

Evaluating continuous training programs using the generalized propensity score*

Jochen Kluge
RWI Essen and IZA Bonn

Hilmar Schneider
IZA Bonn

Arne Uhlendorff
IZA Bonn and DIW Berlin

Zhong Zhao[†]
Renmin University of China and IZA Bonn

23 June 2009

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 10 days to 395 days (i.e. 13 months). 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 data have the advantage that we can consider both the actual and planned training durations as treatment variables: If only actual durations are observed, treatment effect estimates may be biased because of endogenous exits. Our results indicate an increasing dose-response function for treatments of up to 100 days, which then flattens out. That is, longer training programs do not seem to add an additional treatment effect.

JEL Codes: C21, J68

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

Word Count: 9,370 words

* We would like to thank Deborah Cobb-Clark, Dan Hamermesh, Oscar Mitnik, Peter Mueser, Donald Parsons, Ulf Rinne, Gerard van den Berg, participants of the VfS Conference in Munich, the COST meeting in St. Gallen, and seminars at the George Washington University, Renmin University of China, the World Bank and the Washington Statistical Society for valuable discussions and helpful comments and Jan Stuhler for excellent research assistance. Arne Uhlendorff also thanks DIW DC, where part of this research was pursued during his stay in fall 2007. The usual disclaimer applies.

[†] Correspondence: kluge@rwi-essen.de, schneider@iza.org, uhlendorff@iza.org, mr.zhong.zhao@gmail.com

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, LaLonde and Smith 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, Miquel and Wunsch 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 13 months. 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).¹ Participants in these programs learn specific skills required for a certain vocation, like computer aided design for technicians, or receive qualifications that are of general vocational use, like general computer skills. In this paper we compare the impact of being trained within the same program type, but with different durations on the subsequent employment probability.

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 as proposed in Hirano and Imbens (2004). 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

¹ 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.

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). Similar to the binary and multi-valued treatment propensity score methods it is assumed that – conditional on observable characteristics – the level of treatment received can be considered as random. Hirano and Imbens (2004) show that the GPS has a balancing property similar to the balancing property of the "classic" propensity score. This implies that individuals within the same strata of the GPS should look identical in terms of their observable characteristics, independent of their level of treatment. Compared to propensity score methods for multi-valued treatments, the GPS has the advantage that we do not have to discretize the continuously distributed training duration, and are thus able to make use of more comprehensive information. To our knowledge, our paper along with parallel work by Flores-Lagunes, Gonzalez and Neuman (2007) constitute the first applications of the GPS in the context of evaluating active labor market policy.

In implementing the GPS approach, our data have the advantage that we can consider both the actual and planned training durations as treatment variables: If only actual durations are observed, treatment effect estimates may be biased because of endogenous exits. This could be the case, for instance, if observed durations are shorter than the initially planned durations, because people exit from the program early if they find a job. The bias could also point the other way, if a substantial fraction of program participants drops out early. We investigate these issues by taking into account both the actual and planned durations of individual program participants. Our results indicate that in some settings standard regression-type adjustments may not be sufficient to remove all observable bias and may lead to different conclusions than the proposed GPS approach.

The availability of planned training duration variable also allows us to test the exogeneity of observed training duration, since planned training duration was determined before the start of treatment, and can be used as an instrumental variable for observed training duration in order to deal with endogenous exits. This approach does not help to control for potential selection into planned program duration based on unobservables. However, it seems plausible that the planned duration is exogenous once we control for the detailed observed characteristics including previous labor market history. Potential endogeneity could mainly occur due to

reverse causality, i.e., participants leave the program because they received a job offer, or due to the fact that some individuals leave the program because they expect no further benefits from continuing the program.

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 empirical implementation. It discusses the plausibility of the unconfoundedness assumption, it details the GPS estimation, the balancing of covariates and the common support condition, and it presents the results from estimating the dose-response function. Section 5 contains several robustness checks. Section 6 concludes.

2. Bias removal using the Generalized Propensity Score

Research in program evaluation in recent years has made comprehensive use of matching methods². 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. Alternatively we could discretize the

² 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).

continuously distributed treatment variable and apply propensity score methods for multi-valued treatments. The GPS, though, has the advantage that it makes use of the entire information contained in the distribution of treatment duration.

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 $i=1, \dots, N$. For each unit i there exists a set of potential outcomes $Y_i(t)$ for $t \in \mathfrak{T}$, referred to as the unit-level dose-response function. In the continuous case, \mathfrak{T} is an interval $[t_0, t_1]$, whereas in the binary case it would be $\mathfrak{T} = \{0,1\}$. 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 continuous case:

$$(1) \quad Y(t) \perp T \mid X \text{ for all } t \in \mathfrak{T}.$$

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 $r(t, x) = f_{T|X}(t | x)$ 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.

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

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,

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

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

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

which we estimate by ordinary least squares regression (OLS). 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.³ The estimated GPS is calculated as

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

In the second stage we calculate the conditional expectation function of Y_i given T_i and R_i as a flexible function of its two arguments. Our empirical approach uses the following polynomial approximation. In addition to the specification in equation (7) we also implement several other specifications in order to allow for sufficiently flexible functional forms and to test whether our results are robust with respect to a higher degree of flexibility.

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

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

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

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, unconfoundedness assumption and common support condition

Just as in the case of a binary treatment, 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,

³ We alternatively estimated the GPS based on the logarithm of the duration as a robustness check, but using duration instead of the logarithm of the duration turned out to be superior in finding GPS specifications that balance the covariates in our sample.

testing for covariate balance is more difficult with continuous treatments.

Hirano and Imbens (2004) propose *blocking* on both the treatment variable, i.e. length of 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 tests if for each of these blocks the covariate means of individuals belonging to the particular treatment-level group are significantly different from those of 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.

The key assumption for the GPS is the weak unconfoundedness assumption, also known as the assumption of selection on observables. As an identifying assumption, it is not statistically testable. One typical case of violating this assumption is the possibility that treatment duration is endogenous. In our data, besides the actual training duration, we also know the planned training duration. The planned duration is determined prior to the program, which is arguably exogenous once we condition on detailed observed characteristics including previous labor market history. We can use the information on the planned duration to test the endogeneity of the actual treatment duration. If we assume that the planned duration is exogenous, this allows us to take potential endogeneity into account which may occur due to reverse causality - participants may leave the program because they got a job offer - or due to the fact that some individuals leave the program because they expect no further benefits from continuing the program.

Besides using planned training duration as an instrumental variable to test the endogeneity of

observed training duration, we also follow the framework of Heckman and Hotz (1989) testing whether our GPS method produces comparable samples. In this analysis, we focus on pre-training employment status, since employment status is the main outcome variable in our paper. For empirical implementation, we follow the above outlined balancing test approach to adjust the pre-training employment status. Pre-training employment status is not directly controlled in our GPS, but if the GPS estimator is valid in our sample, we would expect that the differences-in-means of pre-training employment status should not be statistically different from zero after we adjust for the GPS properly since nobody received treatment before the start of training program. Mueser, Troske and Gorislavsky (2007) apply a similar approach in the context of binary treatment.

Similar to standard propensity matching methods, common support is also a concern in the GPS application. We propose to test the common support condition as follows⁴: First, following the procedure for testing for the balancing of covariates, we divide the sample into three groups according to the distribution of treatment length, cutting at the 30th and 70th percentile of the distribution. Then we evaluate the GPS at the group median of the treatment duration variable. For example, we evaluate the GPS for the whole sample at the median treatment duration of group 1, and after that we plot the distribution of the evaluated GPS for group 1 vs. the distribution of the GPS for the rest of the sample. Like in the case of binary propensity score matching, by inspecting the overlap of these two distributions we are able to examine the common support condition graphically. In the same fashion, we can test the common support condition of groups 2 and 3 vs. the rest of the sample.

3. Institutional setting and Data

3.1 Public Training Programs in Germany

The most important German government labor market policy relevant to our paper is the Social Code III (Sozialgesetzbuch III) enacted in 1998. The focus group of the Social Code III consists of people who are unemployed or under threat of unemployment. The code has emphasized the use of Active Labor Market Policy (ALMP), and aims to reduce unemployment. The Federal Employment Agency through its 10 regional directorates and 180 local employment agencies (with around 660 branch offices) is responsible for implementation of the federal labor market policy at national, regional and local level. See

⁴ We thank Peter Mueser for suggesting this approach. At a later stage we found out that Flores-Lagunes, Gonzalez and Neuman (2007) independently proposed a similar approach in their paper.

Wunsch (2005) for a detailed description of German labor market policy and related institutions.

Training programs are one of the most important components of ALMP in Germany with an annual budget of around 7 billion Euros (2002 figures, see Eichhorst and Zimmermann 2007). Access to training programs is not a legal entitlement, but based on the decision of the caseworker. If a caseworker has decided that her client needs to go through a training program, the caseworker also specifies the type, the content and the duration of the training and refers the client to a designated training provider. During the process, the factors that the caseworker takes into consideration include the aptitude of her client for a certain job, the likelihood to succeed in a specific training program, the local labor market condition, the cost of training, and to some extent the available training slots in the contracted training institutions.⁵ It is thus reasonable to assume that once we condition on the large set of observable characteristics, including previous labor market outcomes, the decision about the length of the program is independent of the future labor market outcomes of the participants.

In principle, one can distinguish two main types of training programs, programs in which no specific degree is acquired and programs leading to the acquisition of a degree. The latter are not considered here, 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. Among the programs considered here one can distinguish between classroom oriented training programs (type 1) and more practically oriented programs with only a few theoretical parts (type 2). However, the duration as well as the effectiveness of both types is very similar and therefore we pool both programs to increase our sample size.⁶

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 2000, roughly 70% of all participants in training programs were assigned to this type (Schneider and Uhlendorff 2006; IZA, DIW and infas 2007).

⁵ Since 2003, the matching process between trainees and trainers has changed, and the government introduced training vouchers.

⁶ There exists another program type not leading to the acquisition of a degree, training in a “practice” firm, which mainly consists of working in a simulated firm. We do not take this program into account because this type of training differs fundamentally from the types considered here.

3.2 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.

Our sample of participants consists of about 265 male unemployed persons per quarter entering the program during the years 2000, 2001, and 2002, i.e. we observe approximately 3180 program participants. The system of publicly financed training in Germany underwent several alterations during labor policy reforms in 2003 (see Jacobi and Kluve 2007). We therefore restrict our analysis on pre-reform training programs to avoid possible distortions in measuring program effectiveness.⁷ 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 elapsed duration of training varies from approx. 1 week up to 13 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 treatment duration. For all participants we know the initial length of the treatment they were assigned to (i.e. the planned duration), as well as how long they actually stayed in the treatment (i.e. the actual duration).⁸

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 395 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

⁷ The data originate from a research project for the German government evaluating the labor policy reforms. In order to capture trends over time and to separate this from the reform effect in 2003, the original sample consists of entry cohorts over time with stable cohort size per quarter.

⁸ However, we must admit that there is one caveat that we do not have information on the training content, though we restrict our analysis on relatively homogeneous two types of training programs, we cannot rule out that training content could vary with the training duration. Our estimates may reflect the composite effect of training duration and training content. Nonetheless, our paper still has strong policy implications and reflects the reality that training content and training duration are usually offered together as a bundle to the program applicants.

average responses of those individuals that did receive some treatment. Figure 1 shows the distribution of treatment durations, both for the actual and planned durations. Both distributions cover the full range of training durations, and for both distributions two peaks exist at durations of 180 days and 360 days, respectively. Figure 1 also shows that actual durations tend to be slightly shorter than planned 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 2 years after entry into the program. Table 1 presents summary statistics of the two outcome variables and the covariates, for the full sample (columns 1 and 2) as well as for three subsamples, “early exits” (i.e. actual duration < planned duration, columns 3 and 4), “late exits” (i.e. actual duration > planned duration, columns 5 and 6), and “exits as planned” (i.e. actual duration = planned duration, columns 7 and 8). The share of individuals who stayed in the program exactly as long as planned is quite high (68.7%). In the case in which actual and planned durations differ, early exits are much more common than late exits (22.1% and 9.2% of observations, respectively), a pattern already observed in Figure 1 above.

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 detailed information on pre-treatment employment histories, covering the number of days spent in employment and unemployment during the four years preceding treatment, as well as regional indicators. When using nonexperimental estimators, previous studies, e.g. Heckman, Ichimura, Smith and Todd (1998), Diaz and Handa (2006) and Mueser, Troske and Gorislavsky (2007), emphasize the importance of having treated and comparison groups from the same local labor market, applying the same survey instrument for both groups, and having rich information on individuals' recent labor market history. First, given that our data come from one administrative source, the requirement of a homogenous survey instrument is certainly satisfied. Second, the richness of the covariates makes the unconfoundedness assumption entirely plausible. Third, our sample is a random sample of participants from West Germany, such that, in order to control for local labor market conditions, we condition on the local unemployment rate and the regional type, reflecting the general performance of

the regional labor market (for a description of the different regional types see Blien et al. 2004). This and the fact that we analyze a national program ensure that different effects of different program lengths are not driven by different regional administrations or different local labor market conditions.

[Table 1 about here]

Table 1 also shows that the covariate distributions are very similar across all (sub-) samples. Looking at the full sample, the participants are on average 37 years old, around 9% of them are handicapped and 12% 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 (7%) has obtained a university degree.

Before entering training the participants were on average unemployed for 9 months, and their previous employment lasted for about 21 months. The individuals for whom we observe a wage for their last employment earned around 50 Euros 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. There is a clearly increasing trend in the average probability of being unemployed over time as the individuals move closer to enrolment in the program.

Looking at the two outcome variables, both two years after program entry and one year after the program ended around 35% of the participants are employed. Figure 2 contains six panels plotting unadjusted outcomes – i.e. the employment probability two years after program entry as well as the employment probability one year after program exit – against the three treatment variables, i.e. actual, planned, and actual=planned durations. The figures generally show an increasing trend: After an initial dip in employment probability during the first month in the program, employment rates seem to increase with the length of participation.

[Figure 2 about here]

4. Empirical results

4.1 Estimates from a Linear Probability Model

As mentioned in Section 3, in this paper we consider two outcome variables: one is the employment probability at the point in time 2 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 linear probability model (LPM). Table 2, parts a) and b), investigate the relationship between the employment probability at 2 years after entering into the program (part a) and 1 year after exit from the program (part b), respectively, with the treatment duration.

[Tables 2a and 2b about here]

First, the two parts of Table 2 show 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 with or without controlling for additional variables. However, the estimated coefficients of the treatment duration are small, and the explanatory power of the treatment duration is low.⁹ These suggest that the impact of treatment duration on the employment probability is small or negligible.

However, it is worth noting that a regression type analysis such as the LPM models 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. For a discussion of the advantages of matching methods compared to parametric regressions see e.g. Ho et al. (2007).

4.2 Unconfoundedness assumption

The key assumption for the GPS is the weak unconfoundedness assumption, also known as the assumption of selection on observables. As an identifying assumption, it is not statistically testable. In the training program context, there are two types of endogeneity. One is that the decision to participate in training or not can be endogenous. The other is that among the participants, the length of training also can be endogenous, e.g. endogenous attrition from the program. In our paper, we focus on training participants, and investigate the treatment effect

⁹ Cf. the low adjusted R-squared in Panel A of Tables 2a and 2b.

of training duration. So our main concern is the possibility that treatment duration is endogenous. In our data, besides the actual training duration, we also know the planned training duration. The planned duration is determined prior to the program, which is arguably exogenous, and is free from the endogenous attrition problem. We can use the information on the planned duration to test the endogeneity of the actual treatment duration. Tables 3a and 3b are instrumental variables (IV) estimates using planned duration as IV. Comparing these IV estimates to the OLS estimates in Tables 2a and 2b, we find that they are in 15 out of 16 specifications not significantly different (see the results of the Hausman test in Tables 3a and 3b). This suggests that the actual training duration may not suffer strongly from endogeneity.

[Tables 3a and 3b about here]

Another test for the unconfoundedness assumption is based on the approach proposed in Heckman and Hotz (1989). This approach is successfully applied in Mueser, Troske and Gorislavsky (2007) in a binary matching context. In our data, we have pre-training employment status back to more than 3 years ago. In our GPS adjustment process, though we control for pre-training labor market variables, we do not control for the pre-training employment status explicitly. If the GPS estimator is valid, then after adjustment for the GPS, the pre-training employment status should be similar across groups with different training durations since nobody received treatment before the start of the training program.¹⁰ Figure 3 shows the t-statistics following the approach outlined in 2.3 for the actual and the planned duration and for a subsample of individuals with equal planned and actual durations. Before adjustment for the GPS, a large portion of t-statistics for testing equality of means are outside the interval of [-1.96, 1.96]. After adjustment for the GPS, almost all of these t-statistics fall into the interval of [-1, 1], which corresponds to a significance level of about 30%. For example, before adjusting for the GPS, the difference between individuals with a long training duration (group 3) and individuals with shorter training (less than 248 days) is significantly different 20 weeks before entering the program. This significance disappears once we adjust for the GPS. These results strongly suggest that adjustment for the GPS not only balances those covariates directly entering into the estimation of the GPS, but also balances the pre-training employment status, which is not explicitly controlled by our GPS approach.

[Figures 3a, 3b and 3c about here]

¹⁰ This is similar to the binary matching case. If the matching estimator is valid, the pre-treatment outcome variable should not be significantly different between the matched treated and comparison groups, as discussed in Mueser, Troske and Gorislavsky (2007).

4.3 GPS estimation, covariate balance, and common support

Our implementation of the generalized propensity score follows the procedure outlined in Hirano and Imbens (2004) and adapted to our context as presented in sections 2 above. We first estimate the conditional distribution of the length of the training program (treatment) by applying OLS. Table 4 contains the results for the actual and the planned duration and for the subsample with equal planned and actual durations.

[Table 4 about here]

To assess the balancing property of the GPS (cf. section 2.3) we compare the distribution of covariates between three groups, which are defined by cutting the distribution of treatment duration at the 30th and 70th percentiles. We implement this for both the actual and planned durations. For actual durations, group 1 includes individuals with a treatment level between 11 and 137 days, group 2 ranges from 138 to 247 days and group 3 from 248 to 395 days. For planned durations, group 1 includes individuals with a treatment level between 11 and 167 days, group 2 ranges from 168 to 271 days and group 3 from 272 to 395 days. The groups therefore reflect the fact that on average actual durations are shorter than planned durations. The subsample of individuals with equal planned and actual durations the distribution is very similar to the planned duration of the whole sample (group1: 11 to 170, group 2: 171 to 271, group 3: 272 to 395 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 each of the Tables 5, 6 and 7 the corresponding t-statistics are reported. Without adjustment the clear majority of t-statistics are greater than 1.96, indicating a clearly unbalanced distribution of covariates. For example, individuals with longer training programs – trained between 248 and 395 days – are significantly older than the participants who are trained shorter than 248 days¹¹. Individuals with shorter training programs (11 to 137 days) are significantly younger than individuals who are trained more than 137 days. Moreover, the share of disabled unemployed is significantly higher among the longer trained, individuals with the highest schooling degree and individuals with a university degree are more often in the group with short or medium training duration than among the group with longer treatment, and the regional unemployment

¹¹ The t-value of -4.53 describes the difference between the individuals with a training duration of below and above 248 days.

rate tends to go along with a longer training duration. These examples underline the importance of correcting for differences in observables.

[Tables 5, 6 and 7 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 lengths of 84 days, 180 days, and 332 days for the actual duration, at the lengths of 117 days, 184 days, and 348 days for the planned duration, and at the lengths of 117, 187 and 357 days for the subsample of individuals with equal planned and actual training durations. 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 for the actual duration, consisting of individuals with an actual treatment ranging from 11 to 137 days, the GPS evaluated at the median of this group (84 days). The distribution of the GPS $r(84, X_i)$ is then discretized into five blocks using the quintiles of the distribution. For the first group, this leads to the intervals $[0.00005, 0.0017]$, $[0.0017, 0.0025]$, $[0.0025, 0.0030]$, $[0.0030, 0.0035]$ and $[0.0035, 0.0045]$.¹²

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 $r(84, X_i)$ lying in the first interval as well. For each group, this implies five mean differences and five standard errors. The t-statistics reported on the right hand side of each of the Tables 5, 6 and 7 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 observations.

In contrast to the unadjusted sample, we observe no t-statistics larger than 1.96 for the planned duration and for the subsample with equal planned and actual durations (Tables 6 and 7) and only one t-statistic larger than 1.96 for the actual duration (Table 5). These results indicate that the balance of the covariates is clearly improved by adjustment for the GPS.

To test the common support condition for the actual duration, following the approach outlined in section 2.3, we divide the sample into the same three groups that we have used above when

¹² If we specify the GPS based on the logarithm of the training duration the corresponding values are higher and comparable to other studies applying the GPS, e.g. Flores-Lagunes, Gonzalez and Neuman (2007).

testing for covariate balance. We then first evaluate the GPS of the whole sample at the median actual treatment duration of group 1, i.e. 84 days. After that we plot the distribution of this GPS (i.e. evaluated at the median actual duration of the first group) for group 1 and for the other two groups taken together, a procedure which results in the top left panel of Figure 4. We repeat the same procedure for group 2 and group 3, which gives us the top center and the top right panels of Figure 4. Moreover, this common support test is also conducted for the planned duration (the three middle panels) and for the subsample with equal actual and planned durations (the three bottom panels).

The figures show that, with the exception of very few cases in the low tail of the second panel, the common support condition is satisfied. The results are very similar for the actual duration, the planned duration, and the subsample with equal planned and actual durations, i.e. common support is given.

[Figure 4 about here]

4.4 Results from estimating the dose-response function

The final step of our empirical analysis consists in estimating the GPS-adjusted dose-response function. Table 8 contains the estimation results for the dose-response function. Our main results for both outcome variables are presented in Figure 5 (employment probability at time 2 years after entry into the program) and Figure 6 (employment probability at time 1 year after exit from the program), where each figure consists of three parts showing results for a) the actual duration, b) the planned duration, and c) for the subsample of individuals for which actual duration equals planned duration. With respect to observable characteristics, this subsample is very similar to the complete sample, see rows 1 and 7 in Table 1. The Figures also include the non-participant employment probability baseline¹³, which indicates that training effects are generally positive, a finding in line with Rinne, Schneider and Uhlenborff (2007) who also use the IEB data. In addition to the graphs we calculate the pairwise differences in employment probabilities between different durations of training and bootstrap standard errors from 2,000 replications. For example, we calculate the difference in our outcome variables between 18 days of training and 39 days of training, between 18 days of training and 60 days of training etc. This allows us to test whether the impact of different

¹³ Given the richness of our data set we estimate this covariate-adjusted baseline from standard binary propensity score matching methods, i.e. the employment probability of the matched non-participants.

training durations is significantly different from zero¹⁴.

[Table 8 about here]

[Figures 5, 6 about here]

As the figures show, the dose-response functions for both outcome variables considered have similar shapes for all specifications. They generally vary depending on the treatment variable considered: specifications based on the actual duration are rather flat, showing little variation of the outcome with respect to different durations. According to this, we do not find any significant difference in employment probabilities between the different levels of training durations. Specifications based on the planned durations show an increase in employment probability for the short durations of up to around 100 days, a slight dip for durations of about 200 to 250 days, and a final decrease for durations longer than 330 days (where confidence bands, however, are quite large). If we look at the outcome two years after entry into the program (Figure 5), this pattern is supported by the pairwise comparisons of different levels of treatment intensity. We find significant differences between a duration of 18 days and longer training duration up to around 120 days, and these differences are again significant for a training duration between around 270 and 350 days. There is no significant difference between durations of around 100 days and durations of around 300 days. In the alternative specification with the outcome measured one year after program exit we do not find any significant increase in the first 250 training days, and training durations around 300 days lead to significantly higher employment probabilities than shorter training. However, the planned duration is not a “real” treatment and therefore the Figures 5b and 6b are hard to interpret.

The subsample of individuals for which actual duration equals planned duration does not suffer from potentially endogeneous exits from the program. The general profile of the results for the observed durations is confirmed (panels c in Figures 5 and 6, respectively): while it is generally flatter for the long durations, it further emphasizes the increase in the treatment effect for durations of up to 100 days if we look at the employment probability two years after entry into program. We find a significant increase until around 100 days of training, and no further significant increases beyond 100 days. However, for the treatment effect one year after exit, we do not find significant differences between the different levels of treatment intensity.

¹⁴ The detailed Tables are not reported here but available on request.

The GPS-adjusted analysis of the relationship between a continuous training program and the corresponding employment probability of participants thus shows a very interesting pattern. While during the first 100 days of exposure to treatment increasing the dose yields increasing returns in terms of employment prospects two years after entry into the program, further increasing the dose beyond 100 days appears to bring about no additional treatment effect. That is, the human capital enhancing features of training are effective during the beginning period (i.e. the initial doses work), but effectiveness fades out after approximately three months of participation.

We observe the initial increase in the employment probability only for the planned and not for the actual duration, see Figure 5. This result may be driven by individuals who leave the program early because they received a job offer. “Controlling” for early exits by using the planned instead of the actual duration leads to a positive impact – at least on the point estimate – of the first 100 days of training duration. These results are confirmed if we use the sample for which actual duration equals planned duration.

Our findings add interesting additional insights into the effectiveness of training programs for the unemployed in Germany. There exist several papers analyzing effect heterogeneity of training programs with respect to observable characteristics, see e.g. Rinne, Schneider and Uhlenborff (2007). They find only weak evidence for heterogeneous effects with respect to skill levels and age. Other research exploits the difference of long versus short training programs by discretizing the length of training, see e.g. Lechner, Miquel and Wunsch (2004). They show that longer programs have a larger and more sustainable positive impact on subsequent employment prospects. However, in their case the longer duration goes along with a different type of program - substantive training programs leading to a vocational degree are compared with shorter programs not leading to a degree. In this paper, we use the GPS framework to exploit the effect heterogeneity of different levels of training duration within program types not leading to a vocational degree. Our results suggest that an increased duration of training does not automatically go along with higher subsequent employment probabilities, and that after around 3 months of training, additional time spent in programs does not have an impact on subsequent employment prospects. These results indicate that learning specific skills required for a certain vocation (e.g. computer-aided design for a technician/tracer) or receiving qualifications that are of general vocational use (e.g. MS Office, computer skills) – that is, the contents of the programs analyzed here – have positive expected returns compared to job search without training only up to a certain duration.

5. Robustness

In this section, we carry out several sensitivity checks for our main specification. The first check is that we further restrict our sample to the people who went through a training program exactly once. In the second check, we try different specifications for the dose-response functions, and also present estimates from LPM and probit models. Finally, we use planned duration as an instrumental variable for actual treatment duration, and estimate the local average treatment effect (LATE) as developed in Imbens and Angrist (1994).

Figures 5 and 6 above include dose-response functions for a subsample of our data (labeled “dose-response for subsample” in the graphs). The original data contain information on whether a training participant, after having taken part in the course which we analyze here, participated in another training course at some point in time. These are about 7% of individuals in our sample. We therefore include results for the subsample of observations that participated in exactly the one course for which we have data on planned and actual durations.¹⁵ Regarding the shape of the dose-response functions, results for the subsample are very similar to the full sample. It is worth noting though, that the employment probabilities of the matched non-participants are consistently larger for the subsample, and thus the treatment effects are smaller. In particular, the estimated average response of the non-participants is up to 3 percentage points higher, while the response of the participants remains almost the same (cf. Figure 6a).

Our main estimation is based on a cubic specification for the dose-response function. Figures 7 and 8 plot results for the dose-response function for the full sample for quadratic and 4th degree polynomial specifications as well. Like Figures 5 and 6, Figures 7 and 8 consider the two outcome variables and are structured in three panels each, reflecting actual, planned, and actual=planned durations. All six figures show that the general shapes and trends of the dose-response functions remain relatively unchanged under different specifications, though there are some differences in detail. Our central finding that the main body of the dose-response functions is flat, i.e. longer training programs do not seem to add an additional treatment effect, is robust. This main result holds true also if we use the logarithm of the duration as the dependent variable for our GPS (see the thin solid lines in Figures 7 and 8).¹⁶

¹⁵ The balancing of the covariates is very similar for these subsamples and the corresponding tables are available upon request.

¹⁶ The corresponding balancing properties are available upon request. We mainly present results based on the linear specification because the corresponding balancing properties are better.

It is also interesting to compare the results from the GPS with the ones obtained directly from LPM and probit models. In Figure 7, both sets of results are quite similar, but in Figure 8, results from the GPS are rather different from the ones that would be estimated using an LPM and a probit model. This suggests that in the case of Figure 8, a regression-type adjustment may not be sufficient to remove all observable bias, and the GPS provides a valuable alternative approach to control for differences in observables.

[Figures 7, 8 about here]

As we stated earlier, one of the paper's main findings is that longer training programs do not seem to add an additional treatment effect. We carry out another sensitivity check for this statement using an instrumental variable approach.

In our data, for about 31% of participants the actual treatment duration differs from their planned duration. It is possible that the actual treatment duration could be endogenously determined by the participants, either, for instance, by dropping out of the program or by finding a job while still in training. Fortunately, in our data, we also have information on the planned treatment duration, and this variable is decided prior the program, so we can use it as an instrumental variable to control for the possible endogeneity of the actual treatment duration.¹⁷

To apply the approach in Imbens and Angrist (1994), first we discretize both actual and planned treatment duration variables into dummy variables according to the length of treatment. The indicator "1" means that the participants have a shorter duration (actual or planned), and "0" means otherwise. If the treatment duration has little impact on the outcome variables, the IV estimates should not significantly differ from zero, i.e. participants with shorter treatment durations have similar outcomes to the participants with longer treatment duration. In our empirical implementation, we use 5 different cutoff points, respectively, to define the two groups with the shorter vs. the longer treatment duration; i.e. we cut at the 15%, 30%, 50%, 70% and 85% percentiles of the distribution of the actual treatment duration. The same cutoff points are applied to both actual as well as planned treatment duration variables to create dummy variables. After that, we use the dummy variable created from planned

¹⁷ In section 3, we also use planned duration as an instrumental variable, but the objective there differs from here. The one in section 3 is to test the endogeneity of the actual training duration, and the one here is to estimate the effect of training duration on the outcome variable, employment probability.

treatment duration as an instrument for the dummy variable created from the actual treatment duration.

[Table 9 about here]

Table 9 presents the results from this instrumental variable approach, with and without controlling for additional variables. The different models 1 to 5 correspond to the different cutoff points described above, ranging from the 15% percentile to the 85% percentile. The majority of these estimates are insignificant, except for some cases in which the lower cutoff points are used.¹⁸ This provides additional evidence to support our finding from the GPS results that, if treatment duration has an impact at all, it is a weak impact during the first months, and longer durations seem to have no additional impact on the labor market outcomes of the participants.

6. 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 on training duration in days for a period of about 1 week to 13 months. 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. Extensive diagnostics on covariate balance, common support, and the unconfoundedness assumption validate the approach. Moreover, we are able to consider both the planned and actual durations as treatment variables, thus avoiding a potential bias in estimating the dose-response function from endogenous exits, which may play a role if only actual durations are observed. We conduct various robustness checks in order to further solidify our results.

Our findings indicate that the dose-response function that relates continuous training to the corresponding employment probability has a relatively flat shape after an initial increase during the first 100 days of training. Indeed, the first three months of a training program appear to be the most effective period to improve the employment probability and bring about the generally positive effect relative to the non-participant baseline. In the lower segment of the distribution of training durations, additional doses seem to bring about increases in post-

¹⁸ For these cases the estimates are negative, i.e. a shorter treatment duration relates to a lower outcome. This is also consistent with our GPS finding that the dose-responses are upward sloping in the lower treatment duration segment.

training employment prospects.

After approximately 100 days, however, further participation in the program does not seem to lead to an increase in the treatment effect, as the dose-response function is basically flat for higher doses. Whether the effect actually even starts to decrease again for the very long durations (330 days +) cannot be said with certainty, as large confidence bands due to small number of observations exacerbate a precise estimation of this effect. Based on our assessment of the dynamics of the individual employment probability brought about by continuous increases in program duration, there seems to be little justification for training programs in Germany to last longer than about three months.

References

- Augurzky, B. and J. Kluve (2007), "Assessing the performance of matching algorithms when selection into treatment is strong", *Journal of Applied Econometrics* 22, 533-557.
- Behrman, J., Y. Cheng 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.
- Blien, U., F. Hirschenauer, M. Arendt, H.J. Braun, D.M. Gunst, S. Kilcioglu, H. Kleinschmidt, M. Musati, H. Roß, D. Vollkommer, J. Wein, (2004): "Typisierung von Bezirken der Agenturen für Arbeit", *Zeitschrift für ArbeitsmarktForschung*, 37, 146-175.
- Diaz, J.J. and S. Handa (2006), "An Assessment of Propensity Score Matching as a Nonexperimental Impact Estimator: Evidence from Mexico's PROGRESA", *Journal of Human Resources* 41, 319-346.
- Eichhorst, W. and K.F. Zimmermann (2007): "And Then There Were Four... How Many (And Which) Measures of Active Labor Market Policy Do We Still Need?", *Applied Economics Quarterly*, 53, 243-272.
- Flores-Lagunes, A., A. Gonzalez and T.C. Neuman (2007), "Estimating the Effects of Length of Exposure to a Training Program: The Case of Job Corps", *IZA Discussion Paper* 2846.
- Heckman, J.J. and V. J. Hotz (1989), "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training", *Journal of American Statistical Association* 84, 862-874.
- Heckman, J.J., H. Ichimura, J. Smith and P. Todd (1998), "Characterizing Selection Bias Using Experimental Data", *Econometrica* 66, 1017-1098.
- 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 A. Gelman and X. Meng (eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, Wiley.
- Ho, D.E., K. Imai, G. King, E.A. Stewart (2007), "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference", *Political Analysis* 15, 199-236.
- 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.

Imbens, G.W. and J.D. Angrist., "Identification and Estimation of Local Average Treatment Effects," *Econometrica* 62, 467-475.

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

Jacobi, L. and J. Kluge (2007), "Before and After the Hartz Reforms: The Performance of Active Labour Market Policy in Germany", *Journal for Labour Market Research* 40, 45-64.

Kluge, 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.

Mueser P. R., K. R. Troske and A. Gorislavsky (2007), "Using State Administrative Data to Measure Program Performance," *The Review of Economics and Statistics* 89, 761-783.

Rinne, U., M. Schneider and A. Uhlendorff (2007), "Too Bad to Benefit? Effect Heterogeneity of Public Training Programs", *IZA Discussion Paper* 3240.

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* 39, 477-490.

Wunsch, C. (2005), "Labour Market Policy in Germany: Institutions, Instruments and Reforms since Unification," *Discussion Paper No. 2005-06, Department of Economics, University of St. Gallen*.

Table 1. Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Full Sample		Early Exits		Late Exits		Exits as plan.	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	37.22	10.36	36.30	10.54	37.00	10.40	37.55	10.27
<u>Disability</u>								
Disability low degree	0.07	-	0.09	-	0.04	-	0.07	-
Disability medium degree	0.01	-	0.00	-	0.00	-	0.01	-
Disability high degree	0.01	-	0.00	-	0.00	-	0.01	-
<u>Citizenship</u>								
Foreigner EU	0.02	-	0.02	-	0.01	-	0.02	-
Foreigner Non-EU	0.10	-	0.11	-	0.14	-	0.10	-
<u>Educational Attainment</u>								
No graduation	0.12	-	0.14	-	0.09	-	0.12	-
First stage of secondary level	0.48	-	0.53	-	0.48	-	0.47	-
Second stage of secondary level	0.26	-	0.23	-	0.29	-	0.26	-
Advanced tech. college entrance qualification	0.04	-	0.03	-	0.05	-	0.04	-
General qualification for university entrance	0.10	-	0.06	-	0.09	-	0.11	-
<u>Vocational Attainment</u>								
No vocational degree	0.34	-	0.43	-	0.32	-	0.32	-
In-plant training	0.53	-	0.48	-	0.56	-	0.55	-
Off-the-job training, voc. school, tech. school	0.06	-	0.05	-	0.05	-	0.06	-
University, advanced technical college	0.07	-	0.04	-	0.07	-	0.07	-
<u>Employment history</u>								
Previous Unemployment Duration in months	9.38	7.66	9.14	7.55	8.51	7.39	9.57	7.72
Duration of last employment in months	20.74	30.26	17.52	27.22	21.71	32.52	21.65	30.82
Log(wage) of last employment	3.61	1.17	3.59	1.12	3.47	1.32	3.63	1.16
No last employment observed	0.08	-	0.08	-	0.11	-	0.08	-
Share of days in emp., 1 st year before program	0.19	-	0.19	-	0.21	-	0.18	-
Share of days in emp., 2 nd year before program	0.38	-	0.36	-	0.40	-	0.38	-
Share of days in emp., 3 rd year before program	0.43	-	0.41	-	0.41	-	0.43	-
Share of days in emp., 4 th year before program	0.45	-	0.42	-	0.44	-	0.46	-
Share of days in unemp., 1 st year before program	0.67	-	0.68	-	0.64	-	0.67	-
Share of days in unemp., 2 nd year before program	0.39	-	0.43	-	0.36	-	0.39	-
Share of days in unemp., 3 rd year before program	0.34	-	0.37	-	0.33	-	0.33	-
Share of days in unemp., 4 th year before program	0.30	-	0.33	-	0.27	-	0.29	-
<u>Outcome variables</u>								
Employment two years after program entry	0.35	-	0.35	-	0.33	-	0.38	-
Employment one year after program exit	0.34	-	0.35	-	0.34	-	0.33	-
Number of Observations	3162		700		291		2171	

Table 2a. Estimated Effect of Treatment Duration on Employment Probability from Linear Probability Model

Dependent Variable: Employment status at time 2 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
Panel A: Only Control for Treatment Duration								
Constant	0.3354	0.0186	0.2989	0.0335	0.2948	0.0533	0.3052	0.0818
Treatment Duration	0.0001	0.0001	0.0005	0.0004	0.0006	0.0011	0.0002	0.0025
Square of Treatment Duration			-1.11E-06	8.49E-07	-1.71E-06	6.04E-06	2.18E-06	2.38E-05
Cube of Treatment Duration					9.89E-10	9.96E-09	-1.38E-08	8.81E-08
Fourth Power of Treatment Duration							1.87E-11	1.10E-10
Adjusted R Squared	-0.0002		0.0001		-0.0002		-0.0005	
Number of Observations	3162		3162		3162		3162	
Panel B: Control for Treatment Duration and Other Variables								
Constant	-0.1032	0.4742	-0.1466	0.4746	-0.1839	0.4759	-0.1572	0.4791
Treatment Duration	3.72E-05	0.0001	0.0007	0.0003	0.0016	0.0010	0.0006	0.0023
Square of Treatment Duration			-1.52E-06	8.19E-07	-7.42E-06	5.71E-06	3.01E-06	2.24E-05
Cube of Treatment Duration					9.82E-09	9.40E-09	-3.00E-08	8.30E-08
Fourth Power of Treatment Duration							5.02E-11	1.04E-10
Other Control Variables: See Table 1								
Adjusted R Squared	0.1371		0.1378		0.1378		0.1376	
Number of Observations	3130		3130		3130		3130	

Table 2b. Estimated Effect of Treatment Duration on Employment Probability from Linear Probability Model

Dependent Variable: Employment status 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
Panel A: Only Control for Treatment Duration								
Constant	0.3297	0.0186	0.3158	0.0334	0.3504	0.0532	0.3603	0.0816
Treatment Duration	0.00007	0.0001	0.0002	0.0004	-0.0006	0.0011	-0.0009	0.0025
Square of Treatment Duration			-4.25E-07	8.46E-07	4.56E-06	6.03E-06	8.24E-06	2.37E-05
Cube of Treatment Duration					-8.30E-09	9.93E-09	-2.23E-08	8.79E-08
Fourth Power of Treatment Duration							1.77E-11	1.10E-10
Adjusted R Squared	-0.0001		-0.0003		-0.0004		-0.0007	
Number of Observations	3162		3162		3162		3162	
Panel B: Control for Treatment Duration and Other Variables								
Constant	-0.0821	0.4774	-0.1036	0.4780	-0.1056	0.4794	-0.0765	0.4827
Treatment Duration	0.00006	0.0001	0.0004	0.0003	0.0004	0.0010	-0.0007	0.0024
Square of Treatment Duration			-7.53E-07	8.25E-07	-1.06E-06	5.75E-06	1.03E-05	2.25E-05
Cube of Treatment Duration					5.17E-10	9.47E-09	-4.30E-08	8.36E-08
Fourth Power of Treatment Duration							5.49E-11	1.05E-10
Other Control Variables: See Table 1								
Adjusted R Squared	0.1200		0.12		0.1197		0.1195	
Number of Observations	3130		3130		3130		3130	

Table 3a. IV Estimation of Effect of Treatment Duration on Employment Probability from Linear Probability Model

Dependent Variable: Employment status at time 2 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
Panel A: Only Control for Treatment Duration								
Constant	0.3433	0.0236	0.3246	0.0602	0.3426	0.1123	0.0080	0.1887
Treatment Duration	0.0000	0.0001	0.0002	0.0006	-0.0001	0.0020	0.0104	0.0051
Square of Treatment Duration			-4.93E-07	1.40E-06	1.57E-06	1.05E-05	-9.80E-05	4.47E-05
Cube of Treatment Duration					-3.27E-09	1.63E-08	3.57E-07	1.56E-07
Fourth Power of Treatment Duration							-4.40E-10	1.88E-10
Adjusted R Squared	-0.0002		-0.0001		-0.0006		0	
Number of Observations	3162		3162		3162		3162	
Hausman Test: Chi-Squared	0.3000		0.3200		0.2600		9.7100	
Probability>Chi-Squared	0.5829		0.8523		0.8766		0.0078	
Panel B: Control for Treatment Duration and Other Variables								
Constant	-0.0992	0.4746	-0.1995	0.4772	-0.2855	0.4850	-0.5157	0.5051
Treatment Duration	1.87E-05	0.0001	0.0012	0.0006	0.0032	0.0019	0.0108	0.0048
Square of Treatment Duration			-2.82E-06	1.32E-06	-1.33E-05	9.97E-06	-8.57E-05	4.22E-05
Cube of Treatment Duration					1.67E-08	1.55E-08	2.79E-07	1.48E-07
Fourth Power of Treatment Duration							-3.21E-10	1.78E-10
Other Control Variables: See Table 1								
Adjusted R Squared	0.1371		0.1369		0.1359		0.1299	
Number of Observations	3130		3130		3130		3130	
Hausman Test: Chi-Squared	0.0400		1.5700		2.4300		7.3300	
Probability>Chi-Squared	1.0000		1.0000		1.0000		1.0000	

Table 3b. IV Estimation of Effect of Treatment Duration on Employment Probability from Linear Probability Model

Dependent Variable: Employment status 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
Panel A: Only Control for Treatment Duration								
Constant	0.3232	0.0235	0.2993	0.0600	0.3603	0.1120	0.2232	0.1878
Treatment Duration	0.00010	0.0001	0.0004	0.0006	-0.0009	0.0020	0.0034	0.0051
Square of Treatment Duration			-6.27E-07	1.40E-06	6.36E-06	1.05E-05	-3.44E-05	4.44E-05
Cube of Treatment Duration					-1.11E-08	1.63E-08	1.37E-07	1.56E-07
Fourth Power of Treatment Duration							-1.80E-10	1.87E-10
Adjusted R Squared	-0.0002		-0.0005		-0.0005		0	
Number of Observations	3162		3162		3162		3162	
Hausman Test: Chi-Squared	0.2000		0.4200		0.1600		1.4500	
Probability>Chi-Squared	0.6508		0.8120		0.9249		0.4834	
Panel B: Control for Treatment Duration and Other Variables								
Constant	-0.1007	0.4779	-0.1940	0.4811	-0.2227	0.4888	-0.2637046	0.5076307
Treatment Duration	0.00014	0.0001	0.0013	0.0006	0.0019	0.0019	0.0033	0.0048
Square of Treatment Duration			-2.62E-06	1.34E-06	-6.13E-06	1.00E-05	-1.90E-05	4.24E-05
Cube of Treatment Duration					5.58E-09	1.57E-08	5.23E-08	1.49E-07
Fourth Power of Treatment Duration							-5.71E-11	1.79E-10
Other Control Variables: See Table 1								
Adjusted R Squared	0.1198		0.1175		0.1167		0.1158	
Number of Observations	3130		3130		3130		3130	
Hausman Test: Chi-Squared	0.8200		3.5200		3.8800		3.7200	
Probability>Chi-Squared	1.0000		1.0000		1.0000		1.0000	

Table 4. Estimated GPS: Linear Regression of Treatment Level on Covariates

	(1)	(2)	(3)	(4)	(5)	(6)
	Actual Duration		Planned Duration		Actual=Planned Duration	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Previous unemployment duration in months	0.7636	0.3451	0.2429	0.3170	0.6667	0.3900
Age	-2.1639	7.4060	-8.7657	6.8028	-12.2600	8.3077
Age squared	-0.0025	0.1912	0.1927	0.1757	0.2692	0.2129
Age cubic	0.0006	0.0017	-0.0014	0.0016	-0.0019	0.0019
Duration of last employment	0.0023	0.0021	0.0006	0.0019	0.0005	0.0022
No information about last employment	-11.0729	18.4469	4.3886	16.9444	-2.1205	21.2456
No wage of last employment observed	34.2926	23.4434	1.8158	21.5339	3.0140	26.7859
Log(wage) of last employment	4.3508	3.4744	1.4269	3.1914	0.0431	3.8855
Educational Attainment 2	-189.4403	73.3736	-111.4210	67.3971	-194.2237	85.3845
Educational Attainment 3	-209.7043	82.7537	-164.1688	76.0132	-291.6277	94.1019
Educational Attainment 4	-371.8883	143.9889	-401.1327	132.2606	-484.4741	154.6772
Educational Attainment 5	-472.2201	148.2782	-334.0777	136.2005	-462.1233	154.9476
Vocational Attainment 2	63.5927	51.9122	59.1829	47.6838	98.3178	60.2859
Vocational Attainment 3	59.4316	91.8876	65.4984	84.4031	17.7366	98.2616
Vocational Attainment 4	583.0145	186.1424	428.7167	170.9806	432.9606	193.1385
Foreigner EU	-6.0217	13.0374	-4.2181	11.9755	-2.0789	6.7871
Foreigner Non-EU	6.3966	5.9281	1.4674	5.4452	1.7383	14.4846
Share of days in emp., 1 st year before program	-5.2568	2.4743	-3.7461	2.2727	-7.0300	2.6756
Share of days in emp., 2 nd year before program	-1.3409	1.9019	-2.9754	1.7470	-1.3212	2.0579
Share of days in emp., 3 rd year before program	-1.0017	1.9274	0.5746	1.7704	-0.0778	2.1196
Share of days in emp., 4 th year before program	-2.0666	1.6206	-3.1272	1.4886	-3.4186	1.7991
Share of days in unemp., 1 st year before	-1.5590	2.5644	1.0997	2.3555	-2.4781	2.8353
Share of days in unemp., 2 nd year before	-5.0441	1.8696	-4.1941	1.7173	-4.4725	2.0556
Share of days in unemp., 3 rd year before	-0.8236	1.8962	-0.7127	1.7418	-0.9334	2.1055
Share of days in unemp., 4 th year before	-5.4048	1.7663	-4.3744	1.6225	-5.0938	1.9534
Disability low degree	44.5817	21.5007	28.8693	19.7494	19.5723	22.1091
Disability medium degree	20.1793	20.5201	4.6797	18.8487	12.5898	10.9852
Disability high degree	-27.3114	6.2610	-42.4251	5.7510	-40.6165	6.9403
Number of children	-7.1170	3.2875	-5.5518	3.0197	-4.1266	3.6564
Youngest Child < 4 years	5.6853	9.3889	8.0336	8.6242	6.0367	10.4345
Youngest Child < 14 years	1.1432	7.0381	4.1765	6.4648	-3.2098	7.7612
Regional unemployment rate	433.1507	56.5878	387.5052	51.9786	451.8261	60.0793
Regional type 2	0.9976	5.7695	-2.7450	5.2996	-1.3321	6.1091
Regional type 3	-15.7442	7.0030	-20.5547	6.4326	-18.5218	7.5202
Regional type 4	15.2312	10.5725	21.1639	9.7113	24.9902	11.6339
Regional type 5	-10.3845	8.7881	-2.5635	8.0723	2.1099	9.5000
Educational Attainment 2 * age	9.9967	4.2355	5.2211	3.8905	9.4409	4.8708
Educational Attainment 3 * age	10.6215	4.7983	7.8531	4.4074	14.9481	5.4105
Educational Attainment 4 * age	21.9266	7.9930	23.6488	7.3419	27.9872	8.5455
Educational Attainment 5 * age	28.8547	8.0434	19.5861	7.3883	26.3502	8.5103
Educational Attainment 2 * age squared	-0.1259	0.0582	-0.0600	0.0534	-0.1118	0.0662
Educational Attainment 3 * age squared	-0.1230	0.0659	-0.0846	0.0605	-0.1758	0.0739
Educational Attainment 4 * age squared	-0.2846	0.1058	-0.3097	0.0972	-0.3602	0.1122
Educational Attainment 5 * age squared	-0.3853	0.1055	-0.2496	0.0969	-0.3365	0.1128
Vocational Attainment 2 * age	-2.7761	2.9808	-3.1187	2.7380	-5.0758	3.4419
Vocational education 3 * age	-2.3590	5.2393	-3.0691	4.8125	-0.3889	5.6073
Vocational education 4 * age	-28.0278	9.5537	-19.4233	8.7755	-20.6672	9.9959
Vocational Attainment 2 * age squared	0.0320	0.0404	0.0402	0.0371	0.0641	0.0465
Vocational education 3 * age squared	0.0381	0.0700	0.0459	0.0643	0.0160	0.0749
Vocational education 4 * age squared	0.3515	0.1191	0.2392	0.1094	0.2676	0.1256
Constant	220.4666	95.3657	353.6884	87.5979	403.7700	108.0968
Adjusted R Squared	0.1966		0.1999		0.2372	
Number of Observations	3130		3130		2145	

**Table 5. Balance in Covariates with and without Adjustment Based on Actual Duration:
t-statistics for Equality of Means**

Covariate	(1)	(2)	(3)	(4)	(5)	(6)
	Unadjusted			Adjusted		
	[11, 137]	[138,247]	[248,395]	[11, 137]	[138,247]	[248,395]
Age	4.37	0.13	-4.53	0.94	-0.14	0.00
<u>Disability</u>						
No disability	-2.42	-1.39	3.92	-0.46	-0.57	-0.80
Disability low degree	1.70	0.94	-2.71	0.82	0.48	-0.86
Disability medium degree	2.38	0.75	-3.19	0.82	-0.09	0.09
Disability high degree	0.64	1.07	-1.79	-0.03	0.48	-0.19
<u>Citizenship</u>						
German	0.76	1.25	-2.11	-0.31	0.42	-0.42
Foreigner EU	-2.48	0.89	1.53	-0.40	0.35	-0.57
Foreigner Non-EU	0.20	-1.69	-1.61	0.50	-0.59	0.65
<u>Educational Attainment</u>						
No graduation (1)	-2.38	-1.05	3.51	-0.11	0.02	-0.14
First stage of secondary level (2)	-4.35	5.83	10.76	0.16	-1.63	0.83
Second stage of secondary level (3)	2.52	2.12	-4.82	-0.42	0.52	-0.65
Advanced technical college entrance qualification (4)	2.00	1.65	-3.78	-0.02	0.32	-0.05
General qualification for university entrance (5)	4.95	6.82	-12.51	0.56	1.97	-0.10
<u>Vocational Attainment</u>						
No vocational degree (1)	-5.60	-3.39	9.35	-0.30	-0.74	0.72
In-plant training (2)	1.52	-1.44	0.02	-0.20	-0.61	-0.25
Off-the-job training, vocational school, technical school (3)	1.79	3.04	-5.07	-0.25	1.12	0.19
University, advanced technical college (4)	5.95	6.58	-13.29	1.36	1.89	-0.48
<u>Employment History</u>						
Previous Unemployment Duration	2.26	-3.42	1.39	1.14	-1.33	1.12
Duration of last employment	3.00	-0.92	-2.02	0.64	-0.50	-0.05
Log(wage) of last employment	-0.59	-1.17	1.84	-0.19	-0.56	-0.12
No last employment observed	2.04	1.27	-3.42	0.56	0.55	0.07
Share of days in emp., 1 st year before program	-2.79	1.56	1.12	-1.23	0.55	-0.52
Share of days in emp., 2 nd year before program	-2.08	2.56	-0.66	-1.21	1.01	-0.41
Share of days in emp., 3 rd year before program	-0.48	0.82	-0.40	-0.57	0.34	-0.24
Share of days in emp., 4 th year before program	1.38	-0.30	-1.06	0.02	-0.31	-0.23
Share of days in unemp., 1 st year before program	1.58	-2.62	1.22	1.09	-1.04	0.66
Share of days in unemp., 2 nd year before program	-0.14	-3.97	4.40	0.78	-1.47	0.76
Share of days in unemp., 3 rd year before program	-0.31	-3.75	4.33	0.80	-1.38	0.74
Share of days in unemp., 4 th year before program	-2.75	-1.96	4.87	-0.07	-0.45	0.65
<u>Regional indicators</u>						
Regional type 1	6.22	4.74	-11.49	-0.38	1.04	-0.43
Regional type 2	4.22	2.48	-6.93	0.86	0.86	-0.93
Regional type 3	-6.00	-4.53	11.02	-0.18	-1.19	0.98
Regional type 4	1.12	1.22	-2.43	0.01	0.56	-0.06
Regional type 5	-4.75	-3.14	8.18	-0.34	-0.90	-0.90
Regional unemployment rate	8.92	6.75	-16.70	0.48	1.49	-0.89

Bold numbers indicate significance at the 5% level

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

Covariate	(1)	(2)	(3)	(4)	(5)	(6)
	[11, 167]	[168,271]	[272,395]	[11, 167]	[168,271]	[272,395]
	Unadjusted			Adjusted		
Age	2.99	0.40	-3.85	0.70	-0.09	-0.12
<u>Disability</u>						
No disability	-3.36	1.22	2.47	0.01	0.95	-0.52
Disability low degree	2.22	-1.47	-0.90	-0.42	-0.97	0.70
Disability medium degree	3.11	-0.14	3.39	1.10	-0.36	0.01
Disability high degree	1.53	0.65	2.46	0.33	0.19	-0.40
<u>Citizenship</u>						
German	-0.53	-1.21	1.95	0.40	-0.40	0.75
Foreigner EU	1.28	0.95	-2.51	-0.29	0.25	-0.67
Foreigner Non-EU	-2.03	0.50	1.76	-0.22	0.34	-0.13
<u>Educational Attainment</u>						
No graduation (1)	-2.66	-0.85	3.98	-0.47	-0.17	0.59
First stage of secondary level (2)	-5.69	-2.94	9.84	-0.09	-0.61	0.70
Second stage of secondary level (3)	2.37	0.89	-3.69	-0.60	0.06	-0.54
Advanced technical college entrance qualification (4)	2.96	0.35	-3.76	0.46	-0.12	-0.31
General qualification for university entrance (5)	7.12	4.40	-13.27	1.47	1.29	-0.65
<u>Vocational Attainment</u>						
No vocational degree (1)	-6.24	-2.00	9.40	-0.87	-0.34	1.34
In-plant training (2)	0.91	-0.90	-0.04	-0.04	-0.35	-0.65
Off-the-job training, vocational school, technical school (3)	2.84	1.24	-4.63	0.27	0.25	-0.59
University, advanced technical college (4)	7.44	4.48	-13.74	1.81	1.28	-0.45
<u>Employment History</u>						
Previous Unemployment Duration	2.28	-3.40	1.18	1.46	-1.49	1.20
Duration of last employment	2.42	-0.95	-1.69	0.95	-0.50	-0.36
Log(wage) of last employment	-1.51	0.41	1.26	-0.26	0.23	-0.45
No last employment observed	2.57	-0.50	2.37	0.74	-0.32	0.36
Share of days in emp., 1 st year before program	-1.33	-0.11	1.64	-0.31	0.06	-0.38
Share of days in emp., 2 nd year before program	0.16	0.85	0.77	0.04	0.37	-0.96
Share of days in emp., 3 rd year before program	0.40	-0.07	-0.38	0.07	-0.06	-0.46
Share of days in emp., 4 th year before program	1.45	-0.58	-1.01	0.39	-0.37	-0.70
Share of days in unemp., 1 st year before program	0.31	-1.45	1.25	0.49	-0.70	0.92
Share of days in unemp., 2 nd year before program	-2.06	-1.67	4.21	-0.03	-0.52	1.18
Share of days in unemp., 3 rd year before program	-2.11	-1.93	4.56	0.11	-0.64	1.29
Share of days in unemp., 4 th year before program	-3.81	-0.10	4.45	-0.53	0.24	1.08
<u>Regional indicators</u>						
Regional type 1	7.71	2.54	-11.76	-0.31	0.77	-0.57
Regional type 2	3.54	1.69	-5.93	0.67	0.43	-1.51
Regional type 3	-6.07	-3.74	11.21	0.81	0.94	1.32
Regional type 4	1.20	0.71	-2.16	-0.29	0.18	-0.05
Regional type 5	-5.68	-0.50	7.04	-1.15	0.25	0.99
Regional unemployment rate	11.46	2.43	-16.09	1.16	-0.24	-1.22

Bold numbers indicate significance at the 5% level

**Table 7. Balance in Covariates with and without Adjustment Based on Planned=Actual Duration:
t-statistics for Equality of Means**

Covariate	(1)	(2)	(3)	(4)	(5)	(6)
	Unadjusted			Adjusted		
	[11, 170]	[171, 271]	[272,395]	[11, 170]	[171, 271]	[272,395]
Age	3.04	-0.21	-2.85	0.69	-0.52	0.86
<u>Disability</u>						
No disability	-4.79	1.90	2.78	-1.42	1.30	0.10
Disability low degree	3.82	-2.23	-1.44	1.06	-1.28	0.06
Disability medium degree	2.57	-0.47	-2.09	1.05	-0.61	0.09
Disability high degree	1.67	0.95	-2.73	0.45	0.20	-0.56
<u>Citizenship</u>						
German	1.73	1.26	-3.13	0.06	0.14	-0.60
Foreigner EU	-0.76	-0.55	1.37	0.20	-0.01	-0.49
Foreigner Non-EU	-1.53	-1.11	2.77	-0.15	-0.14	0.81
<u>Educational Attainment</u>						
No graduation (1)	-1.36	-1.74	3.26	0.04	-0.35	0.29
First stage of secondary level (2)	-4.09	-4.47	9.12	0.55	-0.82	0.70
Second stage of secondary level (3)	1.04	2.00	-3.23	-1.21	0.26	-0.21
Advanced technical college entrance qualification (4)	2.12	0.61	-2.82	0.15	-0.20	-0.19
General qualification for university entrance (5)	5.15	5.81	-11.79	0.77	1.61	-0.78
<u>Vocational Attainment</u>						
No vocational degree (1)	-4.06	-3.41	7.90	-0.08	-0.65	0.78
In-plant training (2)	-0.75	-0.05	0.82	-0.78	0.07	-0.20
Off-the-job training, vocational school, technical school (3)	2.64	1.13	-3.92	0.13	0.00	-0.17
University, advanced technical college (4)	6.32	5.23	-12.40	1.84	1.24	-0.65
<u>Employment History</u>						
Previous Unemployment Duration	2.71	-4.20	1.79	1.60	-1.86	1.61
Duration of last employment	0.82	0.05	-0.89	0.09	-0.07	0.17
Log(wage) of last employment	-1.47	0.30	1.16	-0.37	0.18	0.09
No last employment observed	1.65	-0.14	-1.52	0.53	-0.19	0.06
Share of days in emp., 1 st year before program	-2.21	0.97	1.19	-0.88	0.62	-0.71
Share of days in emp., 2 nd year before program	-1.08	1.71	-0.75	-0.78	0.72	-0.71
Share of days in emp., 3 rd year before program	-0.80	1.32	-0.62	-0.65	0.59	-0.62
Share of days in emp., 4 th year before program	-0.32	1.13	-0.90	-0.53	0.44	-0.47
Share of days in unemp., 1 st year before program	1.34	-3.02	1.91	1.03	-1.36	1.44
Share of days in unemp., 2 nd year before program	-0.22	-3.18	3.67	0.87	-1.10	1.29
Share of days in unemp., 3 rd year before program	-0.59	-3.24	4.11	0.81	-1.08	1.44
Share of days in unemp., 4 th year before program	-1.88	-1.81	3.87	0.31	-0.47	1.29
<u>Regional indicators</u>						
Regional type 1	5.96	4.45	-11.09	-0.74	0.64	-0.57
Regional type 2	1.44	3.05	-4.79	-0.13	0.93	-1.03
Regional type 3	-5.87	-4.99	11.63	0.42	-0.99	2.03
Regional type 4	1.79	0.28	-2.13	0.26	-0.10	-0.39
Regional type 5	-2.19	-2.47	4.92	0.19	-0.43	-0.05
Regional unemployment rate	8.66	5.02	-14.72	0.26	0.31	-0.88

Bold numbers indicate significance at the 5% level

Table 8. Estimated Dose Response Functions

	(1)	(2)	(3)	(4)	(5)	(6)
	Actual Duration		Planned Duration		Planned=Actual Duration	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Panel A: Outcome Variable: Employment status at time 2 years after entry into the program						
GPS	-198.9665	119.7742	-91.8408	107.6389	34.0560	130.0212
GPS ²	60057.2100	46972.1400	60457.1900	39749.1900	21345.6400	46988.9700
GPS ³	-4442237.0000	5712302.0000	-4925847.0000	4326112.0000	-3043581.0000	5082415.0000
Program Duration	0.0015	0.0015	0.0038	1.9801E-03	0.0025	2.3511E-03
Program Duration ²	-5.56E-06	7.99E-06	-4.22E-06	9.97E-06	-1.94E-06	1.19E-05
Program Duration ³	2.59E-09	1.30E-08	-1.30E-08	1.67E-08	-8.81E-09	1.99E-08
GPS*Program Duration	0.2792	0.5618	-1.4210	0.5881	-0.9737	0.6570
GPS ² *Program Duration	-109.3040	63.0600	-23.6346	56.5314	20.2568	64.9347
GPS*Program Duration ²	0.0006	0.0012	0.0035	0.0013	0.0018	0.0015
Constant	0.4127	0.1063	0.2268	0.1010	0.1620	0.1307
Adjusted R Squared	-0.0002		0.0013		-0.0010	
Number of Observations	3130		3130		2145	
Panel B: Outcome Variable: Employment status at time 1 year after exit from the program						
GPS	-20.4268	119.5165	-102.1634	107.3369	40.6767	129.7814
GPS ²	9059.1920	46871.0700	50623.5700	39637.6800	11637.4800	46902.3300
GPS ³	-732274.6000	5700010.0000	-2130617.0000	4313976.0000	-1266264.0000	5073043.0000
Program Duration	-0.0001	0.0017	0.0020	0.0020	-3.72E-05	0.0023
Program Duration ²	3.49E-06	7.98E-06	4.06E-06	9.94E-06	9.80E-06	1.19E-05
Program Duration ³	-9.01E-09	1.30E-08	-2.34E-08	1.66E-08	-2.48E-08	1.98E-08
GPS*Program Duration	-0.0346	0.5606	-1.2096	0.5864	-0.7756	0.6558
GPS ² *Program Duration	-25.7458	62.9243	-51.5993	56.3728	2.9993	64.8149
GPS*Program Duration ²	0.0003	0.0012	0.0032	0.0013	0.0015	0.0015
Constant	0.3421	0.1060	0.3114	0.1007	0.2727	0.1304
Adjusted R Squared	-0.0020		0.0009		0.0022	
Number of Observations	3130		3130		2145	

Table 9. Instrumental Variable Estimations

	(1)	(2)	(3)	(4)
	Linear Probability Model		Probit Model	
	Treatment Effect	Std. Error	Treatment Effect	Std. Error
Panel A: Outcome Variable: Employment status at time 2 years after entry into the program				
Model 1	-0.0429	0.0456	-0.1192	0.1252
Model 2	0.0293	0.0290	0.0789	0.0778
Model 3	-0.0266	0.0224	-0.0721	0.0606
Model 4	-0.0150	0.0224	-0.0406	0.0606
Model 5	0.0099	0.0281	0.0269	0.0763
(Without Other Control Variables)				
Panel B: Outcome Variable: Employment status at time 1 year after exit from the program				
Model 1	0.0026	0.0455	0.0069	0.1237
Model 2	-0.0101	0.0289	-0.0276	0.0787
Model 3	-0.0371	0.0223	-0.1010	0.0607
Model 4	-0.0210	0.0224	-0.0570	0.0606
Model 5	-0.0019	0.0281	-0.0051	0.0763
(Without Other Control Variables)				
Panel C: Outcome Variable: Employment status at time 2 years after entry into the program				
Model 1	-0.1172	0.0440	-0.4060	0.1393
Model 2	-0.0170	0.0289	-0.0680	0.0895
Model 3	-0.0095	0.0237	-0.0400	0.0730
Model 4	-0.0062	0.0249	-0.0334	0.0763
Model 5	0.0303	0.0300	0.0861	0.0918
(With Other Control Variables: See Table 1)				
Panel D: Outcome Variable: Employment status at time 1 year after exit from the program				
Model 1	-0.0614	0.0444	-0.2033	0.1364
Model 2	-0.0535	0.0291	-0.1808	0.0893
Model 3	-0.0262	0.0238	-0.0867	0.0725
Model 4	-0.0150	0.0250	-0.0509	0.0760
Model 5	0.0133	0.0302	0.0460	0.0917
(With Other Control Variables: See Table 1)				

Bold numbers indicate significance at the 5% level

Note: we use 5 different cutoff points, respectively, to define the two groups with the shorter vs. the longer treatment duration. The different models 1 to 5 correspond to different cutoff points, from the 15% percentile to the 85% percentile

Figure 1. Distributions of Actual and Planned Training Durations

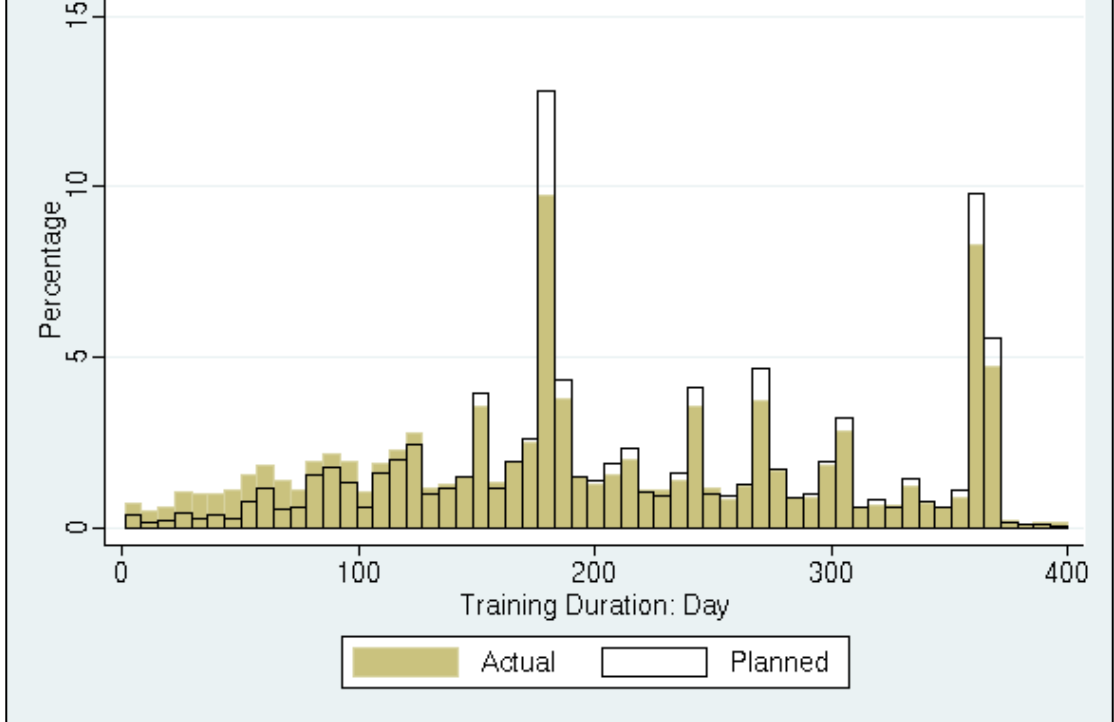


Figure 2a. Unadjusted Employment Probability at Time 2 Years after Entry into the Program Based on Actual Training Duration **Figure 2b. Unadjusted Employment Probability at Time 2 Years after Entry into the Program Based on Planned Training Duration**

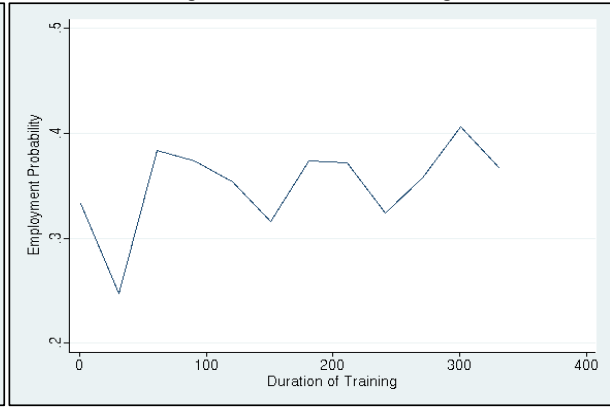
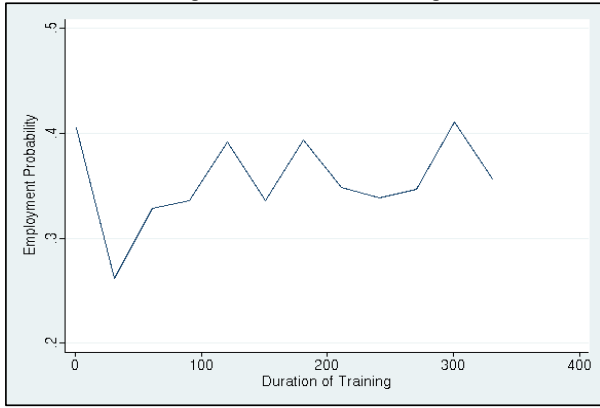


Figure 2c. Unadjusted Employment Probability at Time 2 Years after Entry into the Program Based on Subsample with Actual Training Duration Equal to Planned Duration **Figure 2d. Unadjusted Employment Probability at Time 1 Years after Exit from the Program Based on Actual Training Duration**

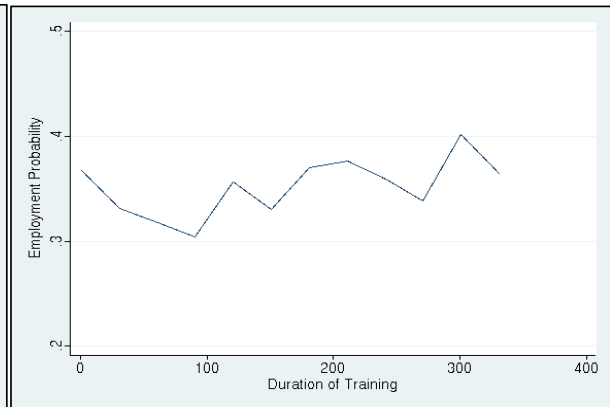
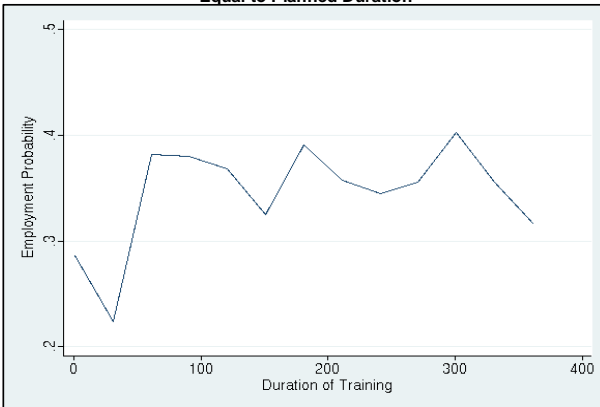


Figure 2e. Unadjusted Employment Probability at Time 1 Years after Exit from the Program Based on Planned Training Duration **Figure 2f. Unadjusted Employment Probability at Time 1 Years after Exit from the Program Based on Subsample with Actual Training Duration Equal to Planned Duration**

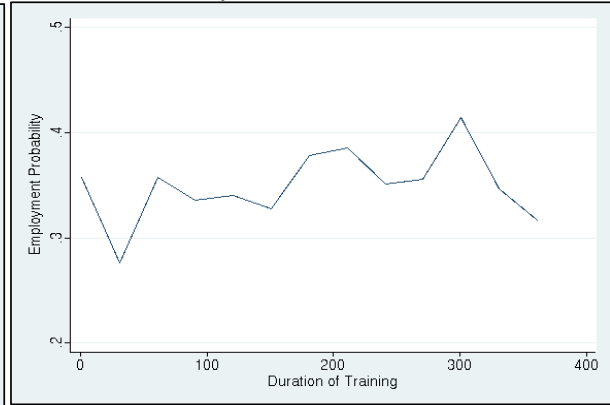
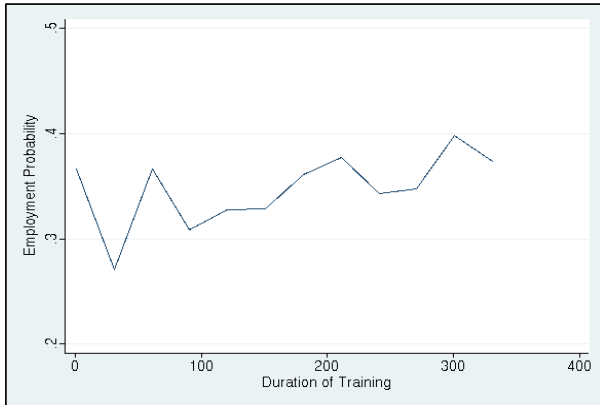


Figure 3a. t-statistics for Equality of Means of Pre-training Employment Status before and After Adjustment for GPS Based on Observed Training Duration



Figure 3b. t-statistics for Equality of Means of Pre-training Employment Status before and After Adjustment for GPS Based on Planned Training Duration

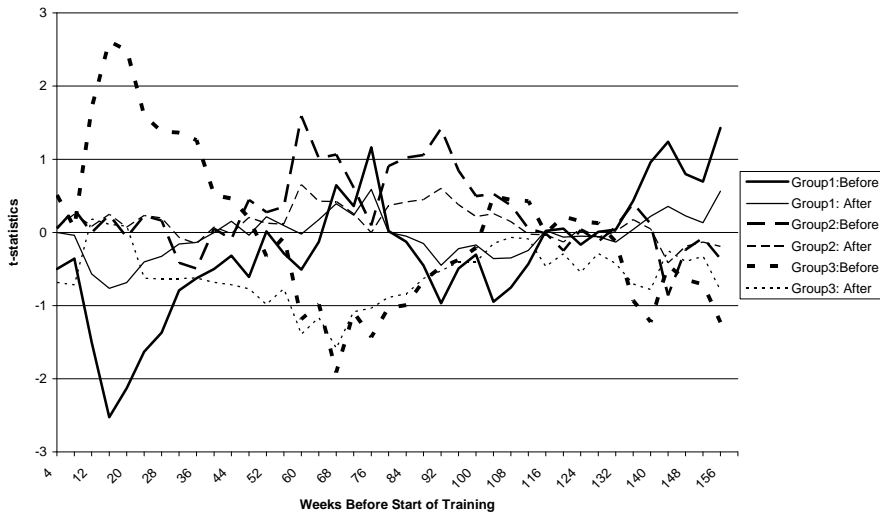


Figure 3c. t-statistics for Equality of Means of Pre-training Employment Status before and After Adjustment for GPS Based on Actual Training Duration Equal to Planned Duration

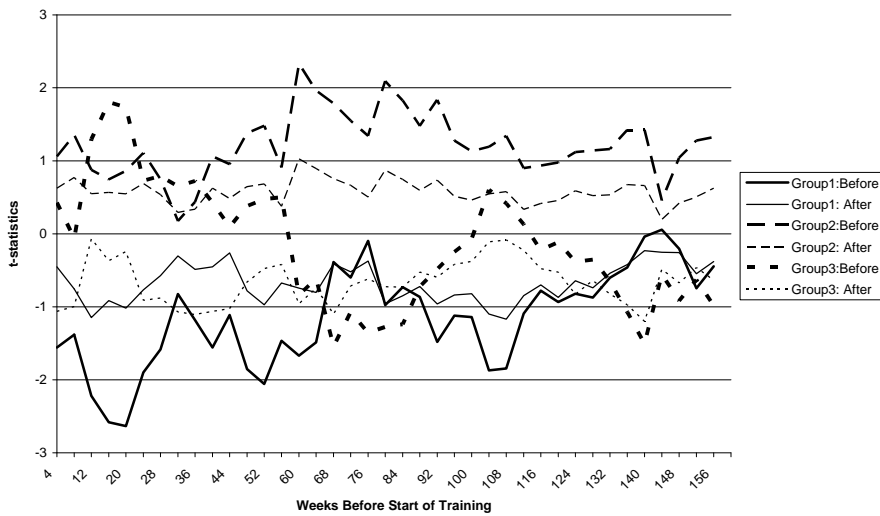


Figure 4. Common Support Condition

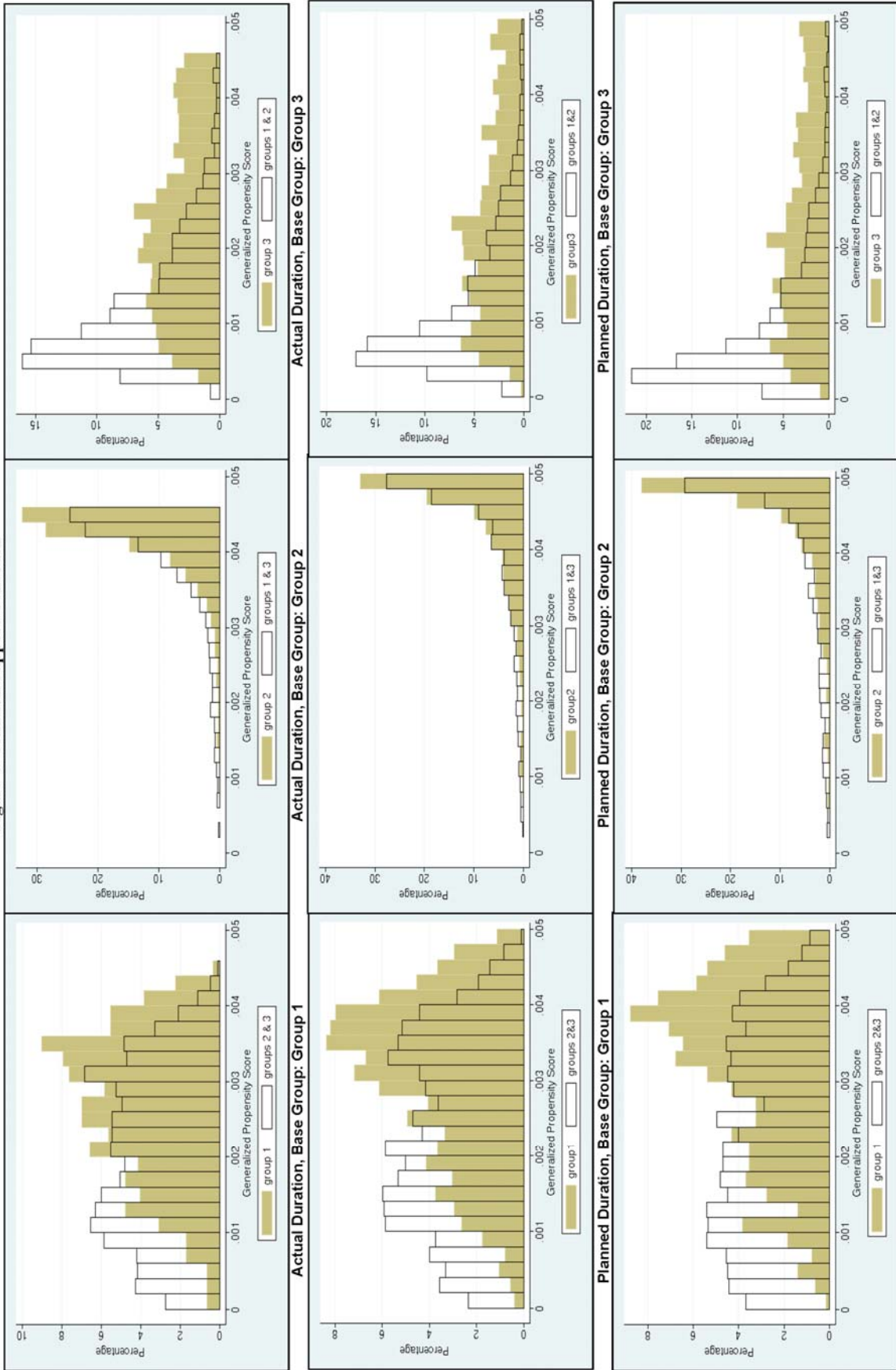


Figure 5a. Employment Probability at Time 2 Years after Entry into the Program Based on Actual Training Duration

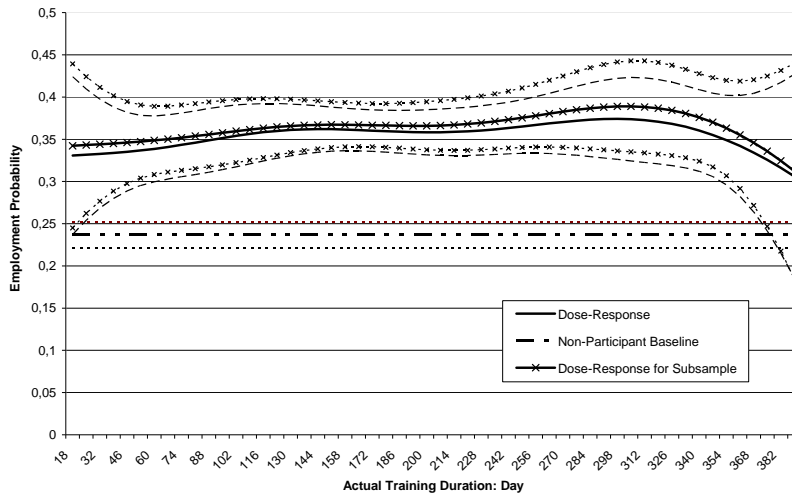


Figure 5b. Employment Probability at Time 2 Years after Entry into the Program Based on Planned Training Duration

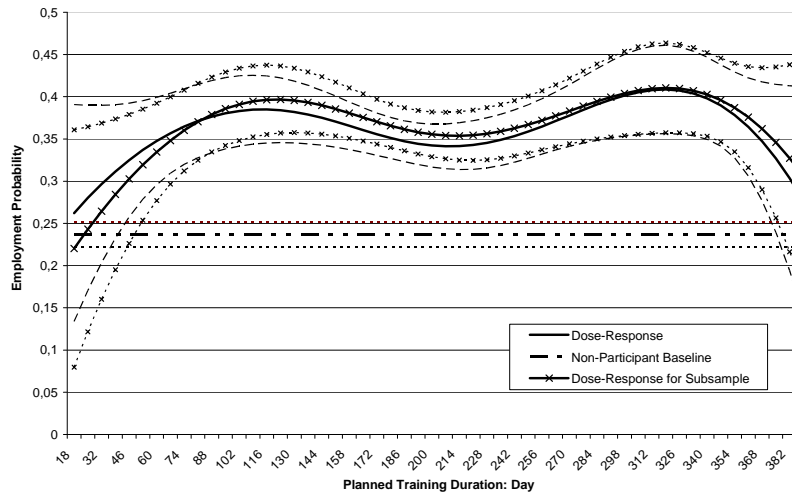
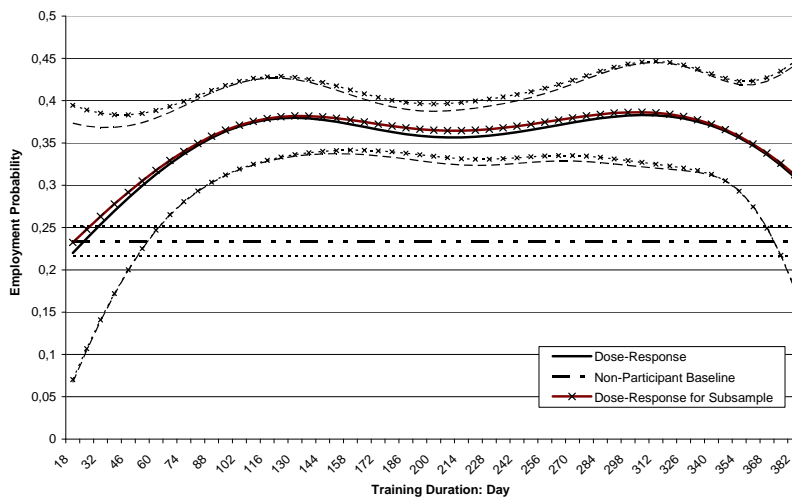


Figure 5c. Employment Probability at Time 2 Years after Entry into the Program Based on Subsample with Actual Training Duration Equal to Planned Duration



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

Subsample refers to the people who went through a training program exactly once.

Figure 6a. Employment Probability at Time 1 Year after Exit from the Program Based on Actual Training Duration

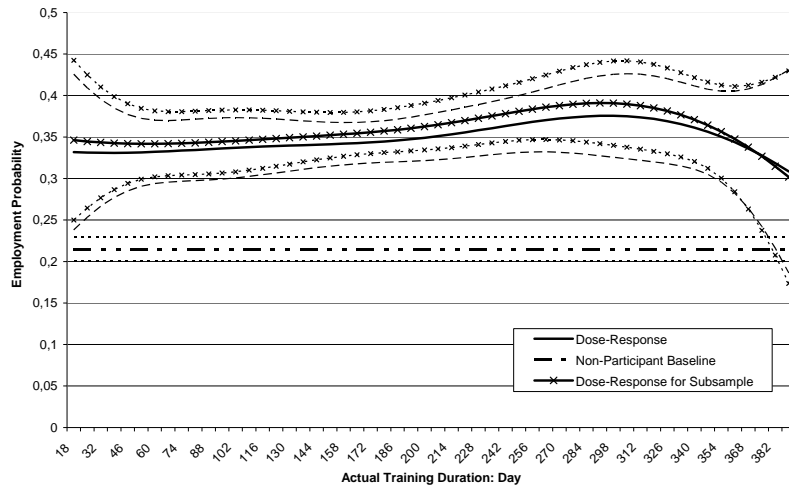


Figure 6b. Employment Probability at Time 1 Year after Exit from the Program Based on Planned Training Duration

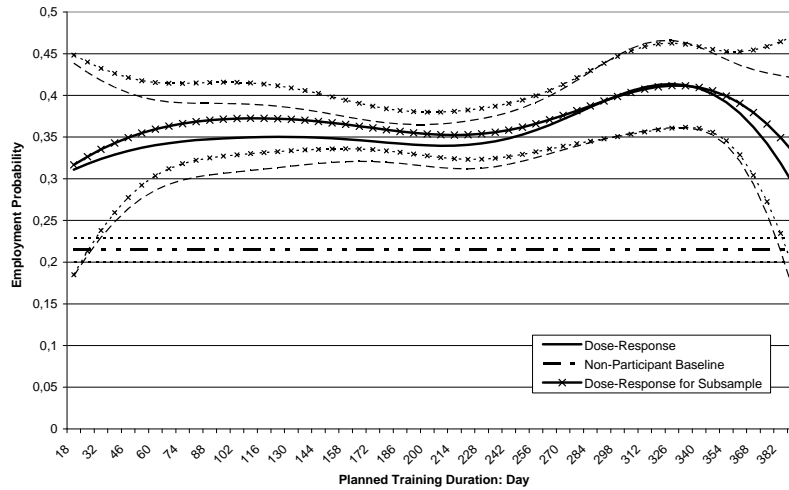
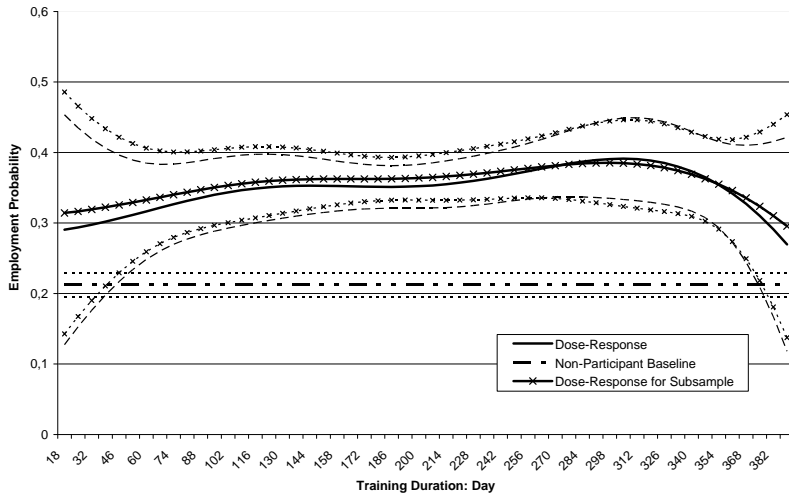


Figure 6c. Employment Probability at Time 1 Year after Exit from the Program Based on Subsample with Actual Training Duration Equal to Planned Duration



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

Subsample refers to the people who went through a training program exactly once.

Figure 7a. Estimated Employment Probability at Time 2 Years after Entry into the Program Based on Actual Duration

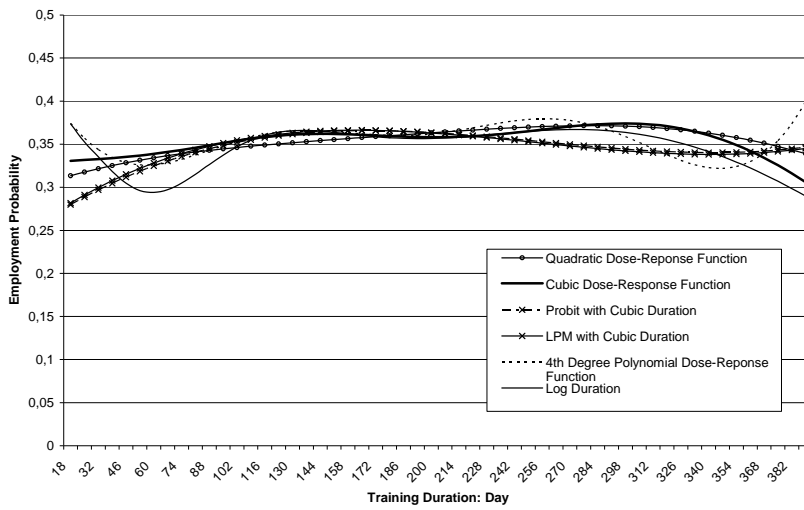


Figure 7b. Estimated Employment Probability at Time 2 Years after Entry into the Program Based on Planned Duration

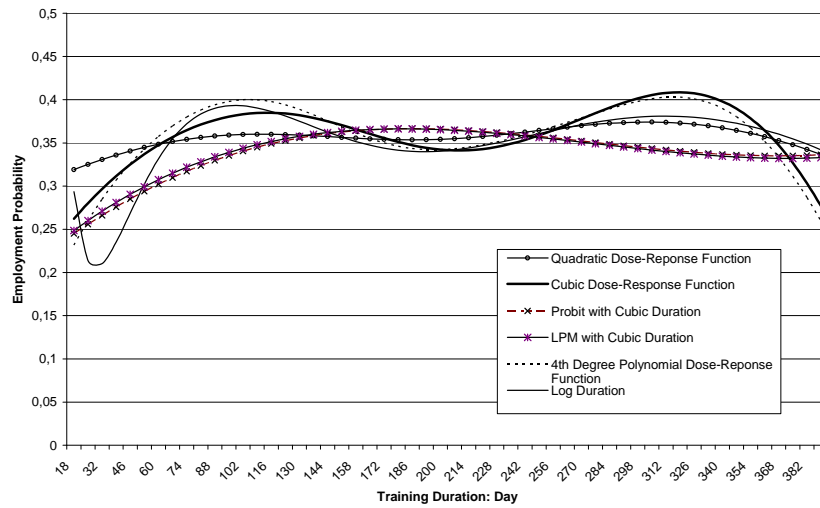
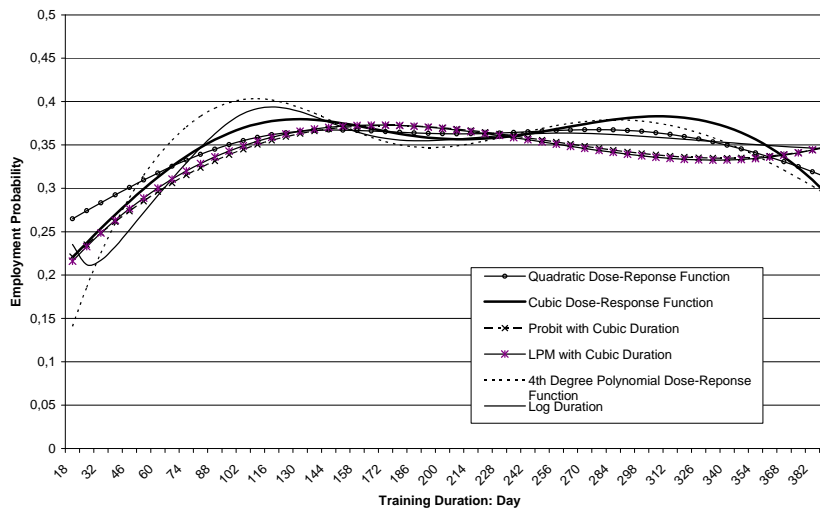


Figure 7c. Estimated Employment Probability at Time 2 Years after Entry into the Program Based on Subsample with Actual Training Duration Equal to Planned Duration



Note: We omit confidence intervals in this figure in order to have a clear picture of the main results.

Figure 8a. Estimated Employment Probability at Time 1 Year after Exit from the Program Based on Actual Duration

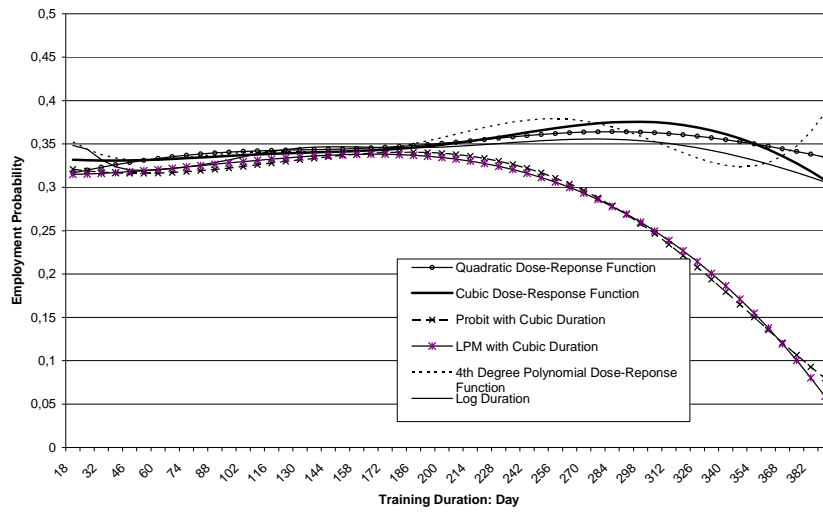


Figure 8b. Estimated Employment Probability at Time 1 Year after Exit from the Program Based on Planned Duration

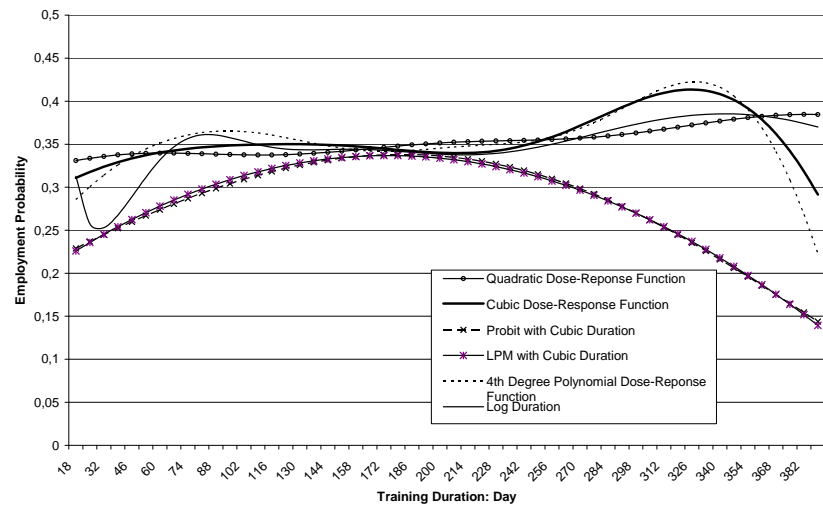
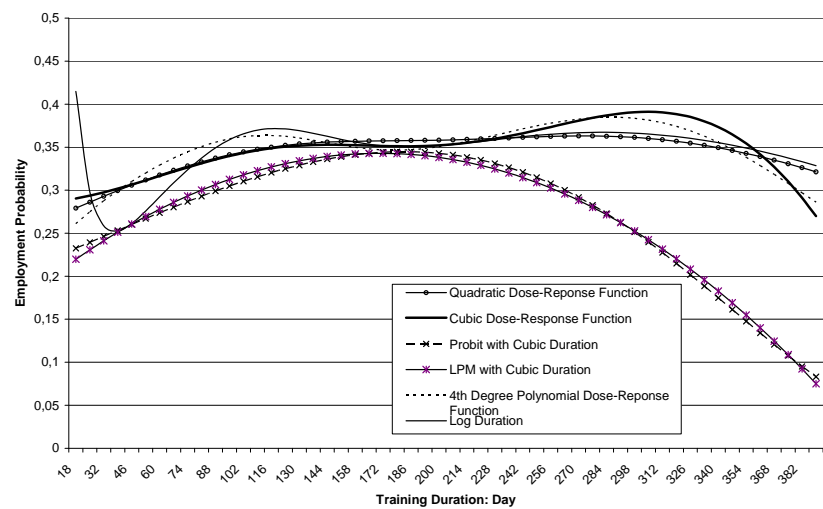


Figure 8c. Estimated Employment Probability at Time 1 Year after Exit from the Program Based on Subsample with Actual Training Duration Equal to Planned Duration



Note: We omit confidence intervals in this figure in order to have a clear picture of the main results.