Conceptual Issues in Developing Return on Investment Estimates of Vocational Rehabilitation Programs

Christopher M. Clapp
University of Chicago

John V. Pepper
University of Virginia

Robert M. Schmidt
University of Richmond

Steven N. Stern
Stony Brook University

Abstract. *We provide an overview of the basic conceptual issues involved in estimating the return on investment (ROI) of state vocational rehabilitation (VR) programs. Our aim is to highlight some of the key issues in ROI evaluations, especially those associated with estimating the benefits and costs of VR. Finally, we discuss different ways of implementing ROI calculations and suggest that rate-of-return type analysis is appealing for VR evaluations where there is no widely accepted discount rate.*

Keywords: Return on investment (ROI), rate of return (ROR), vocational rehabilitation, benefit-cost analysis, discount rates, alternative evaluation designs, long-run benefits

Each year, over 1.3 million disabled adults receive services from public-sector vocational rehabilitation (VR) programs at a cost of around $3 billion per year. These federally mandated state programs are administered by 80 different state-level VR agencies that work to ensure that clients "achieve high-quality employment outcomes" (U.S. Department of Education, 2018). In the last decade, there has been heightened interest in producing credible evaluations of whether these VR programs have been effective in meeting that goal. Several recent reports from the U.S. Government Accountability Office (2005, 2007, 2012) highlight the need for improved data and evaluation methodologies. Additionally, the 2014 Workforce Innovation and Opportunity Act (WIOA), which requires formal reports on VR clients' post-program employment and earnings, has further increased the need for updated and credible data and evaluations. To do this, researchers often turn to a return on investment (ROI) analysis.

Return on investment analyses of state VR programs provide a succinct and useful measure of program efficacy. ROI is a measure of investment performance that compares the amount of financial return or benefit relative to program cost (Hollenbeck, 2019; McGuire-Kuletz & Tomlinson, 2015). For example, one commonly used ROI measure is the benefit cost ratio (BCR) which is the ratio of the present value of selected monetizable program benefits to the present value (PV) of costs:

$$\frac{PV\ of\ Benefits}{PV\ of\ Costs}$$

For every dollar spent on a VR client, the BCR shows how many extra dollars (in present value terms) the client earns as a result. Therefore, if the BCR exceeds one, the ROI is positive This BCR may vary across individual VR clients. One may compute the BCR for each individual or provide a summary measure such as the mean, median, or some other quantile. Another related ROI measure called the rate of return is discussed in Section IV.

While the idea is straightforward, developing a credible ROI estimate is a difficult undertaking. Given the available data, one must first estimate and monetize the present value of the benefits and costs of VR services and then, using basic mathematical formulas such as the BCR, determine the ROI.

In this paper, we provide an overview of the basic conceptual issues involved in estimating the ROI of VR programs. McGuire-Kuletz and Tomlinson (2015), and Hollenbeck (2019) provide a more detailed and technical guide to VR-ROI analyses. Our aim is to highlight some of the key issues in ROI evaluations, not to provide an exhaustive how-to guide. As such, this paper should help VR administrators, state program evaluators, policymakers, and others appreciate the complexities involved in developing a credible analysis and interpreting ROI results. Much of this paper focuses on the central issues involved in estimating the benefits of VR. This is the most critical and demanding part of the ROI analyses. After reviewing the issues involved in estimating the benefits of VR in the first section, we then turn to the more mundane albeit important issues involved in determining the costs of VR and the ROI estimates in the third and fourth sections, respectively. The last section draws conclusions.

Estimating the Benefits of VR on Labor Market Outcomes

As discussed in Stern, Clapp, Pepper, and Schmidt (2019), impact evaluations of VR typically focus on labor market outcomes (i.e., employment, wages and earnings). Employment outcomes are of interest to policymakers, and the primary objective of VR programs is to improve labor market outcomes (U.S. Department of Education, 2018). Moreover, labor market outcomes are easily quantified because they are monetized. VR may also have important effects on other outcomes such as self-esteem and independent living skills but these outcomes are difficult to measure and quantify.

The basic idea of an impact evaluation is simple and appealing. Program outcomes – for example, employment and earnings/wages – are measured and compared to the outcomes that would have resulted in the absence of the program. In practice, however, it is difficult to design a credible evaluation where this comparison can be made. The fundamental difficulty is that client outcomes in the absence of the program are counterfactual and not observable. What would have happened to VR recipients had they not received services?

The data alone cannot answer this question. This fundamental methodological problem, labeled the counterfactual outcomes or selection problem, requires that the evaluation design provide some basis for constructing a credible estimate of the counterfactual outcome. This is difficult in practice because VR clients (or their counselors) choose services based on unobservable confounding characteristics that may bias the counterfactual estimate. For instance, highly motivated individuals may seek out and take full advantage of multiple VR services, then find success in the labor market because of both that training and their motivated attitudes. Had the individuals not received VR services, they still might have enjoyed a good deal of job-market success because of their hard-working ways. In this scenario, the researcher has no way of knowing whether positive labor market outcomes are due to VR services or unobserved client motivation because the counterfactual scenario without VR assistance is unobservable. A positive association between VR services and labor market outcomes may reflect unobserved client attitudes and motivation. This would result in an overstatement of VR benefits (positive selection bias). Alternatively, clients with significant impairments that limit their potential returns in the labor market may attempt to overcome the significance of their disabilities by making use of multiple VR services. If the clients' disabilities would have resulted in poorer than average labor market outcomes in the absence of VR services, the effects of those services will be understated (negative selection bias).

More generally, unobserved characteristics such as ability, attitude towards work (e.g., motivation), health status, family support, local labor market conditions, access to transportation, and support from other related programs may affect both the decision to receive substantial VR services and labor market outcomes. Thus, any observed relationships between VR service receipt and labor market outcomes could be spurious. A *selection problem* results from the facts that a) individuals may *select* themselves into a treated group that receives substantial VR services or an untreated group that does not receive substantial services based on their expectation of the resulting labor market outcome and b) the data alone cannot reveal what the counterfactual labor market outcomes would have been.

In a randomized controlled trial (RCT) research design, concerns about selection are negated by randomly assigning subjects into either a treatment group that receives substantial VR services or a control group that does not receive substantial services. In this setting, the decision to assign services is exogenous or unrelated to the labor market outcomes. In practice, selection bias may also impact an RCT if some individuals assigned to the treatment group do not follow through on treatment (dropout bias) and/or individuals assigned to the control group obtain similar treatment outside of the program (contamination bias). Although a useful benchmark to keep in mind, the RCT design is infeasible in most VR settings where counselors and agencies are reluctant to, or possibly even prohibited from, randomly assigning different VR services to clients.

With administrative data on VR clients (see Stern et al., 2019), conclusions about the counterfactual outcomes depend critically on what data are available and what assumptions the evaluator brings to bear. Although this problem can be resolved if the employment data are combined with sufficiently strong assumptions, there is no established solution to the counterfactual outcomes problem that is valid in all settings. Labor economists have long recognized this as the central problem in addressing the impact of job training programs (Friedlander, Greenberg, & Robins, 1997; LaLonde, 1995). Hotz (1992) provided a framework for the Governmental Accountability Office that laid out several options for evaluation of the public-sector VR program in a non-experimental setting that presents a variety of techniques to control for the problem of selection bias inherent in such voluntary programs. Imbens and Wooldridge (2009) provide a summary of some of the recent developments in program evaluation methodologies. As such, establishing credible estimates of what the outcomes would have been without the program is the most critical and demanding part of an impact evaluation. When those estimates are convincing, the effects found in the evaluation can be attributed to the program rather than to any of the many other possible influences on the outcomes (e.g., unobserved motivation, health issues or functional limitations, imperfect controls for local economic conditions, or unobserved support from other programs). Otherwise, the evaluation may be misleading. For example, a simple comparison of the employment outcomes of treated and untreated clients may not estimate the true impact of VR services. Any differences in labor market outcomes could be due easily to one or more of the aforementioned preexisting differences between the groups. The job of a good impact evaluation design is to neutralize or rule out such problems.

Three Simple Evaluation Designs

To illustrate the counterfactual outcomes problem in a relatively simple setting, we reexamine the data from Dean, Pepper, Schmidt, and Stern's (2018) analysis of the Virginia General VR program on clients diagnosed with physical impairments. Since clients receive services for an average of about two years, we focus on employment outcomes three years after the application quarter. This analysis is based on a pre-WIOA period and uses pre-WIOA data.

Table 1 displays the quarterly employment rates one year before and three years after the application for VR services in state fiscal year (SFY) 2000 for clients who received substantial VR services and those who did not receive substantial services. Manski and Pepper (2018) provide a similar illustration in their analysis of right-to-carry gun laws. Following the literature, we refer to these two groups as the treated and untreated, respectively.

These data may be used to compute three simple estimates of the effect of VR services on employment rates. A "before-and-after" analysis compares employment rates for treated clients, yielding the estimate -0.11 (0.41 – 0.52). This estimate suggests VR reduces the employment probability by 11 percentage points. Contemporaneous comparison of the treated and untreated yields the estimate 0.13 (0.41 – 0.28), suggesting VR increases the employment probability by 13 percentage points. The difference-in-difference (DID) estimate compares the time-series changes in employment rates for the treated and untreated, yielding the estimate 0.14 [(0.41-0.52) - (0.28 – 0.53)]. This estimate suggests VR increases the employment probability by 14 percentage points.

These three estimates yield different empirical findings. Given the validity of certain assumptions, each

Table 1

Quarterly Employment Rates by Application Quarter and Treatment Status, SFY 2000 Virginia General VR Agency Clients with Physical Impairments

	Group[2]	
Period[1]	Untreated	Treated
Pre-Application	0.53	0.52
Post-Application	0.28	0.41

Note:

1. The period is four quarters before (pre) or twelve quarters after (post) the date when the VR clients applied for services in SFY 2000.
2. The treated group received substantial VR purchased services. The untreated group did not.

might appropriately measure the effect of VR on the employment rate of Virginia's clients with physical impairments in SFY 2000. However, the assumptions that justify the interpretations differ across estimates, and there is no guarantee that any of the requisite assumptions are valid. Moreover, even if the underlying assumptions are valid, there are several reasons this type of analysis may not reflect the true social benefits of VR services (Dean et al., 2017). First, these estimates do not account for the potential displacement of non-VR participants, particularly if VR services do not improve the VR participant skills or the job matching process. Second, VR services may lead to improved self-esteem and other social benefits associated with increased attachment to the labor market as well as a resulting reduction in use of the social welfare system. While society does not benefit from reduced transfer payments or increased tax revenues – taxpayer gains exactly offset VR participant losses (except for changes in deadweight loss) – social benefits may result from reduced administrative cost associated with welfare programs and increased VR participant utility due to reduced welfare dependence, improved health status, and access to health care insurance (LaLonde, 1995). At the same time, the deadweight costs of taxation may change if welfare receipt and tax payments change.

The "before-after" analysis is correct if one can credibly assume that no determinant of employment, including health status or the local labor market, changed over the four-years between the pre- and post-application periods except for receipt of substantial VR services. In this illustration, the assumption does not appear to hold, at least for the untreated. The employment rate for the untreated fell from 0.53 one year prior to the application quarter to 0.28 three years after the application quarter. Since the untreated group did not receive substantive VR services, something else must have changed, possibly their health and/or local labor market conditions. This casts doubt on the validity of the "before-after" assumption and analysis.

The contemporaneous comparison of employment rates is correct under the assumption that the treated and untreated had the same employment propensities and faced the same labor market environments except for the fact that the treated received substantial VR services. This is commonly referred to as the exogenous or random selection assumption that is credible in RCTs, but it is not generally credible in observational studies where treatments (i.e., VR service receipt) are self-selected. A particular concern is that the collaboration between counselors and clients in determining a plan for services (i.e., the Individualized Plan for Employment) may be influenced by a client's propensity to find employment. In this case, the observed association would be spurious: treated clients would have higher or lower employment rates regardless, depending on whether the selection bias is positive or negative.

Finally, the DID finding is correct if one can plausibly make the assumption that, in the absence of VR services, the treated and untreated would have experienced the same change in employment rates. As with the before-after analysis, the DID model alone only estimates the effect of VR for treated clients. To use this model to estimate the effect for the full population of clients, one needs to combine the DID assumptions with a homogeneity assumption that the effect of VR on employment is the same for the treated and untreated clients. This often is formalized using a linear mean regression model that assumes the effect is the same for all clients.

Clearly, the credibility of this approach depends on whether the "untreated" are a reasonable comparison group – that is, do the untreated clients provide information on the counterfactual trends in the employment rates for the treated clients? To proxy for those trends, researchers have used a number of different internal comparison groups in practice. Those groups include individuals who apply but drop out of the program after being determined eligible and applicants who are ''screened-out'' (e.g., persons whose disabilities are too significant for them to benefit from VR services or those whose disabilities do not constitute or result in a substantial barrier to employment). As with the contemporaneous comparison analysis, a common concern with this approach is that the treatment decision – whether it is made by the client deciding to drop out or the VR counselor who screens out – may, in part, be based upon beliefs about either a client's propensity to find employment or the efficacy of services for that client.

All three of these research designs are commonly used in the literature on the impact of VR programs, frequently in the same evaluations. To determine the impact of workforce development programs in Texas, King, Tang, Smith, Schroeder, and Barnow (2008), and Smith, Christensen, and Cumpton (2015) use a before-after design to evaluate the effects of low-intensity services and contemporaneous comparison to evaluate the effects of high intensity services (relative to low intensity services). Hollenbeck and Huang (2006), and Maryns and Robertson (2015) use both contemporaneous comparison and DID methods to evaluate Washington state's and Minnesota's workforce programs, respectively, while Uvin, Karaaslani, and White (2004), and Wilhelm and Robinson (2013) use all three methods to evaluate the VR programs in Massachusetts and Utah.

While these three approaches are widely used, it may be difficult to credibly address the selection problem using the internal comparison groups they are all based on. VR services are not likely to be randomly assigned, and any imaginable control group is likely to differ in ways that may lead to spurious correlations in the observed data and biased employment impacts.

One common but potentially problematic approach for addressing this concern is to statistically account for observed factors such as age, gender, disability status and severity, and so forth. In this case, researchers assume that VR service receipt is exogenously or randomly assigned conditional on the set of observed covariates even if it may not be exogenous when excluding such control variables from the analysis. A related approach statistically matches clients to untreated individuals based on observable characteristics to construct the most similar counterfactual group (Hollenbeck & Huang, 2006). Yet, the fact that clients with the same covariates receive different services suggests that confounding unobserved factors may play a role in the selection process.

Other Evaluation Designs

Given concerns that VR services are generally not randomly assigned, other model-based evaluation designs have been applied in the literature assessing VR programs. Dean and Schmidt (2005a), for example, address the selection problem by modeling the joint relationship between earnings and VR service receipt using the Heckman (1979) two-stage selection model. Aakvik, Heckman, and Vytlacil (2005) use similar statistical modelling approaches to evaluate VR programs in Norway. More recently, Dean et al. (2015, 2017, 2018, 2019) combine the basic structure of the DID model of labor market outcomes with a model of VR service receipt decisions. By formalizing and estimating a model jointly describing how treatments are selected and outcomes determined, these studies can evaluate the impact of VR services in the presence of the selection problem. While these nonlinear simultaneous equations models allow researchers to formally model the selection problem, they are theoretically, statistically, and computationally complex. This makes them difficult to estimate and evaluate. In contrast, the before-after, contemporaneous comparison, and DID models in the previous section that take realized treatments as given and only model outcomes are relatively more straightforward. For example, Dean et al. (2015, 2017, 2018, 2019) include three jointly determined equations to reflect the mix of services provided, clients' choices to work, and their earnings conditional on working. Since the selection problem occurs because unobserved characteristics may affect both service and labor market outcomes, the researchers model all three relationships as a function of random, unobserved components or error terms. Using this model, they allow services to be assigned based in part on expected labor market outcomes through those unobserved components.

Finally, a well-established approach to address the selection problem exploits some observed covariate, termed an instrumental variable (IV), which has no direct effect on employment outcomes but does influence VR service receipt. In statistical terminology, the IV is said to be independent of employment outcomes but not service receipt. This type of exogenous variation has been shown to help estimate the impact of the treatment. A number of possible observed variables might serve as credible instruments for evaluating the impact of VR services. For example, a client's distance to a VR field office and service provider capacity in a specific geographic area might be related to whether a VR applicant receives services but unrelated to labor market outcomes. Likewise, an order of selection regime may serve as an instrument that is correlated with service receipt but not labor market outcomes.

Dean et al. (2015, 2017, 2018, 2019) use the propensity of a client's VR counselor to assign specific services as an instrument, arguing that counselor tendencies impact VR service receipt but are not directly related to labor market outcomes. As a simplified but intuitive example to illustrate how this IV addresses the selection problem, one can think of there being two types of counselors with respect to a particular service type: high and low propensity. High-propensity counselors decide that every client requires substantial VR services of that type, and low-propensity counselors decide that no client should receive substantial services of that type. If counselors are randomly assigned to clients, or at least if the assignment is unrelated to future labor market outcomes as the researchers argue, then the unobserved factors associated with the assignment to VR services are effectively exogenous, just as in a RCT (Dean et al., 2015, 2017, 2018, 2019).

Two Other Issues

We highlight two other issues related to impact evaluations that are particularly salient for ROI analyses. First, it is important to recognize there is variation in the types of VR services and the types of impairments of VR clients. Second, there may be differences between the short- and long-run impact of VR.

Accounting for heterogeneity in VR services and in the client population. VR agencies provide a wide range of different services to clients with a wide range of disabilities and other characteristics. The decision of how to account for this variation, or heterogeneity, in services and client circumstances is a key issue in designing an impact evaluation. If the estimated impacts differ by type(s) of service received and the type of limitation, the ROI is likely to vary across services and individuals.

Most evaluations classify clients as either receiving or not receiving substantial VR services. Dean et al. (2002) and Dean et al. (2015, 2017, 2018, 2019) aggregate VR services into six types: (1) diagnosis and evaluation, (2) training, (3) education, (4) restoration, (5) maintenance, and (6) other; and allow these six services to have different labor market effects. Moreover, the authors evaluate the impact of VR services on clients with specific types of impairments (e.g., mental illness, cognitive impairments, and physical impairments) rather than the entire caseload, and account for a number of different observed factors, including age, race, gender, years of schooling, and the severity of the disability. Except for Dean and Dolan (1991), and Dean et al. (2015, 2017, 2018, 2019), the existing state-level evaluations of VR services either ignore differences in limitations entirely (Bua-Iam & Bias, 2011; King et al., 2008; Maryns & Robertson, 2015; Wilhelm & Robinson, 2010) or distinguish among clients with different disabilities only by including dummy variables for type of impairment in regression models (Hollenbeck & Huang, 2006; Uvin et al., 2004).

Measuring long run benefits. VR services are thought to have long-run labor market benefits that may be important to account for in an ROI calculation. Dean and Schmidt (2005b), for example, argue that the 10-year ROI is too conservative since earnings gains may be incurred many years after the program. The problem with conducting a lifetime ROI estimate is that the data used to evaluate VR programs do not include lifetime labor market profiles. The longest panel used in the literature evaluating VR programs is the Dean et al. (2015, 2017, 2018, 2019) analysis of applicants to the Virginia general VR agency in SFY 2000 which uses the quarterly labor market outcomes of clients for ten years post-application. Mann, Honeycutt, Bailey, and O'Neill (2017) track VR client outcomes for up to seven years after service receipt. Without the full lifetime labor market profile, which may be too time consuming and costly to assemble, analysts face the problem of trying to use near-term, observed labor market data to draw conclusions about lifetime, unobserved labor market outcomes. To resolve this problem, researchers impute the longer-run benefits from the shorter-run outcome data. Imputing long-run benefits requires assumptions mapping observed data and benefit estimates to future benefit forecasts. The problem is that there is not a single set of assumptions for the extrapolation problem that credibly applies in all settings (Manski, Newman, & Pepper, 2002). This problem may be mitigated in cases where short- and intermediate-run outcomes imply a high rate of return. In such cases, the discounted longer run outcomes may not matter enough to change the basic qualitative conclusion.

Estimating the Costs of VR

Relative to estimating the impact that VR services have on client outcomes, determining the cost of providing VR is straightforward. ROI studies of VR generally ignore the counterfactual outcomes problem when assessing costs. In this case, one merely assembles the realized costs data on VR services and administration. (Yet, there could be a selection problem if there is heterogeneity in costs related to unobserved client characteristics.) Data from the state agency's client data system and from the Rehabilitation Service Administration's Annual Vocational Rehabilitation Program/Cost Report (also known as the RSA-2) provide the necessary information on the costs of services and administrative costs.

Services are provided to clients in any combination of three ways: (1) as a "purchased service" through an outside vendor using agency funds, (2) as a "similar benefit" purchased or provided by another governmental agency or not-for-profit organization with no charge to the VR agency, and/or (3) internally by agency personnel ("in-house benefits"). The section entitled, "Data on Purchased Services and In-House Services," in Stern et al., 2019, provides more detail. VR administrative data provide actual purchased service costs but may not contain the same detailed information for in-house services or similar benefits. Instead, Dean et al. (2015, 2017, 2018, 2019) measure non-purchased service provision costs and administrative costs using data from the RSA-2.

To be clear, there is some uncertainty about the cost estimates derived using the RSA reports, especially for the costs of in-house and similar benefit services. A more detailed analysis of these costs would be useful. In the absence of these details, Dean, Pepper, Schmidt, and Stern (2015, 2017, 2018, 2019) report a range of ROI estimates under different costs estimates.

Computing ROI

Given estimated benefits and costs of VR services, one can then compute a ROI. The basic computations are well-known and largely standardized. Still, there are number of steps in the process that are worth reviewing.

The first step is to discount the dollar values of future benefits and costs to a present value. Benefits and possibly the costs of VR services may be accrued over many years, and a dollar today is worth more than a dollar tomorrow. Discounting is a way to standardize the units of future dollars so they are comparable with current dollars. This allows for an apples-to-apples comparison that reflects the different periods when benefits and costs may be realized. Importantly, this is not an adjustment for inflation but rather a way to account for the real gains that could be realized by investing a dollar today.

Formally, the present value of money recieved in periods in the future equals

$$PV_0 = F\ V_n/(1 + r)^n$$

where PV_0 is the present value in year 0 (i.e., the base year), FV_n is the value n periods into the future (i.e., the future value of benefits), and r is the discount rate. When future streams of money accrue over multiple periods, one adds the discounted stream of money from each period. See Hollenbeck (2019) for more details.

To illustrate, suppose that, five years from today, one will receive \$15,000. How much is that \$15,000 worth today? If the discount rate equals 0.05, then the present value equals \$11,753 ($15{,}000/(1+.05)^5$). That is, with a five-percent discount rate, \$15,000 in five years (future value of benefits) is worth \$11,753 today (present value of benefits). In other words, investing \$11,753 compounded annually at five percent would yield \$15,000 in five years.

Figure 1 displays the present value of \$15,000 five years from today for a range of discount rates from 0.00 to 0.25. For instance, the figure shows that, with a 0.02 discount rate, the present value of \$15,000 in five years is \$13,586, and, for a discount rate of 0.10, the present value is \$9,314.

Clearly, the present value is sensitive to the choice of the discount rate, r. The discount rate represents the foregone value of money spent today. Stated another way, it is the opportunity cost of not saving or investing capital in the current period. It is chosen by the researcher and is often set to or at least centered on

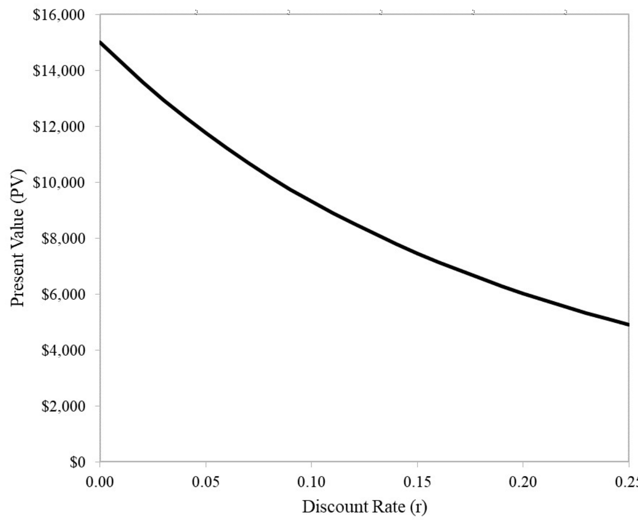

Figure 1. Illustration of How Discount Rate Affects Present Value. For this example, we assume a future value of $15,000 and a time horizon (n) of five years.

some basic interest rate (e.g., savings account interest rate).

After discounting the stream of benefits and costs to the present, a straightforward way to assess the ROI is to compare the present value of benefits to costs. In particular, as noted previously, the BCR equals

$$\frac{PV\ of\ Benefits}{PV\ of\ Costs}$$

If the present value of benefits exceeds the present value of costs, the return to VR services is positive and the BCR > 1. Otherwise, the return to VR services is negative and the BCR < 1. The BCR can be interpreted as the "bang per buck." In the VR context, this means that, for every dollar of VR service provision, the customer earns BCR extra dollars (in present value terms). For example, suppose the present value of the costs of VR services is $10,000 and the present value of the benefits is $11,753. Then, the BCR is 1.18, implying that a dollar of VR services results in $1.18 in additional earnings.

Although the BCR is easy to interpret, it is sensitive to the choice of the discount rate. Lower values of the discount rate make the investment look better, and higher values make it look worse. To illustrate, note that the previous hypothetical example calculates the present value of benefits by assuming that VR results in $15,000 in benefits in five years and the discount rate is 0.05. Yet, if the discount rate is 0.084, then the present value of benefits equals $10,000 and the BCR = 1. If the discount rate is 0.10, then the present value of benefits is $9,314, and the BCR is less than one.

The sensitivity of the BCR to the discount rate may be problematic for evaluating workforce training programs. Businesses typically use some measure of their financing costs (i.e., "cost of capital") as a discount rate when evaluating an investment. By contrast, there is no widely accepted "cost of capital" or discount rate for evaluating workforce training programs. Moore, Boardman, Vining, Weimer, and Greenberg (2004) present a discussion of the issues surrounding the use of discount rates in program evaluation and guidance on how to choose an appropriate rate. The choice is largely arbitrary, and, given the sensitivity of BCR to the discount rate, the rate used can make a program look good or bad.

In this setting, the rate of return (ROR) provides an alternative approach that may be preferred. The ROR is the discount rate that equilibrates the returns from an investment to the cost of the investment. That is, the ROR is the interest rate where the BCR = 1 or the present value of benefits equals the present value of costs. This calculation does not require the choice of an arbitrary discount rate. To illustrate, Figure 2 shows the ROR that results from a range of benefits realized five years after $10,000 of costs were incurred. The figure shows that, if the $10,000 investment returns $15,000 in 5 years (see Figure 1), the ROR is 0.084 $(15{,}000/10{,}000)^{1/5}$-1). That is, for a

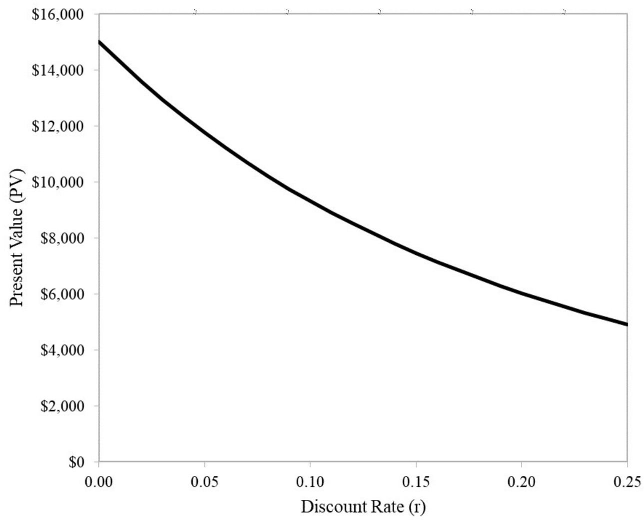

Figure 1. Illustration of How Discount Rate Affects Present Value. For this example, we assume a future value of $15,000 and a time horizon (n) of five years.

discount rate of 0.084, the present value of benefits equals the present value of costs.

The ROR can be compared to that of other government programs or well-known returns in the private sector. For example, current annual returns on money market accounts are 2% or less and the long-run annual rate of return to the U.S. stock market is about 10%. Thinking of a discount rate as the "opportunity cost of capital" and using the ROR of 8.4% from our hypothetical example, purely profit maximizing individuals would choose to "invest" their money in VR instead of a money market account, but would prefer the long-run returns from the stock market to either of the other two investments. Alternatively, the U.S. Office of Management and Budget (OMB) sets guidelines for evaluating public sector programs (OMB, 1992). Those guidelines include discount rates by time horizon that are updated each year. According to OMB (2018), current discount rates vary from 1% for 3-year horizons to 2.6% for 30-year horizons.

Dean et al.'s (2015, 2017, 2018, 2019) recent analyses of applicants to the Virginia general VR agency in SFY 2000 estimate the long run ROR of VR services for each client. They report a median annualized rate of return of 20% for clients with mental illness, 19% for clients with cognitive impairments, and 169% for clients with physical impairments. Thus, by any conventional standard, the ROI of VR services for this cohort is positive and substantial. Schmidt, Clapp, Pepper, and Stern (2019) summarize more recent ROI evaluations.

Conclusion

In this paper, we highlight key conceptual issues involved in ROI evaluations of VR programs. Most notably, we focus on estimating the benefits and costs of VR in light of the counterfactual outcomes problems. This is the most critical and demanding part of ROI analyses. We then discuss different ways of implementing ROI calculations and suggest that the ROR analysis is appealing for VR evaluations where there is no widely accepted discount rate.

There are many other critical steps involved in undertaking such an evaluation. For example, analysts must decide whether to report returns at a client or program level; what outcomes to monetize (e.g., labor market, disability insurance, others); whether the return should be measured for society, the taxpayer, the client, or some other group; what the relevant time period should be; and how to account for statistical uncertainty. These and other issues shape the details of an ROI analysis. Readers interested in a more complete and in-depth analysis of VR-ROI might turn to McGuire-Kuletz and Tomlinson (2015), and Hollenbeck (2019).

References

Aakvik, A., Heckman, J. J., & Vytlacil, E. J. (2005). Estimating treatment effects for discrete outcomes when responses to treatment vary: an application to Norwegian vocational rehabilitation programs. *Journal of Econometrics*, *125*(1-2), 15–51.

Bua-Iam, P., & Bias, T. K. (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.

Dean, D. H., & Dolan, R. C. (1991). Fixed-effects estimates of earnings impacts for the vocational rehabilitation program. *Journal of Human Resources*, *26*(2), 380–391.

Dean, D. H., Dolan, R. C., Schmidt, R., Wehman, P., Kregel, J., & Revell, G. (2002). A paradigm for evaluation of the Federal-State vocational rehabilitation program. In J. Kregel, D. H. Dean, & P. Wehman (Eds.), *Achievements and challenges in employment services for people with disabilities: The longitudinal impact of workplace supports (*pp. 1–40). Richmond, VA: Virginia Commonwealth University Rehabilitation Research and Training Center on Workplace Supports.

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*, 68–88.

Dean, D. H., & Schmidt, R. (2005a). *Determining the 'return on investment' for VR participants. Data on employment of people with disabilities: Deliverable 6.* Richmond, VA: Virginia Department of Rehabilitative Services.

Dean, D. H., & Schmidt, R. (2005b). *An outcome-based assessment of the Chapter 31 Program.* Final Report, U.S. Department of Veterans Affairs, Contract No. 101-Y27247.

Friedlander, D., Greenberg, D. H., & Robins, P. K. (1997). Evaluating government training programs for the economically disadvantaged. *Journal of Economic Literature*, *35*(4), 1809–1855.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161.

Hollenbeck, K. M. (2019). What is ROI? *The Journal of Rehabilitation Administration*, *40*(1), 5–10.

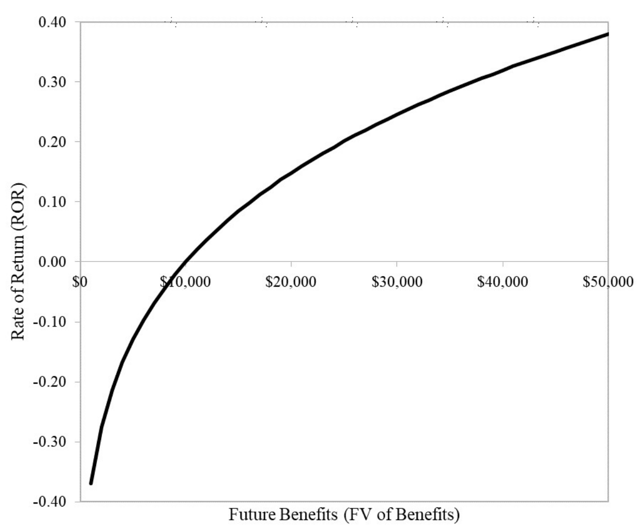

Figure 2. Illustration of How the Future Value of Benefits Affect Rate of Return. For this example, we assume the present value of costs is $10,000 and a time horizon (n) of five years.

Hollenbeck, K. M., & Huang, W. J. (2006). *Net impact and benefit-cost estimates of the workforce development system in Washington state*. Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.

Hotz, V. J. (1992). *Design analysis plan for the GAO study of the vocational rehabilitation program*. Correspondence with the GAO Program Evaluation and Methodology Division, 1–10.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86.

King, C. T., Tang, Y., Smith, T., Schroeder, D., & Barnow, B. S. (2008). *Returns from Investments in Workforce Services: Texas Statewide Estimates for Participants, Taxpayers, and Society*. Austin, TX: Ray Marshall Center for the Study of Human Resources.

LaLonde, R. J. (1995). The promise of public sector-sponsored training programs. *Journal of Economic Perspectives*, *9*(2), 149–168.

Mann, D. R., Honeycutt, T., Bailey, M. S., & O'Neill, J. (2017). Using administrative data to explore the employment and benefit receipt outcomes of vocational rehabilitation applicants years after program exit. *Journal of Vocational Rehabilitation*, *46*(2), 159–176.

Manski, C. F. & Pepper, J. V. (2018). How do right-to-carry laws affect crime rates? Coping with ambiguity using bounded-variation assumptions. *Review of Economics and Statistics, 100*(2), 232–244.

Manski, C. F., Newman, J., & Pepper, J. V. (2002). Using performance standards to evaluate social programs with incomplete outcome data: general issues and application to a higher education block grant program. *Evaluation Review*, *26*(4), 355–381.

Maryns, N., & Robertson, R. (2015). *Smart investments, real results: A net impact evaluation framework for Minnesota's workforce Development System and Initial Findings*. St. Paul, MN: Minnesota Department of Employment and Economic Development (DEED).

McGuire-Kuletz, M., & Tomlinson, P. (2015). *Return on investment and economic impact: Determining and communicating the value of vocational rehabilitation* (Institute on Rehabilitation Issues Monograph No. 38). Washington, DC: The George Washington University Center for Rehabilitation Counseling Research and Education (GW-CRCRE).

Moore, M. A., Boardman, A. E., Vining, A. R., Weimer, D. L., & Greenberg, D. H. (2004). "Just give me a number!" Practical values for the social discount rate. *Journal of Policy Analysis and Management*, *23*(4), 789–812.

Office of Management and Budget. (1992, October 29). *Memorandum for heads of executive departments and establishments: Guidelines and discount rates for benefit-cost analysis of federal programs* [Circular No. A-94 Revised]. Retrieved from: https://obamawhitehouse.archives.gov/omb/circulars_a094/

Office of Management and Budget. (2018, February 8). *Discount rates for cost-effectiveness Analysis of Federal Programs* [Revisions to Appendix C of OMB Circular A-94]. Retrieved from https:// www.federalregister.gov/documents/2018/02/08/2018-02520/discount-rates-for-cost-effectiveness-analysis-of-federal-programs/

Schmidt, R. M., Clapp, C. M., Pepper, J. V., & Stern, S. (2019). Applications of the VR-ROI project: ROI estimates for Virginia and Maryland. *Journal of Rehabilitation Administration, 40*(1), 57–72.

Smith, T., Christensen, K., & Cumpton, G. (2015). *An evaluation of local investments in workforce development: 2014 Update*. Austin, TX: Ray Marshall Center for Human Resources.

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.

U.S. Department of Education, Office of Special Education and Rehabilitative Services. (2018) *Rehabilitation Services Administration Report for Fiscal Years 2014–15, Report on Federal Activities under the Rehabilitation Act of 1973, as Amended*, Washington, DC.

U.S. Government Accountability Office. (2005). *Vocational rehabilitation: Better measures and monitoring could improve the performance of the VR program*. (GAO-05-865). Washington, DC.

U.S. Government Accountability Office. (2007). *Vocational rehabilitation: improved information and practices may enhance state agency earnings outcomes for SSA beneficiaries*. (GAO-07-521). Washington, DC.

U.S. Government Accountability Office. (June 2012). *Employment for people with disabilities: Little is known about the effectiveness of fragmented and overlapping programs*. (GAO-12-677). Washington, DC.

Uvin, J., Karaaslani, D., & White, G. (2004). *Evaluation of Massachusetts' public vocational rehabilitation program: Final report*. Boston, MA: Massachusetts Rehabilitation Commission.

Wilhelm, S. & Robinson, J. L. (2013). The economic impact of Utah's vocational rehabilitation program. *Journal of Disability Policy Studies*, *24*(3), 148–17.

Author Notes

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

We would like to thank Kevin Hollenbeck, Kirsten Rowe and two anonymous reviewers for helpful comments and the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR Grant 90DP0070) for generous funding. All errors are ours.