Evaluating continuous training programs using the generalized propensity score[1]

Jochen Kluve
RWI Essen and IZA Bonn

Hilmar Schneider
IZA Bonn

Arne Uhlendorff
DIW Berlin, IZA Bonn and FU Berlin

Zhong Zhao[2]
IZA Bonn

February 28, 2007
Preliminary and incomplete – please do not quote!

Abstract

This paper assesses the dynamics of treatment effects arising from variation in the duration of training. We use German administrative data that have the extraordinary feature that the amount of treatment varies continuously from 1 day to 720 days (i.e. 2 years). This feature allows us to estimate a continuous dose-response function that relates each value of the dose, i.e. days of training, to the individual post-treatment employment probability (the response). The dose-response function is estimated after adjusting for covariate imbalance using the generalized propensity score, a recently developed method for covariate adjustment under continuous treatment regimes. Our results indicate an increasing dose-response function for treatments of up to 360 days, and a similarly steady decline afterwards.

JEL Codes: C21, J68

Keywords: Training, program evaluation, continuous treatment, generalized propensity score.

---

[1] The data used in this paper originate in the evaluation of continuous training programs as part of the evaluation of the so-called Hartz-Reforms. The corresponding report by IZA et al. 2007 (cf. references) contains details. We wish to thank Ulf Rinne for excellent research assistance. The usual disclaimer applies.
[2] Correspondence: kluve@rwi-essen.de, schneider@iza.org, auhlendorff@diw.de, zhao@iza.org

1. Introduction

Over recent years there has been an increasing amount of research on the effectiveness of labor market training programs in many countries. Training programs represent the "classic" type of so-called active labor market programs, due to their objective of enhancing participants' employment prospects by increasing their human capital. While the evidence on early training programs in the 1970s and 1980s showed relatively optimistic results, the more recent research from the 1990s and 2000s – generally based on much better data and advanced econometric methods – points to the result that training programs seem to be modestly effective at best (Heckman et al. 1999, Kluve 2006). Adding to this general finding, one recent line of research shows that positive treatment effects may only materialize in the long run, and that program effectiveness can show a considerable dynamic ranging from often severe short-term locking-in effects to long-term gains in employment prospects (e.g. Lechner et al. 2004).

In this paper we contribute to the literature on training programs by focusing on the dynamics inherent to the provision of training, i.e. we study the treatment effects that arise from variation in the treatment duration. We implement this analysis on the basis of data on training programs in Germany. The key feature of the data is the fact that the treatment duration varies almost continuously from approximately 1 week duration up to approximately 24 months, i.e. 2 years. We focus on programs in which no specific degree is acquired as part of the program requirements – this is the majority of training programs in Germany (about 70% in 2000, for instance). Training programs leading to the acquisition of a degree are not considered, since the degree requirement generates discontinuities in the distribution of treatment durations, and the objective of the analysis in this paper is to estimate the employment outcomes associated with each level of a continuous treatment.

The evaluation question that corresponds to the continuous administering of training is how effective (relative to each other) are training programs with different durations? This assessment of the dynamics of treatment duration essentially amounts to estimating a dose-response function. In this paper we therefore estimate the responses – i.e. the employment probability – that correspond to specific values of continuous doses – i.e. training of a particular length. In a setting in which doses are not administered under experimental conditions, estimation of a dose-response function is possible using the generalized propensity score (GPS). The GPS for continuous treatments is a straightforward extension of the well-established and widely used propensity score methodology for binary treatments (Rosenbaum and Rubin 1983) and multi-valued treatments (Imbens 2000, Lechner 2001). The GPS methodology is developed in Hirano and Imbens (2004) and Imai and van Dyk (2004). To our knowledge, our paper is the first application of the GPS in the context of evaluating active labor market policy.

The paper is organized as follows. Section 2 describes the methodology of estimating a dose-response function to evaluate a continuous policy measure, adjusting for the generalized propensity score. Section 3 gives details on the data and the treatment we study. The fourth section contains the application and discusses the results of balancing the sample by blocking on the GPS as well as our estimates of the dose-response function. Section 5 concludes.

2. Bias removal using the Generalized Propensity Score

Research in program evaluation in recent years has made comprehensive use of matching methods[3]. In the absence of experimental data, which is largely the case, the popularity of matching is due to its intuitively appealing technique of mimicking an experiment *ex post*. The standard case, which is also appropriate for the majority of applications, considers a binary treatment. One of the key results that have made matching such an attractive empirical tool is developed in Rosenbaum and Rubin (1983), who show that, rather than conditioning on the full set of covariates, conditioning on the propensity score – i.e. the probability of receiving the treatment given the covariates – is sufficient to balance treatment and comparison groups.

Subsequently, the literature has extended propensity score methods to the cases of multi-valued treatments (Imbens 2000, Lechner 2001) and, more recently, continuous treatments (Imbens 2000, Behrmann, Cheng and Todd 2004, Hirano and Imbens 2004, Imai and van Dyk 2004). In this paper, we build on the approach developed by Hirano and Imbens (2004) who propose estimating the entire dose-response function (DRF) of a continuous treatment. This approach fits perfectly with the objective of our paper, since we are interested in the response – i.e. the post-treatment employment probability – associated with each value of the continuous dose, i.e. the days spent in training.

[3] Cf. *inter alia* the overview given in Augurzky and Kluve (2007) and articles in a recent symposium on the econometrics of matching in *The Review of Economics and Statistics* (2004, Vol. 86, No. 1, pp. 1-194), in particular the survey article by Imbens (2004).

2.1 The GPS methodology

Hirano and Imbens (2004) develop the GPS methodology in the context of the potential outcomes model for estimation of causal effects of treatments. In what follows we closely follow their presentation. Suppose we have a random sample of units, indexed by $i=1,\ldots,N$. For each unit i there exists a set of potential outcomes $Y_i(t)$ for $t \in \Im$, referred to as the unit-level dose-response function. In the continuous case, $\Im$ is an interval $[t_0, t_1]$, whereas in the binary case it would be $\Im = \{0,1\}$. Our objective is to estimate the average dose-response function (ADRF) $\mu(t) = E[Y_i(t)]$. For each unit i, we observe a vector of covariates X_i, the level T_i of the treatment that unit i actually receives, with $T_i \in [t_0, t_1]$, and the potential outcome corresponding to the level of treatment received, $Y_i = Y_i(T_i)$. In the remainder of this section the subscript i will be omitted to simplify notation.

The key assumption of Hirano and Imbens (2004) generalizes the *unconfoundedness* assumption for binary treatments made by Rosenbaum and Rubin (1983) to the multi-valued case:

(1) $\quad Y(t) \perp T \mid X \;\; \textit{for all}\; t \in \Im.$

Hirano and Imbens (2004) refer to this as *weak unconfoundedness*, since it only requires conditional independence to hold for each value of the treatment, rather than joint independence of all potential outcomes. Calling $r(t,x) = f_{T|X}(t \mid x)$ the conditional density of the treatment given the covariates, the *Generalized Propensity Score (GPS)* is defined as

(2) $\quad R = r(T, X).$

The GPS has a balancing property similar to the balancing property of the propensity score for binary treatments. Within strata with the same value of $r(t,X)$ the probability that $T=t$ does not depend on the value of X, i.e. the GPS has the property that $X \perp \mathbf{1}\{T=t\} \mid r(t,X)$. Hirano and Imbens (2004) emphasize that this is a mechanical implication of the definition of the GPS and does not require unconfoundedness. In combination with unconfoundedness, however, it implies that assignment to treatment is unconfounded given the GPS. That is, Hirano and Imbens (2004) prove that, if assignment to treatment is weakly unconfounded given covariates X, then it is also weakly unconfounded given the Generalized Propensity Score.

Given this result, it is possible to use the GPS to remove bias associated with differences in covariates in two steps. The first step is to estimate the conditional expectation of the outcome as a function of two scalar variables, the treatment level T and the GPS R, i.e.

(3) $$\beta(t,r) = E[Y \mid T = t, R = r].$$

The second step is to estimate the DRF at each particular level of the treatment. This is implemented by averaging the conditional expectation function over the GPS at that particular level of the treatment,

(4) $$\mu(t) = E[\beta(t, r(t, X))].$$

The procedure does not average over the GPS $R=r(T,X)$, but instead it averages over the score evaluated at the treatment level of interest $r(t,X)$. Hirano and Imbens (2004) also emphasize that the regression function $\beta(t,r)$ does not have a causal interpretation, but that $\mu(t)$ corresponds to the value of the DRF for treatment value t, which compared to another treatment level t' does have a causal interpretation.

2.2 Implementation

In the practical implementation of the methodology outlined in the previous section, we use a normal distribution for the treatment given the covariates

(5) $$T_i \mid X_i \sim N(\beta_0 + \beta_1' X_i, \sigma^2),$$

which we estimate by ordinary least square regression (OLS) (adjusting properly $\hat{\sigma}^2$).[4] The estimated GPS is calculated as

(6) $$\hat{R}_i = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{1}{2\hat{\sigma}^2}(T_i - \hat{\beta}_0 - \hat{\beta}_1' X_i)^2\right).$$

In the second stage we calculate the conditional expectation function of Y_i given T_i and R_i as a

[4] It is possible to assume other distributions than the normal distribution, and estimate the GPS by other methods such as maximum likelihood. The key point here, however, is to make sure that the covariates are balanced after adjusting for the GPS: As long as sufficient covariate balance is achieved, the exact procedure of estimating the GPS is of secondary importance.

flexible function of its two arguments. Our empirical approach uses the following approximation.

(7) $$E[Y_i \mid T_i, R_i] = \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 T_i^3 + \alpha_4 R_i + \alpha_5 R_i^2 + \alpha_6 R_i^3 + \alpha_7 T_i R_i + \alpha_8 T_i^2 R_i + \alpha_9 T_i R_i^2 .$$

For each individual the observed T_i and estimated GPS $\hat{R}_i$ is used, and the equation is estimated by OLS. Given the estimated parameters in the second stage, we estimate the average potential outcome at treatment level *t* as

(8) $$\begin{aligned} E\hat{[Y(t)]} = \frac{1}{N}\sum_{i=1}^{N} (&\hat{\alpha}_0 + \hat{\alpha}_1 t + \hat{\alpha}_2 t^2 + \hat{\alpha}_3 t^3 + \hat{\alpha}_4 \hat{r}(t, X_i) + \hat{\alpha}_5 \hat{r}(t, X_i)^2 + \hat{\alpha}_6 \hat{r}(t, X_i)^3 \\ &+ \hat{\alpha}_7 t \hat{r}(t, X_i) + \hat{\alpha}_8 t^2 \hat{r}(t, X_i) + \hat{\alpha}_9 t \hat{r}(t, X_i)^2) . \end{aligned}$$

The entire dose-response function can then be obtained by estimating this average potential outcome for each level of the treatment. In our application, we use bootstrap methods to obtain standard errors that take into account estimation of the GPS and the α parameters.

2.3 Testing for balancing of covariates

Just as in the case of a binary treatment also in the continuous case it is crucial to evaluate how well adjustment for the GPS works in balancing the covariates, i.e. if the specification for estimation of expression (5) is adequate. Whereas in the binary case the typical approach is to compare the covariate means for the treated and control units before and after matching, testing for covariate balance is more difficult with continuous treatments.

Hirano and Imbens (2004) propose *blocking* on both the treatment variable, i.e. participation in training in our case, and on the estimated GPS. We implement this by first dividing the sample into three groups according to the distribution of treatment length, cutting at the 30th and 70th percentile of the distribution. Within each group we evaluate the GPS at the median of the treatment variable. Then, in a second step we divide each group into five blocks by the quintiles of the GPS evaluated at the median, considering only the GPS distribution of individuals in that particular group.

Within each of these blocks we calculate the difference-in-means of covariates with respect to individuals that have a GPS such that they belong to that block, but have a treatment level different from the one being evaluated. This procedure is testing if for each of these blocks the covariate means of the individuals belonging to the particular treatment-level group are significantly different from those of the individuals with a different treatment level, but similar GPS. A weighted average over the five blocks in each treatment-level group can be used to calculate the t-statistic of the differences-in-means between the particular treatment-level group and all other groups. The procedure needs to be repeated for each treatment-level group and for each covariate. If adjustment for the GPS properly balances the covariates, we would expect all those differences-in-means to not be statistically different from zero.

3. Data

In this paper we use a sample of a particularly rich administrative data set, the Integrated Employment Biographies (IEB) of the German Federal Employment Agency FEA (Bundesagentur für Arbeit). The data contain detailed daily information on employment subject to social security contributions, including occupational and sectoral information, receipt of transfer payments during periods of unemployment, job search activity, and participation in different programs of ALMP. Furthermore, the IEB comprise a large variety of covariates like age, education, disability, nationality and regional indicators.

Training participants in the programs we consider learn specific skills required for a certain vocation (e.g. computer-aided design for a technician/tracer) or receive qualifications that are of general vocational use (e.g. MS Office, computer skills). Numerically, these program types constitute the most important ones among all publicly financed training programs: In 2002, roughly 70% of all participants in training programs were assigned to this type (Schneider and Uhlendorff 2006, IZA et al. 2007).

We focus on men only. Our sample of participants consists of about 265 unemployed persons per quarter entering the program during the years 2000 and 2001, i.e. we observe approximately 2100 program participants. The data allow us to draw conclusions on the average participant starting a program during this time period. The programs comprise both occupation-specific training programs ("berufsbezogene Weiterbildung") and general training programs ("berufsübergreifende Weiterbildung"). The core feature of these training programs is the fact that treatment provision is a continuous variable, since the length of training varies from approx. 1 week up to 36 months. We exclusively focus on programs that do not lead to the acquisition of a degree, as the degree requirement would likely create discontinuities in the distribution of the length of treatment.

We discard observations with treatment duration below 10 days, since such short durations arguably do not imply a serious attempt at finishing the program. Durations above 720 days are also discarded, since only very few observations are available. We do not consider durations of length zero, i.e. no non-treated individuals are included. Instead, we focus on the average responses of those individuals that did receive some treatment. Figure 1 shows the distribution of treatment durations. We observe that the majority of programs last up to 360 days (the median is around 180 days). Two peaks exist at durations of 180 days and 360 days, respectively, while within the first year there is substantial probability mass at all possible treatment durations.

[Figure 1 about here]

The responses, i.e. the outcome variables of interest are (i) the employment probability at time 1 year after exit from the program, and (ii) the employment probability at time 3 years after entry into the program. Table 1 presents summary statistics of the two outcome variables and the covariates. As Table 1 shows, the data contain a large number of covariates. In particular, we can use information on numerous variables that have been identified in the program evaluation literature to be important determinants of selection into a program: This comprises detailed data on citizenship and educational background including vocational education. Moreover, we have very detailed information on pre-treatment employment histories as well as regional indicators. Given the richness of the covariates along with the fact that we focus on participants only, rather than on a treatment vs. no-treatment comparison, the assumption of unconfoundedness seems entirely reasonable.

[Table 1 about here]

The participants are on average 37 years old, around 7% of them are handicapped and 13% do not have the German citizenship. The participants are on average relatively low-skilled: more than 60% did not get further than the first stage of secondary level education, around 35% do not have any vocational degree, and only a minority (6%) has obtained a university degree. Before entering a program the participants were on average unemployed for 9 months, and their previous employment lasted for about 20 months. The individuals for whom we observe a wage for their last employment earned around 50€ per day. For the previous employment history we construct eight variables describing the share of time spent in employment and unemployment, respectively, during each of the four years before entering the program. Three years after program entry as well as one year after the program ended around 35% of the participants are employed.

4. Empirical results

4.1 Estimates from a Probit Model

As mentioned in Section 3, in this paper we consider two outcome variables: one is the employment probability at the point in time 3 years after the participants entered into the program, and the second one is the employment probability at the point in time 1 year after the participants exited from the program.

Before presenting results for the GPS, we explore first the relationship between post-treatment employment probability and the duration of treatment using a Probit model. Table 2 displays the raw relationship between the employment probability at 3 years after entering into the program and treatment duration without adjusting for other variables. It shows that there is a positive correlation between employment probability and treatment duration, and a negative correlation between employment probability and the square of the treatment duration. However, if more higher order terms, such as a cube of the treatment duration, are included into the model, these correlations become insignificant, which is probably because of multicollinearity.

[Table 2 about here]

Table 3 presents results after controlling for more variables, such as educational level, pre-program earnings and employment history, citizenship and regional dummies. Again, we observe a positive correlation between the employment probability and treatment duration and a negative correlation between the employment probability and the square of the treatment duration. These relationships are also not robust to adding cubic and higher order terms to the model.

[Table 3 about here]

Similar to the analyses of the raw relationship between treatment and the outcome 3 years after entry, we can investigate the raw relationship between treatment and the employment probability at time 1 year after exit from the program. In this case, as Tables 4 and 5 show, the Probit models do not find any significant relationship between employment probability and treatment duration, whether controlling for additional variables or not.

[Tables 4 and 5 about here]

However, it is worth noting that a regression type analysis such as this Probit model may rely on extrapolation, compare incomparable observations, and have greater risk of mis-specifying the model. All of these could potentially bias the estimates. Propensity score methods can alleviate these potential problems to some extent.

4.2 Estimation of the GPS and testing for balancing of covariates

Our implementation of the generalized propensity score follows the procedure outlined in Hirano and Imbens (2004) and adapted to our context as presented in section 3 above. We first estimate the conditional distribution of the length of the training program (treatment) by applying OLS.[5]

To assess the balancing property of the GPS we compare the distribution of covariates between different groups. Group 1 includes individuals with a treatment level between 10 and 148 days, group 2 ranges from 149 to 264 days and group 3 from 265 to 720 days. For each of the covariates we test whether the difference in means of one group compared to the other two groups is significantly different. In the left part of Table 6 the corresponding t-statistics are reported. Without adjustment 62 of 122 t-statistics are greater than 1.96, indicating a clearly unbalanced distribution of covariates.

[Table 6 about here]

In the second step, we calculate the corresponding t-statistics for the GPS-adjusted sample. To do this, we evaluate the GPS for each individual at the median of the three groups, i.e. at the length of 86 days, 182 days, and 355 days. For each of the three groups, we discretize the GPS by using five blocks, evaluated by the quintiles of the GPS within each group. In other words, we calculate for the first group, consisting of individuals with a treatment ranging from 10 to 148 days, the GPS evaluated at the median of this group (86 days). The distribution of the GPS $r(86, X_i)$ is discretized into five blocks using the quintiles of the

[5] The estimation coefficients are not reported here but are available from the authors upon request.

distribution. For the first group, this leads to the intervals [0.00018, 0.0015], [0.0015, 0.00238], [0.00238, 0.00283], [0.00283, 0.00316] and [0.00316, 0.00395]. To assess the balancing of the adjusted sample, members of the first group with a GPS in the first range are compared with individuals who are not member of the first group, i.e. who have a different level of treatment, but who have a GPS $r(86, X_i)$ lying in the first interval as well. For each group, this implies five mean differences and five standard errors. The t-statistics reported in the right part of Table 6 correspond to the mean difference for each group. To calculate these mean t-statistics the corresponding differences and standard errors of the five blocks are weighted by the number of observation.

In contrast to the unadjusted sample, we observe no t-statistics larger than 1.96. Moreover, only two t-statistics are larger than 1.645. These results indicate that the balance of the covariates is clearly improved by adjustment for the GPS.

4.3 Estimating the dose-response function

The final step of our empirical analysis consists in estimating the GPS-adjusted dose-response function. The main results are presented in Table 7 and Figures 2 and 3. Table 7 displays the DRF as treatment effects in steps of 30 days of the dose, i.e. duration of the training program. Figures 2 and 3 plot the curve of the DRF in steps of 7 days for the two outcome variables. Standard errors are bootstrap standard errors from 2,000 replications.

[Table 7 about here]

As the two figures show, the dose-response functions for the two outcome variables considered have similar shapes. One important finding is that treatment effects are increasing significantly during the first three months, and after that the treatment effects are somehow flatted out, with a slight indication of a dip between durations of 26 weeks and 52 weeks. The participants who have finished one year of training will gain most in terms of their employment probability. However, after 1 year of training, additional training will not translate into a higher employment probability, as the DRF displays a decrease in the treatment effect. If the training had lasted for roughly 1 year and half, the treatment effects would no longer be significantly different from 0. This finding, however, is partly due to the rather large confidence intervals around treatment effect estimates for durations of more than one year, which in turn result from the small number of observations with long durations.

[Figures 2 and 3 about here]

It is also interesting to compare the GPS results with the Probit estimates. For the employment probability at time 3 years after entry into the program, the result from the GPS estimates is consistent with the one from Probit model; both find an inverse-U relationship between treatment effect and treatment duration. For the employment probability at time 1 year after exit from the program, the result from the GPS differs from the Probit model. While the latter shows no significant relationship between treatment effect and treatment duration, the GPS estimates again point to an inverse-U shaped relationship.

5. Conclusions

In this paper, we study the effect of continuous training programs on the post-treatment employment probability, using a particular data set that contains information in training duration in days for a period of about 1 week to 2 years. In particular, we are interested in estimating the dose-response function at all possible treatment durations. We implement this using the recently developed generalized propensity score for continuous treatments.

Our findings indicate that the DRF has an inversed-U shape. The first three months of a training program appear to be the most effective period to improve the employment probability. After three months, further training can only increase the employment probability marginally, indicating even a slight dip after six months, and the treatment effect in terms of employment probability peaks at around 1 year of training. After that, the employment probability is decreasing strongly with further treatment duration. If training lasts more than a year and half, the gain from the program becomes insignificant, though large confidence bands due to small number of observations exacerbate a precise estimation of this effect.

References

Augurzky, B. and J. Kluve (2007), "Assessing the performance of matching algorithms when selection into treatment is strong", *Journal of Applied Econometrics*, forthcoming.

Behrman, J., Y. Chen and P. Todd (2004), "Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach", *Review of Economics and Statistics* 86, 108-132.

Heckman, J.J., R.J. LaLonde and J.A. Smith (1999), "The economics and econometrics of active labour market programs", in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics* 3, Elsevier, Amsterdam.

Hirano, K. and G.W. Imbens (2004), "The Propensity Score with Continuous Treatments", in: Andrew Gelman and Xiao-Li Meng eds, *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, Wiley.

Imai K. and D.A. van Dyk (2004), "Causal Inference with General Treatment Regimes: Generalizing the Propensity Score", *Journal of the American Statistical Association* 99, 854-866.

Imbens, G.W. (2000), "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika* 87, 706-710.

Imbens, G.W. (2004), "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review", *Review of Economics and Statistics* 86, 4-29.

IZA, DIW, infas (2007), *Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission – Modul 1b: Förderung beruflicher Weiterbildung und Transferleistungen, Endbericht Juni 2006*. BMAS Forschungsbericht (Research report of the Federal Ministry for Labour and Social Affairs), forthcoming.

Kluve, J. (2006), "The effectiveness of European Active Labour Market Policy", *IZA Discussion paper* 2018.

Lechner, M. (2001), "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption", in: M. Lechner and F. Pfeiffer, eds., *Econometric Evaluation of Labour Market Policies*, Physica, Heidelberg.

Lechner, M., R. Miquel and C. Wunsch (2004), "Long-Run Effects of Public Sector Sponsored Training in West Germany", *IZA Discussion paper* 1443.

Rosenbaum, P.R. and D.B. Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika* 70, 41-55.

Schneider, H. and A. Uhlendorff (2006), "Die Wirkungen der Hartz-Reform im Bereich der beruflichen Weiterbildung", *Journal for Labour Market Research*, forthcoming.

Table 1. Summary Statistics

	(1) Mean	(2) SD
Age	37.32	12.23
Disability		
Disability low degree	0.06	-
Disability medium degree	0.01	-
Disability high degree	0.01	-
Citizenship		
Foreigner EU	0.02	-
Foreigner Non-EU	0.11	-
Educational Attainment		
No graduation	0.13	-
First stage of secondary level	0.48	-
Second stage of secondary level	0.25	-
Advanced technical college entrance qualification	0.04	-
General qualification for university entrance	0.1	-
Vocational Attainment		
No vocational degree	0.35	-
In-plant training	0.53	-
Off-the-job training, vocational school, technical school	0.05	-
University, advanced technical college	0.06	-
Employment history		
Previous Unemployment Duration in month	9.24	7.75
Duration of last employment in month	19.5	27.97
Log(wage) of last employment	3.89	0.56
No last employment observed	0.08	-
Share of days in employment, 1st year before program start	0.18	0.28
Share of days in employment, 2nd year before program start	0.36	0.41
Share of days in employment, 3rd year before program start	0.41	0.42
Share of days in employment, 4th year before program start	0.43	0.43
Share of days in unemployment, 1st year before program start	0.67	0.32
Share of days in unemployment, 2nd year before program start	0.41	0.39
Share of days in unemployment, 3rd year before program start	0.35	0.39
Share of days in unemployment, 4th year before program start	0.31	0.39
Outcome variables		
Employment three years after program entry	0.35	-
Employment one year after program exit	0.35	-

Table 2. Probit Estimates of Effect of Treatment Duration on Employment Probability

Dependent Variable: Outcome at time 3 years after entry into the program

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Constant	-0.4897	0.0596	-0.6178	0.0948	-0.5743	0.1339	-0.6891	0.1906
Treatment Duration	0.0005	0.0003	0.0019	0.0008	0.0011	0.0018	0.0042	0.0040
Square of Treatment Duration			-2.97E-06	1.71E-06	2.65E-07	7.25E-06	-0.0000218	2.71E-05
Cube of Treatment Duration					-3.82E-09	8.37E-09	5.35E-08	6.84E-08
The Forth Power of Treatment Duration							-4.72E-11	5.61E-11
Log Likelihood	-1369.7851		-1368.2424		-1368.1366		-1367.7707	
Pseudo R Squared	0.0012		0.0023		0.0024		0.0026	
Number of Observations	2126		2126		2126		2126	

Table 3. Probit Estimates of Effect of Treatment Duration on Employment Probability

Dependent Variable: Outcome at time 3 years after entry into the program

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Constant	-1.2038	0.6162	-1.3298	0.6204	-1.3165	0.6288	-1.4303	0.6375
Treatment Duration	0.0003	0.0003	0.0022	0.0009	0.0020	0.0020	0.0063	0.0044
Square of Treatment Duration			-3.94E-06	1.83E-06	-2.95E-06	7.91E-06	-0.0000344	3.01E-05
Cube of Treatment Duration					-1.17E-09	9.09E-09	8.09E-08	7.65E-08
The Forth Power of Treatment Duration							-6.79E-11	6.34E-11
Other Control Variables: See Table 1								
Log Likelihood	-1202.6758		-1200.2769		-1200.2686		-1199.6578	
Pseudo R Squared	0.115		0.1168		0.1168		0.1173	
Number of Observations	2100		2100		2100		2100	

Table 4. Probit Estimates of Effect of Treatment Duration on Employment Probability

Dependent Variable: Outcome at time 1 year after exit the program

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Constant	-0.3787	0.0592	-0.4678	0.0937	-0.3879	0.1318	-0.3899	0.1849
Treatment Duration	-0.0000327	0.0003	0.0010	0.0008	-0.0004	0.0018	-0.0004	0.0039
Square of Treatment Duration			-2.10E-06	1.72E-06	4.01E-06	7.35E-06	3.63E-06	2.65E-05
Cube of Treatment Duration					-7.33E-09	8.63E-09	-6.35E-09	6.71E-08
The Forth Power of Treatment Duration							-8.16E-13	5.49E-11
Log Likelihood	-1376.4014		-1375.6373		-1375.265		-1375.2648	
Pseudo R Squared	0.0000		0.0006		0.0008		0.0008	
Number of Observations	2126		2126		2126		2126	

Table 5. Probit Estimates of Effect of Treatment Duration on Employment Probability

Dependent Variable: Outcome at time 1 year after exit from the program

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Constant	-2.3133	0.6257	-2.4038	0.6295	-2.3625	0.6375	-2.3611	0.6445
Treatment Duration	-0.0002	0.0003	0.0010	0.0009	0.0003	0.0020	0.0003	0.0043
Square of Treatment Duration			-2.66E-06	1.87E-06	5.73E-07	8.10E-06	9.73E-07	2.92E-05
Cube of Treatment Duration					-3.88E-09	9.50E-09	-4.92E-09	7.38E-08
The Forth Power of Treatment Duration							8.68E-13	6.07E-11
Other Control Variables: See Table 1								
Log Likelihood	-1225.0812		-1224.0414		-1223.9567		-1223.9566	
Pseudo R Squared	0.1006		0.1014		0.1014		0.1014	
Number of Observations	2100		2100		2100		2100	

Table 6. Balance in Covariates with and without Adjustment: t-statistics for Equality of Means

	(1)	(2)	(3)	(4)	(5)	(6)
		Unadjusted			Adjusted	
Covariate	[10, 148]	[149,264]	[265,720]	[10, 148]	[149,264]	[265,720]
Age	**2.5**	0.63	**-3.19**	0.63	0.23	0.03
Disability						
No disability	**-2.9**	-0.36	**3.3**	-0.77	0.06	0.03
Disability low degree	**2.08**	-0.19	-1.89	0.52	-0.17	0.07
Disability medium degree	1.67	0.45	**-2.16**	0.49	-0.08	0.1
Disability high degree	1.57	1.2	**-2.87**	0.47	0.38	-0.25
Citizenship						
German	0.74	1.21	**-2.04**	-0.22	0.16	-0.35
Foreigner EU	**-2.74**	0.32	**2.41**	-0.53	-0.01	0.39
Foreigner Non-EU	0.31	-1.41	1.19	0.45	-0.16	0.23
Educational Attainment						
No graduation	**-2.84**	-1.21	**4.17**	-0.37	-0.07	0.38
First state of secondary level	**-3.48**	**-5.15**	**9.14**	0.82	-1.06	0.32
Second stage of secondary level	**3.22**	**1.98**	**-5.39**	-0.51	-0.06	-0.41
Advanced technical college entrance qualification	0.76	**2.43**	**-3.37**	-0.6	0.79	-0.14
General qualification for university entrance	**3.96**	**5.66**	**-10.22**	0.27	1.56	-0.2
Vocational Attainment						
No vocational degree	**-4.62**	**-3.6**	**8.6**	-0.1	-0.65	1.06
In-plant training	1.74	-0.09	-1.65	-0.16	-0.27	-0.58
Off-the-job training, vocational school, technical school	1.16	**2.45**	**-3.79**	-0.31	0.73	-0.1
University, advanced technical college	**4.47**	**5.04**	**-10.06**	0.9	1.32	-0.52

Table 6. Balance in Covariates with and without Adjustment: t-statistics for Equality of Means (Cont.)

	(1)	(2)	(3)	(4)	(5)	(6)
		Unadjusted			Adjusted	
Covariate	[10, 148]	[149,264]	[265,720]	[10, 148]	[149,264]	[265,720]
Employment History						
Previous Unemployment Duration	0.94	**-2.04**	-1.24	0.87	-0.71	0.95
Duration of last employment	1.38	0.28	-1.69	0.24	0.13	-0.21
Log(wage) of last employment	-0.65	0.46	0.17	-0.33	0.19	0.17
No last employment observed	-1.71	0.16	-1.56	0.63	-0.07	-0.06
Share of days in employment, first year before program start	-0.78	0	0.78	0.03	-0.11	-0.37
Share of days in employment, second year before program start	-0.43	**2**	-1.71	-0.66	-0.53	-0.55
Share of days in employment, third year before program start	-0.18	1.1	-0.99	-0.45	0.16	-0.09
Share of days in employment, fourth year before program start	1.65	-0.77	-0.82	0.14	-0.63	0.1
Share of days in unemployment, first year before program start	0.34	-1.33	1.07	0.57	-0.5	1.05
Share of days in unemployment, second year before program start	**-2.06**	**-2.62**	**4.89**	-0.12	-0.64	1.61
Share of days in unemployment, third year before program start	-1.21	**-2.34**	**3.72**	0.06	-0.53	1.23
Share of days in unemployment, fourth year before program start	**-3.19**	-0.85	**4.12**	-0.74	-0.04	1.67
Regional indicators						
Regional type 1	**2.05**	**2.21**	**-4.45**	-0.28	0.15	0.09
Regional type 2	**4.86**	**2.65**	**-7.79**	0.11	-0.25	0.1
Regional type 3	1.42	1.87	**-3.44**	-0.13	0.5	-0.33
Regional type 4	**4.15**	**2.44**	**-6.84**	0.51	0.03	-0.25
Regional type 5	-0.63	0.41	0.19	0	0.71	-1.13
Regional type 6	**-4.6**	1.91	**2.57**	-0.94	1.32	-1.12
Regional type 7	-0.66	**-2.95**	**3.83**	0.52	-0.68	0.12
Regional type 8	-0.68	**-4.84**	**5.88**	0.82	-1.21	1.18
Regional type 9	1.24	0.61	-1.9	0.14	0.43	-0.49
Regional type 10	**-2.61**	-1.21	**3.93**	-0.27	-0.7	1.13
Regional type 11	**-3.41**	**-2.75**	**6.42**	-0.58	-0.7	1.67
Regional type 12	**-2.33**	1.09	1.18	-0.23	0.81	-0.89

Table 7. Dose-response function: Treatment Effects of Continuous Training

	(1)	(2)	(3)	(4)
	Treatment Effects At Time 3 Years After Entry into the Program		Treatment Effects At Time 1 Year After Exit from the Program	
Days of Treatment	Mean	Std. Error	Mean	Std. Error
30	0.282	0.041	0.338	0.043
60	0.334	0.023	0.354	0.024
90	0.357	0.022	0.356	0.022
120	0.363	0.020	0.352	0.020
150	0.359	0.017	0.351	0.016
180	0.351	0.015	0.354	0.015
210	0.345	0.017	0.361	0.017
240	0.349	0.019	0.371	0.019
270	0.367	0.019	0.379	0.018
300	0.391	0.018	0.382	0.018
330	0.411	0.022	0.374	0.021
360	0.416	0.028	0.355	0.027
390	0.402	0.038	0.330	0.035
420	0.369	0.049	0.301	0.045
450	0.325	0.061	0.275	0.058
480	0.277	0.071	0.252	0.069
510	0.235	0.078	0.235	0.078
540	0.201	0.084	0.222	0.084
570	0.180	0.090	0.212	0.090
600	0.173	0.100	0.206	0.098
630	0.182	0.117	0.202	0.113
660	0.209	0.146	0.201	0.139
690	0.255	0.187	0.203	0.178
720	0.323	0.242	0.207	0.231

Note: Standard errors are bootstrapping standard errors from 2,000 replications.

Figure 1. Distribution of Treatment Duration

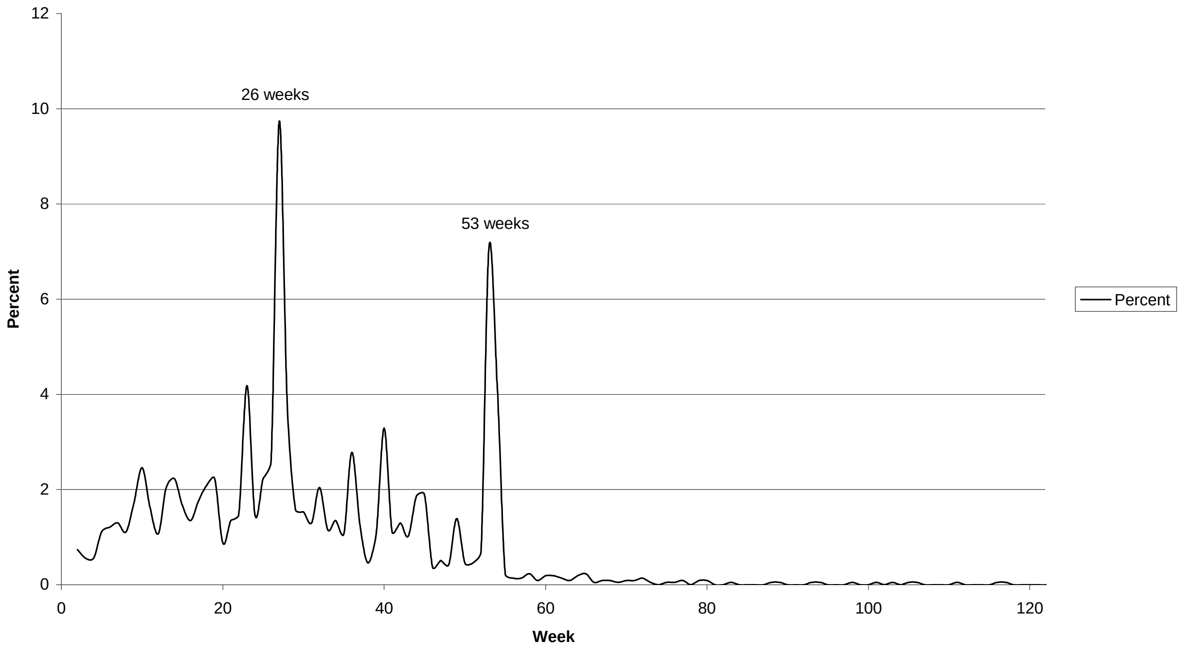


Figure 2. Treatment Effects at Time 3 Years after Entry into the Program

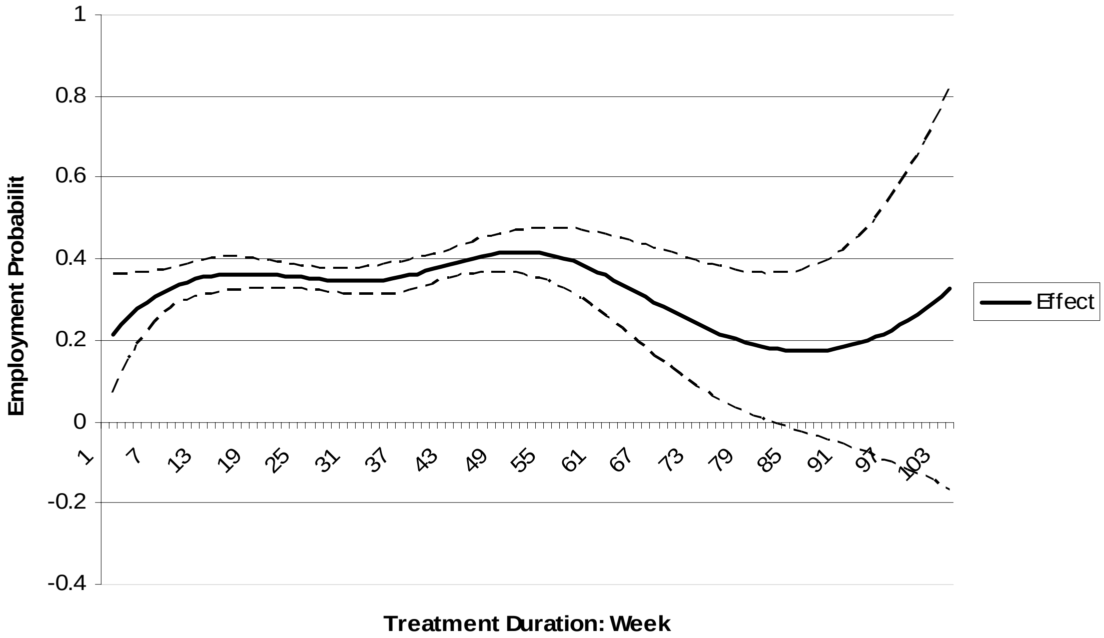
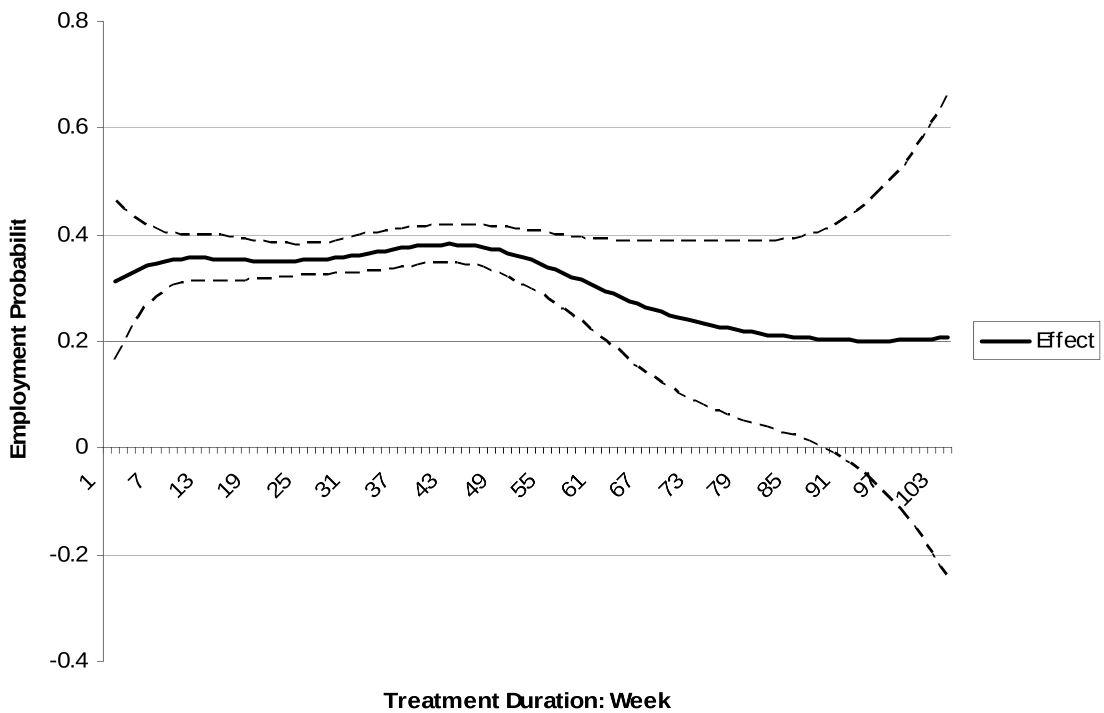


Figure 3. Treatment Effects at Time 1 Year after Exit the Program



Note: Dashed lines are bounds for 95% confidence intervals. These intervals are based on bootstrap standard errors from 2,000 replications.