

Kluve, Jochen; Schneider, Hilmar; Uhlendorff, Arne; Zhao, Zhong

Article — Manuscript Version (Preprint)

Evaluating continuous training programmes by using the generalized propensity score

Journal of the Royal Statistical Society: Series A (Statistics in Society)

Provided in Cooperation with:

RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

Suggested Citation: Kluve, Jochen; Schneider, Hilmar; Uhlendorff, Arne; Zhao, Zhong (2012) : Evaluating continuous training programmes by using the generalized propensity score, Journal of the Royal Statistical Society: Series A (Statistics in Society), ISSN 1467-985X, Wiley Blackwell, Hoboken, Vol. 175, Iss. 2, pp. 587-617, <https://doi.org/10.1111/j.1467-985X.2011.01000.x> , <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-985X.2011.01000.x/full>

This Version is available at:

<https://hdl.handle.net/10419/144165>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Evaluating continuous training programs using the generalized propensity score*

Jochen Kluve
Humboldt University Berlin, RWI Essen and IZA Bonn

Hilmar Schneider
IZA Bonn

Arne Uhlenдорff
University of Mannheim, IZA Bonn and DIW Berlin

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

09 June 2011

Abstract. This paper assesses the heterogeneity 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 120 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.

* We would like to thank the editor Antony Fielding, an Associate Editor and two referees, 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, Xiamen University, the World Bank and the Washington Statistical Society for valuable discussions and helpful comments and Jan Stuhler for excellent research assistance. Arne Uhlenдорff also thanks DIW DC, where part of this research was pursued during his stay in fall 2007. The usual disclaimer applies.

[†] Correspondence: jochen.kluve@hu-berlin.de, schneider@iza.org, uhlenдорff@uni-mannheim.de, 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 2010). 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 2010).

In the previous training literature, the focus is typically on binary treatment effects and there is no differentiation of training programs by their duration. The provision of training, however, can be quite heterogeneous both in terms of training contents and in terms of training duration. In this paper we contribute to the literature on training programs by focusing on the heterogeneity of treatment effects that may 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 vocational degree is acquired as part of the program requirements – this is the majority of training programs in Germany. 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 heterogeneity of treatment effects along the duration dimension essentially amounts to estimating a dose-response function as proposed in Hirano and Imbens (2004). In this paper we therefore estimate the response – i.e. the employment probability – that corresponds 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 be 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, Flores-Lagunes, Gonzalez and Neuman (2010) constitute the first applications of the GPS in the context of evaluating active labor market programs.

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 gives details on the data and the treatment we study. Section 3 describes the methodology of estimating a dose-response function to evaluate a continuous policy measure, adjusting for the generalized propensity score. The fourth section contains the empirical implementation. It discusses the plausibility of the unconfoundedness assumption, it details the GPS estimation, the common support condition and the balancing of covariates, and it presents the results from estimating the dose-response function. Section 5 contains several robustness checks. Section 6 concludes.

2. Institutional setting and Data

2.1 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 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. 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.

Among the programs considered here one can distinguish between classroom oriented training (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. Participants in the programs we consider learn specific skills required for a certain vocation (e.g. computer-aided design for a technician/tracer, "berufsbezogene Weiterbildung") or receive qualifications that are of general vocational use (e.g. MS Office, computer skills, "berufsübergreifende Weiterbildung"). Numerically, these types constitute the most important ones among all publicly financed training programs: In 2000, roughly 70% of all participants in training were assigned to these two types (Schneider and Uhlendorff 2006; IZA, DIW and infas 2007).

Whereas the programs we focus on do not lead to a vocational degree, participants may get a certificate about the type and the content of the training.¹ 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. Programs leading to a vocational degree have a duration of around 2 years – c.f. for example Lechner et al. (2010) for an analysis of these long-term programs.

2.2 Data

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.² The data allow us to draw conclusions on the average participant starting a program during this time period. The core feature of the training programs we analyze is the fact that treatment provision is a continuous variable. 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).³

¹ There exists another program type also 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.

² 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.

³ One caveat is that we do not have detailed information on the training content. Although we restrict our analysis on two relatively homogeneous 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.

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.⁴ 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. 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 A1 in the appendix 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 A1 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, the previous employment states measured at 12 points in time (4, 8, ... , 48 weeks before entering the program) 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 Gorislawsky (2007), emphasize the importance of having treated and

⁴ But we do estimate the baseline counterfactual outcome of non-treated separately to facilitate interpreting our main results.

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 weak unconfoundedness assumption plausible (cf. also below). 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 A1 also shows that the covariate distributions are very similar across all (sub-) samples, which indicates that selection into the different sub-samples based on observable characteristics is not strong. In the remainder of this paper we will therefore focus on presenting results for the actual durations of the full sample, complementing these results with planned durations and the sample of actual = planned durations as appropriate. 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 enrollment in the program. This is also reflected in the decreasing share of individuals who are employed measured at 12 points in time during the year before the program starts.

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 two 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 actual training duration. The figures generally show an increasing trend: After an initial dip in employment probability during the first weeks in the program, employment rates seem to increase with the length of participation.

[Figure 2 about here]

3. Bias removal using the Generalized Propensity Score

Research in program evaluation in recent years has made comprehensive use of matching methods.⁵ 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, 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 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.

⁵ 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).

3.1 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. We have a random sample of training participants, indexed by $i=1, \dots, N$. For each unit i there exists a set of potential outcomes $Y_i(t)$ – the employment state – 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 participant i , we observe a vector of covariates X_i , the duration T_i of the training that individual 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 $r(t, x) = f_{T|X}(t \mid x)$ the conditional density of the treatment given the covariates, the *Generalized Propensity Score (GPS)* is defined as

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

The GPS has a balancing property similar to the balancing property of the propensity score for binary treatments. Within strata with the same value of $r(t, X)$ the probability that $T=t$ does not depend on the value of X , i.e. the GPS has the property that $X \perp 1\{T = t\} \mid r(t, X)$. Hirano and Imbens (2004) emphasize that this is a mechanical implication of the definition of the GPS and does not require weak unconfoundedness. In combination with weak 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 three steps. The first step consists of modeling and estimating the GPS. The second 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.

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

For the estimation of equation (3) one has to assume some functional form of the relationship between the employment state Y , the training duration T , and the GPS R . The third 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.

3.2 Implementation

In the practical implementation of the methodology outlined in the previous section, we use a normal distribution for the training duration 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, or to use nonparametric methods for partial means. 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.⁶ The estimated GPS is calculated as

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

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

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

In addition to the specification in equation (7) we also implement several other specifications in order to allow for sufficiently flexible functional forms. On the one hand we vary the degree of the polynomial specification to test whether our results are robust with respect to a higher degree of flexibility. On the other hand we apply flexible spline models to avoid a parametric specification. Spline models are a non-parametric regression technique. They are not only flexible, but they also have good properties in terms of the mean squared error fit. Another attractiveness is that they are relatively easy to implement, see Keele (2008). Moffitt (2008) applies a spline model in the context of estimating marginal treatment effects.

For each individual the observed training duration 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

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

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, i.e. we bootstrap the entire estimation process.

⁶ 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.

3.3 Assessing the weak unconfoundedness assumption, common support condition, and balancing of covariates

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 potential case of violating this assumption is the possibility of reverse causality. Individuals might leave the training program because they have received an acceptable job offer. In this case, finding a job leads to a shorter duration of the training. The planned duration is determined prior to the program, which is free from endogenous exit and plausibly exogenous once we condition on detailed observed characteristics including previous labor market history. We can use the information on the planned duration as an instrumental variable to test the potential endogeneity due to reverse causality 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 or due to the fact that some individuals leave the program because they expect no further benefits from continuing the program. This approach does not allow to test whether there exist unobservable characteristics (unobserved to researchers, but observable to the program applicants and caseworkers) like motivation or self-esteem which may have an impact on both, the planned and the observed duration of the training program, and which are relevant for subsequent labor market outcomes. However, the detailed information on previous labor market outcomes contained in our data make it very likely that unobservable characteristics such as motivation or self-esteem are captured by the individual labor market history.

Similar to standard propensity score matching methods, common support is also a concern in the GPS application. We propose to test the common support condition as follows⁷: First, 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. Finally, we restrict our final sample to individuals which are comparable across the

⁷ We thank Peter Mueser for suggesting this approach. At a later stage we found out that Flores et al. (2010) independently propose a similar approach in their paper.

three groups simultaneously, i.e. we drop those participants whose GPS is not among the common support region for all three groups.

Finally, in addition to assessing the identifying assumption and common support, in the case of a continuous treatment it is also crucial to evaluate how well adjustment for the GPS works in balancing the covariates, i.e. whether 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. We apply three approaches to test the balancing properties of the GPS.

In the first approach we follow Imai and van Dyk (2004) who propose to evaluate the balancing of the covariates by regressing each covariate on the treatment variable (i) without and (ii) with conditioning on the predicted training duration $E(T|X_i)$. Once we condition on $E(T|X_i)$ we expect that the training duration is not correlated with the covariate if adjustment for the GPS works in balancing the observable characteristics. We use a log transformation of the treatment variable since the training duration is mechanically uncorrelated with each continuous covariate given $E(T|X_i)$ (see App. B. in Imai and van Dyk 2004).

In the second approach we propose to regress each covariate on the treatment variable (i) without and (ii) with conditioning on the distribution of the GPS R_i . R_i is evaluated at different potential values of the treatment level, the 25th, the 50th and the 75th percentile of the training duration. The basic idea is similar to the first approach: Once we condition on the distribution of R_i the training duration should be uncorrelated with the covariate if the GPS properly balances the covariates. For both approaches we employ linear regression models in case of continuous covariates and probit models for binary covariates.

Finally, we follow Hirano and Imbens' (2004) approach of "blocking on the score" and divide 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. 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.⁸

4. Empirical Results

4.1 Estimates from a Probit Model

As mentioned in Section 2, 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. This corresponds to the standard approach in the literature on the effectiveness of training programs: outcomes of participants and non-participants are compared at specific points in time after the program started. However, in our analysis participants leave the program at different points in time. Since the search intensity for a new job might be low during program participation, it seems to be reasonable to alternatively compare individuals with different durations of training at the same point in time since they left the program. This ensures that participants with different training durations have the same time to find a job after leaving the program. Therefore, we additionally consider the employment probability at the point in time 1 year after the participants exited from the program. Before presenting results for the GPS, we explore first the relationship between post-treatment employment probability and the duration of treatment using a probit model. Table 1, 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 1a and 1b about here]

We report marginal effects instead of estimated coefficients for our probit model in order to facilitate the understanding of the relationship between employment probability and treatment

⁸ This test might suffer from “balance test fallacy”, i.e. we might observe for some covariates decreased t-statistics which might be driven by increased variances and not by decreased mean differences (see Imai, King and Stuart (2008) for a general discussion of balance test fallacy in the context of matching approaches). Therefore it is important to apply additional approaches to test the balancing property of the GPS.

duration.⁹ First, the two parts of Table 1 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.¹⁰ These suggest that the impact of treatment duration on the employment probability is small or negligible.¹¹ In Figures O5a – O6b in the online appendix we report the effects from the cubic specification of the probit models, which illustrate the average response levels to the corresponding treatment duration.

However, it is worth noting that a regression type analysis such as the probit or linear probability models may rely on extrapolation, compare incomparable observations, and have greater risk of misspecifying 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 Weak unconfoundedness assumption

The GPS approach is based on weak unconfoundedness as an identifying assumption. The factors which the caseworker takes into account when assigning clients to a training program center around the clients' aptitude for a certain job and the likelihood to succeed in a particular training program (Section 2). The information that the caseworker bases her decision on is largely congruent with the information contained in our data, in particular previous educational attainment and vocational education, and detailed labor market histories capturing all employment and unemployment spells from the previous four years. Controlling for these covariates, we are thus confident that the influence of any remaining unobserved factors is negligible and the application of the GPS justified. Moreover, to investigate the concern that treatment duration may possibly be endogenous, Tables 2a and 2b report instrumental variables (IV) estimates using planned duration as IV. Comparing these IV estimates with OLS estimates based on actual durations (not reported here), we find that they are in 15 out of 16 specifications not significantly different (see the results of the Hausman test in Tables 2a

⁹ The meaning of marginal effect in a probit model differs from the meaning of marginal effect in the treatment effect literature.

¹⁰ Cf. the low Pseudo R-squared in Panel A of Tables 1a and 1b.

¹¹ We additionally estimated linear probability models. The results are similar to the ones of the probit model.

and 2b). This suggests that the actual training duration does not suffer strongly from endogeneity.

[Tables 2a and 2b about here]

4.3 GPS estimation, common support and covariate balance

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.¹² To test the common support condition for the actual duration, following the approach outlined in section 3.3, we divide the sample into three groups, which are defined by cutting the distribution of treatment duration at the 30th and 70th percentiles. 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 3. We repeat the same procedure for group 2 and group 3, which gives us the top center and the top right panels of Figure 3. In a second step we delete those individuals from our sample that fall outside the common support region. For the actual duration we delete around 2 % of our sample. The distributions for the GPS after deleting the non-overlap are reported in the lower part of Figure 3.¹³

[Figure 3 about here]

To assess the balancing property of the GPS (cf. Section 3.3), we first regress every observable characteristic on the logarithm of the training duration and compare the coefficients for specifications without and with conditioning on the predicted individual duration $E(T|X_i)$. The results for the actual training duration are reported in the first four columns of Table A2 in the appendix. Without conditioning on the predicted training duration we observe many significant coefficients. For example, age is strongly and positively correlated with observed training duration. However, once we condition on $E(T|X_i)$, the coefficient for age decreases from 1.17 to 0.11 and is clearly insignificant. This also holds for previous employment outcomes, e.g. the negative correlation between being employed 12

¹² Estimation results are reported in Table O1 in the online appendix.

¹³ Corresponding common support graphs for Planned Training Duration and Actual = Planned Duration are reported in the online appendix.

weeks before entering the program decreases and becomes insignificant. We do not observe any significant correlations for the covariates once we condition on $E(T|X_i)$ and the corresponding coefficients clearly decrease, which indicates that the GPS properly balances the observable characteristics in our sample.

If we compare the regressions of the covariates on the training duration without and with conditioning on the distribution of R_i , which is reported in the last four columns of Table A2, we come to the same conclusion. All previously significant coefficients become insignificant once we condition on R_i and the corresponding coefficients clearly decrease. This picture is the same for the sample with actual=planned duration (cf. Table O2 in the online appendix).

As an additional test for the balancing property of the GPS we apply the approach of "blocking on the score" suggested by Hirano and Imbens (2004). 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 each of the covariates we test whether the difference in means of one group compared to the other two groups is significantly different. Without adjustment the clear majority of t-statistics are greater than 1.96, indicating a clearly unbalanced distribution of covariates. In the second step, we calculate the corresponding t-statistics for the GPS-adjusted sample. In contrast to the unadjusted sample, we observe no t-statistics larger than 1.96 for both the actual duration and for the sample with equal planned and actual durations (reported in Tables O3 and O4 in the online appendix). These results indicate that the balance of the covariates is clearly improved by adjustment for the GPS and confirm the results of the two other balancing tests.

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. The estimation results of the dose-response function for Actual Training Duration are reported in Table A4 in the appendix. The estimated coefficients in the dose-response function have no direct causal interpretation, and whether all the estimated coefficients associated with the GPS terms are equal to zero can indicate whether the covariates introduce any bias, as stressed in Hirano and Imbens (2004). We find that in the case of the outcome variable measured at two years after entry into the program, the coefficients of the GPS terms are more significant than in the case of the outcome variable measured at one year after exit from the program. This implies for the former that it is more important to apply the GPS technique to remove potential bias introduced by the covariates.

Our main results for both outcome variables are presented in Figure 4 (employment probability at time 2 years after entry into the program) and Figure 5 (employment probability at time 1 year after exit from the program), where each figure consists of two parts showing results for a) the actual duration and b) for the sample of individuals for which actual duration equals planned duration. The figures also include the counterfactual non-participant employment probability baseline¹⁴, which indicates that training effects are generally positive. This finding is in line with Rinne, Schneider and Uhlendorff (2010) who implement a binary treatment-comparison training evaluation also using the IEB data.

[Figures 4 and 5 about here]

Figures 4a and 5a show a similar shape for the dose-response function of the two employment outcomes based on actual training duration. Both graphs indicate a slowly and monotonously increasing employment response to different levels of the treatment until around 240 days (i.e. 8 months). While steadily increasing, the shape of the DRF is relatively flat and implies an increase in the post-training employment probability of approximately 2 percentage points from a treatment dose of 50 days (employment response approximately 0.34) to a treatment dose of 200 days (employment response approximately 0.36). The confidence bands, however, suggest that this increase might actually be zero. Training durations larger than 240 days do not seem to lead to a further increase in the employment probability, as indicated by the shape of the DRF and the confidence bands.

Looking at the dose-response function of the sample of participants with equal actual and planned training durations, this pattern becomes more pronounced (Figures 4b and 5b). For the employment outcome at time 2 years after entry into the program in Figure 4b there is a clear indication that the increase in the employment probability induced by the treatment duration occurs mainly during the first approximately 120 days. For longer training durations the DRF is flat and no additional employment impulse seems to be brought about by the treatment. Figure 5b further confirms this general pattern pointing to a core increase in employment probability for programs up to 120-150 days and no additional impact from a maximum of 200 days onward.¹⁵

¹⁴ 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.

¹⁵ Figures O3 and O4 in the online appendix show very similar dose-response functions for planned training duration.

In addition to the graphs displaying the dose-response function 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 training durations is significantly different from zero.

Table A3 in the appendix shows, for example, that for participants with actual equal to planned duration the difference in employment probability at time 2 years after entry into the program between a duration of 18 days and a duration of 102 days is significant at the 10% level. The point estimate suggests an increase of around 15 percentage points.¹⁶ If we compare the training durations of 102 days and 60 days, the estimates indicate an increased employment probability by 5.6 percentage points, which is significantly different from zero at the 5% level. Comparing the employment effect of a training of around 100 days with a training of around 300 days indicates a small and insignificant difference. However, for treatment effects one year after exit, we do not find significant differences between the different levels of treatment intensity. Similarly, for the actual training duration the differential duration impacts also remain largely insignificant, as indicated by the confidence bands in Figures 4a and 5a.

In sum, 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 to 150 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 150 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 a maximum of approximately five months of participation. The initial increase in the employment probability is much less pronounced for the actual training duration (Figures 4a and 5a). 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 sample with planned equal to actual durations leads to estimating a positive impact of the first approximately 120 days of training duration.

¹⁶ We only present Appendix Table A3 as part of the paper, corresponding tables containing point estimates and standard errors for pair-wise comparisons of alternative specifications, i.e. for actual training durations and for the second outcome, are reported in the online appendix, Tables O5 through O7.

These 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 (2010). They find only weak evidence for heterogeneous effects with respect to skill level 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 (2010). 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 the same program types. Our results suggest that an increased duration of training does not automatically go along with higher subsequent employment probabilities, and that after a maximum of around five 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 or receiving qualifications that are of general vocational use – that is, the contents of the programs analyzed here – have positive expected returns only up to a certain duration.

5. Robustness

In this section, we carry out sensitivity checks for our main specification. Some of the individuals in our sample participate in more than one training during their unemployment spell. Since we consider the duration of the first training as the treatment dose, this might not reflect the “true” treatment dose for every unemployed individual in our sample. Therefore the first check is that we 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 in order to investigate whether our results are sensitive to the specification we apply, which include different degrees of polynomial specifications and flexible spline models to check the functional form specifications.

Figures 4 and 5 above also 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.¹⁷ 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.

Our main impact estimates are based on a cubic specification for the dose-response function. We also estimate results for the dose-response function for the full sample for quadratic and 4th degree polynomial specifications as well (see online appendix, Figures O5a,b and O6a,b). These results 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 rather 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 the online appendix figures).¹⁸

Besides experimenting with different degrees of polynomial specifications, we also apply spline models to check the robustness of our results with respect to functional assumptions in our main models. Similar to a nonparametric regression, which requires the researcher to make decisions regarding e.g. the type of kernel and proper bandwidth, in the spline models we need to choose the type of the spline, number of knots and the position of the knots. Among them, the number of knots is the most important one, see Keele (2008). Figures 6 and 7 present results from the spline models based on B-splines with 12 knots for the specification based on actual duration. The results are very similar to the results from our main models. The only difference we observe compared to our main specification is that the results based on the spline model for the outcome at time two years after entry into program indicate a positive increase from around 50 days to around 100 days of training. However, this positive increase is in line with the results based on the sample of participants for which actual duration equals planned duration.

[Figures 6, 7 about here]

¹⁷ The balancing of the covariates is very similar for these subsamples and the corresponding tables are available upon request.

¹⁸ The corresponding balancing properties are available upon request. We mainly present results based on the linear specification because the corresponding balancing properties are better.

As mentioned earlier, the number of knots is important in the spline models; just like the bandwidth, it controls the smoothness of the fitting. Therefore we vary the number of knots from 6 to 12, and the results are quite similar except at the tails.

6. Conclusion

In this paper, we study the effect of the duration of 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 weak 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 training duration to the corresponding employment probability has a relatively flat shape after an initial increase during the first 120 days of training. Indeed, the first three to five 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-training employment prospects.

After approximately 150 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 (300 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 a maximum of about three to five months. This conclusion holds for the type of programs analyzed in this study, i.e. for training programs

which do not lead to a vocational degree. There exists evidence that substantive training programs with a duration of two years and leading to a vocational degree have a large positive impact on subsequent employment outcomes (Lechner et al. 2010).

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, C.A., A. Flores-Lagunes, A. Gonzalez and T.C. Neuman (2010), "Estimating the Effects of Length of Exposure to Instruction in a Training Program: The Case of Job Corps", *The Review of Economics and Statistics*, forthcoming.
- 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 labor 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.
- Imai K. G. King and E.A. Stuart (2008), "Misunderstandings between experimentalists and observationalists about causal inference", *Journal of the Royal Statistical Society Series A*, 171, 481-502.

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 (2004), "Identification and Estimation of Local Average Treatment Effects," *Econometrica* 62, 467-475.

IZA, DIW, ifas (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. Kluve (2007), "Before and After the Hartz Reforms: The Performance of Active Labour Market Policy in Germany", *Journal for Labour Market Research* 40, 45-64.

Keele, Luke (2008), "Semiparametric Regression for the Social Sciences," New York: Wiley.

Kluve, J. (2010), "The effectiveness of European Active Labour Market Policy", *Labour Economics*, 17, 904-918.

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 (2010), "Long-Run Effects of Public Sector Sponsored Training in West Germany", *Journal of the European Economic Association*, forthcoming.

Moffitt, Robert (2009), "Estimating Marginal Treatment Effects in Heterogeneous Populations," *Annales d'Economie et de Statistique*, Fall 2009, 239-261.

Mueser P. R., K. R. Troske and A. Gorislawsky (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 (2010), "Do the Skilled and Prime-Aged Unemployed Benefit More from Training? Effect Heterogeneity of Public Training Programs in Germany", *Applied Economics*, forthcoming.

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 1a. Effect of treatment duration on employment probability at time 2 years after entry into the program – Estimates from a Probit Model

Dependent Variable: Employment status at time 2 years after entry into the program				
	(1)	(2)	(3)	(4)
	Marginal Effect	Marginal Effect	Marginal Effect	Marginal Effect
Variables	<i>Std. Error</i>	<i>Std. Error</i>	<i>Std. Error</i>	<i>Std. Error</i>
Panel A: Only Control for Treatment Duration				
Treatment Duration/100	0.0061	0.0526	0.0630	0.0239
	<i>0.0085</i>	<i>0.0363</i>	<i>0.1062</i>	<i>0.2488</i>
Square of Treatment Duration/100		-0.0112	-0.0175	0.0224
		<i>0.0085</i>	<i>0.0604</i>	<i>0.2374</i>
Cube of Treatment Duration/100			0.0010	-0.0141
			<i>0.0099</i>	<i>0.0879</i>
Fourth Power of Treatment Duration/100				0.0019
				<i>0.0110</i>
Pseudo R Squared	0.0001	0.0006	0.0006	0.0006
Number of Observations	3162	3162	3162	3162
Panel B: Control for Treatment Duration and Other Variables				
Treatment Duration/100	0.0057	0.0797	0.1912	0.0446
	<i>0.0100</i>	<i>0.0385</i>	<i>0.1119</i>	<i>0.2595</i>
Square of Treatment Duration/100		-0.0182	-0.0850	0.0650
		<i>0.0092</i>	<i>0.0636</i>	<i>0.2481</i>
Cube of Treatment Duration/100			0.0111	-0.0461
			<i>0.0105</i>	<i>0.0920</i>
Fourth Power of Treatment Duration/100				0.0072
				<i>0.0115</i>
Other Control Variables: See Table 1				
Pseudo R Squared	0.1321	0.1331	0.1334	0.1335
Number of Observations	3130	3130	3130	3130

Source: Integrated Employment Biographies (IEB) of the German Federal Employment Agency.

Notes: Marginal effects refer to the effect of a change in the training duration on the probability of being employed 2 years after entry into the program, evaluated for the average participant. Sample consists of male participants in training programs in West Germany for the years 2000 to 2002. Additional control variables in Panel B include age, disability, citizenship, educational attainment, vocational attainment, and employment history. Treatment duration in days is divided by 100 to get readable effects.

Table 1b. Effect of treatment duration on employment probability at time 1 year after exit from the program – Estimates from a Probit Model

Dependent Variable: Employment status at time 1 year after exit from the program				
	(1)	(2)	(3)	(4)
	Marginal Effect	Marginal Effect	Marginal Effect	Marginal Effect
Variables	<i>Std. Error</i>	<i>Std. Error</i>	<i>Std. Error</i>	<i>Std. Error</i>
Panel A: Only Control for Treatment Duration				
Treatment Duration/100	0.0069	0.0246	-0.0582	-0.0962
	<i>0.0085</i>	<i>0.0361</i>	<i>0.1056</i>	<i>0.2479</i>
Square of Treatment Duration/100		-0.0043	0.0454	0.0842
		<i>0.0085</i>	<i>0.0602</i>	<i>0.2367</i>
Cube of Treatment Duration/100			-0.0083	-0.0230
			<i>0.0099</i>	<i>0.0876</i>
Fourth Power of Treatment Duration/100				0.0019
				<i>0.0110</i>
Pseudo R Squared	0.0002	0.0002	0.0004	0.0004
Number of Observations	3162	3162	3162	3162
Panel B: Control for Treatment Duration and Other Variables				
Treatment Duration/100	0.0074	0.0485	0.0431	-0.1124
	<i>0.0099</i>	<i>0.0381</i>	<i>0.1106</i>	<i>0.2581</i>
Square of Treatment Duration/100		-0.0101	-0.0069	0.1521
		<i>0.0091</i>	<i>0.0630</i>	<i>0.2468</i>
Cube of Treatment Duration/100			-0.0005	-0.0611
			<i>0.0104</i>	<i>0.0915</i>
Fourth Power of Treatment Duration/100				0.0076
				<i>0.0115</i>
Other Control Variables: See Table 1				
Pseudo R Squared	0.1183	0.1186	0.1186	0.1187
Number of Observations	3130	3130	3130	3130

Source: Integrated Employment Biographies (IEB) of the German Federal Employment Agency.

Notes: Marginal effects refer to the effect of a change in the training duration on the probability of being employed 1 year after exit from the program, evaluated for the average participant. Sample consists of male participants in training programs in West Germany for the years 2000 to 2002. Additional control variables in Panel B include age, disability, citizenship, educational attainment, vocational attainment, and employment history. Treatment duration in days is divided by 100 to get readable effects.

Table 2a. Effect of treatment duration on employment probability at time 2 years after entry into the program – IV estimates from a Linear Probability Model

	(1)	(2)	(3)	(4)
Variables	Coefficient <i>Std. Error</i>	Coefficient <i>Std. Error</i>	Coefficient <i>Std. Error</i>	Coefficient <i>Std. Error</i>
Panel A: Only Control for Treatment Duration				
Constant	0.3433 <i>0.0236</i>	0.3246 <i>0.0602</i>	0.3426 <i>0.1123</i>	0.0080 <i>0.1887</i>
Treatment Duration/100	0.0020 <i>0.0113</i>	0.0237 <i>0.0639</i>	-0.0136 <i>0.2019</i>	1.0427 <i>0.5094</i>
Square of Treatment Duration/100		-0.0049 <i>0.0140</i>	0.0157 <i>0.1049</i>	-0.9798 <i>0.4466</i>
Cube of Treatment Duration/100			-0.0033 <i>0.0163</i>	0.3573 <i>0.1564</i>
Fourth Power of Treatment Duration/100				-0.0440 <i>0.0188</i>
Adjusted R Squared	-0.0002	-0.0001	-0.0006	0.0000
Number of Observations	3162	3162	3162	3162
Hausman Test: Chi-Squared	0.3000	0.3200	0.3400	11.2500
Probability>Chi-Squared	0.8601	0.9563	0.9527	0.0239
Panel B: Control for Treatment Duration and Other Variables				
Constant	-0.0745 <i>0.4759</i>	-0.1768 <i>0.4787</i>	-0.2855 <i>0.4850</i>	-0.5157 <i>0.5051</i>
Treatment Duration/100	0.0014 <i>0.0129</i>	0.1228 <i>0.0603</i>	0.3164 <i>0.1924</i>	1.0302 <i>0.4824</i>
Square of Treatment Duration/100		-0.0278 <i>0.0133</i>	-0.1347 <i>0.0999</i>	-0.8080 <i>0.4226</i>
Cube of Treatment Duration/100			0.0170 <i>0.0156</i>	0.2611 <i>0.1481</i>
Fourth Power of Treatment Duration/100				-0.0298 <i>0.0178</i>
Other Control Variables: See Table 1				
Adjusted R Squared	0.1361	0.1360	0.1350	0.1296
Number of Observations	3130	3130	3130	3130
Hausman Test: Chi-Squared	0.0500	1.4200	2.2600	6.7600
Probability>Chi-Squared	1.0000	1.0000	1.0000	1.0000

Source: Integrated Employment Biographies (IEB) of the German Federal Employment Agency.

Notes: Planned duration is used as an instrumental variable for actual duration, the dependent variable is the probability of being employed 2 years after entry into the program. Sample consists of male participants in training programs in West Germany for the years 2000 to 2002. Hausman tests refer to the differences of the IV estimates and OLS estimates. Additional control variables in Panel B include age, disability, citizenship, educational attainment, vocational attainment, and employment history. Treatment duration in days is divided by 100 to get readable coefficients.

Table 2b. Effect of treatment duration on employment probability at time 1 year after exit from the program – IV estimates from a Linear Probability Model

	(1)	(2)	(3)	(4)
Variables	Coefficient <i>Std. Error</i>	Coefficient <i>Std. Error</i>	Coefficient <i>Std. Error</i>	Coefficient <i>Std. Error</i>
Panel A: Only Control for Treatment Duration				
Constant	0.3232 <i>0.0235</i>	0.2993 <i>0.0600</i>	0.3603 <i>0.1120</i>	0.2232 <i>0.1878</i>
Treatment Duration/100	0.0102 <i>0.0113</i>	0.0379 <i>0.0637</i>	-0.0886 <i>0.2014</i>	0.3443 <i>0.5069</i>
Square of Treatment Duration/100		-0.0063 <i>0.0140</i>	0.0636 <i>0.1046</i>	-0.3444 <i>0.4444</i>
Cube of Treatment Duration/100			-0.0111 <i>0.0163</i>	0.1367 <i>0.1556</i>
Fourth Power of Treatment Duration/100				-0.0180 <i>0.0187</i>
Adjusted R Squared	-0.0002	-0.0005	-0.0005	0.0000
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	-0.1324 <i>0.4792</i>	-0.2332 <i>0.4826</i>	-0.2227 <i>0.4888</i>	-0.2637 <i>0.5076</i>
Treatment Duration/100	0.0151 <i>0.0130</i>	0.1346 <i>0.0608</i>	0.1993 <i>0.1939</i>	0.3405 <i>0.4851</i>
Square of Treatment Duration/100		-0.0274 <i>0.0134</i>	-0.0631 <i>0.1008</i>	-0.1963 <i>0.4250</i>
Cube of Treatment Duration/100			0.0057 <i>0.0157</i>	0.0540 <i>0.1489</i>
Fourth Power of Treatment Duration/100				-0.0059 <i>0.0179</i>
Other Control Variables: See Table 1				
Adjusted R Squared	0.1187	0.1163	0.1154	0.1145
Number of Observations	3130	3130	3130	3130
Hausman Test: Chi-Squared	1.0200	3.8700	4.2100	4.0800
Probability>Chi-Squared	1.0000	1.0000	1.0000	1.0000

Source: Integrated Employment Biographies (IEB) of the German Federal Employment Agency.

Notes: Planned duration is used as an instrumental variable for actual duration; the dependent variable is the probability of being employed 1 year after exit from the program. Sample consists of male participants in training programs in West Germany for the years 2000 to 2002. Hausman tests refer to the differences of the IV estimates and OLS estimates. Additional control variables in Panel B include age, disability, citizenship, educational attainment, vocational attainment, and employment history. Treatment duration in days is divided by 100 to get readable coefficients.

Figure 1. Distributions of Actual and Planned Training Durations

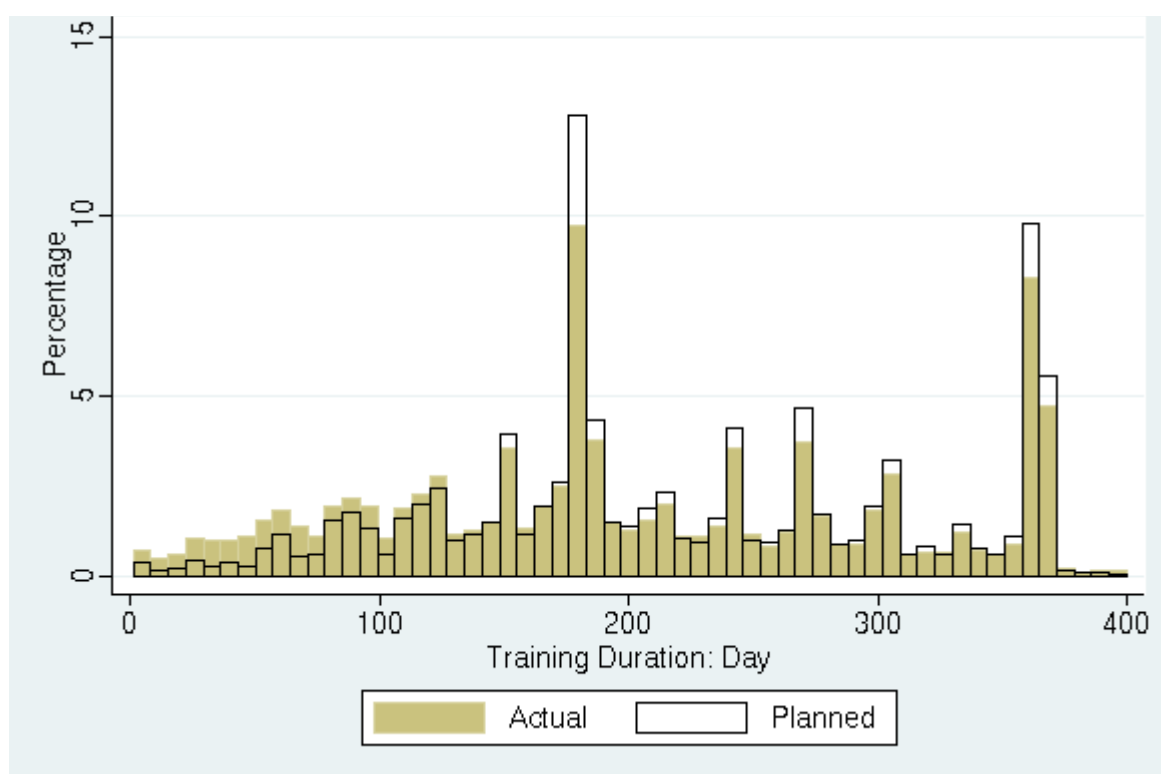


Figure 2a. Unadjusted employment probability at time 2 years after entry into the program – Actual Training Duration

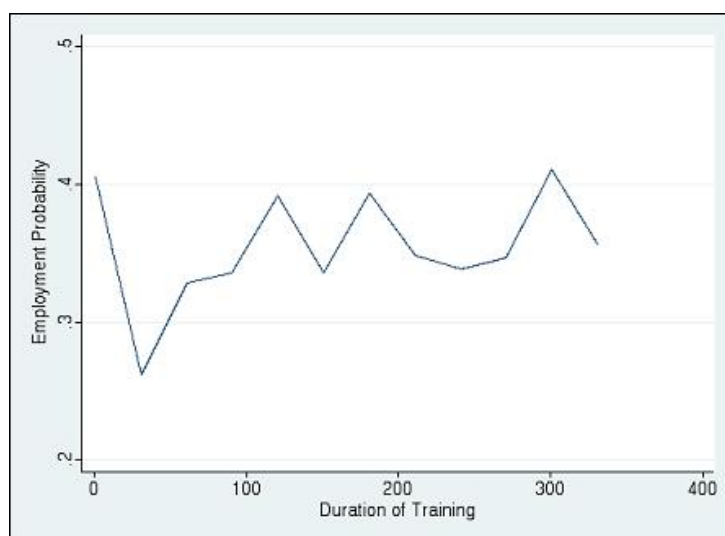


Figure 2a. Unadjusted employment probability at time 2 years after entry into the program – Planned Training Duration

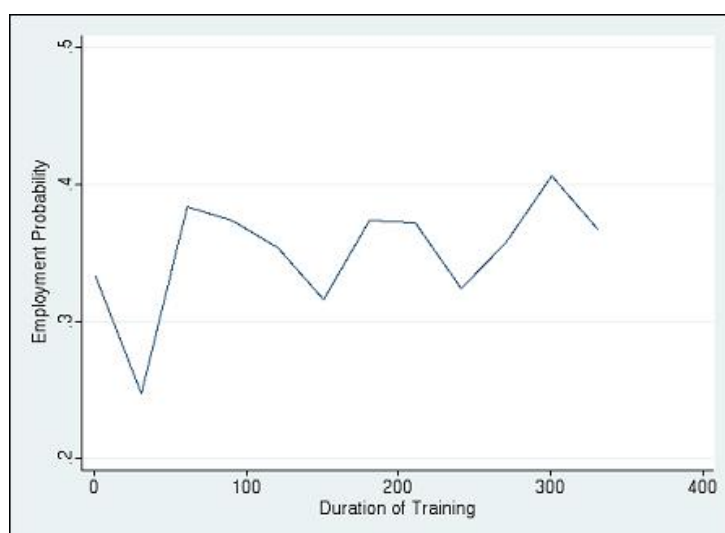


Figure 3. Common Support Condition – Actual Training Duration

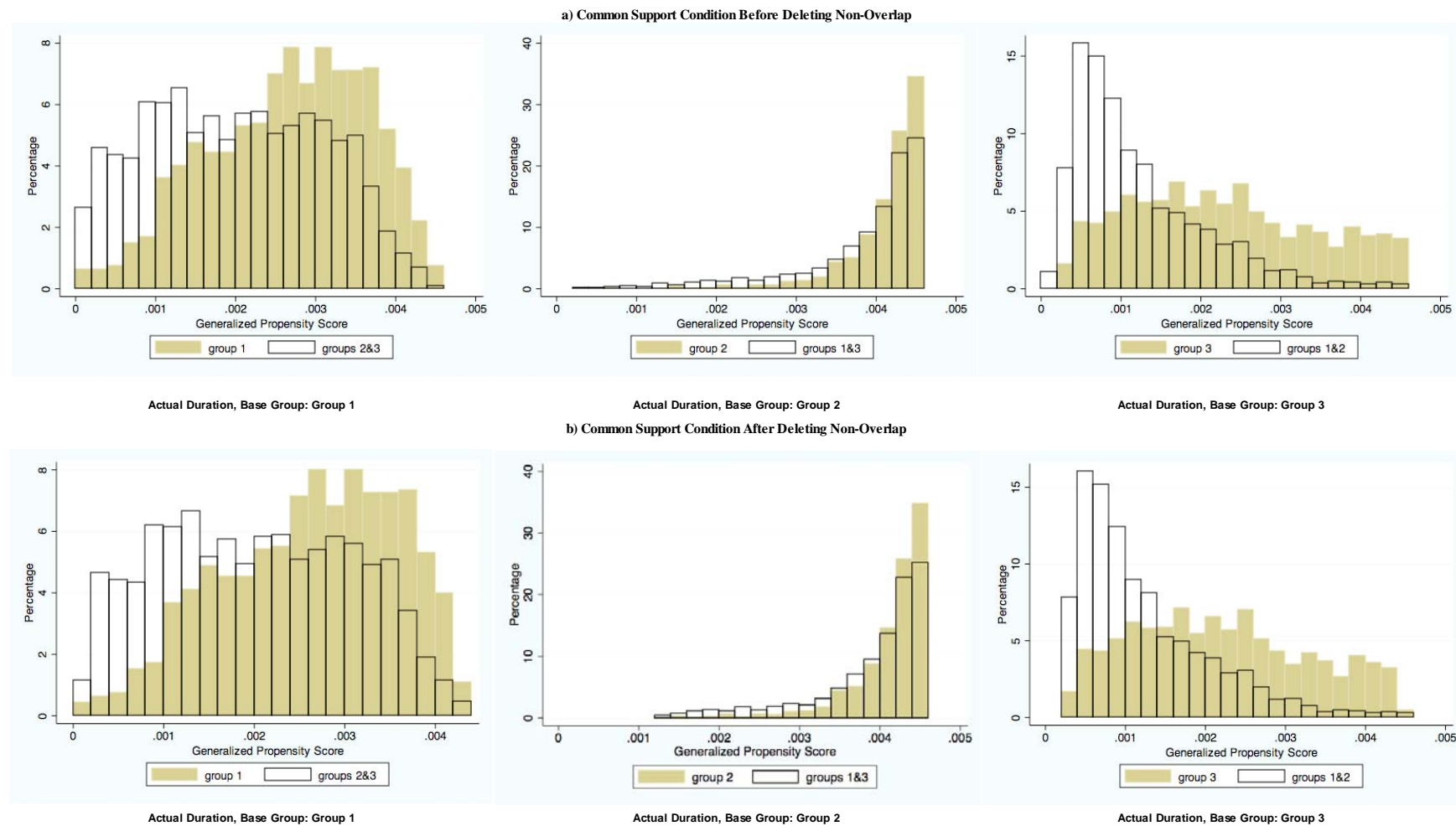


Figure 4a. Dose Response Function of employment probability at time 2 years after entry into the program – Actual Training Duration

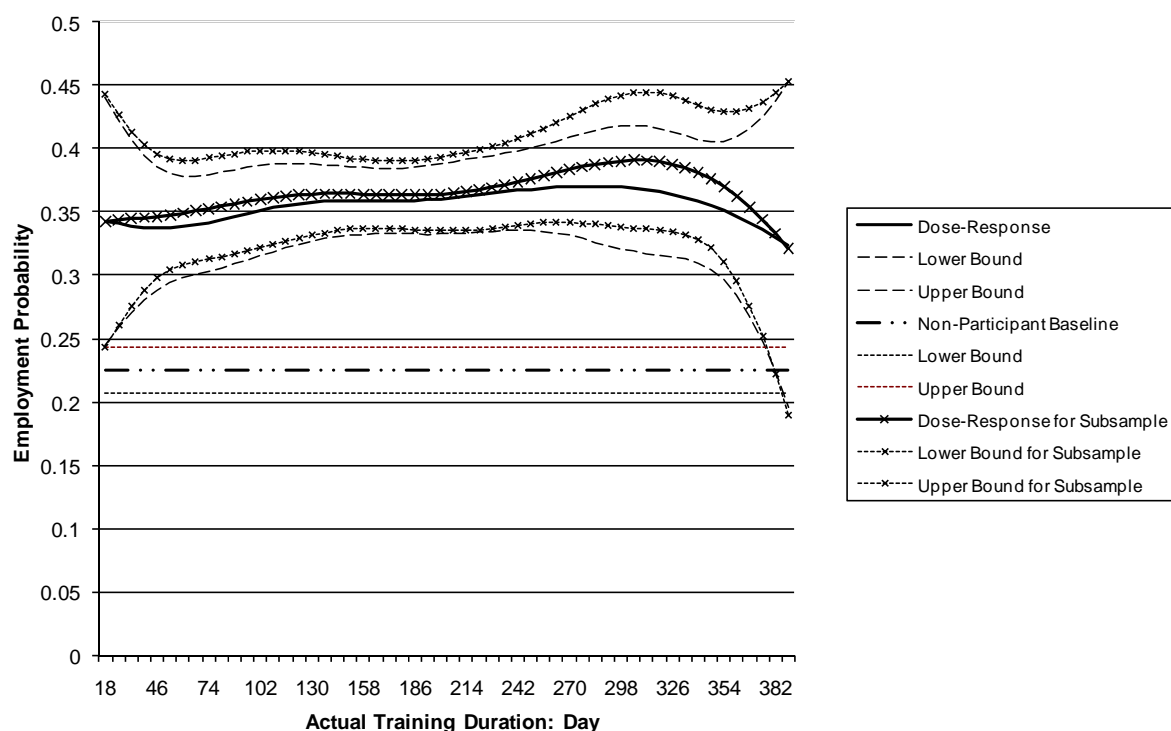
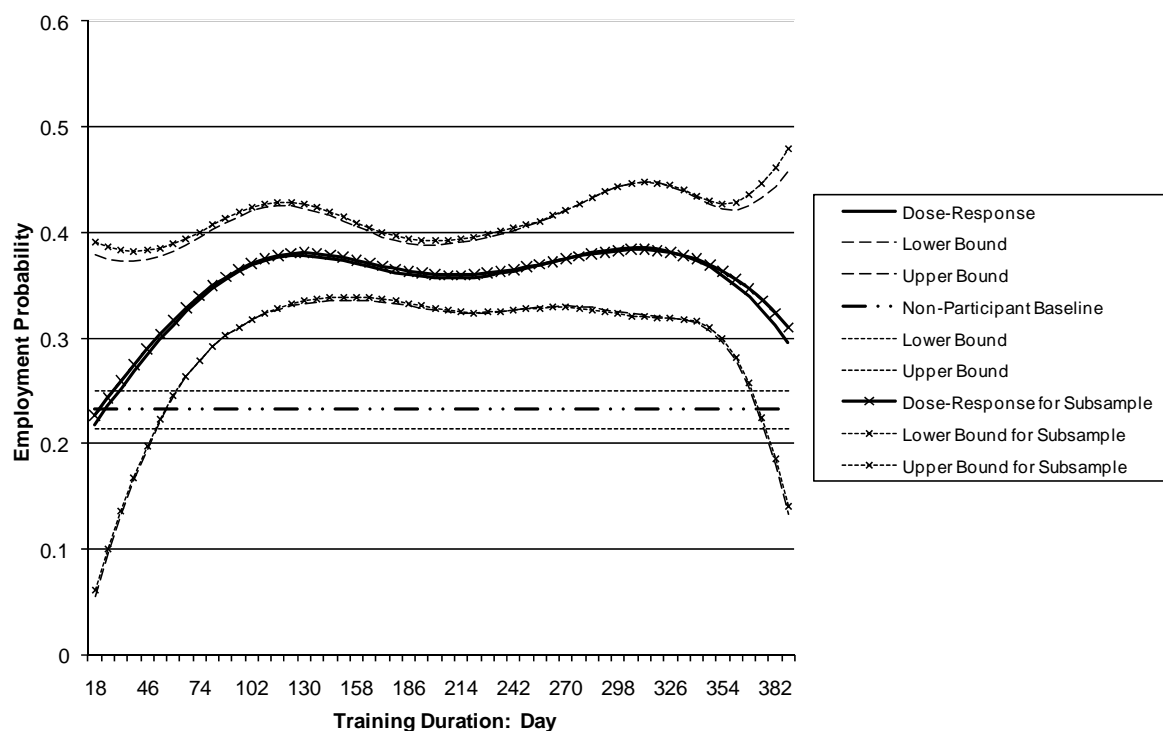


Figure 4b. Dose Response Function of employment probability at time 2 years after entry into the program – Actual = Planned Training Duration



Note: "Dose-Response for Subsample" means dose response function for a subgroup of individuals who went through a training program exactly once.

Figure 5a. Dose Response Function of employment probability at time 1 year after exit from the program – Actual Training Duration

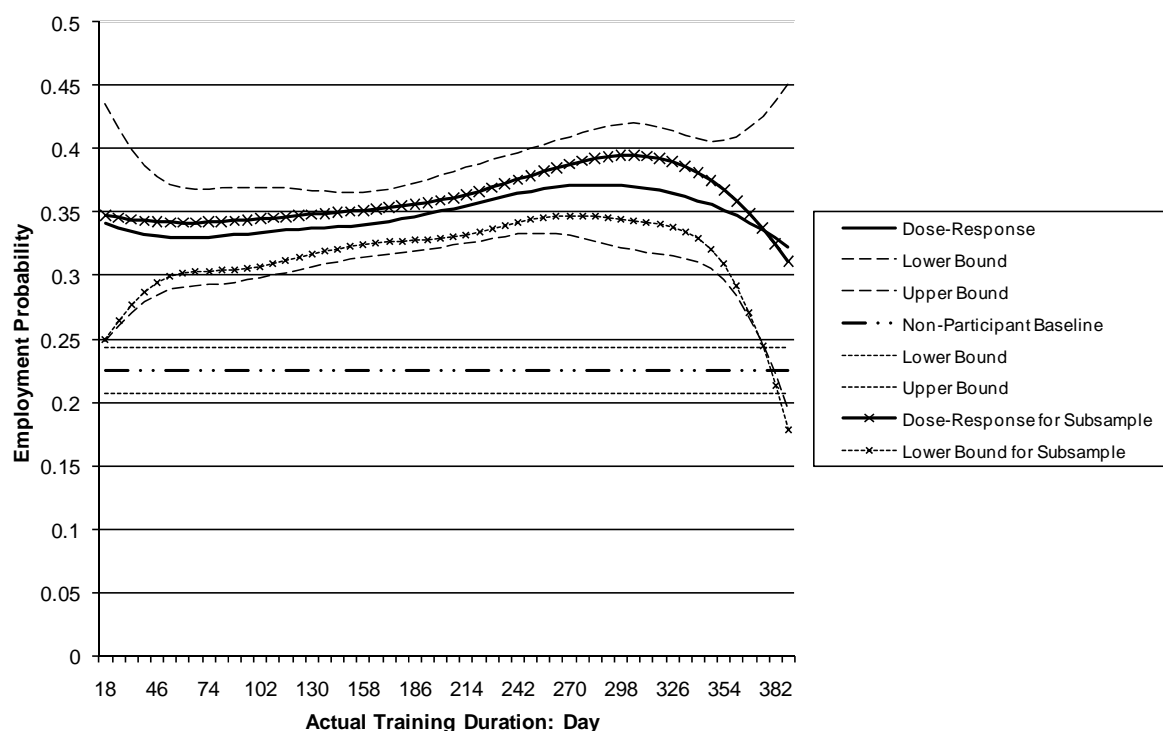
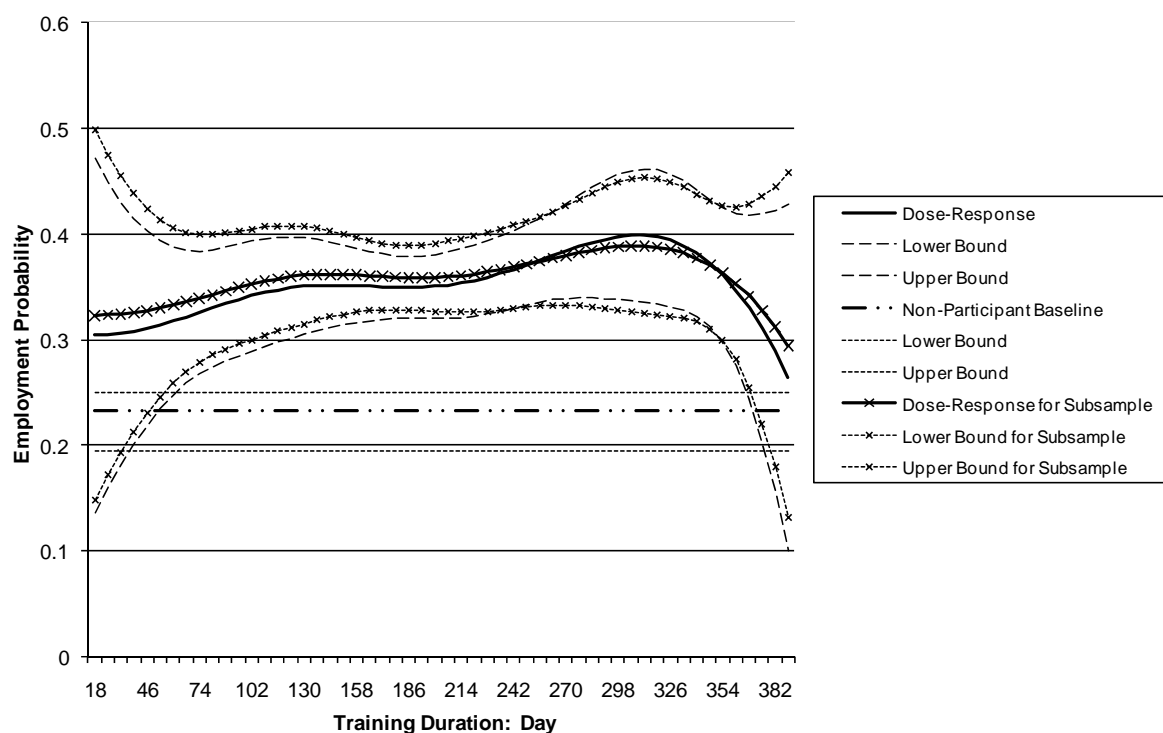


Figure 5b. Dose Response Function of employment probability at time 1 year after exit from the program – Actual = Planned Training Duration



Note: "Dose-Response for Subsample" means dose response function for a subgroup of individuals who went through a training program exactly once.

Figure 6. Spline models: Employment probability at time 2 years after entry into the program – Actual Training Duration

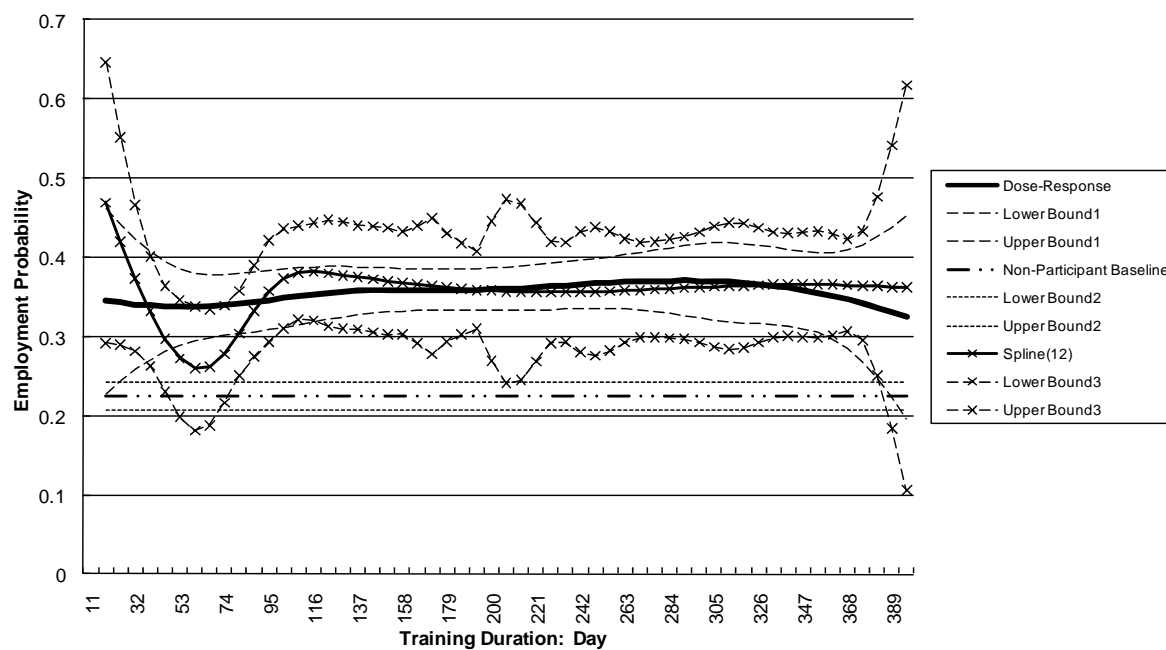
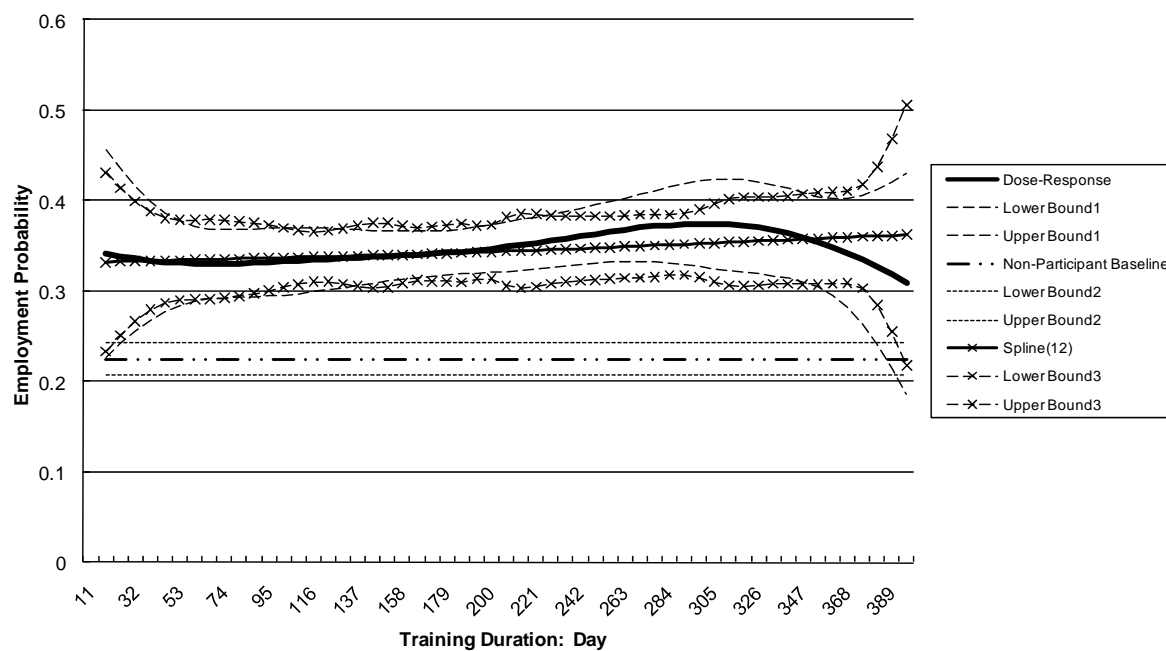


Figure 7. Spline models: Employment probability at time 1 year after exit from the program – Actual Training Duration



Appendix.

Table A1. 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	-
Employment 4 weeks before program entry	0.07	-	0.07	-	0.10	-	0.07	-
Employment 8 weeks before program entry	0.14	-	0.16	-	0.17	-	0.14	-
Employment 12 weeks before program entry	0.20	-	0.21	-	0.21	-	0.19	-
...								
Employment 48 weeks before program entry	0.42	-	0.39	-	0.45	-	0.42	-
<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	

Source: Integrated Employment Biographies (IEB) of the German Federal Employment Agency.

Notes: Sample consists of male participants in training programs in West Germany for the years 2000 to 2002. Subsample “early exists” contains individuals with actual training duration < planned duration, subsample “late exists” refers to actual duration > planned duration and subsample “exits as planned” contains individuals for whom planned duration equals actual duration. All time-varying characteristics are measured at the beginning of the training.

Table A2. Covariate Balance with and without adjustment – Actual Training Duration

Covariate	Unconditional Effect of log(duration)		Effect of log(duration) conditional on E(T Xi)		Unconditional Effect of duration		Effect of duration conditional on Ri	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Age	1.1782	0.2835	0.1134	0.3000	0.8562	0.1883	0.0303	0.2082
<u>Disability</u>								
No disability1	-0.0215	0.0081	-0.0036	0.0083	-0.0145	0.0051	-0.0017	0.0053
Disability low degree1	0.0126	0.0075	0.0030	0.0079	0.0077	0.0048	0.0009	0.0050
Disability medium degree1	0.0063	0.0016	0.0005	0.0006	0.0035	0.0009	0.0000	0.0002
Disability high degree1	0.0021	0.0023	-0.0005	0.0020	0.0018	0.0014	-0.0001	0.0012
<u>Citizenship</u>								
German	0.0146	0.0086	-0.0002	0.0092	0.0127	0.0059	0.0004	0.0065
Foreigner EU	-0.0073	0.0029	-0.0015	0.0026	-0.0057	0.0022	-0.0006	0.0020
Foreigner Non-EU	-0.0063	0.0081	0.0025	0.0087	-0.0070	0.0055	0.0005	0.0061
<u>Educational Attainment</u>								
No graduation (1)	-0.0317	0.0086	-0.0057	0.0090	-0.0239	0.0061	-0.0001	0.0065
First stage of secondary level (2)	-0.0912	0.0138	0.0165	0.0154	-0.0830	0.0093	0.0061	0.0108
Second stage of secondary level (3)	0.0439	0.0122	-0.0003	0.0129	0.0347	0.0079	-0.0024	0.0088
Advanced technical college entrance qualification (4)	0.0187	0.0058	-0.0010	0.0052	0.0135	0.0034	-0.0004	0.0032
General qualification for university entrance (5)	0.0706	0.0083	-0.0054	0.0052	0.0489	0.0046	-0.0017	0.0037
<u>Vocational Attainment</u>								
No vocational degree (1)	-0.0933	0.0130	-0.0013	0.0141	-0.0763	0.0089	0.0010	0.0101
In-plant training (2)	0.0190	0.0137	0.0100	0.0147	0.0084	0.0091	0.0008	0.0104
Off-the-job training, vocational school, technical school (3)	0.0204	0.0067	-0.0067	0.0060	0.0187	0.0041	-0.0015	0.0037
University, advanced technical college (4)	0.0641	0.0063	0.0003	0.0015	0.0376	0.0035	-0.0001	0.0006

Source: Integrated Employment Biographies (IEB) of the German Federal Employment Agency.

Notes: Columns 1 and 2 contain the results of regressing the corresponding observable characteristic on the logarithm of the training duration. Columns 3 and 4 contain the same results conditional on the predicted individual duration E(T|Xi). Columns 5 and 6 report the results of regressing the corresponding observable characteristic on the training duration, while columns 7 and 8 contain the same results conditional on the distribution of the generalized propensity score Ri.

Bold numbers indicate significance at the 5% level. Duration in days divided by 100.

Table A2. (contd.)

Covariate	Unconditional Effect of log(duration)		Effect of log(duration) conditional on E(T Xi)		Unconditional Effect of duration		Effect of duration conditional on Ri	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Employment History								
Previous Unemployment Duration	0.2770	0.2084	0.2234	0.2239	0.0541	0.1385	0.0420	0.1550
Duration of last employment	64.0300	25.0149	11.5619	26.7505	41.5635	16.6230	0.5909	18.5098
Log(wage) of last employment	-0.0408	0.0317	-0.0069	0.0341	-0.0303	0.0211	-0.0054	0.0236
No last employment observed	0.0227	0.0079	0.0056	0.0081	0.0137	0.0049	0.0019	0.0054
Share of days in emp., 1 st year before program	-0.0561	0.0308	-0.0194	0.0330	-0.0308	0.0205	-0.0085	0.0229
Share of days in emp., 2 nd year before program	-0.0301	0.0454	-0.0500	0.0487	0.0118	0.0302	-0.0073	0.0338
Share of days in emp., 3 rd year before program	0.0022	0.0471	-0.0165	0.0505	0.0155	0.0313	-0.0027	0.0350
Share of days in emp., 4 th year before program	0.0742	0.0476	0.0076	0.0511	0.0555	0.0316	0.0030	0.0353
Share of days in unemp., 1 st year before program	0.0108	0.0347	0.0127	0.0373	-0.0034	0.0231	0.0026	0.0258
Share of days in unemp., 2 nd year before program	-0.0904	0.0426	0.0329	0.0453	-0.0988	0.0283	-0.0001	0.0314
Share of days in unemp., 3 rd year before program	-0.1027	0.0422	0.0293	0.0448	-0.1056	0.0280	-0.0012	0.0310
Share of days in unemp., 4 th year before program	-0.1688	0.0411	-0.0043	0.0434	-0.1354	0.0273	-0.0066	0.0300
Employment 4 weeks before program entry	-0.0051	0.0069	-0.0008	0.0074	-0.0024	0.0047	0.0009	0.0052
Employment 8 weeks before program entry	-0.0320	0.0092	-0.0119	0.0098	-0.0188	0.0064	-0.0033	0.0071
Employment 12 weeks before program entry	-0.0257	0.0108	-0.0097	0.0115	-0.0153	0.0073	-0.0041	0.0081
Employment 16 weeks before program entry	-0.0151	0.0119	-0.0028	0.0127	-0.0108	0.0080	-0.0032	0.0089
Employment 20 weeks before program entry	-0.0026	0.0125	-0.0016	0.0134	-0.0011	0.0083	-0.0026	0.0093
Employment 24 weeks before program entry	-0.0067	0.0128	-0.0079	0.0137	0.0002	0.0085	-0.0030	0.0095
Employment 28 weeks before program entry	-0.0055	0.0131	-0.0086	0.0141	0.0014	0.0087	-0.0034	0.0098
Employment 32 weeks before program entry	-0.0004	0.0134	-0.0098	0.0143	0.0062	0.0089	-0.0025	0.0100
Employment 36 weeks before program entry	0.0009	0.0134	-0.0084	0.0144	0.0066	0.0089	-0.0021	0.0100
Employment 40 weeks before program entry	0.0008	0.0134	-0.0104	0.0144	0.0077	0.0089	-0.0010	0.0100
Employment 44 weeks before program entry	-0.0130	0.0135	-0.0174	0.0145	0.0016	0.0090	-0.0015	0.0101
Employment 48 weeks before program entry	-0.0158	0.0135	-0.0151	0.0145	-0.0017	0.0090	-0.0011	0.0101
Regional indicators								
Regional type 1	0.1043	0.0124	-0.0144	0.0125	0.0854	0.0078	-0.0026	0.0059
Regional type 2	0.0589	0.0115	0.0060	0.0119	0.0402	0.0072	0.0026	0.0077
Regional type 3	-0.0991	0.0129	0.0078	0.0140	-0.0885	0.0089	0.0024	0.0100
Regional type 4	0.0069	0.0059	-0.0022	0.0060	0.0067	0.0038	0.0007	0.0039
Regional type 5	-0.0612	0.0101	-0.0008	0.0102	-0.0532	0.0071	-0.0017	0.0076
Regional unemployment rate	0.0180	0.0015	-0.0008	0.0013	0.0152	0.0010	-0.0006	0.0008

Table A3. Duration effects at time 2 years after entry into the program – Actual = Planned Training Duration

Duration	18 days	60 days	102 days	144 days	186 days	228 days	270 days	312 days	354 days
18 days	0.000	-0.095	-0.151	-0.158	-0.143	-0.142	-0.158	-0.168	-0.142
		<i>0.061</i>	<i>0.087</i>	<i>0.090</i>	<i>0.086</i>	<i>0.081</i>	<i>0.082</i>	<i>0.087</i>	<i>0.087</i>
60 days	0.095	0.000	-0.056	-0.063	-0.047	-0.047	-0.063	-0.072	-0.046
	<i>0.061</i>		<i>0.028</i>	<i>0.036</i>	<i>0.038</i>	<i>0.040</i>	<i>0.042</i>	<i>0.047</i>	<i>0.045</i>
102 days	0.151	0.056	0.000	-0.007	0.009	0.009	-0.007	-0.017	0.010
	<i>0.087</i>	<i>0.028</i>		<i>0.014</i>	<i>0.026</i>	<i>0.034</i>	<i>0.036</i>	<i>0.039</i>	<i>0.038</i>
144 days	0.158	0.063	0.007	0.000	0.016	0.016	0.000	-0.009	0.017
	<i>0.090</i>	<i>0.036</i>	<i>0.014</i>		<i>0.016</i>	<i>0.026</i>	<i>0.029</i>	<i>0.035</i>	<i>0.035</i>
186 days	0.143	0.047	-0.009	-0.016	0.000	0.001	-0.016	-0.025	0.001
	<i>0.086</i>	<i>0.038</i>	<i>0.026</i>	<i>0.016</i>		<i>0.013</i>	<i>0.023</i>	<i>0.035</i>	<i>0.035</i>
228 days	0.142	0.047	-0.009	-0.016	-0.001	0.000	-0.016	-0.026	0.000
	<i>0.081</i>	<i>0.040</i>	<i>0.034</i>	<i>0.026</i>	<i>0.013</i>		<i>0.018</i>	<i>0.035</i>	<i>0.037</i>
270 days	0.158	0.063	0.007	0.000	0.016	0.016	0.000	-0.009	0.017
	<i>0.082</i>	<i>0.042</i>	<i>0.036</i>	<i>0.029</i>	<i>0.023</i>	<i>0.018</i>		<i>0.019</i>	<i>0.036</i>
312 days	0.168	0.072	0.017	0.009	0.025	0.026	0.009	0.000	0.026
	<i>0.087</i>	<i>0.047</i>	<i>0.039</i>	<i>0.035</i>	<i>0.035</i>	<i>0.035</i>	<i>0.019</i>		<i>0.033</i>
354 days	0.142	0.046	-0.010	-0.017	-0.001	0.000	-0.017	-0.026	0.000
	<i>0.087</i>	<i>0.045</i>	<i>0.038</i>	<i>0.035</i>	<i>0.035</i>	<i>0.037</i>	<i>0.036</i>	<i>0.033</i>	

Source: Integrated Employment Biographies (IEB) of the German Federal Employment Agency.

Notes: The 1st row and the 1st column are training durations. Table entries are differences in treatment effects from duration in the first column compared to treatment effects from duration in the first row. Treatment effects are based on the estimated dose response function. Sample consists of male participants in training programs in West Germany for the years 2000 to 2002. Standard errors are bootstrapped based on 2000 replications and are reported in italics.

Table A4. Estimated dose-response functions – Actual Training Duration

	(1)	(2)
	Actual Duration	
	Coeff.	Std. Error
Panel A: Outcome Variable: Employment status at time		
2 years after entry into the program		
GPS	-2.4486	1.2408
GPS ²	8.0887	4.8588
GPS ³	-8.0235	5.8881
Program Duration	0.0996	0.1769
Program Duration ²	-0.0420	0.0822
Program Duration ³	0.0022	0.0133
GPS*Program Duration	0.4548	0.6005
GPS ² *Program Duration	-0.9821	0.6654
GPS*Program Duration ²	0.0088	0.1281
Constant	0.4534	0.1105
Adjusted R Squared	0.0000	
Number of Observations	3070	
Panel B: Outcome Variable: Employment status at time		
1 year after exit from the program		
GPS	-0.4982	1.1947
GPS ²	2.3839	4.6814
GPS ³	-3.0652	5.6882
Program Duration	-0.0205	0.1683
Program Duration ²	0.0349	0.0792
Program Duration ³	-0.0082	0.0129
GPS*Program Duration	0.0059	0.5563
GPS ² *Program Duration	-0.1736	0.6275
GPS*Program Duration ²	0.0163	0.1180
Constant	0.3595	0.1067
Adjusted R Squared	-0.0019	
Number of Observations	3069	

Source: Integrated Employment Biographies (IEB) of the German Federal Employment Agency.

Notes: Sample consists of male participants in training programs in West Germany for the years 2000 to 2002. Dependent variable is the probability of being employed 2 years after entry into the program (Panel A) and 1 year after exit from the program (Panel B), respectively. Duration in days divided by 100. The coefficients of the generalized propensity score are reported in Table O1.