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

Jochen Kluve
RWI Essen and IZA Bonn

Hilmar Schneider
IZA Bonn

Arne Uhlendorff
IZA Bonn and DIW DC

Zhong Zhao[2]
IZA Bonn

First Version: February 2007
This Version: December 18, 2007

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.

---

[1] The data used in this paper originate in the evaluation of continuous training programs as part of the evaluation of the so-called Hartz-Reforms. The corresponding report by IZA et al. 2007 (cf. references) contains details. We would like to thank 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, the World Bank and the Washington Statistical Society for valuable discussions and helpful comments. Arne Uhlendorff also thanks DIW DC, where part of this research was pursued during his stay in fall 2007. Jochen Kluve conducted part of this research within the *Collaborative Research Centre* (*Sonderforschungsbereich*) 475 "Reduction of Complexity for Multivariate Data Structures", located at the University of Dortmund and sponsored by the *German Science Foundation* (*DFG*). The usual disclaimer applies.

[2] Correspondence: kluve@rwi-essen.de, schneider@iza.org, uhlendorff@iza.org, zhao@iza.org

1. Introduction

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

In this paper we contribute to the literature on training programs by focusing on the dynamics inherent to the provision of training, i.e. we study the treatment effects that arise from variation in the treatment duration. We implement this analysis on the basis of data on training programs in Germany. The key feature of the data is the fact that the treatment duration varies almost continuously from approximately 1 week duration up to approximately 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). Training programs leading to the acquisition of a degree are not considered, since the degree requirement generates discontinuities in the distribution of treatment durations, and the objective of the analysis in this paper is to estimate the employment outcomes associated with each level of a continuous treatment.

The evaluation question that corresponds to the continuous administering of training is how effective (relative to each other) are training programs with different durations? This assessment of the dynamics of treatment duration essentially amounts to estimating a dose-response function. In this paper we therefore estimate the responses – i.e. the employment probability – that correspond to specific values of continuous doses – i.e. training of a particular length. In a setting in which doses are not administered under experimental conditions, estimation of a dose-response function is possible using the generalized propensity score (GPS). The GPS for continuous treatments is a straightforward extension of the well-established and widely used propensity score methodology for binary treatments (Rosenbaum and Rubin 1983) and multi-valued treatments (Imbens 2000, Lechner 2001). The GPS methodology is developed in Hirano and Imbens (2004) and Imai and van Dyk (2004). 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. To our knowledge, our paper along with parallel work by Flores-Lagunes et al. (2007) constitutes the first application 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.

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

2. Bias removal using the Generalized Propensity Score

Research in program evaluation in recent years has made comprehensive use of matching methods[3]. In the absence of experimental data, which is largely the case, the popularity of matching is due to its intuitively appealing technique of mimicking an experiment *ex post*.

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

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.

2.1 The GPS methodology

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

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

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

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

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

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

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

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

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

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

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

2.2 Implementation

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

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

which we estimate by ordinary least squares regression (OLS).[4] The estimated GPS is calculated as

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

In the second stage we calculate the conditional expectation function of Y_i given T_i and R_i as a flexible function of its two arguments. Our empirical approach uses the following approximation.

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

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

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

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. In addition to the specification in equation (8) we also implement several other specifications in order to allow for sufficiently flexible functional forms.

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

2.3 Testing for balancing of covariates 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, 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.

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 follows[5]: 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

[5] We thank Peter Mueser for suggesting this approach.

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. 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 Active Labor Market Policy (ALMP). Furthermore, the IEB comprise a large variety of covariates like age, education, disability, nationality and regional indicators.

Training participants in the programs we consider learn specific skills required for a certain vocation (e.g. computer-aided design for a technician/tracer) or receive qualifications that are of general vocational use (e.g. MS Office, computer skills). Numerically, these program types constitute the most important ones among all publicly financed training programs: In 2000, roughly 70% of all participants in training programs were assigned to this type (Schneider and Uhlendorff 2006, IZA et al. 2007).

We focus on men only. Our sample of participants consists of about 265 unemployed persons per quarter entering the program during the years 2000, 2001, and 2002, i.e. we observe approximately 3180 program participants. The data allow us to draw conclusions on the average participant starting a program during this time period. The programs comprise both occupation-specific training programs ("berufsbezogene Weiterbildung") and general training programs ("berufsübergreifende Weiterbildung"). The core feature of these training programs is the fact that treatment provision is a continuous variable, since the 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).

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 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. We observe that the same two peaks exist in both distributions, at durations of 180 days and 360 days, respectively.

[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 sub-samples, “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).

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 as well as regional indicators. Given the richness of the covariates along with the fact that we focus on participants only, rather than on a treatment vs. no-treatment comparison, the assumption of unconfoundedness seems entirely reasonable.

[Table 1 about here]

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 a program 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. Looking at the outcomes, two years after program entry as well as one year after the program ended around 35% of the participants are employed.

Figure 2 contains six panels plotting unadjusted outcomes – i.e. employment probability two years after program entry as well as 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), investigates the relationship between the employment probability at 2 years after entering into the program and 1 year after exit from the program, respectively, with the treatment duration.

[Table 2 about here]

From these tables, we have several observations. They 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.[6] 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.

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. 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 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.

[Table 3 about here]

4.2 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 section 3 above. We first estimate the conditional distribution of the length of the training program (treatment) by applying OLS. Table 4 contains the results.

[Table 4 about here]

[6] See low adjusted R-squared in Panel A of Tables 2a and 2b.

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.

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 Tables 5 and 6 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.

[Tables 5, 6 about here]

In the second step, we calculate the corresponding t-statistics for the GPS-adjusted sample. To do this, we evaluate the GPS for each individual at the median of the three groups, i.e. at the lengths of 84 days, 180 days, and 332 days for the actual duration, and at the lengths of 117 days, 184 days, and 348 days for the planned duration. 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]. 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 Tables 5, 6 correspond to the mean difference for each group. To calculate these mean t-statistics, the corresponding differences and standard errors of the five blocks are weighted by the number of observations.

In contrast to the unadjusted sample, we observe no t-statistics larger than 1.96 for the planned duration (Table 6) 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 three groups as we have done above when testing for covariate balance. Then we evaluate the GPS of the whole sample at the median treatment duration of group 1, i.e. 84 days. After that we plot the distribution of the evaluated GPS of group 1 and the same distribution of the rest of the sample in the same figure, which is the first panel of Figure 3. We repeat the same procedure for group 2 and group 3, and these give us the second and the third panels of Figure 3. These 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 last three panels of Figure 3 show results for the planned duration. These are very similar to the ones observed for the actual duration, i.e. common support is given.

[Figure 3 about here]

4.3 Results from estimating the dose-response function

The final step of our empirical analysis consists in estimating the GPS-adjusted dose-response function. Table 7 contains the estimation results for the dose-response function. Our main results for both outcome variables are presented in Figures 4 and 5, 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. The figures also include the non-participant employment probability baseline[7], which indicates that training effects are generally positive. Standard errors are bootstrap standard errors from 2,000 replications.

[Table 7 about here]

[Figures 4, 5 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

[7] This is a covariate-adjusted baseline derived from standard binary matching methods.

of the outcome with respect to different 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). The subsample for which actual durations equals planned ones confirms this profile: while it is generally flatter for the long durations, it emphasizes the increase in the treatment effect for durations of up to 100 days.

4.4 Robustness

In this section, we carry out several sensitivity checks for our main estimation. 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 4 and 5 also plot 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. 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, and thus the treatment effects, are consistently larger for the subsample. In particular, the estimated average response is up to 3 percentage points higher (cf. Figure 5b).

Our main estimation is based on a cubic specification for the dose-response function. Figures 6 and 7 plot results for the dose-response function for the full sample for quadratic and 4^{th} degree polynomial specifications as well. Like Figures 4 and 5, Figures 6 and 7 consider the two outcome variables and are structured in three parts 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.

It is also interesting to compare the results from the GPS with the ones obtained directly from LPM and probit models. In Figure 6, both sets of results are quite similar, but in Figure 7, results from the GPS are rather different from the ones estimated using an LPM and a probit model. This suggests that in the case of Figure 7, 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 6, 7 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, about 31% of participants' actual treatment durations differ from their planned duration. It is possible that the actual treatment duration could be endogenously determined by the participants. 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.

We follow 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.

[Table 8 about here]

Table 8 presents the results from this instrumental variable approach with or without controlling for additional variables. The different models 1 to 5 correspond to different cutoff points, 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.[8] 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.

5. Conclusions

In this paper, we study the effect of continuous training programs on the post-treatment employment probability, using a particular data set that contains information 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. We are able to consider both the planned and actual durations as treatment variables, thus avoiding a potential bias in estimating the DRF 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 DRF 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. After 100 days, however, further participation in the program does not seem to lead to an increase in the treatment effect. Whether the effect actually 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.

[8] 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.

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

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., 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.

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).

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

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

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

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

Schneider, H. and A. Uhlendorff (2006), "Die Wirkungen der Hartz-Reform im Bereich der beruflichen Weiterbildung", *Journal for Labour Market Research* 39, 477-490.

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
Disability								
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	-
Citizenship								
Foreigner EU	0.02	-	0.02	-	0.01	-	0.02	-
Foreigner Non-EU	0.10	-	0.11	-	0.14	-	0.10	-
Educational Attainment								
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	-
Vocational Attainment								
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	-
Employment history								
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., 1st year before program	0.19	-	0.19	-	0.21	-	0.18	-
Share of days in emp., 2nd year before program	0.38	-	0.36	-	0.40	-	0.38	-
Share of days in emp., 3rd year before program	0.43	-	0.41	-	0.41	-	0.43	-
Share of days in emp., 4th year before program	0.45	-	0.42	-	0.44	-	0.46	-
Share of days in unemp., 1st year before program	0.67	-	0.68	-	0.64	-	0.67	-
Share of days in unemp., 2nd year before program	0.39	-	0.43	-	0.36	-	0.39	-
Share of days in unemp., 3rd year before program	0.34	-	0.37	-	0.33	-	0.33	-
Share of days in unemp., 4th year before program	0.30	-	0.33	-	0.27	-	0.29	-
Outcome variables								
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

Variables	(1) Coefficient	(2) Std. Error	(3) Coefficient	(4) Std. Error	(5) Coefficient	(6) Std. Error	(7) Coefficient	(8) 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	
Probablity>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	
Probablity>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	
Probablity>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	
Probablity>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)
	Actual Duration		Planned Duration	
	Coeff.	Std. Error	Coeff.	Std. Error
Previous unemployment duration in months	0.7636	0.3451	0.2429	0.3170
Age	-2.1639	7.4060	-8.7657	6.8028
Age squared	-0.0025	0.1912	0.1927	0.1757
Age cubic	0.0006	0.0017	-0.0014	0.0016
Duration of last employment	0.0023	0.0021	0.0006	0.0019
No information about last employment	-11.0729	18.4469	4.3886	16.9444
No wage of last employment observed	34.2926	23.4434	1.8158	21.5339
Log(wage) of last employment	4.3508	3.4744	1.4269	3.1914
Educational Attainment 2	-189.4403	73.3736	-111.4210	67.3971
Educational Attainment 3	-209.7043	82.7537	-164.1688	76.0132
Educational Attainment 4	-371.8883	143.9889	-401.1327	132.2606
Educational Attainment 5	-472.2201	148.2782	-334.0777	136.2005
Vocational Attainment 2	63.5927	51.9122	59.1829	47.6838
Vocational Attainment 3	59.4316	91.8876	65.4984	84.4031
Vocational Attainment 4	583.0145	186.1424	428.7167	170.9806
Foreigner EU	-6.0217	13.0374	-4.2181	11.9755
Foreigner Non-EU	6.3966	5.9281	1.4674	5.4452
Share of days in emp., 1st year before program	-5.2568	2.4743	-3.7461	2.2727
Share of days in emp., 2nd year before program	-1.3409	1.9019	-2.9754	1.7470
Share of days in emp., 3rd year before program	-1.0017	1.9274	0.5746	1.7704
Share of days in emp., 4th year before program	-2.0666	1.6206	-3.1272	1.4886
Share of days in unemp., 1st year before	-1.5590	2.5644	1.0997	2.3555
Share of days in unemp., 2nd year before	-5.0441	1.8696	-4.1941	1.7173
Share of days in unemp., 3rd year before	-0.8236	1.8962	-0.7127	1.7418
Share of days in unemp., 4th year before	-5.4048	1.7663	-4.3744	1.6225
Disability low degree	44.5817	21.5007	28.8693	19.7494
Disability medium degree	20.1793	20.5201	4.6797	18.8487
Disability high degree	-27.3114	6.2610	-42.4251	5.7510

Table 4. Estimated GPS (Cont.)

	(1)	(2)	(3)	(4)
	Actual Duration		Planned Duration	
	Coeff.	Std. Error	Coeff.	Std. Error
Number of children	-7.1170	3.2875	-5.5518	3.0197
Youngest Child < 4 years	5.6853	9.3889	8.0336	8.6242
Youngest Child < 14 years	1.1432	7.0381	4.1765	6.4648
Regional unemployment rate	433.1507	56.5878	387.5052	51.9786
Regional type 2	0.9976	5.7695	-2.7450	5.2996
Regional type 3	-15.7442	7.0030	-20.5547	6.4326
Regional type 4	15.2312	10.5725	21.1639	9.7113
Regional type 5	-10.3845	8.7881	-2.5635	8.0723
Educational Attainment 2 * age	9.9967	4.2355	5.2211	3.8905
Educational Attainment 3 * age	10.6215	4.7983	7.8531	4.4074
Educational Attainment 4 * age	21.9266	7.9930	23.6488	7.3419
Educational Attainment 5 * age	28.8547	8.0434	19.5861	7.3883
Educational Attainment 2 * age squared	-0.1259	0.0582	-0.0600	0.0534
Educational Attainment 3 * age squared	-0.1230	0.0659	-0.0846	0.0605
Educational Attainment 4 * age squared	-0.2846	0.1058	-0.3097	0.0972
Educational Attainment 5 * age squared	-0.3853	0.1055	-0.2496	0.0969
Vocational Attainment 2 * age	-2.7761	2.9808	-3.1187	2.7380
Vocational education 3 * age	-2.3590	5.2393	-3.0691	4.8125
Vocational education 4 * age	-28.0278	9.5537	-19.4233	8.7755
Vocational Attainment 2 * age squared	0.0320	0.0404	0.0402	0.0371
Vocational education 3 * age squared	0.0381	0.0700	0.0459	0.0643
Vocational education 4 * age squared	0.3515	0.1191	0.2392	0.1094
Constant	220.4666	95.3657	353.6884	87.5979
Adjusted R Squared	0.1966		0.1999	
Number of Observations	3130		3130	

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

	(1)	(2)	(3)	(4)	(5)	(6)
		Unadjusted			**Adjusted**	
Covariate	[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
Disability						
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
Citizenship						
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
Educational Attainment						
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
Vocational Attainment						
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

Bold numbers indicate significance at the 5% level

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

	(1)	(2)	(3)	(4)	(5)	(6)
	Unadjusted			Adjusted		
Covariate	[11, 137]	[138,247]	[248,395]	[11, 137]	[138,247]	[248,395]
Employment History						
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., 1st year before program	**-2.79**	1.56	1.12	-1.23	0.55	-0.52
Share of days in emp., 2nd year before program	**-2.08**	**2.56**	-0.66	-1.21	1.01	-0.41
Share of days in emp., 3rd year before program	-0.48	0.82	-0.40	-0.57	0.34	-0.24
Share of days in emp., 4th year before program	1.38	-0.30	-1.06	0.02	-0.31	-0.23
Share of days in unemp., 1st year before program	1.58	**-2.62**	1.22	1.09	-1.04	0.66
Share of days in unemp., 2nd year before program	-0.14	**-3.97**	**4.40**	0.78	-1.47	0.76
Share of days in unemp., 3rd year before program	-0.31	**-3.75**	**4.33**	0.80	-1.38	0.74
Share of days in unemp., 4th year before program	**-2.75**	**-1.96**	**4.87**	-0.07	-0.45	0.65
Regional indicators						
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

	(1)	(2)	(3)	(4)	(5)	(6)
		Unadjusted			**Adjusted**	
Covariate	[11, 167]	[168,271]	[272,395]	[11, 167]	[168,271]	[272,395]
Age	**2.99**	0.40	**-3.85**	0.70	-0.09	-0.12
Disability						
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
Citizenship						
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
Educational Attainment						
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
Vocational Attainment						
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

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 (Cont.)

	(1)	(2)	(3)	(4)	(5)	(6)
		Unadjusted			**Adjusted**	
Covariate	[11, 167]	[168,271]	[272,395]	[11, 167]	[168,271]	[272,395]
Employment History						
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
Regional indicators						
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. Estimated Dose Response Functions

	(1)	(2)	(3)	(4)
	Actual Duration		Planned Duration	
	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
GPS^2	60057.2100	46972.1400	60457.1900	39749.1900
GPS^3	-4442237.0000	5712302.0000	-4925847.0000	-4925847.0000
Program Duration	0.0015	0.0015	0.0038	1.9801E-03
Program $Duration^2$	-5.56E-06	7.99E-06	-4.22E-06	9.97E-06
Program $Duration^3$	2.59E-09	1.30E-08	-1.30E-08	1.67E-08
GPS*Program Duration	0.2792	0.5618	-1.4210	0.5881
GPS^2*Program Duration	-109.3040	63.0600	-23.6346	56.5314
GPS*Program $Duration^2$	0.0006	0.0012	0.0035	0.0013
Constant	0.4127	0.1063	0.2268	0.1010
Adjusted R Squared	-0.0002		0.0013	
Number of Observations	3130		3130	
Panel B: Outcome Variable: Employment status at time 1 year after exit from the program				
GPS	-20.4268	119.5165	-102.1634	107.3369
GPS^2	9059.1920	46871.0700	50623.5700	39637.6800
GPS^3	-732274.6000	5700010.0000	-2130617.0000	4313976.0000
Program Duration	-0.0001	0.0017	0.0020	0.0020
Program $Duration^2$	3.49E-06	7.98E-06	4.06E-06	9.94E-06
Program $Duration^3$	-9.01E-09	1.30E-08	-2.34E-08	1.66E-08
GPS*Program Duration	-0.0346	0.5606	-1.2096	0.5864
GPS^2*Program Duration	-25.7458	62.9243	-51.5993	56.3728
GPS*Program $Duration^2$	0.0003	0.0012	0.0032	0.0013
Constant	0.3421	0.1060	0.3114	0.1007
Adjusted R Squared	-0.0020		0.0009	
Number of Observations	3130		3130	

Table 8. 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

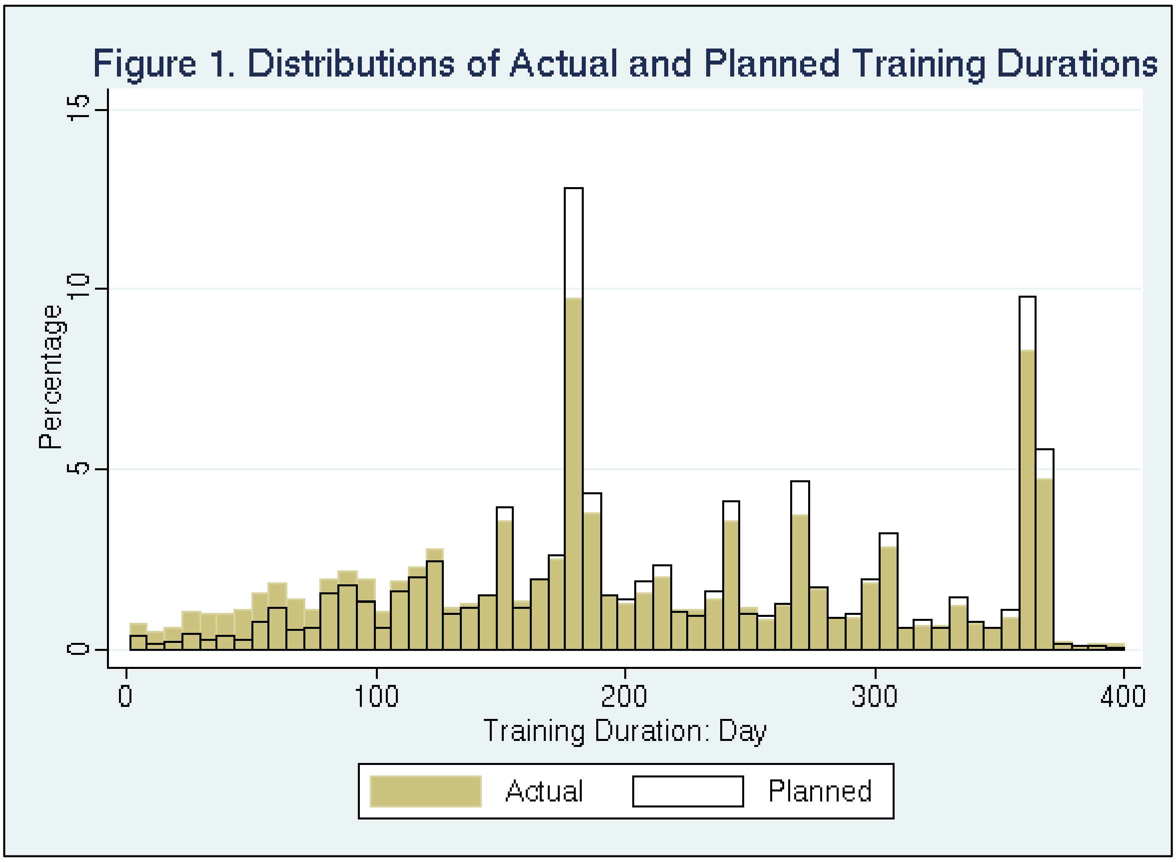


Figure 1. Distributions of Actual and Planned Training Durations
Percentage
0
5
10
15
0
100
200
300
400
Training Duration: Day
Actual
Planned

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

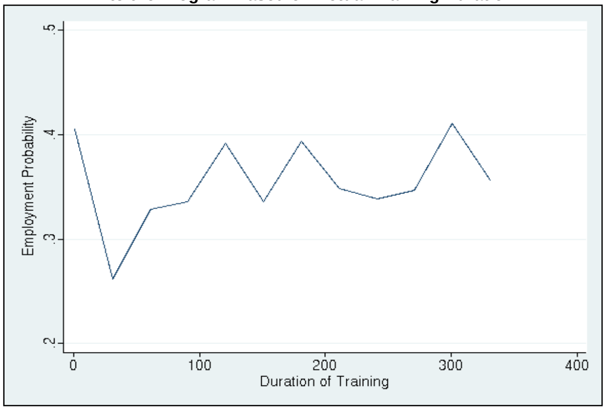


Figure 2b. Unadjusted Employment Probability at Time 2 Years after Entry into the Program Based on Planned Training Duration

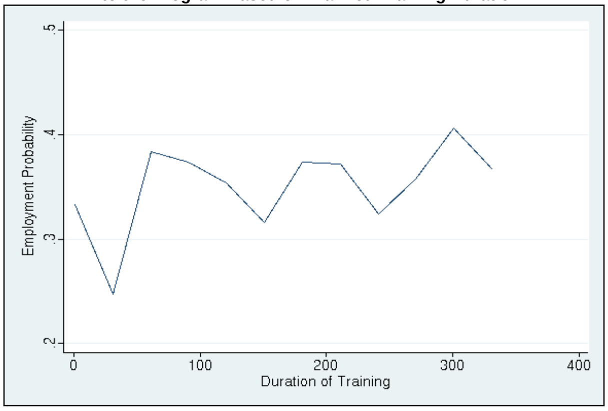


Figure 2c. Unadjusted Employment Probability at Time 2 Years after Entry into the Program Based on Subsample with Actual Training Duration Equal to Planned Duration

Figure 2d. Unadjusted Employment Probability at Time 1 Years after Exit from the Program Based on Actual Training Duration

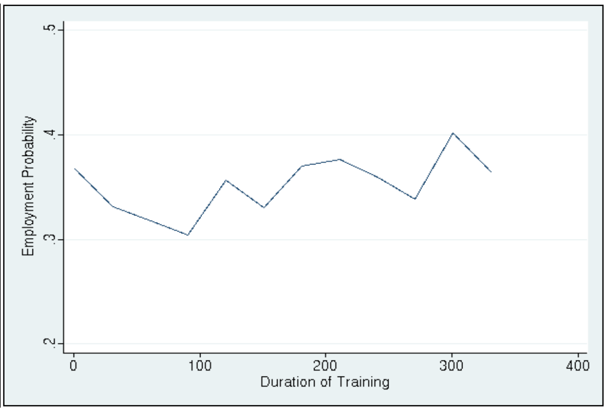


Figure 2e. Unadjusted Employment Probability at Time 1 Years after Exit from the Program Based on Planned Training Duration

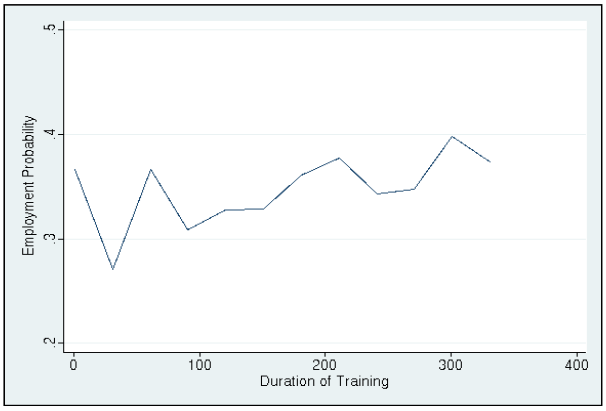


Figure 2f. Unadjusted Employment Probability at Time 1 Years after Exit from the Program Based on Subsample with Actual Training Duration Equal to Planned Duration

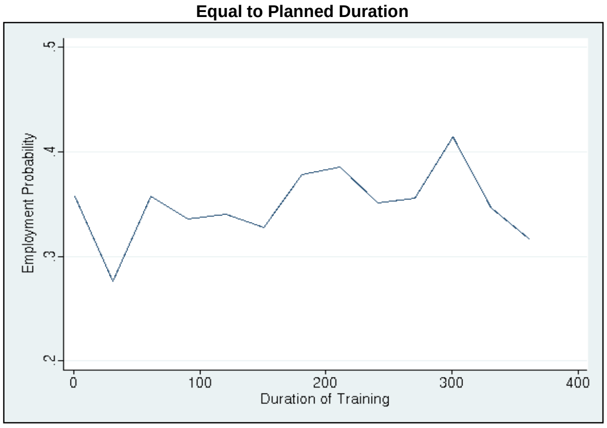


Figure 3. Common Support Condition

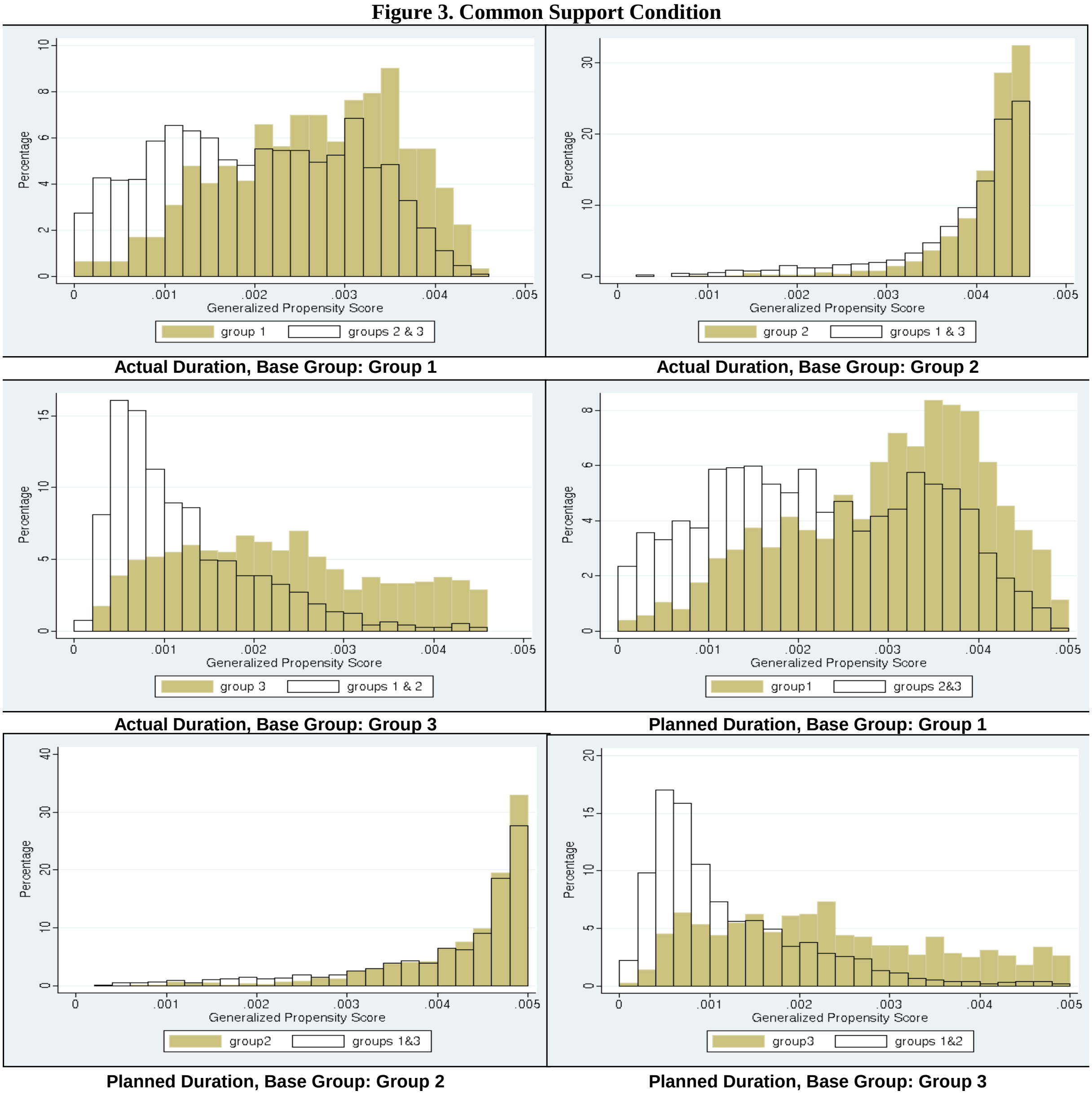
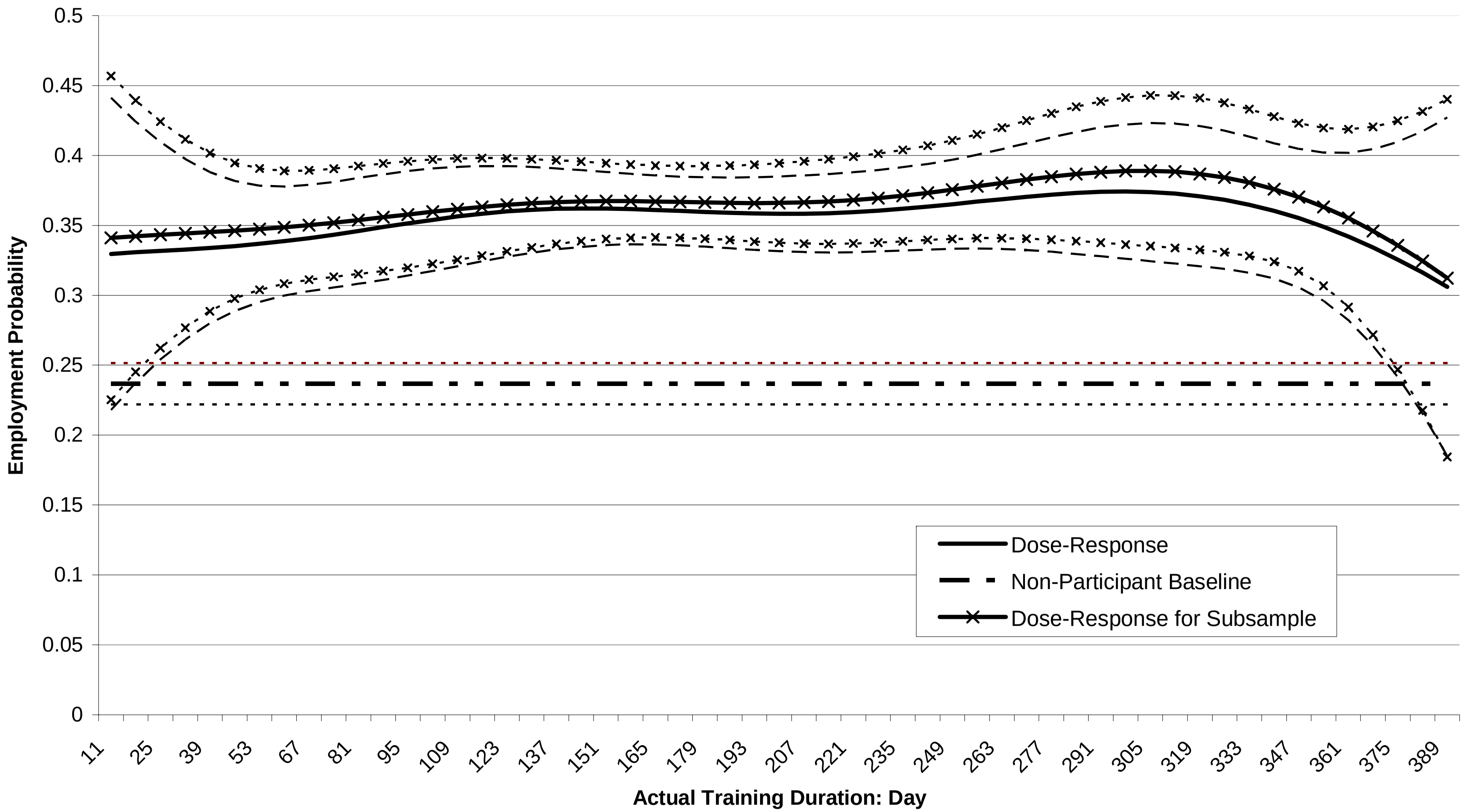
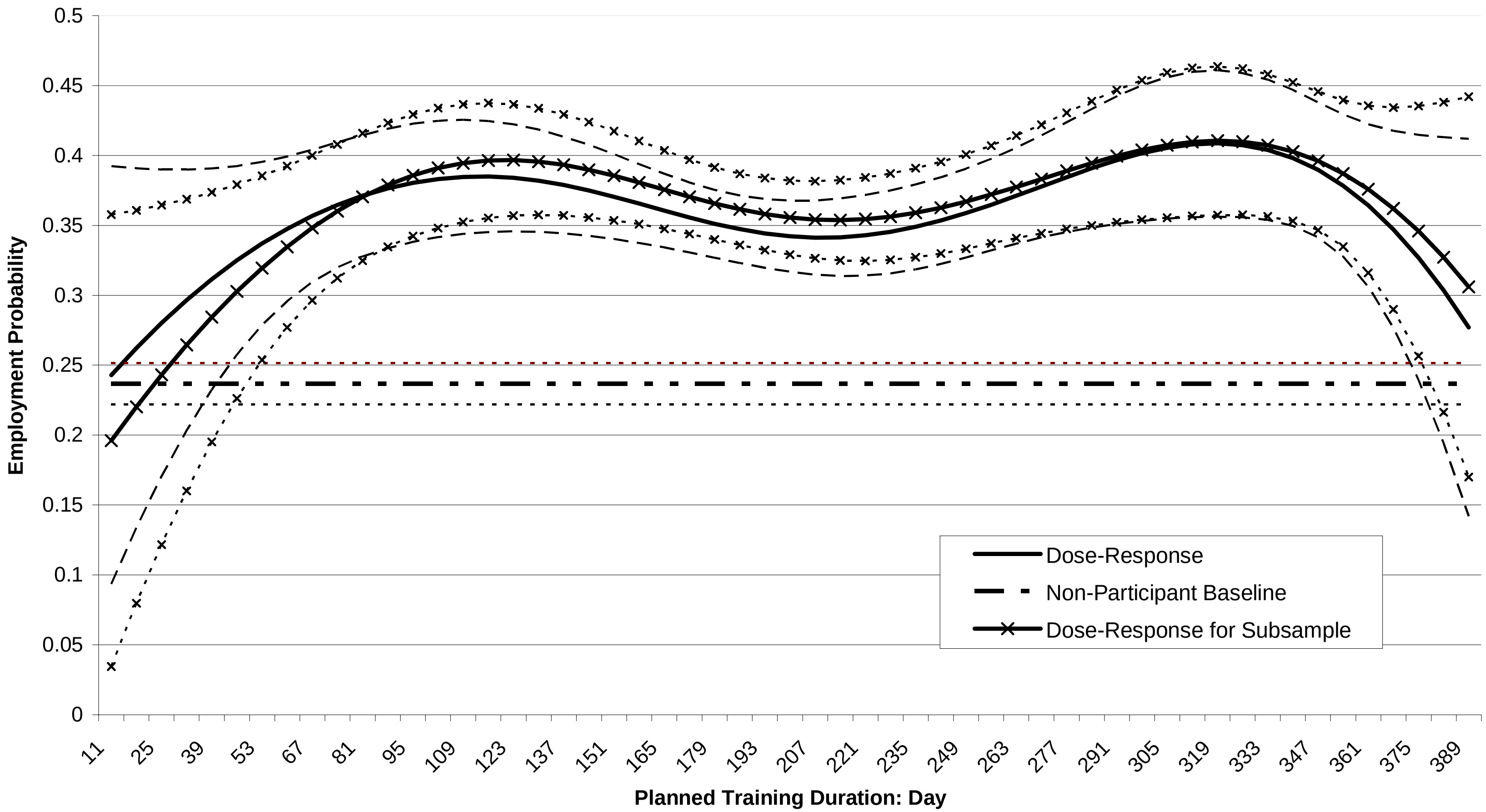


Figure 4a. Employment Probability at Time 2 Years after Entry into the Program Based on Actual Training Duration



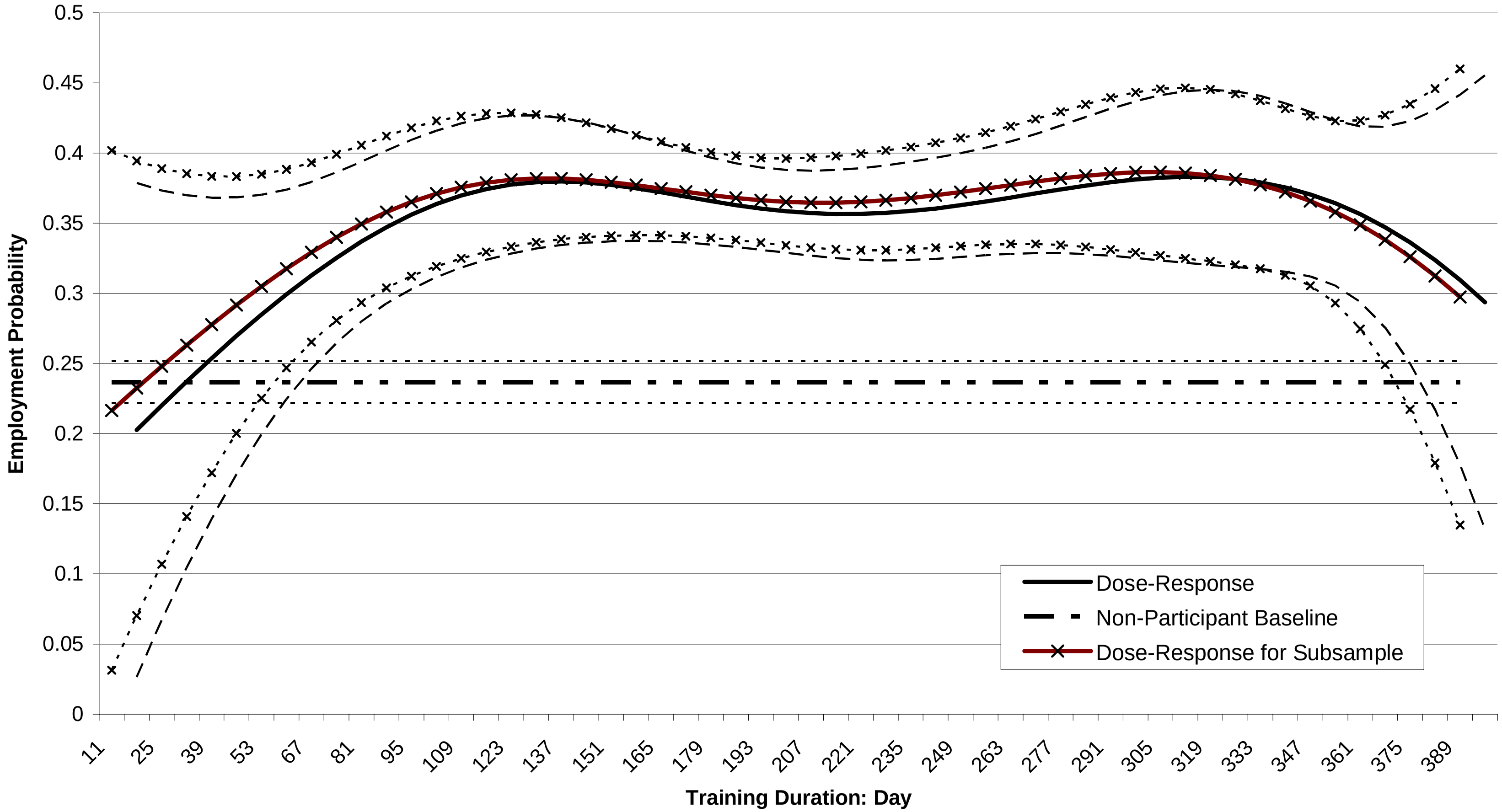
Dashed lines are bounds for 95% confidence intervals. These intervals are based on bootstrap standard errors from 2,000 replications

Figure 4b. Employment Probability at Time 2 Years after Entry into the Program Based on Planned Training Duration



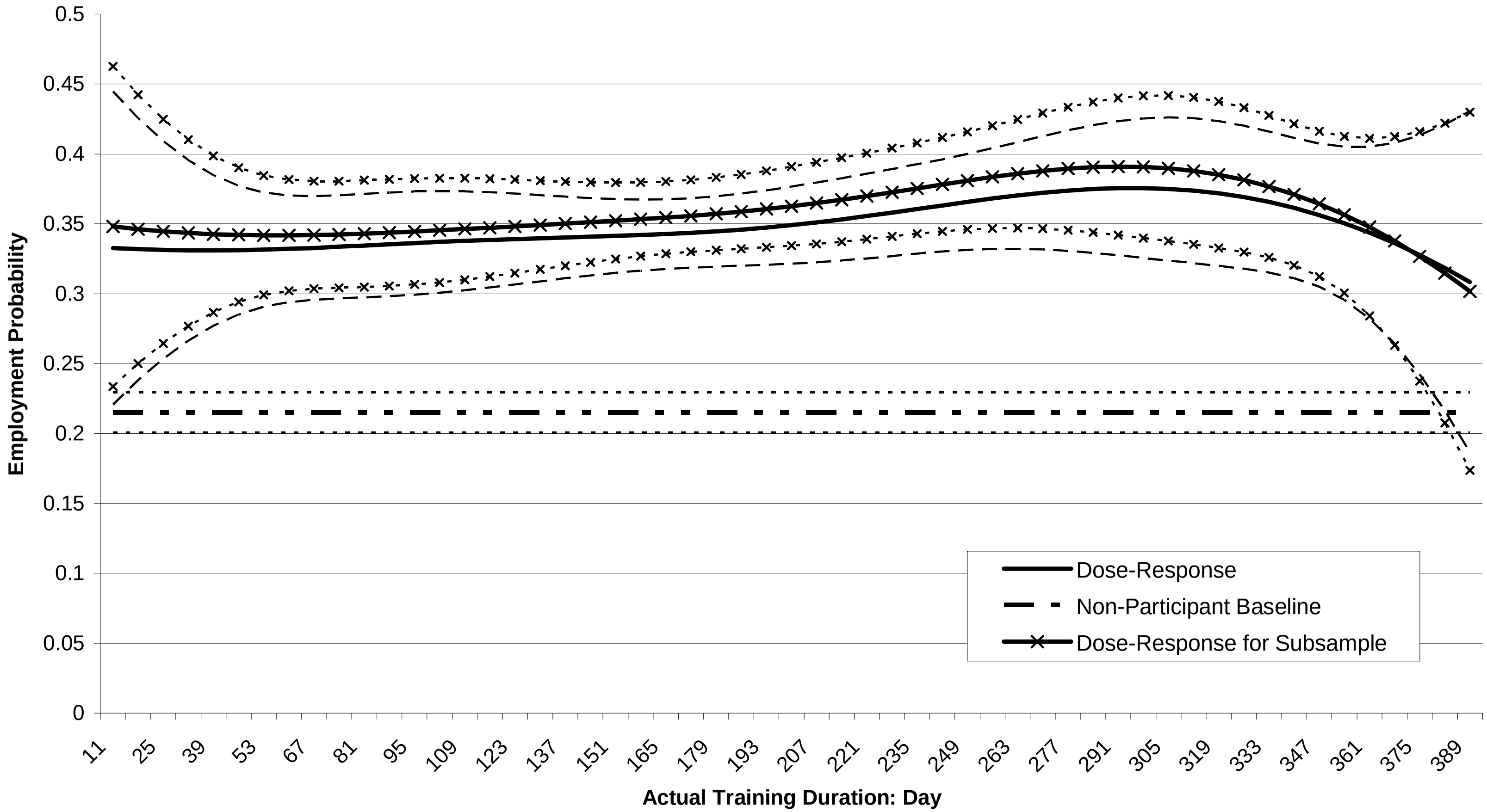
Dashed lines are bounds for 95% confidence intervals. These intervals are based on bootstrap standard errors from 2,000 replications

Figure 4c. Employment Probability at Time 2 Years after Entry into the Program Based on Subsample with Actual Training Duration Equal to Planned Duration



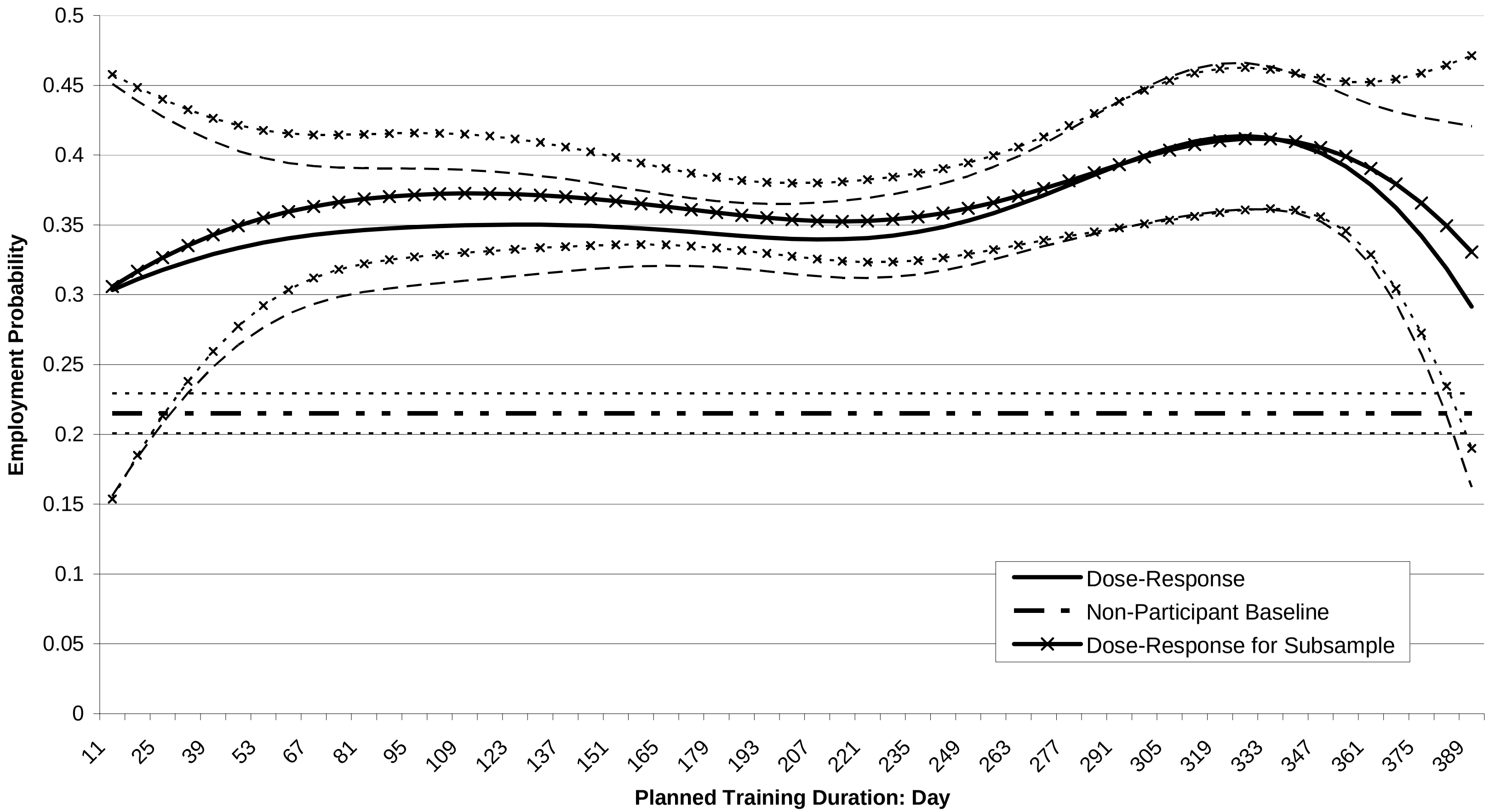
Dashed lines are bounds for 95% confidence intervals. These intervals are based on bootstrap standard errors from 2,000 replications

Figure 5a. Employment Probability at Time 1 Year after Exit from the Program Based on Actual Training Duration



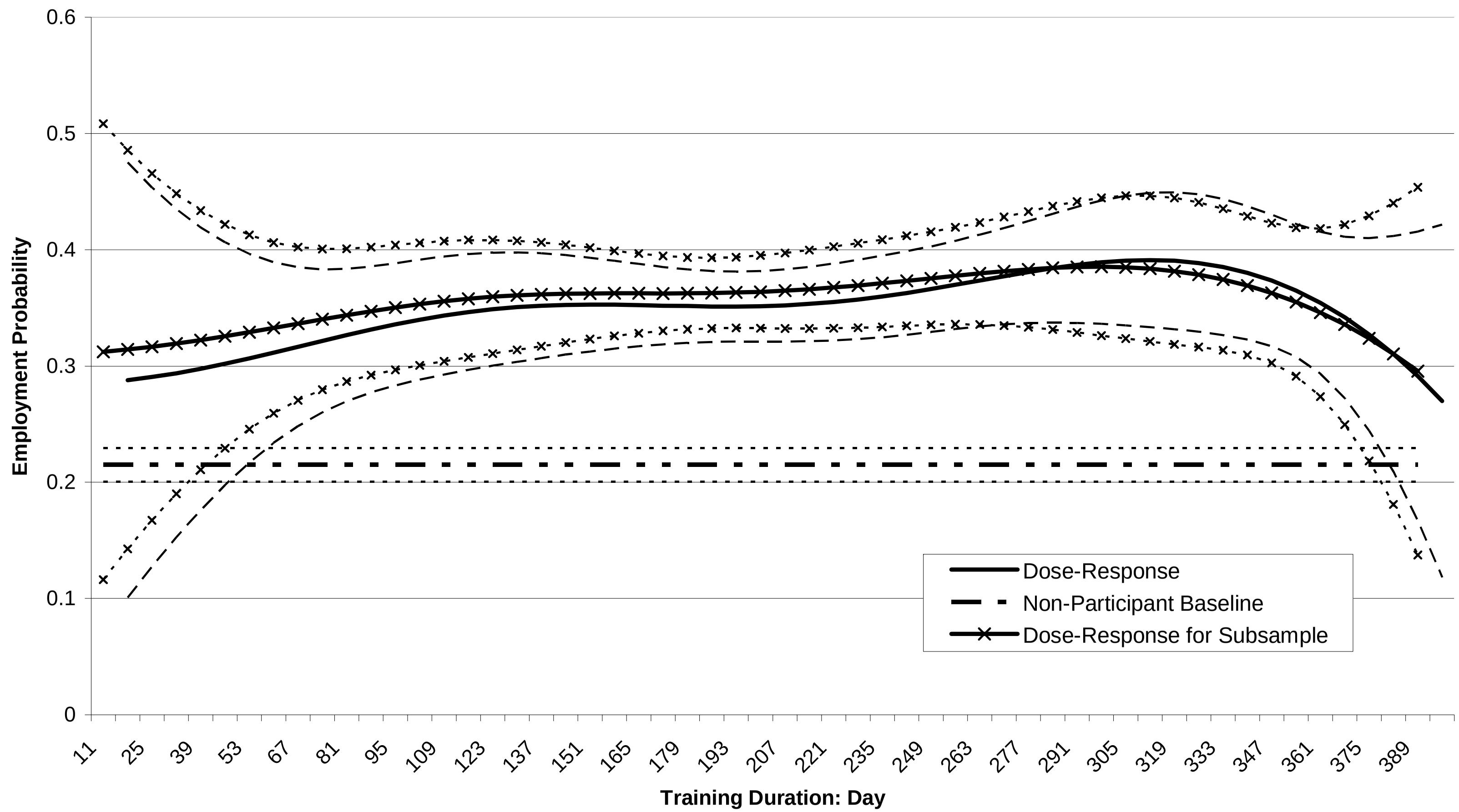
Dashed lines are bounds for 95% confidence intervals. These intervals are based on bootstrap standard errors from 2,000 replications

Figure 5b. Employment Probability at Time 1 Year after Exit from the Program Based on Planned Training Duration



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

Figure 5c. Employment Probability at Time 1 Year after Exit from the Program Based on Subsample with Actual Training Duration Equal to Planned Duration



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

Figure 6a. Estimated Employment Probability at Time 2 Years after Entry into the Program Based on Actual Duration

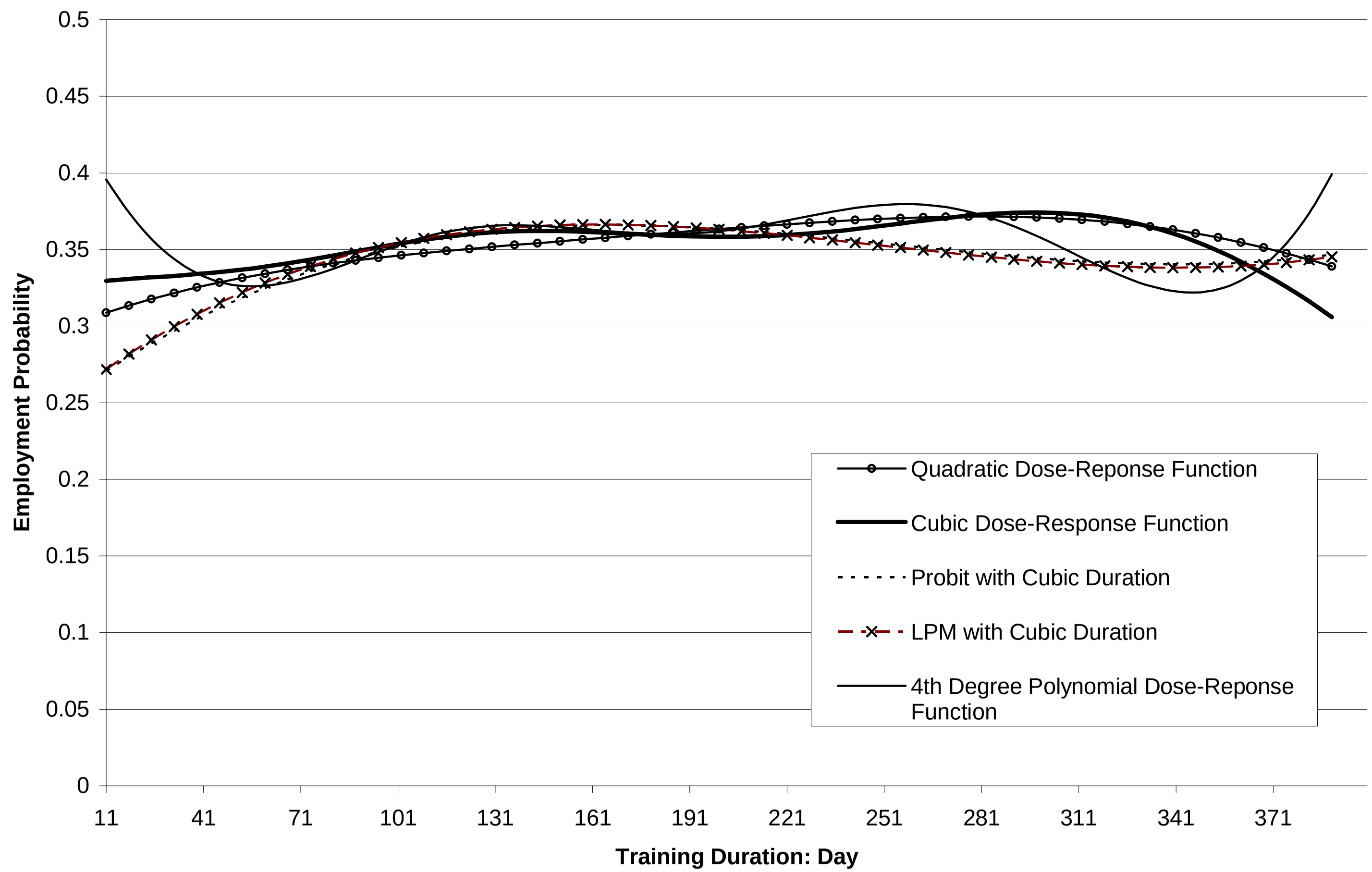


Figure 6b. Estimated Employment Probability at Time 2 Years after Entry into the Program Based on Planned Duration

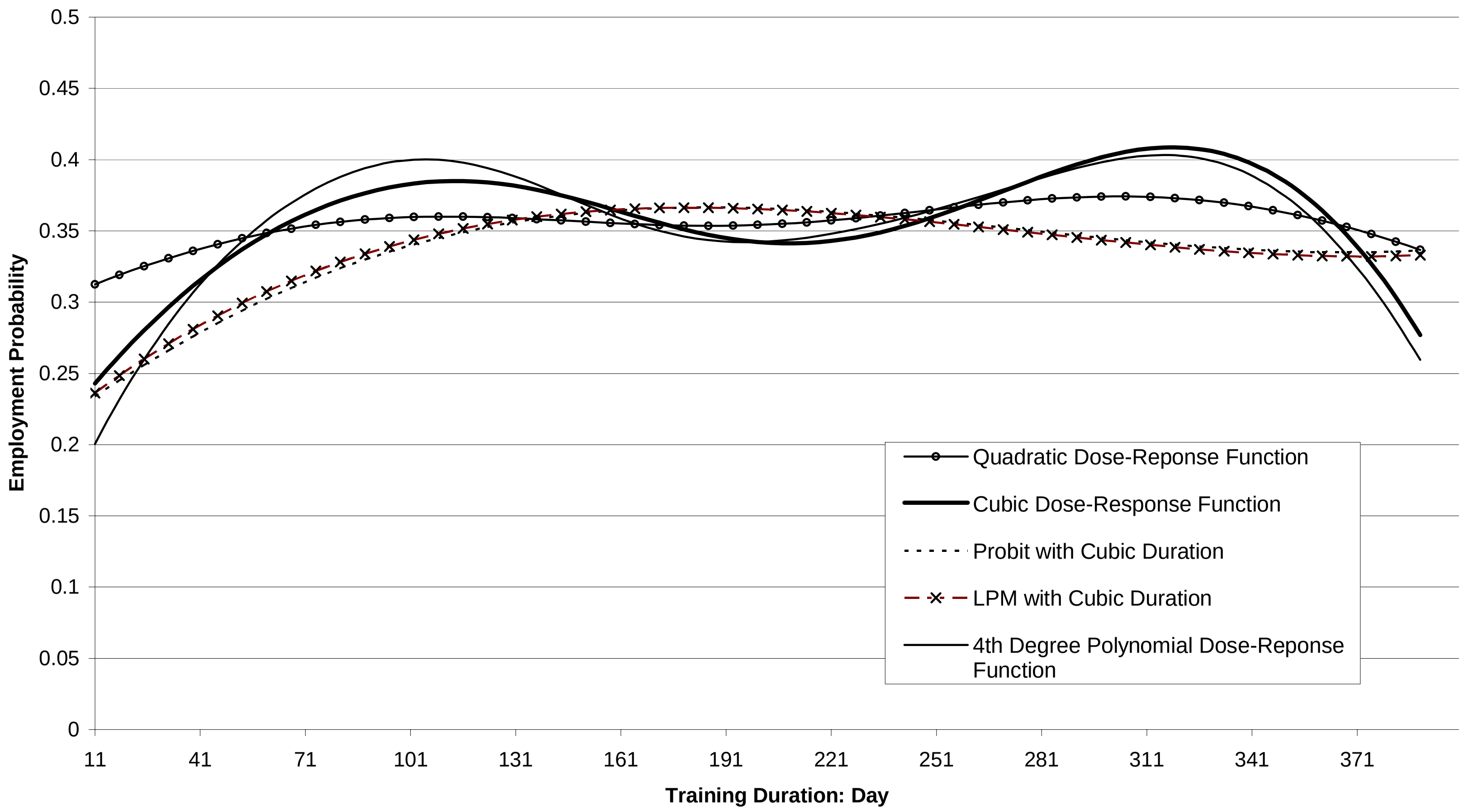


Figure 6c. Estimated Employment Probability at Time 2 Years after Entry into the Program Based on Subsample with Actual Training Duration Equal to Planned Duration

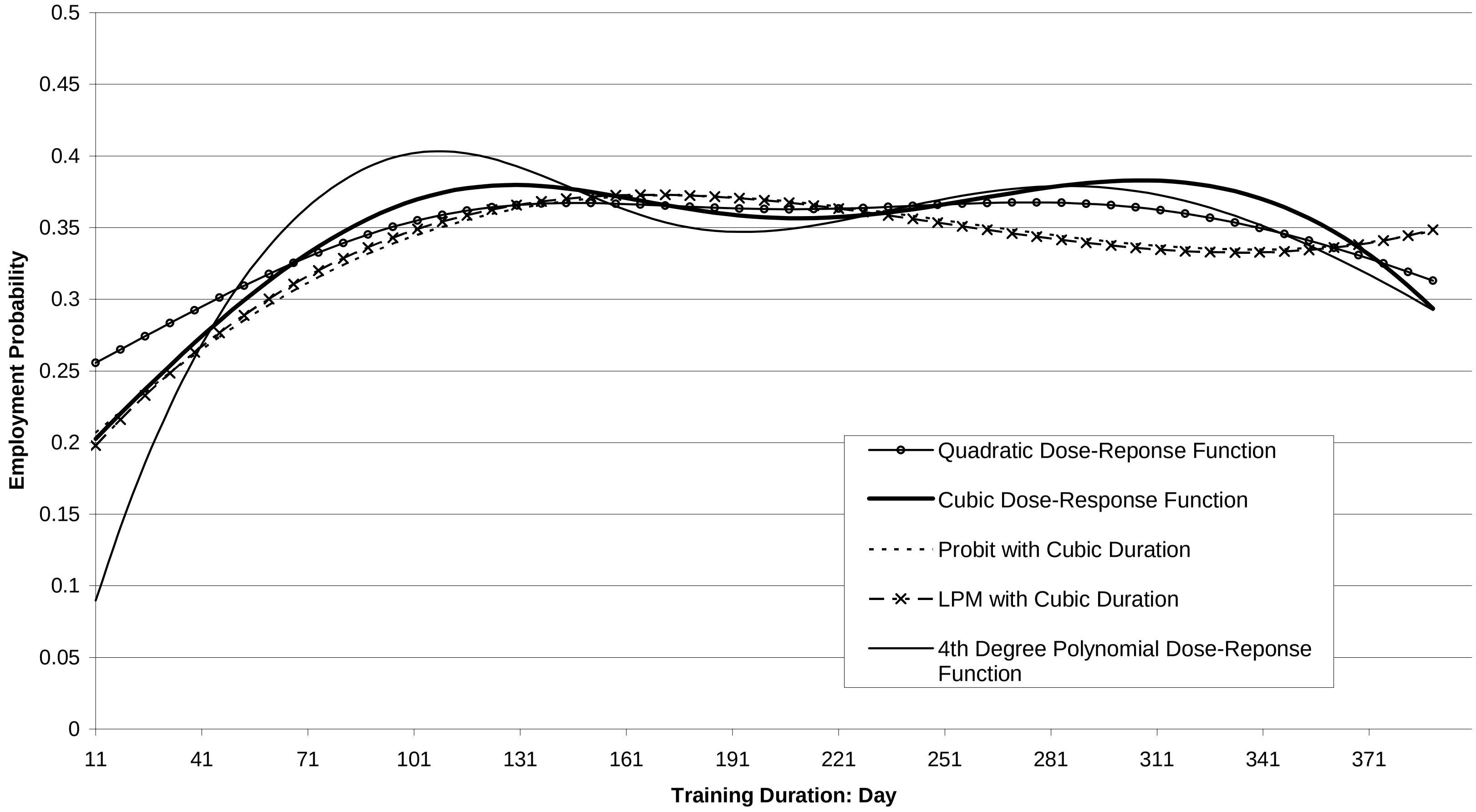


Figure 7a. Estimated Employment Probability at Time 1 Year after Exit from the Program Based on Actual Duration

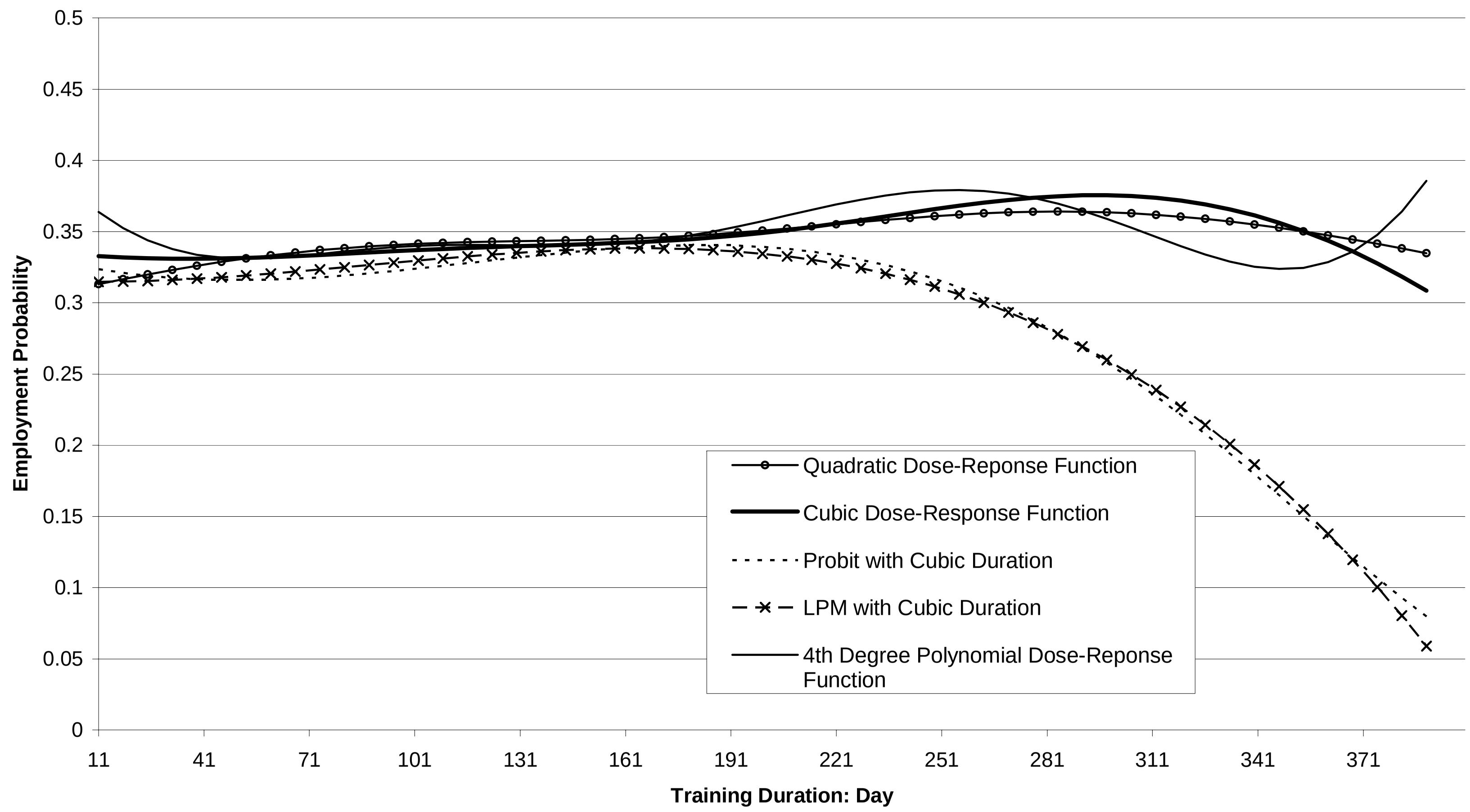


Figure 7b. Estimated Employment Probability at Time 1 Year after Exit from the Program Based on Planned Duration

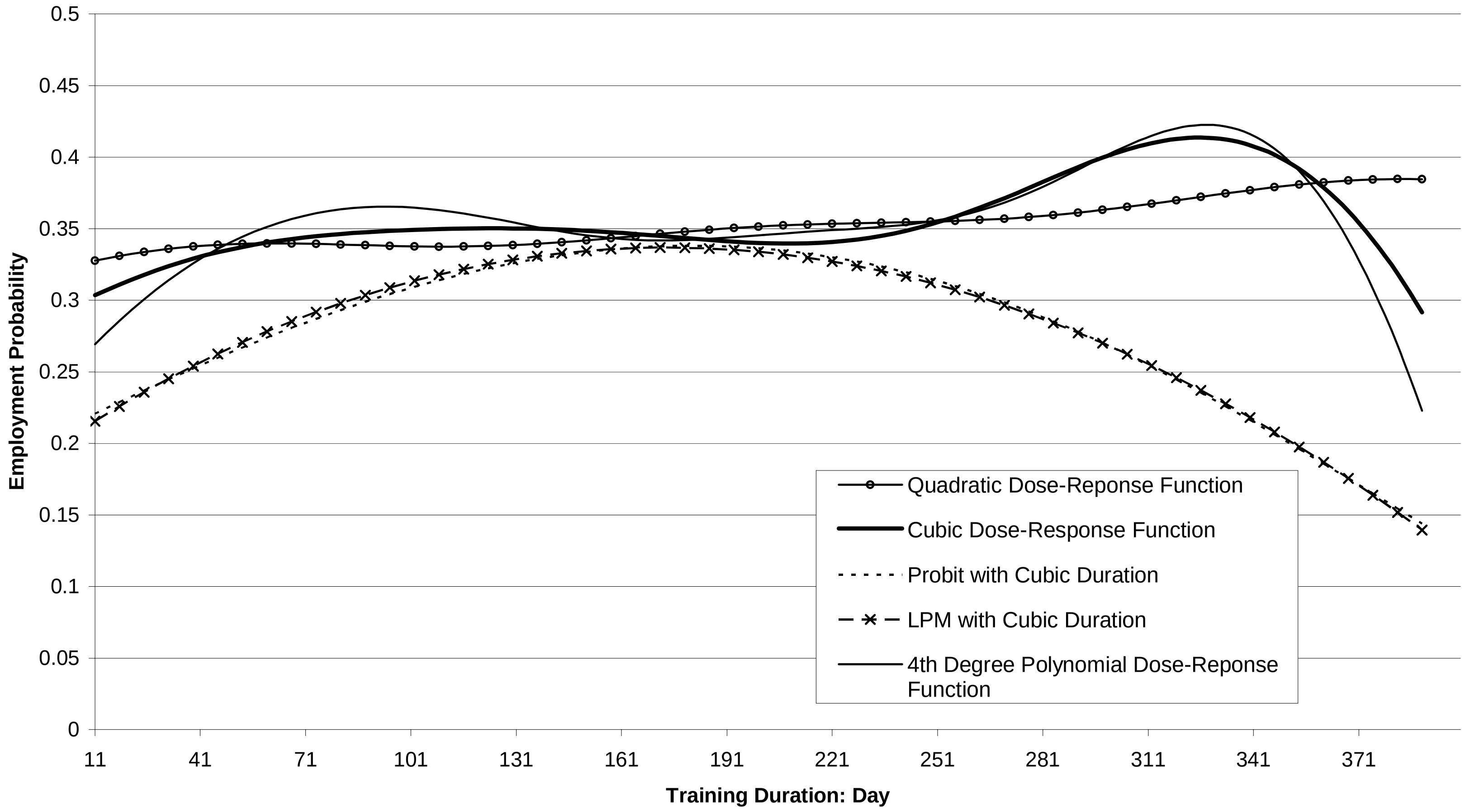


Figure 7c. Estimated Employment Probability at Time 1 Year after Exit from the Program Based on Subsample with Actual Training Duration Equal to Planned Duration