Program Quality and Treatment Completion for Youth Training Programs

Ana C. Dammert
Jose Galdo

DISCUSSION PAPER SERIES

I Z A

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Program Quality and Treatment Completion for Youth Training Programs

Ana C. Dammert
*Carleton University
and IZA*

Jose Galdo
*Carleton University
and IZA*

ABSTRACT

Program Quality and Treatment Completion for Youth Training Programs[*]

This paper analyzes the effects of training quality on the likelihood of treatment completion by estimating dose-response functions via a generalized propensity score. Results show a statistically positive relationship between training quality and treatment completion for youth participants in Peru.

Corresponding author:

Jose Galdo
Department of Economics
and School of Public Policy and Administration
Carleton University
1125 Colonel By Drive
Ottawa, ON K1S5B6
Canada
E-mail: jose_galdo@carleton.ca

---

1. Introduction

Unlike the vast literature on the evaluation of training programs, a relatively few number of studies have explicitly addressed the failure of participants to complete the training dosage (e.g., Heckman et al. 1998, Flores et al. 2012).[1] In fact, dropout is endemic to most voluntary training programs, as it ranges from 5 to 79 percent in developed countries (Heckman et al. 2000) and from 10 to 50 percent in developing countries (Choe et al. 2011). Accounting for dropouts has profound methodological and policy consequences: (i) it demands the application of non-experimental methods even when experimental data is available; (ii) the parameters of interest should carefully distinguish the effects of the program from the effects of training; and (iii) adjusted-dropout average treatment effects might yield very different views on the effectiveness of training programs.[2]

The small stream of literature on dropouts either estimates the causal impacts of training dosage on subsequent outcomes of interest (e.g., Flores et al. 2012, Kluve et al. 2012) or analyses the determinants of dropout behavior (e.g., Waller 2008). This study opens the "black box" of program-length exposure by studying the link between training quality and treatment completion in a developing country. Attempts to incorporate quality measures of training into the evaluation of training initiatives have been severely limited by the availability of data. Training quality affects the timing of the activities, content of the courses, and connection to prospective employers, all of which might affect the likelihood of completing a program. For instance, the average class-size in low-quality courses is much bigger than that for high-quality ones. This feature affects the in-class interaction between trainees and instructors, which might lead the

---

[1] It is important to distinguish the concept of "dropout" from "attrition." The former refers to the case where treatment units drop out from the training program, but they are still in the data. The latter refers to a case where individuals assigned to the treatment group do not remain in the data.

[2] Heckman et al. (2000) show, for instance, that accounting for dropout behavior changes the interpretation of the evidence on the effectiveness of the widely known National JTPA Study.

former to adjust their valuation of benefits and costs of training. Similarly, low-quality training centers have weak relationships with the productive sector which adversely affect trainees' ability to successfully complete, as part of the program's activities, an on-the-job training component.

For our study we used data from the PROJOVEN program, a training initiative that since 1996 has served more than 50,000 disadvantage youth in Peru. Less than half of participants complete the treatment. Following recent developments in the treatment-effects literature with continuous treatments (Hirano and Imbens 2004), we implemented a generalised propensity score (GPS) to estimate dose-response functions as the proportion of fully-treated individuals varies across different percentiles of the training quality distribution. Galdo and Chong (2012) document that sorting into courses of varying quality is greatly ameliorated in this data as treatment assignment is based on a first-come-first-serve basis.

Results show that the estimated likelihood of treatment completion increases from 41 to 65 percent across percentiles of training quality for the full sample. This positive relationship is observed for both subsamples of men and women. Important policy implications for the provision of social programs and the operation of the PROJOVEN program, in particular, emerge from this study.

2. Data and Program Institutions

The PROJOVEN program targets individuals aged 16 to 25 with poor attachment to the labor market. A distinctive feature of this program is the decentralization of the training services through market mechanisms in which public and private training institutions compete for public

funding based on the quality of the training services.[3] The treatment combines in-classroom and on-the-job training in low-skill occupations over a six-month period. PROJOVEN follows a demand-driven approach in which training is offered only for those occupations with assured labor demand from productive firms so that training institutions must set the content of the courses in strict coordination with firms' labor needs. Responsibility for the provision of training falls solely on the training institutions. A detailed description of the institutions in this program is provided in Galdo and Chong (2012) and Diaz and Jaramillo (2006). Both studies reported statistically significant earnings impacts for participants, particularly for women.

The evaluation data was composed of 1,622 individuals from five different cohorts of participants in Lima receiving treatment in 297 different courses from 1996 to 2004.[4] This data shows that only 47 percent of the trainees completed the full 6-month treatment. Large gender differences are observed in the data since 41 percent of men and 51 percent of women completed the treatment. Determinants of dropout behavior in PROJOVEN were reported in De Crombrugge et al. (2009) through probit models estimated over only one cohort of participants. Prior training exposure, formal contracts with private firms and the effectiveness of training, measured as the share of trainees working six months after the program, were reported to be statistically related to dropout behavior.

As the PROJOVEN program uses market-based approaches in the selection of training services through formal bidding processes, we were able to collect data for several variables related to the quality of the training services, including: expenditures per trainee, class size, infrastructure, equipment, teacher characteristics, curricular structure, and market knowledge. This bidding information is generated at the course level, rather than at the school or firm level,

---

[3] Similar programs were implemented in Argentina, Brazil, Chile, Colombia, Panama, and Uruguay.
[4] These data are selected from a stratified random sample of the population of participants corresponding to the first, second, fourth, sixth, and eighth cohorts of participants.

which provides large variation when computing a one-dimensional quality index through first principal component methods. All individuals attending the same training course receive the same quality scores. The estimate scores ranged from -4.87 to 4.22 with 1.75 as the standard deviation. This implies a large separation between high- and low-quality courses. For instance, average expenditures per trainee vary from US$412 to US$303 between courses located in the upper and bottom quartile of the quality distribution, a difference equivalent to almost one-third of the average expenditures per course in the program. Most important, unconditional mean differences show a positive relationship between the estimated quality measure and program length as the proportion of fully treated in the sample goes from 36 to 52 when comparing the bottom and the upper quartile of the quality index. The range of values goes from 25 to 47 and from 43 to 57 for subsamples of men and women, respectively.

The baseline evaluation data provides information on demographics, detailed labor-market variables that include labor force participation, experience, earnings, firm size, social security benefits, formal contract, labor force status transitions, and measures of previous participation in training programs. In addition, the datasets provide detailed information on dwelling characteristics including toilet facilities, house infrastructure, and household density, which are proxies for a household's long-run poverty status. Columns 1 and 2 in Table 1 show basic descriptive statistics for this rich set of variables.

3. Empirical Framework and Results

This study uses data only from PROJOVEN participants, thus we do not require assumptions about the process governing selection into the program. Sorting, however, is the main identification threat as trainees in high-quality courses might be different from their counterparts

attending low-quality courses. Using administrative data about the exact timing or order of sign up into the program, along with standard statistical tests on individual characteristics and baseline outcomes, Galdo and Chong (2012) document extensively that the first-come-first-serve treatment assignment rule essentially randomized individuals across courses of varying quality. Yet, given that we do not have complete control over all factors that might be correlated to treatment completion, we implemented generalised propensity scores (GPS) with continuous treatments (Hirano and Imbens 2004) to evaluate the impact of training quality on the likelihood of program completion.

Let $Y_i(q)$ be the potential outcome of individual i under treatment level Q where in our case q denotes the quality of the treatment received. For any individual only one component of Q can be observed in the data. The data we observe for each unit is therefore *(Y,Q,X)*, with *X* a vector of pre-treatment covariates and $Y_i = Y_i(Q)$ the observed outcome for the level of treatment actually received. The key identifying assumption follows the standard conditional independence assumption used in the binary-treatment literature but this time it is weakly defined at the 'local' treatment level of interest for all q,

$$Q_i \perp \{Y_i(q)\} \mid G_i(q,x) \qquad (1)$$

where $G_i(q,x) = \Pr(Q = q \mid X = x)$ is the generalised propensity score (GPS), or the conditional probability of receiving a particular level of treatment q conditional on the rich set of baseline covariates *X*.

We follow the empirical approach outlined in Hirano and Imbens (2004) and use a normal distribution for the treatment given a rich set of covariates, $Q_i \mid X \sim N(\beta_o + \beta_1 ' X_i, \sigma^2)$

.Thus, the estimated GPS is calculated as $\hat{G}_i = \dfrac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\dfrac{1}{2\hat{\sigma}^2}(Q_i - \hat{\beta}_0 - \hat{\beta}_1 ' X_i)^2 \right)$, where

$[\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2]$ are estimated by ordinary least squares (OLS). The normal distribution assumption is valid as long as the covariates are balanced after adjusting for the estimated GPS values. Next, we estimate the conditional expectation of the outcome, $E(Y_i \mid Q_i, G_i)$, by using a (flexible) parametric regression function:

$$E(Y_i \mid Q_i, G_i) = \beta_0 + \beta_1 Q_i + \beta_2 Q_i^2 + \beta_3 Q_i^3 + \beta_4 G_i + \beta_5 G_i^2 + \beta_6 G_i^3 + \beta_7 Q_i.G_i \qquad (2)$$

Finally, given the estimated parameters in (2), the dose-response function at treatment level q is estimated as

$$\hat{\beta}(q) = \frac{1}{n} \sum_{i=1}^{n} \left[\hat{\beta}_0 + \hat{\beta}_1 q_i + \hat{\beta}_2 q_i^2 + \hat{\beta}_3 q_i^3 + \hat{\beta}_4 G_i + \hat{\beta}_5 G_i^2 + \hat{\beta}_6 G_i^3 + \hat{\beta}_7 q_i.G_i \right] \qquad (3)$$

where q takes 20 different percentiles i.e., 5 percent intervals, corresponding to the sample distribution of the training quality index.

Covariate balance was evaluated by *blocking* on both the training quality and the estimated GPS (Hirano and Imbens 2004). After considering three treatment quality groups corresponding to the top, second and third, and fourth quartiles of the quality distribution, and splitting the individuals in a given quality group in five blocks, defined by quintiles of the estimated probability of being in that group, we cannot reject the null hypothesis of equality of means for *all* variables across *all* quality groups simultaneously. Table 1 shows the weighted average (over the five blocks in each treatment-level group) of the covariate mean differences between the particular treatment-level group and all other groups. By comparing columns 3-5 to columns 6-8 one observes the success of the GPS strategy for comparing comparable individuals simultaneously across all treatment groups. Similarly, we also imposed a joint support region by *blocking* on both the training quality and the estimated GPS (Gerfin and Lechner 2002). Less than four percent of the treatment units are out of the joint support region.

Figures 1A-1C plot the estimated continuous dose-response functions for the full sample and men and women subsamples, respectively. The figures also show 95 percent confidence intervals estimated with 500 bootstraps replications that account for all estimation steps including the estimation of the GPS and the imposition of the common support condition. By looking at Figure 1A one observes a steady result: the likelihood of completing the treatment is an increasing function of the quality of the training. The range of values for the continuous dose-response function increases systematically when moving along the percentiles of training quality from 0.41 to 0.65 in the full sample. Statistically significant effects emerge when comparing individuals in the [20, 40] versus [60, 90] quality intervals of the dose-response function. When analyzing the results by gender in Figures 1B-1C, one observes slightly steeper quality-completion profiles for men than that for women. Likewise, we observe statistically significant differences when comparing trainees in the [20, 40] versus [50, 85] quality intervals for men, and [30, 45] versus [55, 75] quality intervals for women. At the same time, the estimated 95 percent confidence intervals are particularly thick at the extremes of the distribution where the data is sparse, so no statistically significant differences are observed when comparing individuals in the bottom and upper side of the quality distribution.

4. Sensitivity tests

Sensitivity analyses were considered to test the robustness of our results. We implemented alternative higher order terms and interactions in the specification of the GPS model and alternative functional form specifications for the conditional expectation (2). Moreover, we changed the definition of the outcome variable by considering fully treated to those individuals who complete at least 80 percent of the program. All qualitative results hold. In the spirit of

Heckman and Hotz (1989) we also implemented a placebo test to (indirectly) address the adequacy of the identification assumption in the context of this data. We considered training length exposure (in hours) prior to the PROJOVEN program, an outcome that cannot be possibly affected by treatment quality unless selection bias is affecting our methodology. As expected, no statistical significant results emerge across different percentiles of treatment quality.

5. Conclusions

A forthright conclusion of this study is that program quality affects treatment completion in voluntary training programs. This result concurs with the findings in the school literature in developing countries (Hanushek et al. 2008). Training quality might affect a trainee's valuation of expected benefits and costs of training as well as the connection between training centers and productive firms which affect trainees' ability to complete on-the-job training requirements. Improving the quality of the training programs can discourage or mitigate the failure of participants to complete the training programs, and thus ensure the full benefits of completion. An interesting extension to this study would be to analyze the role of particular training attributes such as expenditures and class size on the likelihood of treatment completion.

References

Choe, C., A. Flores-Lagunes and S. Lee. 2011. "Do Dropouts Benefit from Training Programs? Korean Evidence Employing Methods for Continuous Treatments." IZA DP No. 6064.

De Crombrugghe, D., H. Espinoza and H. Heijke. 2009. "Why do youth job training programmes participants drop out? The case of Projoven-Peru." Manuscript.

Diaz, J. and M. Jaramillo. 2006. "An Evaluation of the Peruvian Youth Labor Training Program-PROJOVEN," Inter-American Development Bank, WP# 1006

Flores, C, A. Flores-Lagunes, A. Gonzales, T. Neumann. 2012."Estimating the Effects of Length of Exposure to a Training Program: The Case of Job Corps", *Review of Economics and Statistics*, 94(1): 153-171.

Galdo, J. and A. Chong. 2012. "Does the Quality of Public-Sponsored Training Programs Matter? Evidence from Bidding Processes Data." *Labour Economics,* 19(6): 970-986.

Gerfin, M. and M. Lechner. 2002."A Microeconometric Evaluation of the Active Labour Market Policy in Switzerland," *Economic Journal*, 112(482):854-893.

Hanushek, E. A., V. Lavy, and K. Hitomi, 2008. "Do Students Care about School Quality? Determinants of Dropout Behaviour in Developing Countries," *Journal of Human Capital* 2(1):69-105.

Heckman, J., J. Smith and C. Taber. 1998. "Accounting for Dropouts in Evaluations of Social Programs," *Review of Economics and Statistics,* 80(1): 1-14.

Heckman, J., and Hotz, J. 1989. "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," *Journal of the American Statistical Association*, 84:408, 862-874.

Heckman, J., N. Hohmann, J. Smith and M. Khoo. 2000. "Substitution and Drop Out Bias in Social Experiments: A Study of an Influential Social Experiment," *Quarterly Journal of Economics,* 115(2): 651-694.

Hirano, K, and G. Imbens. 2004. "The Propensity Score with Continuous Treatments", In A. Gelman and X. Li eds., Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives. John Wiley and Sons.

Kluve, J., H. Schneider, A. Uhlendorff, and Z. Zhao. 2012. "Evaluating Continuous Training Programs Using the Generalized Propensity Score," *Journal of the Royal Statistical Society,* Forthcoming.

Waller, M. 2008. "Further Training for the Unemployed: What Can We Learn About Dropouts from Administrative Data?", Manuscript.

Table 1 : Balancing test given the generalized propensity score (GPS) for participants
### PROJOVEN, Lima 1996-2004

Baseline Variables	mean	stdev	Unadjusted Q^l	Unadjusted Q^m	Unadjusted Q^h	GPS Adjusted Q^l	GPS Adjusted Q^m	GPS Adjusted Q^h
Socio-demographic								
sex (1=men)	0.429	0.495	-0.047	-0.003	**0.060**	-0.059	0.000	0.002
age	19.663	2.389	-0.228	0.046	0.193	-0.043	0.056	-0.009
completed high school	0.852	0.354	-0.023	-0.008	0.038	0.000	-0.007	0.013
less than high school	0.088	0.284	0.008	0.004	-0.015	-0.002	0.003	0.002
single	0.912	0.283	-0.021	-0.001	0.028	0.000	-0.000	-0.000
have children	0.143	0.350	0.020	0.002	-0.029	-0.002	0.001	-0.001
Poverty Proxies								
family size	6.248	2.617	0.139	-0.104	0.010	0.096	-0.083	0.100
household density	3.099	1.721	0.046	-0.165	0.182	0.098	-0.136	0.054
earthen floor	0.618	0.486	0.003	0.111	-0.019	-0.007	0.016	-0.015
matted roof	0.640	0.480	0.031	-0.014	-0.015	0.008	-0.014	0.016
matted walls	0.339	0.471	-0.011	**0.055**	-0.051	-0.027	0.045	-0.041
flush toilet	0.648	0.477	-0.027	0.019	0.004	-0.013	0.018	-0.015
Labort-Market Characteristics								
experience (years)	3.221	2.761	-0.104	0.072	0.019	-0.025	0.080	-0.086
training before PROJOVEN	0.226	0.418	-0.000	-0.001	0.001	0.000	-0.004	0.002
duration of prior training (hours)	57.403	189.27	0.867	-3.94	0.665	-0.232	-4.000	1.489
monthly earnings (real, US$)	26.732	4.049	**-4.632**	1.165	3.601	-0.689	0.008	0.284
employment	0.516	0.499	-0.030	0.002	0.030	-0.004	0.000	-0.007
work as an salaried worker	0.273	0.440	**-0.012**	0.027	**-0.030**	-0.001	0.000	-0.001
work as self-employed	0.103	0.304	-0.047	-3.940	-0.030	-0.016	0.024	-0.024
work in large-size firm	0.029	0.169	-0.010	0.005	0.004	-0.006	0.006	-0.005
work with formal contract	0.023	0.151	0.001	0.000	-0.001	0.005	-0.000	-0.006
LaborMarket Transitions								
empl→empl	0.413	0.492	-0.036	0.009	0.031	-0.006	0.007	-0.007
unempl→ unempl	0.163	0.370	0.009	-0.018	0.017	0.010	-0.015	0.005
olf→olf	0.179	0.383	0.046	-0.014	-0.031	0.016	-0.020	0.030
olf→unemp	0.062	0.242	-0.015	0.007	0.003	-0.007	0.011	-0.010
olf→empl	0.078	0.269	0.008	-0.004	-0.010	-0.001	-0.005	-0.000
N	1622	1622	365	802	455	365	802	455

Notes: Bold numbers indicate significance at the 5% level. Balancing test is based on Hirano and Imbens's (2004) method.

Q^l stands for low-quality (1st quartile) , Q^m for medium-quality (2nd and 3rd quartiles), and Q^h for high-quality (upper quartile).

Labor-market transitions are estimated as a pair of statuses. The second is always the status in the month of the start of program

(baseline period), while the first correspond to the status three months before the start of the program.

Figure 1: Dose-Response Function for Training Quality and Treatment Completion
PROJOVEN, Lima, 1996 to 2004

1A: ALL

1B: Men

1C: Women
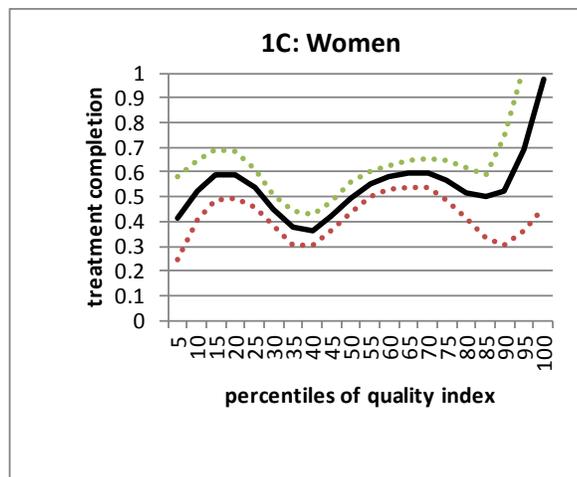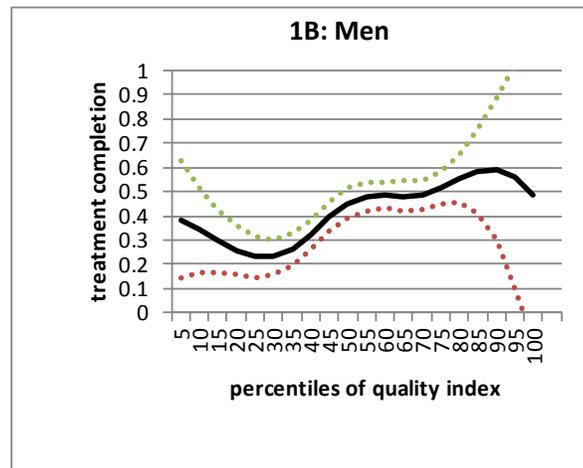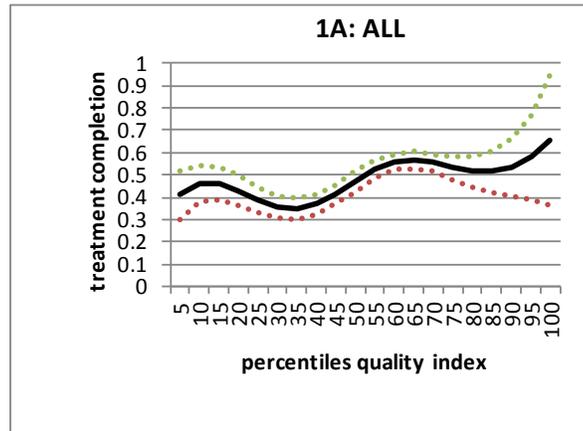
Notes: Dashed lines are bounds for 95% bootstrapped confidence intervals.
N=1622, 697 men, 925 women. The dependent variables is 1 for those who
completed the 6-month treatment, 0 otherwise. The quality index is estimated
by factor analysis methods. The estimation of the dose -response functions
follows Hirano and Imbens (2004) approach.

11