

**Challenges in Conducting Randomized Field Trials: The Experience of the
Wisconsin SSDI Employment Pilot**

**Barry S. Delin
Christopher W. Sell
Stout Vocational Rehabilitation Institute
University of Wisconsin – Stout**

**Anne E. Reither
Utah State University**

**American Evaluation Association Annual Meeting
Baltimore MD
November 8, 2007 (Revised December 2007)**

The authors thank the managers and staff at the Pathways Projects, Office of Independence and Employment, Wisconsin Department of Health and Family Services for their cooperation and support. “Pathways” conducts the Wisconsin SSDI Employment Pilot under contract SS00-05-60008 from the U.S. Social Security Administration. The same contract supports the SSDI-EP’s evaluation. The descriptions and interpretations in this paper are those of the authors and do not necessarily reflect those of the Wisconsin Department of Health and Family Services, the Social Security Administration, or of the institutions where the authors are employed.

Introduction

This paper describes challenges encountered in conducting an evaluation of the Wisconsin SSDI Employment Pilot (SSDI-EP). Inevitably the paper will also describe challenges in program design and implementation, though this paper focuses on the perspectives of the evaluators rather than those of program staff, participants or other stakeholders.

The SSDI-EP is one of four pilot projects that the Social Security Administration has authorized and funded to do preliminary testing of a benefit offset provision for the Social Security Disability Income (SSDI) program. The benefit offset provision is expected to provide an incentive that will encourage better employment outcomes among SSDI beneficiaries. SSA required that the pilots utilize random assignment. While the basic intervention approach and eligibility rules are essentially common across the four pilots, the SSDI-EP is different from the other three pilots in having substantially more decentralized enrollment, service provision, and data collection processes.

Though a cliché, the causes of human behavior, whether individual or collective, are often extremely complex. It can be exceedingly difficult to isolate the role of any putative cause or group of causes. Yet, to the extent that institutions, including governments, seek to design and implement policies for the purpose of changing how people act and, ultimately, the outcomes arising from those actions, it is useful for those institutions (and hopefully their “stakeholders”) to know whether and to what extent policy changes motivated their intended effects.¹ Experimental designs, i.e. those utilizing random assignment, are widely recognized as the best approach for assessing whether changes to policy (and/or program) are “effective” in the sense of producing expected outcomes.

Experimental design involving random assignment is regularly characterized as “the gold standard” for assessing social impacts, though there is a substantial literature discussing both the practical and the methodological limitations of the approach.²

¹ Of course, it can be every bit as important to learn about the unintended consequences, whether positive or not.

² There is a voluminous literature in this area. For a fairly non-technical starting point see St. Pierre, Robert G. 2004. “Using Randomized Experiments” and Hatry, Harry P. and Newcomer, Kathryn E. 2004. “Pitfalls of Evaluation” both in eds. Wholey, Joseph S. et. al. Handbook of Practical Program Evaluation: Second Edition. San Francisco, CA: Jossey-Bass, pp. 150-175 and 547-570 respectively. For a more sustained discussion of the superiority of social experiments utilizing random assignment to other methods see Orr, Larry L. 1998. Social Experiments: Evaluating Public Programs with Experimental Methods. Thousand Oaks, CA: Sage. There is also an extensive literature concentrating specifically on employment issues and thus directly relevant to the benefit offset pilots. For a recent and brief discussion of this literature and of the implications of whether one is interested in measuring effect size or just a decision as to whether to adopt a tested policy or program, see Greenberg, David H., Michalopoulos, Charles and Robins, Philip K. 2006. “Do Experimental and Nonexperimental Evaluations Give Different Answers about the Effectiveness of Government-Funded Training Programs?” Journal of Policy Analysis and Management, 25, No. 3, pp. 522-552. Finally, as an example of a direct challenge to the inherent superiority of experimental design for estimating the outcomes of employment programs, see Heckman, James J. Hotz, V. Joseph and Dabos, Marcelo (1987) “Do We Need Experimental Data to Evaluate the Impact of Manpower Training on Earnings?” Evaluation Review, 11, No. 4, pp. 395-427.

Indeed, those who commission evaluations, especially for governmental purposes, may see even greater virtue in experimental design than those who design and perform evaluations. For example, in a 2004 report released several months prior to SSA's award of contracts to the four state pilots, the Government Accountability Office (GAO) expressed concern over the quality of the evaluations of Social Security Administration (SSA) demonstration projects, in part because of SSA's past practice of not always requiring use of rigorous experimental methodologies. Thus, in its recommendations, GAO charged SSA with using the best available methodology, i.e. experimental design, whenever feasible. In its response to GAO, SSA fully agreed with this recommendation.³ Whether in response to the GAO recommendations or not, SSA required the states implementing a Benefit Offset Pilot to use random assignment and tightly limited the states' discretion over the intervention's core features and eligibility requirements.

Nonetheless, there are very real challenges to implementing social experiments. One very large group of challenges can be termed implementation problems and can affect both the conduct of the experiment and its evaluation. Thus, process evaluations are often designed and conducted in concert with outcomes evaluations to learn (among other things) whether the intervention was sufficiently "present" to allow meaningful evaluation of outcomes. If the intervention is not adequately implemented, random assignment by itself will not provide useful information about the intervention's role in producing observed outcomes.⁴

Within the broad range of possible implementation problems is a class that has proven salient to the SSDI-EP and the other Benefit Offset Pilots, especially as part of the rationale for the pilots has been to inform the development of a national benefit offset demonstration. That class of challenges focuses on the concept of "evaluability," that is whether the intervention itself and the theory as to why that intervention is expected to work are well enough developed so that meaningful outcomes measurement can take place.⁵

We will argue that important elements of the SSDI-EP have not been fully developed and that some of the deficiencies are of a character to threaten the capacity to fulfill evaluation goals. In some cases these issues affect the capacity to complete aspects of the evaluation, though without compromising the ability to perform an overall evaluation of participant outcomes. However, we will identify one area of insufficient program development that has a strong potential to seriously compromise the viability of the SSDI-EP outcomes evaluation and, by implication, those of the other three pilots.

³ Government Accountability Office, November 2004. Improved Processes for Planning and Conducting Demonstrations May Help SSA More Effectively Use Its Demonstration Authority. (GAO-05-19) Washington, DC: Government Accounting Office. In particular, see pp. 18-22.

⁴ Failure to properly implement random assignment is itself an important type of implementation problem.

⁵ For a general discussion of the issues involved, see Wholey, Joseph, F. 2004. "Evaluability Assessment" in eds. Wholey, Joseph S., et al. Handbook of Practical Program Evaluation: Second Edition. San Francisco, CA: Jossey-Bass, pp. 33-62. For a more targeted discussion of the issue of when policy or program can be judged as ready for meaningful evaluation, see Julnes, George and Rog, Debra J. 2007. "Pragmatic Support for Policies on Methodology," New Directions for Evaluation, No. 113, pp. 129-147.

Still, there is no bright line that identifies the border between threats and the fruition of threats that would render the SSDI-EP's outcomes evaluation unproductive; it is a matter of judgment. We will conclude that, with the exception of problems pertaining to the administration and tracking of the benefit offset usage, the implementation challenges are not so severe as to prevent meaningful evaluation of participant outcomes (though these challenges will structure/limit the potential applicability of what can be learned). Moreover, even those challenges with the potential to seriously threaten "evaluability" are not yet beyond effective mitigation. Finally, in the unfortunate event that problems in the administration and tracking of offset usage are not ameliorated, the SSDI-EP, both through its formal process evaluation and the lessons learned by program staff, will provide useful information for future efforts toward improving employment outcomes for persons with severe disabilities.⁶ What will be lost is the ability to get reasonably good estimates of the benefit offset's effects on employment outcomes.

Background: Program and Evaluation

As a prelude to describing and explicating the challenges involved in evaluating the SSDI-EP, it is necessary to provide some background about both the demonstration and the evaluation plan. The SSDI-EP shares a number of features with the pilots in Connecticut, Utah, and Vermont. However, there are also important differences, most notably the SSDI-EP's decentralized structure for enrolling participants, providing or arranging services, and maintaining contact with participants for both operational and research purposes.

SSA Requirements and Options

SSA required that all of the pilots provide the same basic intervention to those participants randomly assigned to the treatment group. The benefit offset consists of a reduction of one dollar in the monthly SSDI benefit amount for every two dollars of earnings over the Substantial Gainful Activity (SGA) level.⁷ Access to the offset was restricted to a period beginning three months after the completion of a treatment group member's Trial Work Period (TWP) through the seventy-second month following TWP completion, irrespective of whether that period occurred during the period the states were contracted to operate the pilots. Those in the treatment group also received protection against loss of SSDI eligibility through suspending scheduled medical Continuing Disability Reviews and, for those past the end of their Extended Period of Eligibility (but still viewed as disabled), restoration of their SSDI cash benefit, subject to the application of the offset provision.

SSA also specified the basic eligibility requirements. Participants had to be volunteers and enrolled through an informed consent process that met SSA required

⁶ In this context "severe disability" refers to impairments that meet or potentially meet eligibility requirements for a Social Security disability program or Medicaid.

⁷ SGA, in 2007, was \$900 a month under standard rules. A SSDI beneficiary who has completed their Trial Work Period will receive no cash benefit in any month when her/his earnings are at or above SGA. For the pilots, SSA decided not to apply the offset to any portion of the SSDI benefit for a treatment group member's dependents.

standards. Enrollment would be limited to adult SSDI beneficiaries who were receiving their benefits as a consequence of their own earnings records.⁸ Beneficiaries eligible for SSI (Supplemental Security Income) benefits were also excluded.⁹ While starting or completing the TWP was not an eligibility requirement, a beneficiary who had completed her/his TWP seventy-two or more months prior to attempting enrollment would not be eligible to enroll. Finally, SSA precluded enrollment of beneficiaries within twenty-four months of an Expedited Reinstatement. Within these constraints, SSA was willing to allow states to have additional eligibility criteria.¹⁰

SSA was far less prescriptive about how the states organized their recruitment, enrollment processes, service provision, and participant contact and tracking processes. SSA did insist that informed consent processes meet certain standards, especially in describing the benefits, risks, and obligations associated with participation in the treatment group. SSA also specified and subsequently amended certain operational procedures, especially those related to administering and monitoring the use of the offset itself. In addition, SSA decided that final decisions for treatment group members concerning project eligibility, continued qualification for SSDI during the pilots, the administration of the offset, and the identification of overpayments would be made in Baltimore, not at SSA local offices.¹¹

Nonetheless, through the contracting process, SSA had the ultimate say as to how states organized their pilot projects. In particular, SSA insisted that the benefit offset pilots had to be operated in a fashion so that no beneficiary would be disadvantaged by his or her participation in the pilot, whether by the offset itself or any other feature of a demonstration.¹²

The four pilots were required to conduct evaluations that would inform the design of the national demonstration as well as indicate the success of each pilot. Through the RFP, SSA identified a number of research questions that evaluation designs were expected to answer and/or contribute to answers that SSA would derive from information

⁸ In particular, this meant that DACs (Disabled Adult Children) and those entitled to DWB (Disabled Widow/Widower Benefits) were excluded from the pilots. This exclusion was added to those listed in the RFP relatively late in the planning process (May 2005).

⁹ However, SSA did not exclude SSDI beneficiaries receiving a state SSI supplement. It left discretion to do so to the states. Wisconsin chose not to exclude otherwise eligible beneficiaries who still received the supplement.

¹⁰ For example, SSA allowed the state projects to require that participants to have started or completed their TWP, to finish the TWP within specified time limits, remain state residents following enrollment, or have a minimum earnings level. Wisconsin chose not to adopt any of these options.

¹¹ Local SSA offices are directly involved in the resolution of overpayments.

¹² It is interesting to note that SSA often limited this injunction to “do no harm” to only the intervention group. For example, see the Social Security Administration (SSA) Solicitation #SSA-RFP-05-1003 “Benefit Offset Pilot Demonstration Project” September 28, 2004, p. 7. The fact the offset is not applied until after the TWP and the grace period is one direct application of the principle.

provided from the four pilots.¹³ Beyond this, SSA gave the states considerable latitude to plan their evaluations, subject to using random assignment and having a sample that “would be drawn from title II disability beneficiaries who are participating in statewide employment support programs.”¹⁴ Again, SSA could use the contracting process to limit the focus or scale of a state’s evaluation.

Wisconsin’s Choices: Program Structure

SSA contracted with the Office of Independence and Employment (OIE) within the Wisconsin Department of Health and Family Services (DHFS) to implement a benefit offset pilot in Wisconsin. OIE houses personnel from the University of Wisconsin system as well as from DHFS to design, test, and encourage the replication of efforts aimed at promoting fuller integration of persons with disabilities into the community. OIE based activities in these areas are branded as the “Pathways Projects” or sometimes just “Pathways.” These terms will be used interchangeably with OIE throughout this paper.

Project decentralization and the role of Pathways

Pathways made a number of choices within the framework of the SSA requirements as to how to organize the SSDI-EP. In many respects they resembled those made by the other states. The one area in which the SSDI-EP is essentially and critically different is the choice to have outreach/recruitment, enrollment, service provision, and significant data collection performed through a network of twenty-two contracted agencies.¹⁵ Most of these agencies are private non-profit entities, though

¹³ SSA’s list of overarching research questions can be found in Social Security Administration (SSA) Solicitation #SSA-RFP-05-1003 “Benefit Offset Pilot Demonstration Project” September 28, 2004, p. 7. SSA wanted information to help it identify (a) effective outreach, recruitment, and enrollment/informed consent processes, (b) effective means to maintain communication with participants and their attachment to the project, (c) effective means to provide employment supports, and (d) the characteristics of participants who benefit most from specific forms or combinations of employment supports. Pages 9-10 of the same document describe the specific areas that SSA wanted the state evaluations to include. These included comparison of differences between the treatment and control groups on a variety of employment outcomes, continued attachment to SSDI, and the impact of the service model. SSA also specified goals for process evaluation activities, some paralleling those identified in the “cross-state” questions and additional ones focused on identification of within state implementation challenges, participant perceptions of the intervention, and the extent participants refuse to cooperate with data reporting or leave the pilots.

¹⁴ Social Security Administration (SSA) Solicitation #SSA-RFP-05-1003 “Benefit Offset Pilot Demonstration Project” September 28, 2004, p. 8. This language, if taken literally, has an interesting implication: that SSA was not interested in using the pilots to assess take-up rates among the universe of qualified beneficiaries and, thereby, the potential fiscal impact of a 1:2 benefit offset.

¹⁵ After Wisconsin, the Utah pilot has the most decentralized structure. However, substantially fewer “partners” are involved and their relationships among them appear to be somewhat different. The Utah pilot appears to have adopted a network structure involving a substantial degree of co-management, though partners may have specific areas of responsibility. There is nothing comparable to the SSDI-EP system where twenty-two contracted agencies perform almost all of functions and activities involving direct contact with participants. To give a key

there are a small number of proprietary and governmental units as well. The key point is that Pathways has no direct authority over these agencies' operations. Though there were important Wisconsin specific reasons for choosing this approach, SSDI-EP's designers felt that a program delivered in a decentralized manner is the most typical pattern for delivery of vocational and other social services in the United States and thus would better model the likely environment in which SSDI beneficiaries would use any statutory offset provision.

Nonetheless, the choice of this decentralized structure for the SSDI-EP reflected both the history of the Pathways Projects and considerations specific to the SSDI-EP. The single most important component of the SSDI-EP's service approach is the provision of work incentive benefits counseling. The Pathways Projects (and the antecedent working group housed in DHFS) had been instrumental in training benefits counselors in the state, particularly in the context of Wisconsin's State Partnership Initiative (SPI) demonstration.¹⁶ From 1999 through 2004, the Pathways Projects had supported training, technical assistance, and, to a significant degree, funding of benefits counseling through the twenty provider agencies that worked directly with SPI participants. In point of fact, there had been little capacity to provide work incentive benefits counseling in Wisconsin before SPI and the capacity that existed was concentrated at organizations that became SPI provider agencies. Because of SPI itself and, later, the training and technical assistance capacity that began in SPI, there had been substantial growth in the number of trained benefits specialists.¹⁷ As a result, much of this capacity has remained at those organizations that had served as SPI provider agencies. It was simply more practical to utilize this existing capacity than to attempt to build it at the central project office in Madison, especially as SSA indicated that the pilots should be able to operate on a statewide basis.

Additionally, the provider agencies during SPI had delivered benefits counseling in the context of a broader person centered vocational planning process (PCP). While Pathways staff did not wish to mandate use of an often costly PCP approach for all SSDI-EP participants, they did want participants to have an opportunity to access such services as they might find useful. Again, this pointed toward giving community based agencies a major role in the pilot. First, the capacity to provide both PCP and benefits counseling was concentrated in such agencies, in particular those that had participated in SPI or had later hired staff who had worked at the SPI provider agencies. Moreover, many of the SPI provider agencies claimed that outcomes for SSDI beneficiaries in that project had been constrained by the lack of a SSDI offset provision.¹⁸ Pathways staff

example, all SSDI-EP agencies provide or arrange for benefits counseling; in Utah almost all benefits counseling is provided through the Utah Office of Rehabilitation.

¹⁶ Wisconsin's SPI project was called "Pathways to Independence." To avoid confusion, this title will not be used again in this paper.

¹⁷ The term "benefits specialist" is used in Wisconsin to denote a person who provides work incentive benefits counseling. In most cases these individuals have gone through intensive training.

¹⁸ The Wisconsin SPI project offered participants who were SSI beneficiaries a waiver that included a more generous offset provision than the "1:2" feature of the existing 1619 provision. Wisconsin had also sought a waiver to offer a benefit offset for SSDI beneficiaries participating in SPI, but did not obtain it.

thought there might be value in looking at whether persons with substantial PCP experience might be in a better position than others to quickly exploit the offset without substantial additional services.

Another significant factor was, with the exception of some ability to fund benefits counseling services, the SSDI-EP would have no ability to pay for participant services.¹⁹ Community agencies, especially those with experience providing vocational services, had established working relationships with the government agencies that typically fund such services for persons with disabilities. Foremost among these is the Wisconsin Division of Vocational Rehabilitation (DVR), though the various Long Term Care programs in DHFS are also an important funding source. Pathways anticipated that these agencies' experience would make it more likely that appropriate individualized service packages could be cobbled together. Indeed, it was hoped that these agencies' existing relationships with consumers and their more visible presence in their respective service areas would make it far easier to recruit potential participants than attempting to do so from a central project office housed at DHFS in the state capital, Madison.

Furthermore, there was a contingency that supported use of community agencies as the setting for direct contact between the SSDI-EP and its participants. In brief, state rules made it easier to contract with entities with which Pathways had an existing contractual relationship than to either solicit new partners or to build the needed statewide capacity at Pathways itself. In most cases, Pathways could enter into contracts with agencies to become SSDI-EP provider agencies as essentially a continuation of the relationship established in SPI.

While this decentralized structure would appear to enhance the reach of the pilot, it also meant that there would be little direct contact between central SSDI-EP staff and most participants. Provider agency staff would be the face of the project for the participants and the SSDI-EP would be highly dependent on agency staff members' understanding of project rules and of the performance of duties entrusted to them. As will be noted later, this condition also applied to the implementation of research tasks such as informed consent processes and the collection and submission of data on a monthly basis.

This decentralized structure places great importance on the capacity of the Pathways staff involved in SSDI-EP operations to create and fine tune working procedures and to provide effective training, technical assistance, contract monitoring, and troubleshooting. The Pathways staff serves as an intermediary between the participants and their benefits specialists on one hand and SSA staff in Baltimore on the other. In particular, the central Pathways operations staff has been involved in the resolution of issues or conflicts involving eligibility, the initiation or end of the offset provision, and overpayments for those assigned to the SSDI-EP treatment group.²⁰

¹⁹ These benefits counseling services were paid out of other monies available to OIE/Pathways, not through the SSDI-EP contract with SSA.

²⁰ An overpayment occurs when SSA pays the beneficiary more than he/she is legally entitled to. Overpayments can occur for many reasons unrelated to participation in the SSDI-EP treatment group.

Thus, it has been necessary that this staff include persons who can function as benefits specialists.

Finally, in addition to the substantial decentralization represented by the use of provider agencies, the SSDI-EP was structured to strongly separate evaluation from other central operations. This was done to facilitate a genuinely independent evaluation. This separation was manifested in at least two important ways. First, data collected for research purposes is, with the exception of a few data elements expressly released by participants for program administration purposes, unavailable for operational uses. Second, during enrollment, there are separate informed consent processes for the pilot and for the research, though to limit participant confusion these are administered concurrently. Though operations and research staff attempt to keep their provision of training, technical assistance, and other contacts with provider agency staff distinct, provider agency staff have some difficulty understanding the division of responsibilities. Perhaps the fact that the research staff is also housed at the Madison office contributes to this, though the co-location with operations staff facilitates co-operation and gives the research staff greater ability to observe the project and perform process evaluation activities.

Intervention and service provision

Pathways decided that it would structure the SSDI-EP so that the availability of the offset provision itself would be the only pilot based difference in what members of the treatment and control groups would experience following random assignment. This statement should not be interpreted as meaning that there was an expectation that their experiences would be the same in a literal sense. It was understood that treatment group members might well have more or better employment opportunities because of the offset and, thus, greater service needs. However, SSDI-EP sought ways to make sure that provider agencies would not deliberately give some participants either a better quality or greater quantity of services simply because of assignment to the treatment group. The aim was to facilitate a comparison of treatment and control group outcomes where any differences would be, at least in theory, attributable to the offset alone.

The SSDI-EP has several policies or standards dealing with service provision designed to support achievement of this goal. The SSDI-EP, with one important exception, does not guarantee participants a specific service package. Provider agencies are expected to make the same effort to determine and arrange for needed services for all participants on an individualized basis that is consistent to the greatest extent possible with the participant's expressed preferences. As noted, funding or in-house resources for services must generally be found on a case by case basis. Agencies were expected to make good faith efforts to locate the resources needed to help all participants achieve their employment goals.

The one area where provider agencies are in some genuine sense required to insure service provision is benefits counseling. The SSDI-EP requires all provider agencies to have or arrange for the capacity to provide work incentive benefits counseling.²¹ However, though all participants are ensured access to needed benefits

²¹ Pathways much preferred that provider agencies had a trained benefits specialist. To encourage this, Pathways put substantial resources into providing for training and ongoing technical assistance. With few exceptions, SSDI-EP provider agencies chose to have benefits

counseling, it is each provider agency that is in the dominant position to interpret what this commitment means. The SSDI-EP central office has not mandated a minimum amount of service, though pilot rules do require that a participant have a full benefits summary when entering the pilot.²² Additionally, provider agencies are expected to arrange for benefits counseling for any SSDI-EP participant when there is a significant change in that person's employment situation or life circumstances. The OIE work incentive benefits counseling grant (OIE grant) provides the means to realize this should there be no other funding source.²³

Provider agencies were expected to enroll any eligible participant, except as limited by three factors. First, the provider agency was not required to enroll an otherwise eligible individual when the agency did not have the capability to serve a person with a particular combination of symptoms or impairments. Second, the provider agency was allowed to refuse participants who were not eligible for agency services because of state or pre-existing agency rules. Finally, provider agencies had designated enrollment and service areas negotiated as part of their DHFS contracts and could choose not to serve individuals who resided outside the boundaries.

In lieu of direct funding for services, the SSDI-EP funds provider agencies chiefly for providing data for both operational and evaluation purposes, but secondarily to support communication with and the involvement of participants and to allow agency staff to participate in pilot related training and technical assistance activities.

Project size and recruitment

The SSDI-EP hoped to enroll 800 participants. The expectation was that approximately half of the participants would be recruited from the 956 persons who had enrolled in Wisconsin's SPI project. However, provider agencies were neither required nor explicitly encouraged to conduct recruitment activities in a manner that would result in enrolling roughly equal proportions of individuals with SPI experience and of those without.

Instead, such a balance, should it occur, was seen as a likely consequence of the primary role that the provider agencies would take in engaging in outreach and recruitment. While provider agencies could not give enrollment preference to individuals with which they had current or past relationships, Pathways anticipated that, as in SPI, the very fact of a relationship between the agency and a potentially eligible consumer would greatly increase the probability of enrollment. As seventeen of the twenty-two

specialists on staff, though several agencies went through periods when they either had no benefits specialist or an inexperienced one.

²² This did not necessarily require doing an assessment de novo. For example, a participant with a full benefits summary completed within six months, sometimes a year, before enrollment would not be seen as automatically needing additional benefits counseling provided a benefits specialist determines that there had been no relevant changes in the consumer's situation.

²³ However, several provider agencies did not apply for the OIE grant until 2007; one agency has still not applied. As a result, until July 2007, there was no way to insure funding for all participants at these agencies unless the agency was willing to absorb the cost. These agencies could have easily qualified for the OIE grant at any time had they chosen to apply.

provider agencies had been SPI agencies, it was thought that it would be relatively easy to do outreach to and ultimately enroll many SPI participants. Indeed, it was hoped that many consumers, based on their SPI experiences, consider themselves good candidates for the SSDI-EP.

The SSDI-EP did not require nor encourage provider agencies to seek potential participants who had either started or completed their TWP. Instead, the pilot's designers wanted a participant population that had some genuine interest in either becoming employed or, if already employed, increasing their earnings or other employment related outcomes. Thus, there was never any expectation that the pilot outcomes would represent the likely outcomes for the universe of all Wisconsin SSDI beneficiaries if a similar benefit offset became law. At most, it was hoped that study outcomes would be indicative of those that would be achieved among persons more likely to use a statutory offset in the current environment.²⁴

Wisconsin's Choices: Evaluation

The evaluation plan focuses on both the measurement of participant outcomes and understanding how the pilot's design, implementation, and the context in which that implementation occurs shapes participant outcomes. This discussion focuses on the outcomes evaluation, especially in the areas of duration and data collection.

The primary participant outcomes of interest are changes in the probability of employment, in the amount of earnings, and in gross income.²⁵ These are followed from the eighth calendar quarter prior to the quarter of study entry through either four or eight quarters following enrollment, depending on how long SSA allows the SSDI-EP to operate. As the last participant enrolled on October 31, 2006, that participant will finish her/his fourth post-enrollment quarter at the end of 2007 and her/his eighth post-enrollment quarter at the end of 2008.

However, the outcomes analysis is complicated by the fact that treatment group members cannot use the offset unless they have completed their TWP plus the three month grace period. As long as most participants have not done so, it will be difficult to

²⁴ 53% of SSDI-EP enrollees reported that they were employed when they enrolled; nearly 60% of those reporting employment claimed to be working at least twenty hours per week. We do not have employment rate or effort data for SSDI only beneficiaries in Wisconsin, though the enrollment form data, if accurate, suggest an employment rate four times higher than that for the state's SSI population (including concurrent beneficiaries). Thus, we are willing to conclude that the SSDI-EP's enrollment is composed of persons more likely to use a statutory offset if one were available. We are not claiming that SSDI-EP participants are necessarily representative of that hypothetical population.

²⁵ Employment outcomes are measured using Wisconsin Unemployment Insurance (UI) records as the primary source of information. These data have exclusions that will result in an underestimation of employment rates and earnings. In addition, these data are reported on a calendar quarter basis that can overstate the continuity of employment, almost certainly resulting in higher employment rates than those at any time point or for any shorter duration. The income proxy measure is the quarterly sum of UI earnings, Social Security payments, and, when applicable, the state SSI supplement. Thus the estimate excludes other sources of income and the amounts of quasi-cash benefits such as Food Stamps and Title 8 housing subsidies.

discern the true impact of the offset provision. Thus the quality of the outcome analysis is contingent on the rate of TWP completion among both the treatment and control groups and whether the evaluation team has adequate information about whether and when each participant completes TWP.

Consequently, the rate and speed of TWP completion becomes in itself an outcome of interest. Other participant outcomes of interest include changes in perceptions and attitudes, the characteristics of jobs and employers, and the reasons jobs end, employment duration, and whether the offset has different impacts on identifiable groups based on factors such as Medicaid Buy-in participation, relative size of the SSDI entitlement, and various socio-demographic, work experience and health and impairment related variables.

Individual level data is collected through multiple sources, including encounter data from the provider agencies, participant surveys, and administrative data from SSA and multiple state agencies. Data from each of these data sources is collected at multiple, but sometimes different time points.²⁶ Additional data are collected through interviews, focus groups, examination of program documents, and direct observation of SSDI-EP activities.²⁷

For the SSDI-EP, the evaluators decided to place greater emphasis on collecting data about participant characteristics through use of administrative data instead of through participant survey or provider agency encounter data. This choice was informed by participant and agency reports during SPI that the surveys and, to a lesser extent, the enrollment forms were too long and burdensome. This has made the evaluation more dependent on the availability of SSA data. This dependence has been amplified by the fact that, unlike SPI, a significant proportion of participants are not DVR consumers. Thus, DVR records cannot be used as a primary source of information about participant characteristics, for example impairment or the severity thereof.

Challenges Faced

Most, some would argue virtually all, evaluations face implementation challenges. The critical issue is whether such challenges seriously compromise or, at worst, invalidate results. Such damage may be limited to a particular component of the evaluation or may make it impossible to get useful findings on the core question(s) the evaluation is meant to address. Indeed, an important purpose of random assignment is to simply avoid a whole class of challenges inherent to non-experimental designs. As we discuss the challenges faced in the SSDI-EP evaluation we will use the term implementation loosely, including some items that might be viewed as deficiencies in planning or analysis that occurred prior to implementation per se. As noted at the outset, though the focus is on evaluation challenges these are often deeply entwined with

²⁶ Participant surveys are completed at enrollment and annually for the subsequent two years. Encounter data are submitted monthly. Administrative data are, depending on source, typically reported on a monthly or quarterly basis, generally with a significant time lag.

²⁷ Direct observation is mainly limited to program level activities, especially those within or sponsored by Pathways. There is no direct observation of the interactions between participants and provider agency staff.

operational challenges, whether at the SSDI-EP's central program office, the provider agencies, or the Social Security Administration.

For reasons of brevity, we concentrate on challenges that either have impaired or retain significant potential to impair the ability to evaluate SSDI-EP participant outcomes as intended.²⁸ Far less attention is given to challenges that were easily met or ameliorated. In a decentralized structure where autonomous organizations engage in the vast majority of pilot-participant interactions and do so largely unobserved by the central project office, almost every aspect of pilot operations may vary across the provider agencies. Yet sometimes that variation may lead to innovation or useful flexibility. That part of the story remains to be told.

Finally, judgments as to whether implementation has been sufficient to support the planned outcomes evaluation are essentially qualitative. Nonetheless, the judgments offered are not arbitrary. Our general approach is to look at what the project was expected to do and to then look at available qualitative and quantitative evidence to assess whether and to what extent those expectations were met. Given the SSDI-EP's decentralized structure we are particularly interested in identifying when more than a handful of the provider agencies were seriously failing to meet program standards. Certainly, an implementation problem that adversely affected even 10% of participants for any extended period might cast doubt on the accuracy of the outcome estimates, even if it did not render the outcome estimates useless.²⁹

Enrollment: Planning and Execution

The SSDI-EP was expected to enroll approximately 800 participants. About half of these participants were expected to be persons who would be newly recruited and who would generally have no or minimal experience with work incentives benefits counseling and person centered vocational planning. The remaining 400 or so participants were expected to be recruited from former SPI participants and thus typically would have had substantial past exposure to the service components associated with SSDI-EP participation.

²⁸ In this paper we do not analyze whether the SSDI-EP and the other benefit offset pilots were appropriate candidates for outcome assessment through use of an experimental or even quasi-experimental designs. We acknowledge that it can be risky to begin such evaluations until there is some evidence that program components are reasonably stable and/or the external environment is reasonably free of other initiatives that would make it virtually possible to identify the effects of the program being examined. As already noted, we are especially concerned that there will be enough time to meaningfully assess participant outcomes. Still, we believe that given SSA's long experience administering the 1619 offset provision and Wisconsin's experience with the basic organizational structure and service provision model the SSDI-EP uses there is a reasonable prima facie case for the use of an experimental design.

²⁹ For example, if the outcome estimates show reasonably large differences between the treatment and the control groups in the hypothesized direction, then the results may be sufficiently good to support policy adoption, even if there is doubt about the preciseness of the estimates.

When SSDI-EP enrollment ended on October 31, 2006, there were 504 enrollees.³⁰ Only twenty-two had been SPI participants. Thus, while the SSDI-EP easily surpassed its enrollment goal for those who had not been in SPI, the SSDI-EP failed miserably in its effort to reach the targeted number of former SPI participants. As a result, a major element of the evaluation plan, the outcomes analysis for those with substantial pre-study receipt of benefits counseling and associated services would not be possible. In this section of the paper we examine the source of this failure. In turn, we also inquire into what may explain the stronger than expected enrollment numbers for participants without SPI experience.

Wisconsin, specifically the DHFS office that was to become OIE (“Pathways”), had sought a SSDI cash benefit offset as part of the SPI project. In conjunction with other states, Wisconsin had continued to press SSA for permission to implement a benefit offset during the late stages of SPI and then after SPI ended. This history is important as during the planning for the SSDI-EP and the lead up to its opening there was a tendency at OIE and among its stakeholders to see the project as an extension of the Wisconsin SPI project.³¹ Wisconsin’s planning was also based on an expectation that, as proposed for its SPI SSDI waiver, concurrent beneficiaries and DACs (Disabled Adult Children) would be eligible for the benefit offset pilots.

Depending on data source, between 620 and 689 SPI participants would have been eligible for the benefit offset pilot if the eligibility rules had been those proposed by Wisconsin.³² Though to get 400 SSDI-EP participants from this group suggested a rather high take-up rate (58% to 65%), there was optimism among program designers that this would be achievable. One of the strongest criticisms of the SPI project by the provider agencies, participants, and other stakeholders had been the inability to get a SSDI waiver. It had been claimed that not having a SSDI waiver had both limited SPI enrollment and had depressed employment outcomes among the SSDI beneficiaries

³⁰ As four of these enrollees were subsequently found ineligible, final enrollment (pending any additional determinations of ineligibility) is 500.

³¹ Before the official end of the SPI project, it had been hoped that any offset pilot would be an extension of SPI and, consequently, eligible SPI participants might enter the pilot by signing an extension to their original informed consent forms.

³² The 620 number is based on participant self-report at enrollment. The 689 estimate comes from records provided to the Wisconsin SPI project from Mathematica Policy Research, Inc. (MPR). Unfortunately, these data did not allow one to unequivocally distinguish between SSDI-only and concurrent individuals. The 689 figure is an estimate because it is based on MPR provided information about the characteristics of the roughly 600 participants who were included in the outcomes analysis rather than all 956. The 689 figure is also likely to be a slight overestimate. SPI participants who enrolled too late in the project to be included in the outcomes analysis self-reported as SSDI beneficiaries at about a 2% lower rate than the overall group. However, it is also likely that some number of SPI participants who entered as SSI only became concurrent through their employment activity following entry into SPI.

Unless otherwise noted, data for the Wisconsin SPI project come from databases maintained by the UW-Stout research personnel located at the Pathways Office. All remaining individual level data for research participants has been de-identified.

who did enroll, particularly those who did so in expectation that a SSDI waiver would be approved.³³

However, by the time SSA had approved the overall plan for the Wisconsin pilot in spring 2005, it was clear that concurrent beneficiaries would not be eligible for enrollment. Based on self-report, only 403 of the SPI participants were SSDI only. In principle, Pathways managers and staff could have acted on this information. There is evidence that the evaluation's designers had discussed this issue, though it is uncertain whether it ever became salient to Pathways managers before either provider agencies were contracted or enrollment started.³⁴ Even if it were, it is possible that for practical reasons the managers would have been resistant to revisit the basic design of the pilot so close to the start-up date in the absence of a SSA request or directive to do so.

Nonetheless, even if the 400 enrollment target was "off the table," the question remains why only twenty-two SPI participants entered the SSDI-EP. A roughly 50% take up rate would probably have supported the planned subgroup analysis; even 20% would have supported use of SPI participation as a control variable in the main analysis. To begin, the 403 number itself was somewhat high, as in the period between SSA contracting with Wisconsin and the SSDI-EPs opening, SSA decided to exclude DACs from the benefit offset pilots. While the proportion of DACs among Wisconsin SPI participants is unknown, it is likely that this had only a marginal impact on the lack of SPI participant enrollment in the SSDI-EP. It is far more likely that the lack of significant SPI based enrollment reflects two factors: the limited information available to provider agencies about who they had served in SPI and the more limited efforts, on average, former SPI agencies made to find and enroll participants compared to those undertaken by SSDI-EP provider agencies that had not been involved in SPI. Both of these factors, while not inevitable results of the SSDI-EP's decentralized structure, would have been irrelevant or mattered far less in a more centralized project structure.

SSDI-EP central operations staff reported that most of the fifteen provider agencies that had taken part in SPI had either not maintained records of who had been SPI participants or could not find those records. Central project staff also reported that staff turnover at the agencies since the end of SPI in 2004 contributed to this problem. Had the former staff member remained, there would have been memories of who had been a SPI participant. Moreover, even when agencies had information about SPI participation, staff attrition meant disruption of existing trust relationships with many of those participants. Central operations staff reported that this disruption of interpersonal, albeit professional, relationships made it less likely that any given consumer would enroll in the pilot.

One could ask why Pathways did not provide the former SPI agencies with the names of their SPI participants. The answer was twofold. The evaluators of Wisconsin's SPI still maintained that information, but could not supply it to Pathways or DHFS under

³³ See Delin, Barry S., Reither, Anne E., Drew, Julia A., and Hanes, Pamela P. 2004. Final Project Report: Wisconsin Pathways to Independence. Menomonie, WI: University of Wisconsin – Stout Vocational Rehabilitation Institute, pp. 124 & 254-55.

³⁴ Two of this paper's authors, Delin and Reither, were responsible for the SSDI-EP research design.

terms of the SPI informed consent forms.³⁵ Pathways itself had not kept SPI participant records that had been available to DHFS based operational staff during that project.

It is also probable that a second factor contributed to the absence of meaningful numbers of SPI participants in SSDI-EP. Ironically, this factor appears to have made a significant contribution to the enrollment of the higher than projected number of non-SPI participants into the pilot. As a group, the former SPI provider agencies seem to have been far less energetic than the “new” agencies added for the SSDI-EP in identifying and then enrolling potential participants. The fifteen former SPI agencies brought forward into the SSDI-EP had served about 85% of all Wisconsin SPI participants. Nonetheless, these fifteen agencies enrolled only 56% of SSDI-EP participants. On average, enrollment at the non-SPI sites was thirty-two versus nineteen at the former SPI sites, despite the former having no official knowledge of who had been a SPI participant.³⁶

While it is not yet possible to give a definitive explanation as to why enrollment levels are typically so much lower at the former SPI agencies, we can provide pertinent contextual information. First, there is no evidence of major differences in staffing levels, staff attrition, access to training and technical assistance, or the availability of external service funding between those agencies that participated in SPI and those that did not. What is clearly different is that during SPI, provider agencies had direct funding to support staff and provide both benefits counseling and vocational services. We conjecture, with support from agency staff interviews, that the former SPI agencies exhibited some reluctance to aggressively recruit participants because SSDI-EP could not provide direct support for participant services. The only direct income flow would be for research reporting and encouraging continuing participant involvement. By contrast, the “new” agencies made their decisions to participate without direct experience of the former, more generous funding environment. We hypothesize that most of the non-SPI agencies sought higher enrollments as part of their “business plans.” They appear to have been more willing to take advantage of economies of scale and spread potential risk from any participants with higher service costs. It is suggestive that participant to staff ratios appear higher at the agencies that did not participate in SPI.³⁷

³⁵ This information was maintained until the end of September 2007. The SPI evaluators were not precluded under terms of the participant consent forms from sharing back with provider agencies encounter data provided by those agencies. However, during SPI the evaluators had limited this sharing to aggregate data, except for the specific purpose of cleaning submitted encounter data. Again, while there is some recollection that the SSDI-EP evaluation designers discussed this possibility following SSA’s decision to exclude concurrent beneficiaries from the pilots, there is no evidence that this possibility was discussed with SSDI-EP managers and operational staff.

³⁶ Two of the “new” agencies were headed by individuals who had worked with SPI participants at their former employers. The agency among these that enrolled a far higher number of participants was headed by an individual who had been employed at a SPI participating agency in another region of the state. Thus, it is unlikely that person’s past contact with SPI participants contributed much to SSDI-EP enrollment at that individual’s new agency.

³⁷ Nevertheless, it is important to remember that agencies can differ in service philosophies. Sometimes this can be a result of legal requirements, organizational choice, and/or needs that arise from the characteristics and circumstances of an agency’s consumers. We have not systematically investigated the extent of differences in these areas between the SPI and “new” agencies, though our impression is that a larger proportion of SPI agencies operate within these

Still, there is a remaining puzzle. When provider agency staff were interviewed in the spring of 2006, respondents from agencies that had been involved in SPI were just as likely as those from the non-SPI agencies to report they actively recruited from both their current and past caseloads. Even if current staff members at the former SPI agencies were not aware of whom among their agencies' current and former consumers had been SPI participants, surely, given consumers' relative geographic immobility and their typical needs for continuing services, recruitment from the general agency caseloads should have led to enrollment of more than twenty-two SPI participants in the pilot?

One part of the puzzle may be that both Pathways central staff and the evaluators overestimated the proportion of enrollment that would be generated from agencies' own caseload. This expectation was based on enrollment patterns observed during SPI. By early 2006 it had become clear that enrollment totals were lagging behind expectations and that the SSDI-EP central office would need to identify a mechanism to supplement agencies' outreach efforts. The mechanism chosen was direct mailings to approximately 10,000 Vocational Rehabilitation (DVR), Medicaid Buy-in, and DHFS Long Term Service Program consumers. The letters were mailed roughly half way through the fifteen month enrollment period.³⁸ The effort appears to have been effective as 60% of SSDI-EP enrollment occurred after the mailings. Additional evidence of the limited ability of agencies to "mobilize" their own consumers comes from the first set of the annual participant follow-up surveys.³⁹ At the time of writing, survey results are available for participants who enrolled before the end of July 2006. While this group includes some participants who enrolled after the mass mailings, it is disproportionately composed of those enrolled before the mailings. Only 21% of respondents reported hearing about the pilot from the agency where they ultimately enrolled.

The second part of the puzzle might be the fact that no beneficiary is eligible to enroll in the Pilot or, if enrolled, to use the benefit offset after the seventy-second month following the end of her/his Trial Work Period. Provider agency staff, Pathways central staff, external observers, and, in focus groups, participants themselves have all asserted that the "seventy-two month rule" depressed enrollment in two ways.⁴⁰ First, it excluded

constraints. One indicator of this is the much larger number of service hours for job coaching and development of on-the-job natural supports reported from one subset of the SPI agencies.

³⁸ These mailings were delayed about two months as not to arrive at the same time as DHFS originated correspondence concerning Medicare D.

³⁹ The annual survey is sent out at approximately the anniversary of each participant's enrollment. The data presented from the first annual follow-up survey is based on 312 responses. There were 165 responses from the treatment group, 147 from the control group. Response rates for the first follow-up survey have been approximately 90% for both the treatment and control groups.

⁴⁰ In particular, provider agency staff reported that about 40% of the consumers indicating a serious interest in entering the pilot were determined ineligible before they could enroll. The main reason for ineligibility was the "seventy-two month rule."

Also, it is important to note that observers are not necessarily or even mainly talking about persons with histories of lengthy spans of above SGA earnings following their TWP. In many cases they are talking about persons who are deliberately holding earnings below SGA, i.e., engaging in what is called "parking." We are not aware of any truly good studies of this

persons who have strong continuous work histories, even when beneficiaries had disabling conditions that unequivocally met SSA listings. Some argued that these were the very people in the best position to make gains under an offset provision. Second, some individuals approaching the seventy-two month limit may not have enrolled because they feared that they would incur high transaction costs during the short period between entrance and exit.⁴¹

We are not in a position to assess the actual impact of the “seventy-two month rule” on SSDI-EP enrollment in general or on former SPI participants. We think it plausible that some SPI participants may have been ineligible for the SSDI-EP for this reason or chose not to enter the pilot because they were approaching the time limit. Certainly many who had made employment gains during the five year SPI project period may have sustained them in the additional two years before SSDI-EP enrollment ended. Unfortunately, the Wisconsin SPI evaluation report lacks information on how many participants completed their TWP or when they did so. While we know that SPI participants made significant gains in employment rates and earnings over time and in comparison to a similar group of Wisconsin Division of Vocational Rehabilitation consumers, it is equally important to remember that employment rates never exceeded 50% in any calendar quarter and even after two years of exposure to the intervention only about 35% of those employed had quarterly earnings at or above the three month equivalent of SGA.⁴² So, without stronger evidence, we would not conclude that the “seventy-two month rule” was a major barrier to enrollment of former SPI participants.

Finally, though we have focused on the overlapping phenomena of the inability to enroll analytically useful numbers of former SPI participants and of the generally higher enrollments of the agencies that had not taken part in SPI, it is important to note that there was significant variation within the groups of former SPI and non-SPI agencies. For example, a few of the former SPI agencies achieved enrollment levels approaching (though not reaching) those they had during SPI; not all of the “new” provider agencies had large enrollments. There was also substantial variation across the provider agencies, independent of SPI involvement, in the reported proportions of interested consumers found ineligible before enrollment. Indeed the results were strongly bimodal.

phenomenon. In part this is a result of lack of a truly good source of monthly earnings information. However the deeper problem is that parking is only parking if someone is deliberately holding their earnings under SGA. Moreover, different individuals may seek to hold their earnings to different levels relative to SGA. Indeed, some beneficiaries may not be aware that the SGA level was raised and indexed in 2000 and, believing that SGA is still at \$500 month, may be parking at roughly 50% of the SGA level. We thank Joseph Entwisle of Health and Disability Advocates for the insights on which this discussion is based.

⁴¹ While such costs can be directly financial, they can also be incurred in time, effort, and anxiety. While such costs might have been viewed as hypothetical during the SSDI-EP’s enrollment period, ongoing difficulties in administration of the benefit offset have made these costs real.

⁴² See Delin, et al., 2004. Final Project Report: Wisconsin Pathways to Independence. Menomonie, WI: University of Wisconsin – Stout Vocational Rehabilitation Institute, pp. 131-35 and pp. 147-50. Employment and earnings data were drawn from Wisconsin Unemployment Insurance records. The Wisconsin SPI project did not use random assignment. Propensity scores were used in regression analyses to “equate” the volunteer comparison group and SPI participants.

About 50% of interviewed provider agency staff who offered an assessment reported ineligibility rates of 25% or less. About 40% reported ineligibility rates of 50% or more.⁴³ There was also substantial variation across agencies on the reported proportions of putatively eligible participants who did not enroll. Though all but one of the thirty-one agency staff providing estimates said that their agency's non-enrollment was less than 50%, the estimates were broadly distributed across that range. Ten of thirty-one said that no eligible consumer refused enrollment. By contrast, eight reported that the non-enrollment rate at their agencies was in the 25% to 50% range. While these differences do not directly challenge the capacity to evaluate the SSDI-EP, it is one more indicator of the potential variation in areas such as the administration of informed consent processes, the provision of equal access to services, and efforts to communicate with participants that may, at some agencies, cast doubt that the versions of the SSDI-EP delivered were close enough to what Pathways intended to support evaluation.

Enrollment: Informed Consent and Random Assignment

SSA's overarching purpose in authorizing the benefit offset pilots was to generate information that would help SSA (and its chosen partners) design and implement a national benefit offset demonstration. One of SSA's specific goals was to learn how to effectively inform potential participants in a national demonstration about the demonstration, the random assignment it would involve, and the implications of assignment on what participants would receive and be obligated to do. As SSA would need participants to consent to their involvement, SSA was also interested in using the pilots' experiences to identify effective ways of obtaining consent that would not unduly bias who participated in the study. The value of random assignment would be lessened to the extent that those agreeing to participation would differ markedly from the intended study population or there was substantially greater attrition of participants from control groups than treatment groups.⁴⁴

In Wisconsin, prospective benefit offset participants received detailed information about the pilot at the twenty-two provider agencies rather than from a central source.⁴⁵ Any discussions about "ability to benefit" took place through interactions at the agency level as did the formal informed consent process. The enrollment process (typically following some earlier contact between the consumer and agency staff) had three stages. First, the potential participant had to read the informed consent materials and have the opportunity to ask any questions. If the individual agreed to participate, he/she would need to sign separate project and research consent forms and complete an initial

⁴³ One agency reported a 100% ineligibility rate and never enrolled a single participant. That agency's involvement with the pilot ended at midyear 2007. A second agency also ended its involvement at that time. Its small number of participants transferred to another agency. Consequently, the loss of these agencies had no significant impact on the SSDI-EP or the ability to evaluate it.

⁴⁴ The draft design for the national demonstration that SSA shared with the state pilots in March 2007 involved multiple treatment and control groups.

⁴⁵ Although agencies varied in whether benefits specialists or other types of staff explained the pilot and/or conducted enrollment, benefits specialists typically assessed eligibility for the pilot.

earnings estimate.⁴⁶ Second, the new participant would complete a baseline survey. Finally, provider agency staff would complete an on-line enrollment form by interviewing the participant.⁴⁷ When this form was submitted, the assignment to a study group would be generated from the central project office in real time. The staff member at the agency would immediately inform the participant of her/his assignment. Subsequently, the SSDI-EP central office would mail the participant a letter confirming enrollment and study group assignment.⁴⁸

Given the complexity of the information that provider agency staff conveyed and, often, of the discussions attendant on it, there is little reason to assume that agency staff were always consistent or accurate in how they implemented the enrollment and informed consent processes. Moreover, there is little reason to assume consistency in the understandings participants took away from these interactions. Uncertainty about what participants in random assignment studies understand about program features and/or their obligations is hardly unique to highly decentralized programs, but the SSDI-EP's decentralization made it more difficult for central project staff to directly identify or to address problems in this area.

A vital step in the enrollment process was to determine whether a consumer was eligible for the pilot. The SSDI-EP placed the responsibility for eligibility determination chiefly on the provider agencies, but did not require agencies to fully verify participant eligibility. Agencies were expected to use the BPQY (Benefits Planning Query) as the main evidence for assessing eligibility. Though provider agency staff reported few problems in getting BPQYs in a timely fashion, many staff members indicated that because BPQYs were often incomplete or included inaccurate information, they supplemented that information with some obtained from local SSA offices and/or consumers. Nonetheless, approximately 5% of those who enrolled in the SSDI-EP were later determined ineligible by SSA.⁴⁹ This finding indicates that at least some agencies

⁴⁶ If the participant was assigned to the treatment group the estimate form would, unless later amended, provide the basic information that SSA would use to calculate the offset amount. Of course, this occurred only if the participant had finished the TWP and a three month grace period and if the estimate indicated monthly earnings above the SGA level.

⁴⁷ Information from the survey and enrollment form went to the evaluation team, though the informed consent materials allowed the evaluation team to provide SSDI-EP operations staff at Pathways limited identifying information needed to operate the pilot. The paper informed consent forms and earnings estimates were mailed to the evaluation team, who then provided those materials to the SSDI-EP operations staff as appropriate.

⁴⁸ Provider agency staff sometimes needed to begin the enrollment process in the field due either to the size of the agency's service area or some consumers' limited ability to secure transportation to the enrolling agency. When internet service was unavailable, there would be delays in submitting enrollment forms and then communicating information about the results of random assignment to those participants.

⁴⁹ SSA staff in Baltimore determined ineligibility for those assigned to treatment, Wisconsin based staff for the control group. For whatever reasons, the SSA unit in Baltimore responsible for the pilots only wanted to know who was in the treatment group. It seemed to us that the conditions of random assignment would be violated if some members of the treatment group were removed from the SSDI-EP for ineligibility, but not members of the control group. Fortunately, Wisconsin based SSA staff was willing to assess enrollees in the control group obviating this potential problem. Curiously, however, the number of those determined ineligible following enrollment was

had difficulties assessing eligibility, though because of SSA follow-up procedures this should have no impact on the outcomes evaluation beyond a slight reduction in study population size.

As noted we had no ability to directly observe how well provider agency staff implemented the enrollment process, including informed consent procedures. However, data from participant surveys and focus groups should indicate whether there have been major problems.

In the first annual follow-up survey, participants were asked whether agency staff had explained the pilot in ways that they could understand. The overall results were strongly positive, with 79% of respondents indicating some degree of agreement and with essentially the same result for both the treatment and control groups. There were differences between the study groups in how strongly they agreed the pilot was well explained. Nearly 60% of respondents from the treatment group strongly agreed that agency staff had been able to explain the SSDI-EP so that it was understood. Only 40% of respondents from the control group indicated strong agreement.

The follow-up survey also included an item that indicated a participant's understanding (or perhaps memory) of whether he/she was assigned to treatment or control. Two findings stand out. Few respondents from either group inaccurately identified their study group (3% for both groups). This reassuring result is, alas, substantially confounded by the second. About 40% of the respondents in both of the study groups indicated they simply didn't know which study group they had been assigned to. This result may seem plausible for those in the control group, who don't receive a direct benefit from the offset.⁵⁰ We find it disconcerting that the "don't know" rate is as high for the treatment group. We speculate that this may be related to the small proportion of treatment group members that appear to have actually used the benefit offset. We also concede the possibility that many participants may not understand the terminology of random assignment, though comments from the participant focus groups suggest that most participants understand the terms "treatment" and "control" and that some have a strong grasp of the logic of random assignment.

In the focus groups we conducted in spring 2007, participants generally expressed less confidence in how well they thought they understood the SSDI-EP than indicated by the survey results. Indeed, respondents were quite evenly split on whether or not they thought they had a good understanding. Though it is possible that these respondents were unrepresentative of participants in general, it is also likely that the detailed discussion of their experiences particularized what in the context of a survey may be a very general assessment. Moreover, the participants at any particular focus group were all enrolled at the same provider agency. This circumstance allowed us to determine that the overall assessment within any focus group about how well the

noticeably higher among the control group (eighteen versus eleven in treatment), suggesting some differences in interpretation across SSA offices. Originally, the divergence was greater, but has closed as SSA in Baltimore has continued to identify new ineligible almost a year after enrollment concluded.

⁵⁰ Control group members might believe they received services that they would not have received otherwise.

participants understood the pilot reflected that group's perception of the quality of benefits counseling they had received.

In the focus groups, participants were also asked their thoughts about random assignment and about the length and intrusiveness of the enrollment process itself. Focus group participants were about equally split on their perceptions of the acceptability of the random assignment process. Those who found it problematic offered two types of objections. Some of the participants objected to the assignment they had been given. Others objected to random assignment on "principled" grounds.⁵¹ Several individuals argued that it "unfairly penalized" SSDI beneficiaries who in good faith sought to improve their standard of living or future employability. Several argued that there was no need to use random assignment given the low "background" rate of employment and SGA level earnings among beneficiaries. Still, these focus group participants had entered and remained in the SSDI-EP despite whatever objects they had to random assignment.

Two other pieces of information tend to corroborate this interpretation. During interviews, agency staff noted that enrolling participants were generally accepting of random assignment and that the failure of eligible consumers to enroll was most often due to fears about benefit loss, privacy loss, and negative consequences after the pilot ends, or were a manifestation of symptoms associated with severe mental illness. Second, only a modest number of participants have withdrawn within a month of enrollment, indicating that participants have been accepting of the results of random assignment. As would be expected, all six of these persons had been assigned to the control group.

Focus group participants generally reported indifference to both the length and complexity of the enrollment process and of the demands made for personal information. More than one participant offered that as persons using public benefit programs they are resigned to both paperwork and intrusiveness. By contrast some provider agency staff (as suggested above) thought that the enrollment process and subsequent reporting obligations were an important, though not the dominant factor, in decisions not to enroll.

To summarize, there is both direct and circumstantial evidence of variation across provider agencies in the conduct of enrollment and informed consent processes. However, there is no evidence that these variations have been of a character that would impact the quality of the SSDI-EP evaluation. Similarly, though many participants appear to have doubts about the fairness and necessity of random assignment, those doubts do not seem to have resulted in meaningful attrition.

Infrastructure and Service Provision

Successful outcome analysis is dependent on whether an intervention has been implemented to some acceptable degree. While technically the benefit offset is the only intervention being tested for its impact on participant outcomes, that test is occurring in a

⁵¹ We did not seek to have equal numbers of treatment and control group members at each focus group. Though we could retrospectively identify the study group assignment of each focus group participant, we have not done so except to generate aggregate descriptions of focus group participant characteristics.

specific framework of service provision. Consequently, it is important to learn whether the intended context in which the offset will be used and assessed is in place.

Our approach to examining this issue will be more quantitative than qualitative, as there will not be a direct examination of the quality of benefits counseling or other activities at the provider agencies either in this section or elsewhere in this paper. However, we can provide some information about participants' perceptions of the appropriateness and value of available services.⁵² Still, from our ongoing discussions and attendance at meetings with managers and staff at the SSDI-EP central office, we are not aware of widespread and serious deficiencies in the performance of benefits specialists and other staff at the provider agencies. Feedback obtained from participant focus groups is consistent with this appraisal.⁵³

In this section, we examine two important aspects of implementation at the provider agencies. The first is whether the front line staff needed to recruit and enroll participants, provide or arrange for participant services, and implement data collection and reporting requirements were in place and then stayed in place. Restated, this examination focuses on the availability of staff, especially the critical resource of work incentive benefits specialists. After all, the one service guarantee that the SSDI-EP makes is that all participants will have access to benefits counseling when it is needed. The absence of a qualified benefits specialist could have adverse consequences in multiple areas ranging from assessing eligibility for enrollment, to updating earnings estimates, calculating the potential gains and losses that a new job or promotion would involve, and helping to deal with problems arising from transition on and off the offset.

We have two sources of data about initial staffing and subsequent continuity of staffing at the provider agencies. One is the records kept by SSDI-EP staff at the central program office, the other is research reporting training records which includes information about the addition and removal of access to the secure online reporting system. Both sources of information indicate that with few exceptions provider agencies maintained adequate staffing throughout the project.⁵⁴

⁵² Though not assessed in this paper, the SSDI-EP provides extensive training and technical assistance activities in both the operational and research domains. These activities provide the SSDI-EP central office with an important means to both guide the decentralized implementation of the pilot and assess performance at the (originally) twenty-two provider agencies. Based on spring 2006 interviews with provider agency staff, this activity is generally valued. However, more respondents place high value on the one to one interactions with SSDI-EP central office personnel than on group events, whether in-person or via conference calls.

⁵³ We have not yet examined participant survey data that should allow assessment of this issue at the agency level, at least for those with enrollments large enough to support the analysis.

⁵⁴ Based on these records, discussions with SSDI-EP operational personnel, and our own observations, we would identify four agencies where the delivery of the intervention was significantly compromised for some protracted period. Fortunately, these problems affected less than 10% of pilot participants.

A complicating factor was that one of these agencies did not initially assign a benefits specialist to the pilot, but contracted out benefits counseling activity to another organization. The contracted organization appears to have been quite slow in delivering benefits counseling services to SSDI-EP participants from this provider agency. There is evidence from a participant focus group held at this provider agency that some participants received no benefits counseling services at all.

Additional information from interviews with provider agency staff tended to reinforce what was learned from SSDI-EP personnel and records. Virtually all respondents reported that there had been no problem identifying existing staff or hiring new staff for the project. Much greater concern was voiced about the ability to identify and hire qualified staff in the case of attrition. Particular concern was raised over potential replacement of benefits specialists. It was noted that new trainings were available only several times a year and that it often took a new benefits specialist months to become proficient and confident at the job.⁵⁵ We conclude, therefore, that across most agencies staffing has been at least adequate. The localized problems that developed have not been of a character to threaten the ability to evaluate participant outcomes for the pilot as a whole.

The second aspect looked at is consistency in the amounts of pilot related services, particularly benefits counseling, that participants assigned to different study groups received. The SSDI-EP was designed to provide all participants, irrespective of study assignment, with equal access to services. Beyond any ethical dimension to this commitment, it was aimed at insuring that the evaluation could get a reasonably clean measure of the benefit offset's impact on outcomes. While no participant, aside from access to necessary benefits counseling, was guaranteed a specific package or amount of services, those assigned to the treatment group were not to receive additional services by virtue of their study group assignment.⁵⁶

Assessing this issue presents several challenges, not all of them resolved. The most important of these is how to determine when differences in the patterns of service provision between treatment and control group members are the result of differential behavior stemming from the availability, whether immediate or in the future, of the offset and its associated Continuing Disability Review protections. If the benefit offset is an effective work incentive, it is reasonable to expect those in the treatment group to seek out or exploit opportunities for working more, having higher hourly pay rates, and/or pursuing career development. Consequently, treatment group members, on average, might be expected to seek out and then receive more benefits counseling or other

There were two other agencies where project start-up and thus enrollment activities were delayed, though this seems not to have seriously affected subsequent operations. Finally, there was an agency that ended its relationship with the SSDI-EP because funding problems unrelated to the pilot required staff reassignment. Participants at this agency agreed to transfer to another provider agency; by all indications, the transfer process went smoothly.

⁵⁵ Training is provided through the Wisconsin Disability Benefits Network. This consortium is chiefly supported by Pathways Projects funding. At the time of writing, benefits specialist training takes ten full days. Substantial (though voluntary) technical assistance has been available after training. Such support appears to be essential. Previous studies of Pathway's projects indicate that it usually takes six to twelve months of on-the-job experience for work incentives benefits counselors to become fully proficient.

⁵⁶ While our ability to examine this question is largely limited to services from the provider agency itself, it is possible that other service providers and funding sources (e.g., DVR, Family Care CMOs, etc) could, to the extent they acquired relevant information from SSDI-EP participants, agency staff, or elsewhere, provide more services to treatment group members than to similarly situated control group members.

employment related services even when provider agencies adhere to the requirement that all participants have equal access to services. Unlike, “discrimination” based upon service provider or funding source knowledge of participants’ study group assignments, these differences in service provision would be “legitimate” within the structure of the experiment.⁵⁷

While we don’t have a method to clearly distinguish between these sources of differences in service provision, we contend that there should be less difference in the patterns of service provision near the time of enrollment than later on. Our claim stems from the reality that it generally takes time to plan how one is going to take advantage of the opportunities that the offset presents. We think two additional factors add purchase to our argument. First, it appears that only a minority of new enrollees have completed or have come close to completing their TWPs.⁵⁸ Second, SSDI beneficiaries are known to be a highly risk adverse population. Beneficiaries have often spent substantial time and effort to establish their eligibility for SSDI benefits. Given the seriousness of their disabling conditions and the possibility those conditions may worsen, many beneficiaries have understandable concerns about replacing both income and access to health care (most typically Medicare) should SSDI eligibility be lost and then being unable to sustain their work effort.⁵⁹

Provider agencies report hours of service provision for every participant on a monthly basis for nine categories of services. To look at differences close to the enrollment date, we chose a period including the calendar month of enrollment and the first three calendar months following enrollment.⁶⁰

Table 1 contains data comparing service provision for the two study groups over participants’ initial period in the SSDI-EP. The typical participant received about the same amount of benefits counseling irrespective of their assignment to treatment or

⁵⁷ In both research interviews and comments made during SSDI-EP technical assistance events, agency staff have reported that they were aware of almost no instances in which DVR or other service funding sources have given more resources to a participant as a result of assignment to the treatment group.

⁵⁸ We cannot verify this assertion until we get access to SSA records. However, given that, to the best of our knowledge, even as late as a year after the final SSDI-EP enrollment only a small number of participants have actually used the offset lends support to this claim.

⁵⁹ We think it important to differentiate between risk adversity as a type of general temperament and that which is acquired by experience and/or observation of the experiences of others.

⁶⁰ As the length of time that a participant spends in the SSDI-EP in their enrollment month can vary from one to thirty-one days, we used a relatively long period to smooth the data. Provider agencies are expected to report all service provision occurring in the enrollment month following enrollment. However, we do not know how much pre-enrollment service, such as benefits counseling to assess eligibility has been included or excluded in the reports for the enrollment month.

One phenomenon that becomes apparent when examining the monthly data within this period is the rather steep declines in average hours of benefits counseling and overall services for both the treatment and control groups. Mean hours for some vocational services exhibit rising or irregular trends over the time period, generally, but not always, to the advantage of the treatment group.

control. Standard deviations are quite similar, suggesting (but not confirming) similar patterns of within group variation. These data strongly support a conclusion that the SSDI-EP had met its commitment to provide equal access to benefits counseling services.

Table 1: Hours of Service Provision, By Treatment and Control Groups, Enrollment Month through the Third Month Following Enrollment

	Benefits Counseling	Any Vocationally Related Services	Any Services
TREATMENT			
Mean Hrs.	3.6	5.0	8.6
Median Hrs.	1.5	0.0	4.0
Standard Deviation Hrs.	4.9	11.8	13.1
CONTROL			
Mean Hrs.	3.8	6.3	10.1
Median Hrs.	1.5	0.0	3.0
Standard Deviation Hrs.	5.3	25.0	25.3

Source: SSDI-EP Research Encounter Data

The data for hours of vocational service provision, as well as the total of all service hours that the provider agencies delivered are still generally consistent with a conclusion that provider agencies complied with the equal access policy. One important finding is the far greater variability in vocational (and as a result, overall) service provision for control group members.⁶¹ The data in Table 2 provides further insight into this result, though it does not explain it.

Table 2: Percentage of Participants Receiving SSDI-EP Related Services from their Provider Agency, Treatment and Control Groups, Enrollment Month through the Third Month Following Enrollment

	Benefits Counseling	Any Vocationally Related Services	Any Services
TREATMENT	60.4%	39.9%	74.3%
CONTROL	60.2%	33.5%	69.7%

Source: SSDI-EP Research Encounter Data

Essentially the same proportion (60%) of participants in both the treatment and control groups received benefits counseling services. By contrast, about 6% more participants in the treatment group received some form of vocationally related services from their provider agency than control group members. Though one could hypothesize that this might be a result of almost immediate gains in treatment group employment rates and earnings compared to the control group, a preliminary analysis of

⁶¹ Across the period of the enrollment through third post-enrollment month, the standard deviation for vocational services was 25.0 hours for the control group, compared to 11.8 hours for the treatment group.

Unemployment Insurance sourced data for early enrollees exhibits no such trend.⁶² It is also interesting to note that when control group members received vocationally related services, they received, on average, more.

While we cannot say for certain whether some portion of the observed differences in service provision were the result of “undeserved” preference for the treatment group, given the relatively small divergence in the proportions getting vocational services and the presence of other explanatory factors, we believe that provider agencies were complying with the “equal access” well enough to not raise serious concerns that any differences between the treatment and control groups will be chiefly due to differential access to services because of study assignment alone. Still, it will be important to see if the early pattern of substantially equal access continues over participants’ full tenure in the pilot.

The annual follow-up surveys include items to elicit participant perceptions about their need for services and whether the services received through the SSDI-EP met those needs. The results described in Tables 3 and 4 are from the first 312 surveys (165 from treatment, 147 from control) received from participants who completed the first annual follow-up. These respondents enrolled in the SSDI-EP before August 2006 and represent over 60% of pilot participants.

Most of these participants, irrespective of study group assignment, tend to agree that they need benefits counseling and vocational services and do so at almost the same magnitudes. However, the data in Table 3 also shows that the percentages of those who “strongly agree” that they need such services is substantially higher for respondents from the treatment group. Given that the surveys were completed roughly a year after entering the SSDI-EP, these findings are consistent with the expectation that treatment group members would be in a better position to exploit available opportunities (or seek new ones) to improve their employment outcomes because of access to the benefit offset. In turn, this suggests having greater need for benefits counseling and other services.

⁶² This analysis was performed on the 116 participants who enrolled by December 31, 2005. There was actually a somewhat higher employment rate for control group members in the enrollment quarter, though post-enrollment quarters one and two exhibited essentially equal rates for both study assignment groups (62% Treatment, 61% of Control in Q1 and 57% for both in Q2). Average earnings were more volatile due to outliers, but again there was no pronounced trend in favor of either group. These analyses were provided to the Social Security Administration on May 8, 2006 via an e-mail communication. Trend analyses including more participants were presented at the November 2007 AEA meeting.

Table 3: Participant Responses to Service Need Items, One Year after Enrollment, by Valid Percentage⁶³

	Strongly Agree	Somewhat Agree	Neither Agree or Disagree	Somewhat Disagree	Strongly Disagree
TREATMENT					
Needed Benefits Counseling	46.0%	22.4%	11.8%	8.7%	11.2%
Needed Job-related Services	40.9%	18.2%	22.0%	5.0%	13.8%
CONTROL					
Needed Benefits Counseling	33.3%	30.5%	12.8%	9.9%	12.8%
Needed Job-related Services	25.2%	22.3%	21.6%	8.6%	21.6%

Source: SSDI-EP Research First Annual Follow-up Survey

Note: Rows may not add to 100% because of “could not code” cases

The data in Table 4 speaks more directly to participant opinions about the value of the services they received (or perhaps more precisely, thought they had received) through the SSDI-EP. Respondents from the treatment group were far more likely to report that they “strongly agreed” that the benefits counseling services they received met their needs than did control group respondents (41% versus 18%). This finding suggests an important question: is the control group’s lesser satisfaction with the benefits counseling services received mainly an assessment of the value of those services or is it related to their dissatisfaction with being assigned to the control group and/or the reduced opportunities that may have resulted from assignment to the control group?

The data reported in Table 4 for job-related services exhibit a somewhat different pattern. The distribution of treatment and control group responses are quite similar, albeit the distribution is slightly tilted toward the “negative” in the case of the control group. Nonetheless, both groups were more likely to report neutral or negative assessments of how well vocational services met their needs than positive ones.

⁶³ The number of non-responses on the items described in Tables 3 and 4 is never greater than 2. Similar non-responses numbers also apply to other survey results presented in this section.

Table 4: Participant Responses to Service Quality Items, One Year after Enrollment, by Valid Percentage

	Strongly Agree	Somewhat Agree	Neither Agree or Disagree	Somewhat Disagree	Strongly Disagree	Did Not Receive Service
TREATMENT						
Received Benefits Counseling that Fit Needs	40.6%	22.5%	13.1%	5.0%	8.8%	10.0%
Received Job-related Services that Fit Needs	17.4%	20.5%	19.9%	9.3%	12.4%	19.3%
CONTROL						
Received Benefits Counseling that Fit Needs	18.4%	29.1%	20.6%	12.1%	7.1%	11.6%
Received Job-related Services that Fit Needs	14.2%	17.0%	20.6%	8.5%	14.2%	25.5%

Source: SSDI-EP Research First Annual Follow-up Survey

Note: Rows may not add to 100% because of "could not code" cases

The results in Table 4 can be further explicated by what participants told us during the spring 2007 focus groups. Participants at all the focus groups, save one, clearly identified benefits counseling services as the main service provided to them by their provider agency. By contrast, when asked to discuss what employment services and supports their agencies had provided, focus group participants usually spoke about services arranged through the Division of Vocational Rehabilitation and other sources. Though sometimes participants would mention that the provider agency had an important role in helping them connect to a service provider or a funding source, just as often participants would talk about their own efforts or those of some third party.⁶⁴

Table 4 also contains information about the percentages of participants who report not getting services. On one dimension the reported findings reinforce the patterns exhibited in the encounter data reported earlier in this section. Again, roughly the same proportion of treatment and control group members received benefits counseling, while the disparities in the proportions getting vocationally related services are of similar magnitude and low enough to be consistent with the SSDI-EP policy of equal access. Nonetheless, there are huge discrepancies between the two data sources

⁶⁴ It is important to note that by chance none of the 2007 focus groups were held at the agencies that, based on monthly encounter data, report comparatively high average hours of employment related services.

in the percentages of participants receiving services. The encounter data suggests about 60% of participants received benefits counseling, the survey data suggests 90%. At first glance, the differences in the time periods contained in the assessment (3-4 months versus a year) would seem to be a likely explanation.

The problem with this explanation is that encounter data from the provider agencies show that few participants receive many hours of additional services from their agencies after the first months in the pilot.⁶⁵ Moreover, the newly reported hours, especially for benefits counseling, are more likely to be for additional than de novo services. It is more likely that participants did not distinguish between services that they received from the agency because of SSDI-EP participation and those they may have received from the agency for some other reason. Our experiences observing SPI and other projects suggest that many consumers do not make that kind of distinction. Moreover, the SSDI-EP did not require a provider agency to perform a full benefits analysis for an enrollee if one had been performed within the previous year and there had been no important changes in the enrollee's circumstances. Thus, some of the 30% difference in the two indicators of whom did not receive benefits counseling services may be due to some participants not making distinctions between services received before and after their entry into the pilot.

Participant Involvement and Data Quality

An inherent challenge in many experimental designs is maintaining participant involvement, particularly of those assigned to a control group. In many experimental designs control group members receive little or nothing of value, yet may bear costs, such as the time and effort involved in providing information for evaluation purposes. Furthermore, in an experiment of long duration the implicit ratio between perceived costs and benefits may be expected to gradually worsen. A likely outcome of these conditions is that control group members either withdraw or reduce their cooperation in significant numbers. It then may not be possible to get a good estimate of intervention effects even when random assignment had been well implemented. Certainly, the Wisconsin pilot is structured in a manner that makes this scenario plausible. Control group members are expected to provide data on a monthly and annual basis and, at minimum, the "active" phase of data collection from participants will last between eighteen and thirty-two

⁶⁵ This fact raises an important implementation issue. Did participants, even in the treatment group, have good access to benefits counseling services throughout the pilot? Our initial look at the data raises concerns as to whether there has been adequate access, as average monthly hours of service provision, including benefits counseling, have been quite low across the project and most providers. We think the phenomenon deserves further investigation because of the possibility that inadequate amounts of support services would seem to have the potential to affect participant outcomes. It would be unfortunate if a "no difference" result was motivated by the absence of benefits counseling and other services that the SSDI-EP's designers argued were critical for effective use of a benefit offset.

months.⁶⁶ Though it is likely that SSDI-EP participation improves access to benefits counseling and other services, it does not directly bestow eligibility.⁶⁷

In addition the continual involvement of those assigned to a treatment group cannot be assumed. Again, the passage of time may be an important factor in lessening involvement. Additionally, involvement could wane because of problems in delivering important features of the intervention or as a result of reduced or ineffectual efforts to maintain communication with participants. Again, substantial reductions in involvement are likely to compromise the ability to estimate impacts. Again, the SSDI-EP and its evaluation are at least potentially vulnerable to these concerns.

In fact, one of SSA's purposes in launching the benefit offset pilots was to look at how well the pilots maintained communication with participants and secured their continued involvement. The lessons learned could then be used to shape the implementation of a national demonstration of one or more versions of an offset and associated support packages. Though all four of the benefit offset pilots are potentially subject to declines in participant involvement, the Wisconsin pilot, due to its decentralized organization, has greater potential for variations in involvement depending on which agency participants are linked with. This decentralization also raises the issue of whether an agency with poor performance in this area would respond to central program efforts to improve that performance, given the agency would be outside Pathways' or DHFS's direct control.

Still, are these potential problems serious? Available data suggest that so far they are not. Though all but one of withdrawals to date (other than for reason of death) came from the control group, this represents only a 5% rate. Nonetheless, available survey data (see Table 5 below) about overall satisfaction with the SSDI-EP indicate noticeably lower satisfaction levels among control group members. A control group participant is about half again as likely to be neutral toward the pilot as a treatment group member and about twice as likely to express clear dissatisfaction. Also, there are data already presented indicating that control group members are less sanguine about the value of SSDI-EP provided or brokered services.

⁶⁶ This range reflects the August 2005 to October 2006 enrollment period. At the time of writing this report, it was known that the SSDI-EP and its provider agencies would be operating through at least March 2008. The authors have no information about whether the pilots will operate beyond spring 2008 or when SSA will make a decision on that possibility.

⁶⁷ As previously described, the only "mandatory" service is benefits counseling. However, neither a decision not to join the SSDI-EP or withdrawal from it affects eligibility per se under even the OIE grant (and certainly not for any external funding source).

Table 5: Participant Overall Satisfaction with the SSDI-EP, One Year after Enrollment, by Valid Percentage

	Very Satisfied	Somewhat Satisfied	Neither Satisfied or Dissatisfied	Somewhat Dissatisfied	Very Dissatisfied
TREATMENT	37.3%	24.8%	21.7%	8.1%	6.8%
CONTROL	14.2%	21.3%	36.9%	13.5%	14.2%

Source: SSDI-EP Research First Annual Follow-up Survey

Note: From surveys returned from participants with one year of participation by July, 31, 2007

Note: Rows may not add to 100% because of “could not code” cases

Whatever the long term effects of the greater dissatisfaction within the control group may prove to be, so far there is little evidence of its manifestation in participant behavior. Survey return rates have been high and nearly uniform at 90% for the treatment group and 89% for the control group, with generally low rates of missing or uncodeable responses.⁶⁸

Another indicator of only modest differences between the involvement levels of treatment and control group members is the respective proportions of monthly participant update forms received within one calendar month of the reporting period. For July 2007 the rate was 93% for the treatment group. The control group rate was 89%. Though these rates are not greatly different, there has been a gradual increase in the amount of divergence that suggests somewhat decreasing cooperation among those in the control group. Only nine months before in November 2006, the rates were the same. Yet, some caution is advised in viewing these results as suggesting even modest declines in control group involvement, as completion rates reflect both provider agency activity and participant involvement.

Therefore, it is important to look at what is known about provider agency efforts to maintain contact with participants, a factor that could potentially vary in both intensity and method across provider agencies. Agencies are required to submit two types of encounter data forms to the evaluation team on a monthly basis. The first, a participant update about employment and living situation, requires contacting participants and eliciting their cooperation in providing relevant information. The second instrument, called the case-noting form, documents the hours of services that the agency has provided to each SSDI-EP participant. No participant contact is needed to provide this data.⁶⁹ So except for circumstances such as staff attrition or serious illness, the monthly completion rate for the case-noting forms should be 100%. However, completion rates (for within a month of the reporting period) have been essentially the same as those for

⁶⁸ There has been a slow decline in return rates over the successive waves of surveys that have been mailed. We will be watching this trend closely as the second annual follow-up survey rolls out, both for overall response rates and those for the two study groups.

⁶⁹ Gathering the data may require some staff coordination and/or maintaining contact with external parties who have been contracted to provide benefits counseling or some other service. It is likely that some agencies pay more attention to doing this than others. There is some information from the 2006 provider agency staff interviews that identifies these differences.

the update forms, exhibiting the same slowly increasing difference between treatment and control group submission rates. For example, the July 2007 rates were 94% for the treatment group and 89% for the control group.

The first participant follow-up survey included a question that asked participants whether agency staff contacted them regularly to talk to them about their employment status, progress toward attaining employment goals, etc. Participants when entering the pilot were informed that they would be asked for certain information on a monthly basis and participant responses at the focus groups indicated widespread understanding of this. Yet Table 6 exhibits a response pattern that suggests that the regularity of this contact appears to be significantly less than implied by the forms completion rates.⁷⁰

Table 6: Participant Reports of the Regularity of Contacts by Agency Staff about Job-Related Activities, One Year after Enrollment, by Valid Percentage

	Strongly Agree There Is Regular Contact	Somewhat Agree There Is Regular Contact	Neither Agree or Disagree There Is Regular Contact	Somewhat Disagree There Is Regular Contact	Strongly Disagree There Is Regular Contact
TREATMENT	40.4%	16.1%	10.6%	14.3%	18.6%
CONTROL	27.0%	22.7%	12.1%	14.9%	22.7%

Source: SSDI-EP Research First Annual Follow-up Survey

Note: From surveys returned from participants with one year of participation by July, 31, 2007

Note: Rows may not add to 100% because of “could not code” cases

The differences between the treatment and control group response distributions are substantial compared to those for some other survey questions about participant experience in the pilot. The pattern suggests that control group members, accurately or not, feel that they don't have regular contact with staff on employment related issues. However, perhaps the more striking pattern is that a significant minority of those in the treatment group (33%) don't feel they have regular contact either. It is possible that participants see the question in a somewhat different context than we expected. These results would be more intelligible if it could be established that many of the survey respondents understood “regular contact about job-related activities” as referring to contacts involving some form of benefits counseling or discussions of service provision and not contacts aimed at getting information for the monthly update form. Indeed, any single staff/participant contact can involve both elements.

Focus group results cast additional doubt on whether the approximately 90% update form return rate accurately describes the saliency or importance of participant contacts.⁷¹ Participant feedback obtained from the five spring 2007 focus groups

⁷⁰ We cannot know what timeframe participants had in mind when they answered the survey question as we used the word “regular” rather than “monthly.”

⁷¹ Restated, not all contacts may be important to participants. It may well be the frequency reported is accurate, but not all types of or occurrences of contacts matter as much for developing or maintaining involvement. For example, the delivery of effective benefits counseling

suggests the possibility of substantial variation across the twenty plus agencies. For example, at two focus groups, all participants had indicated that there was generally monthly contact, but qualified it by saying that in some instances they had initiated the contact and, in fact, were comfortable doing so. Our impression from the sessions was that in the majority of cases a participant thought the agency would have initiated contact within a short time period if he/she had not initiated contact, but that this was not universally so. Several participants at one agency indicated going several months without contact.

Provider agency staff members have often told a different story, one in which multiple attempts to reach some participants met with no useful response. Repeated attempts to contact participants take time away from actual service provision and are poorly compensated if successful, uncompensated if not.⁷² Some provider agencies take actions to contain these costs such as turning data collection over to clerical staff either not familiar with the participant they are trying to contact, the SSDI-EP, or with the terms or categories used on the update form. More recently, agencies are using contact methods such as e-mail. Though using e-mail may help agencies to control costs, that method usually makes it impossible to ask participants clarifying questions in real time. In any case, we have no way to assess the relative impact of the limited effort at least some provider agencies give to maintaining regular contact with participants compared to the reluctance of some participants to respond to agency contacts.

Nonetheless, from an evaluation perspective the real question is whether provider staff members are submitting accurate information, irrespective of whether it is submitted within a month or substantially later.⁷³ The evaluation team is highly

services when a participant gets an overpayment or CDR notice may do far more to maintain involvement than a monthly check-in to get information for the participant update form.

⁷² Provider agencies are compensated for providing a pair of encounter forms, i.e., the update and the case-noting. This may be one reason that submission rates for case-noting forms are about the same as those for the participant update forms.

⁷³ To the extent that the quality of information collected on participant update form is dependent on the accuracy of participant recall, we are concerned that delayed collection of information means less accurate information. Given the cognitive and/or affective conditions that some participants deal with, these concerns are strengthened. While provider agency staff members are in some position to verify these data, the time and effort involved make it less likely that such verification will be done for participants, especially control group members, outside a context involving establishing or maintaining eligibility to some cash benefit or service. It is unlikely that the timing of these efforts will match monthly reporting schedules.

Beginning in September 2006, agencies were able to indicate on the update forms that they had been unable to contact the participant. Agencies were obligated to provide accurate retrospective information if and when contact was reestablished. This procedure brought two (likely) benefits. First it reduced the likelihood that an agency would report there were no updates when there had in fact been no contact. Second, it encouraged submission of an update form that would give the research team a more truthful report of the situation, i.e. the data was missing and some indication of why that was so (e.g. refusal of the participant to respond, lack of a current address or phone number). Still, the “unable to contact” option is potentially subject to gaming. Nevertheless, though we expected a gradual rise in the numbers of “unable to contact” reports, the monthly number stabilized at about forty-five. Incidence for the control group occurs at only a marginally higher rate than for treatment group.

dependent on the provider agencies to assure the quality of the encounter data, as the evaluation team has very limited ability to perform independent assessments in this area. To the extent we can check on the quality of data submitted, we have found reasons for concern. The level of the concern varies both by agency and, when more than one staff member does reporting, by staff member.

Every research form submitted by a provider agency is checked for missing items and inconsistencies. Sometimes other problems can be identified from agency staff comments added to the form or based on the evaluation team's familiarity with participant history, area employers, etc. Evaluation staff meets weekly to discuss identified problems and how they might be resolved. When corrections can be made based upon information on hand, they are made. Often we need to contact the provider agency for clarification or more information. In turn, the provider agency may need to contact the participant. Most problems are resolved within days; in a few instances resolution has taken months. We have not tracked the total number of changes made through this process, though the number must be well into four figures. Yet almost all of these corrections still depend on the reliability of what is reported to us.⁷⁴ Once basic completeness and consistency issues are settled, we can do little more to assess data quality. What follows are two examples where we have been able to look more deeply into the quality of the encounter data.

The SSDI-EP evaluation uses employment data obtained from Wisconsin Unemployment Insurance (UI) records and from encounter data, specifically items on the enrollment and monthly participant update forms. The UI data has considerable virtues, but significant flaws as well. The most important of these is that certain employment does not need to be reported, including important categories such as self-employment and jobs at employers located outside Wisconsin. By contrast, the self-reported data from the provider agencies has one important virtue. In principle, all employment may be captured.

Table 7 presents information about how often employment information derived from UI and form data match. Employment status reflects any report of employment, irrespective of duration, within the relevant time period. In this case, the time period is the first post-enrollment calendar quarter for all participants.⁷⁵ The upper left and lower right cells report data where the two data sources are consistent. As expected, these two cells contain most of the cases. The lower left cell presents cases of employment reported through the forms, but not through UI. We anticipated that this cell would be populated. The number and percentage reported seem plausible given exclusions from the UI system. However, it is the cell on the upper right that is of most interest: The UI system captured twenty-four instances of employment in the quarter that for whatever reason (lack of contact, participant unwillingness to reveal, poor interview technique)

⁷⁴ This is also true for administrative data from state agencies and SSA itself. However, most administrative agencies have auditing procedures that can be expected to significantly reduce the probability of errors.

⁷⁵ Ideally, we would have liked to do this analysis for later quarters of participation. However, at present, we can only use the first post-enrollment quarter. The most recent UI data in hand at the time of writing are those from the first quarter of 2007. The last SSDI-EP enrollees entered the pilot in the fourth quarter of 2006.

were not reported on the forms. Interestingly, there is no significant difference in the patterns found in Table 7 when data from the treatment and control group are separately examined. From this we infer that whatever the reason(s) cases of UI employment aren't captured in encounter data, the occurrence is not an artifact of participant or provider agency responses to random assignment.

Table 7: Cross-tabulation of UI Reported and Encounter Form Reported Employment Status by Number and Proportion of All for Participants, First Calendar Quarter following the Enrollment Quarter

	Reported Employed by Encounter Forms	Not Reported Employed by Forms
Reported Employed by UI Records	229/46.6%	24/4.9%
Not Reported Employed by UI Records	80/16.2%	158/32.2%

Source: SSDI-EP Research Encounter Forms and Wisconsin UI Data

The actual failure rate of encounter forms to capture UI reported employment is certainly underestimated in Table 7. Later in this section, we describe a follow-up data cleaning procedure they we perform on roughly an annual basis to augment the weekly process. This “work history exercise” resulted in a significant number of additional corrections. The first (and to date only) work history exercise was limited to encounter data reported through late February 2007. Thus, only the SSDI-EP participants enrolled in October 2006 had any part of their first post-enrollment calendar quarter excluded from the work history exercise.

For the present, we have no calendar quarter of data that has not been at least partially corrected through the work history exercise. However, if we look at data organized as in Table 7, but for the first quarter of 2007, irrespective of when participants enrolled, we can explore whether our expectation that data uncorrected through the work history exercise will contain a higher proportion of cases where the forms data misses UI reported employment. This expectation is confirmed. There are forty-one of these reports, constituting 8.4% of all cases in the first quarter 2007 data.

Another concern that arises from our ongoing review of encounter forms is that personnel at the agencies often do not apply data element definitions correctly or seem simply unaware of them. Of course, this is an expected challenge in any research when one is dependant on data collection by persons who are not researchers and have other important responsibilities to study participants. Our ongoing attention to training, technical assistance, and data cleaning is how we attempt to address this challenge. Yet there is another factor that may be at work for which it can be difficult to find direct evidence. Some agency staff members appear to want to put the most positive face on the situations they are reporting.⁷⁶ This appears to happen when reporting reasons for

⁷⁶ During SPI, it was clear that at least a few agency staff doubted that the evaluation team could or would refrain from giving individual data to DHFS, DVR, or even participants' employers. There was a concern that “negative” information might affect access to services or future employment

job losses or the characteristics of a new position. Whatever our suspicions may be, we cannot generally determine whether they are justified.

Table 8 provides data that suggests that our concerns have a genuine basis. When an agency staff member reports a job for the first time, he/she is expected, based mainly on participant supplied information, to assign the job to one of several job classification categories. Given combinations of information we already have about participant education and the new information about (imputed) hourly wages and the type of industry where an individual is employed, we are sometimes doubtful about the veracity of the reported job classification. For example, when a participant's new job is classified as a skilled craft in a manufacturing setting and the hourly wage is \$6.00, we suspect the staff member is either exaggerating or has little understanding of what a skilled craft involves.⁷⁷ Nonetheless, it is always possible that any single outlier might be an accurate report.

However, we find it more difficult to believe that multiple outliers are credible. Most of our job classification categories do not have explicit education or training requirements or they may be quite hard to align with the categories we use to measure educational attainment.⁷⁸ One of the categories does. One criterion for classifying a job as a "professional" on the enrollment or update forms is that the position normally requires a baccalaureate or graduate degree. Twenty-seven (45%) of the sixty reported professional jobs are held by persons without a baccalaureate. This finding is simply inconsistent with the definition that staff members are required to use. With somewhat less confidence, we assert that the results are inconsistent with our understanding of the overall job market and the oft voiced claim that college trained persons with serious disabilities have particular difficulties getting the professional positions for which they had prepared.

opportunities. To date, no similar comments have been offered by agency staff working with SSDI-EP participants.

⁷⁷ Examples of a "skilled craft" in a manufacturing situation might include a tool and dye maker or an electrician. The instructions for the enrollment and update forms provide substantial guidance as to the level of capability a skilled craft entails and the amount and types of training and experience generally associated with that capability.

⁷⁸ These categories and to a large degree the others used to capture employment data on the enrollment and update forms have their origins in the categories Virginia Commonwealth University developed for use across the SSA funded SPI projects. These were somewhat modified for use in Wisconsin's SPI project and then for the SSDI-EP. This continuity was less about the inherent worth of the VCU data categories and their definitions as to allow some comparability between SPI and SSDI-EP data. The intent was to be able to link data collected in SPI with data collected during the SSDI-EP for former SPI participants (for those who gave consent to do so). As reported, not enough of the former SPI participants enrolled to fulfill this intent.

Table 8: Number and Percentage of Participant Jobs Classified as Professional, by Number and Percentage, through September 11, 2007

Reported Education Level	Number of Positions Classified as Professional	Percentage of Positions Classified as Professional
12 Years or less without High School Diploma or GED	0	0%
High School Diploma or GED	3	5%
Some College, no Degree	9	15%
Completion of Vocational or Technical Training or Two Year Degree Program	15	25%
Four Year College Degree	26	43%
Some Graduate School or Graduate Degree	7	12%

Source: SSDI-EP Research Encounter Forms. Total number of jobs classified as professional equal sixty.

We conclude this section by presenting the results from the previously mentioned work history exercise conducted in spring 2007. Briefly, the process began with a review of all participants' form reported employment records with the aim of identifying missing or conflicting information about when jobs started, when and why jobs ended, changes in monthly work hours, and the status of reported layoffs and leaves of absences. Provider agencies were then given records of every one of their participants. While agencies were asked to provide or seek information to clarify 351 highlighted issues, they were also asked to make sure that there were no unreported jobs and to double check the status of jobs for which no change had been reported for at least one year. While agencies with larger numbers of participants generally had larger numbers of highlighted issues, the number and potential seriousness of these issues disproportionately clustered at agencies where there had either been staff attrition or where the evaluation team had made a disproportionately higher number of contacts in the context of routine data cleaning.

The work history exercise was "completed" in July 2007, with four of the identified issues remaining unresolved. We made changes to 105 encounter forms, including the addition of twenty-five previously unidentified jobs and the end of thirty-six positions. Since July, we have received reports allowing us to make nine additional corrections to the encounter data for before March 2007. Only two of these were for the unresolved cases. Thus, agency staff found seven new issues that neither the evaluation team nor they had identified through the work history exercise.

To summarize, we have identified both global and agency specific challenges in both maintaining the contacts that support participant involvement in a long project and in the quality of encounter data provided through those contacts. To date any problems do not seem to be of a scale that seriously threatens the capacity to evaluate the Wisconsin pilot. Most participants continue to complete surveys and provide encounter data, though attitudinal data suggests lower levels of satisfaction among control group

members. There have been some problems in provider agencies' data collection, but current data cleaning processes appear to be assuring that at least encounter data about the timing of job starts and endings will be quite accurate, if information about other job characteristics less so.

Offset Administration

Reasonably successful implementation of the SSDI-EP in the areas of recruitment, enrollment, service provision, and participant contact are essential conditions for assessment of whether the benefit offset motivates improved employment outcomes. Yet, even perfect implementation in all these areas would not insure "evaluability." For the SSDI-EP, it is only the impact of the benefit offset itself that is being explicitly tested. So unless the offset is adequately implemented there will not be a "clean test." The SSDI-EP is designed to keep everything else subject to program control as constant as possible to make the clean test of the offset possible. That is the main reason this paper focused on the consequences of the SSDI-EP's decentralized structure, given its potential to result in variation in the performance of functions that the pilot is ultimately accountable for.

The SSDI-EP, however, does not administer the benefit offset. It does not decide when a SSDI benefits check should be adjusted or whether a scheduled disability review (CDR) should be postponed until a treatment group member's participation has ended. Those jobs are performed by the Social Security Administration. As participants in the control group have their checks processed and continuing eligibility determined no differently than had they not volunteered for the pilot, any problems involving their benefit checks or SSDI eligibility after pilot entry do not directly affect the capacity to evaluate the offset.⁷⁹

By contrast, SSA had to create new processes to monitor the status of treatment group members and to take appropriate actions under the rules it had established for the four state pilots. Beyond creating "organizational space" for a novel, small scale task to be performed within a large, highly routinized bureaucracy, these rules and procedures were aimed at insuring that an important goal would be met: that no beneficiary would be disadvantaged by his or her participation in a benefit offset pilot. SSA decided that offset administration would be best performed at agency headquarters in Baltimore, using staff in the Office of Central Operations (OCO) who would not be supervised by the staff at the Office of Program Development and Research. It is these staff who are directly involved in overseeing the pilot projects.

The "evaluability" of all four pilots would depend on the capacity of OCO staff to make timely and accurate decisions about benefit payments and other matters where assignment to the treatment group potentially changed SSDI program rules. Poor implementation could, for example, result in some individuals going months without any cash benefit if the OCO did not promptly recognize that a treatment group member past her/his TWP had earnings above SGA. Similarly, in cases where a treatment group member past TWP had reported SGA earnings but OCO did not promptly adjust the

⁷⁹ It is plausible that treatment group member perceptions of the "background" rates of problems around check calculation and CDRs attendant to work activity may color their perceptions of how well the offset is being implemented.

benefits check, the participant might get a large overpayment and then face the possibility of having substantially reduced benefit payments and/or be expected to make a considerable repayment to SSA.⁸⁰ Finally, failure to correctly respond to changes reported on modified earnings estimates creates potential for incorrect payments either lower or higher than what the treatment group member should have received.

Such individuals might conclude that the offset wasn't worth the trouble and thereafter strive to keep their earnings safely below SGA. Other treatment group members learning of such implementation problems might decide not to complete or initiate their TWP's to avoid ever coming under the pilot rules governing benefit payments.⁸¹ Should poor offset administration lead to these behavioral effects, it would imply that there would be substantial underestimates of the benefit offset's "true" impact. It also would mean that some participants would have been disadvantaged by their participation. Though we have little doubt that SSA would eventually "make good" on underpayments and take action to mitigate the impacts of many overpayments, damage will still have been done.

The evaluation team has heard reports of serious problems in offset administration. Much of the evidence is anecdotal and second hand. Available documentation is rarely in quantitative form, though its consistency and duration is convincing. In reviewing the monthly status reports that the SSDI-EP has provided to SSA, we find ongoing reference to the same issues among which the following have been especially prominent:⁸²

- Delays and errors in determining TWP usage
- Delays and errors in applying the offset for eligible participants with over SGA earnings
- Delays in the approval of work subsidies or Impairment Related Work Expenses (IRWEs)
- Overpayments and their resolution

Similar patterns of issues can also be identified from the notes for the monthly calls among the four states operating benefit offset pilots and the materials that each state provided to SSA and Abt Associates in March 2007.⁸³

⁸⁰ Participants could also have overpayments from periods before the pilot.

⁸¹ It is also possible that reports of widespread problems in offset administration would affect the behavior of benefits specialists and others who advise or provide services with treatment group members, some of whom would work at entities not connected with the SSDI-EP (or the pilots in the other three states).

⁸² There has been gradual change in this list over time. For example problems about eligibility determination or resumption of SSDI-EP payments for (qualified) treatment group members between the end of their EPE and month seventy-two were more prominent in earlier monthly reports.

⁸³ Abt Associates was contracted to design the national demonstration, conduct its evaluation, and, it appears, to have a major role in the management of the demonstration itself.

Nonetheless, we have little direct information from either SSA or treatment group member that allow us to assess the scope of any offset administration problems for Wisconsin treatment group members, let alone their behavioral consequences. In the absence of SSA administrative data the evaluation team does not even know the total number of treatment group members that have completed their TWP. We do not have reliable information about either the current number of persons with SSDI checks being adjusted because of the benefit offset or the number for whom the offset has been applied at any time.⁸⁴

Informants from SSDI-EP central operations recently estimated the number of individuals currently in offset status at twenty-nine. The estimated number of treatment group members who had completed their TWPs and would, if earning above SGA, be able or soon able to utilize the benefit offset was fifty-five.⁸⁵ This information was cobbled together from multiple sources, but that about offset participation was gathered in the context of the identification of problems about offset administration from provider agency staff or directly from participants and of subsequent efforts by SSDI-EP staff at Pathways to resolve these.⁸⁶ Thus, it is conceivable that the twenty-nine is a substantial undercount because it excludes cases where OCO's application of the offset was both timely and accurate and, thus, no problem occurred that would bring these cases of offset participation to the SSDI-EP central office's attention. While recognizing this possibility, our informants expressed doubt that there could have been many such cases. Benefits specialists at the SSDI-EP provider agencies would have been aware of most, if not all, of such cases.⁸⁷

⁸⁴ We do not yet know whether it is possible to confirm offset application and its amount from specific data elements in SSA records. It may need to be inferred from a combination of TWP completion data, payment records, and other information from a participant's Master Beneficiary Record. At the time of writing, we have not seen any SSA administrative data. The process of getting such records has been long and involves, among other things, having state level security plans that meet SSA's understandably stringent standards.

⁸⁵ The offset is not actually applied to the SSDI benefit until three months after the end of the TWP. That is because under regular SSDI rules beneficiaries entering their EPE and earning above SGA would get their full benefit amount during this period. This is another manifestation of SSA's commitment that no one in the treatment group should be worse off than he/she would have been under the regular rules.

⁸⁶ To improve "trouble shooting" of offset administration problems and to monitor treatment group employment and earnings reporting to SSA, SSDI-EP operations staff in Madison have been in the process of constructing a database. Though the evaluators will have some access to this information, it has limited value as an alternative to SSA records for evaluation purposes because it does not include information for the control group.

⁸⁷ It is also likely that most, if not all, such cases would come to the attention of either agency or SSDI-EP central benefits specialists in the context of SSA reconciliation activities. Unfortunately, these activities take place toward the end of the year following the year to be reconciled. As of early October 2007, SSA has not provided the SSDI-EP with the list of 2006 cases that require reconciliation by gathering W2s and other earnings information. The SSDI-EP has urged provider agencies to gather such information for treatment group members by the April 15 tax filing date. Until SSA initiates reconciliation procedures it will not be known how consistently provider agencies performed this task. (Note: only five reconciliations needed to be done for 2005.)

Instead, benefit specialists at the agencies and SSDI-EP central staff report that alterations to or resumptions of benefits checks typically take months to be processed, resulting in either substantial overpayments or underpayments. They report some recent improvement in processing time, though delays of three months or more are said to remain common. Moreover, application of the offset does not mean that a check will be correct. Our informants report that most checks are initially incorrect, including those in the handful of offset cases that agency personnel reported were processed more rapidly than the norm.

Though we have no ability to directly observe SSA operations, we are aware that SSA has taken substantial steps to improve offset administration. Early in the project OCO staff members performed offset administration on a very temporary basis before rotation to other duties. Staffing became more permanent in early 2007 and this summer formal efforts began to achieve greater coordination among these staff. OCO also tightened up the timing of reminders to beneficiaries to return critical documents needed to assess continued disability status following the completion of the trial work period (particularly the "821" form that beneficiaries use to report work activity). Nonetheless, these efforts do not appear to have had their intended effect. Moreover, SSDI-EP staff (and those in other states) still report problems about getting timely information about SSA communications to treatment group members and/or SSA actions. These conditions date back to project start-up and have not yet been effectively addressed.⁸⁸ Another issue is that OCO often sends treatment group members standard correspondence that make no useful reference to the relationship between pilot rules and any decision or reporting requirement the letter is meant to inform the participant about.⁸⁹

However, it needs to be made clear that treatment group members can also act or fail to act in ways that make it difficult for SSA to administer the offset. In particular, some of these participants may fail to modify their earnings estimates or complete 821 forms accurately or in the required timeframes. Benefits specialists at the SSDI-EP provider agencies may differ in how well they encourage or facilitate compliance.

To get a sense of these issues, the evaluation team has included an item to track problems in offset administration on the annual follow-up surveys. Treatment group members were (and will continue to be) asked to identify any problems with their SSDI benefits check in the prior year. Currently, we do not have any ability to distinguish between the responses of treatment group members who had or should have had their payments changed because of the offset from those who were either not yet eligible for application of the offset or never had sufficient earnings to trigger the offset's application. As we do not have any useful estimate about the cumulative number of those with some period of offset use, we have no basis for estimating the numbers of such participants among the 165 treatment group members enrolled by July 31, 2006 who returned their first follow-up survey. However, on the returned surveys 51% of the responses indicated that respondents believed that no changes were made or needed to be made to their

⁸⁸ In the past, SSA argued that certain legal requirements make it impossible to provide correspondence to state project staff. We have been told that SSA has gradually expanded the range of beneficiary notices it has shared with the state pilots.

⁸⁹ There have also been reports of discrepancies between the payment information provided on SSA notices to treatment group members and the actual check amounts.

checks. Another 27% of responses indicated that whatever changes were made had been perceived as both correct and on time.⁹⁰

This leaves a residual 22% of responses where some problem or combination of problems was reported. As expected, problems involved issues like delay in payments or inaccuracy in payment amounts, with the issue of overpayment having particular saliency. The analytical challenge that cannot be dealt with until there is individual data that confirm or support inference of offset usage is whether the incidence of offset related problems are different from the background rates experienced by other beneficiaries who engage in work activity.

In our focus groups, some participants talked about having current or past problems with overpayments and other issues affecting payment. They often complained about SSA's role in causing problems or not responding to their own, their benefits specialist's and/or their case-manager's efforts to resolve problems. Other participants indicated they were familiar with such problems, even if they had not been personally affected. Across all of the focus groups, participants raised concerns about SSA's ability to keep accurate records and, as a consequence, whether transition from the offset at the end of the seventy-two month eligibility period would go smoothly. Another dominant theme was SSA's inability to communicate with beneficiaries in a useful and respectful fashion. Nevertheless, there was no instance of a focus group participant clearly stating that a problem was related to the application of the benefit offset.⁹¹ Given the (apparently) small number of persons using the offset statewide, it is very likely that none of the twenty-six participants who took part in the spring 2007 focus groups had ever had the offset applied to their benefit check.

Still, we think these survey and focus group findings are important for thinking about the potential difficulties that poor offset administration pose for evaluation of the SSDI-EP, the other pilots and, by implication, a much larger national demonstration that may not provide equivalent access to benefits counseling services. In particular, there is significant evidence that the trust relationships built up in continuing one-on-one relationships can help SSA beneficiaries and recipients to overcome their understandable risk averseness to increased work effort and thereby encourage the use of work incentives such as a benefit offset provision.⁹² To state the obvious, even the perception of widespread problems in offset administration is likely to reinforce existing risk averseness among many SSDI beneficiaries and their loved ones.

⁹⁰ Respondents had the option of providing multiple responses; seven exercise this option. Thus the percentages reported are of responses rather than of respondents. The non-response rate for this item was atypically high at 7%.

⁹¹ However, one respondent explicitly linked his overpayment to being in the treatment group. There is some circumstantial evidence that suggests the benefit offset was applied to this participant's check.

⁹² Focus group participants spoke about the importance of the trust relationships with their benefits specialists. Without this trust, it appears to be hard to make use of information that suggests that there would be benefit from employment or increased work effort. Also see Delin, et al., 2004. *Final Project Report: Wisconsin Pathways to Independence*. Menomonie, WI: University of Wisconsin – Stout Vocational Rehabilitation Institute, pp. 123-25 for discussion this point in the context of Wisconsin's SPI demonstration.

We conclude this section with a discussion of SSDI-EP activities that appear to have some influence on SSA's ability to perform benefit offset administration. Some of these activities involve efforts to get information and/or process forms that SSA needs to administer the offset, for example the earnings estimate or the 821 form. Other activities are undertaken to respond to problems flowing out of SSA actions. It must be understood that from a participant's perspective whether a problem attributed to SSA is a consequence of offset administration in any technical sense is probably irrelevant. SSDI-EP participants will probably seek help from their benefits specialist for an overpayment problem whether they are using the offset, a treatment group member not yet eligible to utilize the offset, or a member of the control group.

SSA decided to use annual earnings estimates as the principal source of information for making adjustments to the SSDI benefits checks of those qualified to have the offset applied. It followed that SSA required earnings estimates only for pilot participants assigned to a treatment group. In addition, SSA made provision for within year adjustments, retrospective annual monitoring, and for a modest amount of benefit overpayment to be ignored. SSA did not set up a monthly reporting system similar to that available for administration of the statutory (1619) SSI offset, probably because of the massive resource commitment and data system changes that would be required. SSA allowed each state to design its own means to collect the estimates, though by 2007 SSA became more prescriptive to insure more consistent reporting across the states.

Wisconsin has implemented several versions of an earnings estimate form. Most changes were made in response to inconsistencies in how provider agency staff interpreted instructions for earlier versions, though some changes were made in response to technical issues identified by the agencies. It was not until the May 2007 version was implemented that inconsistent reporting ceased to be a significant problem. Until then it was likely that some of the offset related and other payment problems attributed to SSA had their origin in how agencies completed the estimate forms.

However, a remaining problem is that treatment group members' earnings estimate forms often were not updated in a timely fashion as participant circumstances changed.⁹³ Failure to promptly amend an estimate can be the underlying cause of an "error" at SSA. We have no method for assigning responsibility for delayed earnings estimates between agency staff and participants in the treatment group, but we have noticed differences in the incidence of problems across agencies after taking account of differences in their size. The SSDI-EP central office has tried to ameliorate this problem through periodic reminders to provider agency staff, especially at the start of each calendar year.

For SSDI-EP staff, whether at the provider agencies or the central office, responding to participant issues arising out of SSA actions or communications can take substantial time. Informants tell us that just tracking down basic information usually requires two to three hours of work. We were initially surprised to learn that benefit specialists at the Pathways Project offices estimate that they perform about 70% of the work needed to assess and, hopefully, resolve issues arising out of offset administration. However, this finding seems more credible once several factors are considered. First,

⁹³ The scale of this problem is subject to retrospective estimation using UI earnings data, but would be still underestimated because of exclusion of earnings from the UI records.

SSDI-EP central staff has ongoing communication with SSA staff in Baltimore and access to a secure e-mail system to transfer confidential information. It is likely that the central staff would often need to serve as intermediaries even when provider agency staff continued to be actively involved. Second, participants can and do directly contact the SSDI-EP office. It is likely such contact has increased as treatment group members have moved to locations out of state, beyond their agency's service area, or, for whatever other reasons, contact with an agency has ended or become strained.

Third, the provider agencies often have strong working relationships with staff at local SSA offices. Recall that agencies are expected to provide all participants, including control group members, with benefits counseling services. So it is possible that there is an informal division of labor, with the agency staff concentrating on those combinations of cases and issues that can be worked on at the local level, whether with SSA or other public agencies. It is important to remember that even when SSA actions are timely and accurate, they can have unintentional side effects that may threaten beneficiaries' access to other public benefits and services, including some needed to exploit the benefit offset.⁹⁴

Conclusion

In the introduction to this paper and in much of the subsequent material we emphasized the possible role that the highly decentralized enrollment, delivery and reporting structure of the SSDI-EP may have in producing implementation challenges that might make it more difficult to evaluate participant outcomes, despite utilizing the recognized "gold standard" of random assignment. It is with some irony that we have found that the greatest challenges to the "evaluability" of the Wisconsin pilot come not from the decentralized delivery structure, but from the decisions and actions of the pilot's "unitary" actors, Pathways and the Social Security Administration.

The first major challenge, that of not enrolling into the pilot a substantial number of those who had participated in Wisconsin's SPI demonstration, means there can be no useful analysis of the differences of outcomes between those with substantial previous experience with a combination of intensive benefits counseling and person centered vocational planning services and new participants who could not have experienced these services, at least not in significant proportions. This implementation challenge can no longer be addressed. Nonetheless, this failure, one of planning as well as implementation, has no impact on the ability to estimate employment related outcomes attributed to the offset or to perform any of the secondary analyses identified in the evaluation plan for the SSDI-EP.

The second major challenge is that of the potential for major problems in the administration of the benefit offset to discourage members of the treatment group participants from using the offset provision to increase their employment rates, earnings, and other work related outcomes to the extent presumed to be possible in the absence of major problems. It is also possible that widespread problems in offset administration

⁹⁴ For example, premiums for Wisconsin's Medicaid Buy-in are increased more by an additional dollar of SSDI payment than an additional dollar of earnings. As a consequence, for a treatment group member beyond the TWP, the increased net income resulting from the offset may in some cases generate a premium increase that can, excuse the expression, more than offset the offset.

or perceptions thereof could negatively affect participant willingness to remain involved with the pilot, leading to difficulties in data collection or even outright attrition from the treatment group. The consequences of this challenge, if it is not adequately addressed, are potentially fatal to both the estimation of outcomes and the identification of specific “best practices” to apply to a national demonstration.⁹⁵ However, given the relatively small numbers of treatment group members currently in offset status in Wisconsin and, as we understand it, elsewhere, there is still time for remediation. We are pleased to have seen changes at SSA to improve administrative capacity for offset administration and the emerging cooperation with the four state pilots’ Policy and Operations Task Force.

None of this is to say that the decentralized structure had no consequences in the failure to enroll a useful number of former SPI participants or the ongoing difficulties in the administration of the benefit offset. The action or lack of action by each provider agency often increased and sometimes ameliorated the challenges that had their origin mainly with what Pathways or SSA did. It remains to be seen whether cross agency variation in implementing recruitment and enrollment processes, service provision and referral, activities supporting participant involvement, and research reporting are at a scale that calls into question whether an overall intervention that is acceptably close to what Pathways specified was in fact provided. As noted, we have some concerns about the quality of some of the data reported by provider agency staff, especially when a data element was not included among those checked through the work history exercise. However, our tentative judgment is that in most areas the range of variation across provider agencies is within that to be expected in a multi-site demonstration, especially one where the sites are not under the control of a single organization. Given this, Pathways’ decision to organize the SSDI-EP in ways that mirror systems, in Wisconsin and most other states, with multiple providers and multiple funding sources for providing vocational and support services to persons with severe disabilities appears fully justified.

Finally, we want to highlight an important evaluation challenge that was identified, but only briefly discussed, in the short discussion of Wisconsin’s evaluation choices. We do not know how long the pilot will continue and thus whether there will be adequate time to amass and analyze the desired length of participant data. For most purposes, this period is specified as eight calendar quarters of participant data following the calendar quarter of enrollment.⁹⁶ The last participants enrolled will reach this milestone on December 31, 2008. Receipt of administrative data and the analysis itself will take considerable additional time.

This is not to say that the SSDI-EP will not have had substantial value if its length is truncated at some earlier point. There has already been useful process information and there will be more. We are approaching the time when there will be some useful outcomes information, though to a large degree that is contingent on the proportion of SSDI-EP participants who can complete their TWP a half year or more prior to the pilot’s

⁹⁵ There would still be “best practices” information in the sense of having lessons about what not to do.

⁹⁶ For most analyses not using encounter or survey data, the data period extends back to the eighth calendar quarter prior to the quarter of entry. Analyses using encounter or survey data (except for those data elements treated as “static”) will necessarily be limited to the enrollment period forward.

end. Nonetheless, if the pilots are ended too soon, most of the potential value of what, at least in Wisconsin, appears to have been a more than adequate implementation of both random assignment and of basic pilot infrastructure and processes will have been squandered.