Data Issues in Developing Valid ROI Estimates

Steven N. Stern
Stony Brook University

Christopher M. Clapp
University of Chicago

John V. Pepper
University of Virginia

Robert M. Schmidt
University of Richmond

September 2019

The following manuscript was published as:

Stern, S. N., Clapp, C. M., Pepper, J. V., & Schmidt, R. (2019). Data issues in developing valid ROI estimates. *The Journal of Rehabilitation Administration*, 40(1), 35-46.

Abstract

This paper discusses issues associated with using readily available administrative data in estimating ROI for vocational rehabilitation services. It starts with a discussion of longitudinal outcomes data. The discussion is divided up into labor market outcomes data, other types of outcomes data, necessary sample sizes (power analysis), and ways to deal with people systematically excluded from the outcomes data. Next, the paper focuses on services data. The topics covered include the need for control groups, using service cohort data, different sources of service, and merging service data with outcomes data. Finally, the paper moves to the need for other controlling explanatory variables including discussions of inclusion of demographic explanatory variables and data from local labor markets. Two online appendices to this paper provide additional details through (a) an example of a power analysis to illustrate sample size issues and (b) a discussion of Institutional Review Board issues associated with conducting empirical investigations using administrative data.

Keywords: Return on investment (ROI), vocational rehabilitation, longitudinal data, unemployment insurance data, control group, VR service data, personally-identifiable data, explanatory variables

An important step for return on investment/rate of return (ROI/ROR) analysis of state vocational rehabilitation (VR) programs is constructing and cleaning data. Significant time, care, and expense must go into planning for and execution of the process for collecting, merging, and cleaning of the data. Unfortunately, data construction is expensive, sometimes in acquisition costs and always in time spent merging and cleaning. There are no easy and appropriate ways to avoid the investment necessary to have good data. Furthermore, without good data, estimation and analysis are meaningless. Therefore, it is important for any administrator who is considering an ROI/ROR study to make sure that there are adequate resources to develop appropriate data.

ROI/ROR essentially involves a comparison of benefits and costs of a proposed investment. For state VR programs, the benefits are derived mainly from participants' labor market outcomes, and the costs are borne by the state agency providing services to them. A key feature of VR services is that there are many different types of investments. The first section discusses one of the two key data pieces necessary for estimation: longitudinal outcomes data. The section is divided up into discussions of labor market outcomes data, other types of outcomes data, necessary sample sizes (power analysis), and ways to deal with people systematically excluded from the outcomes data. The second section discusses the other key data piece, services data. The section is divided up into discussions of the need for control groups, using service cohort data, different sources of service, and merging service data with outcomes data. The third section discusses the need for other controlling explanatory variables. This includes discussions of inclusion of demographic explanatory variables and data from local labor markets. Measuring cost is outside the scope of this article; however, one can see Clapp, Pepper, Schmidt, and Stern (2019) for a discussion of necessary cost measurement issues.

Longitudinal Outcomes Data

One of the two key data components for an empirical analysis of the effectiveness of a VR program is data on outcomes that are potentially influenced by that program. These data must be linked or merged with data on VR services received by each person. Once merged, the combined data can be used to estimate how receipt of VR services affects the outcomes of interest over a long enough period to produce solid ROI/ROR estimates. Without longitudinal data, one must assume that the labor market effects are constant over time, and this assumption is rejected by the data. In the next section, entitled “Service Receipt Data,” we discuss the necessary characteristics of service data and issues concerning merging. In this section, we limit our discussion to issues associated with the outcomes data.

Labor Market Outcomes

A key set of outcomes for policy makers is labor market outcomes. This arises partially because the stated goal of VR services is to help people with disabilities prepare for, develop skills for, and gain access to the labor market. Four important labor market outcomes are employment, hours worked, wage, and earnings. Employment is a binary measure of involvement in the labor market; one is either employed or not. Hours worked per week, per quarter, or per year provide information on intensity of involvement in the labor market. Wage is a measure of how much one is paid for each fixed unit of time worked. Labor economists argue that wage is a measure of productivity in the labor market (e.g., Ehrenberg and Smith, 2012, chapter 3). One of the goals of a VR program might be to help participants become more productive and therefore increase their wage. Earnings, defined as the wage multiplied by the units of time working, are the outcome that policymakers tend to focus on the most.

Maybe the best available source of data on labor market outcomes that can be merged with information on service receipt is unemployment insurance earnings (UI) data. Each state's UI program collects data from most business establishments on the quarterly earnings of each of their employees. The main purpose of the collection is to determine eligibility and benefit amounts for unemployment insurance benefits. However, as has been recognized (Dean and Dolan, 1991; Hollenbeck and Huang 2006; Wilhelm and Robinson, 2010; Dean, Pepper, Schmidt, and Stern, 2015, 2017, 2018, 2019), the data also can be used as a longitudinal panel for all workers in the state who are eligible for unemployment insurance benefits.

There are three prominent problems with UI data: a) most state UI programs do not collect information on wages or hours worked; b) a significant number of people in each state are not covered by their state's UI program; and c) the researcher cannot distinguish among exit from the labor market, exit from a covered job to an uncovered job, exit from the state, and death. There are no good solutions to (a) and (c). Using only UI data for outcomes prevents the researcher from saying anything about wages (and, therefore, productivity) and hours worked. Both of these pieces of information would be valuable to observe but are not available in most states, nor can they be imputed from the available data. In addition, when a person disappears from the UI data, the researcher cannot determine the reason. This is unfortunate as it would be useful to know the circumstances leading to attrition; different causes of attrition have very different policy implications (e.g., de Graaf, Bijl, Smit, Ravelli, and Vollebergh, 2000; Chang, Yang, Tang, and Gianguli, 2009). For example, death would imply a loss of the accumulated skills, while migration would not. With respect to (b), there are two large groups not covered in a single state's UI data: (1) federal government workers; and (2) people who cross state lines to work, who would be covered in the state where they work but not in the state whose program is being analyzed. The planned State Wage Interchange System (U.S. Department of Labor [USDOL], 2016), which will enable state VR agencies to access other states' UI data on their VR program participants and largely alleviate the issue of people crossing state lines for work. The Federal Employment Data Exchange System (FEDES; Jacob France Institute, 2018) previously provided data on federal employment to participating states to help them meet their reporting requirements and could potentially address the UI data gap for federal workers. However, the FEDES was suspended in January 2018 (USDOL, 2017) while the U.S. Department of Labor assesses its feasibility. In the section entitled, "Dealing with People not Covered by Unemployment Insurance," we discuss some alternative methods to adjust for these current data gaps.

The big advantages of the UI data are that i) the researcher can observe multiple quarters of data both before and after service receipt; and ii) the data are high quality because it has an important administrative purpose. With respect to (i), Dean et al (2015, 2017, 2018) used three years of UI data prior to VR service receipt and ten years after service receipt. As seen in Table 1, the results in Dean et al (2015, 2017, 2018) show that labor market outcomes right after service receipt are not predictive about longer-term effects of VR. The short-run estimates include the first two years after service receipt, and the long-run estimates are for quarters after two years. In almost all cases, the short-run and long-run estimates, both for employment and earnings, are quite different. For example, for people with mental illness, quarterly earnings (if employed) are 5.5% lower in the short run than before service but 13.6% higher in the long run.

Other potentially useful data sets that do not rely on UI coverage have their own problems. One possibility is to use specially provided Social Security Administration (SSA) earnings data (see, for example, Schley, Weathers, Hemmeter, Hennessey, and Burkhauser, 2011;

Honeycutt, Thompkins, Bardos, and Stern, 2015, 2017). However, gaining access to such data is difficult and must be used under restrictive rules. National surveys typically collect self-reported earnings data that is known to be inaccurate. For example, Bricker and Engelhardt (2007) report standard deviations of measurement error in reported earnings data of 5.9% (for men) and 6.7% (for women) in the Health and Retirement Survey. Kim and Tamborini (2014) report systematic measurement error by gender and race in the Survey of Income and Program Participation. Both of these used administrative data to assess the measurement error in earnings. Kapteyn and Ypma (2006) provide a survey of the research on this topic and suggest that administrative data might have errors as well. However, overall, administrative earnings data are usually taken as “the truth.”

Other Outcomes

Besides labor market data, there might be other VR outcomes of interest to policy makers, including job-connected benefits, especially health benefits, and receipt of SSI and SSDI (and other government transfers). Job-connected health benefits are not observable in UI data. They are frequently observed in other national data sets, but these data lack information on receipt of VR services. Information on SSI and SSDI receipt can be acquired from SSA, but at very large cost with significant binding restrictions. For example, Dean et al (2017) used individual SSI/SSDI data attained from SSA on Virginia VR program participants. However, it took about eight years to get the data after initial application, and we were no longer allowed to use them about five years later. Thus, while including such outcomes would provide valuable information, getting meaningful estimates of the outcomes is difficult and expensive.

Other potential VR outcomes include independent living skills, community integration, and emotional health. These are all outcomes that both can be affected by VR services and are of significant value to individuals with disabilities. They should also be important to policy makers even if they are concerned only with government fiscal health. For example, a person with good independent living skills requires less help from expensive government services. However, no administrative body collects data on any of these non-labor market outcomes, and the cost and effort to acquire them through expansion of administrative data systems or survey research would be prohibitive. In addition, by their very nature, measures of independent living skills, community integration, and emotional health are not as objective as earnings and employment and are difficult to monetize.

Necessary Sample Sizes

The first step in ROI/ROR analysis is to estimate statistically significant effects of services on outcomes of interest. The bigger the sample, the more likely the estimates will be statistically significant. A reasonable rule of thumb is to require at least 1000 observations and have a strong preference for approximately 2000 observations. Online Appendix A (available at: https://scholarship.richmond.edu/economics-faculty-publications/55/) provides some detail in a simplified example of a power analysis. In some analyses for people with small prevalence rates (e.g., people who are blind or people with autism), the researcher can increase sample size by using multiple cohorts of service recipients (see the section entitled, “Using a Service Receipt Cohort”).

Dealing with People not Covered by Unemployment Insurance

As previously mentioned, there are two large groups of employed people who are not covered by UI and therefore not included in a given state’s UI data: a) people who work for the federal government; and b) people who commute across state lines to work. The numbers of VR participants who are not included in their state’s UI records can be substantial. Online Appendix

7 for Dean et al (2017) report on an analysis of over 9,000 VR participants in Virginia with records in both the Virginia UI system and SSA earnings files. In 12% of those cases, the SSA showed earnings in 2001 when UI did not. Plausibly, the majority of these instances are due to federal employment and cross-state commuting. This proportion may be larger than in most states because Virginia has many military installations (leading to many jobs with the federal government), and shares borders with Washington, DC (leading to many jobs with the federal government) and Maryland (leading to many jobs in that state).

Both of these problems have two possible solutions. The first (and best) solution, if possible, is to collect labor market data for the missing people. Hollenbeck and Huang (2006) used this approach, and Wilson (2005) proposes using this approach. The second solution is to use statistical techniques to control for the missing data. One can use publicly available data from the federal government on the number of people in each county who work for the federal government and on the number of people who commute across state lines. These can be turned into per capita numbers and then used as regressors to control for variation in prevalence of federal government jobs and across-state-lines commuting (Schmidt, Clapp, Pepper, and Stern, 2019).

Service Receipt Data

While having outcomes data is essential to estimating the effects of VR service receipt, so is having service receipt data that vary across program participants. The service data should reflect the choices made by VR participants and their counselors in a parsimonious way without losing the sense of flexibility in state VR programs. In this section, we begin by discussing why variation in service receipt is essential. Then we discuss issues associated with choosing the source of service data and merging the service data with the outcomes data.

Need for a Control Group

In order to say anything about the effects of VR services on labor market outcomes, one needs a control group; a subset of people who did not receive the services for which the effects are being estimated. It is only through comparison of outcomes for people who did and did not receive the service that the researcher can make any statements about causation (or even correlation). Without a control group, the researcher has no information about the difference in outcomes between those who received service and those who did not; although Bua-Iam, Hampton, Sink and Snuffer (2013) argued in this journal that features unique to state VR programs make it unreasonable to use control groups. Bua-Iam and Bias (2011) estimated effects of VR programs essentially assuming that VR participants would have earned nothing without the VR services. This is an extreme assumption and certainly not as good as using a flawed but reasonable control group such as those constructed by Dean et al (2015, 2017, 2018, 2019).

One of the interesting features of state VR programs is that there are multiple types of services. The existence of people choosing from and sometimes using multiple services allows the researcher to identify the effect of each service even if there are no people using no services (see Clapp et al, 2019). For example, Dean et al (2015, 2017, 2018) aggregated the set of services offered by the Virginia Department of Aging and Rehabilitative Services (DARS) into six separate service groups: diagnosis & evaluation, training, education, restoration, maintenance, and other services. They modeled service choice as well as service impact, allowing for any combination of service categories including none at all. Other authors acknowledge the existence of multiple services but seldom address the issue directly. For example, Aakvik, Heckman, and Vytlacil (2005) note that there are multiple service types but do not model them. Rather, they allow for heterogeneous treatment effects of an observed binary treatment variable. Frolich, Heshmati, and Lechner (2004) observe the existence of multiple programs but focus on comparing them. By contrast, Hollenbeck and Huang (2006) evaluate multiple services separately.

In fact, many VR applicants receive no VR services. These people are either declared ineligible for services or choose to withdraw. One might argue that this group is not a good control group as their non-receipt of services is probably correlated with their labor market outcomes (Dean and Dolan, 1991). The potential correlation causes a bias in the estimate although the direction of the bias is unclear. For example, one might argue that the people who choose to utilize a particular service are those who would benefit most from it. If this were the case, then the observed labor market outcomes for those who used the service would provide an upwardly biased estimate of the service's effect for those who chose not to use it (or for a randomly chosen person). Alternatively, one might argue that the people who choose to utilize a particular service are at greatest need for any service because they have poor market skills. If this were the case, then the observed labor market outcomes for those who used the service would provide a downwardly biased estimate of the service's effect for those who chose not to use it (or for a randomly chosen person). These endogeneity concerns (or "selection" issues) are almost surely valid and are the concern behind much of the economic literature on estimating treatment effects (Dean and Dolan, 1991; Heckman, Ichimura, and Todd, 1998; Heckman, LaLonde, and Smith, 1999; Aakvik et al, 2005; Doyle, 2007). However, the same concern applies to all recipients of any service as well, and the solution is to find or construct instrumental variables (Dean and Dolan, 1991; Heckman et al, 1999; Aakvik et al, 2005; Doyle, 2007; Clapp et al, 2019). An instrumental variable is correlated with the likelihood of VR service receipt but not with employment and earnings, except through the effect of that service on employment and earnings. For example, closing certain order-of-selection categories influences the likelihood of those individuals receiving services, but its only impact on their employment is (1) if it precludes them from getting VR service or (2) if order of selection is correlated with labor market conditions not already controlled for.

Using a Service Receipt Cohort

There are various ways to organize VR administrative data. Dean et al (2015, 2017, 2018, 2019) used observations from the cohort of VR participants who applied for VR services during the fiscal year (FY) 2000 disaggregated into three large disability groups: people with (1) cognitive impairments, (2) mental illness, and (3) physical impairments. They further limited their analysis to people whose first VR case was in FY 2000, because Dean et al (2015) showed that there was bias associated with using people who had had prior cases (as one would expect).

There are two reasons a researcher might want to deviate from such a strategy. First, given changes in VR priorities, the proportion of individuals who are youth in transition has increased significantly. In fact, the Workforce Innovation and Opportunity Act (WIOA) of 2014 requires VR agencies (see section 419 of WIOA) to reserve at least 15% of their allotted funds for pre-employment transition services to students with disabilities. With such participants, using only the first case results either in no labor market outcomes (because they are too young) or deleting a high proportion of them from the estimation sample (because their first service case occurred while they were in school). Thus, other statistical adjustments must be used as an alternative to dealing with the bias suggested in Dean et al (2015).

Second, there are some disability groups with small enough prevalence so that there would not be enough observations if the researcher limited analysis to one cohort year. One way to address this problem is to use multiple year cohorts from the same state. Another is to use single year cohorts across multiple states. For example, in on-going work, Clapp et al (2018b) use applicant cohorts from FY 2007 from Maryland, Oklahoma, and Virginia to estimate the effect of VR service receipt on labor market outcomes for people who are blind, and Clapp et al (2018c) use applicant cohorts from multiple fiscal years in Virginia (FY 2000 - FY 2007) to estimate the effect of VR service receipt on labor market outcomes for adults with autism.

Another interesting possibility is to use available data on multiple cases for each individual rather than exclude individuals with prior cases as was done in Dean et al (2015, 2017, 2018). While including individuals with prior cases increases sample sizes, there are serious data and statistical issues that need to be addressed. First, and most importantly, the researcher must think less in terms of discrete VR cases that begin with application and end with closure, and more in terms of "service spells," each of which potentially could encompass multiple VR cases. Thus, one must take a stand on when a service spell begins and ends. One approach is to use service receipt dates to define a spell. However, our understanding is that dates recorded for services suffer from serious measurement error. Frequently, an authorization record for a specific service includes the earliest date a service can begin as well as the date when the service is expected to be completed (which can be well after the case closure date). Neither identifies the precise delivery date. A second approach is to use case application and closure dates to define service stints. A service stint could be defined as encompassing all cases where a case's application date is within a researcher-defined period (e.g., within the same quarter) of the prior case's closure date. Also, there are interesting econometric issues associated with using this approach. Dean et al (2015) found significantly different service impacts for individuals with prior case(s) versus those for whom the case was their first. Thus, this heterogeneity must be accounted for when including both in a single analysis. Additionally, one might check, for example, Lancaster (1992) for a discussion of issues when using data with varying service spell lengths.

Data on Purchased Services and In-House Services

VR provides services to participants through any combination of four sources -- internally from their VR counselors or other VR staff, through comparable benefits, externally through purchased services, and/or through a state-operated comprehensive rehabilitation facility. This section begins by considering each in turn.

In the past, VR staff typically maintained written case notes about the services they provided in lieu of entering such information into VR administrative databases. Because ROI/ROR analyses tend to use retrospective cohorts and readily available administrative data, the extent and nature of in-house services have rarely been included in such analyses. Similarly, the provision of and precise nature of comparable benefits have seldom been available to the researcher for previous ROI analyses. However, as VR data systems grow in sophistication, better information on in-house services and comparable benefits may become available.

By contrast, state VR agencies have long tracked purchased services for both accounts payable and required federal reporting purposes. Agencies typically record a rich array of information for every purchase, including authorization date, date paid, vendor, service type (generally, with hundreds of categories), dollar amount, and so forth. From an ROI/ROR perspective, the key missing item in these records is the date of service delivery. Although agencies typically record the beginning and ending dates for service delivery, these are set to provide flexibility in their provision and do not provide a clear indication of the actual service date. As a result, researchers must make their own decision/assumption with respect to the period during which services were provided. In the case of Dean et al (2015, 2017, 2018, 2019), services were presumed to begin (potentially) at application. When they ended was not important to determine and was left as an open question.

Eight states also provide services through a state-operated comprehensive rehabilitation facility (e.g., the Maryland's Workforce Technology Center and Virginia's Wilson Workforce and Rehabilitation Center). Generally, these facilities provide services as part of an Individualized Plan for Employment and in coordination with the participant's counselor. In our experience, all of the items recorded for purchased services are also recorded for these facilities with one exception that relates to the cost of the service rather than its provision. For purchased services, the recorded dollar amount reflects the actual price paid by the agency for the specific service. For services provided by these facilities, either no dollar amount is recorded for individual services, or a "charge" for the service is recorded based on a schedule that identifies either a blanket charge or a per-unit charge that is coupled with the number of units provided to the participant.

Service Aggregation

Dean et al (2015, 2017, 2018, 2019) show that different types of VR services are likely to have very different labor market effects. They classify services into six types: diagnostic, training, education, restoration, maintenance, and other. Schmidt et al (2019) reference recent work that replaces the "other" type with two new types, placement and job supports. Irrespective of the service types chosen, mapping the agency's detailed service categories is one of the most important tasks the researcher performs. The rationale by which ROI/ROR analyses group services differs from that by which VR staff think about service types. For example, consider the diagnostic category. From a VR perspective, receipt of diagnostic services suggests eligibility determination; from an ROI/ROR perspective, receipt of diagnostic services identifies other services that would improve functioning in labor markets. Thus, diagnostic services would include eligibility determination as well as medical diagnostics and vocational evaluation. Performing the crucial task of aggregating service categories well requires the combined efforts of VR staff with agency knowledge and a researcher with an ROI/ROR mindset.

An alternative source for identifying service provision is the standardized Case Service Report (also known as the RSA-911) submitted by state VR agencies to the federal Rehabilitation Services Administration (RSA). Over time, RSA has required the reporting of increasing amounts of service-related data. Conceptually, the use of the RSA-911 has an important advantage over the above-described data sources: it includes a code for provision of each service type by agency personnel. As a practical matter, however, if information on in-house services resides strictly in case notes, then the validity of these data relies upon VR counselors to enter this information accurately. To use this source, the researcher must also address how well the RSA-911 categories work for an ROI/ROR analysis. For example, the RSA-911 combines both medical diagnostics and treatment into a single category. However, as discussed above, diagnostic services may have different labor market effects than treatments.

Merging Process

UI data and VR administrative data are not useful for estimating labor market outcomes of VR service receipt unless the researcher can merge the two data sets together. In other words, for each person in the VR administrative data, the researcher must find the labor market earnings history of that person in the UI data. The key to doing this involves merging by Social Security number (SSN) since it is included in both data sets. The level of complexity is increased by the need for maximum possible anonymity required by VR agencies as well as Institutional Review Board (IRB) rules (see online Appendix B, available at:

https://scholarship.richmond.edu/economics-faculty-publications/55/) for a discussion of IRB issues associated with conducting empirical investigations using state VR program data).

A good merging methodology used for SSA data merger in Dean et al (2017) is:

1. VR agency staff member draws one large-digit random number (PIN) for each person in the VR administrative data set, then creates and stores a crosswalk table connecting each SSN in the VR administrative data to the assigned PIN.
2. VR agency staff member strips each VR administrative data observation of the SSN, replaces it with the PIN for that observation (from step 1), removes other personally identifiable information (PII), and provides the amended data to the researchers.
3. VR agency staff member provides SSA staff member with the crosswalk connecting each SSN in the VR administrative data to the assigned PIN for that observation (from step 1).
4. SSA staff member gathers the SSA administrative data observations with the identified SSNs, strips each SSA observation of its SSN and other PII, replaces it with the PIN for that observation (from step 1), and provides the constructed SSA data to the researchers.

Note that no individual in this process sees everything. The VR agency staff member sees none of the SSA data, the SSA staff person sees none of the VR data, and the researchers see no Social Security numbers. Other algorithms could be used, but an alternative algorithm should have the same features: maximum anonymity and minimum complexity with merged, anonymous data going to the researcher at the end of the process. For data needs in Dean et al (2015, 2017, 2018, 2019), a similar but less rigorous method was used for merging DARS and UI data. DARS sent SSNs to a UI staff member and got back the quarterly UI earnings records. DARS then replaced the SSNs with the PINs provided to the researcher in step 1. It is useful to note that, because of the program performance reporting requirements of WIOA, state VR agencies are increasingly coordinating with their UI agencies to get earnings data.

A question of increasing interest motivated by WIOA is the possibility of merging VR data and UI data with administrative data from other agencies providing complementary services, such as the other core WIOA programs. There has been a push to accommodate such data merging partially because of the large number of workforce-related programs existing in many states (e.g., Ganzglass, Reamer, Roberts, Smith, and Unruh, 2010; Jenkins and Harmon, 2010). In principle, this can work as long as there is a common merging identification code across such agencies, the most obvious of which would be the SSN. However, many agencies are cautious about providing Social Security numbers in data available to outside entities, and, in fact, many are moving away from using SSNs even internally. Without such a common merging code, it is impossible to combine agency databases. With such a code, there remain confidentiality issues that can be addressed using carefully constructed merging algorithms as suggested above.

Other Explanatory Variables

Individual Characteristics

It is important to control for other explanatory variables that might influence service choice or labor market outcomes. The online Appendix 6 from Dean et al (2017) as well as Schmidt et al (2019) show that controlling for other explanatory variables is critical for estimation and in particular, has a large effect on estimates of VR service effects. Fortunately, VR administrative data sets provide a wealth of information about each individual. Using such data, Dean et al (2015, 2017, 2018) controlled for gender, race, education, age, marital status, number of dependents, measures of transportation availability, a rich description of the person’s disabilities, and labor market history prior to service receipt. Frolich, Heshmati, and Lechner (2004) controlled for age, gender, citizenship, occupation, some disability variables, and some local community characteristics. Hollenbeck and Huang (2006) controlled for gender, race, age, disability, education, veteran status, English proficiency, and labor market history prior to service receipt.

Controlling for such explanatory variables reduces omitted variables bias (e.g., Stock and Watson, 2007, Chapter 6.1; Wooldridge, 2010, Chapter 4.3). It also may help reduce bias caused by service receipt choices being related to unobserved characteristics directly affecting labor market outcomes. In much of the literature, this is called using propensity scores (Heckman et al, 1998; Frolich et al, 2004; Hollenbeck and Huang, 2006), which help only if they include an instrumental variable. Finally, it improves the precision of parameter estimates associated with the effect of service provision on labor market outcomes as discussed in the section above entitled, "Dealing with People not Covered by Unemployment Insurance."

Local Labor Market Conditions

Another set of useful variables to collect, merge, and use are measures of local economic activity. VR administrative data provide information about the location of residence of each program participant. This information can be combined with county-level measures of economic activity (primarily available through federal government data sources) such as employment rates, unemployment rates, levels of income, and demographic composition. For example, Dean et al (2015, 2017, 2018) used information on ratios of number of jobs to population in each county from the Bureau of Economic Analysis, and Dean et al (2019) used Virginia data on the number of youth with Individualized Education Programs (IEPs) by county. Another example of such data is described in the section above entitled, “Dealing with People not Covered by Unemployment Insurance.”

Summary

Constructing quality data is a necessary step in any good ROI/ROR analysis. Doing it well requires skill, knowledge, and care. However, available data provide one with the opportunity to learn much about the operation and effectiveness of the VR agency. Such an analysis is useful both for providing evidence to policy makers interested in the effectiveness of the programs being funded and for use in informing decision-making and continuous improvement methods by agency administration. Our experience is that administrative VR agency data and UI earnings data are very high quality. It is still the user’s responsibility to ensure that data are collected in an ethical manner, data elements are considered appropriately, and data are protected.

References

Aakvik, A., Heckman, J., & Vytlacil, E. (2005). Estimating treatment effects for discrete outcomes when responses to treatment vary: An application to Norwegian vocational rehabilitation programs. *Journal of Econometrics*, *125*, 15-51.

Bricker, J. & Engelhardt, G. (2007). Measurement error in earnings data in the health and retirement study. Unpublished manuscript.

Bua-Iam, P. & Bias, T. (2011). Economic impacts of West Virginia Division of Rehabilitation Services on consumers with significant disabilities: Realistic return-on-investment models for state-federal VR programs. *Journal of Rehabilitation*, *77*(3), 25-30.

Bua-Iam, P., Hampton, J., Sink, T., & Snuffer, D. (2013). Critical issues unique to the selection of a realistic Return on investment method for public vocational rehabilitation programs. *Journal of Rehabilitation Administration*, *37*(1), 25-33.

Chang, C., Yang, H., Tang, G., & Ganguli, M. (2009). Minimizing attrition bias: A longitudinal study of depressive symptoms in an elderly cohort. *International Psychogeriatrics Journal*, *21*(5), 869--878.

Clapp, C. M., Pepper, J. V., Schmidt, R. M., & Stern, S. N. (2018a). Identification with multiple treatments. Unpublished manuscript.

Clapp, C. M., Pepper, J. V., Schmidt, R. M., & Stern, S. N. (2018b). Ongoing work on effects of vocational rehabilitation for people who are blind.

Clapp, C. M., Pepper, J. V., Schmidt, R. M., & Stern, S. N. (2018c). Ongoing work on effects of vocational rehabilitation for adults with autism.

Clapp, C. M., Pepper, J. V., Schmidt, R. M., & Stern, S. N. (2019). Conceptual issues in developing return on investment estimates of vocational rehabilitation programs. *The Journal of Rehabilitation Administration*, forthcoming.

Dean, D. H. & Dolan, R. C. (1991). Fixed-effects estimates of earnings impacts for the vocational rehabilitation program. *Journal of Human Resources*, *26*, 380-391.

Dean, D. H., Pepper, J. V., Schmidt, R. M., & Stern S. N. (2015). The effects of vocational rehabilitation for people with cognitive impairments. *International Economic Review*, *56*(2), 399-426.

Dean, D. H., Pepper, J. V., Schmidt, R. M., & Stern, S. N. (2017). The effects of vocational rehabilitation for people with mental illness. *Journal of Human Resources*, *52*(3), 826-858.

Dean, D. H., Pepper, J. V., Schmidt, R. M., & Stern, S. N. (2018). The effects of vocational rehabilitation for people with physical impairments. *Journal of Human Capital*, *12*(1), 1-37.

Dean, D. H., Pepper, J. V., Schmidt, R. M., & Stern, S. N. (2019). The effects of youth transition programs on labor market outcomes. *Economics of Education Review*, *68* (February), 68-88.

de Graaf, R., Bijl, R., Smit, F., Ravelli, A., & Vollebergh, W. (2000). Psychiatric and sociodemographic predictors of attrition in a longitudinal study: The Netherlands mental health survey and incidence study (NEMESIS). *American Journal of Epidemiology*, *152*(11), 1039-1047.

Doyle, J. (2007). Child protection and child outcomes: Measuring the effects of foster care. *American Economic Review*, *97*, 1583-1610.

Ehrenberg, R. & Smith, R. (2012). *Modern Labor Economics: Theory and Public Policy. 11th edition*. Boston: Pearson Education.

Frolich, M., Heshmati, A., & Lechner, M. (2004). A microeconometric evaluation of rehabilitation of long-term sickness in Sweden. *Journal of Applied Econometrics*, *19,* 375-396.

Ganzglass, E., Reamer, A., Roberts, B., Smith, W., and Unruh, R. (2010). Recommendations for incorporating postsecondary and workforce data systems in statewide longitudinal data systems. Retrieved from http://www.clasp.org/resources_and_publications/publication?id=0734&list=publications

Heckman, J., Ichimura, H., and Todd, P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies*, *65*, 261-294.

Heckman, J., LaLonde, R., and Smith, J. (1999). The economics and econometrics of active labor market programs. *Handbook of Labor Economics Volume 3*. (eds.) O. Ashenfelter and D. Card. Amsterdam: North-Holland,1865-2097.

Hollenbeck, K. and Huang, W. (2006). *Net Impact and Benefit-Cost Estimates of the Workforce Development System in Washington State*. Kalamazoo: W.E. Upjohn Institute for Employment Research.

Honeycutt, T., Thompkins, A., Bardos, M., and Stern, S. (2015). State differences in the vocational rehabilitation experiences of transition-age youth with disabilities. *Journal of Vocational Rehabilitation*, *42*, 17-30.

Honeycutt, T., Thompkins, A., Bardos, M/. and Stern, S. (2017). Youth with disabilities at the crossroads: The intersection of vocational rehabilitation and disability benefits for youth with disabilities. *Rehabilitation Counseling Bulletin*, *60*(3). Retrieved from http://journals.sagepub.com/doi/abs/10.1177/0034355215621897

Jacob France Institute, 2018. Retrieved from http://www.ubalt.edu/jfi/fedes/

Jenkins, D. and Harmon, T. (2010). Using state data to promote continuous improvement of workforce programs: Guidance for states preparing applications to the U.S. DOL Workforce Data Quality Initiative. Retrieved from https://ccrc.tc.columbia.edu/media/k2/attachments/using--state-data-workforce.pdf

Kapteyn, A. and Ypma, J. (2006). Measurement error and misclassification: A comparison of survey and register data. Unpublished manuscript.

Kim, C. and Tamborini, C. (2014). Response error in earnings: An analysis of the survey of income and program participation matched with administrative data. *Sociological Methods and Research*, *43*(1), 39-72.

Lancaster, T. (1992). *The Econometric Analysis of Transition Data*. Econometric Society Monograph.

Schley, S., G., Weathers, W. R., Hemmeter, J., Hennessey, J., and Burkhauser, R. (2011). Effect of postsecondary education on the economic status of persons who are deaf or hard of hearing. *Journal of Deaf Studies and Deaf Education*, *16*(4), 524-536.

Schmidt, R. M., Clapp, C. M., Pepper, J. V., & Stern, S. (2019). Applications of the VR-ROI project: ROI estimates for Virginia and Maryland. *The Journal of Rehabilitation Administration*, forthcoming.

Stock, J. and Watson, M. (2007). *Introduction to Econometrics*. Boston: Addison-Wesley.

United States Bureau of Labor Statistics (2018). National Longitudinal Surveys. https://www.nlsinfo.org/content/cohorts/nlsy97; accessed September 2018.

U.S. Department of Labor, 2016. Retrieved from https://wdr.doleta.gov/directives/attach/TEGL/TEGL_26-16_Acc.pdf

U.S. Department of Labor, 2017. Retrieved from https://wdr.doleta.gov/directives/attach/TEN/TEN_3-17_Acc.pdf

W, S. and Robinson, J. (2010). Utah State Office of Rehabilitation economic impact study. *Journal of Disability Policy Studies*, *24*(3), 148-157.

Wilson, B. (2005). Integrated performance information for workforce development. Retrieved from https://www.nga.org/files/live/sites/NGA/files/pdf/0603IPIBLUEPRINT.pdf

Wooldridge, J. (2010). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Workforce Innovation and Opportunity Act (2014). Retrieved from https://www.congress.gov/113/bills/hr803/BILLS-113hr803enr.pdf

Table 1: Short-Run and Long-Run Changes in Employment and Earnings Due to Receipt of Training Services

	Cognitive Impairment	Mental Illness	Physical Impairment
Change in Employment Rate			
Short Run	10.3%	9.4%	5.0%
Long Run	6.8%	8.0%	6.4%
% Change in Quarterly Earnings (if employed)			
Short Run	20.9%	-5.5%	0.9%
Long Run	28.5%	13.6%	17.2%

Notes:
1) Short-run numbers are for the first two years after service, and long run numbers are after

2) All numbers come from Dean et al (2015, 2017, 2018). Employment Rate numbers translate the reported results into rates.

3) Both short-run and long-run numbers above are the difference in estimates after service receipt minus estimates prior to service receipt. The change in the outcome is the effect of the service.

Author's Note

Steven Stern, Ph.D., is a full Professor of Economics at Stony Brook University. Christopher Clapp, Ph.D., is a Consultant for the VR Return on Investment Project and a Senior Research Associate in the Kenneth C. Griffin Department of Economics at the University of Chicago. John Pepper, Ph.D., is a researcher for the VR Return on Investment Project and the Merrill H. Bankard Professor of Economics at the University of Virginia. Robert Schmidt, Ph.D., is a co-Principal Investigator for the VR Return on Investment Project and an Emeritus Professor of Economics at the University of Richmond.

We would like to thank Joe Ashley, Robert Froehlich, Maureen McGuire-Kuletz, Kevin Hollenbeck, and especially Kirsten Rowe for thoughtful assistance and comments and the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR Grant 90DP0070) for generous funding. All errors are ours.