demonstration project. The consultants think that this type of design should not be used to study induced entry. A demonstration that provides the treatment in 200 counties appears to be better than any alternative yet considered. However, the large numbers of localities and people involved, the potentially high increment to program costs, and some important unresolved defects in this design may mean that conducting such a demonstration is not worthwhile. Further, the size of this design would severely tax SSA's administrative capacity.

Implementation and Evaluation of a National Program. Implementing a $1-for-$2 benefit offset on a nationwide basis would provide a real-world test but would have several disadvantages. First, if it is implemented for a time-limited period, persons may behave differently than if it is permanent. Second, the implementation of a new national program could be very costly. Finally, because this approach would not provide a contemporaneous control group, one could not determine whether changes in program costs and in beneficiary behavior stemmed from the change in the SSDI program or from other socio-economic changes.

Conclusions

The consultants think that valuable information on the impact of induced entry may be acquired at relatively modest cost through the use of dynamic modeling of individual behavior and through responses to hypothetical question in a survey. Although these methods pose some issues and must be carefully interpreted, the consultants think that SSA should continue to pursue these methods, even if it is decided to study induced entry through a demonstration project.

The consultants agree that the classical experimental designs considered have very serious defects and should not be used to study induced entry. Designs based upon localities cannot be dismissed as easily as the classical experiment designs and could be considered. However, given the large numbers of localities and people involved, and given some key design problems that may not have good solutions, a demonstration study of induced entry may not be worth its cost. The consultants are uncertain whether such a design, despite its advantages relative to other designs, and despite its potentially huge costs, would yield estimates of induced entry that would be sufficiently accurate and reliable to meet policy-makers' needs.

Approaches to Evaluating Induced Entry
Into a New SSDI Program with
A $1 Reduction in Benefits for Each $2 in Earnings

When new legislation significantly changes an existing program or policy, people often respond by altering their behavior. Indeed, inducing people to change their behavior is often an aim of new legislation. Consequently, in evaluating the impact of a new program in general and its incremental cost in particular, one must consider changes in people's behavior. It is not safe to assume that people will continue to behave as they have in the past.

A further complication arises because one needs to consider the behavioral response of both those directly affected by the existing program ("current participants") and others seemingly unaffected by the existing program ("current nonparticipants"). Of the current nonparticipants, one must consider especially "induced entrants": new participants who enter the new program but who would not have entered the existing program.

"Induced entry" refers to a general and widespread phenomenon in which policy changes (e.g., changes in benefit levels or rules of eligibility for benefits) alter incentives for participation in some activity, so that individuals *who would not have participated under the former policy* are "induced" (i.e., encouraged) to participate under the new policy. Some form of induced entry may result from changes in taxation; the minimum wage; salaries of military personnel and civil servants; and subsidies, loans, and other forms of assistance to businesses, including farmers.[1] However, induced entry is rarely mentioned in such contexts. Rather, it is discussed mainly in relation to participation in programs that assist individuals who have difficulty in supporting themselves solely through their own efforts (e.g., see Moffitt 1992).

This paper focuses on induced entry into the SSDI program, the Disability Insurance program of the Social Security Administration (SSA), that would result if the existing program were changed in certain ways. The basic change under consideration is to allow SSDI participants (i.e., beneficiaries) to earn more than under the existing program and still receive some disability benefits: SSDI benefits would be reduced "$1-for-$2" above some disregarded level of a beneficiary's earnings, commonly called the "earnings disregard." Necessary work expenses, which vary across individuals, may also be disregarded (i.e., deducted from a beneficiary's gross earnings). One should keep in mind that beneficiaries' earnings are subject to federal, state and local income taxes, and that beneficiaries are required to contribute from their earnings to the Social Security retirement and Medicare funds.

Study of the impact of this possible change in the SSDI program is mandated by Section 302 of P.L. 106-170, the Ticket to Work and Work Incentives Improvement Act (TWWIIA) of 1999. This legislation states that:

---

[1]For example, lowering the minimum wage may raise entry into the work force of people with few skills.

> "The Commissioner of Social Security shall conduct demonstration projects for the purpose of evaluating, through the collection of data, a program for title II disability beneficiaries ... under which benefits payable ... are reduced by $1 for each $2 of the beneficiary's earnings that is above a level to be determined by the Commissioner."

The legislation also requires, among other things, that the demonstration projects are "to be of sufficient duration, sufficient scope and are to be conducted on a wide enough scale to permit a thorough evaluation of the project to determine ... the effects, if any, of induced entry into the project and reduced exit from the project." According to the legislation, the evaluation is to be broad enough in its coverage that findings can be used to generalize to the entire country.

Evaluations of program changes in the past have almost always focused on the response of current participants. This emphasis is usually appropriate because current participants' responses tend to have greatest effect on a new program's overall impact and cost. Fortunately, methods for estimating current participants' responses to a new program are highly developed (though rarely easy, quick, or cheap). Unfortunately, methods for estimating induced entrants and their responses are not yet well developed.

STUDY OF INDUCED ENTRY: MAJOR PROBLEMS

It is extremely difficult and costly to estimate the number and basic characteristics of induced entrants into a new program, and then to assess their behavior under the new program. There are two main reasons.

First, current participants are usually easier to locate, and their future behavior is easier to predict, because existing administrative records give some information about them and their past behavior. In contrast, nonparticipants are a more heterogeneous and often more geographically-dispersed population. Usually, little is known about nonparticipants, especially characteristics of them that can be used to predict how they will respond to a new program.

Second, evaluating induced entry is often very difficult because nonparticipants are far more numerous than current participants, and because current participants are more numerous than the likely induced entrants. Consequently, studying induced entrants resembles searching for a proverbial needle in a haystack, only one is looking for multiple needles in a very large haystack.

For example, in December 1999, of the approximately 277 million people in the United States, roughly 136 million (i.e., about half) were "disability insured," that is, potentially eligible for SSDI benefits, assuming they applied for SSDI and met the disability criteria (Table 4.C2).[2] A much smaller number, nearly 4.9 million, received SSDI benefits in 1999 (Table 5.A4). An even smaller number, about 0.62 million individuals, had applied to and entered the SSDI program

[2]Tables cited in this paper are in the *Annual Statistical Supplement to the Social Security Bulletin*, 2000. The rough estimate of the U.S. population in December 1999 was calculated by interpolating between the U.S. populations reported in the 1990 and 2000 U.S. Census of Population.

during 1999 (Table 6.C7); a similar number had applied but not been accepted into the SSDI program. Somewhat fewer, about 0.43 million people, left the SSDI rolls that year (Table 6.F2).[3]

Thus, SSDI beneficiaries are about 3.6 percent of insured workers (i.e., about 1.8 percent of the U.S. population). New entrants to the existing SSDI program in a year are about 0.46 percent of insured workers (i.e., about 0.23 percent of the U.S. population). Only about 0.32 percent of the insured worker population (i.e., about 0.16 percent of the U.S. population) leave the SSDI program in a year. These percentages fluctuate slightly from year to year but are typical.

Without collecting new data in a well-designed study, one can only roughly estimate the number of induced entrants to an SSDI program with a $1-for-$2 benefit reduction. The vast majority of current nonparticipants are not impaired and would not be eligible for SSDI, even if they applied. But current nonparticipants do include some people who are qualified for the SSDI program but who have either not applied for SSDI benefits, or applied but not been awarded SSDI benefits (e.g., because their application is still under review). These potentially eligible persons would serve as the main pool for induced entrants for the first few years after a new SSDI program starts. It is unclear how many current nonparticipants in SSDI are potentially eligible, and why some people do not participate in an existing program that would clearly benefit them.[4] Without collecting new data on the number of severely impaired people who do not receive disability benefits, one cannot confidently put even an upper bound on the number of induced entrants.

McLaughlin (1994) in SSA's Office of the Chief Actuary (OCACT) estimated that the likely number of induced entrants to the SSDI program resulting from a $1-for-$2 reduction in benefits due to monthly earnings above $85, the legal lower limit on disregarded earnings, would be roughly 400,000 in ten years, with the increase phased in at roughly one-tenth per year – that is, about 40,000 per year.[5] His estimate implies that, under a new $1-for-$2 program as compared

---

[3]Of those whose SSDI benefits were terminated in 1999, 38 percent died. Another 48 percent turned 65 years of age or left because their SSDI benefits were less than Social Security benefits. Only 13 percent were terminated because they no longer met the medical standards. A mere 2 percent were terminated for "other reasons," which apparently includes having earnings that were high enough and enduring enough to cause an individual to be terminated from the SSDI program.

[4]Because most severely impaired individuals cannot support themselves solely through their own earnings and savings, one expects that most individuals eligible for SSDI benefits would have actually applied for SSDI benefits, given the degree of impairment required to become a SSDI beneficiary. Brehm and Rush (1988) found, however, that some people who would qualify for SSDI benefits based on the severity of their impairments still work above the level defined as "Substantial Gainful Activity" (the SGA level) and therefore are ineligible for SSDI benefits. To become a SSDI beneficiary, they would need to stop working for at least five months while their application to SSDI was being considered. Some such individuals may think that five months without earnings are not worth the SSDI benefits that they might receive in the future, or they may not have a way to support themselves during the five-month period of joblessness needed to become eligible for SSDI benefits.

[5]He also considered a $1-for-$2 benefit reduction program in which the earnings disregard equalled the SGA level, which was $500/month then but was $740/month in fiscal year 2000. He estimated that the number of induced entrants would be about the same whether the earnings disregard was $85 or the SGA

with the existing SSDI program, about .029 percent more of the insured worker population (or about .015 percent more of the U.S. population) would apply for and be awarded SSDI benefits. Obtaining a good and reliable estimate of this increase is very difficult because it involves a very small change in a small percentage of a very large population. It requires study of a rare change for those who have experienced an already rare event – a disabling impairment. Moreover, it requires study of an effect that is likely to take a long time to reach its ultimate level. It will take time for eligible nonparticipants to learn of any changes in the SSDI program, decide what course of action would be best for them, apply to the new program, and go through at least a five-month waiting period between applying for and being awarded SSDI benefits.

Such a small increase in the entry rate could, however, substantially increase the cost of the SSDI program. The average benefit of current SSDI beneficiaries in 1999 was $754/month, or $9,048 per year (Table 5.E2). If average benefits of induced entrants resemble those of current beneficiaries, the addition of 40,000 induced entrants every year would increase the cost of the SSDI program by about $360 million every year, ignoring increases in Medicare costs. OCACT has estimated that the cost of the new SSDI program could rise by $4 billion in its first five years, excluding increased Medicare costs, if there were to be as many induced entrants as McLaughlin (1994) estimated.[6] If induced entrants remain SSDI beneficiaries for as long as current SSDI beneficiaries (or longer), or if they had higher average benefits than current beneficiaries (as some have speculated), the incremental cost due to induced entry could be even higher. Further, the incremental cost per period (in real dollars) could rise over time under certain scenarios.

On the other hand, if induced entrants are fewer than estimated or rather quickly become earners with reduced, below-average SSDI benefits, the cumulative impact of induced entry on incremental costs of the SSDI program could be much less. Indeed, the incremental cost due to induced entrants might be offset by reductions in SSDI benefits to current beneficiaries who start earning more than they do under the existing SSDI program. It is even possible that the cost of the new SSDI program would be less than the existing program, if enough beneficiaries' benefits decline because of their increased work efforts and earnings.

It should be emphasized that in making his estimates of the number of induced entrants and of incremental program costs, McLaughlin (1994) admits that the actual number could be appreciably higher or lower. He states that, "Considering the uncertainty about the size of the population that could qualify for benefits in the absence of earnings, it is clear that the actual effect ... could be substantially different from our estimate."

Thus, without new data collection, there is considerable uncertainty about incremental costs of a SSDI program with a $1 reduction in benefits for a $2 increase in beneficiaries' earnings. This is the reason why one might want to collect new data in a well-designed study in order to learn how many people would be induced to enter the new SSDI program and how they would behave under it in the long run. To what extent would more people with severe impairments enter

---

level, but that the incremental cost of the SSDI program would be roughly 2.5 times higher if the earnings disregard equalled the SGA level.

[6]Weathers et al. (2000, p. 74, footnote 1) explain briefly how this estimate was obtained.

the new SSDI program? To what extent would SSDI beneficiaries work under the new program? How quickly (or slowly) would SSDI beneficiaries return to work? How much would they earn, and what would be their SSDI benefits? How long would they remain SSDI beneficiaries? These are some of the questions that need to be answered so that the impact of induced entry on the cost of a change in the SSDI program can be estimated.

Clearly these are important questions. But an equally important question is: What kind of study can provide good answers to these questions? How much would a credible study cost? Would the likely benefits of such a study exceed the costs of the study?

INDUCED ENTRY: HOW TO STUDY?

Although social scientists recognize many basic problems in studying induced entry, good solutions to most of the problems are not readily available. An honest assessment is that the methods for reliably estimating the level of induced entry into any new program (not only a new SSDI program), and the subsequent behavior of induced entrants, are not yet well developed.[7] Consequently, it is unusually hard for SSA to develop ways to satisfy its mandate under Section 302 of P.L. 106-170 to evaluate induced entry resulting from a change in the SSDI program to include a $1-for-$2 benefit reduction. It is especially hard to perform an evaluation that would yield reliable estimates of induced entry and its impacts. In this context "reliable" means that the estimates would be "on target" (i.e., not biased) and that similar estimates would be reached if the study were repeated many times (i.e., the confidence interval around the estimate is narrow).

The Office of Research, Evaluation, and Statistics (ORES) in SSA is responsible for recommending to the Social Security Commissioner how to comply with the legislative mandate to "determine ... the effects, if any, of induced entry" resulting from a $1 reduction in benefits for each $2 in earnings. In preparation for making its recommendations, members of ORES have consulted with the 12-member Work Incentives Advisory Panel (WIAP)[8] and a team of seven consultants with expertise relevant to evaluating the proposed change in the SSDI program.

APPROACHES TO ASSESSING INDUCED ENTRY

There are five basic approaches to assessing induced entry resulting from a change in the SSDI program. SSA's ORES has discussed each type with its team of consultants and with the members of the WIAP. It has, however, concentrated mainly on approaches that involve

---

[7]As noted earlier, methods for studying the response of current participants to a program change are well developed, having evolved over more than 30 years. Social scientists have much experience in using such methods to make reliable estimates of current participants' response to a program change.

[8]The current members of the WIAP are Richard V. Burkhauser, Kristin E. Flaten, Thomas P. Golden, Frances Gracechild, Christine M. Griffin, Larry D. Henderson, Jerome Kleckley, Stephanie Smith Lee, Bryon R. MacDonald, Sarah Wiggins Mitchell (Chair), Stephen L. Start, and Susan Webb. Brief biographical information on WIAP's members may be found at http://www.ssa.gov/work/panel/.

"demonstration projects for the purpose of evaluation, through data collection," as mandated by Section 302 of P.L. 106-170. This paper also gives more attention to this approach. It is useful, however, to keep a broader set of approaches in mind so that the relative merits of different approaches come into sharper focus. The five basic types are summarized below in the order of their probable cost, starting with the least expensive approach.

1. Prediction Based on the Aggregate Response to Previous Program Changes

In this approach, one reviews information about people's responses to previous changes in the existing program (i.e., SSDI), in other programs aimed at a similar target population (e.g., the SSI disability program), and in other programs involving similar behavioral adaptations (i.e., changes in work effort and earnings). Applications of this approach can range from making rough calculations to using highly sophisticated mathematical models and/or actuarial projection techniques that use existing administrative records as their input.

This approach is usually the least expensive because there is no new data collection. But the quality of the resulting estimates is hard to assess. The quality varies because the aggregate-level model may not reflect very well the actual, combined responses of many individuals to the program change. If the new program differs substantially from the existing program, extrapolating from the aggregate response to minor program changes in the past can be misleading. A priori, one cannot be sure if prediction based on the aggregate response to previous program changes over- or under-estimates the aggregate response to the proposed new program.

The proposed $1 reduction in SSDI benefits for each $2 increase in earnings above some level is a major change in the SSDI program. The proposed new program implicitly changes the definition of "disabled." Under the existing SSDI program, non-aged adults are disabled if they have an impairment that prevents them from earning above the SGA level. Under the proposed new SSDI program, the disabled would also include non-aged adults who have a severe impairment that limits gainful work activity but that does not fully prevent it.

Estimates based on aggregate responses to previous program changes, such as those made by McLaughlin (1994), may be somewhat useful because they may suggest whether induced entry should be studied more thoroughly and whether there is a need to collect new data.

2. Prediction Based on a Dynamic Model of Individual Behavior

Another approach is to forecast people's behavior using a dynamic model that has been calibrated to predict well under circumstances that have previously been studied and for which data already exist. In this instance, the dynamic model would first be adjusted (i.e., calibrated) to predict behavior of current SSDI beneficiaries. A *good* model is one that predicts average behavior well, not only for all beneficiaries, but also for important subgroups (e.g., individuals who are alike in gender, age, place of residence, type of impairment, length of time as a SSDI beneficiary). A *very good* model is one that also predicts well the distribution (i.e., variability) of behavior, as well as the average.

Building a good dynamic model from scratch would be very expensive. However, previous

investments in basic research on people's behavior under various circumstances have provided social scientists with much knowledge that sometimes allows such a dynamic model to be developed at relatively low cost. Moreover, a dynamic model can be improved steadily over time. When a model is found to lead to an inaccurate prediction, a search for what is wrong with or missing from the dynamic model follows. Once social scientists understand the reasons for prediction errors, they can work to modify the dynamic model so that it yields better predictions under a wider range of conditions in the future.

Data on how current SSDI beneficiaries respond to changes in the program – whether they increase their work effort and earnings, and whether they leave the SSDI rolls less often – will be gathered in the demonstration study of "reduced exit" mandated by Section 302 of P.L. 106-170, which is currently being planned by ORES (discussed below). The new data collected in this study can be used in conjunction with a dynamic model to gain additional insight into the responses and costs associated with proposed changes in the SSDI program. Findings about the responses of current SSDI beneficiaries may, however, not necessarily give good guidance about the level of induced entry or about the long-term work efforts and earnings of induced entrants. As compared to current SSDI beneficiaries, induced entrants have better prospects for earning above the SGA level.

3. Predictions Based on Responses to a Survey with Hypothetical Questions

A third approach involves surveying members of the population at risk of induced entry and eliciting responses to questions about their behavior under various hypothetical scenarios. The hypothetical situations start with a description of a new program. Respondents are then asked what they would be likely to do, if the new program were adopted. Since many people may know little about the existing program, they are asked what they know about it. If necessary, key features of the existing program are explained to them. They are then asked what they are likely to do in the future, if the existing program is retained.

The survey may also ask respondents to predict their behavior under both the new and existing programs in hypothetical personal situations that may affect their behavior under the new and existing programs. They are also asked about their characteristics, circumstances, and behavior under the existing program.

Responses to the survey are analyzed using appropriate statistical methods. Results of the statistical analyses, in conjunction with information on characteristics of the population at risk, are then used to estimate the likely impact of the new program on people's behaviors and on the incremental cost of a nation-wide program. The statistical results may be used in conjunction with a dynamic model (see the second approach, described above), or separately.

To yield good estimates, a survey with hypothetical questions must explain a new program sufficiently well that most respondents understand it and how it would affect them. In addition, the accuracy of predictions based on the survey responses depends critically on people's ability to assess their reaction to a new program. It is hard, however, for people to make realistic assessments of their future behavior under a new, hypothetical program on the spur of the moment in a survey situation in which they answer numerous questions in a brief period of time. Ideally,

the survey should also test whether the respondents really understand the programs that are described to them. There should be questions that allow analysts to evaluate whether respondents' predictions of their behavior in various hypothetical situations are logical and consistent.

Although the third approach offers no silver bullet to the problems of evaluating induced entry, the consultants think that using it in combination with other approaches is worthwhile. For example, it can be useful to compare current SSDI beneficiaries' predictions of their response to the new SSDI program in the survey with current SSDI beneficiaries' actual response to a demonstration of the new program. This kind of comparison helps one to learn how well people can anticipate their actual response to a new program. Many individuals who are potentially eligible for SSDI but are not current SSDI beneficiaries may have difficulty either in understanding the new (or even the existing) SSDI program or in predicting their behavior. To the extent that this is true, a survey with hypothetical questions about people's reactions to a new SSDI program will not yield good estimates of induced entry.

ORES and its team of consultants have discussed the addition of certain questions to the National Study of Health and Activity (NSHA), which was already being planned (for a summary, see Appendix A). It is hoped that responses to the additional questions will yield information that can be used to improve prediction of responses to a change in the SSDI program.

First, the NSHA will gather data that will yield a greatly improved estimate of the number of people who are not currently SSDI beneficiaries but who are medically eligible for SSDI benefits. Obtaining a good estimate of this number is extremely important because it is the upper limit on the number of induced entrants in the short run. Information on the impairment histories of these individuals will be useful in gauging the upper limit on the long-term increase in induced entrants, in contrast to a short-run increase due to "pent-up demand" for a SSDI program that allows beneficiaries to both earn and receive partial benefits for an extended period of time.

In addition, NSHA will also include questions concerning people's knowledge of the existing SSDI program (and other programs overseen by SSA). Questions about a person's hypothetical response to a new SSDI program with a $1-for-$2 benefit reduction are also being planned.

ORES and its team of consultants are discussing the benefits and design of a follow-up survey to NSHA. The follow-up survey would permit the collection of more detailed information about individuals with disabling impairments. These discussions are at an early stage, but it is planned to ask respondents to NSHA to consent to being recontacted in a follow-up survey.

A survey with hypothetical questions is clearly much less expensive than an actual demonstration. The unresolved issue, of course, is whether such a survey approach will yield satisfactory estimates of induced entry and other responses to changes in the SSDI program.

4. Evaluation of Responses to a Demonstration (Field Experiment)

The approach generally expected to yield the most reliable estimates of response to a new program is a demonstration or "field experiment." One key element to the successful use of this approach is comparison of the behavior of individuals exposed to the new program (the "treatment"

group) to the behavior of other individuals exposed to the existing program (the "control" group). Behavior of members of both groups is often compared with the behavior of the same individuals at an earlier time before the "treatment," and possibly also at a later time after the "treatment" ends. However, if the behavior of those in the treatment group is not compared to the behavior of those in the control group at the same time, one cannot be confident that behavioral differences between the treatment and control groups are due to the program change rather than to some other change (e.g., a change in the rate of economic growth or in other social policies).

Another key element to the successful use of this approach is random assignment to the treatment and control groups. If individuals are permitted to choose between the new and existing programs (i.e., between the treatment and control situation), they will rationally select the one that is better for them. No matter how similar members of the two groups may appear to be at the outset, they inevitably differ in ways important to the outcome because one group prefers the new program and the other group prefers the existing program. Indeed, results of the study would also be compromised if political leaders of geographical areas were permitted to choose between the new and existing programs for their areas: these leaders would also tend to choose the program that is better for their constituents. Then the areas in the treatment and control groups would inevitably differ in ways that are important to the outcome. Neither scientists nor anyone else should be permitted to choose who is in which group. Even with the best intentions, assignment to the treatment and control groups based on anyone choosing (rather than on some random assignment process) is likely to bias the findings. For these reasons, scientists think that random assignment to the treatment and control groups is <u>essential</u> for a study of the impact of a new program or treatment on people's behaviors.

Although random assignment is crucial, various attributes of members of the treatment and control groups can be controlled to get somewhat better estimates at a lower cost. For example, individual attributes that one might want to control include age, gender, and nature of the disability; attributes of areas that one might want to control include population size, population density, and the unemployment rate. Various techniques may be used to control these and other factors that affect the outcome. But, controlling attributes of individuals and of areas does not obviate the basic need for, and value of, random assignment to the treatment and control groups.

Various methods of selecting geographical areas to get good estimates of induced entry at a somewhat lower cost, and the relative advantages of these various methods, have been considered by SSA's ORES in considerable depth. We summarize their analyses and review their conclusions about these various methods later in this paper.

5. Implementation and Evaluation of a National Program

Finally, legislation could introduce a new SSDI program with a $1-for-$2 benefit reduction. It might be introduced as a new permanent program or as one that would run for a certain number of years, with continuation of the new program after an initial expiration date subject to Congressional and Presidential approval.

A major advantage of this approach is that the program implemented and studied would have the features preferred by Congress and the President. Further, it would be a "real world" test of

the entire program, including both administration of it and dissemination of information about it. In particular, people's knowledge about the new program would be acquired in a typical way: Researchers would not need to develop special, atypical ways to explain the new program to participants in a study. Another big advantage is that the new program would be backed by the full force and authority of the U.S. government. Actual and potential beneficiaries would take the new program seriously, and the experts on whose opinions they rely (e.g., physicians and others providing services to people with disabilities), would advise potentially eligible people about the new program. Lastly, actual and potential beneficiaries could not move to another location within the U.S. to benefit from or to avoid the new program.

This approach has, however, several noteworthy disadvantages. First, it would be the most costly because the new SSDI program would be implemented nationwide. Second, it would not furnish results for a "control" group against which results for the "treatment" group could be compared in a straightforward way. (The entire U.S. population would be in the "treatment" group.) If the cost of the SSDI program increased, it could be because the SSDI program changed or, for example, because the economy had gone sour. On the other hand, if the cost of the SSDI program decreased, it might be because the SSDI program changed or because the economy had improved so much that employers had become more willing to hire people with disabilities. Many other exogenous environmental changes could occur to make it unclear whether the program change per se caused program costs to vary. It would be clear fairly soon, however, whether the cost of the SSDI program was rising, and if so, whether the rate of increase was higher than in the recent past under similar economic conditions.

Third, if legislation put a time limit on the new national program, then people might behave differently than they would under a national program intended to be permanent. Some would want to "hurry and apply" to get the maximum benefits of the new program; others might not bother to apply to the new program at all, thinking that temporary benefits from it would not be worth the trouble. Alternatively, introducing the new program as a permanent one would avoid these problems. But, it would then be necessary for Congress and the President to have the political fortitude and clout to change the SSDI program yet again, if the $1-for-$2 program should turn out to have undesirable consequences (e.g., unexpectedly high program costs due to induced entry).

> Consultants' Conclusion: Using several different types of approach to evaluate induced entry under a SSDI program with a $1-for-$2 benefit reduction would give increased confidence in the conclusions about the impact of induced entry. Of the five approaches mentioned above, serious consideration should be given to a dynamic model of individual behavior and a survey with hypothetical questions, as well as a demonstration.

Since Section 302 of P.L. 106-170 mandates a demonstration, this paper focuses on that type of approach below.

THE DEMONSTRATION APPROACH TO STUDY OF INDUCED ENTRY: BASIC ISSUES

Before describing specific demonstration designs and assessing their merits, we review general issues and problems pertinent to all of the designs. Most of these issues and problems are

discussed by Weathers et al. (2000), usually in greater detail than in the present summary.

1. Similarity of Demonstration Program Studied to a New National Program

The reliability of conclusions derived from any demonstration depends critically on the similarity between the program studied and the national program eventually implemented by new legislation (Weathers et al., 2000, p. 85). If key features of the programs studied in the demonstration differ from the corresponding features specified by new legislation, generalizing from the study's conclusions to the national program is highly problematic.

Section 302 of P.L. 106-170 states that a $1-for-$2 benefit reduction is to be studied. It does not permit SSA to study any other benefit reduction. But, Congress could conceivably consider and choose some other rate of benefit reduction (e.g., $1-for-$3 or $2-for-$3). A study of a $1-for-$2 benefit reduction will provide little guidance about the impact and cost of a new SSDI program with some other rate of benefit reduction.

Section 302 does not specify other potentially important program characteristics:

(1) Section 302 states that the Commissioner of Social Security is to establish the level of earnings at which a benefit reduction would start (the "earnings disregard"). The level of the earnings disregard determines which individuals will benefit from the new program and how much they will benefit.

(2) The existing SSDI program permits a nine-month period of "trial work" before a beneficiary suffers any loss of benefits. The length of the "trial work period" could have a substantial impact on response to a new SSDI program. The longer the trial work period, the more likely that beneficiaries will try to work and earn.

Those responsible for designing the demonstration must make decisions about the program parameters not specified by Section 302 of P.L. 106-170, despite imperfect information about the parameters that Congress and the President would be likely to choose for a national program. For various scientific and other reasons, SSA may also need to study program parameters that would not be considered for a national SSDI program.

Decisions about the lengths of the trial work period to be studied in the demonstration illustrate this point. A trial work period of indefinite length is highly unlikely in a national program but would give very useful scientific information: It would help analysts to determine the upper bound on the level of induced entry and on the long-run behavior of induced entrants. In contrast, a trial work period of zero length (i.e., elimination of a trial work period) has been considered by SSA (e.g., McLaughlin 1994), and it might well be attractive to Congress and the President in a national program. If elimination of trial work seems likely in a national program, it should be studied in the demonstration. Studying it in a demonstration raises ethical concerns, however, because elimination of trial work would disadvantage participants in the demonstration as compared to those in the control group.

Additionally, Section 302 mandates that the new "Ticket to Work" program is to be included

in the demonstration. But the Ticket program has not been implemented everywhere, and it is unclear how long full implementation will take. Even where the Ticket program has begun, it is often just getting off the ground. Some time is likely to elapse before the Ticket program settles into a routine, steady-state situation throughout the country. For this reason, features of the Ticket program affecting beneficiaries' work efforts are likely to vary over time as well as across geographical areas. Even if a SSDI demonstration has program parameters identical to those chosen for a new national program, the relationship of the Ticket program to a new national program and to the demonstration could differ because the Ticket program is evolving and because response to a $1-for-$2 program could depend on how the Ticket program is operating.

A further problem with combining the Ticket program and the $1-for-$2 program arises because providers of training under the Ticket program are paid a substantial amount only if a trainee completely leaves the SSDI benefit rolls. But SSDI beneficiaries with earnings are less likely to leave the SSDI rolls completely under the $1-for-$2 program than under the existing program. Consequently, the incentives for training providers to help SSDI beneficiaries to work and earn are smaller under the $1-for-$2 program than under the existing SSDI program.

The choice of the earnings disregard(s) to be studied also presents a dilemma: On the one hand, if the earnings disregard equals the minimum allowable by law, $85, many beneficiaries who currently work would be worse off the under the $1-for-$2 program than under the existing SSDI program. Making participants in a study worse off raises ethical concerns. To avoid this, it has been suggested that people chosen to participate in a demonstration of the $1-for-$2 program should be allowed to choose to remain under the existing SSDI program. Allowing study participants to choose whether they are in a treatment or control group would greatly complicate data analysis. Models that seek to correct for the selection biases that result when people can choose their treatments would be needed. One is never sure, however, whether a model used to correct for selection biases makes the desired corrections effectively.

On the other hand, if the earning disregard equals the SGA level, no beneficiary who works would be worse off under the $1-for-$2 program than under the existing SSDI program. Hence, the ethical issue raised by a lower earnings disregard would vanish, and there would be no reason to allow study participants to choose whether they are in a treatment or control group. However, if the earnings disregard equals the SGA level, program costs could be considerably higher than if it equals $85 because many beneficiaries may be able to earn something (e.g., a few hundred dollars per month), but not more than the SGA level. Moreover, one would not know how people would behave in a program with a lower earnings disregard.

2. Scale Effects

A new national program and a demonstration always differ in important ways. People take a national program more seriously than a demonstration and are more likely to attend to it. They also usually regard a national program as more permanent. Some may hurry to take advantage of a demonstration project while it exists; others may ignore it because it is temporary.

A national program has a comprehensiveness that a demonstration project can never achieve. Unlike a demonstration, a national program is not subject to "migration" effects, in which people

move from one area to another to benefit from the demonstration program or to avoid it.

In addition, a national program has subtle but possibly important "macro-level" effects that are impossible, for all practical purposes, to measure in a demonstration. First, SSA's administration of a national program and of a demonstration would almost surely differ, and these administrative differences might be consequential for both beneficiaries' behavior and for program costs. In addition, under a national program, employers might modify their hiring practices vis-a-vis disabled workers, and states might alter their disability-related programs. In contrast, a demonstration project is extremely unlikely to lead employers or states to act differently.

Finally, much learning about a national program that is an essential precursor to becoming a participant occurs informally – through people's interactions with relatives, friends, neighbors, opinion leaders, and relevant experts in their local community. Modes of learning in a social scientific study tend to operate differently. "Saturation" studies in which a demonstration applies to an entire community (e.g., county or state) attempt to overcome this limitation. However, the time limits on a demonstration and the need to report results quickly rarely allow the differences in the ways people learn about a national program and about a demonstration to be overcome fully.

3. Hawthorne Effects

In an early study of behavior of workers in the Hawthorne plant, Roethlisberger and Dickson (1939/1961) found that people acted differently simply because their behavior was being carefully observed. When people know that their behavior is being closely monitored and, especially, that important decisions will be made on the basis of researchers' observations, they often behave differently than they would otherwise. This effect on individuals of their being studied came to be called the "Hawthorne effect." It is sometimes possible to observe people's behaviors without them knowing that they are being observed, but this is not possible in any large-scale demonstration of a new social program. One way to avoid a Hawthorne effect is to implement a national program, observe what happens, and change the national program if it has undesirable consequences. Another way is to have two control groups, one that is monitored as closely as the treatment group and another that is monitored using only normal administrative procedures (a so-called "silent control group"). If behavior in the two control groups differs significantly, it means that people are partly responding to being carefully monitored.

4. Duration and Timing Effects

People's responses to both a new national program and a demonstration project vary over time. It takes time for people to learn of a new program, understand it, and decide how it might affect them; therefore, their response tends to be less initially than later on. Once they understand the program change, there may be an unusually large response for a while due to "pent up demand" among those who like the new program but did not like the previous one. As more time passes and "pent up demand" has been met, the aggregate-level response may again decrease to a level that can be considered to be the long-term response. These differences are sometimes discussed in terms of the effect of a "stock" versus a "flow" (Weathers et al., 2000, pp. 113-114).

An additional source of variation over time occurs in a demonstration, especially one with

a clear termination date. If participants expect the demonstration program will apply to them for only a few years, their response may be different than it would be if it were permanent. Some may rush to take advantage of the program; others may decide that it will not last long enough to bother with it. Those who are motivated to participate in it may anticipate its end and start to adapt to their anticipated situation afterwards, although the demonstration is still on-going. To avoid these problems, participants in a demonstration may be promised that the demonstration program will apply to them indefinitely, even after data collection ends. Ethical issues arise unless this promise is kept, but participants in a study may not believe this promise in any case.

The response to both a national program and a demonstration also vary over time because people's response depends on the macro-economic situation (e.g., the unemployment rate). Data collection in a demonstration lasts a finite amount of time and occurs during a certain phase of the economic cycle. A long-lasting national program will eventually encompass both economic booms and busts. It may be unwise to generalize from data gathered during the first few years of a demonstration to a permanent national program, especially for phases of the economic cycle that did not occur during the data-collection period of the demonstration.

Various statistical models may be used to try to correct for duration and timing effects. But inevitably, they introduce considerable uncertainty into whatever conclusions are reached.

5. A Tolerable Increase

Evaluating change in any program raises a key issue. What is an acceptable, or at least tolerable, increase in program costs and other adverse consequences due to the program change? While everyone hopes that the cost of the program will not increase (and perhaps even decrease) and that there will be no other adverse effects, such hopes may not be fulfilled. At what point does an increase in costs or other detrimental consequences become "too much" – not "tolerable"? As the size of a tolerable increase goes up, the size of the study needed to estimate the actual increase goes down. (If <u>any</u> increase in program costs would be tolerable, one would not need to estimate the effect of the program change ahead of time.)

As mentioned, McLaughlin (1994) projected that a SSDI program with a $1 reduction in benefits for a $2 increase in a beneficiary's earnings above <u>either</u> $85 <u>or</u> the SGA level might lead to about 40,000 induced entrants per year (i.e., an increase of about 6.5 percent). Based on his projection, OCACT has estimated that the incremental cost of SSDI with a $1-for-$2 benefit reduction could be $4 billion in five years, excluding increases in Medicare costs (Weathers et al., 2000, p. 74). Based on reactions to these projected increases in the numbers of SSDI beneficiaries and their associated costs, ORES concluded that such an increase is regarded as <u>not</u> tolerable.

Consequently, in evaluating various designs, ORES has considered zero, two, five, and ten percent increases in the number of entrants into SSDI per year and then considered sample sizes and other design features that would be likely to detect such increase in a reliable and cost

effective way.[9] An increase of ten percent is probably above the tolerable level, but it seems reasonable to the team of consultants that an increase of two or five percent might be tolerable. (Neither Section 302 of P.L. 106-170 nor the Social Security Commissioner has said what increase in entry rates or in program costs might be tolerable.)

ORES's analyses of various designs start with the number entering the existing SSDI program each year. As stated earlier, about 0.62 million people (0.456 percent of insured workers and 0.224 percent of the U.S. population) enter the SSDI program each year. Two and five percent increases in entry rates imply approximately 12,000 and 31,000 induced entrants per year, respectively. ORES has considered the features of a research design necessary to detect such increases in the number of people entering the SSDI program each year, on average. These figures mean that the percentage of insured workers entering the SSDI program each year would rise from 0.456 percent to 0.465 percent if there are 12,000 induced entrants per year, and to 0.479 percent if there are 31,000 induced entrants per year. Thus, the entry rate per year would increase by 0.009 percentage points in the first case and by 0.023 percentage points in the second case. Although both percentage-point increases are extremely small numerically, even the first case could add more than a $1 billion to the long-run costs of the SSDI program under plausible (though by no means certain) assumptions about beneficiaries' behavior and earnings under a SSDI program with $1-for-$2 benefit reduction (L. Scott Muller, personal communication).

6. Statistical Issues

In planning any study in which conclusions will be based on statistical analyses of data on participants in the study, a key issue is the size of the sample - that is, the number of participants in the study. In determining the sample size needed, ORES has envisioned that the data collected in the study will be used to conclude whether the $1-for-$2 program leads to any increase in entry into the SSDI program (and whether it increases program costs because of induced entry). This procedure consists of performing a statistical test of a null hypothesis - in this instance, that there is no induced entry. In such a test, analysis of the data from a particular sample can lead to two possible errors. One is to conclude that there is a change (i.e., induced entry) when in fact there is no change; the second is to conclude that there is no change (i.e., no induced entry) when there actually is a change. (These are termed type I and type II errors, respectively.) There is no guidance in P.L. 106-170 about the levels of these errors that are acceptable. Following scientific tradition, when calculating the required sample sizes, ORES has chosen a five percent chance of the first kind of error, and either a ten or twenty percent chance of the second type of error.

Researchers must also decide if both a positive and negative change are of concern; if so, they should use a "two-sided" test. If only an increase or a decrease is of concern, they should use a "one-sided" test. In a study of induced entry under a $1-for-$2 program, the sole concern is a possible increase in the rate of entry into the SSDI program. (Only an increase in the entry rate could increase the cost of the SSDI program. Moreover, economic theory unambiguously predicts that a $1-for-$2 benefit reduction program would not decrease the entry rate because no

[9]Note that this number of induced entrants might not increase costs of the SSDI program appreciably, if many beneficiaries increase their work efforts and earnings and therefore receive lower SSDI benefits.

one would be worse off under this new program than under the existing program.) Hence, ORES has appropriately chosen to use one-sided tests in making sample size calculations.

An alternative approach, not used by ORES, is to calculate a sample size that will yield an estimate of the quantity of interest that is sufficiently close to the true but unknown value of this quantity. In this instance, one would like to estimate the difference between the entry rate under the $1-for-$2 program and the entry rate under the existing program. One would like the deviation between the estimated difference and the true difference not to exceed a specified value, e.g., x percent. One can then use statistical theory to select a sample size so that, with a specified probability p, the deviation will not exceed x percent. In other words, we can ensure that, with probability p, the estimated difference in entry rates under the new and existing programs will deviate from the true (but unknown) difference in entry rates by no more than x percent – if a sample of that size chosen in a way that satisfies the assumptions of statistical theory (e.g., concerning randomness).

> Consultants' Conclusions: A somewhat larger chance of the first kind of error (e.g., ten percent rather than five percent) might be acceptable. The chance of the second kind of error might be reduced (e.g., from ten or twenty percent to five percent). Sample sizes might be calculated on the basis of one-sided confidence intervals rather than on the basis of one-sided tests of the hypothesis that the entry rate does not rise under a $1-for-$2 benefit reduction program. Various other, slightly different assumptions might be made in calculating the necessary sample sizes. The necessary sample sizes for a particular design would then be somewhat different from those calculated by ORES but probably not dramatically different.

7. Relationship to the Demonstration Study of Reduced Exit

Section 302 of P.L. 106-170 mandates that there be a demonstration study of "reduced exit" as well as a demonstration study of "induced entry." Although the study of reduced exit largely lies outside the scope of this paper (see Appendix B for a short summary), its relationship to the study of induced entry does need to be considered because the two studies could be conducted sequentially or simultaneously (and jointly).

If the two are conducted sequentially, it would be sensible to study reduced exit first because it would be a less costly study. One would then be able to estimate the impact on program costs of reduced exit and beneficiaries' work behavior under the $1-for-$2 program. If the new program generates costs savings, it would be clearer how much induced entry could be occur without increasing the cost of the SSDI program above its current level. If program costs rise under the $1-for-$2 program, solely on the basis of reduced exit, it would be obvious that a national $1-for-$2 program would cost more than the existing program. (Economic theory unambiguously predicts that any level of induced entry will increase program costs, unless there are compensating cost savings because beneficiaries earn more and have lower life-time benefits.) Under this scenario, the more difficult and expensive study of induced entry might then not be implemented.

One downside of sequential studies is that final results could take twice as long to obtain (e.g., ten years rather than five years). There are some other good reasons for conducting the two studies simultaneously. It is as important for the study of reduced exit as for the study of induced

entry that the demonstration $1-for-$2 program be as realistic as possible. Conclusions from both studies could be faulty otherwise. In particular, dissemination of information about the $1-for-$2 benefit reduction needs to be realistic in both studies. Current as well as potential beneficiaries need to learn and form opinions about the $1-for-$2 program in a manner similar to what there would be in a national program. Conducting the two studies simultaneously, especially if the demonstrations are implemented in a sample of areas, as discussed in the next section, would greatly increase the realism of the $1-for-$2 program for the study of reduced exit as well as the study of induced entry. People learn and form opinions about social programs from others on those programs, not simply from written and oral explanations by researchers and SSA staff.

A demonstration study of reduced exit from the SSDI rolls would gather detailed data on a sample of beneficiaries (e.g., by surveying them), whether it occurs before or at the same time as the study of induced entry. Aggregate-level information on SSDI participation and program costs for the areas covered by the demonstration would also be collected. The quality of the estimates of the incremental costs of a national $1-for-$2 program are likely to be better if the two demonstration studies are collected simultaneously in a sample of areas.

> Consultants' Conclusions: The greater realism of a demonstration that studies both reduced exit and induced entry simultaneously and jointly needs to be weighed carefully against the potential savings in study costs (including incremental program costs) associated with sequential demonstration studies of reduced exit and induced entry.

SPECIFIC DEMONSTRATION DESIGNS AND THEIR MERITS

SSA's ORES has reviewed in some detail three main types of design involving implementation of a new SSDI program or demonstration, plus several variations of one of them.

The first is a classical experimental design with random assignment to a treatment (i.e., some variation of a new SSDI program) and the control group. As Weathers et al. (2000) explain, a classical experimental design has many well-known advantages in most instances, but it has notable disadvantages for studying induced entry onto a new SSDI program. It is useful to review this design, nonetheless, as it provides a useful baseline against which to compare other designs.

A second design resembles the first, but with one important difference. Random assignment to a treatment or to the control group is limited to the subpopulation of people with serious physical, mental, or health impairments who are not currently SSDI beneficiaries.

Finally, ORES has considered several variations of a third design in which a demonstration SSDI program is mounted in selected geographical areas (e.g., states or counties). Behavior in the areas with the demonstration is then compared to behavior in geographical areas maintaining the existing SSDI program.

1. Classic Experimental Design with Random Assignment

This design would involve random assignment of persons 18-64 years old to an experimental

condition – either a treatment group subject to a new SSDI program, or the control group subject to the existing SSDI program. Using this design to study induced entry into a new SSDI program presents several serious problems (see Weathers et al., 2000, pp. 96-98).

First, Weathers and his co-authors calculated that one would need a sample of 9 million people in each condition (i.e., in each variation of the new SSDI program plus the control group) in order to detect a two percent increase in new SSDI beneficiaries per year with a statistical power of .80.[10] Greater statistical power (e.g., .90) would require a larger number in each experimental condition. Their calculations imply that the total number of people studied in all conditions would be very large. Those studied would inevitably be widely dispersed geographically. It would be difficult and costly to study such a large and geographically-dispersed group of people and to follow them over time.

Second, it would be necessary to explain the new SSDI program to every person in every treatment group in order to believe that different estimates for a treatment and control group are due to a change in the SSDI program. Not only would this require a massive effort at information dissemination, but people's understanding of the SSDI program could be better in a treatment group than in the control group. Consequently, one would need at least two control groups. The existing SSDI program would be explained to one control group and not explained to another control group. The two control groups would allow the effects of being given information about the SSDI program to be disentangled from the effects of the actual changes in the SSDI program. (To control for a possible Hawthorne effect, there might need to be yet a third, "silent" control group that would also not receive information about the existing SSDI program and that would be studied using only administrative records.)

Third, the manner of disseminating information about the demonstration program is likely to differ greatly from that in a national program. The effects of the difference in information dissemination might conceal the real differences between the response to the demonstration program and the response to an actual national program.

> Consultants' Conclusion: A classical experimental design with random assignment to experimental conditions has major disadvantages for studying induced entry resulting from a new SSDI program with $1-for-$2 benefit reduction. In particular, dissemination of information about SSDI programs would differ greatly from that in a national program and would also be very costly. Data collection using a classical experimental design would be also difficult (and therefore very expensive) because the individuals who are studied would be highly dispersed geographically. The consultants think that this design should not be used.

2. Experiment Targeted at Persons with Health Limitations

This design resembles the classical experiment described above. The primary difference is

[10]They calculated that a sample of 1.44 million people in each group would be needed in order to detect a five percent increase in the number of new SSDI beneficiaries with a statistical power of .80.

that the people who are randomly assigned to an experimental condition (treatment or control group) would not be selected from the general population but from a targeted subpopulation: those 18-64 years old who have health limitations but are not currently receiving disability benefits. (For a description and discussion of this design, see Weathers et al., 2000, pp. 106-108.)

Several past surveys have found that roughly 20 percent of the working-age population in the U.S. report that they have some health impairment (see Bound and Burkhauser 1999). About half of these (i.e., about 10 percent of the working-age population) say that their health impairment limits the kind of work that they can do. Viewed differently, 90 percent of the working-age population does not report any work limitation due to a health impairment. Smaller numbers of people would need to be randomly assigned to each experimental condition if one could select them from a high-risk subpopulation.

Although far fewer people would need to be studied in this design than in the first design, the defects of the first and second designs are similar in most other respects. A further practical problem with the second design is that it requires a list of individuals in the targeted, high-risk subpopulation (or a random sample of such individuals). No such list exists. Hence, in order to use this design, one would first need to develop such a list (e.g., by first conducting screening interviews of the general population). In principle, a list could be constructed from responses to the "long form" of the 2000 U.S. Census of Population, which asked about health limitations and was given to 1/6 of U.S. residents in April 2000. However, this option is not feasible because it would violate the U.S. Census Bureau's rules governing confidentiality of census information.

> Consultants' Conclusions: This design has major disadvantages. As in the classic experimental design, dissemination of information about SSDI programs would differ greatly from that in a national program: it would require explaining the program to every participant in the study (except those in certain control groups). Participants would be highly dispersed geographically, making both information dissemination and data collection unusually costly. In addition (and unlike the classic experimental design), it requires a list of people in the targeted, high-risk subpopulation from which study participants could be chosen. No such list exists; developing one, even for a random sample of individuals, would be very difficult and costly. As a result, an experimental design targeted at people with health limitations might be more expensive than the classic experimental design, even though the number of people studied would be much smaller. The consultants agree that this design should not be used.

3. Demonstrations in Geographical Areas

Weathers et al. (2000) consider several versions of a design in which a demonstration (i.e., new SSDI program) is implemented in selected geographical areas; Muller et al. (2001) analyze one version of this design in greater depth. In this type of design, behavior in the areas chosen for the demonstration is compared to behavior in other areas having the existing SSDI program.

An important advantage of implementing a demonstration program in a number of areas is that information about the new program can be disseminated in more cost-effective ways than in the individualized approach required in the first two designs. For example, the new program could

be announced in the mass media (e.g., TV and radio spots, newspaper articles); then detailed written information about it could be mailed to physicians and others in the local community who provide services to people likely to be eligible for the SSDI program. These methods of disseminating program information are not only relatively inexpensive but also much more similar to the methods likely to be used in a national program. Disseminating information in these ways would greatly enhance the realism of the demonstration. Consequently, other things being equal, results will be more credible if based on a demonstration in selected areas than if based on a design that requires explaining SSDI programs to study participants individually.

Another advantage of mounting a demonstration in selected areas is that SSA's administration of the SSDI program in those areas would be much more like that in a national program. It would therefore be more realistic in this respect as well. In addition, mounting a demonstration in selected areas would permit researchers to examine differences (if any) between administration of a SSDI program with $1-for-$2 benefit reduction and administration of the existing SSDI program.

A potential disadvantage must be noted, however. Disseminating information in an area in "typical" ways, while more realistic, is likely to prolong the length of time for people to respond to a new SSDI program. A longer period of data collection would be needed to learn how people adapt to the new program over time and to assess the long-run response to the new program.

A demonstration approach in geographical areas can vary in three main design features. First, one can vary the kind of area where the demonstration is implemented; states and counties are the two main kinds of area that are usually considered. Second, one can use different methods of selecting the areas that are to have (or not have) the demonstration. Areas may be chosen strictly randomly or randomly within certain strata (i.e., categories) defined on the basis of certain attributes (e.g., population size). Third, one can vary the number of areas in which each experimental condition is implemented.

Kind of Area: Comparing counties rather than states has clear advantages, when it is possible to implement a program in a county. Conditions affecting the behavioral response are usually more similar in different counties of a given state than in different states. For example, if entry into a new SSDI program demonstrated in Wyoming differs from entry into the existing program in Delaware, one does not know if the difference is because the new and existing SSDI programs differ or because these two states have different laws, different programs, and different social and economic conditions. If a new SSDI program is demonstrated in several states, and if entry rates are compared in those states and other states having the existing program, one still cannot be sure whether measured differences in the entry rates are due to program differences or due to other differences between the two sets of states.

Socio-economic conditions are likely to be fairly similar when one compares counties within the same state. In addition, state laws, state programs, and state systems of taxation are the same in different counties within a given state. Differences in entry rates into the new SSDI program and the existing program are more likely to be due to the change in the SSDI program itself than

to differences between the areas if one can compare counties within the same state.[11]

Selecting counties rather than states as the areas also has potential disadvantages. There are difficulties in implementing one program for residents of one county and another program for residents of an adjacent county, especially in densely-populated counties that belong to a large urban system. In such areas, both mass media and those providing services to people with disabilities often communicate with residents of more than one county. The general public may find it hard to comprehend that one version of SSDI applies in one county and another version in a neighboring county.

If people do understand the differences in the SSDI programs in neighboring counties, some people with disabilities may migrate from one county to another to take advantage of their preferred version if it is offered in a nearby area but not their present area (see Weathers et al., 2000, pp. 94-95).

The migration problem could be mitigated if eligibility for the new SSDI program is based on a person's place of residence when the new program starts rather than on his or her place of residence at the time of application for SSDI benefits. However, this limitation, while solving the migration problem, would be an additional difference from the existing SSDI program. It also would not avoid confusion resulting from mass media and service providers saying that different versions of the SSDI program apply in different counties. The consequences of basing eligibility on place of residence when the study starts might be small, but one cannot know for sure a priori.

Another way to deal with the mobility problem (but not the confusion problem) would be to collect data on place of residence when the demonstration starts as well as on place of residence at the time of applying to SSDI, and then to investigate to what extent people with disabilities move among counties with different versions of the SSDI program. Ideally, the rate of intercounty moves of disabled applicants to the SSDI programs would be compared to those for the general population (e.g., by using CPS data on the latter).

Stratifying Areas: There are also clear advantages to stratifying areas on characteristics likely to affect the behavioral response to the new program and then randomly assigning the experimental treatment (i.e., the new program) and the control situation (i.e., the existing program) to areas within a given stratum. For example, one might stratify counties on the basis of the state that they are in and their population size. Stratifying helps to improve the precision of estimates without increasing the sample size. Equivalently, it allows one to reduce the sample size in an experimental condition while maintaining a given level of precision in the estimates. Random assignment of the treatment and control group to areas within certain strata is clearly superior to random assignment without stratification in terms of statistical power.

---

[11]The Congressional Budget Office proposed that results of a demonstration of the new SSDI program in five small states be compared with results in five other small states having the existing SSDI program (see Weathers et al., 2000, pp. 104-106). This design suffers from a variety of defects. It is based on a comparison of states rather than counties (which has the difficulties noted above), requires the assumption that behavior in large states would resemble that in small states, and employs too few areas to yield reliable results. The consultants concur with ORES that findings from this design would not be reliable.

Number of Areas: Finally, designs vary in the number of areas in which a particular experimental condition (i.e., a specific new program) is studied. Increasing the number of areas tends to increase both the cost of the study and the reliability of the results. The bottom-line issue is how much one is willing to pay to achieve results that can be trusted.

One could mount a demonstration of a new program in one area and compare it with the existing program in another area. There is likely to be skepticism about the generalizability of the results, whatever they turn out to be. Various random events could have occurred in either area to make the results different from what they would be "normally." If one compares results for two sets of areas, with and without the new program, confidence in the overall results grows.

Weathers et al. (2000) and Muller et al. (2001) used pseudo-random (so-called "Monte Carlo simulation") techniques to gauge how the number of states or counties in a specific experimental condition affects the estimates of induced entry. They examined a great many variations in the type of area, the method of selecting areas, and the number of areas. They concluded that, of the variations they analyzed, the best one involves demonstrating a given experimental condition in 200 counties having a population over 25,000. (The control groups would be chosen from the counties in the same size category that were not assigned to an experimental program.)

Muller et al. (2001) estimated that the total population of 200 counties having a population over 25,000 would be about 34 million people, or about 12 percent of the U.S. population. Obviously this is a huge number of people to make eligible for a new SSDI program. However, researchers would not provide detailed explanations of the new SSDI program to everyone in an area (but would presumably use the mass media to give brief general information, as mentioned above) and also would not interview everyone in an area. Detailed written information about the relevant SSDI program(s) would be given to service providers and would be available in local offices of the SSA. Administrative records (possibly extended to give somewhat more detail than is standard) on all applicants and all awardees in an area would be analyzed. In addition, some subset of the applicants and awardees would probably be surveyed to try to understand both the factors that led to their application and the award of SSDI benefits to them (or not). A further subset of those awarded SSDI benefits would probably be surveyed over time in more detail to learn how they behave while on the SSDI program. A very important aspect of awardees' behavior would be whether and when they return to work and, if they do, how many hours they work, what kind of work they do, and how much they earn.

If eligibility for the new program is limited to those who apply for SSDI after the demonstration starts and before some termination date, *and* if entry rates are the same as under the existing SSDI program, the expected number entering the demonstration SSDI program in 200 counties with a population over 25,000 would be about 74,400 people per year or 6,200 per month. *If* the number entering under the new program rises by five percent, the expected number of induced entrants in the demonstration would be about 3,720 more people per year or 310 more per month.[12] In a five-year study in which applicants to the demonstration program can apply only

[12]This estimate assumes that only a single set of program parameters is studied. The number would go up in a study of several variations (e.g., different levels of disregarded earnings).

during the first 48 months but are followed until the end of the sixtieth month (to see if SSDI beneficiaries return to work after they enter the SSDI program),[13] the incremental cost of induced entry into the demonstration program would be roughly $353 million,[14] if average SSDI benefits do not change. These estimated costs exclude the cost of administering the study, collecting data, and analyzing them; they also do not consider the extra cost of Medicare benefits for the induced entrants. If the number of induced entrants should rise by only two percent, the incremental costs would be 40 percent as great (i.e., roughly $141 million). Similarly, if the number of induced entrants should rise by ten percent, the incremental costs would be twice as great (i.e., about $706 million). It should be kept in mind that these are incremental program costs for the five-year period of the demonstration and do not include Medicare costs, administrative costs, study costs, or any costs incurred in fulfilling promises to participants about their future benefits. Further, a demonstration of this size would severely tax and possibly exceed SSA's administrative capacities.

For induced entry rates in this range, the incremental program costs would unquestionably be very large. However, the incremental cost of a demonstration would be small in comparison to the extra $4 billion in five years that SSA's OCACT judged might result from induced entry under a national SSDI program with a $1-for-$2 benefit reduction. Policy-makers will have different opinions about whether it is worthwhile to study induced entry at this potential incremental cost. The assessment of the value of such a study will depend on one's prior assessment of the likely level of induced entry, one's prior evaluation of the potential benefits to SSDI beneficiaries of reducing beneficiaries' benefits by $1 for each $2 in earnings above some specified level, and one's willingness to change the national SSDI program, despite the potential risk in misjudging its long-term consequences.

> Consultants' Conclusion: If a demonstration study of induced entry is mounted, a design in which roughly 200 counties having populations exceeding 25,000 are in each treatment group appears to be better than any alternative design based on states or counties that has yet been considered. Despite its advantages relative to other designs, this design still has important defects:
>
> > 1) A large number of counties comprising a sizable fraction of the U.S. population would need to be in the study in order to detect and measure induced entry reliably. With even a small increase in the number of entrants (e.g., two

[13]A study shorter than five years would not give good estimates of the long-run behavioral response.

[14]If, during each of the 48 months of intake, 310 more persons enter the demonstration program than the existing SSDI program, there would be 364,560 (= 310*49*48/2) more person-months of benefits awarded during the 48 months when new awards would be made. Assuming every applicant awarded SSDI benefits enters the SSDI program immediately after the five-month waiting period, and assuming everyone remains on the SSDI program until the study ends, 14,880 more people (310 people * 48 months of intake) would be receiving SSDI benefits during the final seven months of the demonstration program. This leads to an additional 104,160 person-months of benefits (14,880 people for 7 months). Assuming the average benefit in the demonstration remains at $754/month yields an estimated incremental program cost for the demonstration of $353 million. This does not include Medicare costs. Given an estimate of the monthly Medicare cost, one could estimate that, too, by multiplying it by 468,720 (364,560+104,160).

percent), the incremental program costs would be huge, about $70 million per year of a five-year study.

2) This design raises serious problems concerning the dissemination of information about the demonstration and existing SSDI programs to residents of adjacent counties, and the potential mobility of some people with disabilities to gain access to a preferred version of the SSDI program.

The consultants are uncertain whether such a design, despite its advantages relative to other designs, and despite its potentially huge costs, would yield estimates of induced entry that would be sufficiently accurate and reliable to meet policy-makers' needs.

CONCLUSION

This paper focuses on the study of a change in the Social Security Disability Insurance (SSDI) program in which benefits are reduced by $1 for each $2 in a beneficiary's earnings above a "disregard" level set by the Social Security Commissioner; it is commonly called a "$1-for-$2" benefit reduction. Section 302 of P.L. 106-170 mandates that the Social Security Administration (SSA) study how a $1-for-$2 benefit reduction would impact the cost of the SSDI program through its effect on people's behavior. In particular, Section 302 mandates that there be a demonstration study whose results could be reliably generalized to a national $1-for-$2 SSDI program.

A $1-for-$2 benefit reduction could lead to "reduced exit" from the SSDI rolls if some beneficiaries work and earn more but stay on the SSDI program longer while receiving partial benefits. It could also foster "induced entry" because insured workers with serious impairments who are not presently on SSDI might be motivated to apply for SSDI benefits under the new program. The present paper pays primary attention to the impact of a $1-for-$2 benefit reduction on induced entry into the SSDI program. The possibility of linking a demonstration study of induced entry to a demonstration study of current SSDI beneficiaries is also briefly considered.

This paper describes five main approaches that could be used to estimate the impact of a program change on participation in the program, behavior of program participants, and program costs. Estimates may be based on: 1) the aggregate response to other program changes in the past, 2) a dynamic model of individual behavior using results of previous studies, 3) individuals' responses to a survey with hypothetical questions, 4) individuals' responses to a demonstration (field experiment), and 5) implementation of a national program coupled with a study of people's responses to it. The five types are listed in ascending order of the probable study costs.

SSA's OCACT has already produced estimates using the first approach (see McLaughlin 1994). ORES's team of consultants believes that valuable information on the impact of induced entry may be acquired through the second approach (dynamic modeling of individual behavior) and the third (responses to hypothetical questions in a survey) at a relatively modest cost. While these methods pose some issues and must be carefully interpreted, they have been used with some success to assess behavioral impacts in a variety of other contexts (e.g., welfare reform). The team of consultants thinks that SSA should continue to pursue such methods to assess the impact of

induced entry, even if it is decided to study induced entry through a demonstration project.

Even the fifth approach – implementing a national program – presents problems, quite aside from its cost (which could be large enough to threaten the Social Security Trust Fund). In particular, there would be no contemporaneous control group. Consequently, it would not be possible to say whether changes in program costs and in SSDI beneficiaries' behavior result from the $1-for-$2 program or from other socio-economic changes. Except for the aforementioned issues and its potential cost, this approach would, however, solve virtually all other problems.

The rest of the paper deals with the fourth approach, a demonstration, because Section 302 mandates that this type of study be used. To set the stage for a review of specific demonstration designs, seven general design issues are first discussed: 1) the similarity of the demonstration program to a new national program, 2) scale effects, which arise because a demonstration does not cover every place and may not even cover everyone in a given place, 3) Hawthorne effects, which occur because people often act differently simply because they are being monitored, 4) duration and timing effects, which result because data collection for a demonstration covers a relatively brief period of time while people are still adjusting to the program change and further because data collection may occur under atypical socio-economic conditions, 5) the size of a tolerable change, 6) various statistical issues, and 7) the relationship of a demonstration study of induced entry to a demonstration study of reduced exit. These issues, except for the seventh, are general ones relevant to demonstration and other empirical studies of changes in a wide variety of programs.

The challenges and dilemmas presented by each of the seven issues have been discussed above as concisely as possible and cannot be summarized again here. In the case of almost every one of these issues, those planning a demonstration must make choices that affect both study costs and the quality of the results. Not surprisingly, there is usually a trade-off between quality and cost: the higher the quality, the greater the cost. But there are other important trade-offs. A choice that helps to solve one design problem sometimes amplifies another design problem. Indeed, design questions rarely have easy answers, even if the costs of the study are ignored.

Finally, the paper evaluates three basic types of demonstration designs that SSA's ORES has analyzed in some detail: 1) a classic experiment with random assignment of individuals to specific SSDI programs, 2) a modified classic experiment with random assignment of individuals with physical, mental, or health impairments to specific SSDI programs, and 3) demonstrations in randomly-selected areas (e.g., states or counties) that may be stratified on certain characteristics.

The first and second types of demonstrations have very serious defects. Both lead to unmanageable problems in implementing a demonstration that would disseminate program information adequately and in a manner sufficiently similar to a national program that estimates based on the study could be reliably generalized to a national program. The second type presents an additional problem. It requires prior development of a list of individuals with physical, mental, or health impairments from which experimental subjects could be selected. Creation of such a list would be very difficult and expensive. Consequently, ORES and the team of consultants agree that the first two types of designs should not be used to study induced entry.

ORES has evaluated several variations of the third type of demonstration design. It presents

problems but cannot be dismissed as easily as the previous two types of demonstration design. Based on its analyses of various instances of the third type of design, ORES has reached a number of conclusions. The team of consultants agrees with most of these,[15] in particular:

a) It would be better to choose *counties* (or other small areas) rather than states to take part in the demonstration. This choice has some disadvantages: It would exacerbate problems in disseminating information about different SSDI programs in nearby sites, especially in densely-populated urban areas. It would also generate somewhat higher levels of mobility between sites with different SSDI programs. But there are important counterbalancing advantages: It would allow more areas to be studied, permit a greater diversity of local environments and of states, and permit states' characteristics (e.g., laws, programs) to be controlled.

b) The *number of areas* needed to assess whether a $1-for-$2 program leads to a tolerable increase in induced entrants (tolerable in terms of increased program cost) is likely to be large, and the population of these areas is likely to be very large. The incremental program and study costs could be huge.

c) *Stratifying* areas on the basis of key characteristics would yield somewhat more precise estimates at no additional cost.

d) Given the large numbers of areas and people involved in a demonstration study of induced entry, and, given that some key design problems may not have good solutions, *a demonstration study of induced entry may not be worth its cost. Ignoring its cost, it is, however, superior to the alternatives considered to date.*

If a demonstration study of induced entry into the SSDI program (in particular, one mounted in randomly-chosen areas) proceeds, attention needs to be given to many other important issues, a few of which are summarized in Appendix C. To satisfy the mandate of Section 302 of P.L. 106-170, these other issues need to be resolved because they are not independent of the design issues discussed in this paper and because they will also affect study costs.

[15]Some of the consultants are uneasy with the necessary sample sizes that ORES calculated. Based on their experience, ORES's estimated sample sizes for some specific designs seem too large. There is no disagreement, however, that the sampled number of counties would need to be large, that their populations would be very large, and that the cost of a demonstration study of induced entry could be huge.

APPENDIX A
The National Study of Health and Activity

The National Study of Health and Activity (NSHA), sponsored by the Social Security Administration, will conduct a survey designed to give information on the health, health impairments, and disabling conditions of adults living in the U.S. The data collected in this survey will furnish estimates of the number of adults living in the U.S. who are currently disabled or who have health limitations that may cause them to become disabled in the foreseeable future. It will provide information on how the disabilities and health limitations of U.S. adults affect their activities at work and at home, and in their community. It will also give information on people's access to health care. All of this information will improve SSA's ability to answer pressing policy questions about the nature and extent of disability in the U.S. and to plan for the future, particularly concerning the SSDI program.

The NSHA will initially screen about 100,000 households. From this screening, researchers will select for further study a sample of about 5,500 individuals who fall into one of four categories: (1) severely impaired and a Social Security beneficiary (under either the SSDI or SSI program); (2) severely impaired but not a Social Security beneficiary; (3) moderately impaired; and (4) not impaired. Individuals in this sample of 5,500 people will be given a second, comprehensive interview. They will also be asked to participate in a medical examination and some tests of their physical and mental functioning. A team of medical professionals will conduct the medical exams in a Mobile Examination Center (MEC), designed specifically for the NSHA.

NSHA will combine self-reported measures of health limitations and impairments from the comprehensive survey, the NSHA medical examination given in the MEC, and medical records from the immediately preceding three years. The combined information will be used to create hypothetical SSA disability folders for each person in the study who is given the medical examination. Regular disability examiners will then be asked to make disability decisions based on the medical information in these folders. Those individuals whom the examiners find medically eligible for social security benefits (ignoring work and other nonmedical reasons for disqualification) will be considered "disabled" for Social Security purposes. These individuals comprise the pool of respondents to the NSHA who would be eligible for SSDI benefits if they applied for SSA disability benefits and were also qualified on nonmedical grounds.

Just as important as the number in the second category (i.e., those who are currently eligible for Social Security benefits but not a current beneficiary) is the number who are "in the pipeline," who might qualify for disability benefits at some future point due to a progressive condition (e.g., ALS, MS). These individuals are on the threshold of a severe disability; they are likely to be medically eligible for disability benefits in the foreseeable future because they have a progressive degenerative disease or for other condition leading their health to deteriorate. NSHA will estimate the number of these borderline cases using the MEC physician's prognosis, together with an actuarial projection based on the person's health information.

APPENDIX B
A Demonstration Study of Reduced Exit

An important aim of a new SSDI program with a $1-for-$2 benefit reduction is to promote working by SSDI beneficiaries. When people with disabilities work and have earnings, it tends to enhance both their psychological and economic well-being.

Another key policy goal is to minimize, or at least to contain, the costs of the SSDI program. Under the existing SSDI program, beneficiaries may work and have any level of earnings for nine months of "trial work." If, after 36 months or later, they have earnings above the SGA level, they are terminated from the SSDI rolls. In contrast, with a $1-for-$2 benefit reduction, many SSDI beneficiaries could work indefinitely above the SGA level and still receive partial SSDI benefits. Consequently, exit from the SSDI rolls might be lower under the $1-for-$2 program than under the existing program. This phenomenon is referred to as "reduced exit" from the SSDI program.

At most two percent of the 0.43 million people leaving the SSDI rolls in 1999 (i.e., 8,600 people) may have left because their earnings exceeded the SGA level 36 months or more after their trial work period ended (see Table 6.F2). This number is very small considering the size of the U.S. population, or even the total number of SSDI beneficiaries. But costs could rise substantially if even this many were on the SSDI rolls much longer under the $1-for-$2 program than currently.

It is not known to what extent SSDI program costs would increase because of "reduced exit." Consequently, Section 302 of P.L. 106-170 mandates that there be demonstration projects to determine "... the effects, if any, ... of reduced exit from the project" as well as "... the effects, if any, of induced entry" into it. A study of "reduced exit" usually focuses on current program participants, which tends to be easier and cheaper (though still hard and expensive) than a study of induced entrants.

Of the seven basic issues in a demonstration study of induced entry, the first six of these apply as well to the study of reduced exit from the SSDI rolls and of beneficiaries' work behavior. A satisfactory solution to these six issues sometimes takes a slightly different form, but to a large extent, the solutions are similar and are therefore not reviewed again here.

A demonstration study of "reduced exit" involves drawing a sample of SSDI beneficiaries and assessing their behavior over time. Sample sizes that yield reliable estimates have been judged to be much smaller for a demonstration study of reduced exit than for a demonstration study of induced entry. ORES has been considering drawing samples of about 5,000 SSDI beneficiaries for each treatment condition in a demonstration study of reduced exit and of beneficiaries' work behavior. Further, it has discussed stratifying this sample on the basis of a beneficiary's duration on the SSDI rolls and, in particular, concentrating this empirical study on those who have been on the rolls a comparatively short time, under the assumption that those who have been on the SSDI rolls for many years are unlikely to return to work. ORES's team of consultants thinks that this kind of stratification is sensible but has not yet formed an opinion about the details. The team of has also not yet formed an opinion about the necessary sample sizes for a demonstration study of reduced exit. Too many other relevant factors remain unspecified to draw conclusions about this point. The consultants will be giving the study of reduced exit greater attention in the future.

APPENDIX C
Additional Issues

A number of other issues that have not yet received much attention need to be resolved before starting a demonstration study of a $1-for-$2 benefit reduction. They include:

1) What kind of data needs to be gathered on everyone in the study sites? Since the number of people eligible for the program will be very large, it seems likely that only a (possibly extended) version of SSA's administrative records could feasibly be collected on everyone in the study sites who are in Social Security's data base.

2) Since information more detailed than that in administrative records will be needed, what subpopulations in the study sites should be studied in more detail? How will they be selected?

3) What should be the size of samples of the subpopulations studied in detail?

4) What information should be collected on the people in these subsamples? To understand the factors prompting induced entrants (and current beneficiaries) to return to work as well as those factors affecting their work efforts and earnings, some people who enter the SSDI program will need to be surveyed. What topics should be covered in the survey? For comparative purposes, it may also be desirable to survey some individuals who are not on the SSDI program, or who have applied to the SSDI program but not been awarded benefits.

5) How frequently will those chosen for detailed study be surveyed? Periodic surveys will almost surely be needed because return to work involves a prolonged and complex process of deciding to seek work, searching for work, considering alternatives, starting to work, and possibly then leaving the work force or switching to another job.

6) Will individuals other than the beneficiaries themselves be surveyed? Family members and others who support people with disabilities, or provide assistance to them, might provide useful information in understanding the work behavior of people with disabilities. It might also be useful to survey some of their employers to learn why and how employers try to accommodate the work limitations of those workers.

REFERENCES

Bound, John, and Richard V. Burkhauser. 1999. "Economic Analysis of Transfer Programs Targeted on People with Disabilities." Pages 3417-3528 in *Handbook of Labor Economics*, Volume 3C, edited by Orley C. Ashenfelter and David Card. Amsterdam: North Holland Press.

Brehm, Henry P., and Thomas V. Rush. 1988. "Disability Analysis of Longitudinal Health Data: Policy Implications for Social Security Disability Insurance." *Journal of Aging Studies* 2 (4): 379-399.

McLaughlin, James R. June 13, 1994. "Estimated Increase in OASDI Benefit Payments That Would Result from Two 'Earnings Test' Type Alternatives to the Current Criteria for Cessation of Disability Benefits—Information." Memorandum, SSA Office of the Actuary.

Moffitt, Robert. 1992. "Evaluation Methods for Program Entry Effects." Pp. 231-252 in *Evaluating Welfare and Training Programs*, edited by Charles F. Manski and Irwin Garfinkel. Cambridge, MA: Harvard University Press.

Muller, L. Scott, Robert Weathers, John Hennessey, and Fred Bellemore. June 12, 2001. "Induced Entry Demonstration Designs. $1 for $2 Benefit Offset: Exploring the Feasibility of Measuring Induced Entry and Evaluating Potential County-Level Project Designs." 23 pages.

Roethlisberger, Fritz K., and Dickson, William J. (1939) 1961. *Management and the Worker: An Account of a Research Program Conducted by the Western Electric Company, Hawthorne Works, Chicago*. Cambridge, MA: Harvard University Press.

Social Security Administration. 2000. *Annual Statistical Supplement to the Social Security Bulletin*. Washington, DC: Office of Research, Evaluation, and Statistics, Social Security Administration, U.S. Department of Health and Human Services.

Weathers, Robert, L. Scott Muller, and John Hennessey. September 14, 2000. "Induced Entry Demonstration Designs. $1 for $2 Benefit Offset: Exploring the Feasibility of Measuring Induced Entry and Evaluating Potential State-Level Project Designs." 72 pages.