TECHNICAL REPORT

Research Designs for Estimating Induced Entry into the SSDI Program Resulting from a Benefit Offset

Nicole Maestas • Kathleen J. Mullen • Gema Zamarro

Sponsored by the Social Security Administration

RAND LABOR AND POPULATION

This research was funded by a contract from the Social Security Administration and was undertaken within RAND Labor and Population.

Published 2010 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

To support the Social Security Administration (SSA) in fulfilling its legislative mandate under the Ticket to Work Incentive and Work Incentives Improvement Act (Ticket Act), this report has the principal aim of providing SSA with a set of research design options for estimating induced entry effects of a proposed $1-for-$2 benefit offset for its Social Security Disability Insurance (SSDI) program. Although the Ticket Act included induced entry effects among the set of effects to be evaluated with a demonstration project, SSA has determined that a demonstration project aimed at estimating induced entry effects is not feasible. Having determined that a demonstration project would not produce credible estimates of induced entry at reasonable cost, SSA must now determine an alternative method for fulfilling its mandate under the Ticket Act. In service of that goal, this report is designed to provide SSA with two carefully selected research design options and the information needed to evaluate each design against several important criteria.

This research was funded by a contract from SSA. The opinions expressed and conclusions drawn in this report are the responsibility of the authors and do not represent the official views of SSA, other agencies, or the RAND Corporation.

This research was undertaken within RAND Labor and Population. RAND Labor and Population has built an international reputation for conducting objective, high-quality, empirical research to support and improve policies and organizations around the world. Its work focuses on labor markets, social welfare policy, demographic behavior, immigration, international development, and issues related to aging and retirement with a common aim of understanding how policy and social and economic forces affect individual decisionmaking and the well-being of children, adults, and families.

For more information on RAND Labor and Population, contact Arie Kapteyn, Director, RAND Labor and Population, RAND Corporation, 1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138, (310) 393-0411 x7973, Arie_Kapteyn@rand.org.

Contents

Figures and Tables

Figures

Tables

Summary

To support the Social Security Administration (SSA) in fulfilling its legislative mandate under the Ticket to Work Incentive and Work Incentives Improvement Act (P.L. 106-170; the Ticket Act), this report has the principal aim of providing SSA with a set of research design options for estimating induced entry effects of a proposed \$1-for-\$2 benefit offset for its Social Security Disability Insurance (SSDI) program. Whereas under current program rules, SSDI beneficiaries who have completed their Trial Work Period (TWP) and who earn more than the threshold for Substantial Gainful Activity (SGA)—currently set at \$1,000 per month—are ineligible to receive benefits (that is, they lose their benefit entirely), under the proposed benefit offset policy these individuals would retain \$1 in benefits for every \$2 earned above the SGA. Thus, a benefit offset may induce entry because it makes participation more attractive for individuals who are medically eligible for SSDI benefits but able to earn more than the SGA. The size of the population of induced entrants is a critical input into an analysis of the effects of a benefit offset policy on overall program costs.

Although the Ticket Act included induced entry effects among the set of effects to be evaluated with a demonstration project, SSA has determined that a demonstration project aimed at estimating induced entry is not feasible (Tuma, 2001). As a result, SSA must now determine an alternative method of fulfilling its mandate under the Ticket Act. In service of that goal, this report is designed to provide SSA with two carefully selected research design options to estimate induced entry under the proposed benefit offset policy, as well as the information needed to evaluate each design on several dimensions: internal validity, external validity, flexibility, economy (cost), and speed.

To accomplish these objectives, we first performed an extensive literature search and prepared a list of candidate research designs. In January 2010, we convened a meeting with SSA and our Technical Advisory Group (TAG), consisting of experts on the SSDI program and research methods for estimating entry effects, where we presented this list of candidate designs. After consultation with the TAG and SSA stakeholders in attendance, two research designs were identified as the most promising:

- a research design using **stated preferences (SP)**
- a research design using **past policy (PP) changes** in a simple structural framework.

In the remainder of this summary, we highlight key findings with respect to the study objectives.

Theoretical Framework for Induced Entry

One important contribution of this report is to develop a simple theoretical framework for understanding the mechanisms that give rise to induced entry. This framework provides a rigorous yet intuitive starting point for an analysis of induced entry. The model is dynamic, forward-looking, and yields steady-state conditions for SSDI claiming and employment behavior. The model identifies several factors, such as health insurance, that figure into the decision to apply for SSDI benefits. More importantly, the model identifies a group of disabled, nonbeneficiary workers who would be better off claiming SSDI under the benefit offset policy but not under the current policy, and thus could be induced to enter the SSDI program. This group consists of individuals with earnings in a particular range defined by the SGA threshold, the benefit offset rate, and the individual's monthly benefit amount. This condition plays an important role in defining the sample frame for the research design using stated preferences.

Research Design Using Stated Preferences

A promising method for estimating potential entry effects as a result of a \$1-for-\$2 benefit offset is a research design using the stated preferences method. In this method, one administers a series of stated choice experiments designed to reveal respondents' preferences for claiming disability under varying program rules and economic conditions. In particular, the SP approach consists of presenting respondents with a set of scenarios describing different states of the world and asking them to rate, rank, or choose among different possible actions (e.g., continuing to work versus claiming disability under varying conditions). The scenarios are characterized by either real or hypothetical attributes (or a mix of both), such as a benefit offset rate or earnings disregard level, and allow one to estimate the impact of a *hypothetical* policy that has never been experienced by respondents.

Sampling Plan

An important feature of the SP design is that it requires new data collection. Because the target population—potential induced entrants—has unknown characteristics and is likely small relative to the general population, a critical issue is how to sample and screen respondents. As noted above, economic theory offers a useful guide for winnowing down the sampling frame to individuals in a particular range of earnings, who are most likely to make up the target population. We identified two potential sampling frames as promising candidates for an SP-based research design:

- We identified the **SSA administrative database** of U.S. workers as an ideal sampling frame, since it includes every worker insured for SSDI benefits in the United States and their history of earnings and benefit receipt. Since health information is not available in SSA's administrative data, individuals cannot be sampled on the basis of their likelihood of medically qualifying for the SSDI program. A health screener, such as the 26-item screener developed by Westat in 2002 for the National Study of Health and Activity (NSHA), could identify respondents with health conditions that may qualify them for the program. However, because the medically eligible population is very small, at most one-quarter of those screened would likely be included in the final survey.

- We also identified the **American Community Survey (ACS)** as a potential sampling frame. The ACS surveys roughly 1.5 million individuals ages 25–64 per year. Linking the ACS to SSA administrative data would allow one to narrow the sample frame to individuals with earnings in the appropriate range who are eligible for but not receiving SSDI benefits. The advantage of the ACS is that it already includes six questions on disability that can be used to pre-screen respondents. An additional health screener, such as the Westat screener, could be used to further refine the sample. However, additional research is needed to determine the fraction of individuals likely to pass the screen.

Experimental Design

As noted above, the SP design consists of asking respondents to imagine their behavior under a series of hypothetical scenarios. We identified three variations on the SP design that could be used to estimate induced entry:

- The **baseline** approach is the simplest approach, designed to yield an estimate of induced entry under a $1-for-$2 benefit offset in an otherwise unchanged program environment. This approach consists of describing the benefit offset to currently disabled nonbeneficiaries and asking whether they would apply for SSDI benefits under the new policy. This approach is by far the most expensive to implement, on the order of $2.1 million, assuming a sample frame based on SSA administrative data (excluding pilot testing and other survey-design activities). Moreover, the baseline design does not offer any flexibility to estimate responses to variations of the benefit offset policy.
- A **baseline plus** approach goes a step further and specifies a statistical model for SSDI claiming as a function of proposed program parameters (e.g., offset rate, disregard level) and current program parameters (e.g., the SGA level), known as attributes. This design is extremely flexible and reduces costs substantially by imposing modest structure on the estimation problem with few additional assumptions. Additionally, it is possible to conduct *randomized* choice experiments by randomly varying hypothetical attributes over respondents, which maximizes statistical power by setting the correlation between attributes to zero. If respondents each rate several profiles (scenarios consisting of different attributes), then sample size can be further reduced. Estimated implementation costs for a baseline plus design varying two attributes, each with 3–4 levels, and asking respondents to rate 6–12 profiles range between $381,000 and $632,000.
- Finally, we propose an **alternative** method that achieves cost savings by eliminating the need to screen out 75 percent of the sample based on health. We do so by recasting health itself as an attribute to be specified explicitly in the hypothetical scenarios presented to respondents. Introducing health as an attribute has the added advantage of allowing one to control for health in a uniform way by designing scenarios specifically based on SSDI medical eligibility criteria. Estimated

implementation costs for an alternative design asking respondents to rate 5–15 profiles range between $204,500 and $354,000.

Summary Evaluation

The SP approach is extremely flexible and allows program parameters and other attributes of the scenarios to be varied easily. In addition, it requires very few assumptions regarding specification of the individual's decision environment. The key assumption for identification is that respondents are able to accurately forecast their behavior under new and unfamiliar policy conditions. In addition, using health as an attribute of choice scenarios and surveying individuals who are not medically eligible requires the additional assumption that individuals can forecast their behavior under different *health* conditions. Although new data collection is costly, SSA administrative records and the ACS provide inexpensive yet comprehensive sample frames. In addition, presenting respondents with multiple, randomized scenarios leads to impressive reductions in sample size without sacrificing statistical power.

Research Design Using Past Policy Changes in a Simple Structural Approach

This research design leverages past changes in the SGA threshold to estimate key behavioral parameters that could be used to forecast entry behavior. The SGA threshold is a fundamental program parameter, determining both initial eligibility and ongoing entitlement to SSDI benefits, and it figures directly into the current work rules. These past policy changes are closely related to the introduction of a proposed benefit offset, as they both modify the shape of the budget constraint that potential entrants face. This research design is composed of two parts: (1) a reduced form analysis of the impact of SGA changes on SSDI applications/enrollment, which provides a potential test of whether one might expect any induced entry under a benefit offset, and (2) a simple structural analysis, which relates the reduced form estimates to induced entry under a specific $1-for-$2 benefit offset or a range of offset policies.

Reduced Form Analysis

There is significant variation in real SGA levels over time, including increases, decreases, and periods of relative stability. SSA has increased the (nominal) SGA threshold several times in past decades, including large increases in 1990 and 1999. At the same time, inflation has led to real declines in the SGA before and in between these increases. Since December 2000, the SGA threshold has been indexed to a measure of annual average wages for all employees in the United States. In addition, the SGA level is relatively more generous in areas with lower costs of living and/or lower average wages. Therefore, there is considerable variation in the SGA level across time and *space* (e.g., states or counties) when considered in relative terms. One can construct a measure of real, relative SGA levels by dividing the Consumer Price Index–adjusted national SGA level by a state- or county-level index of average wages. Using SSA administrative data, one can then regress SSDI application and/or enrollment rates at the state-year level on real, relative SGA levels along with controls for changes in macroeconomic conditions (i.e., state and year fixed effects, and such variables as state-level unemployment rates). Since the SGA threshold is such a fundamental program parameter—

affecting both entry and ongoing entitlement—failure to detect an effect of past SGA changes on SSDI entry might lead one to expect little or no change in entry in response to a proposed benefit offset.

Structural Analysis

While raising the SGA threshold is not *equivalent* to introducing a benefit offset, both policy changes affect approximately the same area of the budget constraint faced by potential entrants. As a result, one could relate the SGA-induced entry effect to the benefit offset setting using a simple structural framework. Specifically, one could specify a utility maximization problem where individuals jointly determine their labor supply and SSDI program participation. Once one assumes a functional form for labor supply (or, equivalently, the indirect utility function), specifies a role for observable individual characteristics, and assumes a distribution for unobservables, the model could be estimated using maximum likelihood or method of moments. Once one has obtained estimates of the utility function parameters, one could apply them to the hypothetical budget constraint under the proposed benefit offset to simulate who would apply for SSDI under the new program. An estimate of induced entry can then be obtained by subtracting the number of applicants under the current policy from those under the proposed benefit offset policy. In this framework one could make use of three sources of identification:

- The **discontinuity** in the budget constraint arising from the presence of the SGA threshold—in principle, past policy *changes* are not necessary to identify the parameters of the model, as revealed preference under a nonlinear budget set in cross-section is sufficient to identify the model, with certain assumptions (Moffitt, 1990). One assumption is that observed nonlabor income is exogenous.
- Bringing sufficiently large **SGA changes** (across time or space) into the analysis allows one to relax assumptions about the income elasticity of program participation with respect to benefits. Intuitively, individuals who previously earned more than the old SGA but less than the new SGA experience a local outward shift in their budget constraint, as they are now eligible for SSDI benefits. The receipt of SSDI benefits increases their total net income (earnings plus benefits) without affecting their net wage rate (the amount they can keep as income if they work an additional hour). This allows one to identify the income elasticity without additional assumptions about nonlabor income. Incorporating SGA changes allows one to explicitly link the proposed reduced form and structural analyses. Specifically, one could estimate the model using a method of moments strategy and include the estimated reduced form effect as a moment to be matched.
- Finally, the induced entry project is only part of a portfolio of projects funded by SSA to estimate potential impacts of a benefit offset. Another such project is the **Benefit Offset National Demonstration** (**BOND**) project, which is investigating the effect of the benefit offset on the labor supply of current beneficiaries. The BOND gives one the opportunity to observe actual responses to the exact change in budget constraint under the benefit offset. A critical drawback of the BOND for the purpose of estimating induced entry is that it only provides information on labor supply; that

is, by design, it does not include information on the *participation* decision. However, the BOND does provide a valuable opportunity to incorporate additional variation into the estimation (e.g., by including an additional moment to be matched) or to test the ability of the model to make out-of-sample predictions.

Summary Evaluation

This research design is fast and inexpensive. It utilizes data on the group most likely to approximate marginal entrants under a benefit offset, actual SSDI applicants. In addition, it provides a fair amount of flexibility in that it would be easy to modify the budget constraint in the decision problem used to simulate behavior under hypothetical rules. A drawback of the approach is that it relies heavily on distributional and functional form assumptions. However, opportunities abound for testing and relaxing some of these assumptions by exploiting the "natural experiments" arising from past SGA changes as well as an actual randomized experiment, the BOND project.

Both research designs were determined to be capable of providing SSA with credible estimates of induced entry into SSDI resulting from a benefit offset with relatively small sets of assumptions. In addition, both approaches offer a great deal of flexibility and allow for a range of estimates that would provide valuable insight into how potential SSDI applicants make decisions regarding program participation. While both research designs produce partial-equilibrium "steady-state" estimates of induced entry, they yield parameter estimates that could be used to forecast entry over time, accounting for changing economic and demographic conditions (e.g., trends in population aging, health, and labor demand). In a head-to-head comparison, there is no clear winner, as both research designs are strongest on different criteria. Whereas the SP design may offer slightly greater flexibility and require fewer and weaker assumptions, the PP design is cheaper, faster, and uses data on individuals who most closely approximate marginal entrants. As SSA has stressed a strong desire for a *range* of plausible induced entry estimates, one promising avenue for further research is to implement both research designs and compare the results.

Acknowledgments

We would like to thank Bob Weathers, the SSA task manager, for his valuable input during the process of planning and conducting the research presented here. Other SSA stakeholders, including Richard Balkus, Stephen Goss, Eli Donkar, L. Scott Muller, Debbie Daugs, Jamie Kendall, Jeff Hemmeter, Chris Silanskis, Debbie Engler, Debbie Harkin, and Michael Wiseman, provided useful feedback during the planning phase of the project. In addition, we benefited from the technical advice and comments provided by a group of independent advisers to the project: David Autor (Massachusetts Institute of Technology), Richard Burkhauser (Cornell University), Hilary Hoynes (University of California, Davis), and Robert Moffitt (John Hopkins University).

We also thank our RAND colleagues, Bing Han for providing assistance with the power analyses, and Beverly Weidmer for advice regarding fielding surveys and assistance with the cost estimates for the stated preference research design. Sarah Kups and Liz Voss provided excellent research assistance.

Abbreviations

ALJ	administrative law judge
ACS	American Community Survey
BOND	Benefit Offset National Demonstration
CDR	Continuing Disability Review
EPE	Extended Period of Eligibility
EXR	Expedited Reinstatement
HRS	Health and Retirement Study
NHANES	National Health Activity and Nutrition Examination Survey
NSHA	National Study of Health and Activity
PIA	Primary Insurance Amount
PP	past policy
RFC	residual functional capacity
RP	revealed preference
SGA	Substantial Gainful Activity
SIPP	Survey of Income and Program Participation
SSA	Social Security Administration
SSDI	Social Security Disability Insurance
SP	stated preferences
TAG	Technical Advisory Group
Ticket Act	Ticket to Work Incentive and Work Incentives Improvement Act
TWP	Trial Work Period

CHAPTER ONE

Introduction

Background on the SSDI Program and the Benefit Offset

The U.S. Social Security Disability Insurance (SSDI) program was designed to provide income replacement to workers who are no longer able to work because of a long-lasting health condition. Unlike disability systems in many other countries, it is not a temporary disability system, nor does it prorate benefits for partial disabilities. Consequently, the system operates from an underlying presumption that SSDI recipients are largely unable to work. Applicants must demonstrate that they did not perform substantial gainful activity (SGA) for at least five months after disability onset, and once enrolled in the program the first dollar earned above the SGA threshold following a nine-month Trial Work Period (TWP) and a three-month Grace Period results in suspension of benefits. For the average SSDI beneficiary receiving a monthly benefit of $1,053 in 2007, this amounts to a staggering 100,000 percent implicit marginal tax rate on earnings.[1]

The presumption that SSDI recipients are unable to work has long been questioned, and recent legislative reforms have been oriented around the idea that many SSDI recipients *could* potentially work if offered targeted employment support services. The Ticket to Work Program, established in 1999 under the Ticket to Work Incentive and Work Incentives Improvement Act (P.L. 106-170; the Ticket Act) makes available a voucher (or "ticket") that can be used to obtain vocational rehabilitation and employment services within an approved network of public and private providers. The program was phased in between 2001 and 2004.[2] The Ticket Act also extended Medicare coverage to beneficiaries who return to work by offering them continuation of premium-free Medicare Part A for 93 months and the option to purchase Medicare Part B.

Under the Ticket Act, Congress also directed the Social Security Administration (SSA) to study the effects of a policy proposal to cut the implicit marginal tax rate on earnings to 50 percent. Under a new "benefit offset" policy, the SSDI benefit would be reduced by $1 for every $2 of earnings above a disregard amount, following the TWP and Grace Period.

[1] This contrasts with implicit tax rates in other social programs, such as Temporary Assistance to Needy Families, which range from 6 percent to 30 percent (Coe et al., 1998).

[2] New regulations in 2008 aimed to expand ticket eligibility, attract more service providers, and expand the range of rehabilitation and employment services offered.

While a benefit offset policy would undoubtedly affect the labor supply, earnings, and program exit rates of current SSDI beneficiaries, such a policy could affect program entry as well. In particular, there may exist a group of "marginal" individuals who under the current policy regime choose not to apply for SSDI benefits (despite a high probability of being eligible), but who *would* apply for benefits under a new benefit offset policy. This marginal group is distinct from two other groups of high-probability eligibles: those who would apply under *either* policy ("always takers") and those who would apply under *neither* policy ("never takers"). While the "always takers" and "never takers" are unaffected by the policy change, the marginal group is induced to apply for SSDI solely because of the policy change; hence the effect of interest is referred to as an "induced entry" effect. The size of the population of induced entrants is a critical input into an analysis of the effects of a benefit offset policy on overall program costs.

Most Salient Program Rules Affecting the Value of SSDI Participation

In this section, we provide a brief review of the most important SSDI program rules governing application, disability determination, program participation, and termination. This section draws heavily on the Red Book (Social Security Administration, 2009) and personal communications with SSA staff.

Eligibility. Workers are eligible for SSDI benefits if they are fully insured and have recent work activity.[3] In order to be fully insured, an individual must have accumulated at least one calendar quarter of work[4] in covered employment for every year elapsing since age 22, up to a maximum of 40 quarters.[5] The recency requirement requires that at least half of those quarters be earned within the last 10 years.

Definition of Disability. Disability is defined as the inability to engage in any Substantial Gainful Activity (SGA) because of a medically determinable physical or mental impairment that is expected to result in death, or that has lasted or is expected to last for a continuous period of not less than 12 months.

Substantial Gainful Activity. Work is considered substantial if it involves doing significant physical or mental activities or a combination of both. A gainful work activity is work performed for pay or profit, or is of a nature generally performed for pay or profit, or is intended for profit, whether or not a profit is realized. The threshold defining SGA is $1,000 per month in 2010 and increases annually with the cost of living. The SGA threshold is higher for blind beneficiaries ($1,640 per month in 2010).

Waiting Period. The applicant may not work above the SGA threshold during a five-month waiting period beginning with the first full calendar month following the date of disability onset. If benefits are awarded, payments begin after completion of the waiting period. Back payments are made to beneficiaries whose application process takes longer than five months.

---

[3] Blind workers need only be fully insured; the recency requirement does not apply.

[4] The amount of earnings required for a quarter of coverage in 2010 is $1,120; the amount increases annually with the national average wage index.

[5] Those who become disabled before age 24 need a minimum of six quarters earned during the past three years.

Disability Determination Process. SSDI applications are reviewed in five sequential steps. In step one, applicants earning more than the SGA threshold are denied, and those earning less than the SGA threshold proceed to step two. At step two, the application is evaluated for severity of medical impairment(s). Applications for impairments deemed obviously nonsevere or temporary (e.g., pregnancy) are denied, and the rest proceed to step three. Step three evaluates the upper tail of the severity distribution; applications for impairments meeting codified criteria in the Listing of Impairments are allowed at this step, while all others move to steps four and five, which add vocational factors in addition to the medical factors considered in the previous steps. In step four, applicants' residual functional capacities (RFCs) are determined, and applicants are denied if they are deemed able to meet the physical and/or mental requirements of their past jobs. Finally, applicants who reach step five are evaluated according to their ability to do any work in the national economy compared with individuals of similar age, education, and work experience. Applicants judged unable to work are allowed, and the remaining applicants are denied.

Reconsideration and Appeals. Applicants can appeal initial rejection. There are four stages in the appeals process. In the first stage, denied applicants can apply for *reconsideration* within 60 days of denial, and the reconsideration is performed by the same Disability Determination Service (DDS) office that made the initial determination. Applicants denied at reconsideration have 60 days to appeal to an administrative law judge (ALJ), who must consider the application using the same steps and in the same order as the initial review process. Applicants denied by an ALJ can further appeal to the SSA Appeals Council, and finally to Federal Court, which may choose to hear the case or remand the case back to the ALJ level.

Continuing Disability Review. Once an individual is receiving benefits, SSA periodically reviews his or her case to obtain information about the disability and any work activity in order to determine whether SSA should continue disability payments. This review is called a Continuing Disability Review (CDR). The length of time between CDRs varies from one case to another, depending on the likelihood of medical improvement.

Trial Work Period. After completing the waiting period, benefits commence and the beneficiary begins his/her Trial Work Period. The TWP allows the recipient to test his ability to work for at least nine months. During the TWP, full SSDI benefits are paid regardless of how high the recipient's earnings are. The TWP continues indefinitely until the recipient accumulates nine months of earnings above a threshold, which is lower than the SGA threshold (currently $720), during a rolling five-year period. These nine months do not need to be consecutive. At the end of the rolling TWP (once nine months of work above the threshold have been accumulated), there is a Work Continuing Disability Review, in which it is determined whether earnings are above or below the higher SGA threshold. If earnings are between the two thresholds (i.e., $720–$1,000), the beneficiary is issued a continuance of benefits until earnings exceed the higher SGA threshold. If earnings are above the $1,000 SGA threshold, benefits stop after a Grace Period, which includes the cessation month plus two additional months.

Extended Period of Eligibility (EPE). The EPE is a period of 36 consecutive months following the end of the TWP (i.e., if the individual has resumed work) when, if the disabled individual qualifies, he can restart his SSDI benefits without a new application, disability determination, or waiting period. The EPE begins the month after the TWP ends. During this period, benefits are paid for months of earnings below the SGA threshold as long as the recipient continues to have a disabling impairment. Benefits are suspended for months of earnings above the SGA threshold. A new application is not required to restart the benefits if, in a month of work, the recipient's earnings are below the SGA threshold. At the end of the EPE, benefits will be terminated if earnings are above the SGA threshold.

Expedited Reinstatement (EXR). After the 36th month of the EPE, if the individual is earning below the SGA threshold and receiving benefits, benefits continue. If not, benefits are terminated. If benefits are terminated because of earnings above the SGA threshold (as opposed to medical improvement) and earnings fall below the SGA threshold at any point within five years of when benefits stopped, then, under Expedited Reinstatement, benefits can be started again without a waiting period for a comparable or related medical condition.

Medicare Eligibility. SSDI beneficiaries receive Medicare coverage automatically once they have been enrolled in SSDI for two years after the disability onset date. If benefits are terminated because earnings exceed the SGA threshold, disabled individuals will continue to receive at least 93 consecutive months of Medicare Hospital Insurance (Part A), Supplemental Medical Insurance (Part B), and Prescription Drug coverage (Part D), after the TWP. Premiums are not paid for Part A. The 93-months count begins with the month after the last month of the TWP.

Previous Estimates of Induced Entry

There exist at least two estimates of the magnitude of potential induced entry under a new benefit offset policy. McLaughlin (1994) estimated that a $1-for-$2 benefit offset would increase the number of SSDI disabled worker beneficiaries by about 6.4 percent over a 10-year period (1995–2004). The estimate was constructed by first estimating the size of the pool of potential induced entrants (i.e., the nonbeneficiary population with earnings over the SGA threshold who would be medically eligible) using the 1978 Survey of Disability and Work, then assuming that 20 percent of this group would actually apply for benefits. The 20 percent take-up rate was an estimate that could be further refined.

In a related paper, Hoynes and Moffitt (1999) simulated the financial impacts of a number of potential reforms, including a $1-for-$2 benefit offset, on current and potential SSDI recipients. They conclude that the financial incentives for entering SSDI under a benefit offset policy may be substantial: Part-time (20 hours/week) workers earning the median wage could more than double their income if they enter SSDI under the new rules, and even full-time workers could increase their earnings by 35–46 percent by entering SSDI. Hoynes and Moffitt did not provide an estimate of the size of the pool of potential induced entrants or the percentage of induced entrants who would actually apply for benefits.

More recently, Benitez-Silva et al. (2006) use data from the Health and Retirement Study (HRS) to calibrate a life-cycle model of labor supply and SSDI claiming in order to estimate induced entry from a $1-for-$2 benefit offset. The model assumes that individuals choose whether and when to claim disability (and old-age benefits) in order to maximize the present discounted value of lifetime utility, which depends on consumption and leisure in each period. Onset of disability affects individuals' disutility of work (i.e., makes work more unpleasant), and the decision to apply for SSDI incurs a fixed cost, which includes any "stigma" or "hassle" (e.g., dealing with bureaucracy, waiting). They found that a $1-for-$2 benefit offset above the SGA threshold would increase SSDI applications by 2.2 percent and SSDI entrants by 3.2 percent.

SSA already possesses some knowledge of the strengths and weaknesses of different methods for estimating induced entry, but most of what is known in any detail pertains to experimental demonstration designs; SSA's formal knowledge of other potential designs is fairly general. For example, Tuma (2001), with assistance from a panel of independent advisers, identifies several methods that could be used to estimate induced entry, but most of her analysis is devoted to experimental designs; this is because the Ticket Act specifically proposed a demonstration project to study induced entry. Randomized controlled experiments are considered by many to be the most credible way of estimating the effect of policy changes, but it is not clear that they can be cost-effectively used to estimate induced entry. The main concern is that induced entry is a rare phenomenon, and thus very large samples are required to obtain reliable estimates. Additional concerns include the inability to control for informational and social interaction effects in an experimental design. This is important because the decision to apply for SSDI depends in part on knowledge of program rules and procedures. Information about these rules might come from sources other than SSA (e.g., relatives, friends, attorneys, advocates), and these channels may operate differently in an experimental context.

Tuma and advisers find that a demonstration study of induced entry, even implemented by randomizing at the county level, would likely yield very high costs and uncertain benefits relative to other approaches. They conclude by recommending that SSA consider alternative, more cost-effective approaches. This is the course SSA has chosen to pursue. The purpose of this report is to inform that deliberation. In particular, we present two detailed research designs, each capable of yielding credible estimates of induced entry under a benefit offset policy. These designs improve on previous estimates of induced entry in at least two ways. Relative to McLaughlin (1994) and Hoynes and Moffitt (1999), the designs offer a method of estimating likely take-up under the policy. In McLaughlin, take-up was an assumption and not informed by behavioral estimates. Similarly, in Hoynes and Moffitt, take-up was not estimated in the simulations of who would benefit under the policy. And while the designs here are not without their identifying assumptions, relative to the complex dynamic programming model in Benitez-Silva et al. (2006), the two methods presented here use simpler modeling assumptions. Where simple structural modeling is proposed in our second design, we also propose to leverage exogenous policy variation to improve identification of key parameters.

CHAPTER TWO

A Theoretical Framework for Induced Entry

In this chapter, we develop a theoretical framework designed to capture the most salient aspects of an individual's decision to apply for SSDI under the current program and under an alternative benefit offset policy. The model is dynamic, forward-looking, and yields steady-state conditions for individual application and employment behavior. It also yields a condition defining the group of disabled workers who would have a *new* incentive to apply for SSDI under the benefit offset policy—in other words, those who could be induced to enter the program by the new policy.

Suppose at any moment in time a disabled individual can choose to be in one of four discrete states: "Employed and Not on SSDI" (*E*), "Employed and on SSDI" (*ED*), "Not Employed and Not on SSDI" (*N*), and "Not Employed and on SSDI" (*ND*). Let $u(t)$ denote flow utility at time *t* as follows:

$$(2.1)\quad u(t)=\begin{cases} w(t)+h(t)-e(t) & \text{if } E \\ w(t)+\max\left[h(t),m(t)\right]-e(t)+b(t)-b(t)\cdot 1\left[w(t)\geq SGA\right]-g(t) & \text{if } ED \\ c(k) & \text{if } N \\ b(t)+m(t)-g(t) & \text{if } ND \end{cases}$$

where $w(t)$ represents earnings, $h(t)$ is the value of any employer-provided health insurance,[6] $e(t)$ is a monetized utility cost of working that captures the cost of exerting effort at work or the degree of employer accommodation, $b(t)$ is the SSDI cash benefit, $m(t)$ is the value of the SSDI Medicare benefit (and is equal to zero before the two-year waiting period is satisfied), and $g(t)$ is a monetized utility cost of applying for and participating in SSDI that may include stigma, foregone human capital, opportunity costs (which may vary with the business cycle), uncertainty in the award decision, or costs associated with the five-month waiting period. The utility cost is expressed as a recurring cost of program participation, but it could be split into up-front application costs and recurring participation costs; utility costs could also be made dependent on past

[6] Employer-provided health insurance could also come through a spouse's employer. In this case, $h(t)$ would also enter flow utility in the two nonemployment states.

employment and participation decisions and the severity of disability. The term $c(k)$ is a consumption floor, representing the individual's consumption level when neither employed nor on SSDI; it varies with household composition, k, to account for the presence of other family members who could support the individual or who themselves need support.

According to Equation 2.1, if the individual is employed and not on SSDI, utility at time t is just earnings plus the value of any employer-provided health insurance less the cost of effort. If the individual is employed but on SSDI, utility at time t consists of earnings plus the value of health insurance (either through the employer or Medicare, whichever is greater) less utility costs, plus the SSDI cash benefit. As specified under current policy, the term $-b(t)\cdot 1\left[w(t)\geq SGA\right]$ represents the deduction of the entire SSDI benefit if earnings exceed the SGA threshold ($1[\cdot]$is the indicator function). Note that when this occurs, only the cash benefit is lost; Medicare entitlement is maintained.[7] If the individual is not employed and not on SSDI, utility at time t is equal to the consumption floor, which could be achieved through the income of other family members or by participation in means-tested welfare programs. And finally, if the individual is not employed and on SSDI, he or she receives utility from the SSDI cash and Medicare benefits less utility costs.

Using dynamic programming methods, one can derive the expected value of discounted lifetime utility in any state from the present moment forward. This will equal flow utility in the current state *plus* the expected value of lifetime utility in the future. The expected value of lifetime utility in the future depends on a set of transition probabilities that govern the likelihood of moving from one state to another. For example, employed disabled individuals may separate from jobs with probability per unit time q, enter the SSDI program with probability per unit time a, and be terminated from SSDI with probability per unit time x.[8] The inclusion of the job separation probability, q, allows economic shocks to reduce the value of employment relative to program participation. Evaluating the individual's decision problem in this way yields a set of expressions describing the *return* on lifetime utility in each state. The easiest way of understanding them is to think of lifetime utility as an asset paying a rate of return equal to the subjective discount rate ρ . For the asset to be held, the return on the asset must equal any dividends plus expected capital gains or losses. Lifetime utility while employed pays a "dividend" at time t equal to $w + h - e$, and there is a probability per unit time q of a capital loss associated with job separation, $V_E - V_N$.

(2.2) $\rho V_E = w + h - e - q\left(V_E - V_N\right).$

[7] We abstract from the TWP, during which time the individual can earn above the SGA threshold without penalty. The deduction of benefits when earnings exceed the SGA threshold, as expressed in Equation 2.1, applies after the individual has exhausted the nine-month TWP and three-month Grace Period. Although benefits stop, Medicare entitlement is maintained for at least 93 months after the end of the TWP. The existence of the TWP means that costs associated with benefit reductions after the TWP will be borne in the distant future, and thus are discounted. This will tend to reduce their influence in the application decision rules that follow. Similarly, the existence of the two-year waiting period for Medicare coverage reduces the influence of the Medicare benefit in the application decision rule.

[8] In practice, only a small fraction of SSDI beneficiaries are ever terminated from SSDI on the basis of medical recovery.

On the other hand, lifetime utility while employed and on SSDI pays a dividend equal to its flow utility minus expected capital losses from possible program termination (*x*) and job separation (*q*):

(2.3) $\rho V_{ED} = w + \max\left[h,m\right] + b - b \cdot 1\left[w \geq SGA\right] - g - e - x\left(V_{ED} - V_{E}\right) - q\left(V_{ED} - V_{ND}\right)$.

Lifetime utility while not employed and not on SSDI pays a dividend equal to its flow utility plus expected capital gains associated with the probabilities of entering the SSDI program (*a*) or receiving a job offer (*j*):

(2.4) $\rho V_{N} = c(k) + a\left(V_{ND} - V_{N}\right) + j\left(V_{E} - V_{N}\right)$.

And finally, lifetime utility while not employed and on SSDI pays a dividend equal to its flow utility plus an expected capital gain associated with possibly receiving a job offer (*j*) and an expected capital loss from possible program termination (*x*):

(2.5) $\rho V_{ND} = b + m - g + j\left(V_{ED} - V_{ND}\right) - x\left(V_{ND} - V_{N}\right)$.

An employed disabled individual will choose to apply for SSDI if $V_{ED} \geq V_{E}$. Using Equations 2.2 through 2.5 to solve for the four V terms yields a condition describing the application decision in terms of earnings, the value of health insurance, the disutility of work effort, the cash benefit amount, opportunity and stigma costs, and the probabilities of job loss, program allowance, program termination, and receiving a new job offer. Differencing Equations 2.3 and 2.2 and assuming for simplicity that expected capital gains and losses have only second-order effects and can be ignored, a simplified expression for the SSDI application rule under current policy is:

(2.6) *Apply if*: $\max\left[h,m\right] - h + b - b \cdot 1\left[w \geq SGA\right] > g$.

Equation 2.6 says that the individual applies for SSDI when the gain in health insurance coverage, $\max[h,m] - h$, plus the net cash benefit, exceeds the utility costs of application and participation. The gain in health insurance coverage is maximized (and equal to m) for an individual with no employer-provided health insurance and minimized (and equal to zero) for an individual with employer-provided health insurance of greater value than Medicare. Note that an individual with no (or less-generous) employer-provided health insurance and low application/participation costs could benefit from program participation *even if* his entire cash benefit were offset for earnings above the SGA threshold.

In this framework, it is easy to see how the proposed \$1-for-\$2 benefit offset policy would affect the application decision. The benefit offset policy would change the flow utility or "dividend" associated with employment while on SSDI; the return on lifetime utility in that state becomes:

(2.3')

$$\rho V'_{ED} = w + \max\left[h,m\right] + b - \frac{1}{2}\left(w - SGA\right)\cdot 1\left[w \geq SGA\right] - g - e - x\left(V_{ED} - V_E\right) - q\left(V_{ED} - V_{ND}\right),$$

where now instead of losing the full benefit if earnings exceed the SGA threshold, benefits are reduced by $1 for every $2 of earnings above the SGA threshold. The application decision rule becomes:

(2.6') *Apply if:* $\max\left[h,m\right] - h + b - \frac{1}{2}\left(w - SGA\right)\cdot 1\left[w \geq SGA\right] > g$.

Although the net cash benefit is different under the benefit offset policy, the potential gain in health insurance coverage for the individual is the same as under current policy, since Medicare entitlement would be unaffected by the benefit offset.

The individual would be better off under a benefit offset policy if $V'_{ED} > V_{ED}$. Differencing Equations 2.3' and 2.3 (or alternatively differencing Equations 2.6' and 2.6) and assuming capital gains and losses are of second order, we find that the return to employment while on SSDI would rise under the new policy only for individuals with earnings in a well-defined range:

(2.7) $SGA \leq w < 2b + SGA$.

To a first approximation, this expression identifies the marginal group of disabled workers who would be better off under a new benefit offset policy and thus *could be* induced to enter the SSDI program: those earning between the SGA threshold and an amount equal to twice their SSDI benefit plus the SGA amount. Given an SGA threshold of $1,000 per month and a median monthly benefit of about $1,000, Equation 2.7 implies that disabled individuals with countable annual earnings[9] between $12,000 and an upper threshold that varies with the individual's potential benefit but is centered around $36,000 could be induced to enter the program under a benefit offset policy.[10] Those earning below the SGA threshold would experience no change in their incentive to apply for SSDI benefits; thus, if they chose not to apply under current policy, there is no reason to expect that the benefit offset policy would induce them to apply. Those earning above their upper threshold are similarly unaffected by the policy change; if they chose not to apply under current policy, there is no reason to expect that the benefit offset policy would induce them to apply. Equation 2.7 also shows that the size of the population of potential induced entrants is sensitive to the benefit offset rate. For example, if the benefit offset were lowered to $1 for

[9] For people with Impairment Related Work Expenses, countable earnings are earnings minus these expenses.

[10] The upper threshold is individual-specific because it depends on the individual's potential benefit. Using the distribution of annual benefits in the 2008 SSDI beneficiary population as an approximation to the distribution of potential benefits among potential induced entrants, about 50 percent of potential induced entrants will have an upper threshold below $36,000, 75 percent will have an upper threshold below $45,000, 90 percent will have an upper threshold below $55,000, and 99 percent of potential induced entrants will have an upper threshold below $62,000.

every \$3, the population of induced entrants would expand to include those earning up to a threshold centered on three times the benefit amount plus SGA (or about \$48,000). Importantly, Equation 2.7 only identifies those who would be strictly better off under a new benefit offset policy; it does not indicate the take-up rate among this group—that is, how many of them would actually apply. To understand take-up, the application rule in Equation 2.6' applies to this group just as it does in general.

Equation 2.7 does not depend on the value of Medicare coverage. While the value of Medicare does have an important effect on the individual's application decision, this effect is the same under current policy as under the benefit offset policy. Since Medicare coverage does not interact with the benefit offset in any way, it does not figure into the condition defining the marginal group of disabled workers who would have a new incentive to apply under the benefit offset policy.

Moreover, among those with a new incentive to apply, there will still be variation in the *magnitude* of the incentive. The application rule in Equation 2.6' suggests that, among those with a new incentive to apply, take-up rates would be higher among those with more to gain—those who could keep a greater portion of their benefit, individuals without employer-provided health insurance coverage, and those with low application costs. Since those with lower earnings would keep more of their benefit (less would be offset), are less likely to have employer-provided health insurance coverage, and face lower opportunity costs of application, the likelihood of take-up among the group of potential induced entrants is higher among those with lower earnings.

As noted above, Equation 2.7 is a first approximation. The lifetime values of all states will change in equilibrium, and this may in turn have second-order effects on the expected capital losses. Perhaps more importantly, this framework assumes a discrete labor supply choice—people choose to either work or not work. In reality, some can also adjust the number of hours they work, and if this is the case, some of those working just beyond the SGA threshold may increase their utility by reducing their hours just enough to qualify for the program (i.e., just so their total earnings do not exceed the SGA threshold). Allowing for labor supply adjustments on the hours margin or imperfect information about potential benefits would increase the upper earnings threshold in Equation 2.7 for some individuals. A further complication could arise if people negotiate down their wages (or strategically take lower-wage jobs) in order to attain or maintain qualification for the program.

A number of other simplifying assumptions could be relaxed to make the model more realistic. For instance, we could introduce risk aversion by relaxing the assumption of quasi-linear utility. This might be particularly important given the substantial uncertainty in SSDI award decisions, and it would reduce the expected net gain from program participation. More careful modeling of the TWP and Medicare waiting period would illustrate how some program costs and benefits arise up front while others are deferred into the future. The fact that benefit reductions under a benefit offset would begin at least 12 months after entitlement (even longer for those who take a long time to complete their TWP) means that this particular class of costs would be discounted in the application decision rule. On the other hand, the expected gain from the Medicare benefit would also be discounted.

One could also model the ease with which suspended benefits may be restarted during the Extended Period of Eligibility or under Expedited Reinstatement by incorporating a state variable recording prior benefit receipt. One could model departures from "rationality" by allowing for myopia, hyperbolic discounting, or even behavioral norms (which could change with the introduction of the benefit offset). Imperfect knowledge of SSDI program rules could be incorporated by assuming a population distribution of beliefs about program rules, and this distribution could be affected by the behavior and knowledge transmission of professionals, such as disability law attorneys and advocates. Finally, the framework is flexible enough to incorporate the behavior of other agents; for example, the SSDI entry probability, a, is in part a policy lever controlled by SSA through the application denial rate, which could be allowed to differ if disability examiners interpret the benefit offset policy as a change in SSDI's definition of disability. A final elaboration of the model might be with respect to dependent benefits; depending on how the benefit offset would interact with dependent benefits, the incentives for program participation could be different for people with dependents.

CHAPTER THREE

Research Designs

In this chapter, we present two research designs that could be used to estimate induced entry arising from a proposed benefit offset. The first is a stated preferences research design, in which one administers to a sample of respondents a series of stated choice experiments designed to elicit preference for claiming SSDI benefits under current and hypothetical circumstances (e.g., a benefit offset). The second research design uses past policy variation arising through variation in the SGA threshold over time and space to estimate a minimum of structural parameters necessary to forecast behavior under a benefit offset. Many more research designs were considered but, upon consultation with the Technical Advisory Group (TAG), were determined not worth pursing. Before proceeding to a detailed presentation of the selected research designs, we begin with a brief overview of the designs we considered but did not pursue, followed by a description of the criteria used to assess the designs.

Research Designs Considered But Not Pursued

Other research designs that were considered but later rejected included the use of pure dynamic lifecycle structural models (e.g., dynamic programming models). Such models enable one to flexibly model the decisionmaking environment and can be formulated to capture forward-looking behavior, uncertainty, and even information imperfections. They involve specifying equations describing utility streams in different decision states (e.g., employment versus SSDI participation), the information set on which decisions are based (e.g., health status, current earnings), and a set of transition probabilities describing how individuals move between states. Once the parameters defining utility are estimated or calibrated, a dynamic structural model can be used to simulate behavioral responses to policy changes that affect the relative value of each state. Structural models have been used in many settings, including the SSDI application decision. Recent examples include Burkhauser, Butler, and Gumus (2004), Benitez-Silva et al. (2006), and Kreider and Riphahn (2000). Overall, the TAG advised us that dynamic lifecycle structural models were too restrictive and hard to test; a better alternative was to consider more-flexible approaches in which policy variation would be combined with a simple structural model to derive implications for the benefit offset. The TAG stressed the importance of taking an adaptive approach in which one would start with a few key parameters (e.g., sufficient to define income and substitution elasticities) and see how far conclusions could be taken with a minimum of assumptions. This is the approach we propose for the research design using past policy changes in a simple structural framework.

We also investigated the use of economic variation as a natural experiment to identify induced entry effects. This approach is based on the idea that one could estimate the elasticity of the number of applications or accepted beneficiaries attributable to exogenous changes in economic circumstances that alter the relative return to continued work versus program participation. Examples of this approach are found in Black, Kermit, and Sanders (2002), who use temporal variation in coal prices and spatial variation in coal reserves in four states (Kentucky, Ohio, Pennsylvania, and West Virginia) to estimate the impact of local economic booms and busts on SSDI and SSI payments, and in Autor and Duggan (2003), who use state variation in labor demand and average replacement rates in order to estimate the elasticity of labor force participation of lower-skilled workers and SSDI program participation with respect to benefit generosity. The TAG advised us not to consider economic variation as the main source of identification for the model, primarily because the opportunity cost of application for economically displaced workers is much lower than it would be for employed individuals under a benefit offset policy; consequently, economic variation would tend to induce more entry than would ever arise under a benefit offset policy.

Finally, we investigated the existence of useful policy experiments from other countries. Many countries in recent decades have undertaken reforms in their disability insurance systems (see, e.g., Organisation for Economic Co-operation and Development [2003] for a summary of recent policy reforms). However, most of the reform efforts have been oriented toward the expansion of employment-related integration measures and the tightening of access to disability insurance benefits. In fact, only a few countries have considered the introduction of various forms of work incentives. Even if useful policy experiments did exist, a central challenge confronting this approach would be external validity. Direct comparison to the U.S. setting would be hampered by differences in institutional setting and history; for example, in some countries, disability insurance systems are used as if they were unemployment insurance or early retirement systems.

Criteria for Evaluating the Research Designs

We established five key criteria on which to evaluate each research design. The criteria were selected based on SSA's need to identify a method that can produce an unbiased, reasonably precise estimate of program entry by insured individuals who would qualify for SSDI, at reasonable cost and in a reasonable period of time. The five criteria are internal validity, external validity, flexibility, economy, and speed. The criteria are defined as follows:

- **Internal Validity.** Internal validity refers to how well an estimator approximates the true entry effect. We list and evaluate the plausibility of key assumptions necessary for identification, as well as sensitivity to deviations from the assumptions (robustness).
- **External Validity.** External validity refers to the generalizability of the estimate to the SSDI induced entrant population. For example, a method based on data from the SSDI eligible population would be considered to approximate near-perfect external

validity, whereas a method based on data from the general U.S. population, or even a population from another country, may have low external validity.

- **Flexibility.** Flexibility refers to the extent to which the design offers the possibility of varying policy parameters, such as the benefit offset rate or SGA amount, and/or relaxing assumptions about the decision environment.
- **Economy.** Economy refers to the cost of implementing the research design with sufficient statistical power to detect an entry effect of plausible size. High-cost designs obtain a poor economy rating, while low-cost designs obtain an excellent rating. Costs may also include any pilot testing, data collection, and/or computing resources.
- **Speed.** Speed measures the time it would take to produce an estimate with the research design, which will vary depending on factors such as whether the design involves new data collection and technical complexity.

Each research design section below begins with a description of the method and how it measures induced entry. We cite key findings from the literature that support the use of the method in this context, and we list possible data sources and/or collection techniques needed to implement the approach. Finally, we discuss the strengths and limitations of each research design according to each criterion.

Research Design Using Stated Preferences

A promising method of estimating potential entry effects as a result of a \$1-for-\$2 benefit offset is a *stated preferences* (SP) approach. Under this approach, one administers to a sample of respondents a series of stated choice experiments designed to elicit preferences for claiming disability benefits under different circumstances. Typically, respondents are presented with a set of scenarios describing different states of the world and asked to either rate, rank, or choose among different possible actions (e.g., continuing to work versus applying for SSDI benefits). The scenarios may have either real or hypothetical attributes (or a mix of both), which facilitates elicitation of preferences under hypothetical circumstances, such as a benefit offset policy or a hypothetical health state. Moreover, it is possible to conduct *randomized* choice experiments by randomly varying hypothetical attributes (e.g., the offset rate or SGA threshold) over respondents. Randomization increases variation in the design and facilitates measurement of a dose-response relationship between features such as the offset rate or disregard level and application behavior, improving efficiency of the estimates and reducing costs.

The SP approach has been used successfully in a number of settings. Of particular interest to this setting is a RAND study by van Soest, Kapteyn, and Zissimopoulos (2006) that used SP data to analyze preferences for full and partial retirement under hypothetical pension claiming rules, and a study by Delavande and Rohwedder (2009), who investigated preferences for spending and labor supply in response to a hypothetical 30 percent reduction in Social Security benefits. Louviere, Hensher, and Swait (2002) give a complete description of the method and review several studies comparing preference parameter estimates based on SP versus traditional revealed preference methods, finding that the two methods are usually quite close. However, as they stress, "SP data can capture a wider and broader array of preference-driven behaviors than revealed preference data on actual behavior, allowing for experiments with choice opportunities that do not yet exist in the market."

In this section, we present a detailed research design for implementing the SP approach. We begin with a statistical overview of the components of induced entry that will guide the discussion of how the method should be implemented. Next we outline a sampling plan and method for screening respondents. We follow with a presentation of our proposed SP scenarios and stated choice questions. This is followed by a methodological discussion of how response data from the SP scenarios could be analyzed to obtain an estimate of induced entry. Finally, we assess the design according to the five key design criteria (internal and external validity, flexibility, economy, and speed), highlighting the strengths and limitations of the SP design. Throughout the section, we note when it would be beneficial for SSA to engage in pilot-implementation activities (e.g., pilot testing) in order to enhance implementation success.

Components of Induced Entry

Let B = 1 if the benefit offset policy is in place and B = 0 otherwise (i.e., the current policy). The vector X denotes characteristics determining program qualification, such as health, age, education, and work history, and $F(X)$ is the cumulative distribution of X in the population. Then, assuming the acceptance criteria are the same under both policies, we have

$$(3.1)\quad \textit{Induced Entry} = \int \Pr(\textit{Accepted} \mid \textit{Apply}, X)\left[\Pr(\textit{Apply} \mid X, B=1) - \Pr(\textit{Apply} \mid X, B=0)\right] dF(X)$$

This gives the rate of induced entry into the SSDI program for a given population. The total number of induced entrants is the induced entry rate multiplied by the size of the population (N). Alternatively, if we condition on the set of individuals who did not apply under the baseline policy, then the induced entry rate is

$$(3.2)\quad \textit{Induced Entry} = \int \underbrace{\Pr(\textit{Accepted} \mid \textit{Apply}, X)}_{(III)} \underbrace{\Pr(\textit{Apply} \mid X, B=1)}_{(II)} \underbrace{dG(X)}_{(I)}$$

where $G(X) = F(X \mid \textit{Apply} = 0, B = 0)$.

Thus, an estimate of the rate of induced entry consists of three components: the population "at risk" of qualifying for SSDI benefits under the new benefit offset policy (I), the application or take-up rate (II), and an acceptance probability (III).

In an ideal world, if we were able to perfectly condition on the set of eligible applicants (I) and there were no classification errors, then the probability of acceptance for these applicants (III) would be 1. As we expand the sample beyond this set of applicants, the average probability of acceptance will fall as we include more people with a lower likelihood of qualifying for the program. Thus, components (I) and (III) are inextricably linked, and will tend to impact cost (through the size of the sampling frame) and external validity (through the accuracy of the screening criteria to identify people who would truly qualify for SSDI) independently of take-up (II). We discuss each of these components in turn below.

Sampling Plan

Implementing the SP design requires new data collection. Because the target population—potential induced entrants—has unknown characteristics and is likely small relative to the general population, a critical issue is how to sample from the target population in a cost-effective way. This issue has surfaced repeatedly in the induced entry context, and it was one reason why a panel of expert consultants to SSA concluded that a national demonstration project would not be a cost-effective way of measuring induced entry (see Tuma, 2001). The sampling plan we propose leverages insights from economic theory and advantages unique to SSA and to the SP approach; these insights and advantages enhance cost-effectiveness.

Sampling Frame

The first step is to identify a sampling frame. A sampling frame is an enumeration of individuals in the target population, from which one draws a random sample. Sampling frames can be built from scratch or, in some cases, derived from frames used by existing surveys. SSA already has an ideal sampling frame—an administrative database identifying every worker insured for SSDI benefits in the United States who has not previously applied for SSDI. The database contains key pieces of information, such as date of birth, sex, earnings history, and mailing address. Mailing address is critical—a survey organization could be enlisted to reverse match addresses to phone numbers, so that sampled individuals could then be contacted by either mail or phone.

By definition of its being the universe of insured workers, the SSA sampling frame includes, among other workers, the entire target population. But because the exact characteristics of the target population are unknown, it is not possible to identify *ex ante* exactly which workers belong to the target population. This is the issue of cost-effectiveness: The less precisely that members of the target population can be identified, the more members of the general population need to be sampled in order to guarantee that "enough" of them belong to the target population—enough, that is, for meaningful statistical inference.

Fortunately, economic theory offers some useful guides for winnowing down the sampling frame. According to the theoretical framework presented in Chapter Two, only a subset of the currently employed would ever be better off participating in SSDI under a $1-for-$2 benefit offset, relative to current policy: those with annual earnings between $12,000 (the annualized SGA threshold) and an individual-specific upper threshold centered around $36,000 (twice the annual benefit plus the annualized SGA threshold). The variability in the upper threshold comes from variability in the annual benefit amount, which is itself a function of the individual's earnings history. People earning beyond their upper threshold would be worse off participating in SSDI under a benefit offset compared with current policy, while people below the lower threshold would be neither worse nor better off compared with current policy.

Accordingly, the sampling frame could be narrowed to include only individuals with annual earnings between $12,000 and an individual-specific upper threshold. Because the SSA administrative data include annual earnings and the annual benefit amount (essentially the Primary Insurance Amount [PIA]), should the individual become entitled in a given year, the upper threshold could be easily computed for each person. Moreover, since, as we showed in Chapter Two, the threshold is an inverse function of the offset rate, one could maximize flexibility by using the smallest potential offset rate to compute each individual's upper earnings threshold. For instance, if an offset rate of $1 for $3 is the most generous offset rate policy that would be considered, then the upper earnings threshold for each person could be set at roughly the annualized SGA threshold plus three times the individual's annual benefit amount if he or she were to become entitled in that year.[11] Then, if an individual's annual earnings fall within the lower and upper thresholds, the individual would be part of the sampling frame.

The earnings range above is defined for disabled workers, and thus it represents earnings *after* disability onset. Below, we discuss an alternative approach to the baseline SP design that allows one to sample individuals who have never been disabled and ask them to imagine how they would react to a change in SSDI program rules if they were to become disabled. Since earning power is likely to decrease significantly once one becomes disabled, the range of earnings from which the sample should be drawn should be shifted upward to capture *pre*-onset earnings. For example, if individuals' earnings drop 20 percent post-onset, then the $12,000-to-$36,000 range of post-onset earnings would correspond to a pre-onset earnings range of $15,000 to $45,000. Previous research shows that individuals may experience significant and persistent earnings declines after the onset of disability (see, e.g., Mok et al., 2008). However, this research does not explicitly account for the fact that individuals who qualify for SSDI have incentives to keep their earnings below the SGA. Moreover, the group affected by the benefit offset contains individuals who would earn *above* the SGA even if they were to qualify for SSDI. Thus, a more targeted study of the relationship between pre- and post-onset earnings for marginal SSDI entrants may be needed.

An alternative to using the SSA administrative database as a sampling frame would be to pursue access to the respondents to an existing nationally representative survey. A potentially attractive possibility is the American Community Survey (ACS), which surveys roughly 1.5 million people ages 25–64 per year.[12] Of these, about 45,000 are likely to be medically eligible nonparticipants and thus ideal candidates for a study of induced entry. If linked to SSA administrative data, the same narrow sample definition based on earnings, PIA, insured status, and non-SSDI participation could be used. We discuss how a sample based on the ACS would affect administration of a health screener, as well as resulting impacts on sample size, respondent burden, and cost, below.

Probability Sampling

Once the sampling frame is defined, individuals could be sampled from it using conventional probability sampling methods. Available variables such as age, sex, race, earnings, geographic

[11] The threshold could also include an additive factor to account for labor supply adjustments on the hours margin as well as the participation margin. See Chapter Two for a brief explanation of why the "breakeven" earnings threshold would be higher if we account for continuous changes in hours.

[12] We also considered the Survey of Income and Program Participation (SIPP), the Health and Retirement Study (HRS), and the National Health Activity and Nutrition Examination Survey (NHANES). Because these surveys are much smaller than the ACS and not targeted to potentially disabled individuals, it would be necessary to implement the SP design using health as an attribute (discussed further below) in order to achieve sufficient statistical power.

location, and even prior SSDI program participation or application could be used to conduct stratified sampling.

As discussed in Chapter Two, the expected gain to participating under the benefit offset varies with one's earnings. Theoretically, one could gain power by using a higher sampling rate for individuals with higher expected take-up rates if one is willing to make assumptions about the relationship between take-up and earnings. However, if these assumptions were wrong, then the design could be left underpowered. In the discussion below and in our power analysis, we take a conservative approach and assume a uniform sampling rate within the target earnings range.

Screening

In an ideal world, the sampling frame would be defined by both earnings and health status. If information about health status were available, one could sample only individuals with a high probability of belonging to the population of potential induced entrants. Since health status is not available in SSA's administrative data, individuals cannot be sampled on the basis of their likelihood of qualifying for the SSDI program. Rather, they must first be sampled irrespective of health, then subsequently screened for health status prior to administering the survey. Thus, restricting the survey to individuals with poor health status necessitates the sampling of a very large number of individuals, since prior evidence suggests that only 3 percent of the nonbeneficiary population would likely qualify for SSDI benefits (Dwyer et al., 2001). In a lower-earnings population, such as that defined by the sampling frame, this number is likely to be bigger, though still essentially small.

In 2002, SSA contracted with Westat to develop a questionnaire that could classify respondents as possibly or likely disabled for use as a screener for the then-proposed National Study of Health and Activity (NSHA). Westat developed a 26-item survey measuring functional limitations (e.g., reaching, walking, sitting), activities of daily living (e.g., bathing, dressing, eating), instrumental activities of daily living (e.g., preparing meals, doing light housework), and a number of other items (e.g., whether the respondent has a condition that limits basic physical activities, or uses a walking device). The scales of most items were modified to a binary (0 or 1) scale, with two exceptions that were modified to a range of 0–2. Thus, adding up all 26 items produced a score ranging between 0 and 28. Respondents scoring 0–2 were classified as not likely disabled; those scoring between 3 and 4 were classified as possibly disabled; and those with scores greater than 5 were classified as likely disabled.

A good health screener has a high probability of classifying disabled individuals as likely beneficiaries (power) and a low false positive rate. Westat performed a validation study of its screener, comparing its performance to that of four other screening algorithms. It used three validation groups. The first group consisted of a sample of actual beneficiaries who were determined to be disabled at some point by SSA. The second and third groups consisted of nonbeneficiaries for whom the disability determination process was mimicked (that is, folders were completed and reviewed by a disability examiner who determined whether the case met the medical eligibility requirements for SSDI). The second group consisted of nonbeneficiaries who would be allowed if they were to apply for benefits, and the third group consisted of nonbeneficiaries who would be denied. Table 3.1 presents classification rates for each of these three groups based on the screener ultimately selected by

Westat. The Westat screener has a "true" positive rate of 85.2 percent and a false positive rate of 24.2 percent, which was enough to outperform other screening algorithms.[13] Note that, given the very low prevalence of qualifying disability in the population, a low false positive rate is crucial to the ability of the screener to identify potential induced entrants. For example, assuming 3 percent of the sample frame would be medically eligible for disability benefits, administering the strict version of the Westat screener (i.e., including only those who are "likely disabled," with scores greater than or equal to 5) would result in a sample of roughly 26 percent (= 81.4% × 3% + 24.2% × 97%) of initial respondents. Among these, only about 9.4 percent (= 81.4% × 3% ÷ 26%) of individuals would be allowed onto the program if they applied. Even if as many as 10 percent of the sample frame would qualify for benefits if they applied, only about 27 percent of the screened sample would qualify for the program.

Table 3.1
Classification Rates for Beneficiaries, Allowed Nonbeneficiaries, and Nonallowed Nonbeneficiaries from Westat Validation Study

Classification	Beneficiaries	Allowed Nonbeneficiaries	Nonallowed Nonbeneficiaries
Not likely disabled	4.1%	4.3%	60.2%
Possibly disabled	10.7%	14.3%	15.4%
Likely disabled	85.2%	81.4%	24.2%
Sample size	467	70	255
Source: Frey et al. (2002)			

Note that the fraction of individuals passing through the health screen would immediately provide us with an estimate of the size of the population at risk for induced entry (i.e., before considering the take-up element). The fraction of respondents classified as likely disabled equals the fraction who are medically eligible (*p*) times 81.4 percent (the probability of correctly classifying allowed nonbeneficiaries) plus (1– *p*) times 24.2 percent (the probability of incorrectly classifying nonallowed nonbeneficiaries). Solving this equation for *p* yields the fraction of nonbeneficiaries who are insured and medically eligible for SSDI, with earnings in the range of those who would be better off participating under the benefit offset policy.

Finally, even if one were to use the ACS as a sampling frame, the need to screen based on health status is not necessarily completely eliminated. The ACS contains six questions on disability, which could be used to narrow the initial sample frame, potentially reducing the length of the screening questionnaire and increasing the rate of useable respondents. However, without conducting a validation study using the ACS questions linked to SSA administrative data, one would not know how well a screener based on the six ACS questions—alone or in parallel with the Westat screener—would perform relative to the 26-item Westat screener described above.

---

[13] The Westat screener also outperforms alternative classifications based on (1) a simple affirmative response to the question "Does your health or condition prevent you from working?" ("true" positive rate = 53 percent, false positive rate = 57.7 percent) and (2) the preferred method developed by Dwyer et al. (2001) ("true" positive rate = 67.2 percent, false positive rate = 30.6 percent).

Demographic Characteristics

Demographic information used to assess qualification for SSDI benefits but not routinely available in the administrative earnings data, such as education, skill type, or work experience (e.g., based on the work history form used by disability examiners), could also be collected. For applicants who do not qualify solely on medical criteria, this information is used by disability examiners to apply the medical-vocational "grid" that guides award decisions. This information would be useful to (1) refine prediction of the likelihood of acceptance into the program and (2) assess whether estimated entry patterns under the benefit offset policy would vary with demographic characteristics.

Experimental Design

The SP approach consists of presenting respondents with real or hypothetical scenarios and asking them to choose what action they would take (e.g., applying for disability benefits or continuing to work). Each scenario describes all relevant features or "attributes" of the state of the world, including policy parameters (benefit offset rate, disregard amount, any time limit on benefits), program parameters (SGA threshold, waiting period), and possibly individual characteristics, such as health, which may be the respondents' actual health or a hypothetical health state posed to them. Table 3.2 presents a potential list of attributes that could be specified in the hypothetical scenarios.[14]

Table 3.2
Example Choice Attributes

Proposed Program Parameters	Current Program Parameters	Individual Characteristics
Offset rate Time limit on benefit offset Earnings disregard	SGA threshold Waiting period	Health status

The text box below presents an example of such a scenario (Scenario A). Scenario A describes the current state of the world, in which the benefit offset policy does not exist and beneficiaries earning over the SGA threshold lose all of their benefits. The format of the scenario is based on lessons from a series of cognitive interviews conducted by Westat in 2002, with questions designed to measure induced entry into SSDI resulting from a benefit offset. In particular, it is based on show cards developed for the third round of interviews. Two of the attributes listed in Table 3.2 are highlighted in bold. The scenarios can easily be personalized to display each respondent's actual monthly SSDI benefit (in brackets in the text box, to denote that it is preloaded into the survey instrument), computed based on his PIA. Note that Scenario A is unlikely to yield useful answers if it is asked of respondents in fairly good health, so it is reasonable to restrict its use to respondents whose health

[14] SSA expressed interest in estimating responsiveness to a time limit on the benefit offset. This is meant to be analogous to the extended period of eligibility. In the current Benefit Offset National Demonstration (BOND) project, beneficiaries are eligible for the benefit offset for a period of 60 months, after which they would no longer be entitled to SSDI benefits. Because the Westat study found that respondents had a very difficult time with complex concepts, we omit discussion of a time limit in the description of the offset policy. Incorporating a time limit would likely entail a fair amount of pilot testing to develop usable questions.

measured on the screener described above indicates they are likely (or possibly) disabled. We describe how one could incorporate healthier respondents in the analysis further below.

Scenario A (Baseline)

In this scenario, assume the following:

To QUALIFY for disability benefits…

- You *must* have a health problem that affects your ability to work AND
- You must have no (or very limited) earnings from a job for at least **5** months.

IF you qualify…

- You get a monthly disability benefit of [$1,000].
- If you earn more than **$1,000** per month from working, then you lose your benefit.

For this scenario, assume that your health is the same as it is now.

Note that the SP design depends on how well the scenarios are able to mimic the decisionmaking process of potential applicants. This includes how well they present information on which applicants base their decisions to apply. This implies that two types of preliminary data would be critical to the success of the SP design. First, little is known about what applicants actually know about SSDI program rules, or even the likely application costs (monetary and otherwise), when they apply. Information about SSDI might come from sources other than SSA (e.g., relatives, friends, attorneys, advocates), and thus respondents may have very different ideas about how SSDI works when they apply to the program.[15] A survey of current applicants would be a way to obtain this information and incorporate it into the scenarios. While research exists on individuals' knowledge of other social insurance programs (e.g., Social Security; see Leibman and Luttmer, 2009), we are not aware of any rigorous studies on this issue in the context of SSDI.[16]

Second, careful pilot testing of the presentation of information to SP respondents would be necessary. In their interviews, the Westat researchers found that respondents frequently were overwhelmed by long questions and tended to have a lot of trouble with dollar amounts. Many respondents misunderstood the actual program rules and particularly did not understand restrictions on earnings during the waiting period and while receiving disability benefits. The researchers found that visual aids (show cards) helped respondents understand the questions, as long as the wording was consistent with the verbal questions posed by the interviewers. Responses to detailed examples of how the offset policy worked were inconsistent; some respondents (in the first round) thought they would help, while others (in the second round) found them more confusing than helpful. However, careful

[15] In addition, potential applicants may weigh information from different sources differently. The SP design could attempt to mimic this feature by attributing various sources to statements about program rules. Again, gathering of preliminary data on the information sources of SSDI applicants and careful pilot testing would be needed.

[16] Westat piloted some questions regarding program awareness and knowledge in their cognitive interviews; however, their focus was on developing useable questions rather than analyzing the responses. Additionally, only 8–10 potentially nonrepresentative individuals participated in each of three rounds of interviews.

pilot testing could potentially be used to construct an example that would aid respondents in understanding current program rules and the proposed benefit offset policy. Pilot testing could also be used to incorporate information that may be relevant but potentially too complex for respondents to understand without a careful presentation, such as the existence of and rules surrounding the TWP.

After respondents are presented with a scenario, such as Scenario A above, they would be asked to state whether they would apply for disability benefits in that case. An example of such a *stated choice question* is the following:

> If this were the case, would you apply for disability benefits, yes or no?

A simple yes-or-no question avoids confusion over what responses such as "very likely" or "very unlikely" mean, especially since this may differ across respondents. (The Westat report found that respondents found definitions in terms of percentages confusing.)

In order to assess induced entry from the proposed benefit offset policy, respondents can also be presented with Scenario B, presented in the text box below. Thus, a simple estimate of the effect of a $1-for-$2 benefit offset policy in the current program (and economic) environment is simply the difference between the fraction of respondents who answer yes to the stated choice question when presented with Scenario B versus Scenario A (times the population size). Note that, in this case, Scenario A serves as a coherency check on respondents' answers; in theory, if no previous SSDI applicants are in the sample, then no one should respond that they would apply for disability benefits when presented with Scenario A. Pilot testing should thus incorporate this coherency check to ensure that respondents are understanding the information presented in the scenarios.

Scenario B (Benefit Offset)

In this scenario, assume the following:

To QUALIFY for disability benefits…
- You *must* have a health problem that affects your ability to work AND
- You must have no (or very limited) earnings from a job for at least **5** months.

IF you qualify…
- You get a monthly disability benefit of [$1,000].
- If you earn more than **$1,000** per month from working, you lose *part* of your benefit:
    - Your benefit is reduced by $1 for every $**2** you make over $**1,000** a month.
    - But overall, the more you earn from a job, the higher your total income for the month (job earnings *plus* reduced disability benefits).

For this scenario, assume that your health is the same as it is now.

Modeling Take-Up

While only these two scenarios (and technically only one) are needed to answer the question of who would be induced to apply for SSDI benefits under the specific proposed benefit offset policy, the SP approach can be harnessed to provide much more flexible estimates under varying circumstances. For example, the effect of a $1-for-$3 benefit offset policy could easily be studied by substituting "$3" in place of "$2" in Scenario B. Indeed, with some assumptions about the structure of preferences, stated choice responses under the original and modified Scenario B can be used to estimate the effect of a policy never even posed to respondents, such as a $1-for-$4 offset. Consider the following simple logit model of SSDI application:

$$(3.3)\ \Pr(Apply \mid Z) = \frac{\exp(Z\gamma)}{1+\exp(Z\gamma)},$$

where Z is a vector of attributes, such as the benefit offset rate, and γ is a vector of taste weights associated with each attribute. Given variation in Z, one can estimate the taste parameters γ by maximum likelihood. In a simple model where the benefit offset rate enters linearly, it is easy to extrapolate the effect of a new offset rate by plugging it into the formula above along with the coefficient estimates γ.

Indeed, any of the attributes in Table 3.2 can be randomized across respondents in order to estimate a general model of SSDI application behavior. Note that the values of the attributes should be randomized so that the components of Z are generally orthogonal to each other. That is, if one varies the benefit offset rate and disregard level, they should be varied independently so that the model can be identified. If, for example, one would like to present two offset rates (1/2, 1/3) and two disregard levels ($500, $1,000) then one should construct at least three different scenarios (e.g., (1) offset rate = 1/2 and disregard level = $500, (2) offset rate = 1/2 and disregard level = $1000, and (3) offset rate = 1/3 and disregard level = $500 (or $1,000)) so that the effects of the two attributes can be distinguished from each other statistically. Note, however, that not every respondent needs to receive the same set of scenarios, as long as the scenarios are randomized across respondents. This allows a great deal of flexibility in the number of attributes (and interactions) the model can accommodate.

The model above can easily be expanded to allow for influences from individual characteristics such as gender, age, health status, health insurance, and employment (as measured by the screener) by including these in Z. As such, it is possible to use estimates based on the sample to extrapolate how a hypothetical population might respond to a benefit offset policy. For example, if the policy is enacted several years later, trends in population aging, health, and health insurance coverage can be incorporated by predicting entry conditional on individual characteristics, weighted by the frequency with which they are presumed to occur in the population, and summing the predicted probabilities.

However, the ability of the model to predict responses accurately conditional on health status or other characteristics will depend on the observed frequency of such characteristics in the sample. Thus, a weakness of the above approach is that its success depends on one's ability to identify and survey disabled workers who could be induced into entering the program under the new regime. The "marginal" SSDI applicants for a $1-for-$2 offset are those who meet the medical requirements for disability insurance and who earn (net of any earnings disregard) more than the SGA threshold but less than roughly three times their expected benefit. Since the earnings restriction is fairly straightforward, the difficult part is identifying potential applicants who would be likely to pass the medical screening for disability. As mentioned above, Dwyer et al. (2001) estimate that just under 3 percent of the general population (ages 18–64) was medically eligible but not receiving disability benefits as of 1992. One solution is to screen on health to select those most likely to qualify. Another is to *introduce health as an attribute* to be varied explicitly in the hypothetical scenarios presented to respondents. Introducing health as an attribute serves two purposes: (1) it allows one to control for health uniformly across respondents and (2) it reduces the number of people that need to be sampled to obtain enough members of the target population by asking all respondents to respond *as if* they were members of the target population (i.e., as if they had a health problem that qualified them for SSDI), regardless of whether they actually are part of the target population.

Introducing Health as an Attribute

The text box below describes a scenario in which respondents are asked to consider how they would respond if they experienced severe back pain. The example is taken from Kapteyn, Smith, and van Soest, 2007, who use vignettes to adjust self-reports of work disability to account for reporting differences across countries. While vignettes provide a useful starting point for incorporating hypothetical health into the stated preference scenarios, many of the vignettes developed in the existing literature are aimed at general work disability and do not necessarily conform to the definition of disability used by the SSDI program (namely, the inability to perform work for substantial gainful activity for at least 12 months). In particular, many vignettes focus on how individuals *feel*, rather than what individuals can *do*.

Scenario C (Back Pain)

In this scenario, assume the following:

To QUALIFY for disability benefits…

- You *must* have a health problem that affects your ability to work AND
- You must have no (or very limited) earnings from a job for at least **5** months.

IF you qualify…

- You get a monthly disability benefit of [$1,000].
- If you earn more than **$1,000** per month from working, you lose *part* of your benefit.
  - Your benefit is reduced by $1 for every $**2** you make over $**1,000** a month.
  - But overall, the more you earn from a job, the higher your total income for the month (job earnings *plus* reduced disability benefits).

For this scenario, imagine that your health is **different from what it is now. Instead, assume that you have back pain that makes changes in body position while working very uncomfortable. You are unable to stand or sit for more than half an hour. Medicines decrease the pain a little, but it is there all the time and interferes with your ability to carry out day-to-day tasks at work.**

One way to develop hypothetical health attributes is to mimic the disability determination process. Allowances are roughly evenly divided between medical and vocational allowances. Medical allowances are based on listings of impairments, which are divided into body systems (e.g., musculoskeletal system). Hypothetical health attributes could be developed in consultation with a physician to describe work-related symptoms of common conditions that meet the listings for a particular body system code (e.g., major dysfunction of a joint within the musculoskeletal system). Health attributes could be constructed for common conditions, such as musculoskeletal disorders.

Similarly, vocational allowances are based on the residual functional capacity (RFC) assessment used by SSDI examiners to evaluate disability. Questions from the RFC could be mapped into a likely assessment of maximum sustained work capacity (sedentary, light, or medium) by regressing outcomes from SSDI application data on answers to the RFC. The questions contributing most to the functional capacity assessment would then be used to develop a short description of health, such as the example in the text box below. The descriptions could be validated by surveying disability examiners and asking them to rate likelihood of an assessment (or disability award) based on the description. Since vocational allowances depend on age, education, and work experience in conjunction with a determination of maximum sustained work capacity, it may be useful to restrict their use to respondents who would likely qualify for the program based on the scenario provided. For example, a semi-skilled high school graduate restricted to "sedentary" work would qualify for the program if he were older than age 50, but not if he were younger.

Example of Hypothetical Health Attribute Based on RFC (Sedentary)

For this scenario, assume that your health is **different from what it is now. Instead, assume that you can stand and/or walk (with normal breaks) for a total of less than 2 hours in an 8-hour workday. You frequently have trouble stooping and crouching, and occasionally have trouble balancing. Additionally, you are limited in handling and fingering objects.**

A crucial assumption is that individuals can forecast their behavior under unfamiliar health conditions. A wide literature has established that the "framing" of questions can have a significant impact on survey responses. For example, Smith et al. (2006) provide evidence that individuals can be primed to focus on a specific health condition, which then becomes salient when answering assessments about their life satisfaction. In their study, they randomized the introduction to a survey of Parkinson's disease patients to focus on either a population of Parkinson's disease patients or the general population. The authors found that respondents who were given the introduction that focused on Parkinson's disease rated their life satisfaction lower than those given the general introduction when life satisfaction was assessed after health satisfaction. Similarly, in the vignette literature, Hopkins and King (2010) show that switching the question order so that self-assessments follow hypothetical vignettes results in more-standardized response scales across respondents. Thus, respondents use the hypothetical vignettes to inform their interpretation of later questions. One test of whether respondents are able to imagine their reactions under different health circumstances is to pose the corresponding profile with hypothetical health information omitted (e.g., Scenario B) to respondents with the *actual* health limitation, and compare the responses of those with the same, but actual versus imagined, health limitations. However, given the scarcity of such respondents, it may be difficult to carry out this type of validation exercise.

As discussed above, one of the consequences of disability onset is reduced earnings power. Thus, another coherency check similar to the one proposed above is to ask individuals what they imagine their earnings would be if they continued to work given their hypothetical health status. Those who would apply under the benefit offset but not under the current rules should hypothesize that their post-onset earnings fall within the range identified by economic theory (e.g., $12,000 to $36,000 for an individual with the median monthly benefit of $1,000 under a $1-for-$2 offset). Additional follow-up questions could ask individuals to imagine whether they would remain employed (assuming they are currently employed) or be forced to drop out of the labor force due to the onset of the hypothetical limitation, and what would happen to their health insurance. These questions would allow one to test whether employment status or health insurance are related to induced entry.[17]

Estimating Induced Entry

Baseline Approach

As discussed above, the simplest method to estimate induced entry screens respondents who are likely disabled and asks them the stated choice question—"If this were the case, would you apply for disability benefits, yes or no?"—based on Scenarios A and B. The estimate is the difference in the proportion answering yes to these two scenarios, weighted by the probability of acceptance conditional on application to the program. One could construct individual-specific acceptance weights based on auxiliary regressions of the award decision on the health index plus individual characteristics, such as age or education, for a sample of SSDI beneficiaries. A concern is that induced entrants might differ from current applicants on unobservable dimensions; indeed, the fact that under current policy induced entrants choose to work instead of applying for SSDI suggests they may have less severe medical conditions or unusually accommodating employers. If the former, their true acceptance probabilities would be lower than those of applicants under the current rules. This is essentially a selection problem, and there are a number of possible solutions, including parametric selection models and computation of best- and worse-case bounds.

A sample size of roughly 8,400 respondents would enable one to detect an induced entry effect as small as 2 percentage points using a two-sided t-test with 5 percent significance and 80 percent power and assuming a baseline take-up rate of 70 percent. Assuming a response rate of 60 percent, and assuming that only about one-quarter of respondents make it through the health screen, at least 56,000 potential respondents would need to be contacted and 33,600 administered the health screen to obtain a sample size of 8,400. The estimated cost of administering even the most basic survey to likely disabled respondents only is quite high: $2.1 million, assuming that respondents are contacted by telephone with a mail-in visual aid, that they are paid $10 for completing the screener and $5 for completing the SP questions if eligible, and that it takes them an average of 12 minutes to complete the screener and 4 minutes (2 minutes for each SP question) to

[17] Note that the theoretical framework outlined in Chapter Two suggests that only employed individuals would be better under the benefit offset. The theory also suggests that health insurance should not play a role in induced entry because Medicare is the same under the current and proposed policies. An additional test of the role of health insurance would be to randomize information about Medicare to respondents and examine whether estimated application rates differ under the current and proposed policies. As the provisions of the recently enacted healthcare reform take effect, understanding the role of health insurance in take-up decisions becomes even more important.

complete the survey.[18] This does not include necessary pre-survey activities, such as carefully designing and pilot testing the questions, or post-survey activities (i.e., analyzing the data).

Note that all cost estimates are based on using SSA administrative data to construct the sampling frame. As noted above, the ACS contains six questions on disability, which could be used to narrow the sample frame, reduce the length of the screening questionnaire, and increase the rate of useable respondents, all of which would reduce costs. However, a validation study of a screener based on the ACS questions would be needed to estimate the cost savings.

Baseline Plus Approach: Modeling SSDI Take-Up

Combining the SP approach with a statistical model, like the logit model outlined above, markedly reduces the number of respondents that need to be sampled. Intuitively, imposing a particular structure for preferences over program participation (e.g., a logit model) harnesses new information and allows one to improve the *precision* of the estimates, thereby allowing reductions in sample size without losing power, assuming the model is not (grossly) mis-specified. Furthermore, each respondent can be asked to evaluate *multiple* independent scenarios (profiles), leading to greater reductions in sample size. The set of scenarios presented to each respondent, as well as the order in which they are presented, could be randomized to eliminate any framing effects.

In this case, the power analysis is more complicated than the simple t-test above, and no analytical solution exists. A Monte Carlo simulation reveals that, in the case of two attributes (offset rate and disregard level) that are allowed to take on the values presented in Table 3.3, stratifying by health status, each profile must be evaluated approximately 190 times by a given type of respondent in order to detect an induced entry effect as small as two percentage points, assuming 5 percent significance level, 80 percent power, and a baseline take-up rate of 70 percent. Note that a full factorial design produces $4 \times 3 = 12$ possible profiles (i.e., there are 12 unique combinations of offset rate [4 values] and disregard level [3 values]). If we assume that health status of respondents can be grouped into five categories, then 11,400 ($5 \times 12 \times 190$) evaluations are needed. Therefore, if each respondent evaluates 6 profiles, then the sample size is 1,900 ($11{,}400 \div 6$); if each respondent evaluates all 12 profiles, the sample size is reduced to 950. Table 3.4 presents estimates of sample size, respondent burden (in terms of survey length), and cost of administering the survey if respondents were asked to evaluate 6 and 12 profiles each, respectively, and we assume that respondents completing the more involved SP questionnaire are paid $15 (in addition to $10 for the screener).

Table 3.3. Attributes and Levels Assumed in Power Analysis: Baseline Plus Approach

Attribute	Levels
Offset rate	1 (Baseline), 1/2, 1/3, 1/4
Disregard level	$0 (Baseline), $500, $1000

---

[18] These cost estimates were prepared by RAND's Survey Research Group and are meant as ballpark estimates only. Note that if SSA were to pursue this design, competitive bidding might result in lower costs.

Table 3.4. Estimates of Sample Size, Respondent Burden, and Cost of Administering Survey

Approach	Attributes Varied	No. Respondents Completing		Profiles/ Person	Avg. Length in Minutes		Cost Estimate
		Screen	Profiles		Screen	Profiles	
Baseline	None	33,600	8,400	2	12	4	$2,100,000
Baseline plus, short	Offset rate, disregard	7,600	1,900	6	12	12	$632,190
Baseline plus, medium	Offset rate, disregard	3,800	950	12	12	24	$380,983
Alternative, short	Offset rate, disregard, health	2,280	2,280	5	12	10	$353,994
Alternative, medium	Offset rate, disregard, health	1,140	1,140	10	12	20	$241,885
Alternative, long	Offset rate, disregard, health	760	760	15	12	30	$204,515

Note: Number of respondents is based on power analysis assuming baseline probability = 70%, effect size = 2%, and power = 80%. Cost assumes telephone mode with a mail-in visual aid and response rate of 60%. We assume the following payments to compensate respondents for their participation in the survey: $10 to complete the screener and $15 ($5) to complete the SP questions if eligible under the baseline plus and alternative approaches (baseline approach).

To estimate induced entry, one first administers the stated preferences questions to respondents. Using these data, one then estimates the logit model specified in Equation 3.3, where Z includes the attributes, benefit offset rate, and disregard level, as well as a given level of health limitation (e.g., sedentary). Using the estimated coefficients $\hat{\gamma}$, one can then construct

$$(3.4)\quad \hat{P}_0 = \frac{\exp(Z_0\hat{\gamma})}{1+\exp(Z_0\hat{\gamma})} \text{ and } \hat{P}_1 = \frac{\exp(Z_1\hat{\gamma})}{1+\exp(Z_1\hat{\gamma})},$$

where Z_0 contains the *current* program rules (i.e., benefit offset rate = 1, disregard level = \$0) and individuals' actual health limitation, and Z_1 contains the proposed program rules (i.e., benefit offset rate = 1/2, disregard level=\$1,000) and the same level of health limitation. Then the estimate of induced entry for someone with a given level of health is simply $\hat{E} = \hat{P}_1 - \hat{P}_0$. Note that we subtract baseline entry in order to measure *induced* entry to the benefit offset; however, a significant estimate for $\hat{P}_0$ would suggest problems with the design of the SP questionnaire that should have been remedied at the pilot testing stage. To obtain a population-level estimate for the rate of induced entry, one should take the weighted sum of the individual $\hat{E}$ estimates, where the weights represent probability of acceptance into the program, as discussed above. The number of induced entrants is simply the rate times the number of medically eligible individuals in the population, estimated using the fraction of individuals classified as likely disabled based on the health screener, as described above. One can obtain estimates of take-up $\hat{E}$ *conditional on a given health class* by aggregating the responses of those individuals with the same health status. Thus, it is straightforward to take into account trends in population health (or analogously, population aging, etc.) by reweighting the conditional take-up estimates for any given distribution of health.

Alternative Approach: Introducing Health as an Attribute

Introducing health as an attribute that can be controlled and varied across even healthy respondents further reduces costs. Savings are achieved by posing SP questions to *all* willing respondents, in contrast with the baseline approach, which screens out roughly three-quarters of respondents.[19] In addition, since the health descriptions are constructed explicitly to conform to SSA's disability determination criteria, acceptance weights are not needed since the probability of acceptance is 100 percent by construction.

Including health as an attribute increases the total number of profiles to be evaluated by respondents. For example, if we assume five levels for health (e.g., two scenarios based on

[19] If one worried that healthier respondents might be unable to imagine having a health impairment, then one could administer a weaker screen than that proposed in the baseline approach, e.g., by screening out only those who are not likely disabled (and keeping those who are possibly disabled). Taking estimates of type I and II errors from Table 3.1, this approach would screen out approximately 59 percent of the sample. Thus, to achieve a final sample of 2,280 respondents, approximately 11,382 potential respondents would need to be contacted (in contrast with 4,667 assumed in the report) and approximately 6,829 would consent to be screened. Additional costs would result from contacting 6,715 additional respondents and administering the 12-minute health screening questionnaire to 4,549 additional respondents.

listings and three based on RFC), then the number of possible profiles is 60 ($5 \times 4 \times 3$) (where the number of levels for the offset rate [4 values] and disregard level [3 values] are given in Table 3.3). However, now that health is an attribute, each profile can be evaluated by all respondents and need not be stratified across respondents according to actual health status. As before, each profile must be evaluated approximately 190 times in order to detect an induced entry effect as small as two percentage points. Table 3.4 presents estimates of the sample size, respondent burden, and cost if respondents were asked to evaluate 5, 10, and 15 profiles, respectively, with health as an attribute.

As before, take-up estimates $\hat{E} = \hat{P}_1 - \hat{P}_0$ can be constructed for a given health class, where $\hat{P}_1$ is estimated entry under the proposed benefit offset and $\hat{P}_0$ is estimated entry under the status quo. Note that, in the case where health is an attribute, it is *necessary* to subtract baseline entry in order to estimate induced entry, since respondents' preferences under the status quo are not revealed by their actual decisions (as was the case where responses were conditioned on *actual* rather than hypothetical health). To obtain a population-level estimate for induced entry, take the weighted sum of the $\hat{E}$ estimates conditional on health, where the weights now represent the share of potential applicants with health conditions of similar severity. Estimates of the prevalence of such conditions can be taken from the health screening questionnaire administered at the beginning of the survey.

Evaluative Criteria

Internal Validity:

- The SP approach relies on respondents being able to accurately forecast their behavior under new and unfamiliar (policy) conditions.
- However, the SP approach requires much weaker assumptions about the unobserved expectations and opportunities of potential entrants than traditional revealed preference (RP) methods. RP methods must impose structure to extrapolate how potential entrants might respond to the new policy from observations of actual behavior under existing conditions. In contrast, SP experiments allow one to trace out disabled workers' preferences under hypothetically varying conditions—including conditions that do not yet exist—without the need to characterize correctly the entire environment.
- Pilot testing by Westat of a similar survey design revealed that respondents may have a difficult time processing questions relating to application decisions, especially when the question includes "too many" numbers. On the one hand, SP responses may mimic real-life decisions made under limited comprehension of the program. On the other hand, the SP approach does not allow respondents much time to process the information, or to consult family members or others to help them make decisions as they might in real life.
- The baseline approach may suffer in terms of accuracy if it is hard to identify potential entrants. By relaxing this screening restriction, the alternative approach (adding health as an attribute) allows for a larger potential sample frame.

External Validity:

- External validity for the baseline approach, which poses questions to non-applicant respondents who have been screened according to SSDI eligibility criteria, is quite high.
- External validity for the alternative approach, which varies health as an attribute, requires the additional assumption that potential applicants can forecast their behavior under unfamiliar *health* conditions.

Flexibility:

- Proposed program parameters, such as the benefit offset rate or earnings disregard amount, can easily be varied.
- Other attributes of the scenarios can be varied to produce a range of estimates based on varying degrees of information available to respondents. SSDI rules and departures from the current rules can be presented clearly, as in the example above, in order to allow respondents to make more informed decisions, or they can be stated more vaguely to capture respondents' actual uncertainty over program rules. For example, by calling out the five-month waiting period, the example above raises the saliency of that particular aspect of the application process.
- The scenarios could be made dependent on the actual working hours of respondents (or their earnings), or respondents could be asked to evaluate scenarios with hypothetical work hours or earnings amounts. They could depend on an average probability of acceptance, or they could depend on respondents' perceived probabilities of acceptance. Work hours and perceived probabilities could be elicited earlier in the survey.

Economy:

- New data collection is costly; however, SSA administrative records provide an inexpensive yet comprehensive sample frame. The ACS may also provide an ideal sample frame.
- Combining the SP approach with a statistical model and administering scenarios based on hypothetical health attributes allows one to decrease the sample size while still maintaining statistical power.

Speed:

- Collecting new data, including pilot testing the survey instrument, will add time relative to methods based on existing publicly available or administrative data. The time cost may be lower if one can take advantage of an Internet-based platform for either pilot testing or final data collection.

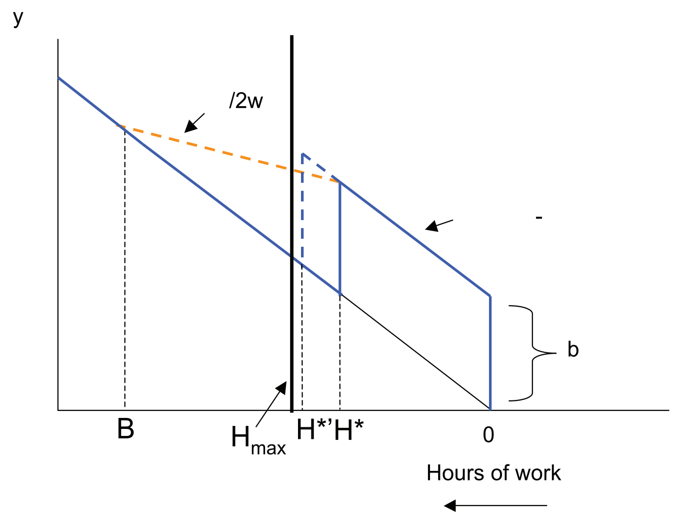
Research Design Using Past Policy Changes in a Simple Structural Framework

While there exist no past policy changes that are exactly equivalent to introducing a benefit offset, changes in the SGA threshold are relatively close. In this section, we propose a research design that would exploit changes in the SGA threshold as a way to learn about possible induced entry effects arising from a benefit offset. The research design consists of two parts. In the first part, we present a reduced form approach that would provide an estimate of induced entry using policy variation in the SGA level over time and relative to average wages across states. The second part proposes a simple structural framework that could be used, along with the SGA policy variation, to estimate key behavioral parameters of the SSDI application decision. These parameter estimates, obtained from behavioral responses to changes in the SGA threshold, could then be used to simulate induced entry under a hypothetical benefit offset policy.

The SGA threshold is a fundamental program parameter, determining both initial eligibility and ongoing entitlement to SSDI benefits. It figures directly into the current work rules, interacting with the current "full benefit offset" in much the same way as it would interact with a partial benefit offset, and thus changes in the SGA threshold offer a potentially instructive natural experiment.[20] The solid blue line in Figure 3.1 represents the SSDI budget constraint at the end of the TWP, under current policy. The dotted blue line shows how the budget constraint would be affected by a change in the SGA, while the dotted orange line shows how the budget constraint would change with the introduction of a benefit offset. Under current policy, those earning above the SGA threshold, after their TWP, are ineligible for benefit payments. Participants' net income Y is reduced by the full cash benefit amount b if they choose to work more than H^* hours, where for a given wage w, H^* defines the point at which earnings, wH^*, equal the SGA level. The loss of the entire benefit at this point creates a discontinuous drop in income at the SGA threshold. An increase in the SGA level to $wH^{*\prime}$ would lead to higher net income for those working between H^* and $H^{*\prime}$ hours, since they would now be eligible to receive benefits. Under the benefit offset, individuals could work even more than $H^{*\prime}$ and still receive benefits, although benefits now would be reduced by 1$ for each $2 increase in earnings.

[20] In the Technical Advisory Group (TAG) meeting, past SGA changes were identified as a particularly informative natural experiment. As one TAG member put it, if there is no effect of an increase in the SGA on entry, then there is unlikely to be a major effect of a benefit offset on entry.

Figure 3.1
SSDI Budget Constraint Before and After a Change in the SGA Threshold and Before and After Introduction of a Benefit Offset



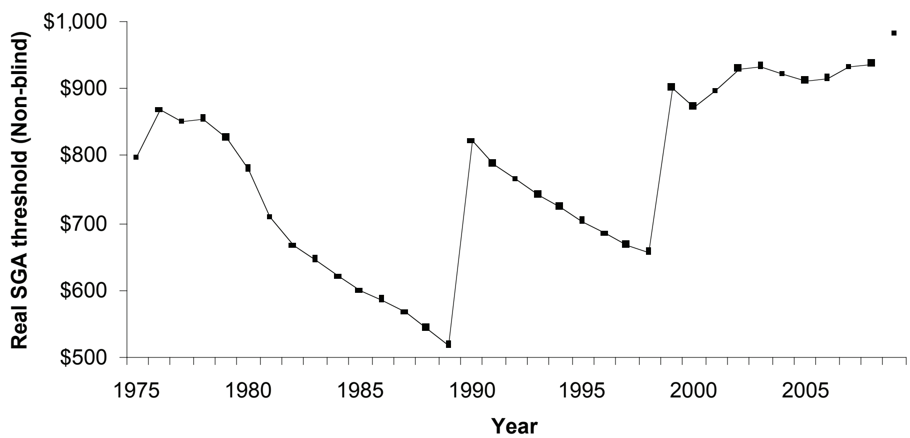
For low-wage individuals, a large change in the SGA level and the introduction of a benefit offset would affect the budget constraint over the same range of hours. For instance, consider an individual earning just over $5 an hour, the minimum wage in 1998. In 1998, the SGA threshold for this individual corresponded to H^* = 100 hours per month. When the monthly SGA level rose from $500 to $700 in 1999, his new SGA threshold became $H^{*\prime}$ = 140 hours per month. In comparison, a hypothetical benefit offset would apply to work hours between H^* = 100 and his benefit offset breakeven point, B, the point at which his disability benefit would be fully offset. His breakeven point under a $1-for-$2 benefit offset would correspond to B = 380 hours per month (assuming a potential monthly benefit in 1999 of $700), well beyond maximum workable hours of 160 hours per month (indicated by the solid line at H_{max} in Figure 3.1). In other words, the change in the SGA in 1999 affected a minimum-wage earner working between 100–140 hours per month, while the benefit offset would have affected a minimum-wage earner working between 100–160 hours per month—nearly the same range of hours. Importantly, as an individual's wage rate increases, H_{max} shifts leftward, and the benefit offset extends over a greater range of hours relative to an increase in the SGA. The 1999 increase in the SGA affected individuals earning twice the minimum wage ($10/hour) only if they were working 50–70 hours per month, whereas the benefit offset would have affected them over the range of 50–160 hours per month. Smaller changes in the SGA affect a smaller range of hours relative to a benefit offset for both types of workers.[21]

[21] An additional difference between the effect of an SGA increase and a benefit offset, not represented in Figure 3.1, is that for people with earnings between the old and the new SGA threshold, the increase in the SGA would delay their completion of the TWP. In contrast, the benefit offset policy would have no effect on the duration of the TWP.

Under the Social Security Act, the Commissioner of Social Security has the authority to set the SGA level for disabled individuals other than the blind. SSA has increased the nominal SGA threshold several times in past decades, and since December 2000 the threshold has been indexed to a measure of annual average wages in the United States. While raising the SGA threshold is not equivalent to introducing a benefit offset, the impact of a benefit offset on the individual budget constraint depends critically on the SGA threshold, since earnings below SGA would likely be disregarded from the offset.[22] Like the benefit offset, an increase in the SGA level may prompt some beneficiaries to venture into the workforce if the higher threshold makes available new options for combining work with benefit receipt; but the availability of new options may also make the SSDI program more attractive to new applicants who are disabled but currently working.

Figure 3.2 shows the evolution of SGA levels for the nonblind since 1975, expressed in real terms.[23] Periods during which the real SGA amount was declining correspond to periods in which the nominal SGA amount was flat. Increases in the nominal SGA, before its indexation in the year 2000, took place in the years 1976, 1977, 1978, 1979, 1980, 1990, and 1999, with the largest increases in 1990 and 1999.

Figure 3.2
Real SSDI SGA Amount, by Year (2009 dollars)



Although the SGA level is set nationally, it is relatively more generous in areas with lower costs of living and lower average wages, compared with areas with higher costs of living and higher average wages. In lower-wage areas, applying for SSDI might be more attractive if individuals are still able to work in a variety of occupations while still receiving benefits. Similarly, *absolute changes* in the national SGA amount will induce different relative changes in different areas of the country. To illustrate this, Figure 3.3 shows the density function of relative changes in the SGA amount between 1998 and 1999, by state, as a percentage of the state average annual wage measured in 1998. Figure 3.4 provides an alternative view of the distribution of relative changes in the SGA amount in 1999, showing its geographical

[22] In principle, the disregard need not be set at the SGA level.

[23] Real SGA amounts are obtained using the Consumer Price Index–All Urban Consumers.

distribution across states. In 1999, the nominal SGA amount rose from $500 to $700 per month, an average relative change of 8.6 percent of average annual wages. Importantly, there is considerable variation in the *relative change* across states; the coefficient of variation for the distribution is 13.2 percent.[24]

Figure 3.3
Density of Relative Changes in SGA, by State, in 1999 (as a percentage of state average annual wage)

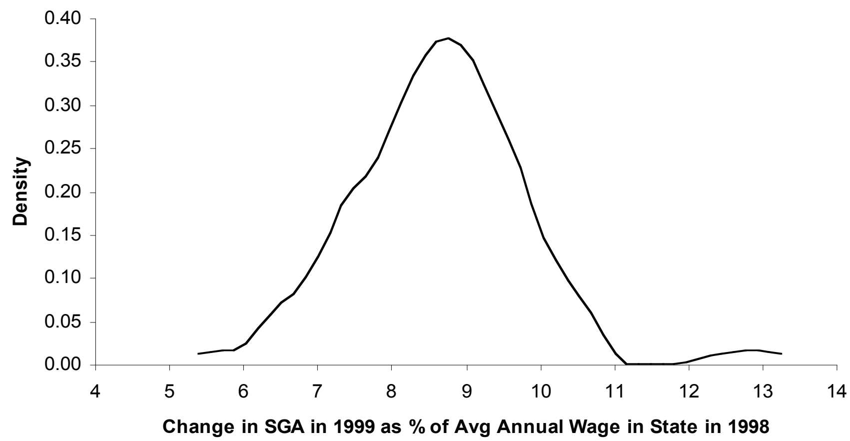
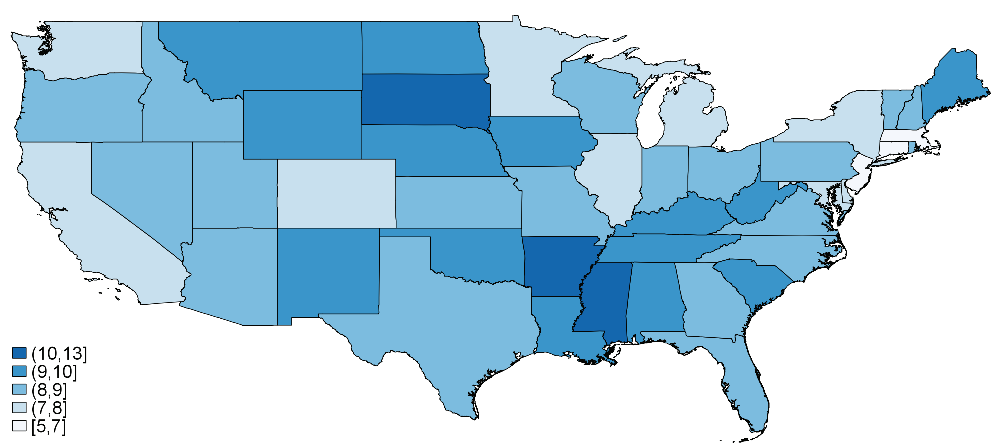


Figure 3.4
Density of Relative Changes in SGA, by State, in 1999 (geographical distribution)



In principle, any of the many changes in the SGA level over time constitutes a potential natural experiment that could be exploited to assess how applications respond to changes in work incentives. One could make use of SSA administrative data to compare

---

24 A related literature has used variation in replacement rates arising from the interaction of variation in regional wage levels with the progressivity of the benefit formula to estimate the effect of SSDI benefit levels on the labor supply of disabled individuals (see Autor and Duggan, 2003).

application and/or enrollment rates across states with different average wage levels, before and after changes in the real SGA level.[25] Specifically, one could estimate models of the following type to assess how changes in the SGA level affect the fraction of individuals applying for or receiving SSDI benefits:

$$(3.5)\quad SSDI_{st} = \alpha_t + \mu_s + \beta SGA_{st} + \varepsilon_{st}\,,$$

where $SSDI_{st}$ is the fraction of individuals applying for and/or enrolling in SSDI benefits in state *s* at time *t*. The term α_t is a year effect capturing common factors such as macroeconomic conditions that influence SSDI applications and/or allowances in each year. In the same way, μ_s is a state effect and controls for fixed, state-specific components of applications and allowances. In addition, one could include state-level unemployment rates over time to control for variation in business cycles across states. The key explanatory variable is SGA_{st}, which denotes the real relative SGA level in state *s* at time *t* ; in other words, the real national SGA level divided by the average annual wage in state *s* at time *t.* In order to avoid idiosyncratic fluctuation in real relative SGA levels due to transitory changes in mean wages over time, one could compute the average wage over a period of several years. Real relative SGA levels exploit two sources of variation in the SGA—variation over time and variation across states due to differences in average wages. The parameter of interest, β, measures the reduced form effect of the SGA level on application or allowance rates.[26]

A potential problem with an estimate of β obtained from Equation 3.5 is that it could be contaminated by unobserved state-specific trends—trends that might drive variation in both SSDI application rates and the real relative SGA (perhaps through average wages). In order to minimize the impact of such state-specific trends, one could consider an alternative specification that would relate the *change* from year $t-1$ to year t in applications and/or allowances ($\Delta SSDI_{st}$) to the *change* from year $t-1$ to year t in the real relative SGA level in state *s* (ΔSGA_{st}), while still controlling for state fixed effects. In addition, one could estimate the model separately for different time periods and compare the results.[27]

Estimating the model for different time periods would be of interest in and of itself. For example, by estimating the model for years with bigger changes in nominal SGA levels and comparing the results to estimates based on smaller changes, we could assess to what extent the results are driven by large changes in the SGA amount, and by changes in which years. This would help address concerns about identification, given that the two largest

---

[25] States could alternatively be compared on the basis of average education level.

[26] By comparing the separate effects of SGA variation on application and acceptance rates, one can also learn about the marginal entrants into the SSDI program. For instance, if application rates increase after a change in SGA but acceptance rates decrease, then this would suggest that marginal entrants might be on average in better health than current beneficiaries. On the other hand, if both application and award rates increase following a change in the SGA level, this will suggest that marginal entrants and current beneficiaries might not differ significantly in health status.

[27] As an alternative approach to dealing with state-specific confounding trends that might be correlated with average wages, one could use state variation in average education, as noted above. This information could be used to classify states into treatment and control groups and perform a difference-in-differences analysis.

changes in the nominal SGA amount occurred just prior to recessions. In addition, the declining value of the real SGA over the late 1970s and 1980s may have reduced applications by forward-looking individuals inclined to combine program participation and work over the longer run. This effect may have changed once the SGA amount was indexed to inflation beginning in 2000.

Finally, Equation 3.5 could be equivalently estimated at the individual level using SSA administrative earnings records linked to application data. Given that the relative SGA level varies at a more aggregated level (state-time level as opposed to individual level), the estimated parameter of interest (β) would be the same as in the aggregated specification. However, this approach would offer the flexibility to interact individual characteristics (e.g., age, PIA) with the real relative SGA in order to estimate conditional take-up rates.

Relating SGA-Induced Entry Effects to the Proposed Benefit Offset

The parameter β in Equation 3.5 provides an estimate of induced entry arising from variation in the SGA level. If $\beta = 0$, one may conclude that SGA-induced entry has been negligible; this suggests that induced entry arising from a benefit offset would also likely be negligible. On the other hand, $\beta > 0$ points to some degree of SGA-induced entry and indicates a need for further analysis to relate the SGA-induced entry effect to the benefit offset setting. This requires placing some structure on the individual's decision problem under current program rules in order to estimate key behavioral parameters that could then be applied to hypothetical program rules.

Conceptually, in an economic utility-maximization framework in which individuals choose consumption and leisure subject to a budget constraint, the SSDI program directly affects the budget constraint through two potential channels. The first is through the provision of a disability benefit that increases income. The additional income enables an individual to increase both consumption and leisure—the latter achieved by reducing hours of work, in some cases by enough to qualify for the program—and is called an income effect (or income elasticity). The second potential channel arises if program rules alter the price of leisure or, conversely, the return to work. Current program rules apply a very high implicit marginal tax rate on the first dollar earned above the SGA threshold after the TWP and Grace Period—indeed, benefits are suspended, equivalent to a full benefit offset.[28] A $1-for-$2 benefit offset would impose a much lower implicit marginal tax rate (50 percent) on a range of earnings above the SGA threshold, offsetting only part of the benefit. The presence of an implicit marginal tax on earnings reduces the effective return to work and causes the individual to either increase leisure by reducing hours worked (termed a substitution effect or substitution elasticity) or decrease leisure by increasing hours worked in order to make up for reduced consumption (an income effect). The total effect of the SSDI program on hours worked and program participation can be decomposed into income and substitution elasticities, which themselves are functions of the behavioral parameters of the utility maximization problem.

---

[28] Benefits resume for months in which earnings fall below the SGA threshold during a period of 36 months (the Extended Period of Eligibility).

Although the precise substitution and income effects on hours of work associated with the program are different under a $1-for-$2 benefit offset than they are under the current full benefit offset, the underlying behavioral parameters governing individual preferences for consumption and leisure should be the same. Therefore, with estimates of these parameters obtained by using behavioral responses to the level of and changes in the SGA, one can forecast the implied behavioral responses to a benefit offset. In this context, because of the structure imposed on the problem, observational data from a single cross-section of time would be sufficient to identify the income and substitution elasticities; however, as will be explained below, an advantage of our approach is that policy *variation* due to SGA changes is an important additional source of identification that would improve estimates of the key behavioral parameters.

Alternatively, one could think about using income and substitution effects already estimated in the existing literature to simulate the impact of the benefit offset. The main concern with this approach is external validity. For instance, elasticities derived using business cycle variation likely relate to a different marginal applicant, in particular, one with a very low opportunity cost of application. In the same way, estimates from other transfer programs may not reflect the behavior of the marginal SSDI applicant under a benefit offset.

Consider a simplification of the theoretical framework described in Chapter Two, where we abstract from the dynamic aspects of the model, and assume that individuals decide how many hours to work and whether to participate in the SSDI program. Utility is a function of hours of work (*H*) (a "bad," since work hours are inversely related to leisure), consumption/income (*Y*), and participation in SSDI (*P*). Following Moffitt (1983), we incorporate a utility cost or stigma parameter (ϕ) to take into account the fact that some eligible individuals may decide not to apply for SSDI benefits even if it makes them "better off" in an economic sense.[29] In particular, we assume that this stigma parameter is separable and depends on severity of the disability *d*:

(3.6) $U(H,Y,P) = U(H,Y) - \phi(d)P$.

In addition to capturing disutility derived from participating in the program, the stigma parameter may also include disutility arising from a number of other barriers to program participation, such as application costs, lost human capital while in the program, or lack of information about the program. Additionally, a significant component of the decision to apply for disability benefits is uncertainty of acceptance (see, e.g., Halpern and Hausman, 1986).

Since it is not possible to separate empirically the influence of true stigma from risk of denial, this is also included in ϕ. Since both are likely decreasing in severity of the disability, we model stigma as a function of *d.* Note that in Equation 3.6 the stigma cost is a fixed component that arises merely from enrollment in SSDI. It could also be modeled with a

[29] Dwyer et al. (2001) find that approximately 3 percent of the nonbeneficiary population ages 18–64 would meet SSA's medical criteria for disability, and among these, two thirds have earnings below the SGA level.

variable component that depends on the size of the benefit payment received as in Moffitt (1983).

Individuals choose hours of work (*H*), their level of consumption or income (*Y*), and whether to participate in SSDI (P = 1) in order to maximize their utility function subject to the following budget constraint:

$Y = wH + N$ if not on SSDI, or
$Y = wH + N + b - b \cdot 1[wH \geq SGA]$ if on SSDI,[30]

where *N* represents other nonwage income, *w* represents the individual's hourly wage, and *b* is the benefit amount if on SSDI.

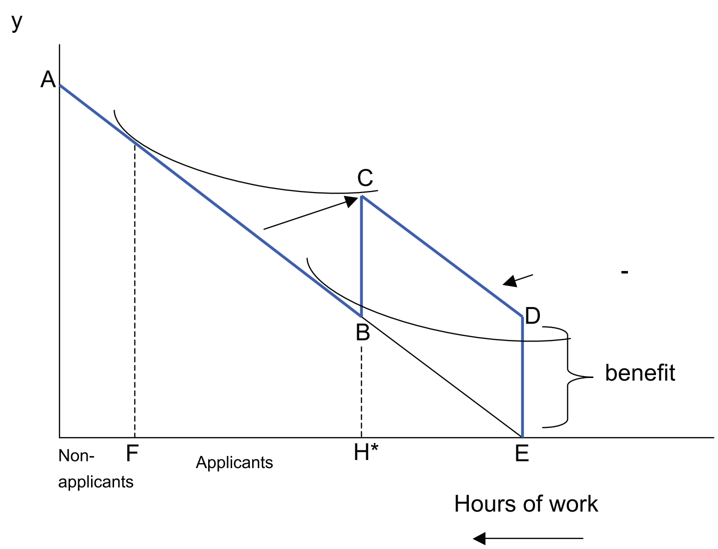
This simple model predicts that participation and labor-supply responses to changes in the net return of SSDI participation will depend crucially on the "stigma" of SSDI participation. Assume that individuals are heterogeneous in both taste for work (i.e., marginal disutility of work) and taste for SSDI participation (stigma). Then the probability of participating in SSDI is greater for individuals with greater distaste for work and/or lower stigma. This implies two types of nonparticipation predicted by the model. The first type includes medically eligible individuals with earnings below the SGA threshold who decide not to apply. These individuals are characterized by high distaste for both work and SSDI. A second group of nonparticipants includes *non*eligible individuals who do not participate in SSDI (that is, individuals who are medically eligible but who do not qualify based on earnings since they are above the SGA threshold). These individuals have low distaste for work but high distaste for participating in SSDI.

To obtain individuals' optimal choices, we first solve for the utility-maximizing choices of hours of work (*H*) and income (*Y*) assuming P = 0 and P = 1, respectively. Then, the optimal choice for participation is obtained by comparing indirect utility when on and off SSDI. Figure 3.5 shows the budget constraint when on and off SSDI and indifference curves for hypothetical people. As can be seen in this figure, an individual can choose to locate on one of three areas of the budget constraint: (1) He can decide not to participate in SSDI and locate on segment A-E, (2) he can decide to participate in SSDI and locate on the interior of segment C-D, or (3) he can participate in SSDI and locate at the discontinuity at point C. The individual who is indifferent between locating at the discontinuity C (and applying for SSDI) and locating on the segment AE (and not applying for SSDI) will determine the amount of work hours associated with a breakeven point; in the case where there are no stigma costs, the breakeven point occurs at point F in the figure. In general, the breakeven point will be a function of the stigma cost, ϕ. All individuals between the breakeven point F(ϕ) and the SGA threshold (*H**) have an incentive to reduce their hours of work to exactly the SGA threshold (*H**) and apply for SSDI benefits. Thus, the model

[30] This is the budget constraint that applies once the individual has completed the TWP. During the TWP the budget constraint would be: $Y = wH + N + b$ if on SSDI.

predicts some amount of pooling at the discontinuity at point C.[31] Individuals with large stigma costs, however, would locate on segment AB close to H^* (or even on segment BE).

Figure 3.5
SSDI Budget Constraint Under the Current Policy



Given the budget constraint, we can write the conditional labor supply equation as

$$(3.7) \qquad H = H_0(w,N) \text{ if } P^* \leq 0$$
$$H = \max(H^*, H_1(w,N,b)) \text{ if } P^* > 0,$$

where H^* is the number of hours that, given the individual's wage rate, corresponds to earnings exactly at the SGA level. Individuals will participate in SSDI (P = 1) if $P^* > 0$ and will not participate (P = 0) if $P^* \leq 0$, where

$$(3.8) \quad P^* = V[1,w,N,b] - V[0,w,N]$$

and *V[P,.]* is the indirect utility function, evaluated at the optimal choices of hours of work and income given participation in SSDI (P).

Equations 3.7 and 3.8 can be estimated using maximum likelihood methods once one assumes a functional form for either the labor supply or the indirect utility function (since one implies the other), specifies a role for observable characteristics such as age or race, and

[31] Schimmel, Stapleton, and Song (2010) do not find strong evidence of pooling just under the SGA threshold. However, pooling is a phenomenon that is hard to detect in administrative data. This is because earnings are measured at the annual level, but it may be the *monthly* SGA level that is relevant for beneficiaries. In that case, beneficiaries who earn just under the SGA for several months (but not the whole year) would not appear to pool in annual data. In addition, beneficiaries may have limited control over their earnings, which would limit the amount of observed pooling.

assumes a distribution for unobservable characteristics. Specifically, the log likelihood function would be

$$
\begin{aligned}
L = & \sum \log[\Pr(P^* > 0, h = H_1)] \text{ (participants working below SGA)} \\
& + \sum \log[\Pr(P^* > 0, h = H^*)] \text{ (participants working at SGA level)} \\
& + \sum \log[\Pr(P^* \leq 0, h = H_0)] \text{ (nonparticipants).}
\end{aligned}
$$

The income and substitution elasticities of program participation with respect to benefits will depend on the estimated structural parameters of the utility function. For example, Moffitt (1983) chooses a linear labor supply equation and allows the participation equation and labor supply equations to each have additive and normally distributed error terms. Both the constant term in the labor supply equation and the stigma parameter (ϕ) are functions of individual observed characteristics. Similar assumptions have been adopted in the literature analyzing piecewise-linear constraints (see Moffitt, 1986, for a review).

Estimates of the model could be obtained using data on hours of work, wage rates, SSDI participation, and individual characteristics from the Survey of Income and Program Participation (SIPP), which is linked to SSA's 831 File.[32] Alternatively, one could express the problem in terms of earnings instead of hours; in this case, one could use only SSA administrative data without matched survey data. This second approach has the advantage that it takes into account that individuals may be limited in their choice of hours. At the same time, individuals may find it advantageous to negotiate their wage slightly downward in order to satisfy the SGA requirement. This phenomenon would be difficult to capture in a model with hours of work, as it is considered the only choice variable for labor force participation.

The maximum likelihood approach has been criticized for its dependence on distributional and functional form assumptions. In particular, MaCurdy, Green, and Paarsch (1990) argue that, in order to achieve coherency of the model, the maximum likelihood approach limits the range of elasticities that can be obtained. This is especially problematic in the presence of non-convex preferences and raises the concern that estimated policy effects may be driven by the constraints of the model rather than by the data. As a result, alternative estimation approaches have been proposed; Blundell, Duncan, and Meghir (1998) use a difference-in-differences approach, and Saez (2010) develops an estimator exploiting the amount of bunching (analogous to pooling in our case) at discontinuities. However, in using these reduced form techniques, one would sacrifice the ability to do counterfactual policy evaluations and simulations. An alternative option could be to

[32] The SIPP is a series of national panels, with sample sizes ranging from approximately 14,000 to 36,700 interviewed households, beginning in the years 1984–1993, 1996, 2001, 2004, and 2008. Each SIPP panel runs between 2.5 and 4 years. During 1984–1993, a new panel of households was introduced each year in February. A 4-year panel was introduced in April 1996; a 3-year panel was started in February 2000 but canceled after 8 months for budget reasons; a 3-year panel was introduced in February 2001; and a 2.5-year panel was started in February 2004. The SIPP includes ample coverage of the large change in the nominal SGA in 1990, and more sparse coverage of the period surrounding the large change in 1999. The change in 1999 occurs near the end of the 4-year panel begun in 1996, although an additional 8 months' worth of data can be obtained from the panel begun in early 2000.

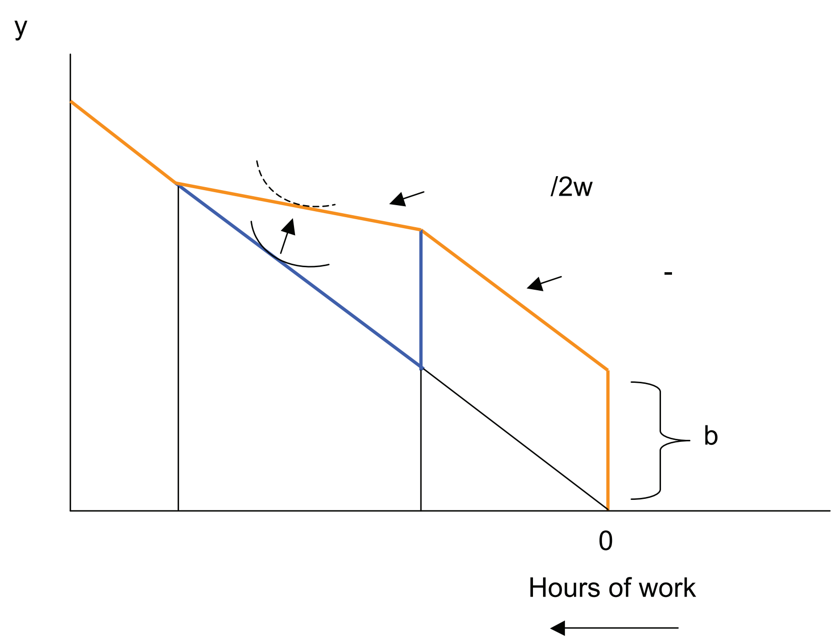
combine nonparametric regression techniques with maximum likelihood analysis, as in van Soest, Marcel, and Gong (2002).

Once one has obtained estimates of the utility function parameters, including ϕ , one can apply them to the hypothetical budget constraint under the proposed benefit offset to simulate who would apply for SSDI under the new program. The new budget constraint, represented in orange in Figure 3.6, will have the following form:

$$Y = wH + N \text{ if not on SSDI, or}$$
$$Y = wH + N + b - \frac{1}{2}(wH - SGA)\cdot 1[wH \geq SGA] \text{ if on SSDI.}$$

An estimate of induced entry can then be obtained by differencing application rates under the current policy and under the proposed benefit offset.

Figure 3.6
SSDI Budget Constraint Under the Benefit Offset



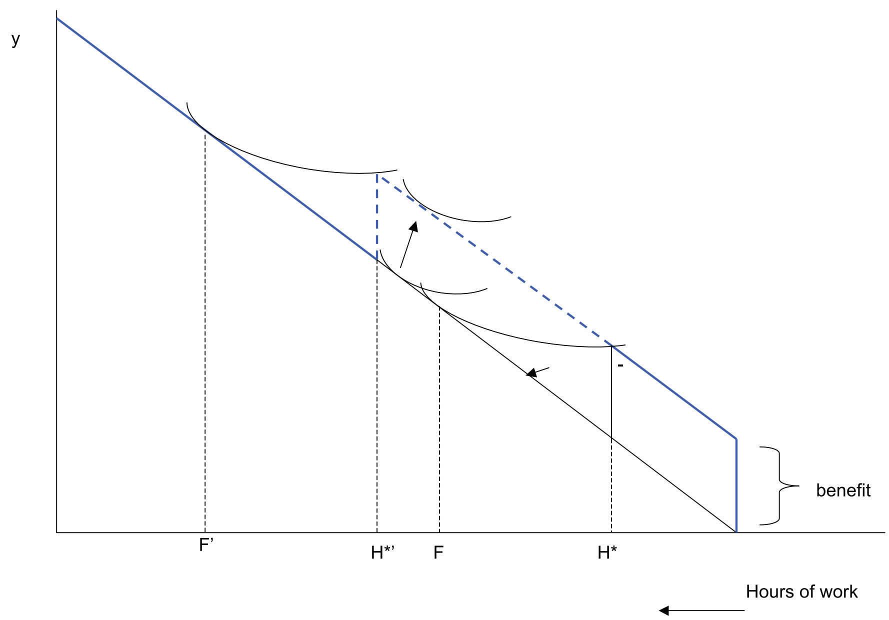
Incorporating SGA Changes into the Research Design

As stated above, SSA has increased the SGA threshold several times during the past decades, and relative SGA levels vary considerably across states. This offers an opportunity to check the external validity of the model estimates by doing out-of-sample prediction of actual induced entry effects derived from SGA changes and comparing them with the predicted effect from Equation 3.5 (i.e., the estimated parameter β). In addition, one can make use of this information about the effect of changes in the SGA to update the estimates and relax some of the assumptions required in the maximum likelihood approach. Maximum likelihood estimates of Equations 3.7 and 3.8 use individual variation in nonwage income

(N) to identify income effects. This variation, however, can be correlated with the unobservables in the model and is therefore potentially endogenous.

As we can see in Figure 3.7, with sufficiently big changes in the SGA level, such that the new SGA level (H*') is above the old breakeven point (F), one can isolate the income effect for those individuals who previously located between the new SGA level (H*') and the old breakeven point (F). These individuals were not previously program participants but were induced to enter the program as a result of the change in the SGA. However, since the tradeoff between leisure (work hours) and consumption (income) remains the same under the old and new regimes, any observed labor supply change is due only to the additional income provided by the receipt of disability benefits. In contrast, observed labor supply changes for those who previously located between the new breakeven point F' and the new SGA level H*' reflect a combination of *both* income and substitution effects. Subtracting the pure income effect from this combined effect gives the substitution effect.

Figure 3.7
SSDI Budget Constraint Before and After a Large Change in the SGA Threshold



Using information on the income effect provided by changes in the SGA threshold allows one to relax the dependence on variation in nonwage income. One method for exploiting these policy changes is to maximize an extended version of the likelihood function including contributions of individuals who face different budget constraints, in different years and states. Alternatively, consider the estimated effect of SGA changes in applications or enrollment rates derived from the estimates of the parameter β in Equation 3.5; one could estimate the system of Equations 3.7 and 3.8 using a method of simulated moments (or indirect inference) estimation strategy that includes this parameter as a moment to match directly. This latter method exploits the fact that the maximum likelihood estimator is equivalent to a method of moments estimator that minimizes the expected scores of the likelihood function. Thus, the moment provided by the parameter β can be combined with the moments derived from the likelihood. A similar strategy was followed by Manoli, Mullen, and Wagner (2009) in their study on pension benefits and retirement decisions.

Incorporating Results from the BOND Project

SSA is currently conducting a benefit offset demonstration project on a sample of SSDI beneficiaries, with the objective of determining possible labor supply effects of the $1-for-$2 offset. The demonstration project is currently designed as a two-stage project. In the first stage, SSDI beneficiaries participating in the demonstration will be randomly assigned to one of three groups. The first group will consist of beneficiaries who will be enrolled into the benefit offset program; their benefits will be reduced by $1 for every $2 earned above the annualized SGA earnings level after they have completed the TWP and Grace Period. The second group will be offered the chance to volunteer in a second stage of the project, in which intensive benefits counseling services will be provided. The third group will serve as the control group for stage one of the project.

Although the demonstration project was not designed to measure possible induced entry (as the benefit offset demonstration will only be applied to current SSDI beneficiaries), any observed changes in labor supply of demonstration participants would be informative about the very income and substitution elasticities needed to estimate induced entry. Therefore, these estimated elasticities would provide additional information to estimate the system of Equations 3.7 and 3.8 using a method of simulated moments estimation strategy, as described above. Note, however, that this additional information would only complement the research design outlined above, as it will not provide information about the decision to *participate* in SSDI. Data on actual application decisions is necessary to properly identify the stigma parameter (which governs take-up) in the model.

Evaluative Criteria

Internal Validity:

- To forecast the possible entry effects arising from the introduction of a benefit offset, one would need to rely on distributional and functional form assumptions. There are, however, opportunities for relaxing some of these assumptions by exploiting the information provided by SGA changes and the benefit offset demonstration project.

External Validity:

- Although changes in the SGA threshold are similar to the proposed benefit offset in some ways, it is not obvious whether they are similar enough. For example, because the SGA threshold figures prominently in the criteria for program eligibility and ongoing entitlement, applicants may be more aware of and hence more responsive to changes in the SGA than they would be to a change in the benefit offset amount.

Flexibility:

- It would be easy to test responsiveness to other benefit offset rates by modifying the budget constraint in the decision problem used to simulate behavior under hypothetical rules.

Economy:

- Estimates could be obtained at low cost, given that this method would rely mostly on administrative data records that are already available. The key component of cost is therefore labor of research staff.

Speed:

- Estimates could be obtained in a relatively short period of time.

CHAPTER FOUR

Recommendations and Conclusions

The goal of this study has been to provide the Social Security Administration with a set of research design options for evaluating the effect of a change in the Social Security Disability Insurance program resulting from a benefit offset policy. Under the current program rules, SSDI benefits are eliminated completely if, after their Trial Work Period, beneficiaries earn more than the threshold of Substantial Gainful Activity, which is currently set at $1,000 per month. Under the proposed benefit offset policy, SSDI benefits would be reduced by $1 for every $2 of earnings above a disregard amount. The Ticket Act identified induced entry as an important component of potential costs that could arise from the introduction of a benefit offset, and mandated that SSA provide an estimate of induced entry before such a policy could be enacted.[33]

There are a number of methodological challenges to developing a credible estimate of induced entry. Because the population of SSDI applicants, as well as medically eligible nonbeneficiaries, is very small, identifying marginal applicants in an analysis of induced entry is difficult. In addition, because the benefit offset may represent a small change in the value of program participation to many potential applicants, for whom working is difficult by definition, the expected impact of the offset is also likely to be small. Previous estimates suggest that it may be no higher than 3–4 percent and as low as 2 percent. As a result, the sample necessary to identify such an effect with statistical precision will generally be very large. Finally, since such a policy has never been enacted, one cannot simply extrapolate from estimates of induced entry from previous policy changes.

We examined a wide variety of potential research designs and identified two designs that are especially promising to estimate induced entry in this context. These include a research design using stated preferences (SP) and a research design using past policy (PP) changes in a simple structural framework. We evaluated the research designs on the basis of five criteria: internal validity, external validity, flexibility, economy (cost), and speed. Both designs have clear strengths as well as some weaknesses. We briefly summarize the essential features of each design, and then conclude with a comparison of their performance on these five criteria.

---

[33] The other components include cost savings from reduced benefits paid to current beneficiaries who increase work effort, and incurred costs associated with the possibility that some recipients, those who otherwise would be terminated for work above the SGA threshold, may now remain in the program.

Research Design Using Stated Preferences

This method consists of presenting respondents with a set of scenarios describing different states of the world (i.e., different sets of program rules) and asking them to state whether they would apply for SSDI benefits under these varying conditions. The scenarios are characterized by either real or hypothetical attributes (or a mix of both), such as the benefit offset rate or earnings disregard level, and it is possible to conduct *randomized* choice experiments by randomly varying hypothetical attributes over respondents. Respondents may be asked to consider multiple scenarios, or profiles. The SP design requires new data collection, but economic theory offers guidance as to how to target the sample frame to identify marginal applicants who would benefit economically from the benefit offset policy. In addition, by specifying a statistical model for SSDI take-up as a function of proposed program parameters (e.g., offset rate, disregard level), current program parameters (e.g., the SGA level), and individual characteristics (e.g., health) and asking respondents to consider multiple scenarios, or profiles, one can reduce costs substantially with few additional assumptions. Introducing health as an attribute that can be specified and varied across respondents further reduces costs by reducing the number of respondents who need to be sampled to attain sufficient statistical power.

Research Design Using Past Policy Changes in a Simple Structural Framework

This research design leverages past changes in the SGA threshold. If there is evidence of induced entry from past policy changes, one could relate the SGA-induced entry effect to the benefit offset setting using a simple structural framework. Specifically, one could specify a utility maximization problem where individuals jointly determine their labor supply and decide whether to participate in the SSDI program. Estimates of the key behavioral parameters of such a model could be obtained using maximum likelihood methods once one assumes a functional form for labor supply (or, equivalently, the indirect utility function), specifies a role for observable individual characteristics, and assumes a distribution for unobservables. Once estimates of the utility function parameters have been obtained, they can be applied to the hypothetical budget constraint under the proposed benefit offset to simulate who would apply for SSDI under the new program. Past policy changes offer valuable opportunities to relax some of the assumptions required in the maximum likelihood approach. Similarly, the Benefit Offset National Demonstration (BOND) project may provide additional, complementary information that can be used to test and estimate the model.

Comparing the Two Approaches

Table 4.1 summarizes and compares the performance of the two approaches based on five criteria: internal validity, external validity, flexibility, economy (cost), and speed. We discuss each criterion in turn below.

Internal Validity

The ability of a research design to approximate the true induced entry effect depends on the plausibility of the identifying assumptions. The key assumption of the SP approach is that respondents are able to forecast their behavior under new and unfamiliar conditions. However, a significant advantage is that the *researcher* is not forced to make assumptions about how respondents make decisions or how they perceive (accurately or not) the state of the world (e.g., opportunities for employment). By contrast, the PP design relies on strong distribution and functional form assumptions characterizing individuals' opportunity sets and preferences. While past policy changes provide plausibly exogenous variation in program rules that can be exploited to relax some of these assumptions, it is still necessary to impose some structure in order to translate past participation changes into *hypothetical* participation changes due to a completely new policy. A drawback of both approaches is that neither provides a satisfactory way of dealing with potential social multiplier effects that may amplify individual-level responses (i.e., if individuals are more likely to apply for SSDI because they know more people on SSDI, such as family, friends or neighbors). While the reduced form analysis of the PP design captures social multiplier effects in the estimated entry effect, it is unclear how to disentangle these effects in the structural analysis.[34] Generally, since the SP design requires fewer and weaker assumptions, we determined that it dominates the PP approach on this criterion.

External Validity

External validity describes the extent to which the estimate is applicable to the SSDI induced entrant population. The PP design uses data on actual SSDI applications, while the SP design must necessarily target a broader population. Although past SSDI applicants are not equivalent to the marginal applicants under a benefit offset—in particular, past SGA changes relate to both program eligibility and ongoing entitlement, while the benefit offset relates only to ongoing entitlement—they are probably the closest possible match. Thus, unless pilot testing of the SP approach were to show that respondents can *exactly* forecast their behavior under unfamiliar health conditions, the PP design edges out the SP design on this criterion.

Flexibility

Both approaches offer a great deal of flexibility in estimating induced entry for benefit offset policies with different offset rates and/or earnings disregard levels. The SP approach allows one to present respondents with a range of hypothetical profiles, chosen randomly, and allows estimation of a statistical model that even allows one to forecast behavior under scenarios not explicitly asked about. Similarly, imposing some structure on potential applicants' preferences for labor supply and program participation allows one to estimate

---

[34] One potential solution to this problem in both approaches is to use estimates from the literature to apply a social multiplier to the induced entry estimate. For example, Rege, Telle, and Votruba (2009) empirically investigate social interaction effects in disability insurance participation in Norway. They find that a 1 percentage point increase in the participation rate of previously employed neighbors increases the subsequent four-year entry rate of older workers by about one-half a percentage point.

entry under a virtually unlimited array of budget constraints. However, the PP design is not as flexible as the SP design in that it cannot vary attributes that do not affect the budget constraint. For example, whereas an SP question could be developed to consider a time limit on the benefit offset, this cannot be incorporated into the static model specified in the PP approach. Additionally, it is not clear what elements of program rules, including eligibility criteria and rules governing benefit receipt once accepted to the program, are known by potential SSDI applicants, and how precisely they are understood. While the SP design can incorporate varying levels of information disclosed to respondents to produce a range of estimates, the PP design assumes that respondents' information set can be fully characterized in the budget constraint. In this sense, the SP approach dominates the PP approach.

Economy and Speed

In contrast to the PP design, which makes use of readily available survey and/or administrative data, the SP approach requires new data collection. Pilot testing and fielding a survey imposes significant added costs (in terms of both time and money) in addition to the labor cost of research staff common to both designs. In this respect, the PP design is less expensive and will produce estimates more quickly than the SP design.

Summary

As noted earlier, both research designs are capable of providing SSA with credible estimates of induced entry into the SSDI program resulting from a benefit offset. In addition, both approaches offer a great deal of flexibility and allow for a range of estimates that would provide valuable insight into how potential SSDI applicants make decisions regarding program participation. While both research designs produce partial-equilibrium "steady-state" estimates of induced entry, they yield parameter estimates that could be used to forecast entry over time accounting for changing economic and demographic conditions (e.g., trends in population aging, health, and labor demand). In a head-to-head comparison, there is no clear winner, as both research designs are strongest on different criteria. Whereas the SP design may offer slightly greater flexibility and require fewer and weaker assumptions, the PP design is cheaper, faster, and uses data on individuals who most closely approximate marginal entrants. As SSA has stressed a strong desire for a *range* of plausible induced entry estimates, one promising avenue for further research is to implement both research designs and compare the results.

Table 4.1. Summary and Comparison of Research Designs

Criterion	Research Design Using Stated Preferences	Research Design Using Past Policy Changes in a Simple Structural Framework	Which Research Design Performs Better?
Internal Validity—evaluates plausibility and robustness of assumptions necessary for identification	Assumes respondents are able to accurately forecast their behavior under new and unfamiliar (policy) conditions Does not require correct characterization of entire decision environment	Relies on distribution and functional form assumptions However, there are opportunities for relaxing some assumptions (e.g., exogeneity of nonwage income) by exploiting SGA changes and BOND	The stated preferences (SP) design requires fewer and weaker assumptions on unobserved expectations and opportunities of potential entrants Winner: SP
External Validity—applicability to the SSDI induced entrant population	Requires pilot testing to determine the extent to which non–medically eligible respondents can forecast their behavior under unfamiliar *health* conditions	Makes use of matched survey-administrative data, or administrative data alone, on past SSDI applications	The past policy (PP) design uses data on actual SSDI applications, compared with less precisely targeted sample required for SP Winner: PP
Flexibility—extent to which the design offers the possibility of varying policy parameters and relaxing assumptions about the decision environment	Can easily vary proposed policy parameters Can vary other attributes of the scenarios (e.g., detail of information presented) to relax assumptions about decision environment	Can easily vary proposed policy parameters, provided they affect the budget constraint Cannot easily relax assumptions about decision environment	The SP design offers increased flexibility to vary policy parameters and assumptions about decision environment Winner: SP
Economy—cost of implementing the research design with sufficient statistical power to detect an entry effect of plausible size	Requires new data collection, although SSA administrative records provide inexpensive sample frame Imposing statistical structure and defining health as attribute reduces necessary sample size	Makes use of administrative data, which is inexpensive	Collecting new data adds costs Winner: PP
Speed—the time it would take to produce an estimate	Requires new data collection, including pilot testing of survey instrument	Makes use of administrative data, which is readily available	Collecting new data adds time Winner: PP

References

Autor, David, and Mark Duggan. 2003. The Rise in the Disability Rolls and the Decline in Unemployment. *The Quarterly Journal of Economics* 118(1): 157–205.

Benitez-Silva, Hugo, Moshe Buchinsky, and John Rust. 2006. Induced Entry Effects of a $1 for $2 Offset in SSDI Benefits. Manuscript, State University of New York–Stony Brook.

Black, Dan, Daniel Kermit, and Seth Sanders. 2002. The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Bust and Boom. *The American Economic Review* 92(1): 27–50.

Blundell, Richard, Alan Duncan, and Costas Meghir. 1998. Estimating Labor Supply Responses Using Tax Reforms. *Econometrica* 66: 827–861.

Burkhauser, Richard, J. S. Butler, and Gulcin Gumus. 2004. Dynamic Programming Model Estimates of Social Security Disability Insurance Application Timing. *Journal of Applied Econometrics* 19: 671–685.

Coe, Norma B., Gregory Acs, Robert I. Lerman, and Keith Watson. 1998. Does Work Pay? A Summary of the Work Incentives Under TANF. Policy Brief A-28. Washington, D.C.: The Urban Institute.

Delavande, Adeline, and Susann Rohwedder. 2009. Changes in Spending and in Labor Supply in Response to a Social Security Benefit Cut: An Analysis Based on Stated Choice Data. Michigan Retirement Research Center Working Paper 2008-182.

Dwyer, Debra, Jianting Hu, Denton R. Vaughan, and Bernard Wixon. 2001. Counting the Disabled: Using Survey Self-Reports to Estimate Medical Eligibility for Social Security's Disability Programs. ORES Working Paper No. 90. Washington, D.C.: Social Security Administration, Office of Policy, Office of Research, Evaluation, and Statistics.

Frey, William, Holly Hollomon, Jarnee Riley, Rene Gonin, Judy Kasper, and Graham Kalton. 2002. National Study of Health and Activity Pilot Study. Evaluation Report 3: Screening for Disability. WESTAT Report.

Halpern, Janice, and Jerry Hausman. 1986. Choice Under Uncertainty: A Model of Applications for the Social Security Disability Insurance Program. *Journal of Public Economics* 3(2): 131–161.

Hopkins, Daniel J., and Gary King. 2010. Improving Anchoring Vignettes: Designing Surveys to Correct Interpersonal Incomparability. Mimeo.

Hoynes, Hilary W., and Robert Moffitt. 1999. Tax Rates and Work Incentives in the Social Security Disability Insurance Program: Current Law and Alternative Reforms. *National Tax Journal* 52(4): 623–654.

Kapteyn, Arie, James P. Smith, and Arthur van Soest. 2007. Vignettes and Self-Reports of Work Disability in the United States and the Netherlands. *The American Economic Review* 97(1): 461–473.

Kreider, Brent, and Regina T. Riphahn. 2000. Explaining Applications to the U.S. Disability System: A Semiparametric Approach. *The Journal of Human Resources* 35: 82–115.

Leibman, Jeffrey B., and Erzo Luttmer. 2009. The Perception of Social Security Incentives for Labor Supply and Retirement: The Median Voter Knows More than You'd Think. As of September 9, 2010: http://www.rand.org/content/dam/rand/www/external/labor/seminars/adp/pdfs/2009_luttmer.pdf

Louviere, Jordan, David Hensher, and Joffre Swait. 2002. *Stated Choice Methods: Analysis and Application.* Cambridge University Press. New York.

MaCurdy Thomas, David Green, and Harry Paarsch. 1990. Assessing Empirical Approaches for Analyzing Taxes and Labor Supply. *Journal of Human Resources* 25: 415–490.

Manoli, Dayanand, Kathleen Mullen, and Mathis Wagner. 2009. Pension Benefits and Retirement Decisions: Income vs. Price Elasticities. Mimeo.

McLaughlin, James R. 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. SSA Memorandum.

Moffitt, Robert. 1983. An Economic Model of Welfare Stigma. *The American Economic Review* 73(5): 1023–1035.

Moffitt, Robert. 1986. The Econometrics of Piecewise-Linear Budget Constraints: A Survey and Exposition of the Maximum Likelihood Method. *Journal of Business & Economic Statistics* 4(3): 317–328.

Moffit, Robert. 1990. The Econometrics of Kinked Budget Constraints. *Journal of Economic Perspectives* 4(2): 119–139.

Mok, Wallace K. C., Bruce D. Meyer, Kerwin Kofi Charles, and Alexandra C. Achen. 2008. A Note on "The Longitudinal Structure of Earnings Losses Among Work-Limited Disabled Workers." *Journal of Human Resources* 43(3): 721–728.

Organisation for Economic Co-operation and Development. 2003. Transforming Disability into Ability: Policies to Promote Work and Income Security for Disabled People. Paris: OECD Publishing.

Public Law 106-170. Ticket to Work and Work Incentives Improvement Act of 1999. December 17, 1999.

Rege Mari, Kjetil Telle, and Mark Votruba. 2009. Social Interaction Effects in Disability Pension Participation: Evidence from Plant Downsizing. UIS Working Papers in Economics and Finance No. 2009/30. University of Stavanger, Norway.

Schimmel, Jody, David Stapleton, and Jae Song. 2010. How Common Is "Parking" Among Social Security Disability Insurance (SSDI) Beneficiaries? Evidence from the 1999 Change in the Level of Substantial Gainful Activity (SGA). Ann Arbor, Mich.: Michigan Retirement Research Center, University of Michigan.

Saez, Emmanuel. 2010. Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy* 2(3): 180–212.

Social Security Administration. 2009. 2009 Red Book: A Summary Guide to Employment Supports for Individuals with Disabilities under the Social Security Disability Insurance and Supplemental Security Income Programs. As of September 9, 2010:
http://www.ssa.gov/redbook/eng/redbook.pdf

Smith, Dylan M., Norbert Schwartz, Todd R. Roberts, and Peter A. Ubel. 2006. Why Are You Calling Me? How Study Introductions Change Response Patterns. *Quality of Life Research* 15: 621–630.

Tuma, Nancy. 2001. Approaches to Evaluating Induced Entry into a New SSDI Program with a $1 Reduction in Benefits for Each $2 in Earnings. Manuscript, Stanford University.

van Soest, Arthur, Das Marcel, and Xiaodong Gong. 2002. A Structural Labor Supply Model with Flexible Preferences. *Journal of Econometrics* 107: 345–374.

van Soest, Arthur, Arie Kapteyn, and Julie Zissimopoulos. 2006. Using Stated Preferences Data to Analyze Preferences for Full and Partial Retirement. DNB Working Paper 081/2006, De Nederlandsche Bank, Amsterdam.