The Effects of Vocational Rehabilitation for People with Cognitive Impairments

David Dean  
University of Richmond

John Pepper  
University of Virginia

Robert Schmidt  
University of Richmond

Steven Stern*  
University of Virginia

October 7, 2013

Abstract

The public-sector Vocational Rehabilitation (VR) program is a $3 billion federal-state partnership designed to provide employment-related assistance to persons with disabilities. There is, however, relatively little known about the long-term efficacy of VR programs. This paper utilizes unique and detailed administrative and employment data to examine both short and longer-term employment impacts for all persons diagnosed with cognitive impairments who applied for VR services in the state of Virginia in State Fiscal Years 1988 and 2000. These data provide quarterly information on VR services and employment outcomes for many years before and after the application quarter. We improve the literature on this topic by a) allowing for a richer set of service choices beyond the usual binary choice, b) use long-run labor market data, c) model behavior allowing for a much richer set of policy analyses, and d) construct a reasonable set of instrumental variables. Estimates from our model of service provision and labor market outcomes reveal that VR services generally have positive long-run labor market outcome effects that appear to substantially exceed the cost of providing services.

1 Introduction

The deinstitutionalization of persons with intellectual disabilities in the United States that started in the late 1960’s led to an increased demand for vocational

*We would like to thank Joe Ashley, John Phelps, Kirsten Rowe, Ann Stanfield, Vlad Mednikov, and Jim Rothrock from Virginia Department of Rehabilitation Services (DRS), Jeff Smith, and participants at seminars at the University of Virginia, APPAM, ASSA, Econometric Society, and University of Maryland for excellent help and advice and Rachel Fowley for excellent research assistance. The DRS, the Virginia Department of Medical Assistive Services, the National Institute for Disability and Rehabilitation Research, and the University of Virginia Bankard Fund for Political Economy provided generous financial support. All errors are ours.
services for such individuals who were in need of expanded employment opportunities (Gidugu et al., 2011). Legislation extending vocational rehabilitation (VR) assistance culminated in the landmark reauthorization of the Vocational Rehabilitation Act of 1973. VR services, previously designed for disabled veterans and persons with occupationally-related impairments, were mandated to be made available to a broader spectrum of persons with disabilities, including those with intellectual disabilities. In this paper, we provide the first evaluation of the long-term efficacy of VR service provision in the United States on the employment-related outcomes of persons with cognitive impairments.

The federal-state VR program partnership, administered by the federal Rehabilitation Services Administration (RSA), currently gives approximately $3 billion annually to state agencies to provide a wide variety of vocational rehabilitation services to individuals with a broad spectrum of disabling conditions. During the past decade, state VR agencies have closed an average of over 600,000 cases annually, with a very stable 1/9 of these being cases with a diagnosis of intellectual or developmental disability.¹ These individuals receive a service regimen that includes human capital development ranging from pre-vocational training through supported employment. Some may even receive post-secondary education (Migliore and Butterworth, 2008). Yet, the last evaluation of the U.S. public-sector VR program published in an economics journal, Dean and Dolan (1991), is from over 20 years ago.²

In this paper, we provide an updated and innovative evaluation of the impact of VR services on clients with cognitive impairments using unique panel data on all persons who applied for services in the state of Virginia in State Fiscal Years 1988 or 2000.³ For each applicant cohort, the data reveal a long panel of employment records and VR services as well as information on the client’s limitations and other characteristics at the time of application. The data on the 2000 cohort, for example, provide quarterly employment and earnings information as well as detailed information on VR services from 1995 to 2008.

Focusing attention on the important group of clients with cognitive impairments, we are able to make a number of substantive contributions to the VR evaluation literature. At the most basic level, we evaluate the short and long run labor market effects of VR services and examine the impact of specific types of services rather than just a single treatment indicator. To address the selection problem, we formalize and estimate a structural model of endogenous service provision and labor market outcomes. We identify the parameters of this model using instrumental variables that are assumed to impact service receipt but not

¹Source: Rehabilitation Services Administration, RSA-911, reported in Butterworth et al. (2011). Table 8.
²Evaluation of the efficacy of such employment services for persons with cognitive impairments began in earnest in the early 1990s with a series of randomized, controlled experiments involving transitional/supported employment training to recipients of federal Supplemental Security Income (SSI) disability payments (e.g., Prero and Thornton, 1991). Subsequent research using small-scale demonstrations evaluate the effectiveness of training services on short-term employment outcomes (see Howarth et al., 2006).
³Dean et al. (2013a) conduct a similar analysis focused on clients with mental illnesses.
the latent labor market outcomes and pre-program labor market outcomes that control for differences between those who will and will not receive services.\footnote{While these models have not been used to evaluate the VR program in the United States, several studies of the European active labor market programs for persons with disabilities have applied such methodologies (e.g., Frolich, Heshmati, and Lechner, 2004; Aakvik, Heckman, and Vytlacil, 2005).}

In addition, with data on the 1988 and 2000 applicants, we are able to examine cohorts from different epochs where the services provided and clients served have notably changed.\footnote{From a data-gathering perspective, 1988 was the first year in which a “relational” data collection system was implemented within the Virginia DRS (Vocational Rehabilitation Information System, VRIS) and this allowed, for the first time, a linkage between the purchased services by the state VR agency with the individual receiving those services.} Over this period, there has been a dramatic shift in the types of training procured by VR agencies which, at one level, can be viewed as a movement away from sheltered to integrated supported employment and, at another level, as the movement from VR agencies purchasing pre-vocational and work-adjustment training in a “train and place” environment at a comprehensive rehabilitation facility to purchasing job coach training and supported employment services in an integrated work environment. Rusch and Braddock (2004) note a more than doubling in the share of supported employment participants of total day/work program participants from 1988 to 2002.

The paper proceeds as follows: Section 2 describes the multivariate discrete choice model for service provision choices and labor market outcomes used throughout this paper. We allow for correlation of errors among all of the equations. In Sections 3 and 4, we describe the data used in our analysis and the methodology used to estimate the parameters of the discrete choice structural model. Estimation results are presented in Section 5, and a rate-of-return analysis is presented in Section 6. The paper ends with conclusions.

2 Model

In this section, we describe the model of behavior to be estimated. It follows directly from and uses the same notation as Dean et al. (2013a). Given the high proportion of individuals receiving more than one service type, we allow for the possibility of receipt of multiple services in our model; i.e., we have a multivariate binary choice model for service provision rather than a polychotomous discrete choice model. In particular, let $y^*_{ij}$ be the value for individual $i$ of participating in service $j$, $j = 1, 2, ..., J$, and define $y_{ij} = 1 (y^*_{ij} > 0)$ to be an indicator for whether $i$ receives service $j$. Assume that

$$ y^*_{ij} = X^y_{ij} \beta_j + u^y_{ij} + \varepsilon_{ij}, \tag{1} $$

$$ \varepsilon_{ij} \sim Logistic $$

where $X^y_{ij}$ is a vector of exogenous explanatory variables, and $u^y_{ij}$ is an error whose structure is specified below. Next, let $z^w_{it}$ be the value to $i$ of working at...
quarter $t$, and define $z_{it} = 1(z_{it}^* > 0)$. Assume that

$$z_{it}^* = X_{it}^* \gamma + \sum_{k=1}^{K} d_{ik} \sum_{j=1}^{J} \alpha_{jk}^z y_{ij} + u_{it}^z + v_{it}^z \quad (2)$$

where $X_{it}^*$ is a vector of (possibly) time-varying, exogenous explanatory variables, $d_{ik}$ is a dummy variable equal to one if the amount of time between last quarter of service receipt and $t$ is between $\tau_k$ and $\tau_{k+1}$, and $u_{it}^z$ is an error whose structure is specified below. We use four time nodes: a) 2 or more quarters before service, b) 1 quarter before service, c) from 1 quarter after service to 8 quarters after service, and d) 9 or more quarters after service.

Next let $w_{it}$ be the log quarterly earnings of $i$ at $t$, and assume that

$$w_{it} = X_{it}^w \delta + \sum_{k=1}^{K} d_{ik} \sum_{j=1}^{J} \alpha_{jk}^w y_{ij} + u_{it}^w + v_{it}^w \quad (3)$$

where variables are defined analogously to equation (2). Finally, assume that

$$u_{ij}^w = \lambda_{ij1}^y e_{i1} + \lambda_{ij2}^w e_{i2},$$

$$u_{it}^z = \lambda_{i1}^z e_{i1} + \lambda_{i2}^z e_{i2} + \eta_{it}^z,$$

$$u_{it}^w = \lambda_{i1}^w e_{i1} + \lambda_{i2}^w e_{i2} + \eta_{it}^w,$$

$$\eta_{it}^z = \rho_{y} \eta_{it-1} + \zeta_{it}^z,$$

$$\eta_{it}^w = \rho_{y} \eta_{it-1} + \zeta_{it}^w,$$

$$\begin{pmatrix} \zeta_{it}^z \\ \zeta_{it}^w \end{pmatrix} \sim iidN \left(0, \begin{pmatrix} \sigma_{\zeta}^2 & \rho_{\zeta} \\ \rho_{\zeta} & 1 \end{pmatrix} \right),$$

$$\begin{pmatrix} e_{i1} \\ e_{i2} \end{pmatrix} \sim iidN [0, I],$$

$$v_{it}^z \sim iidN [0, 1],$$

$$v_{it}^w \sim iidN [0, \sigma_{v}^2].$$

We include the $(e_{i1}, e_{i2})$ to allow for two common factors affecting all dependent variables with factor loadings $(\lambda_{ij1}^y, \lambda_{ij2}^z, \lambda_{ij2}^w)_{k=1}^K$. We also allow for serial correlation and contemporaneous correlation in the labor market errors $(\eta_{it}^z, \eta_{it}^w)$. The covariance matrix of the errors implied by equation (4) is provided in an online appendix.\(^6\)

3 Data

We use three main sources of data: a) the administrative records from the Virginia Department of Rehabilitative Services (DRS) for a cohort of applicants in 2000, b) the administrative records from a cohort of DRS applicants in 1988, c) the administrative records from a cohort of DRS applicants in 1988.

\(^6\)See www.people.virginia.edu/~sns5r/resint/vocrehstf/vocrehciappendix.pdf.
Table 1: Missing Value Analysis for 1988 and 2000 Cohorts

<table>
<thead>
<tr>
<th>Cause</th>
<th>1988 Cohort</th>
<th>2000 Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># Obs Lost</td>
<td>Proportion of Total</td>
</tr>
<tr>
<td>Applicants in SFY</td>
<td>11596</td>
<td></td>
</tr>
<tr>
<td>Missing or Questionable SSN</td>
<td>14</td>
<td>0.001</td>
</tr>
<tr>
<td>Died While in Program</td>
<td>70</td>
<td>0.006</td>
</tr>
<tr>
<td>Missing Gender or Date of Birth</td>
<td>10</td>
<td>0.001</td>
</tr>
<tr>
<td>Not Residing in Virginia</td>
<td>107</td>
<td>0.009</td>
</tr>
<tr>
<td>No Cognitive Impairment</td>
<td>9488</td>
<td>0.818</td>
</tr>
<tr>
<td>Missing Primary Disability</td>
<td>35</td>
<td>0.003</td>
</tr>
<tr>
<td>Missing Secondary Disability</td>
<td>5</td>
<td>0.000</td>
</tr>
<tr>
<td>Initial Service Spell before SFY 2000</td>
<td>822</td>
<td>0.080</td>
</tr>
<tr>
<td>Age Younger than 17 Years</td>
<td>117</td>
<td>0.011</td>
</tr>
<tr>
<td>Neither VR Service nor Employment Record</td>
<td>57</td>
<td>0.006</td>
</tr>
<tr>
<td>Number Remaining in Sample</td>
<td>1907</td>
<td>0.164</td>
</tr>
</tbody>
</table>

and c) administrative records from the Virginia Employment Commission from the third quarter of 1984 to the fourth quarter of 2009 for those people in the DRS data. We also merge these records with data from the Bureau of Economic Analysis on county-specific employment patterns. Each of these is described in turn in the discussion below.

3.1 DRS 2000 Applicant Cohort

3.1.1 Sample Frame

We begin with the administrative records of the Virginia DRS for the 10323 individuals who applied for VR services in SFY 2000 (July 1, 1999 - June 30, 2000). Table 1 provides information about sequential selection into our sample. The major cause of selection out of the sample is that the individual does not have a cognitive impairment. We define someone as having a cognitive impairment if either their primary or secondary diagnosis was cognitive impairment in any of the service episodes observed in the data.\(^7\) Approximately 78% of the original sample is excluded because of this restriction. We exclude another 8% of individuals who had a DRS service episode prior to 2000. We do this to avoid left-censoring issues discussed in Heckman and Singer (1984). In particular, an individual participating in a subsequent service episode may be doing so because she was (endogenously) unsuccessful in the labor market or because she found the first service episode (endogenously) unusually productive. In either case, inclusion of such individuals causes estimation bias. Later, in Section 5.2, we provide some measures of the size of this bias. There are a number of other selection criteria, listed in Table 1, that have minor impacts on the final sample. After selection, we have a sample size of 1009 individuals.

\(^7\)In 2000 RSA coding, the equivalent of having a cognitive impairment was having a disability “cause” of mental retardation.
3.1.2 Service Provision

New clients to DRS are assigned a counselor who determines whether the individual is eligible for services. This step usually involves the counselor recording a diagnosis for the individual. At this point, the individual may be administratively closed because the counselor determines that the individual’s disability is insufficiently severe or that it is too severe to benefit from VR services. Also, at this point, the individual may withdraw from further consideration of service receipt. If the individual is determined to be eligible for VR services and does not withdraw, then the counselor and individual together develop an individualized plan for employment (IPE) which specifies the array of services to be provided.

Services are provided a) internally by DRS personnel, b) as a “similar benefit” purchased or provided by another governmental agency or not-for-profit organization with no charge to DRS, c) as a “purchased service” through an outside vendor using DRS funds, or d) as a combination of (a), (b), and (c). We are not able to observe services provided in-house or through similar benefits. Based on the 1988 cohort discussed in Section 3.2, it appears that most of the in-house services were for diagnosis and evaluation.

Purchased service data come from the administrative records of DRS. These data contain the information required of each state to supply to the US Department of Education’s Rehabilitation Services Administration as the RSA-911 Case Service Report. Available data for purchased services include the beginning and ending dates of service provision, expenditures, and types of purchased services. Throughout the analysis, the unit of time is a quarter (because we observe quarterly labor market data). To avoid issues associated with varying lengths of time of service receipt and issues associated with how to interpret quarters when the individual is both receiving services and participating in the labor market, we offset earnings from the quarter in which an individual applied for service. We then separate post-service earnings periods into a short run period of up to eight quarters post-application and a long run period of nine quarters or more. For the 2000 cohort, this rule results in 16 to 19 quarters of pre-service earnings quarters and 28 to 31 post-service quarters.\(^8\)

For most of the paper, we focus exclusively on the receipt of purchased services. However, when performing a cost/benefit and rate-of-return analysis in Section 6, we impute a value of total service costs. Also, in our model,

---

\(^8\)While this is a simple and appealing way to define the date of service receipt, there are potential limitations of this measure (see Dean et al., 2013a). Most notably, since some clients receive services over multiple quarters, the estimated short-run effects of VR services may be downward biased. Using a Lagrange Multiplier test, however, we find no evidence that service duration is statistically significant (see the online Appendix for further details). Still, our specification distinguishes between outcomes 8 or fewer quarters after service and 9 or more quarters after service. For this sample of VR clients, service receipt lasts for 3 quarters or less for 97.2% receiving diagnosis & evaluation, 78.1% receiving training, 92.1% receiving education, 96.2% receiving restoration, 91.8% receiving maintenance, and 84.8% receiving other services, and it lasts for 9 quarters or more for 2.6% receiving education and less than 0.3% for all other services. Thus, for the most part, one can interpret the results for 9 or more quarters as being post-service receipt.
we focus on binary measures of service receipt and ignore information about either the length of service receipt or the service expenditure. We do this because we are missing some data on service expenditure and provision dates and the standard approach for evaluating labor market training and VR programs is to focus on binary indicators of service provision (see, for example, Dean and Dolan, 1991; LaLonde, 1995; Friedlander, Greenberg, and Robins, 1997; Heckman, LaLonde, and Smith, 1999; and Imbens and Wooldridge, 2009).

There are 76 separate services provided by DRS, other state agencies, and 465 vendors. Since we can not estimate a set of coefficients for so many different services, nor would it be particularly useful to do so, we aggregate services, following Dean et al. (2002), into six service types listed in Table 2. Diagnosis & evaluation are provided at intake in assessing eligibility and developing an IPE. Training includes vocationally-oriented expenditures for on-the-job training, job coach training, work adjustment, and supported employment. Education includes tuition and fees for a GED (graduate equivalency degree) program, a vocational or business school, a community college, or a university. Restoration covers a wide variety of medical expenditures including dental services, hearing/speech services, eyeglasses and contact lenses, drug and alcohol treatments, psychological services, surgical procedures, hospitalization, prosthetic devices, and other assistive devices. Maintenance includes cash payments to facilitate everyday living and covers such items as transportation, clothing, motor vehicle and/or home modifications, and services to family members. Other services consists of payments outside of the previous categories such as for tools and equipment.

Table 2 shows that diagnosis & evaluation and training are the two most popular purchased services. One should note that a much higher proportion of clients receive diagnosis & evaluation, but they receive it in-house. The clients who receive purchased diagnosis & evaluation may have unusually difficult cases to diagnose and thus may be different than other clients in important unobserved ways (see Dean et al., 2013a). After diagnosis & evaluation and training, restoration, maintenance, and other services are received by approximately 25%.

---

Table 2: Proportion Receiving DRS Purchased Services by Type

<table>
<thead>
<tr>
<th>Variable</th>
<th>1988 Cohort</th>
<th>2000 Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis &amp; Evaluation</td>
<td>0.896</td>
<td>0.447</td>
</tr>
<tr>
<td>Training</td>
<td>0.519</td>
<td>0.426</td>
</tr>
<tr>
<td>Education</td>
<td>0.023</td>
<td>0.034</td>
</tr>
<tr>
<td>Restoration</td>
<td>0.202</td>
<td>0.208</td>
</tr>
<tr>
<td>Maintenance</td>
<td>0.372</td>
<td>0.251</td>
</tr>
<tr>
<td>Other Service</td>
<td>0.462</td>
<td>0.249</td>
</tr>
</tbody>
</table>

---

We put variable names in a different font to avoid confusion.
of clients, and then education is received by only 3.4% of clients. It should be noted that 5.1% of applicants with cognitive impairments in 2000 are not accepted into the program, and another 18.5% drop out after acceptance but before receiving substantive services. Thus, there are many clients who receive no services or who receive only diagnosis & evaluation.

Finally, while some clients receive no services, others receive multiple services. For example, 86 clients receive only diagnosis & evaluation (d), while 50 receive a combination of diagnosis & evaluation (d), training (t), and maintenance (m).

3.1.3 Explanatory Variables

For each individual in our sample, we observe a rich set of explanatory variables listed in Table 3. Besides the usual demographic variables such as gender, race, and age, we observe some disability measures not commonly included in other data sets and some other variables particularly relevant for this population. We have dummy variables for hearing/speech disability, musculo/skeletal disability, internal disability, learning disability, and mental illness. Also, we have a measure of severity of the individual’s disability, evaluated by the counselor, recorded as either not significant disability (the reference case), significant disability, or most significant disability.

Besides the usual demographic characteristics, we also observe whether the individual received special education services. Given our population of interest, about 33% received such services. A significant number of observations had missing information about education. Rather than delete them, we include a dummy variable for when education is missing. Finally, we observe variables such as marital status, two transportation variables, and a dummy variable for receipt of government assistance.\textsuperscript{10}

\begin{table}[h]
\centering
\caption{Moments of Explanatory Variables for 2000 Cohort}
\begin{tabular}{lcc|lcc}
\hline
Variable & Demographic Variables & Disability Variables &  \\
\hline
Male & 0.506 & 0.500 & Type & 0.041 & 0.197 \\
White & 0.557 & 0.497 & Hearing/Speech Disability & 0.067 & 0.251 \\
Education & 6.770 & 5.423 & Musculo/Skeletal Disability & 0.061 & 0.240 \\
Special Education & 0.332 & 0.471 & Internal Disability & 0.061 & 0.240 \\
Education Missing & 0.033 & 0.178 & Learning Disability & 0.046 & 0.209 \\
Age (Quarters/100) & 1.004 & 0.409 & Mental Illness & 0.183 & 0.387 \\
Married & 0.041 & 0.197 & Other Disability & 0.058 & 0.235 \\
Dependents & 0.366 & 0.922 & Extent &  \\
Transportation Available & 0.460 & 0.498 & Significant Disability & 0.573 & 0.495 \\
Has Driving License & 0.174 & 0.379 & Most Significant Disability & 0.401 & 0.490 \\
Receives Government Assistance & 0.218 & 0.255 &  \\
\hline
\end{tabular}
\end{table}

\textsuperscript{10}For most analyses, receipt of government assistance is endogenous because the rules associated with receipt depend critically on involvement in the labor market. However, for our population, most of the time, an individual can participate in the labor market to some degree without losing their benefits.
We are concerned about possible endogeneity of service provision. As described in much of the literature, individuals may have unobserved characteristics causing them to perform poorly in the labor market and to receive rehabilitation services. However, it might be that individuals have unobserved characteristics causing rehabilitation services to be unusually productive, thus leading to higher propensities to use the service and good labor market outcomes afterwards. We ameliorate this problem by using a rich set of observed covariates and by controlling for employment patterns and quarterly earnings prior to vocational rehabilitation service receipt. But, most importantly, we address this problem by using two instrumental variables for each of the six binary service provision variables. The first is the proportion of other clients of the individual’s counselor who were provided with the service. The second is the proportion of other clients of the individual’s VR field office who were provided with the service. For both field offices and counselors, we observe significant variation in the proportion of clients who receive each of the six service types. For example, while 19% of field offices provide training to no more than 21% of their clients, 79% provide training to no less than 34% of their clients. More importantly, there is significant dispersion in service provision across field offices and counselors. Using a likelihood ratio test, we reject the null hypothesis that the joint density of services within offices does not vary across offices with a test statistic of 407.8 (with 240 df and normalized value of 7.66). We also can test the null hypothesis that each office provides each service in the same proportion, one at a time, using a likelihood ratio test. The test statistic is 527.5 (with 288 df and a normalized value of 9.98). For counselors, the analogous test statistics are 957.3 (with 765 df and a normalized value of 40.92) and 2366.7 (with 918 df and a normalized value of 33.81). The fact that there is significant variation in the provision of services across offices and counselors make our instrument viable.

3.2 DRS 1988 Applicant Cohort

In addition to evaluating the clients from the 2000 cohort, we also evaluate data from applicants in 1988. Over this period, there have been important changes in the classification of persons with cognitive impairments and in the provision of VR services to these persons. The 1988 cohort included only persons with intellectual disabilities, while, by 2000, the general definition of cognitive impairment had broadened to include persons with learning disabilities. To maintain consistency in selection rules across the two years, we restricted the samples based on the 1988 definition in both years.\footnote{In section 3.2.3, we discuss how we control for learning disabilities in the 2000 cohort using a dummy explanatory variable.} Between 1988 and 2000, several changes occurred for persons with cognitive impairments. The Americans with Disabilities Act of 1990 opened up greater employment opportunities (see, for example, Hotchkiss 2003; and Jolls 2004 for a discussion on the effect of the ADA on labor market outcomes for people with disabilities). Indeed,
the overall civil rights movement for persons with disabilities led to significant movement away from provision of sheltered employment towards competitive supported employment in integrated community-based settings. Additionally, in 1997, federal Medicaid spending for supported employment was implemented, and the 1998 amendments to the Rehabilitation Act of 1973 authorized the provision of services to individuals with significant disabilities specifically with an emphasis on high-quality competitive employment which eliminated sheltered employment as a “successful rehabilitation” outcome.

3.2.1 Sample Frame

We begin with the administrative records of the Virginia DRS for the 11596 individuals who applied for VR services in 1988. Table 1 provides information about selection into our sample. As was the case with the DRS 2000 Cohort, the major cause of selection out of the sample is that the individual does not have a cognitive impairment. Approximately 82% of the original sample is excluded because of this restriction versus 78% for SFY 2000. However, a larger share of the original cohort remains in the SFY 1988 final sample (16.4%) than for SFY 2000 (9.8%). There are two reasons for this difference. Most importantly, unlike the DRS 2000 Cohort, we cannot observe DRS service episodes prior to SFY 1988. So we cannot delete observations to avoid left-censoring problems. (Later, in Section 5.2, we provide some evidence about the size of the bias caused by left-censoring.) Secondly, other selection criteria that have minor impacts on the final sample have less of an impact on the SFY 1988 Cohort (eliminating 1.7% of the total cohort) than they did for the SFY 2000 Cohort (eliminating 4.1% of the cohort). After selection, we have a sample size of 1907 DRS clients with cognitive impairments.

3.2.2 Service Provision

As with the 2000 applicant cohort, services are aggregated into 6 different groupings based on observed purchased services. For the 1988 cohort, however, we also observe similar benefits provision, which are almost entirely concentrated as diagnosis & evaluation and other services. Table 2 shows the proportion of clients receiving each service type in 1988 and 2000. Provision of diagnosis & evaluation and other services are cut in half, but this is due mostly to having information in 1988 on services provided as a similar benefit. There are smaller reductions of service for training and maintenance, and these reductions are actually in purchased services. There is no change in provision of restoration, and education services increase. With respect to the composition of services within these broad aggregates, the most important change is with training where, as noted previously, we observe a shift away from work adjustment training usually provided as a prelude to sheltered employment (21.7% of training services in the 1988 cohort to 13.5% in the 2000 cohort) and toward job coach training services
or supported employment (from 21.3% in 1988 to 32.1% in 2000). As was true in the 2000 cohort, many clients receive multiple services. Relative to the 2000 cohort, diagnosis & evaluation is a much more common service component in the 1988 cohort (because we have similar benefits data), and multiple service provision is more common.

3.2.3 Explanatory Variables

Table 4 provides information about the moments of explanatory variables in the 1988 cohort and how they differ from the 2000 cohort. The significant changes are in education which increases by 0.476 years of schooling, transportation availability (0.068), # dependents (0.086), the prevalence of mental illness (0.110), and the mix of disability severity which changed from providing services to more clients with moderate, not significant disabilities in 1988 to almost all clients with significant or most significant disabilities in 2000. This latter change is most likely in response to the federal RSA mandate to serve persons with more significant disabilities. In fact, starting in 2000, RSA performance standards established in the 1998 amendments to the Rehabilitation Act of 1973 required state agencies to serve certain minimum percentages of their caseloads with significant disabilities. Over the 1988 – 2000 period, there also was a national trend toward diagnosing people with a learning disability rather than with a cognitive impairment. Over this time period, in our sample, the prevalence of learning disability doubled from 2.0% to 4.6%. While there were not enough observations with learning disability in 1988 to use in estimation, there were in 2000.

3.3 VEC Data

Our next source of data is the administrative records of the Virginia Employment Commission (VEC). The VEC data provide information about individual quarterly earnings prior to, during, and after service receipt for the universe of applicants for DRS services in SFY 1988 and SFY 2000. The SFY 1988 Cohort includes a minimum of 12 quarters of pre-application employment history and a minimum of 60 quarters post-application history. The SFY 2000 Cohort includes more pre-application quarters (16) but fewer post-application quarters (38). Despite the longer coverage period for the 1988 Cohort, the percentage

\[ \text{The composition of services within each of the remaining aggregated service categories are relatively stable with a few exceptions. For diagnosis & evaluation, exams and visits were the dominant service (65\% of diagnosis & evaluation) in 1988, and vocational evaluations (32\%) and situation assessment/supported employment (39\%) are dominant in 2000. For education, vocational/business school constituted 70\% of services in 1988, and services were pretty evenly divided among college, secondary education, and vocational/business school in 2000. For restoration, psychological services is the dominant service in both years, and for maintenance, transportation support is the dominant service both years. Job guidance is the dominant service in both years for other services.} \]

\[ \text{These changes are also subject to the potential biases caused by including people with prior episodes of service in 1988. However, our intuition suggests that the bias in 1988 overstates the proportion of clients with significant and most significant disabilities.} \]
Table 4: Moments of Explanatory Variables for 1988 Cohort

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std Dev</td>
<td>Mean</td>
<td>Std Dev</td>
</tr>
<tr>
<td>Male</td>
<td>0.552</td>
<td>0.497</td>
<td>0.563</td>
<td>0.496</td>
</tr>
<tr>
<td>Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>0.563</td>
<td>0.496</td>
<td>0.563</td>
<td>0.496</td>
</tr>
<tr>
<td>Hearing/Speech Disability</td>
<td>0.023</td>
<td>0.150</td>
<td>0.023</td>
<td>0.150</td>
</tr>
<tr>
<td>Musculo/Skeletal Disability</td>
<td>0.070</td>
<td>0.256</td>
<td>0.070</td>
<td>0.256</td>
</tr>
<tr>
<td>Special Education</td>
<td>0.317</td>
<td>0.465</td>
<td>0.319</td>
<td>0.465</td>
</tr>
<tr>
<td>Mental Illness</td>
<td>0.073</td>
<td>0.261</td>
<td>0.073</td>
<td>0.261</td>
</tr>
<tr>
<td>Age (Quarters)/100</td>
<td>1.026</td>
<td>0.369</td>
<td>1.026</td>
<td>0.369</td>
</tr>
<tr>
<td>Married</td>
<td>0.060</td>
<td>0.237</td>
<td>0.060</td>
<td>0.237</td>
</tr>
<tr>
<td>Dependents</td>
<td>0.280</td>
<td>0.718</td>
<td>0.280</td>
<td>0.718</td>
</tr>
<tr>
<td>Transportation Available</td>
<td>0.392</td>
<td>0.488</td>
<td>0.392</td>
<td>0.488</td>
</tr>
<tr>
<td>Has Driving License</td>
<td>0.180</td>
<td>0.384</td>
<td>0.180</td>
<td>0.384</td>
</tr>
<tr>
<td>Receives Gov't Assistance</td>
<td>0.301</td>
<td>1.456</td>
<td>0.301</td>
<td>1.456</td>
</tr>
</tbody>
</table>

Notes:
1) Variables used in 2000 but not in 1988 include internal disability, learning disability, and other disability.
2) Variables used in 1988 but not in 2000 include disability level unknown.
3) Receives Gov't Assistance is a continuous measure in 1988 and a binary measure in 2000.

of individuals with at least one quarter of “covered” employment is similar at 86.5% for 1988 and 88.6% for 2000. The remaining individuals were, for the entire interval, either a) unemployed or out of the labor force and/or b) employed in jobs that are not covered by the VEC (e.g., were self-employed or worked out of state, for federal employers, for very small-sized firms, or at contingent-type jobs that do not provide benefits).14

The VEC data provide us with two labor market outcome variables: employment and log quarterly earnings. The employment variable is a binary measure of working in a particular quarter in the labor market, and it corresponds to $z_{it}$ associated with equation (2). The log quarterly earnings variable corresponds to $w_{it}$ in equation (3). Unfortunately, we cannot decompose quarterly earnings into wage level and hours.

Table 5 provides information on the moments of the two labor market variables for both cohorts. The log quarterly earnings moments are conditional on employment. First, note that the large number of quarterly observations allows us to estimate precisely a rich model of labor market outcomes even after allowing for time dependence. Next, note that, in both cohorts, mean employment rates and mean earnings increase from pre-service to post-service. Also, the increases are larger in the 2000 cohort than in the 1988 cohort. This may be due to real changes in the VR program, real changes in the composition of VR applicants, and/or changes in the sample selection criteria. We explore these possibilities in Section 5.2.

Figures 1 and 2 display the time series trends in the mean quarterly labor market outcomes – employment and real earnings conditional on employment – for the 2000 applicant cohort. Normalizing the application quarter to equal zero, the figure illustrates how the labor market outcomes vary before and after

14Dean et al. (2013a) provide evidence that indicates that there is a VEC “coverage gap” of about 12% compared to employment reported at the federal level in Social Security earnings data.
the application quarter in SFY 2000 between applicants receiving substantial VR services – the treated group – and those that did not receive substantive services, the untreated. Prior to the application quarter, employment rates and average quarterly earnings of the treated and untreated are nearly identical. For example, one year prior to the application quarter the employment rates for both groups are about 27% and average quarterly earnings are around $1290. Thus, there is little evidence of selection into VR service receipt based on prior labor market outcomes. Shortly after the application quarter, however, the labor market outcomes of the treated and not-treated begin to diverge. In particular, Figure 1 shows that the employment rates for the treated increase relative to the rates for the untreated, leading to an employment gap of about 7% that last for 20 quarters, while Figure 2 shows that average quarterly earnings of the treated fall about $500 below the earnings of those not treated. Thus, the data reveal that VR treatment services are associated with a notable and sustained increase in employment but a drop in quarterly earnings. Although interesting descriptive statistics, the results displayed in Figures 1 and 2 should not be interpreted as measuring the impact of VR services. These models do not control for any observed covariates, do not account for selection into VR service receipt, and do not allow for heterogeneity in the types of services.

3.4 BEA Data

Labor market outcomes may be influenced by local labor market conditions. Though there are no measures of local labor market conditions in either the DRS data or the VEC data, the DRS data contain geographic identifiers so that we can match each DRS client with their county of residence. The Bureau of Economic Analysis (BEA) provides information on population size and number of people employed, disaggregated by age and county (BEA, 2010). Using these data, we construct a measure of the county log employment rate (see Dean et al. (2013a) for further details).
4 Econometric Methodology

4.1 Likelihood Function

The parameters of the model are $\theta = (\theta_y, \theta_z, \theta_w)$ where

$$\theta_y = (\beta_j, \lambda_j^{y1}, \lambda_j^{y2})_{j=1}^J,$$
$$\theta_z = (\gamma, \lambda_j^z, \lambda_j^z, \rho, \sigma^2, \rho, [\alpha_j^z]_{j=1}^J),$$
$$\theta_w = (\delta, \lambda_j^w, \lambda_j^w, \rho, \sigma^2, [\alpha_j^w]_{j=1}^J).$$

We estimate the parameters of the model using maximum simulated likelihood (MSL). The likelihood contribution for observation $i$ is

$$L_i = \int L_i (u_i) dG (u_i | \Omega)$$

where

$$L_i (u_i) = L_i^y (u_i^y) \prod_{t=1}^T L_i^{zw} (u_{it}^z, u_{it}^w),$$

$$L_i^y (u_i^y) = \prod_{j=1}^J \frac{\exp \{ X_i^y \beta_j + u_{ij}^y \}}{1 + \exp \{ X_i^y \beta_j + u_{ij}^y \}}.$$
\[ L_{it}^{zw} (u_{it}, u_{it}^w) = [L_{it}^0 (u_{it}, u_{it})]^{1 - z_{it}} [L_{it}^1 (u_{it}, u_{it})]^{z_{it}}, \]  
\[ L_{it}^0 (u_{it}, u_{it}) = 1 - \Phi \left( X_{it}^z \gamma + \sum_{k=1}^{K} \sum_{j=1}^{J} \alpha_{jk} y_{ij} + u_{it}^z \right), \]  
\[ L_{it}^1 (u_{it}, u_{it}) = \frac{1}{\sigma_w} \Phi \left( \frac{w_{it} - X_{it}^w \delta - \sum_{k=1}^{K} \sum_{j=1}^{J} \alpha_{jk} y_{ij} - u_{it}^w}{\sigma_w} \right), \]

and \( G (u_i | \Omega) \) is the joint normal density with covariance matrix \( \Omega \) described in the online appendix. While, in general, it is difficult to evaluate the multivariate integral in equation (5), it is straightforward to simulate the integral using well-known methods described in Stern (1997). The functional form of the conditional likelihood contribution associated with observed VR service choices, \( L_{i}^{y} (u_{it}^y) \) in equation (6), follows from the assumption in equation (1) that the idiosyncratic errors are iid logit. The functional form of the conditional likelihood contribution for labor market outcomes, \( L_{it}^{zw} (u_{it}^z, u_{it}^w) \) in equations (7), (8), and (9), follow from the normality assumption for \( (z_{it}^z, w_{it}^w) \) and the bivariate normality assumption for \( (z_{it}^z, w_{it}^w) \) in equation (4). The log likelihood function is

\[ L = \sum_{i=1}^{n} \log L_i. \]

In theory, the parameter estimates are consistent only as the number of independent draws used to simulate the likelihood contributions goes off to infinity. However, Börsch-Supan and Hajivassiliou (1992) show that MSL estimates perform well for small and moderate numbers of draws as long as good simulation methods are used, and Geweke (1988) shows that the simulation error occurring in simulation-based estimators is of order \( (1/n) \) when antithetic acceleration is used.

### 4.2 Identification

We face two types of identification issues. The first is the typical issue of identification in non-linear models. As is usual, covariation in the data between dependent variables and explanatory variables identifies many of the model parameters. For example, the \( \beta \) terms in equation (1) are identified by covariation between explanatory variables in the equation and observed VR service choices. Similar arguments apply to the \( \gamma \) terms in equation (2) and the \( \delta \) terms in equation (3). For example, the \( \beta \) terms in equation (1) are identified by covariation between explanatory variables in the equation and observed VR service choices. Similar arguments apply to the \( \gamma \) terms in equation (2) and the \( \delta \) terms in equation (3).
equation (3). Second moment parameters such as $\sigma^2$ and $\rho_c$ in equation (4) are identified by corresponding second sample moments.

The second issue concerns identification problems caused by endogeneity of the VR service treatment choices. We address endogeneity issues in two ways. First, we control for pre-treatment labor market differences between those who do and do not receive services. (Meyer, 1995; Heckman, LaLonde, and Smith, 1999, Section 4). Second, we include two instruments for each binary service variable. The instruments for service $j$ are the propensity of an individual’s VR counselor to assign other clients to service $j$ and the propensity of an individual’s VR field office to assign other clients to service $j$. Doyle (2007), Arrighi et al. (2010), Dean et al. (2013a, 2013b, 2013c), Clapp et al. (2010), and Maestas, Mullen, and Strand (2013) use a similar instrument. We use a non-linear transformation of these variables, described in the online appendix, to improve their performance.

In order for these variables to be valid instruments, they must a) be correlated with service receipt, b) not belong directly in equations (2) and (3), and c) be exogenous. By construction, (a) is satisfied as long as there is enough independent variation in the two instruments (see Section 3.2.3). In fact, the correlation between the instruments for each service is approximately 70% and, in Section 3.2.3, we show that the instruments significantly influence service receipt. There is no reason to think that (b) is a problem. Finally, while condition (c) is more difficult to assess, we believe it is a credible assumption given that we are able to condition the analysis on the client’s observed limitations and county level unemployment rates. Importantly, DRS clients have limited ability to select their field office or counselor; the field office is determined by the residential location of the client, and counselors are randomly assigned conditional on the client’s observed limitations. So, unless clients relocate to take advantage of the practices of particular field offices, the assignment to offices and counselors is effectively random conditional on the observed limitations of clients. The assumption might be violated if counselors and/or field offices made service provision decisions based on idiosyncratic features of the local labor market. For example, a particular field office may not provide education services because there are very few local jobs that would benefit from improved education. Including measures of local labor market conditions directly in equations (2) and (3) should ameliorate this problem. It also could be a problem if, for example, high quality counselors both tend to choose particular services and are better than average at placing their clients in good jobs. However, despite these concerns, along with Aakvik, Heckman, and Vytlacil (2005) and Dean et al. (2013a, 2013c), this is one of the first studies to identify the impact of VR services on labor market outcomes using both a history of pre-program earnings and plausibly exogenous instrumental variables.
5 Estimation Results

In this section, we present estimation results for the 2000 cohort and compare these results to the findings from the 1988 cohort. We divide up the discussion into different parts to focus the discussion. First, we discuss estimates directly associated with the effect of purchased DRS services on labor market outcomes. Then we discuss estimates of the determinants of service choice followed by a discussion of the effect of other explanatory variables on labor market outcomes. Next, we present and interpret covariance term estimates. We finish with a discussion of how estimates from the 2000 cohort differ from estimates from the 1988 cohort.

5.1 Estimates for 2000 Cohort

5.1.1 Effect of Purchased DRS Services on Labor Market Outcomes

Tables 6 and 7 report estimates of the effects of purchased DRS services on employment propensity and log quarterly earnings, respectively. Service effects are allowed to vary across VR service types and across the different time periods. We specify 4 different time periods: 2 or more quarters before service, the immediate quarter before VR service, the 8 quarters after service (short run), and 9 or more quarters after service (long run). Consider the effect of training in Table 6. The first column for training implies that, holding all else constant, training lowers pre-service employment propensity by 0.330. This result should be interpreted as a selection effect; i.e., individuals with unusually low employment propensity are more likely to receive training. The third and fourth columns imply that, holding all else constant, training increases employment propensity by 0.364 in the short run and 0.128 in the long run. Thus, our estimates of the short- and long-run effects of training on employment propensity are 0.364 + 0.330 = 0.694 and 0.128 + 0.330 = 0.458, respectively. Employment and earnings are known to drop in the periods just before individuals apply to many labor market training programs. To account for this Ashenfelter dip (Ashenfelter, 1978; Heckman, Lalonde and Smith, 1999), we explicitly allow service effects to vary in the immediate quarter before service. For the most part, these results are similar to the pre-service estimates displayed in column 1, yet are statistically insignificant.

Overall, we find that all services other than restoration and maintenance result in both short- and long-run increases in employment propensity. The variation in effects over services and over time points out the value of modelling and estimating a richer model of service provision than the usual binary service provision model.

Table 7 provides similar estimates for the effects of each purchased service on log quarterly earnings. For example, the short- and long-run change in log quar-

\footnote{All F-statistics testing for the joint significance of the short-term and long-term employment effects and log quarterly earnings effects relative to the effect prior to program participation are statistically significant with p-values less than 0.0001.}
<table>
<thead>
<tr>
<th>Variable</th>
<th>Diagnosis &amp; Evaluation</th>
<th>Training</th>
<th>Education</th>
<th>Restoration</th>
<th>Maintenance</th>
<th>Other Services</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prior to Service</td>
<td>Quarter</td>
<td>First 2</td>
<td>More than 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participation</td>
<td>Prior to</td>
<td>Years After</td>
<td>Years After</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Service</td>
<td>Service</td>
<td>Service</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Participation</td>
<td>Participation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosis &amp; Evaluation</td>
<td>-0.056 **</td>
<td>-0.033</td>
<td>0.366 **</td>
<td>0.254 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.018)</td>
<td>(0.147)</td>
<td>(0.026)</td>
<td>(0.014)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>-0.330 **</td>
<td>-0.141</td>
<td>0.364 **</td>
<td>0.128 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.018)</td>
<td>(0.155)</td>
<td>(0.029)</td>
<td>(0.014)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.633 **</td>
<td>0.693</td>
<td>0.755 **</td>
<td>1.171 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.069)</td>
<td>(0.549)</td>
<td>(0.097)</td>
<td>(0.040)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restoration</td>
<td>-0.223 **</td>
<td>-0.173</td>
<td>-0.554 **</td>
<td>-0.681 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.020)</td>
<td>(0.161)</td>
<td>(0.031)</td>
<td>(0.016)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintenance</td>
<td>-0.079 **</td>
<td>-0.035</td>
<td>-0.203 **</td>
<td>-0.138 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.019)</td>
<td>(0.162)</td>
<td>(0.030)</td>
<td>(0.015)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Services</td>
<td>-0.117 **</td>
<td>0.200</td>
<td>0.266 **</td>
<td>0.302 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.019)</td>
<td>(0.163)</td>
<td>(0.028)</td>
<td>(0.015)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
1. Standard errors are in parentheses.
2. Single-starred items are statistically significant at the 10% level, and double-starred items are statistically significant at the 5% level.

Weekly earnings from training are $0.331 - 0.122 = 0.219$ and $0.333 - 0.046 = 0.287$, respectively. Table 7 shows positive effects on log quarterly earnings for all services other than restoration and maintenance. Dean and Dolan (1991) also find evidence of positive earnings effects in their earlier evaluation of VR services, although in some cases, especially for men, the results are not statistically significant. After using an instrumental variable to address the selection problem, Aakvik et al. (2005) find no evidence of employment effects of VR services in Norway.

Tables 6 and 7 provide information on the effects of purchased services on the dependent variables in equations (2) and (3). While employment propensity and log quarterly earnings were appropriate dependent variables for estimation, the variables with policy interest are employment probability and quarterly earnings, each a nonlinear function of its corresponding dependent variable. Following Dean et al. (2013a), Figure 3 uses the employment propensity effects from Table 6 and the log quarterly earnings effects from Table 7 to compute the average marginal effect of each service type on labor market outcomes. This figure shows positive short- and long-run improvements in both employment probabilities and quarterly earnings for all services except restoration and maintenance. This figure reveals that both the short- and long-run mean labor market effects are estimated to be positive for diagnosis & evaluation, training, education and other services, but negative for restoration and maintenance.

Figure 3 has two limitations. First, it shows only mean effects while the model implies a distribution of effects caused by interactions between nonlinearity associated with equations (2) and (3) and variation in explanatory variables in the two equations. Second, it provides information separately on short- and long-run effects but provides no information on long-term discounted benefits. Figure 4 provides information about the median discounted benefits of
### Table 7: DRS Purchased Service Participation Effects on Log Quarterly Earnings

<table>
<thead>
<tr>
<th>Variable</th>
<th>Prior to Service Participation</th>
<th>Quarter Prior to Service Participation</th>
<th>First 2 Years After Service Participation</th>
<th>More than 2 Years After Service Participation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis &amp; Evaluation</td>
<td>-0.157 **</td>
<td>0.143 **</td>
<td>-0.150 **</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.037)</td>
<td>(0.016)</td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>-0.331 **</td>
<td>-0.122 **</td>
<td>-0.046 **</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.039)</td>
<td>(0.019)</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.296 **</td>
<td>0.389 **</td>
<td>0.851 **</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.103)</td>
<td>(0.113)</td>
<td>(0.058)</td>
<td></td>
</tr>
<tr>
<td>Restoration</td>
<td>-0.160 **</td>
<td>-0.468 **</td>
<td>-0.401 **</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.041)</td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>Maintenance</td>
<td>-0.051</td>
<td>-0.302 **</td>
<td>-0.120 **</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.042)</td>
<td>(0.019)</td>
<td></td>
</tr>
<tr>
<td>Other Services</td>
<td>0.063 *</td>
<td>0.371 **</td>
<td>0.287 **</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.042)</td>
<td>(0.019)</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
1. Estimates are effects on log quarterly earnings conditional on employment.
2. Standard errors are in parentheses.
3. Single-starred items are statistically significant at the 10% level, and double-starred items are statistically significant at the 5% level.

![Figure 3: DRS Purchased Service Effects on Labor Market Outcomes](image_url)
each service and a 95% quantile range for the benefits of each service measured in $1000.\textsuperscript{17} Consistent with Figure 3, all services have positive present value except restoration and maintenance. Median long-run discounted benefits are around $10000 for training, diagnosis & evaluation, and other services, and they are $36000 for education. The figure displays significant skewness in returns; in general, for those services with positive median returns, the upper bound of the 95% quantile range is approximately 20 times further away from the median than is the lower bound; and, for those services with negative median returns the lower bound of the quantile range is approximately 20 times further away from the median than is the upper bound. The skewness is caused by the convexity of the exponential function associated with transforming \( w_{it} \) in equation (3) into quarterly earnings. Note also that, for each service, all discounted benefits are on the same side of zero; this follows from the additively separable effect of services in equations (2) and (3). The wide variation in benefits points to the importance of allowing other exogenous variables to affect labor market outcomes when measuring the direct effect of the treatment.

These basic findings on the labor market effects of VR are similar to the results reported in other evaluations of Virginia’s program. In particular, using a similar model to study VR clients with mental illnesses, Dean et. al. (2013a) find large positive employment and earnings effects for most service types except for diagnosis & evaluation. For people with cognitive impairments, we find no such negative effect as the diagnosis is more straightforward than for mental illness and such individuals with developmental disabilities have likely been extensively evaluated through special education programs prior to applying for VR services. Dean and Dolan (1991) also find evidence of positive earnings effects in their earlier evaluation of VR services, although in some cases, especially for

\textsuperscript{17}In Figure 4, we arbitrarily use a 0.95 quarterly discount factor and a 10-year time horizon. Later, in Section 6, we present information on the distribution of rates of return.
men, the results are not statistically significant. Evaluations of European programs that use similar methodologies to address selection into treatment find more muted effects of VR programs. For example, Aakvik et al. (2005), who use a latent variable selection model with unobserved factor loadings and an instrumental variable, find no evidence of employment effects of VR services in Norway. Likewise, using propensity score matching methods to evaluate VR programs in Sweden, Frolich, Heshmati, and Lechner (2004) find heterogeneity in the impact of different services but conclude that VR services do not increase the likelihood of employment.

5.1.2 Effect of Explanatory Variables on Purchased DRS Service Receipt

The model in Section 2 implies a latent service value \( y_{ij} \) that depends on the explanatory variables in Table 3 and the instruments discussed below Table 3. Most of the explanatory variables from Table 3 have no statistically significant effect on service provision. Exceptions include the positive effects of receives government assistance on diagnosis & evaluation, training, and maintenance, the negative effect of musculo/skeletal disability on training, and the positive effects of age, # dependents, has driving license and musculo/skeletal disability on restoration.\(^{18}\)

On the other hand, the instruments are very influential in explaining service provision. We allow counselor and field office effects to enter into equation (1) in two ways. First, the transformed instruments for counselor and field office effects described in the on-line Appendix directly enter into equation (1). We impose the restriction that the coefficients on counselor effects and field office effects do not vary across services. Second, as discussed in an online appendix, because of the existence of counselors and field offices with very few clients, we include dummy variables for those cases where counselor and/or field office effects can not be computed. There is too much multicollinearity to include all of the missing value variables, so we include only missing counselor effects. Table 8 displays estimates for counselor and field office effects and missing value effects. While the field office effect is statistically insignificant, the counselor effect is statistically significant and explains much of the variation in service provision. We are already controlling for a large set of person-specific observed variables, so it is unlikely that the counselor effects are capturing correlation of characteristics with a counselor caseload. In general, the missing values variables are statistically significant and negative, suggesting that counselors with small caseloads are hesitant to provide services.

5.1.3 Effect of Explanatory Variables on Labor Market Outcomes

The model in Section 2 implies labor market outcome specifications depending on service provision (see Tables 6 and 7), the explanatory variables listed in Table 3, and the local labor market condition described in Section 3.4. Table

\(^{18}\)A full set of estimates and standard errors is presented in the online appendix.
Table 8: Counselor and Office Effects on Service Receipt

<table>
<thead>
<tr>
<th>Instruments</th>
<th>Estimate</th>
<th>Std Err</th>
<th>Missing Counselor Effects</th>
<th>Estimate</th>
<th>Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counselor Effect</td>
<td>0.451 **</td>
<td>0.128</td>
<td>Diagnosis &amp; Evaluation</td>
<td>0.309</td>
<td>0.332</td>
</tr>
<tr>
<td>Office Effect</td>
<td>-0.403</td>
<td>0.319</td>
<td>Training</td>
<td>-3.810 *</td>
<td>2.241</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Education</td>
<td>-1.413 **</td>
<td>0.403</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Restoration</td>
<td>-0.751 **</td>
<td>0.334</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Maintenance</td>
<td>-0.810 **</td>
<td>0.332</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other Services</td>
<td>0.876 **</td>
<td>0.092</td>
</tr>
</tbody>
</table>

Notes:
1. Single-starred items are statistically significant at the 10% level, and double-starred items are statistically significant at the 5% level.
2. Other than those reported, missing counselor and field office effects parameters were excluded because of multicollinearity problems.

9 provides estimates for the explanatory variables excluding service provision. There are two estimated equations, one for employment propensity and one for log quarterly earnings. Because of the large sample sizes, almost all of the estimates are statistically significant. Some of the estimates for demographic variables have expected signs such as the positive effects of male, education, age, transportation available, and has driving license on both outcomes, and the positive effect of white on employment propensity. On the other hand, counterintuitive estimates include the negative effect of white on log quarterly earnings and the positive effect of special education on both outcomes. For disability measures, internal disability, learning disability, mental illness, and other disability have negative effects on both outcomes, and musculoskeletal disability has a negative effect on log quarterly earnings. Only hearing/speech disability has a statistically significant positive effect on either labor market outcome (log quarterly earnings). The disability severity measures work as expected: the more significant the disability, the larger the negative effect it has on both labor market outcomes. The measure of local labor market employment rate has a counterintuitive sign. In fact, in Dean et al. (2013a, 2013b, 2013c), we have consistently found no meaningful effect for local labor market conditions.

5.1.4 Covariance Terms

In general, we allow for covariation among unobservables in two ways in the model. Equation (4) specifies the existence of two latent factors affecting service provision, employment propensity, and log quarterly earnings. Table 10 presents the estimates of the factor loadings for each latent factor. Note that both factors have similar factor loading structures; i.e., with one exception, factor loadings for service provision are statistically insignificant, and the factor loadings for the two labor market outcomes have opposite signs. This suggests two factors capturing a measure of desire to work. The larger the factors, the lower the individual’s reservation wage, causing employment probability to rise and quarterly earnings to fall. The fact that essentially none of the service pro-
### Table 9: Labor Market Effects

<table>
<thead>
<tr>
<th>Variable</th>
<th>Employment Estimate</th>
<th>Employment Std Err</th>
<th>Log Quarterly Earnings Estimate</th>
<th>Log Quarterly Earnings Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.863 **</td>
<td>0.047</td>
<td>4.748 **</td>
<td>0.080</td>
</tr>
<tr>
<td>Male</td>
<td>0.212 **</td>
<td>0.010</td>
<td>0.122 **</td>
<td>0.012</td>
</tr>
<tr>
<td>White</td>
<td>0.005 **</td>
<td>0.010</td>
<td>-0.083 **</td>
<td>0.011</td>
</tr>
<tr>
<td>Education</td>
<td>0.041 **</td>
<td>0.002</td>
<td>0.072 **</td>
<td>0.003</td>
</tr>
<tr>
<td>Special Education</td>
<td>0.547 **</td>
<td>0.027</td>
<td>0.852 **</td>
<td>0.039</td>
</tr>
<tr>
<td>Education Missing</td>
<td>-0.710 **</td>
<td>0.041</td>
<td>-0.326 **</td>
<td>0.044</td>
</tr>
<tr>
<td>Age/100</td>
<td>1.158 **</td>
<td>0.015</td>
<td>1.045 **</td>
<td>0.019</td>
</tr>
<tr>
<td>Married</td>
<td>-0.922 **</td>
<td>0.033</td>
<td>-0.494 **</td>
<td>0.030</td>
</tr>
<tr>
<td># Dependents</td>
<td>-0.010 *</td>
<td>0.006</td>
<td>-0.090 **</td>
<td>0.007</td>
</tr>
<tr>
<td>Transportation Available</td>
<td>0.335 **</td>
<td>0.010</td>
<td>0.463 **</td>
<td>0.012</td>
</tr>
<tr>
<td>Has Driving License</td>
<td>0.612 **</td>
<td>0.013</td>
<td>0.562 **</td>
<td>0.015</td>
</tr>
<tr>
<td>Receives Gov Assistance</td>
<td>0.152 **</td>
<td>0.021</td>
<td>-0.359 **</td>
<td>0.026</td>
</tr>
<tr>
<td>Hearing/Speech Disability</td>
<td>0.081</td>
<td>0.057</td>
<td>0.304 **</td>
<td>0.048</td>
</tr>
<tr>
<td>Musculoskeletal Disability</td>
<td>0.003</td>
<td>0.022</td>
<td>-0.202 **</td>
<td>0.024</td>
</tr>
<tr>
<td>Internal Disability</td>
<td>-0.516 **</td>
<td>0.023</td>
<td>-0.764 **</td>
<td>0.025</td>
</tr>
<tr>
<td>Learning Disability</td>
<td>-0.044 **</td>
<td>0.019</td>
<td>-0.218 **</td>
<td>0.019</td>
</tr>
<tr>
<td>Mental Illness</td>
<td>-0.513 **</td>
<td>0.014</td>
<td>-0.498 **</td>
<td>0.015</td>
</tr>
<tr>
<td>Other Disability</td>
<td>-0.913 **</td>
<td>0.052</td>
<td>-1.136 **</td>
<td>0.040</td>
</tr>
<tr>
<td>Disability Significant</td>
<td>-0.300 **</td>
<td>0.034</td>
<td>-0.198 **</td>
<td>0.060</td>
</tr>
<tr>
<td>Disability Most Significant</td>
<td>-0.571 **</td>
<td>0.025</td>
<td>-0.340 **</td>
<td>0.061</td>
</tr>
<tr>
<td>Local Employment Rate</td>
<td>-0.003</td>
<td>0.103</td>
<td>-0.421 **</td>
<td>0.136</td>
</tr>
</tbody>
</table>

Note: Single-starred items are statistically significant at the 10% level, and double-starred items are statistically significant at the 5% level.

### Table 10: Covariance Factor Loadings

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor 1 Estimate</th>
<th>Factor 1 Std Err</th>
<th>Factor 2 Estimate</th>
<th>Factor 2 Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis &amp; Evaluation</td>
<td>1.089 **</td>
<td>0.339</td>
<td>-0.399</td>
<td>0.356</td>
</tr>
<tr>
<td>Training</td>
<td>-0.592 *</td>
<td>0.340</td>
<td>0.125</td>
<td>0.153</td>
</tr>
<tr>
<td>Education</td>
<td>0.274</td>
<td>0.582</td>
<td>-0.098</td>
<td>0.126</td>
</tr>
<tr>
<td>Restoration</td>
<td>-0.170</td>
<td>0.580</td>
<td>-0.100</td>
<td>0.127</td>
</tr>
<tr>
<td>Maintenance</td>
<td>-0.096</td>
<td>0.112</td>
<td>-0.038</td>
<td>0.100</td>
</tr>
<tr>
<td>Other Services</td>
<td>-0.214 *</td>
<td>0.120</td>
<td>-2.405</td>
<td></td>
</tr>
<tr>
<td>Employment</td>
<td>0.708 *</td>
<td>0.007</td>
<td>0.085 **</td>
<td>0.008</td>
</tr>
<tr>
<td>Log Quarterly Earnings</td>
<td>-1.276 **</td>
<td>0.004</td>
<td>-1.362 **</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Notes:
1. Single-starred items are statistically significant at the 10% level, and double-starred items are statistically significant at the 5% level.
2. The element in Factor 2 for other services is restricted to ensure that the two factors are orthogonal with respect to the six factor loadings associated with the services available.

Vision factor loadings are statistically significant suggests that endogeneity of service provision may not be a large issue. Equation (4) introduces a few more covariance terms, and their estimates are displayed in Table 11. The correlation terms $\rho_\eta$ and $\rho_\xi$, allowing for serial correlation in labor market outcomes and contemporaneous correlation between the two labor market outcomes, are both statistically significant and positive. Note that the serial correlation estimate ($\rho_\eta = 0.977$) is similar in magnitude to estimates in the literature for non-disabled populations (e.g., Macurdy, 1982; Abowd and Card, 1989; Topel, 1991; and Meghir and Pistaferri, 2004). The positive contemporaneous correlation estimate ($\rho_\xi = 0.727$) suggests the existence of unobserved ability affecting employment and earnings in the same direction.
Table 11: Other Covariance Terms

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std Err</th>
<th>Variable</th>
<th>Estimate</th>
<th>Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_1$</td>
<td>0.977 **</td>
<td>0.000</td>
<td>$\rho_1$</td>
<td>0.727 **</td>
<td>0.006</td>
</tr>
<tr>
<td>$\sigma_\zeta$</td>
<td>0.163 **</td>
<td>0.00010a</td>
<td>$\sigma_w$</td>
<td>1.096 **</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Notes:
1. Double-starred items are statistically significant at the 5% level.
2. Correlation terms are estimated using the transformation,
   \[ \rho = \frac{2 \exp(\zeta)}{1 + \exp(\zeta)} - 1 \]
   where $\zeta$ is estimated to ensure that $-1 < \rho < 1$. Standard deviations are estimated using the transformation, $\sigma = \exp(\zeta)$ where $\zeta$ is estimated to ensure that $\sigma > 0$. Standard errors for both are derived using the delta method.

5.1.5 Specification Tests

Goodness-of-fit tests for service provision and employment, and Lagrange Multiplier tests on a number of potential extensions of the model are presented in an online appendix. Overall, the estimated model fits service provision well but over-predicts the observed variation in employment probabilities. This apparent misspecification seems to reflect, in part, the fact that there are no interesting dynamics in the model. In terms of the specification of the model, the LM tests reveal no evidence of missing interactions among the covariates or the provision of different services, or that the length of service provision affects labor market outcomes.

5.2 Estimates of Change from 1988 to 2000

As noted above, there have been important changes to the mix of clients and service provisions over time. Recall, the 1998 amendments to the Rehabilitation Act of 1973 mandated that state agencies serve a more significantly disabled clientele. Also, state agencies could no longer take credit for closures into sheltered workshops, i.e. extended employment services; rather, they altered their service packages to include supported employment provided in an integrated community setting. To assess how these changes impact the evaluation of DRS services, we re-estimate the models using data from the 1988 cohort. We have 1988 cohort data analogous in structure to the 2000 cohort data that allow us to perform such an exercise. However, an important concern is that, while for the 2000 cohort data, we are able to control for left-censoring (see Section 3.1.1) by excluding observations with service episodes prior to 2000, the 1988 cohort data does not provide us with the information necessary to use the same exclusion criterion. In the interest of understanding (and possibly controlling for) the bias introduced by left-censoring, we estimate the model with three samples:

a) the 1988 cohort including left-censored observations ($N = 1907$),

b) the 2000 cohort including left-censored observations ($N = 1788$), and

c) the 2000 cohort excluding left-censored observations($N = 1009$).

The difference in estimates associated with samples (b) and (c) provides us with some information about the direction and magnitude of bias caused by left-censoring and, to the extent that this bias is stable over time, the difference
in estimates associated with samples (a) and (b) provides us with information about the change in parameters over time.19

Figure 5 provides information about point estimates associated with short- and long-run service effects for each service and for each sample. Each bar represents the point estimate of the difference between either a short- or long-run effect minus the corresponding pre-service effect.20 For example, the (relevant) estimates for the effect of training on employment for sample (a) are $-0.162$ in the pre-service period and $-0.010$ in the first two years after service; thus the estimated short-run effect is $0.162 - 0.010 = 0.152$, the height of the first bar among the training bars, which is for the short-run effect using sample (a). The pattern for short-run employment effects for training is that the estimates for samples (b) and (c) are very close, and both are significantly larger than for sample (a). This strongly suggests that, for the short-run employment effect of training, left-censoring bias is not very large, and the effectiveness of training, with respect to its effect on employment, increased significantly between 1988 and 2000. The same pattern holds true for the long-run effect of training on employment. On the other hand, for the long-run effect of education on employment, the estimates for samples (a) and (b) are close and significantly less than for (c), suggesting that, for this case, left-censoring bias explains almost all of the variation in the change in the effect. In terms of the other services, Figure 5 shows that diagnosis & evaluation increases employment in both cohorts. The effect is smaller in the 2000 cohort, and left-censoring bias hides some of the loss between the two years. For education, all effects are positive in both cohorts. In the short-run, the positive effect declined, and, in the long-run, it increases but probably only because of left-censoring bias. Restoration has negative effects on employment in both cohorts, and they seem to be declining with time. The effects for maintenance are generally small. Finally, other services has positive effects in both cohorts, and the apparent improvement over time probably is due to left-censoring bias. In general, left-censoring biases appears to play a more substantial role in the parameters associated with less frequently used services such as education than the more commonly utilized services such as training.

Figure 6 provides analogous information for the effects of services on log quarterly earnings. For this case, diagnosis & evaluation, training, and education have uniformly positive effects. The improvement over cohorts for long-run education probably are due to left-censoring bias, while the improvement over cohorts for long-run training and the decline over cohorts for short-run education

19Changes in the estimates for diagnosis & evaluation and other services may partially reflect differences in the way service receipt is measured in the two cohorts. Recall that similar benefits provisions are included in the service receipt indicators of the 1988 cohort but are not observed in 2000 (see Section 3.2.2). As a result, the estimates in Table 2 show that much larger fractions of the 1988 caseload are classified as receiving diagnosis & evaluation (89.6% versus 44.7%) and other services (46.2% versus 24.9%). However, it is not obvious how the change would bias estimates.

20The only estimates that are not statistically significant at the 5% level are long-run maintenance on employment and long-run training on log quarterly earnings in sample (a); short-run education on employment in sample (b); and short-run education on employment and log quarterly earnings and long-run maintenance on log quarterly earnings in sample (c).
Diagnosis & Evaluation; Training; Education; Restoration; Maintenance; Other Services

Estimated Employment Effects Across Samples

1988 Incl, Short-Run
2000 Incl, Short-Run
2000 Excl, Short-Run
1988 Incl, Long-Run
2000 Incl, Long-Run
2000 Excl, Long-Run

Figure 5: Estimated Employment Effects Across Samples

Estimated log Earnings Effects Across Samples

1988 Incl, Short-Run
2000 Incl, Short-Run
2000 Excl, Short-Run
1988 Incl, Long-Run
2000 Incl, Long-Run
2000, Excl, Long-Run

Figure 6: Estimated log Earnings Effects Across Samples

and long-run diagnosis & evaluation are real. Similar to employment effects, restoration generally has negative effects with unclear changes over cohorts. Maintenance and other service effects are difficult to interpret.  

To illustrate the differences in VR service efficacy across the two cohorts, it is useful to assume that left-censoring bias is stable over cohorts (but still possibly varying over specific parameters). Given the changes to the VR program and clientele over these two periods (described above), this assumption may be somewhat restrictive. Still, it is worthwhile to perform this exercise since, under this assumption, the difference in estimates between samples (a) and (b) identifies the real change in service effectiveness over cohorts. The results are displayed in Figure 7. The estimates imply that training became more effective in both the short- and long-run between 1988 and 2000. For example, the first

21 Using a Hausman test statistic, we reject the null hypothesis that left-censoring bias does not exist. See the on-line Appendix for details.
bar associated with training has a height of 0.574, implying that, after controlling for left-censoring bias, the estimated change in the short-run employment effect of training is 0.574. To some degree, the training estimates may be due to the reorientation of training from sheltered employment services to supported employment services, resulting in a switch from sub-minimum wage payments (which may not have been covered in VEC records) to more competitive employment with payments greater than minimum wage. Meanwhile, diagnosis & evaluation and education became less effective for both employment and earnings. As with training, these differences may partially reflect two factors: first, there were substantial changes in the composition of services provided within the broad aggregated categories on diagnosis & evaluation and education (see Section 3.2.2, footnote 12); and second, 1988 service receipt indicators include similar benefits provisions while these provisions are not observed in 2000 (see Table 2 and footnote 19). Restoration became less effective for employment but more effective for earnings. Overall, these results suggest that there are important differences in the efficacy of the VR services across the two cohorts.

6 Rate of Return

Figure 4 provides information about the distribution of long-run discounted benefit of each of the VR services. In this section, we add cost to the analysis and measure rates of return. As is true for Figure 4, we present information about the distribution of rates of return especially because there is wide variation across DRS clients caused by variation in characteristics affecting labor market outcomes. We focus this analysis on the 2000 cohort.

In general, we can think of the net benefit of service provision as the long-run discounted benefits minus costs. We simulate the private labor market benefits to DRS clients using the structural model estimates summarized in
Section 5.\textsuperscript{22} In particular, we compute the present discounted expected value of the provided services relative to receiving no services using both a 5- and 10-year post-treatment observation period for each individual who received some service. The estimated mean discounted benefits across DRS clients are $10979 with a standard deviation of $17987 using the 5-year window and $21175 with a standard deviation of $35684 using a 10-year window. Thus, the mean long-run discounted benefits are approximately twice the mean short-run benefits, reflecting the effect of adding an extra 5 years and the fact that the long-run (over 2 years) labor market effects generally are estimated to be much larger than the short-run effects and are assumed for the purpose of this analysis to last throughout the 10-year window.

One should note that private benefits might differ from social benefits in a number of ways. First, to the degree that DRS services are helpful in providing a client with a job, they may be preventing another individual from getting that job. Training programs for low skilled workers, however, have not been found to result in this type of labor market displacement (LaLonde, 1995). Second, there may be some social benefits associated with DRS service. Some, such as increased tax revenue and reduced welfare payments, have no real social benefit because any improvement for the government is a loss of equal size to the DRS client. However others, such as reduction in deadweight loss, reduced administrative cost of welfare program participation, and increased client happiness associated with having a job, have true social value but are much more difficult to measure. Thus we ignore them in this analysis and note the downward bias in estimated rates of return associated with this decision (see LaLonde, 1995).

Next, we focus on cost estimation. Information about purchased service expenditures is summarized in Table 12. Mean expenditures conditional on service receipt for \textit{training} ($2855) account for just over 70\% of the total average cost of purchased services. Note that the average cost for \textit{training} is substantially less than the median long-term discounted marginal benefit\textsuperscript{23} of $9909 (see Figure 4). Overall, the mean conditional costs of purchased services among all 1009 clients (33\% of whom receive no purchased services) equals $2418 with a standard deviation of $2933. These mean conditional cost estimates have not been discounted, and thus will be inflated to the extent the purchased services are provided over long periods.

The cost estimates for purchased services from Table 12 grossly underestimate total cost because they do not include a cost for DRS-provided services, similar benefits, and administrative costs. To estimate these costs, we use information on DRS spending by fiscal year as reported to the RSA. These reports provide information on aggregate administrative costs, DRS-provided counseling, guidance, and placement service costs, purchased service cost, and size of the caseload for each fiscal year. Unfortunately, no information is provided on the costs associated with similar benefits. While there is some variation in the

\textsuperscript{22}This simulation has a similar structure to the one used to compute marginal effects in Section 5 (see Figure 4). But here we compute the present discounted value of the actual set of treatments provided by DRS rather than a conjectured treatment for single service, \( j \).

\textsuperscript{23}Median benefits are being used because a few large outliers significantly increase the mean.
### Table 12: Conditional Moments of Expenditure Data on Purchased VR Services for the 2000 Applicant Cohort

<table>
<thead>
<tr>
<th>Service</th>
<th>Mean Conditional Expenditure</th>
<th>Std Dev</th>
<th>% with Positive Expenditures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis &amp; Evaluation</td>
<td>$747</td>
<td>$756</td>
<td>36%</td>
</tr>
<tr>
<td>Training</td>
<td>$2,855</td>
<td>$2,776</td>
<td>41%</td>
</tr>
<tr>
<td>Education</td>
<td>$733</td>
<td>$752</td>
<td>2%</td>
</tr>
<tr>
<td>Restoration</td>
<td>$361</td>
<td>$770</td>
<td>21%</td>
</tr>
<tr>
<td>Maintenance</td>
<td>$441</td>
<td>$1,149</td>
<td>25%</td>
</tr>
<tr>
<td>Other Services</td>
<td>$226</td>
<td>$210</td>
<td>3%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$2,418</strong></td>
<td><strong>$2,933</strong></td>
<td>67%</td>
</tr>
</tbody>
</table>

Notes: Moments do not include cost of in-house services or similar benefits.

The distribution of costs across years, in general, non-purchased service and administrative costs account for 55% of total expenditures, reflecting an average cost per client of roughly $200 per month.

We use two methods to estimate non-purchased service costs:

1. Focus on the fact that purchased services account for 45% of total VR costs. Since purchased service costs for our sample average $2418 per client, fixed costs are estimated to be $2000 (\( \approx 0.67 \times (2418/0.45 - 2418) \)) per client; or

2. Focus on the fact that the average costs of administration and non-purchased services is $200 per client-month. Since the average service episode length is 7 quarters, costs are estimated to be $4200 (200 \times 7 \times 3) per client.

There are some potential problems with both of these methods. First, it may be the case that service costs for individuals with cognitive impairment differ from those for individuals with other disabilities (e.g., compare our estimates with those for clients with mental illness in Dean et al., 2013a). If, for example, cases of individuals with cognitive impairments have low average purchased services relative to non-purchased service costs, the first approach would be downward-biased. If instead, such cases have relatively low average costs associated with administration or non-purchased services, the second approach would be upward-biased. Second, we are not allowing for any variation in cost across individuals. We could use actual expenditure data from our administrative DRS data, or we could estimate an expenditure equation using that data. In either case, we would be able to include heterogeneity in cost into our analysis. We choose not to do so because, in the model, there is no heterogeneity in service provision other than that captured by the six service types. We think it is important for the model of service expenditure to be consistent with the model of service provision.

Comparing estimated costs and benefits reveals that DRS services provided to people with cognitive impairments have substantial positive returns, both in the short run and long run. Mean benefits range from $10979 for the short run.
to $21175 for the long run, while mean costs range from $3632 to $5832. Even under the most conservative assumptions about non-purchased services costs, the long-run mean benefit is estimated to exceed cost by a factor of 3.6.

Figure 8 provides information about the distribution of rate of return under a number of different assumptions. For each sample individual receiving some service, we compare the expected flow of benefits he would get with the service package he received relative to the flow of benefits with no services. We approximate cost as

\[ f + \sum_{j=1}^{J} y_{ij} c_j \]

where \( f \) is a combination of administrative costs and average (unobserved) in-house service and similar benefits costs, \( y_{ij} \) is an indicator for receipt of service \( j \) by person \( i \) (as defined in equation 1), and \( c_j \) is the average cost associated with service \( j \) computed as the ratio of “mean expenditure” and “% with positive expenditure” in Table 12.

Figure 8 shows the distribution of quarterly rates of return under four scenarios: two with \( f = $2000 \) and two with \( f = $4200 \); and, for each assumption about \( f \), we consider a 10-year horizon and a 5-year horizon. The figure clearly illustrates the importance of accounting fixed (non-purchased service) costs and for earnings flows in years 6 through 10, at least for conventional rates of return. Thus, it is important to use long panels of earnings data such as ours when estimating rates of return.24 Focusing on the distribution curves associated with a 10-year horizon, one sees that 21.3% of clients with cognitive impairments have negative rates of return if \( f = $2000 \) and 27.7% have negative rates of return if \( f = $4200 \) (i.e., there is no positive discount rate that will justify the cost of services relative to the flow of future benefits). At the same time, even if \( f = $4200 \), the median rate of return is quite high at 4.6% quarterly (19.7% annually), and 20% of rates of return are above 12.9% quarterly (62.5% annually); if \( f = $2000 \), the median rate of return is 7.7% quarterly (34.5% annually), and 20% of rates of return are above 19.2% quarterly (101.9% annually). The proportion with negative returns increases significantly when focusing on the distribution curves associated with a 5-year horizon. In fact, if \( f = $4200 \), nearly half the DRS clients are estimated to have negative rates of return.25

24 Estimated rates of returns for non-VR government training programs aimed at economically disadvantaged people also tend to be sensitive to short versus long horizons, and vary widely across programs, demographics, and studies. In some cases, these training programs are found to have average rates of return that are negative. But, in many others, the average annual rates of return are in excess of 100% (Friedlander, Greenberg, and Robins, 1997; and LaLonde, 1995).

25 It should be noted that the variation in rates of return here are due solely to variation in observable characteristics of individuals and variation in the set of services they receive; it is not due to randomness inherent in labor market experience.
7 Conclusions

Over the last few years, a number of state-level return-on-investment evaluations of VR services produced by economic consulting firms or university research bureaus (e.g., Hollenbeck and Huang, 2006; Kisker et al., 2008; and Wilhelm and Robinson, 2010) have compared outcomes of a “treated” and “untreated” group, as we do in Figures 1 and 2. These studies tend to find large positive returns to VR services, but they have a number of serious shortcomings including problems caused by censored data, the selection problem, and unaccounted-for heterogeneity among clients and in the services provided. An evaluation of Utah’s VR program, for example, found that the public benefits of the program, measured in dollars, exceed the cost by a factor of 5.64 (Wilhelm and Robinson, 2010). As in Dean et al. (2013a, 2013b, 2013c), our analysis of the Virginia VR program addresses important limitations of these recent studies by evaluating a long panel of labor market outcomes before and after the provision of services; by formally accounting for the possibility that selection into the treatment is endogenous; by focusing on clients with a specific limitation, namely cognitive impairments; and by examining specific types of services rather than just a single treatment indicator. All of these contributions are found to be important for drawing inferences about the impact of VR services. A final contribution is that we explicitly address the biases arising from left-censoring by restricting our sample to first time applicants. Without this restriction, the rate of return analysis is biased. For example, note the effect of left-censoring on the estimated long-run effect of education on employment in Figure 5.

Overall, this innovative model and detailed administrative data reveal a much more nuanced picture of the VR program. On the one hand, as in these
earlier evaluations, we find VR services have large positive returns: for cognitively impaired people, the mean long-run benefits of over $21000 exceed the mean costs by 4 to 6 times, a factor that is similar to the estimate reported in Wilhelm and Robinson (2010). At the same time, we find striking evidence of substantial heterogeneity in the efficacy of VR services across clients (see Figure 8), types of services (see Figure 4), and time periods (see Figure 7). Return on investment analyses ignoring these heterogeneities are nearly certain to present an incomplete and misleading picture of the efficacy of VR services.

References


