INITIAL IMPACTS OF THE TICKET TO WORK PROGRAM: ESTIMATES BASED ON EXOGENOUS VARIATION IN TICKET MAIL MONTHS

David Stapleton
Senior Fellow and Director, Center for Studying Disability Policy
Mathematica Policy Research
1100 First St NE, 12th Floor
Washington, DC 20002-4221

Arif Mamun[1]
Senior Research Economist and Associate Director of Research
Mathematica Policy Research
1100 First St NE, 12th Floor
Washington, DC 20002-4221
Email: amamun@mathematica-mpr.com

Jeremy Page
Senior Programmer Analyst
Mathematica Policy Research
1100 First St NE, 12th Floor
Washington, DC 20002-4221

ABSTRACT

This paper presents results from an impact analysis of the Ticket to Work (TTW) program, as implemented by the Social Security Administration (SSA) from 2002 through 2007. For new, young Social Security Disability beneficiaries, we use exogenous variation in the month of Ticket mailing to rigorously estimate impacts of TTW on beneficiary outcomes over a 48-month period following the start of Ticket mailings in the beneficiary's state. We find substantial impacts on service enrollment, but no consistent evidence of impacts on the number of months in which beneficiaries did not receive benefits because of work, or on other outcomes.

JEL classification: H55, I38

Keywords: Ticket to Work, Social Security disability benefits, impact analysis, voucher, employment services.

---

[1] Corresponding author.

I. INTRODUCTION

Social Security Disability (SSD) benefits are available to workers who experience long-lasting medical impairments that prevent work at a substantial level (disabled workers), as well as to Disabled Adult Children (DAC) and Disabled Widow(er)s of other Social Security beneficiaries.[2] In 2011, more than 9.8 million people received SSD benefits.[3]

Many SSD beneficiaries are able and willing to work at some level; most of those who work earn little enough that they do not lose their benefits. Recognizing this, the Ticket to Work and Work Improvement Incentives Act of 1999 (Ticket Act) put into place a number of new policies and programs designed to encourage beneficiaries' return-to-work efforts. The leading initiative is the Ticket to Work (TTW) program. Initially, the Social Security Administration (SSA) mailed each eligible disability program beneficiary a "Ticket" that he or she could assign to either a state vocational rehabilitation agency (SVRA) or to a prequalified local rehabilitation service provider, called an employment network (EN), in exchange for employment placement, job training, and other services.[4] SSA promised to pay the provider on the basis of earnings and benefit outcomes for the beneficiary. TTW was designed to expand the service options available to beneficiaries and create greater incentives for providers to help beneficiaries earn enough to forgo benefits.

TTW was rolled out in three phases. A first set of states completed the TTW rollout in 2002 (Phase 1), a second set in 2003 (Phase 2), and a final set in 2004 (Phase 3). In July 2008, SSA significantly changed the regulations governing TTW to attract more providers and reflect a more flexible return-to-work concept; hereafter, we call the pre-2008 program the "original" program.

Previous attempts to estimate impacts of TTW provide inconclusive evidence. The earlier analyses were, in essence, based on annual trends in differences for mean service enrollment, earnings and benefit outcomes across the three phases (Thornton et al. 2007; Stapleton et al. 2008). Results were inconclusive, because methodological issues made it impossible to discriminate between potentially very small, yet important impacts of TTW and pre-existing trends in the differences across phases for earnings and benefit outcomes. A number of alternative strategies were attempted in recent years to estimate Ticket impacts, but were also found to be inadequate.[5]

In this article, we present results from a rigorous new analysis of the impact of the introduction of the original TTW program, incorporating multiple innovations relative to earlier efforts. The analysis exploits a feature of the initial TTW rollout in each phase: just before the start of the rollout,

---

[2] DAC receive benefits on the basis of a parent's entitlement as a "primary beneficiary"—a parent who is a disabled worker, retirement beneficiary, or deceased worker. The DAC must be deemed unable to work as of the age of 22 under the same medical criteria applied to disabled workers, he or she is not entitled to benefits until the parent is entitled. Each disabled widow(er) beneficiary (DWB) receives benefits on the basis of the entitlement of a deceased spouse; the DWB must be at least 50 years old as well as meet the same medical criteria as disabled workers. DAC and DWB benefits are paid out of the Social Security Disability Insurance (SSDI) Trust Fund if the primary beneficiary is a disabled worker, or out of the Old Age and Survivors Insurance (OASI) Trust Fund if the primary beneficiary is a retiree or deceased. See SSA (2012) for further details.

[3] Because the analysis presented in this report includes SSDI disabled worker beneficiaries as well as DAC and DWB, and benefits for most of the latter are not paid from the SSDI Trust Fund, we use SSD to encompass all three groups.

[4] SSA no longer mails tickets to beneficiaries. Instead, the beneficiary can approach a provider and the provider may contact SSA to verify eligibility.

[5] A brief discussion of these alternative strategies is available in Stapleton et al. 2013

SSA selected the month in which it intended to mail each eligible beneficiary's Ticket in an essentially random fashion. We use variation in the intended mail month to rigorously estimate how the timing of Ticket mailing affects beneficiary outcomes over the following 48 months, then use the estimates to draw inferences about impacts of TTW (versus no TTW) over the same period. The new analysis also takes advantage of improvements in the measurement of work related outcomes from administrative data, namely a monthly indicator of benefit suspension or termination for work (STW) and a count of months in nonpayment status after STW (NSTW months) and before returning to current-pay status, attainment of the full retirement age (FRA), or death (NSTW months).[6] This article focuses on impacts for NSTW months as well as for two intermediate outcome variables: enrollment for employment services with an SVRA or EN, and an event that must precede STW: completion of the trial work period (TWP).[7, 8]

The analysis focuses on the subgroup of beneficiaries for whom we expect impacts to be easiest to detect: young (ages 18 to 39 at award), new SSD-only beneficiaries—that is, those awarded SSD recently who are not also receiving Supplemental Security Income (SSI). Recent research suggests that TTW had the most promise of having impacts on work outcomes for this group (Liu and Stapleton 2011, Mamun et al. 2011, Stapleton et al. 2008). We assumed that if we found no substantial evidence of impacts on key outcomes for this group, we could be reasonably confident that there would be no evidence of impacts if we applied the same approach to other groups. If instead we found evidence of positive impacts for this group, we could extend the methodology to see if the same approach produced positive evidence for other groups. The focus on new beneficiaries also allowed us to exclude those who received awards before July 1999, when there was a significant change in how earnings affect benefits.

The findings reported here directly address the following primary research questions related to the three outcomes. Each question concerns the impact of duration from the month before the rollout start in the beneficiary's state to the month in which SSA mailed a Ticket to the beneficiary (the beneficiary's "mail month") on outcomes over the 48 months after the rollout start.

- Was enrollment for employment services and completion of TWP less likely to occur as of 12, 24, 36, and 48 months after rollout start the longer the duration from rollout start to mail month?

- Was the number of NSTW months as of 12, 24, 36, and 48 months after rollout start smaller the longer the duration from rollout start to mail month?

We then use the findings to indirectly answer the question of most interest to policymakers:

---

[6] "Current-pay" status means that the individual is eligible for a cash payment for the current month.

[7] During the TWP, SSD beneficiaries are permitted to work and earn at any level without loss of benefits, provided that they continue to meet the medical eligibility requirements. The TWP consists of 9 months, which need not be consecutive—any 9 months in a 60-month rolling window are counted. After completing the TWP, beneficiaries enter an extended period of eligibility (EPE). Except for a 3-month grace period, individuals who engage in substantial gainful activity (SGA) in any of the next 36 months have their benefits suspended for that month. The beneficiary is entitled to full benefits during any month of this period when he or she is not engaged in SGA, provided that benefits have not been terminated for medical recovery or some other reason. After 36 months, SSD benefits are terminated in the first month of SGA after use of any remaining grace period months.

[8] We also analyzed two other outcomes—starting the TWP, and first month of benefit suspense or termination for work (STW)—but their results are not essential to understand the key impacts of the TTW program. The results for these outcomes are available in Stapleton et al. 2013.

- What was the impact of mailing Tickets as of 48 months later versus not mailing Tickets at all?

We also assess whether TTW was self-financing by 2007, before the new regulations took effect.

II. DATA AND METHODS

A. Ticket Research File

We used data from the 2007 Ticket Research File (TRF07). The TRF is a set of analytic administrative data files constructed for the TTW evaluation. The TRF07 contains current and historical information on more than 22 million SSD beneficiaries or SSI recipients who received a benefit in at least one month from January 1996 through December 2007 (Hildebrand et al. 2009).[9] For the purpose of this study, we constructed an analytic file for those awarded benefits from 1999 through 2003, based on the month that SSA first paid a benefit to the awardee.[10]

B. Analytic Samples

1. Sample Selection

The sample includes beneficiaries first awarded SSD benefits from July 1999 through October 2003. For the analysis, we followed each beneficiary for 48 months starting with the first month of the rollout in the beneficiary's state. As the Phase 3 rollout started in November 2003, the last month in the sample is October 2007. We limit the analysis to this period because of factors external to the introduction of TTW. We started with July 1999 SSD awardees because this is the month in which the non-blind substantial gainful activity (SGA) level was increased from $500 to $700. We end the follow-up period in 2007 because of the severe recession that started in the last quarter of 2007 and because SSA made substantial changes to TTW regulations in 2008 that may have affected beneficiary outcomes in 2008 and later.

The results reported here use analysis samples that consist of young (ages 18 to 39 at award) SSD-only awardees who were first paid SSD benefits no earlier than July 1999 and were selected for the initial rollout of the TTW program on two of the three initial Ticket selection dates: October 26, 2002 (Phase 2), or October 18, 2003 (Phase 3).[11] It was SSA's intent to mail Tickets to every beneficiary in these samples during a subsequent rollout month (hereafter, the "intended mail month" [IMM]), to be determined by the terminal digit of the beneficiary's SSN. Table 1 provides the sample sizes for each phase by IMM. Both phases follow identical 11-month schedules, except

---

[9] Extracts from several Social Security administrative files were merged to create the TRF, including the Master Beneficiary Record, Supplemental Security Record, Numerical Identification System (Numident) file, the 831 and 832/33 Disability files, the Disability Control File, monthly snapshot files, and files from the payment history update system.

[10] The first payment month (that is, the award month) is that in which the first payment was actually made, which is usually after the first month for which the beneficiary is entitled to a benefit (that is, the entitlement month). The latter is often used in SSA's statistics to classify beneficiaries by entry year (for example, SSA 2009). We use the award month instead because our focus is on the activities of beneficiaries once they become informed of their award and are entitled to use the DI work incentives.

[11] SSA determined all beneficiaries who were eligible to receive a Ticket and who resided within the phase's states as of the phase's selection month. Almost all SSD beneficiaries and SSI recipients over age 18 were eligible; the main exceptions were (1) new beneficiaries with a status of medical improvement expected (MIE) who had not yet had their first medical continuing disability review (medical CDR) and (2) child SSI recipients who had reached age 18 and were waiting for redetermination as adults.

separated by 12 months, and mailings were uniformly distributed across 10 of the 11 rollout months, with the second month being the exception. As shown in Table 1, SSA mailed the vast majority of these Tickets on the IMM. For each phase, we treat the samples defined by IMM (hereafter, the "IMM samples") as randomly assigned samples of those included on the phase's selection date. As described later, the methodology also addresses the fact that some Tickets were not mailed on the IMM.

We also produced results for young, new SSD-only awardees selected for TTW rollout on January 12, 2002 (Phase 1), but two features of the Phase 1 rollout substantially limit their value. The first such feature is that the Phase 1 sample had to be split into two relatively small samples because an operational issue led to different rollout schedules for New York (NY) and the rest of the Phase 1 states: the respective sample sizes in these two sample groups were 12,023 and 43,080, compared to 77,161 in Phase 2 and 114,657 in Phase 3. A second reason is that the rollout periods in Phases 2 and 3 (11 months in each) were substantially longer than in either part of Phase 1 (nine months in NY and five months in the rest of Phase 1). The larger samples and longer rollout periods in the later phases contribute substantially to the ability of the methodology to detect small impacts. The Phase 1 findings do not contradict or illuminate the findings reported here, so have been omitted for brevity.[12]

Table1. Sample Sizes by Intended Mail Months in Phases 2 and 3

Rollout Month	Phase 2				Phase 3			
	Calendar Month	Intended Mail Month	N	% Actually Mailed on IMM	Calendar Month	Intended Mail Month	N	% Actually Mailed on IMM
1	Nov-02	10	7,573	97.46	Nov-03	21	11,531	99.85
2								
3	Jan-03	12	7,733	95.93	Jan-04	23	11,328	94.13
4	Feb-03	13	7,679	95.86	Feb-04	24	11,539	93.89
5	Mar-03	14	7,745	95.36	Mar-04	25	11,569	93.22
6	Apr-03	15	7,743	94.56	Apr-04	26	11,523	93.04
7	May-03	16	7,778	94.33	May-04	27	11,533	92.39
8	Jun-03	17	7,598	94.04	Jun-04	28	11,434	92.21
9	Jul-03	18	7,826	93.05	Jul-04	29	11,519	91.67
10	Aug-03	19	7,681	92.98	Aug-04	30	11,253	91.42
11	Sep-03	20	7,805	92.68	Sep-04	31	11,428	90.79
Total			**77,161**				**114,657**	

2.   Intended and Actual Mail Months

Although SSA actually mailed Tickets on the IMM for most beneficiaries, for a small fraction the actual mail month (MM) did not correspond to the IMM. The TRF records include the actual mail date, making it possible to determine the MM. Across the two phase-samples, in 93 to 95 percent of the cases the MM is the IMM. Although the fraction of Tickets mailed on the IMM was very high in each month of the rollout, it did decline in successive months. One reason for the decline is a provision of the regulations called "Ticket on demand;" beneficiaries in each phase could request a Ticket in advance their mail date, and beneficiaries assigned to later IMM had more opportunity to make such requests a Ticket than those with early IMM. In addition, as the rollout progressed, SSA identified some beneficiaries who had died or were no longer in current-pay status, and consequently did not mail these beneficiaries their Tickets (see Table 2). Because mailing a

---

[12] The Phase 1 findings are reported in Stapleton et al. (2013).

Ticket on demand, mortality, and loss of current pay status for some other reason are likely predictive of the outcome variables, we made adjustments to the methodology to avoid confounding the correlation of these factors with the outcomes with the impacts of mailing the Ticket, as described in Section II.C.

One other issues is that, in each phase, for a small share of beneficiaries (about 4.7 percent in each phase) the state of residence for the beneficiary obtained from the TRF was not among the states included in the phase's rollout (Table 1). We do not know detailed reasons, but there are several possibilities: SSA included people in neighboring states that were served by a field office located in a state within the phase group; the state shown in the data reflects an address that is not the beneficiary's own; or the beneficiary at some point moved to a non-phase state, but SSA did not know of this move on the selection date. Because we are aiming to retain as much of the original IMM sample as possible for the analysis, and because we found little variation in the percentage of the sample in each of these states across the mail months within phase, we did not exclude these cases.[13]

Table 2. Correspondence of Actual Mail Months (MM) and Intended Mail Months (IMM)

	Phase 2		Phase 3	
Actual Mail Month Is—				
Before Rollout	0	0.00%	0	0.00%
Before IMM, During Rollout (Ticket on Demand)	3,355	4.35%	6,307	5.50%
Corresponds to the IMM	73,008	94.62%	106,938	93.27%
After IMM, During Rollout	224	0.29%	39	0.03%
After Rollout	51	0.07%	173	0.19%
Missing Mail Date				
Status as of IMM	523	0.68%	1,200	1.06%
Deceased	433	0.56%	973	0.85%
Suspense/termination for work	2	<0.01%	3	<0.01%
Other Suspense/termination	85	0.11%	204	0.18%
Current pay	3	<0.01%	20	0.02%
Total	77,161	100.00%	114,657	100.0%
Not resident of phase state	3,615	4.69%	5,418	4.72%

3. Beneficiary Characteristics and Tests of Statistical Equivalence

In Table 3, we present characteristics of the beneficiaries in the two phase samples. Almost all of the characteristics are defined as of the beneficiary's Ticket selection date. The exceptions are the primary disabling conditions, measured at SSD award date; the primary insurance amount, which is the earliest recorded value; and the indexed monthly earnings, also the earliest recorded value. The beneficiary populations vary somewhat across phases, as reflected in modest differences in means. Compared to the Phase 2 sample, the Phase 3 sample has relatively fewer African Americans (16 percent versus 21 percent), more Hispanics (12 percent versus 3 percent), higher indexed monthly earnings ($1,125 versus $1,090) and Primary Insurance Amount

---

[13] With a small number of cases in the "out-of-phase" states in each phase, the random variation at the state level may explain a substantial fraction of the variation in some characteristics across IMM samples. We address this issue in footnote 14 in the next section.

(PIA)[14] ($643 versus $632), and more beneficiaries with major affective disorders (18 percent versus 16 percent). Many differences reflect the fact that the Phase 3 rollout started 12 months after the Phase 2 rollout, so beneficiaries in Phase 3 had aged a year between the Phase 2 Ticket selection date and their own selection date, and more new awardees were added to the Phase 3 sample during the same period. Compared to those in the Phase 2 sample, as of the selection date, they were older (mean of 34.3 versus 33.7) and had been on the rolls longer (mean of 26 months versus 18 months). In addition, those in the Phase 2 sample were more likely to have: previously enrolled for services, started the TWP, completed the TWP, experienced a month of suspension or termination for work, and become eligible for Medicare. In addition, some differences are expected between Phases 2 and 3 because of differences between the economic, policy and cultural environments for states in each phase.

Table 3 also presents tests of the statistical equivalence of the IMM samples within each phase. The statistical equivalence tests for each phase's sample were conducted by running linear regressions of each characteristic on a set of IMM indicators for the months within that phase, without an intercept. For each regression, we conducted a joint test (an F-test) for the hypothesis that all of the population coefficients are equal. In conducting the test, we treated each state in the phase as a cluster and allowed for heteroscedasticity in the regression disturbance.[15] The F-tests show that we would reject the null hypothesis of "no difference" across IMM samples within phase for a large number of characteristics. Substantively, however, even when a baseline characteristic is found to be statistically different across IMM samples within a phase, variation in the means across the IMMs is not substantial. For example, in Phase 2, we found significant differences in mean beneficiary age at Ticket selection date across IMMs, but the difference between the maximum and minimum mean is 0.19 years around a mean for the phase of 33.70. There are also significant differences for some baseline values of the outcome variables. Most notably in Phase 2, 19.1 percent of beneficiaries had previously been found eligible for SVRA services, and the range of this percentage across the IMM was 1.8 percent. The distribution of the sample across states in each phase is not shown in Table 3 for brevity, but there were no statistically significant differences in state of residence by IMM for either phase.[16]

---

[14] The PIA is an amount based on past earnings that determines the benefit amount. For disabled worker beneficiaries, the PIA is the worker's benefit amounts, and qualified dependents receive an additional fraction of this amount. For DAC and DWB, the PIA is based on the parent or spouse's past earnings and the benefits are a fraction of the amount.

[15] As noted earlier, because a small sample of beneficiaries are residents of "out-of-phase" states, each of which is treated as a cluster, it is conceivable that they might be substantially influencing the results of the joint tests of statistical equivalence. To explore this, we conducted the joint tests without correcting for state-level clustering (but adjusting for heteroscedasticity) and found that we would reject the null hypothesis for far fewer characteristics. We suspect that with a small number of cases in the nontargeted states in each phase, the random variation in the cluster component of the model's error term explains so much variation in some characteristics that tiny differences across IMM groups are found to be significant. But this is just conjecture, and we are not aware of any technical problem with including a set of clusters with very small samples along with clusters that are much larger.

[16] These statistics are available in Stapleton et al. (2013).

Table 3. Beneficiary Characteristics: Means by Phase and Intended Mail Months

	Phase 2			Phase 3		
	Mean	Range[a]	Test[b]	Mean	Range[a]	Test[b]
Gender						
Male	0.521	0.020		0.523	0.015	***
Gender missing	0.000	0.000		0.000	0.000	
Age at Selection Date	33.697	0.192	***	34.31	0.21	**
Age at Disability Onset						
Age	27.963	0.343	**	27.96	0.29	***
Age missing	0.000	0.000		0.000	0.000	
Race/Ethnicity						
Asian	0.009	0.005	***	0.026	0.004	
African American (non-Hispanic)	0.209	0.012		0.156	0.011	***
Hispanic	0.032	0.006		0.119	0.011	***
American Indian/Hawaiian	0.006	0.002		0.005	0.002	**
White [Ref]	0.728	0.016	**	0.667	0.013	**
Other	0.004	0.002		0.010	0.003	***
Missing	0.012	0.004		0.017	0.004	
Education at Selection Date						
Less than high school [Ref]	0.187	0.016	***	0.152	0.014	***
High school graduate	0.391	0.015		0.374	0.013	***
More than high school	0.185	0.018	***	0.204	0.012	***
Missing	0.238	0.013		0.270	0.014	***
Expectations about Medical Improvement at Selection Date						
Expected	0.039	0.009		0.028	0.004	
Possible	0.548	0.015	***	0.601	0.014	***
Not expected	0.180	0.016	***	0.196	0.011	*
Missing	0.233	0.013		0.175	0.014	***
Medicare Eligibility at Selection Date [Ref = not eligible]						
Eligible	0.646	0.011		0.779	0.015	***
Eligibility missing	0.031	0.006		0.018	0.004	***
Number of Dependent Beneficiaries at Selection Date						
No dependent [Ref]	0.455	0.017	***	0.478	0.017	**
1	0.171	0.011	**	0.163	0.010	***
2 or more	0.265	0.012		0.249	0.013	***
Missing	0.108	0.012	**	0.110	0.007	***
VR Services Before Selection Date						
Determined eligible for VR services	0.191	0.018	***	0.207	0.010	
VR service eligibility missing	0.761	0.024	***	0.760	0.011	
Outcome Achieved Before Ticket Selection Date						
TWP start before Ticket selection	0.063	0.010	***	0.084	0.010	**
TWP completion before Ticket selection	0.038	0.007	***	0.048	0.007	**
STW before Ticket selection	0.018	0.004		0.025	0.005	**
Months Between SSD Award and Selection Date	18.496	0.224		26.352	0.540	***
Primary Insurance Amount (PIA, $)						
Mean PIA	626.9	18.8	*	643.3	12.3	***
PIA missing	0.125	0.011	**	0.123	0.009	***
Indexed Monthly Earnings (IME, $)						
Mean IME	1089.6	45.8		1124.9	28.4	***
IME missing	0.192	0.009		0.185	0.008	**
Primary Disabling Conditions at SSD Award						
Major affective disorders [Ref]	0.155	0.018	***	0.184	0.015	
Other psychiatric disorders and mental retardation	0.241	0.019	**	0.241	0.017	***
Back disorders and musculoskeletal system	0.120	0.009	**	0.111	0.011	***
Other physical disabilities	0.482	0.027	***	0.463	0.017	***
Missing	0.001	0.001		0.001	0.001	
SSD Award Year						

	Phase 2			Phase 3		
	Mean	Range[a]	Test[b]	Mean	Range[a]	Test[b]
1999	0.097	0.005		0.077	0.009	***
2000	0.328	0.012		0.258	0.009	
2001	0.300	0.012	*	0.293	0.014	
2002	0.250	0.013		0.232	0.016	***
2003	0.015	0.004	***	0.128	0.015	***
2004	0.009	0.003	*	0.012	0.003	***
State Unemployment Rate						
Mean in 6 months around IMM	0.079	0.843	***	-0.255	0.264	***
Change in 6 months around IMM	5.710	0.385	***	6.007	0.511	***
Anomalous Sequence of Events						
Ticket selection before SSD award	0.036	0.008	**	-	-	
TWP completion before start	0.000	0.001		0.000	0.000	***
STW before TWP start	0.000	0.000		0.000	0.000	***
STW before TWP completion	0.001	0.002	***	0.001	0.001	***

Source: Authors' calculations using data from Ticket Research File 2007

Note: "Ref" indicates the reference category for the discrete variable in the multivariate regression models.

[a]"Range" is the difference between the minimum and maximum mean across IMM in each sample.
[b]"Test" shows the results from the test of the null hypothesis that the means are the same across IMM in the sample; *, **, and *** indicate significantly different at the 0.10, 0.05, and 0.01 levels, respectively.

We consider the IMM samples in each phase to be substantively very similar even though the joint tests indicate that they are not statistically equivalent in many regards. The differences likely reflect the fact that the method SSA used to assign IMM was not purely random, particularly with respect to age and factors that are associated with age. For this reason, it is important to control for these characteristics in the analysis—most critically, for the occurrence of the outcome events prior to Ticket selection date.

4. Outcome Measures

The outcome measures are based on the 48 months starting with the first rollout month for the phase (month zero is the pre-rollout month). This period ends September 2006 for Phase 2, and September 2007 for Phase 3. For each individual in the sample we report results for:

- Two binary "event" variables. We determined when in the 48 months following start of rollout each of the following events occurred, if at all: (1) enrolled for employment services (assigned their Ticket to an EN or were determined eligible for services by an SVRA); and (2) completed their last TWP month. In the analysis of whether an event has occurred as of a specified rollout month (month 12, 24, 36, or 48), we define a binary variable for each event that is equal to one if the event occurred after the rollout start and before that month, and zero otherwise.

- NSTW months, a count of the number of months in nonpayment status following STW that occurred during the 48-month period. NSTW months include all months after benefits are suspended or terminated for work until the first of the following events occurs: (1) return to current-pay status, (2) suspension or termination for some other reason, or (3) the end of the 48-month period. Beneficiaries are not necessarily engaged in SGA during all NSTW months; we know only that they are not receiving benefits.

Means for the outcome variables in the IMM samples as of month 48 are presented in Table 4. The overall mean for Phase 2 is higher than Phase 3 for service enrollment, but the opposite is true for the other two outcomes. These differences reflect state differences in beneficiary characteristics, the labor market, and the support system, as well as the 12-month difference in time period.

The variation in service enrollment rates across IMM within each phase is consistent with a negative impact of duration to the mail month on enrollment as of month 48. For instance, for Phase 2, service enrollment declines from 7.25 percent for November 2002 to 7.07 percent for September 2003; the corresponding figures for Phase 3 are 5.90 and 5.57 percent. These differences are not statistically significant, however. The variation in means across IMM within each phase for other outcomes is not clearly consistent with negative impacts for those outcomes.

Table 4. IMM Sample Percentages Experiencing Four Events by End of Month 48 After Rollout Start

| | Service Enrollment | | TWP Completed | | NSTW Months | |
	Mean (%)	SE	Mean (%)	SE	Mean (Months)	SE
Phase 2	**6.78**	**0.09**	**8.08**	**0.10**	**1.460**	**0.024**
Nov-02	7.25	0.30	7.88	0.31	1.412	0.074
Dec-02						
Jan-03	6.91	0.29	8.33	0.31	1.470	0.075
Feb-03	6.72	0.29	7.96	0.31	1.597	0.079
Mar-03	6.75	0.29	7.64	0.30	1.394	0.072
Apr-03	6.70	0.28	8.69	0.32	1.483	0.075
May-03	6.78	0.28	8.15	0.31	1.472	0.074
Jun-03	6.69	0.29	8.08	0.31	1.542	0.079
Jul-03	6.45	0.28	8.01	0.31	1.448	0.074
Aug-03	6.46	0.28	8.06	0.31	1.357	0.072
Sep-03	7.07	0.29	7.99	0.31	1.423	0.075
Phase 3	**5.67**	**0.07**	**8.24**	**0.08**	**1.686**	**0.021**
Nov-03	5.90	0.22	8.33	0.26	1.645	0.065
Dec-03						
Jan-04	5.73	0.22	8.76	0.27	1.750	0.068
Feb-04	5.84	0.22	8.22	0.26	1.617	0.064
Mar-04	5.67	0.22	8.17	0.25	1.641	0.065
Apr-04	5.39	0.21	8.17	0.26	1.668	0.066
May-04	5.54	0.21	8.45	0.26	1.756	0.067
Jun-04	6.09	0.22	8.26	0.26	1.733	0.068
Jul-04	5.55	0.21	7.78	0.25	1.675	0.067
Aug-04	5.46	0.21	8.17	0.26	1.672	0.066
Sep-04	5.57	0.21	8.12	0.26	1.703	0.067

Note:"Mean" shows the mean percent experiencing the event or the mean NSTW months over the first 48 rollout months; SE is the standard error of the estimated mean.

C. Estimation Approach

1. Identification Strategy

For each phase of the TTW rollout, SSA selected the IMM for all beneficiaries who were eligible on the phase's Ticket selection date—approximately one month before rollout began. SSA used the terminal digit of the beneficiary's SSN to determine the rollout month in which SSA would mail the beneficiary a Ticket. Because the last four digits (the serial numbers) of SSNs are considered

to be random after conditioning on age,[17] this strategy essentially led to random assignment of the eligible beneficiaries to IMMs, after controlling for age. Consequently after controlling for age we assume that variation in the duration from the rollout start to the IMM is exogenous to each of the outcome variables (that is, independent of other unobserved factors that might affect outcomes). This provides the foundation for estimating the impacts of the duration from Ticket rollout start to the IMM on later beneficiary outcomes.

We used the exogenous assignment of IMMs to identify the impacts of delaying actual Ticket mail month (MM) on beneficiary outcomes while accounting for a limited number of non-random deviations of the MM from the IMM. We hypothesize that the longer the duration from rollout start to the MM the lower the expected value for each outcome variable—enrollment in vocational services, completion of the TWP, and the number of NSTW months. The estimated impact of delaying the MM is expected to be different from the direct, intent-to-treat (ITT) impacts of delaying the IMM, and is likely to be of greater interest to policymakers.[18] The difference might be substantial because we are relying on random variation in duration from rollout start to the IMM to identify impacts, and the later a beneficiary's IMM, the greater the likelihood of an adjustment to the actual MM. To produce these estimates, we use the IMM variables as instrumental variables (IV) for the MM variables.[19]

2. Instrumental Variables Estimation

To estimate the impact of actually mailing the Ticket on each outcome, we applied an IV approach to the following model:

Equation (1) $\qquad MM_i = \theta\ IMM_i + \tau X_i + v_i$

Equation (2) $\qquad E_{it} = \beta_t'MM_i + \gamma_t'X_i + u_{it}$

where MM_i is a vector of dummies for the actual mail month; IMM_i is a vector of dummies for the IMM; E_{it} is a dummy variable for whether the event has occurred for beneficiary i as of month t following the rollout start, or is the number of NSTW months experienced by beneficiary i as of month t following the rollout start; X_i is a vector of baseline control variables, v_i and u_{it} are zero mean, residual terms; and t=12, 24, 36, and 48 months following the rollout start. Because MM_i is a vector, Equation (1) represents a set of equations, one for each mail month, and θ and τ are both matrices. There is no intercept, because X_i contains an exhaustive set of state indicators. To avoid exact collinearity with the exhaustive set of MM indicators, we restricted the coefficients of the MM indicators to sum to zero. This normalization means that the coefficient for each MM indicator is the predicted impact of Ticket mailing in that MM relative to mailing on the average MM during the rollout period (approximately 6.4 for both samples).

---

[17] The serial numbers in SSNs are considered random only after conditioning on age because they are historically assigned in sequence.

[18] Estimates of direct impacts of delaying the IMM on beneficiary outcomes are available in Stapleton et al. 2013.

[19] As stated earlier, in order to address the fact that some Tickets were never mailed because benefits were suspended or terminated prior to the IMM, we coded the MM for those observations as if the Tickets were actually mailed on the IMM. We had previously verified that, with almost no exceptions, termination or suspension of benefits had occurred for reasons other than work—most commonly mortality. Our reasoning for this modification is that mailing the Ticket to these beneficiaries during any month of the rollout would almost certainly have had no impact on their employment outcomes, in which case essentially all outcomes would have been the same as those observed had SSA mailed all of these Tickets in their IMM.

Although we call MM_i the actual mail month indicator vector, it does not indicate the actual mail month for every observation, because a small share of Tickets were never mailed, and an even smaller share were mailed shortly after the rollout window (that is, after the last month represented in MM_i). In coding MM_i, we had to choose one of the rollout period mail months for each of these observations in order to keep them in the sample. For each case, we chose a virtual month that is assumed to be consistent with the individual's actual behavior. For the bulk of such cases—those never mailed a Ticket because of benefit termination prior to their IMM—we used the actual IMM on the assumption that had SSA proceeded to mail their Ticket on their IMMs, their behavior would have been the same as their actual behavior. That seems very likely, because they would have received their Ticket after they could no longer use them, and a large majority were deceased. For the very small share of cases in which the Ticket was mailed a few months after the rollout ended (0.07 percent for Phase 2 and 0.19 percent for Phase 3, as shown in Table 2), we chose the last rollout month; that is, we coded these late mailing cases as if their Tickets were mailed a few months earlier than they were actually mailed.

Two assumptions must be satisfied for IMM_i to be a set of valid instruments. First, conditional on X_i, IMM_i must be uncorrelated with the disturbance terms in equations 2 and 3. Second, again conditional on X_i, IMM_i must be correlated with the MM_i. (Angrist et al. 1996). Both assumptions are satisfied in our case. The first assumption is satisfied because SSA assigned IMM in a fashion that was exogenous with respect to the individual's characteristics after conditioning on age; thus, by design, IMM_i is independent of any unobserved individual characteristics. Further, the IMM selected could have no effect on the outcomes of interest except through its effect on the actual mailing of Tickets. The second assumption is satisfied because the vast majority of Tickets were mailed on the IMM (see Table 1). Hence, taken together, the IMMs constitute valid instruments for estimating the impact of the actual MM on beneficiary outcomes. Further, they are a very strong set of instruments, in that the correlation between the IMM and MM variable for each mail month is quite high, reflecting the fact that the IMM and MM are identical in the vast majority of cases. This is also reflected in the very large F-statistic from the first stage of IV estimation for each endogenous MM, which ranged between 537 and 1,507 for Phase 2, and between 484 and 971 for Phase 3 for the MMs in each phase.

We assume that the residual terms in Equations 1 and 2 have a state component that is independent across states and an individual component that is independent across individuals. Hence, we estimated each equation using adjustments for clustering at the state level. These estimates also adjust for heteroskedasticity, which is expected because of the nature of the dependent variables. To test the null hypothesis that duration to MM has no impact on an outcome, we tested the hypothesis that all of the mail-month coefficients are zero. For each outcome, we also tested the hypothesis that the marginal impact of delaying the mailing of the Ticket an additional month was the same throughout the rollout period (that is, that there is a linear relationship between duration to mail month and the expected outcome). We used a chi-square test in each case.

III. FINDINGS

We present the IV estimates for the impacts of duration from rollout start to MM on service enrollment, TWP completion, and NSTW months graphically in Figure 1. Each panel in the figure shows the estimated impacts for each MM for the samples in Phase 2 (left panel) and Phase 3 (right panel). The mean value of the 10 point estimates in each phase is zero, by design; each point estimate measures the expected outcome for the sample mailed a Ticket in the corresponding month relative to the overall mean outcome for all those in the phase sample after accounting for pre-rollout characteristics and the fact that not all Tickets were mailed on schedule. In addition, we have plotted 95 percent confidence intervals around each point estimate (the short vertical line through each X) as well as the line obtained by constraining the IV estimates to fall on a straight line. The slope of the trend line indicates how much each one-month delay in mailing the Ticket affected the outcome. We also present two test statistics at the bottom of each panel: the first test statistic shows the result for the joint test of the null hypothesis that all the estimated IV coefficients are zero, and the second is for the linear restrictions on the coefficients.

A. Clear Evidence of Impacts on Service Enrollment

As seen from the top panel of Figure 1, in both phases we found strong evidence of negative effects of duration to MM on service enrollment at 12 months following rollout start in Phase 2 and Phase 3 (that is, the longer the duration, the lower the proportion enrolled). For both phases, the confidence intervals for the point estimates are narrow, the estimates steadily decline with duration, the data strongly reject the hypothesis that they are all zero, and fail to reject the hypothesis that the true values lie on a straight line. For Phase 2, the slope of the trend line is -0.084, indicating that each one-month delay in mailing the Ticket reduced the percentage enrolled in services as of month 12 by an estimated 0.084 percentage points, with a 95 percent confidence interval of ±0.033. Extrapolating to 12 months, the estimated impact on service enrollment at month 12 of mailing the ticket in month one versus not mailing it at all−−which is equivalent to the projected impact of a delay from month one to month 13−−is 1.0 percentage points (-0.084 x 12). The estimates suggest that the impact on service enrollment for Phase 3 was somewhat smaller than for Phase 2. The point estimate of the slope of the trend line is -0.066 (±0.025), and the projected impact on service enrollment at month 12 versus not mailing it at all is 0.8 percentage points. The difference between the Phase 2 and Phase 3 slopes is not statistically significant, however.

We found more limited evidence of impacts on service enrollment at 24, 36, and 48 months following rollout start. Estimates at each observation point showed a negative relationship between duration and service enrollment, but progressively weaker with the duration from rollout start to the observation month (see Stapleton et al. 2013 for details). This is consistent with expectations; as each month goes by, those mailed Tickets late in the rollout period have more time to catch up to those mailed Tickets earlier in terms of service enrollment. Thus, it appears that, on average, an early MM accelerated the beneficiary's entry into service enrollment relative to a later MM, but service enrollment for those with later MM had, by the end of the observation period, largely caught up to enrollment for those with earlier MM.

[INSERT FIGURE 1 HERE]

B.   Unclear Evidence of Impacts on TWP Completion

The middle panel in Figure 1 plots the instrumental variable estimates for impacts on the likelihood of TWP completion at 48 months after the start of rollout, along with their 95 percent confidence intervals and estimated trend lines. Estimates for TWP completion at 12, 24 and 36 months appear in Stapleton et al. 2013 and are no stronger in terms of evidence of impacts than those at month 48. For both phases, the monthly estimates are jointly significant at the 5 percent level, but the patterns of monthly coefficients in each phase do not support the conclusion that their joint significance reflects an impact of duration to MM on TWP completion. In each phase, the null hypothesis of zero coefficients is rejected primarily because estimates for two months have relatively large magnitudes (March and April for Phase 2 and January and July for Phase 3). For Phase 2 the pattern of the estimates is clearly inconsistent with a negative effect, and the slope of the trend line is very small and statistically insignificant. For Phase 3, the slope of the trend line is negative and significant at the 5 percent level, but the hypothesis that the impacts are linear is rejected at the 10 percent level of significance. This leaves open the distinct possibility that the estimates for Phase 3 are simply due to chance difference for the January and July samples rather than to a negative impact of duration to MM on TWP completion.

C.   Unclear Evidence of Impacts on NSTW Months

On the bottom panel in Figure 1, we plot the monthly IV estimates for impacts of duration to MM on the number of NSTW months completed as of 48 months after rollout start. The evidence from Phase 2 for NSTW months is marginally indicative of a substantive impact when viewed in isolation, but in the context of all of the findings—including lack of evidence for an impact in Phase 3—it seems equally plausible that the Phase 2 results simply reflect chance.

For Phase 2, the effects of duration to MM on NSTW months as of month 48 are jointly significant at the 5 percent level (bottom left panel). We do not reject the null hypothesis that the impact of duration to MM is linear over the rollout period. However, the slope of the trend line is not very large--it is -0.0059 NSTW months per month of duration to MM, which translates into a mean effect of -0.07 months for a 12-month delay in mailing the Ticket, over a 48-month period--and is not statistically significant. For Phase 3 (bottom right panel), we reject the hypothesis that all monthly values are zero at the 1 percent level, but the slope of the fitted line is *positive* (0.0055) effects, rather than negative, although not statistically significant. For Phase 3 we also reject the hypothesis that the impact of duration to MM is linear over the rollout period at the 5 percent level. Thus, it seems likely that the Phase 3 results simply reflect chance rather than an impact of duration to MM, and chance might equally well explain the Phase 2 results.

Interpretation of the Phase 2 results for NSTW months as indicative of impacts is also undermined by the evidence of impacts on TWP completion in the Phase 2 and 3 samples, described previously. We would not expect a negative impact on NSTW months unless there is a negative impact on TWP completion, as NSTW months cannot start until the TWP is completed. As described earlier, the evidence for an impact on TWP completion is very weak and, if anything, stronger for Phase 3 than for Phase 2. If we interpret the point estimates as impacts, we must conclude that an essentially zero impact on TWP completion for Phase 2 translated into a modest negative impact on NSTW months, while a modest negative impact on TWP months in Phase 3 translated into a modest *positive* impact on NSTW months. An alternative explanation of these inconsistent results is that they are all due to chance. The analysis of total impacts, presented in the next section, reinforces the conclusion that the marginally significant impacts on NSTW months found for Phase 2 are simply the result of chance.

D. Projections of Total Impacts of TTW

In this section we present projections of the impact of mailing Tickets in the first rollout month versus never mailing them, as of 12, 24, 36 and 48 months later. These are derived from the estimates for the impacts of the duration to MM at 12, 24, 36 and 48 months under two important, but quite plausible and partially verifiable assumptions. We call these estimates projections because they rely on two maintained assumptions.

The first assumption is the "linearity" assumption: that the marginal impact of delaying the mailing of the Ticket on each outcome as of month 12, 24, 36, or 48 is linear through month 13 of the 48-month observation period for each sample. This assumption is clearly consistent with the acceptance of the linearity restrictions for the service enrollment estimates as of month 12 (shown earlier) as well as analogous restrictions for months (24, 36 and 48—not shown). Linearity is sometimes rejected for TWP completion and NSTW months, but in these cases it appears that rejection is due to one or two outlier estimates; the results for TWP completion at 48 months in Phase 2 and NSTW months in Phase 3 shown above are illustrative.

The second assumption is the "impact only delayed" assumption: that the impact of mailing the Ticket on each outcome for those mailed Tickets in month 13 is always exactly 12 months behind the impact on enrollment for those mailed Tickets in month one. For instance, the impact of mailing Tickets in month 13 as of month 24, 36, or 48 is exactly the same as the impact of mailing the Ticket in month one as of month 12, 24, or 36, respectively. This assumption is clearly consistent with the service enrollment point estimates for months 24, 36 and 48, which are progressively smaller than impacts at month 12, and which suggest that service enrollment for those mailed Tickets late in the rollout had essentially caught up to service enrollment for those mailed Tickets earlier (see Stapleton et al. 2013). This assumption is also not contradicted by the evidence for other outcomes, although for those outcomes there is not consistent evidence of impacts.

Figure 2 illustrates how these assumptions are used to derive the projections for total impacts over the 48-month period after Ticket mailing on an outcome at 12, 24, 36, and 48 months after rollout start, using estimates for service enrollment impacts for Phase 2. Given the assumptions, the solid line traces the impact of delaying the mailing from month 1 to month 13 at each 12-month interval of the 48-month period. The dashed line represents the total impact of mailing the Ticket in month 13 as of month 24, 36, and 48. The length of the vertical double arrow represents the impact of delaying the mailing from month 1 to month 13 at each observation point. The sum of the estimated impacts of mailing the Ticket at month 13 instead of month 1 as of months 12, 24, 36, and 48 (illustrated by the lengths of the four vertical arrows) is the total impact as of month 48 of mailing the Ticket in month 1 versus not mailing it at all.

[INSERT FIGURE 2 HERE]

We applied this approach to all outcome variables in both phase samples. As illustrated in Figure 2 and shown in Table 5, the projected impact on service enrollment at month 48 for Phase 2 is 2.3 percentage points and statistically significant. The corresponding projection for Phase 3 is a more modest 1.2 percentage points, but also significant. Another feature of the service enrollment projections is that the point estimates increase with the projection month in each phase—reflecting the maintained assumptions and the fact that the restricted IV estimates of all coefficients in the

duration to MM models are positive. Further, for Phases 2 and 3, the increment to the projection diminishes with each 12-month period, as we would expect.

None of the projections for total impacts on other outcome variables are significant at even the 10 percent level as of any observation point. The NSTW-months estimates stand in stark contrast to those for service enrollment—the latter with uniformly positive point estimates and significant at the 0.10 level or better. These projections reinforce our earlier conclusion that there is no evidence of a substantial impact on any outcomes other than service enrollment.

Table 5. Projected Total Impacts on Service Enrollment, TWP Completion and NSTW Months as of 12, 24, 36 and 48 Months, by Phase

Sample and Observation Month	Service Enrollment	TWP Complete	NSTW Months
Phase 2			
12	1.004***	-0.070	-0.017
z-stat.	4.895	-0.326	-1.114
24	1.736***	-0.058	-0.033
z-stat.	4.395	-0.125	-0.661
36	2.038***	-0.016	-0.009
z-stat.	3.514	-0.021	-0.087
48	2.319***	0.064	0.062
z-stat.	3.047	0.058	0.329
Phase 3			
12	0.791***	-0.202	-0.005
z-stat.	4.674	-1.070	-0.353
24	1.028***	0.040	-0.035
z-stat.	3.153	0.098	-0.783
36	1.137***	0.408	-0.074
z-stat.	2.369	0.609	-0.769
48	1.210*	0.899	-0.140
z-stat.	1.876	0.936	-0.824

Note: "z-stat" is the standard normal test statistic for the hypothesis that the projected impact is zero; *, **, and *** indicate significant at the 0.10, 0.05, and 0.01 levels, respectively.

F. Assessment of Whether TTW Was Self-Financing by 2007

The fact that we did not find statistically significant impacts on NSTW months does not by itself rule out the possibility that TTW under the initial regulations had impacts on these outcomes that were sufficiently large for the program to be "self-financing"—that is, for savings from a net reduction in benefits to be sufficient to pay for TTW payments to providers and all administrative costs attributed to the program. Thornton (2012) suggests that only a very small impact—an increase of 3,000 or so in the number of all beneficiaries experiencing suspension or termination for work (STW) for the first time in each year—might be sufficient for the program to be self-financing. Because this issue is critical to policymakers, in this section we assess in more detail whether the estimates are consistent with the self-financing hypothesis.

An impact of 3,000 is quite small relative to the number of first-time STW cases actually observed in any recent year. Based on findings in Schimmel et al. (2013) and additional tabulations of their data, we estimate that an impact of 3,000 first STW cases is about five percent of the number of first STW cases in 2007 that would have occurred in the absence of TTW. An increase in STW of five percent would be sufficient for TTW to be self-financing only if NSTW increases by at

least the same relative amount; if instead those who attain STW as the result of TTW return to the rolls quickly rather than accumulating NSTW months, reductions in benefits would be minimal.

Under certain strong assumptions, we could conclude that a TTW impact of five percent or greater on NSTW at 48 months for new, young SSD-only beneficiaries would be large enough to have made the program self-financing in 2007.[20] The most problematic assumption is that 2007 can be interpreted as a steady state with respect to the number and characteristics of beneficiaries and their work activity. That assumption is required to interpret a cross-sectional impact of five percent in 2007 as the equivalent to a longitudinal impact of five percent for recent program entrants. In fact, 2007 was far from a steady state, primarily because the number of program entrants in the period leading up to 2007 was far larger than the number exiting the program. Because recent entrants are much more likely to enter STW and start to accumulate NSTW months than those who have been on the rolls for many months (Liu and Stapleton 2011), we conclude that the longitudinal percentage impact on NSTW months for recent entrants would have to be substantially higher than five percent in order to achieve a cross-sectional impact of five percent in 2007.

For now, however, we treat the five percent figure as a lower bound and test the following hypothesis: the mailing of Tickets to young, new SSD-only beneficiaries increased the number of NSTW months as of month 48 after the mailing by at least five percent versus the alternative hypothesis that the impact was less than five percent. We then consider how the results would change if the minimum percentage impact consistent with self-financing was larger than five percent, as it might well be.

We use the projected total impacts in Phase 2 and 3 separately, and then, to increase power, we pool the results for the two phases on the assumption that the true relative impacts for the two phases are the same. The pooled projection is the minimum variance projection under the assumption that percentage impacts were the same for Phases 2 and 3.[21] Because of the inequalities in the null and alternative hypotheses, a one-tailed test is appropriate. Results appear in Table 6. We also show tests for the null hypothesis of "no impact" versus the one-tailed alternative of "positive impact."

The statistical power of the projections for NSTW months is insufficient to rule out the possibility that TTW had impacts of at least five percent for Phases 2 and 3 pooled, but at the same time the evidence from these projections more consistent with zero or negative impacts than an impact of 5 percent or more. The percentage projections themselves are all *smaller* than 5 percent, and both the Phase 3 and pooled projections are negative (-8.0 percent and -3.4 percent, respectively). We cannot, however, reject the null hypothesis of a 5 percent impact based on the pooled sample (p-value of 0.14). Note, though, that the p-value for that test is much smaller than the p-value for the test of the null hypothesis that the true impact is zero or negative (0.67 percent). That is, the evidence is more consistent with the hypothesis of a zero or negative impact than with an impact of at least 5 percent.

---

[20] These assumptions are described and assessed in more detail in Stapleton et al. (2013, Appendix D).

[21] The minimum variance estimate is a weighted mean of the estimates for the two phases where the weights have been chosen to minimize the variance of the estimate. More weight is given to the Phase 3 estimate for each impact because the Phase 3 estimate has lower variance than the Phase 2 estimate.

Table 6. Projected Relative Impacts on NSTW Months at 48 Months After Mailing

	Phase 2	Phase 3	Pooled
NSTW Months			
Projected relative impacts at 48 months	4.1%	-8.0%	-3.4%
Standard error of relative impacts	12.5%	9.7%	7.7%
P-value for test of "no impact (or negative impact)" versus "positive impact"	0.371	0.795	0.673
P-value for test of "impact of 5.0% (or more)" versus "impact less than 5.0%"	0.471	0.090	0.136

Notes: The relative projected impacts were calculated by comparing the projected total impacts on NSTW as of 48 months and the estimated means in the absence of TTW (counterfactual). For each outcome, the counterfactual mean was estimated by subtracting the weighted mean of the Phase 2 and 3 impact estimates at 48 months from the actual mean for the phase. The p-values are for one-tailed tests, reflecting the inequalities in the hypotheses.

As indicated above, the fact that 2007 followed a period of rapid program growth leads us to conclude that the smallest percentage impact for young SSD-only entrants that is consistent with self-financing is larger than five percent. If we had used a larger value in the tests above, the results would clearly be less favorable to the hypothesis of self-financing. For instance, a value of nine percent would lead to rejection of the hypothesis of self-financing at the five percent significance level using the pooled data. That is, if self-financing required at least a nine percent impact on NSTW months—a plausible value—we would have to reject the hypothesis that TTW was self-financing as of 2007.

IV. CONCLUSION

We find clear evidence that the mailing of Tickets during the rollout period did increase service enrollment. The most important findings are captured in their implications for the impact of Ticket mailing, versus no Ticket mailing, on service enrollment over the next 48 months. The Phase 2 and 3 findings are very significant and consistent with each other. The Phase 2 point estimates imply that the impact of mailing Tickets is 1.0 percentage points 12 months later, and 2.3 percentage points 48 months later. The corresponding estimates for Phase 3 are 0.8 and 1.2 percentage points. All of these estimates are very significant statistically. They are also large relative to what service enrollment would have been in the absence of ticket; the 48-month estimates imply relative impacts on the order of 50 percent and 25 percent for the two phases, respectively.[22] The point estimates are quite comparable to results from earlier impact analysis for SSD-only beneficiaries under age 40: a 0.6 percentage point increase in service enrollment by the end of the rollout year and a 1.5 percentage point increase at the end of the following year (Thornton et al. 2007; Stapleton et al. 2008). Another feature of the findings is that, by month 48 after rollout start, service enrollment for those mailed Tickets late within each rollout period had essentially caught up with service enrollment for those who were mailed Tickets earlier in the rollout period.

The analysis provides no consistent evidence of impacts on other outcomes. Some estimates for Phase 2 are suggestive of an impact, but it seems likely that they are due to chance. Specifically, marginally significant Phase 2 point estimates for NSTW months imply that a 12-month delay in

---

[22] We used 4.5 percent as the counterfactual value for both phases, based on the following calculations. The percentages enrolled at 48 months for the two phases are 6.8 and 5.7 percent. If we assume that as of month 48 the impacts for those mailed Tickets in later rollout months had caught up with impacts for those mailed Tickets in the first month, then these percentages would have both been 4.5 percent if the Ticket had never been mailed (6.8 – 2.3 and 5.7 – 1.2, respectively). The actual values are 51.1 and 26.7 percent larger than 4.5 percent, respectively.

mailing a Ticket decreases the number of NSTW months as of month 48 by an average of 0.07 months—approximately a five percent decrease.[23] There are, however, substantial reasons to believe that these results are simply due to chance. The fundamental reason is the multiple comparisons problem; whenever an evaluation produces impacts for many outcomes, there are bound to be a few statistically significant findings by chance alone even if the intervention has absolutely no impacts. We have produced impact estimates for many different outcomes (not all independent), so we would expect to find that some estimated impacts beyond those for service enrollment would be statistically significant even if there are no impacts on these outcomes. Hence, to assess whether the Phase 2 results for NSTW months reflect real impacts or simply chance, it is important to consider them in the context of all the estimates produced—are the latter consistent with real impacts for these outcomes in Phase 2?[24] In brief, the Phase 3 point estimate for the impact on NSTW months as of 48 months is positive, that is in the opposite direction found for Phase 2, and just as large. It is very hard to understand why comparable impacts on service enrollment in the two samples would translate into such different impacts for NSTW months. Further, the point estimates for the impact on TWP completion—a necessary precursor to accumulation of NSTW months—at 48 months is essentially zero for phase 2 and negative for Phase 3. Examination of the plots of coefficients for individual MM in Section III reveals that outlier estimates for early and late rollout months appear to explain the estimated relationship between duration to MM and TWP completion for the two phase samples, rather than the impacts of Ticket mailing.

Although we have considerable confidence that the methodology provides unbiased estimates and can detect impacts as small as those found for service enrollment, we have also illustrated that its power is not sufficient to rule out the small positive impacts on NSTW months that would be required for TTW to have been self-financing in 2007. Overall, however, the evidence is more consistent with no impact on NSTW months than with an impact large enough to make the program self-financing.

It is important to keep in mind that these estimates are for TTW under the original regulations. The 2008 changes to the TTW regulations clearly stimulated provider interest and the number of beneficiaries assigning their Tickets (Schimmel et al. 2013). In principle, those changes could have had a positive impact on NSTW months. However, it appears impossible to rigorously measure any such impact because the new regulations were implemented nationally, without a test, and also because implementation of the new regulations occurred during a recession that was deeper and longer lasting than any recession that has occurred since the 1956 inception of DI.

The analysis provides a lesson for SSA and other agencies when, in the future, they are asked to make a significant change to a large national or state program—including significant future changes to TTW. Inasmuch as such a change often requires a lengthy rollout period, the agency should consider the knowledge that might be gained by implementing a rollout in which program participants are randomly assigned an implementation month over a period of 12 months or so. This approach has its limits, however; it will not necessarily have sufficient power to identify substantively important impacts if such impacts are very small. Power can be increased if the program participants most likely to be affected by the change can be identified in advance, the rollout period can be lengthened, or a more extreme version of the change could be applied to randomly chosen

---

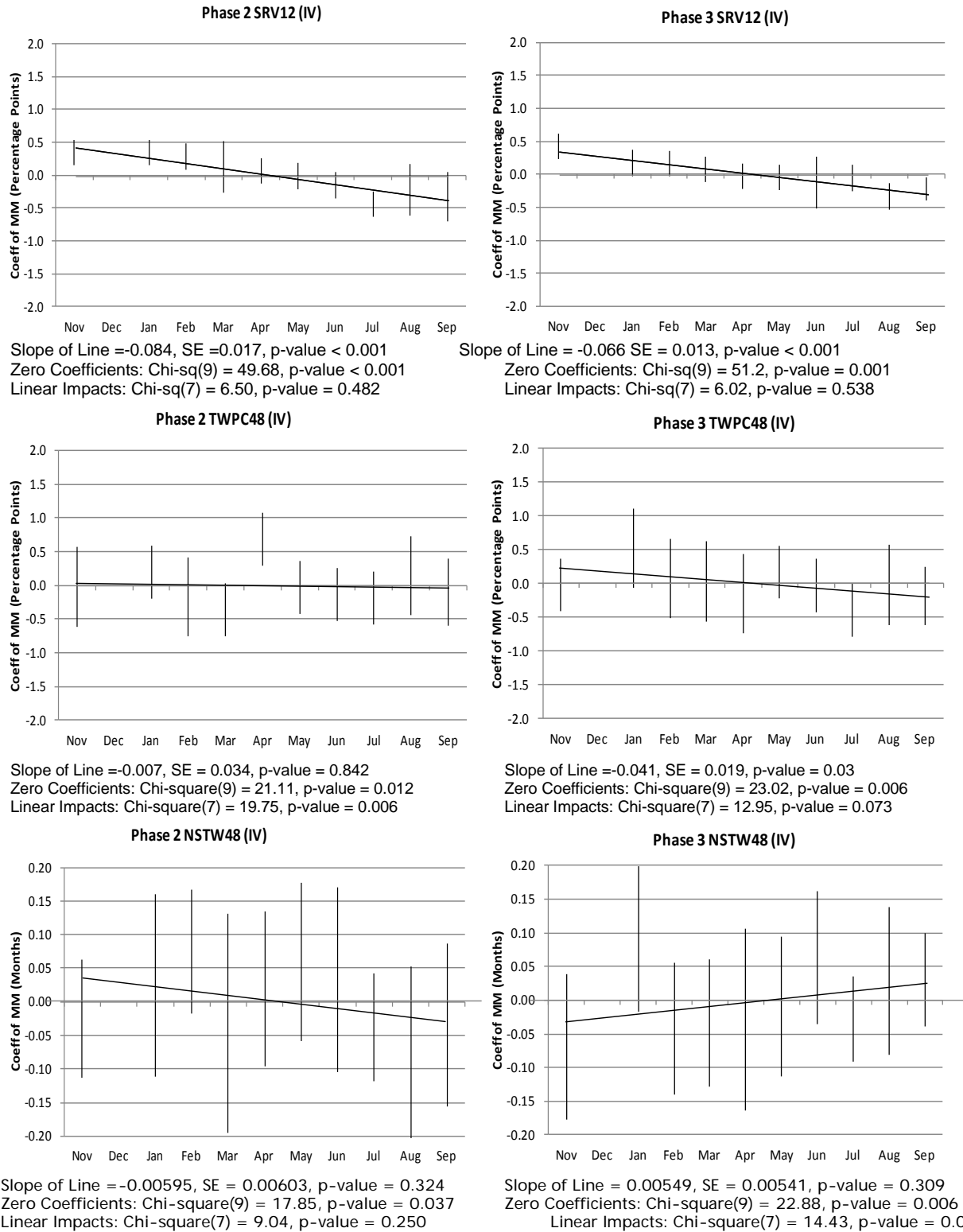[23] The mean of NSTW months as of month 48 in Phase 2 is1.46 months.

[24] There are formal ways to address the multiple comparison problem (see Schochet 2008, 2009). We have not conducted a more formal analysis because so few estimates other than those for service enrollment are even marginally significant.

participants. Such enhancements make this approach more like the approach that would be best from a purely methodological perspective: a randomized control trial.

REFERENCES

Angrist, J., G. Imbens, and D. Rubin. "Identification of Causal Effects Using Instrumental Variables." Journal of the American Statistical Association, vol. 91, no. 434, 1996, pp. 444–455.

Liu, Su, and David Stapleton. "Longitudinal Statistics on Work Activity and Use of Employment Supports for New Social Security Disability Insurance Beneficiaries." Social Security Bulletin, vol. 71, no. 3, August 2011, pp. 11–75.

Mamun, Arif, Paul O'Leary, David Wittenburg, and Jesse Gregory. "Employment Among Social Security Disability Program Beneficiaries: 1996-2007." Social Security Bulletin, vol. 71, no. 3, 2011, pp. 11–34.

Social Security Administration. Annual Statistical Report on the Social Security Disability Insurance Program, 2011. Washington, DC, July 2012.

Schimmel, Jody, David Stapleton, David Mann, Sarah Prenovitz, and Dawn Phelps. "Participant and Provider Outcomes since the Inception of Ticket to Work and an Assessment of the Effect of the Changes to the Regulations in 2008." Draft report submitted to the Social Security Administration. Washington, DC: Mathematica Policy Research, June 2013.

Schochet, Peter Z. "Guidelines for Multiple Testing in Impact Evaluations of Educational Interventions." Final report submitted to the U.S. Department of Education, Institute of Education Sciences. Princeton, NJ: Mathematical Policy Research, May 2008.

Schochet, Peter Z. "An Approach for Addressing the Multiple Testing Problem in Social Policy Impact Evaluations." Evaluation Review, vol. 33, no. 6, December 2009.

Stapleton, David, Arif Mamun, and Jeremy Page. "Initial Impacts of the Ticket to Work Program for Young New Social Security Disability Awardees: Estimates Based on Randomly Assigned Mail Months." Washington, DC: Mathematica Policy Research, 2013.

Stapleton, David, Gina Livermore, Bonnie O'Day, Craig Thornton, Robert Weathers, Krista Harrison, So O'Neil, Emily Sama Martin, David Wittenburg, and Debra Wright. "Ticket to Work at the Crossroads: A Solid Foundation with an Uncertain Future." Report submitted to the Social Security Administration. Washington, DC: Mathematica Policy Research, September 2008.

Thornton, Craig, Robert Weathers, David Wittenburg, Thomas Fraker, David Stapleton, Jesse Gregory, and Arif Mamun. "Initial Impacts of the Ticket to Work Program on Social Security Disability Beneficiary Service Enrollment, Earnings and Benefits." Journal of Vocational Rehabilitation, vol. 27, no. 2, 2007, pp. 129–140.

Thornton, Craig. "Can the Ticket to Work Program Be Self-Financing?" Washington, DC: Mathematica Policy Research, 2012.

Figure 1. IV Estimates for Impacts of Duration to MM on Service Enrollment at 12 Months, TWP Completion at 48 Months, and NSTW Months at 48 Months Following Rollout Start, Phases 2 and 3

Phase 2 SRV12 (IV)

Slope of Line =-0.084, SE =0.017, p-value < 0.001
Zero Coefficients: Chi-sq(9) = 49.68, p-value < 0.001
Linear Impacts: Chi-sq(7) = 6.50, p-value = 0.482

Phase 3 SRV12 (IV)

Slope of Line = -0.066 SE = 0.013, p-value < 0.001
Zero Coefficients: Chi-sq(9) = 51.2, p-value = 0.001
Linear Impacts: Chi-sq(7) = 6.02, p-value = 0.538

Phase 2 TWPC48 (IV)

Slope of Line =-0.007, SE = 0.034, p-value = 0.842
Zero Coefficients: Chi-square(9) = 21.11, p-value = 0.012
Linear Impacts: Chi-square(7) = 19.75, p-value = 0.006

Phase 3 TWPC48 (IV)

Slope of Line =-0.041, SE = 0.019, p-value = 0.03
Zero Coefficients: Chi-square(9) = 23.02, p-value = 0.006
Linear Impacts: Chi-square(7) = 12.95, p-value = 0.073

Phase 2 NSTW48 (IV)

Slope of Line =-0.00595, SE = 0.00603, p-value = 0.324
Zero Coefficients: Chi-square(9) = 17.85, p-value = 0.037
Linear Impacts: Chi-square(7) = 9.04, p-value = 0.250

Phase 3 NSTW48 (IV)

Slope of Line = 0.00549, SE = 0.00541, p-value = 0.309
Zero Coefficients: Chi-square(9) = 22.88, p-value = 0.006
Linear Impacts: Chi-square(7) = 14.43, p-value = 0.044

Notes: IV using IMM as instruments for MM. All estimates are constrained to sum to zero. The vertical line on each coefficient shows the 95 percent confidence interval. The line shows the estimates with linear restrictions imposed on the coefficients. The first test statistic is for the test of the null hypothesis that all the coefficients are zero, and the second is for the linear restrictions on the coefficients.

Figure 2. Illustrative Computation of the Projection for the Total Impact of Mailing the Ticket on an Outcome in Month 48 (Service Enrollment, Phase 2)