

Job Search, Conditional Treatment and Recidivism: The Employment Services for Ex-Offenders Program Reconsidered*

Herman J. Bierens[†]

Department of Economics and CAPCP,[‡] Pennsylvania State University

José R. Carvalho[§]

CAEN, Universidade Federal do Ceará, Fortaleza, Brazil.

November 8, 2010

*Forthcoming in *The B.E. Journal of Economic Analysis & Policy*.

[†]Correspondence address: Herman J. Bierens, Department of Economics, Pennsylvania State University, 608 Kern Graduate Building, University Park, PA 16802. E-mail: hbierens@psu.edu. This paper was presented by Herman Bierens at Lehigh University and the Tinbergen Institute, Amsterdam. The comments of Jaap Abbring, Gerard van den Berg, and two referees are gratefully acknowledged.

[‡]Support for research within the Center for the Study of Auctions, Procurements, and Competition Policy (CAPCP) at Penn State has been provided by a gift from the Human Capital Foundation (<http://www.hcfoundation.ru/en/>).

[§]E-mail: josecarv@ufc.br. The financial support of CAPES Foundation, Brazil, is gratefully acknowledged.

Abstract

The objective of this paper is to re-evaluate the effect of the 1985 "Employment Services for Ex-Offenders" (ESEO) program on recidivism, in San Diego, Chicago and Boston. The initial group of program participants was split randomly in a control group and a treatment group. The actual treatment (mainly being job related counseling) only takes place conditional on finding a job, and not having been arrested, for those selected in the treatment group. We use interval-censored proportional hazard models for job search and recidivism time, where the latter model incorporates the conditional treatment effect, depending on covariates. We find that the effect of the program depends on location and age. The ESEO program reduces the risk of recidivism only for ex-inmates over the age of 27 in San Diego and Chicago, and over the age of 36 in Boston, but increases the risk of recidivism for the other ex-inmates in the treatment group.

Key words: Job search, Recidivism, Conditional treatment, Program evaluation, Duration models, Interval-censoring.

JEL Codes: C31, C34, C41, C51, J46, J64, K42

1 Introduction

Accordingly to Bloom (2006), more than 600,000 people are released from prison each year in the United States. By any standards this fact raises many concerns about the likely consequences, both social and economic, of such a massive influx of ex-convicts into society. A prominent issue nowadays is the high rates of recidivism prevalent in the country, despite the huge efforts by public authorities to make sure that the rehabilitative role of prisons works properly. In fact, recidivism rates measured by the Bureau of Justice Statistics (1994) show that two-third of released prisoners are re-arrested and one-half are re-incarcerated within three years after release. Freeman (2003) asserts that, after a decade from release, up to 80% of prisoners are re-arrested.

Most ex-offenders have low educational levels, medical problems, less work experience than non-offenders. Therefore, the prospects of ex-offenders in the job market are almost always worse than for non-offenders with comparable characteristics (Freeman 1999 and Western 2002), although this is also true before incarceration. This means that incarceration may not be the only reason for low job market performance; the characteristics that caused the conviction could have caused the bad performance in the labor market as well.

The idea that having a job diminishes the chances of recidivism is well established in the criminological literature. See for example Harer (1994), Sampson and Laub (1997) and Uggen (2000). A job provides the necessary means for survival, improves self-esteem, increases the attachment to a community and develops the sense of belonging to a group. Therefore, even though finding a job is a difficult task for ex-offenders, a policy that help these people to find a job will likely decrease the chances of recidivism.

During the 1970's and 1980's policy makers in the United States have sponsored evaluations of both in-prison job training programs, as well as post-release (community-based) employment interventions. As to the first type of programs, Visser et al. (2005) assert that these evaluations provide at best mixed results concerning the effectiveness of those interventions. Because of these mixed results, the attention has turned to community-based employment programs. However, as outlined by Bloom (2006), "there appear to have been few rigorous studies of employment-focused re-entry models". Moreover, Farrington and Welsh (2005) state that very few evaluations in criminology can be classified as rigorous, in that they are not randomized

experiments. Furthermore, it appears that in the criminological literature a randomized experiment is considered synonym for a program with randomized treatment.

During the last decades, sociologists, criminologists and, to a lesser extent, economists have been devising programs to ease the difficult transition faced by ex-offenders during the period of time between release and reintegration into society. These programs are basically of two types: post-release programs and in-prison programs. While the former type of program offers assistance after the individual has been released, the latter type of program starts helping the individual while he/she is in prison. As experience has accumulated, a fundamental goal to a complete reintegration turned out to be job placement. A good job would be necessary not only to provide the basic needs for survival in the short run but also as a key element to secure self-esteem, security and sense of integration in the society as whole.

The Life Insurance for Ex-offenders and the Transitional Aid for Ex-offenders are two early examples of employment services for ex-offenders. Both programs offered financial assistance as well as job placement services. The two programs reached similar conclusions: while financial assistance appeared to decrease the recidivism rate, job placement had little or no effect on reducing criminal activity, unless for those who succeeded in securing a job for a long time. The lack of follow-up after placement was conjectured by Milkman et al. (1985) as the main obstacle to the complete success of such programs.

The new paradigm of employment services for ex-offenders have resulted in the appearance of programs that had a strong preoccupation with the post-placement of their clients. These programs have designed follow-up strategies to overcome the major criticism of past experience. Among various programs, three deserve recognition for both being successful and having similar structures: the Comprehensive Offender Resource System (COERS) in Boston, the Safer Foundation (SF) in Chicago and Project JOVE in San Diego. Not surprisingly, the US Department of Justice saw this as an opportunity for assessing the efficacy of employment services programs that contained a follow-up component. See Phipps et al. (1999) for a review of these programs.

During the period of 1980 – 1985, the National Institute of Justice sponsored a controlled experiment to evaluate the impact of reemployment programs for recent released prisoners. This evaluation was performed by the Lazar Institute in McLean, Virginia. The three aforementioned programs,

COERS in Boston, JOVE in San Diego and SF in Chicago, were chosen to participate in the Employment Services for Ex-Offenders Program, henceforth ESEO. A total of 2045 prisoners who volunteered to participate in the program were randomly assigned to either a treatment group or a control group. Those in the first group received, besides the normal services (orientation, screening, evaluation, support services, job development seminar, and job search coaching), special services which consisted of an assignment to a follow-up specialist who provided support during the job search and the 180 days following job placement. The control group received only normal services.

Using OLS regressions, Milkman (2001) found that the effect of the ESEO special services program is negligible. However, this evaluation did not account for the conditional feature of the treatment. The timing of the treatment was completely neglected, which is a very important characteristic of the program under evaluation.

The objective of our paper is to re-evaluate the effect of the ESEO program on recidivism using duration models for job search and recidivism (the latter being defined as the time between release and the first arrest), with the conditional treatment incorporated in the latter duration model. These two durations are interval censored, though.

The notion of job search duration does not need any explanation, but the meaning of "recidivism" is not unambiguous. There are two ways to measure recidivism outcomes in the ESEO program, via count data or duration data. Detailed data on the number of arrests from date of released to the end of the program for all clients was gathered in the respective state police departments. That was the type of data used in the original evaluation made by Milkman et al. (1985), as well as in almost all literature regarding the evaluation of employment programs for ex-offenders. See Visher et al. (2005) and Seiter and Kadela (2003).

In the criminology literature three possible definitions of recidivism are considered: rearrest, reconviction and re-incarceration. It seems that rearrest has been proven to be the most reliable among the three possible measures, as reported in Beck and Shipley (1989) and Maltz (1984). The latter is what we will use as the duration of recidivism. Of course, it would be better to distinguish between rearrest with subsequent conviction and re-incarceration, and rearrest without conviction, but the ESEO data does not contain sufficient information about convictions and re-incarcerations.

As said before, it is well-documented in the criminology literature that

employed ex-offenders have a lower recidivism rate than unemployed ex-offenders, which is the very reason for the existence of employment assistance programs like the three programs mentioned above. Since during the job search period the ex-offender is unemployed and therefore has a higher risk of recidivism than after finding a job, it seems plausible that the longer the job search takes the higher the risk of recidivism and thus the shorter the time between release and rearrest.

However, one can also make a case for independence of the recidivism and job search durations for the individuals in the control group, possibly conditional on covariates, as during the job search stage the participants in the ESEO program receive assistance services such as food, transportation, and clothing, which may offset the unemployment effect. Actually, this is what we find in the current study! More precisely, the job search duration appears to be exogenous for the recidivism duration. So why bothering about the job search duration? The reason is twofold. First, in the case of the ESEO program the duration that triggers the treatment is the job search. Second, and more importantly, both durations are interval censored so that the actual durations are unobserved. In order to determine the probabilities that these durations fall in these intervals we need to integrate their joint density over the intervals involved. If the exact job search durations were observed there would be no need for a job search model, as then we may simply condition on job search duration.

Carvalho and Bierens (2007) have studied the effect of the ESEO program on recidivism using a bivariate mixed proportional Weibull hazard model for job search and recidivism with common heterogeneity, where both durations are assumed to be independent conditional on the covariates and the heterogeneity variable. The latter is assumed to be Gamma distributed, and integrated out to make the two durations dependent conditional on the covariates only. However, the problem with this approach is that then a positive relation is forced between the job search and recidivism durations. This seems unlikely. Moreover, the incorporation of treatment in the model involved is not well founded in the treatment literature.

Therefore, in the this paper we take a completely different approach, by incorporating treatment in the conditional distribution of the recidivism duration given the covariates and the job search duration. This approach is inspired by (but not exactly equal to) the approach of Abbring and Van den Berg (2003). These authors propose a bivariate mixed proportional hazard model, where one duration (job search in the ESEO case) is the timing of

the treatment and the other one is the duration of interest (recidivism in the ESEO case) that is affected by the treatment.

Note that the idea of using the duration of recidivism, defined as the time between release and rearrest, is not new. Uggen (2000) uses the same measure of recidivism. The main difference of the National Supported Work Demonstration Project (NSWDP) program analyzed by Uggen (2000) and the ESEO program is that in the case of the NSWDP program the ex-offenders in the treatment group were offered minimum-wage jobs, and the ones in the control group had to find work on their own, whereas in the ESEO program the actual treatment of the ex-offenders in the treatment group is conditional on finding a job. The selection in the treatment group and control group in the NSWDP program was done randomly, as in the ESEO case. Therefore, the design of the NSWDP program is in tune with the standard practice in medical and pharmaceutical research, whereas the ESEO program is not.

This paper makes two main contributions to the literature. First, we find that contrary to previous studies the ESEO program has an effect on recidivism but this effect is heterogeneous. In particular, the recidivism duration in the case of the COERS program in Boston is shorter than in the two other participating programs, but in either case recidivism reduces with age. Second, we make a methodological contribution by showing that with careful econometric modeling and some data mining it is possible to extract useful information from bad data.

As to the latter, the main limitations of the ESEO data set are that the job search and recidivism durations are only measured in the form of a few intervals, and the presence of substantial attrition. The interval censoring problem prevents us from taking possible unobserved heterogeneity into account in a semi-nonparametric way, because it is shown in Bierens (2008) that then semi-nonparametric identification of the unobserved heterogeneity distribution requires that at least one covariate has the whole real line as support, whereas all the observed covariates in the ESEO data set are discrete. The attrition rate is more than 50%, and in all but a few cases attrition takes place straight after release. It is therefore not possible to link attrition directly to the job search and recidivism durations via observed covariates, although it is conceivable that there is a link via common or correlated unobserved heterogeneity. However, the interval censoring problem prevents us from taking this possible unobserved heterogeneity link into account. Consequently, our only option is to conduct the duration analysis conditional on absence of attrition.

The setup of the paper is as follows. In Section 2 we discuss the ESEO program in more detail, and we describe the data set and the variables involved. Moreover, in this section we also discuss the attrition problem. In Section 3 we conduct a preliminary data analysis by nonparametrically estimating and comparing the interval probabilities for both durations without using covariates, in order to check for evidence of dependence of recidivism on job search, and evidence of a treatment effect. We find neither, and therefore we conclude that at least for the control group the job search and recidivism durations are independent, and that if there is a treatment effect then it will likely work via covariates. Section 4 deals with model specification and estimation. In Section 4.1 we briefly review the standard treatment approach and the proper notion of treatment effect, and in section 4.2 we discuss the Abbring and Van den Berg (2003) model and its relation to our model. In Section 4.3 we estimate and test an interval-censored proportional hazard model for the job search, where in first instance the integrated hazard is left unspecified. Only the location (Chicago, San Diego or Boston) seems to matter for the job search duration, and the results indicate that a Weibull baseline hazard is appropriate for job search. Similarly, in Section 4.4 we estimate and test an interval censored proportional hazard model for the recidivism duration of the control group only. It appears that none of the covariates matter and that the baseline hazard is constant, hence the distribution involved is exponential and independent of the job search duration. In Section 4.5 we extend this exponential model for recidivism to a model that incorporates treatment via covariates, depending on the job search duration, and merge it with the job search model, which are then estimated jointly by maximum likelihood. The results are presented in Section 4.6. It appears that a treatment effect is present, but its magnitude and direction depends on the location and on age. In Section 4.7 we compute the actual treatment effect, in terms of expected increase or reduction of the recidivism duration in months, on the basis of our estimations results. In Section 4.8 we compare the estimation results with the preliminary data analysis, and explain why we did not find in first instance evidence of dependence of the job search and recidivism durations and the presence of a treatment effect. In Section 5 we summarize our results and make recommendations for future setup and evaluations of employment programs for ex-offenders. Section 6 is an appendix containing the details of the model selection procedures.

2 The ESEO program and data set

2.1 The ESEO Program

In the ESEO program, after being assigned randomly to either the control or treatment group, the clients enter the intake unit, where they receive initial orientation, screening and evaluation by an intake counsellor. While still in this first phase, to secure survival up to the job search phase, the intake counsellor offers assistance services such as food, transportation, and clothing. After intake, the clients enter the second phase that will prepare them to develop job search skills. In particular, a brief job development seminar is offered which deals with issues like appropriate dress and deportment, typical job rules, goal setting, interviewing techniques, and job hunting strategy. It is assumed that the time spent in the first and second phases are negligible compared to the job search phase. The next and final phase before possible treatment is the job search assistance. This is the traditional job search assistance type of service, as described by Bloom (2006) and Heckman et al. (1999). The job search assistance in the ESEO program is offered equally to both control and treatment groups.

The actual treatment starts upon the first job placement. The people in the control group do not receive help after placement, whereas the people in the treatment group receive follow-up help. The follow-up special services consist mainly of crisis intervention, counselling and, whenever necessary, re-employment assistance, during a period of six month after the first placement. See Timrots (1985), Milkman et al. (1985) and Milkman (2001) for the details of the programs and treatment. Thus, even though the split of the sample in treatment group and control group is random, the actual treatment is the follow-up special services after finding a job and is therefore not randomized.

2.2 The ESEO data

The ESEO data set used in our analysis comes from the Inter-University Consortium for Political and Social Research, henceforth ICPSR, University of Michigan, under the study number 8619. This data set consist of 2045 individuals¹ who participated in one of the three programs: 511 in Boston,

¹In the official codebook that ICPSR made available it is written on page 11 that there are 2045 observations. But, on page 20 (FILE STRUCTURE) of the same document it is stated that there are 1215 observations, as was also noticed by Milkman et al. (1985) and

934 in Chicago and 600 in San Diego. However, the ICPSR only made available 1074 usable observations: 325 in Boston, 489 in Chicago and 260 in San Diego.

A first important empirical issue is related to the characterization of the population being sampled. Unless very special assumptions are made, the validity of our findings can not be extrapolated beyond the population involved. In order to be eligible to participate in the ESEO program an individual must have the following background:

1. Participants voluntarily accepted program services;
2. Participants had been incarcerated at an adult Federal, State, or local correctional facility for at least 3 months and had been released within 6 months of program participation;
3. Participants exhibited a pattern of income-producing offenses.

From the eligibility criteria it is clear that our population is a special, indeed a very special, subset of the population of ex-offenders. Also, since participation is voluntary and there is no information on non-participants (those who did not choose to participate even though they fulfilled requirements 2 and 3), it is not possible to assess the potential bias induced by this selection scheme.

Given the initial sample, the individuals were randomly assigned to either the treatment or control group. The control group received the standard services and the treatment group received, in addition to that, emotional support and advocacy during the follow-up period of 180 days after placement. Two durations are of great importance, time spent searching a job and recidivism time. These two variables are interval-censored, however; they are only observed in the form of intervals. Moreover, there is a substantial number of individuals in the sample who do not show up for the common part of the program, i.e., the assistance with the job search. Thus, the endogenous variables are:

A : Indicator for attrition. $A = 1$ means the individual is either a “no show” or a “drop-out”, $A = 0$ otherwise;

T_s ($s = search$) is the duration of the job search, i.e., the time between the

Timrots (1985). Thus, the data set contains only 1215 observations. After removing some missing values, we end up with 1074 observations.

date of release and the date of placement in the first job after release;
 T_c ($c = crime$) is the recidivism time, i.e., the time between the date of release and the date of the first arrest after release.

However, as will be argued in the next subsection, the attrition indicator is actually an exogenous variable.

The point of departure for the choice of the covariates of the recidivism duration model is Schmidt and Witte (1988): age at release, time served for the sample sentence, sex, education, marital status, race, drug use, supervision status, and dummies that characterize the type of recidivism. However, we have also paid close attention to the criminological literature on recidivism, for instance Gendreau et al. (1996).

The literature on unemployment (and job search) duration has been refined since the 70's. Nowadays, it has a status of a complete theory of unemployment, as it appears in Pissarides (2000). Its empirical contents has been developed since the late 70's and this first wave of empiricism is characterized for being concerned with "reduced form" type models. A good account of this first phase can be found in Devine and Kiefer (1991). A final wave of literature is characterized by advocating a "structural" approach. An updated account of that appears in Van den Berg (1999, 2000). There has been also studies close to ours that try to measure the effect of programs in a context of a model of unemployment and job search duration. See for instance, Abbring and Van den Berg (2003), Abbring et al. (2005), Eberwein et al. (1997) and Van den Berg et al. (2004).

In view of the recidivism and job search duration literature mentioned above, and the limitations of imposed by the ESEO data set, the following set of covariates has been singled out, next to the group assignment indicator:

$G = 1$ if selected in the treatment group, $= 0$ if selected in the control group;
DRUGS: Indicator for drugs use during the last 5 years;
WHITE: Race indicator;
MALE: Gender indicator;
EDUC_L = 1 if years of schooling ≤ 8 , $= 0$ otherwise;
EDUC_H = 1 if years of schooling > 12 , $= 0$ otherwise;
AGE: Age in years;
AGE1ARR: Age of first arrest, in years;
SANDIEGO: Indicator for San Diego;
CHICAGO: Indicator for Chicago.

In the econometric model to be developed the job search and recidivism durations T_s and T_c , respectively, are in first instance modeled as continuous latent variables, where treatment kicks in for people in the treatment group if $T_c > T_s$. The treatment is effective in reducing recidivism if, given the same job search duration T_s and the covariates, the recidivism duration T_c for the treatment group is larger than for the control group. Note that, since $T_c < T_s$ is also possible, some people in the treatment group may end up not receiving the treatment at all.

Finding a job may postpone recidivism, but will likely not eliminate it. As Freeman (2003) asserts, "Getting and ex-offender a job does not mean that they will eschew a criminal opportunity if it arises". Thus, in our model the probability that $T_c = \infty$ is assumed to be zero. The use of continuous latent variables advocated by us also appears to conform better with what is noted by Fagan and Freeman (1999) and Freeman (2003), namely that the boundary between crime and work is thin and very diffuse for many young men just after release. The two clocks, i.e., time to recidivism and time to get a job, are about to be pushed down every day for these young men.

Although the latent variables T_s and T_c will be modeled as a joint continuous distribution, conditional on covariates, we only observe them in the form of intervals: $T_s \in (a, b]$, $T_c \in (p, q]$, where $(a, b], (p, q] \in \{(0, 1], (1, 6], (6, 12], (12, \infty)\}$. The unit of measurement for these durations is months. These events are only observed if $A = 0$. Given the joint conditional distribution of T_s and T_c , the probabilities of these discrete events can be computed. The latter are then used to form a likelihood function.

As said before, the ex-convicts in the sample have been randomly assigned to either a control group ($G = 0$) or a treatment group ($G = 1$). Both groups get standard assistance with the job search. Treatment consists of extra help after finding a job, and is therefore conditional on the job search duration T_s . The purpose of this study is to determine whether this treatment has an effect on the risk of recidivism, given the covariates listed above.

2.3 Attrition

As to attrition, one could argue that "no show" and "drop-out" deserve separate treatment, as pointed out by Bloom (1984) and Heckman et al. (1998). However, our data set does not allow for this disaggregation.

With a few exceptions, attrition occurs straight away after release. Therefore it will be assumed that conditional on the covariates listed above the attrition indicator A is an exogenous variable for T_s and T_c because in almost all cases the attrition decision $A = 1$ is made before T_s or T_c are realized. Since our focus is on the effect of the treatment on the duration T_c , conditional on T_s and observed covariates, and T_s and T_c are not observed if $A = 1$, our analysis has to be conducted conditional on absence of attrition ($A = 0$) by using the subsample of individuals for which $A = 0$.

If data on recidivism and job search for the ex-offenders quitting the ESEO program were available, we would be able to make attrition endogenous, but we don't have this data. Another way to link the attrition decision to the durations T_s and T_c is via unobserved heterogeneity.² However, as said before, the interval censoring of T_s and T_c prevents us from going this route. Admittedly, conditioning on absence of attrition makes it difficult to extrapolate our findings to other intervention programs, but there is no alternative for this problem.

Thus, the question what the determinants of attrition are is a side issue for the present study. Nevertheless, this question is of intrinsic interest and will therefore be addressed in a separate Appendix, Bierens and Carvalho (2010), via Logit and Probit analyses. As to the results, the attrition rates of the initial 1074 participants in the ESEO program depend on gender, location and group assignment. It appears that males have a higher attrition rate than females, but only if selected in the control group. Moreover, the attrition rates in Chicago are significantly higher for the control group than for the treatment group, and the same applies to San Diego. For the treatment group the attrition rates in Chicago and San Diego are about the same, but for the control group the attrition rate in Chicago is higher than in San Diego. Furthermore, the attrition rates in Chicago and San Diego are much higher than in Boston, in particular for the control group.

3 Preliminary data analysis

The durations T_c and T_s are only observed in the form of interval indicators, for the intervals $(0, 1]$, $(1, 6]$, $(6, 12]$ and $(12, \infty)$. Table 1 presents the number of observations in each interval and combination of intervals for the treatment group ($G = 1$) and the control group ($G = 0$).

²As suggested by a referee.

Table 1. Observations per interval ($A = 0$)**Treatment group ($G = 1$)**

$T_s \setminus T_c$	(0, 1]	(1, 6]	(6, 12]	(12, ∞)	Total
(0, 1]	9	39	32	105	185
(1, 6]	5	35	30	81	151
(6, 12]	1	4	5	18	28
(12, ∞)	0	1	1	3	5
Total	15	79	68	207	369

Control group ($G = 0$)

$T_s \setminus T_c$	(0, 1]	(1, 6]	(6, 12]	(12, ∞)	Total
(0, 1]	3	17	11	47	78
(1, 6]	1	9	9	31	50
(6, 12]	0	3	1	2	6
(12, ∞)	0	0	0	0	0
Total	4	29	21	80	134

Observe from the lower triangular part in the panels in Table 1 that the number of individuals for which the job search duration exceeds the recidivism duration is relatively small: 12 out of 369 in the treatment group and 4 out of 134 in the control group. Apparently, these individuals have been rearrested but not incarcerated (for long), as otherwise the job search would have stopped. The number of ex-offenders for which the job search and recidivism durations fall in the same interval is 52 out of 369 in the treatment group and 13 out of 134 in the control group. For these individuals it is unknown whether $T_c > T_s$ or $T_c \leq T_s$, hence it is unknown whether these individuals have been incarcerated or not.

By dividing the entries in rows 1-4 in the two panels of Table 1 by the corresponding row totals we get nonparametric estimates of the conditional probabilities $\Pr [T_c \in (p, q] | T_s \in (a, b)]$, and in the last rows the unconditional probabilities $\Pr [T_c \in (p, q)]$. These estimated conditional probabilities, times 100%, are presented in Table 2.

Comparing the entries in rows 1-4 of the panels in Table 2 with the corresponding entries in row 5, it appears that for both groups all but one of the estimates of $\Pr [T_c \in (p, q] | T_s \in (a, b)]$ are close to the estimates of $\Pr [T_c \in (p, q)]$. The exception is the estimate of $\Pr [T_c > 12 | T_s \in (6, 12]]$ for the control group, but this estimate is based on only two observations.

Table 2. Estimated conditional probabilities $\Pr [T_c \in (p, q] | T_s \in (a, b)] \times 100\%$ **Treatment group** ($G = 1$)

$T_s \setminus T_c$	(0, 1]	(1, 6]	(6, 12]	(12, ∞)
(0, 1]	5	21	17	57
(1, 6]	3	23	20	54
(6, 12]	4	14	18	64
(12, ∞)	0	20	20	60
(0, ∞)	4	21.5	18.5	56

Control group ($G = 0$) (? = undefined)

$T_s \setminus T_c$	(0, 1]	(1, 6]	(6, 12]	(12, ∞)
(0, 1]	4	22	14	60
(1, 6]	2	18	18	62
(6, 12]	0	50	17	33
(12, ∞)	?	?	?	?
(0, ∞)	3	21.6	15.7	59.7

For our analysis only the probabilities $\Pr [T_c \in (p, q] | T_s \in (a, b)]$ for $p \geq b$ are relevant, as in the treatment group the individuals involved have received the actual treatment. In the Appendix we present the results of tests of the null hypothesis that $\Pr [T_c \in (p, q] | T_s \in (a, b)] = \Pr [T_c \in (p, q)]$ for $p \geq b$. In all cases the null hypothesis is not rejected at any conventional significance level! Therefore, it seems that the events $T_c \in (p, q]$ and $T_s \in (a, b]$ for $b \leq p$ are independent. However, if there is a treatment effect one would expect that in the case $G = 1$ these events are dependent.

On the other hand, even if the events $T_c \in (p, q]$ and $T_s \in (a, b]$ for $b \leq p$ are independent, it is possible that conditional on covariates they are dependent. To see this, observe that

$$\begin{aligned}
\Pr [T_c \in (p, q] \text{ and } T_s \in (a, b)] &= \int_a^b \Pr [T_c \in (p, q] | T_s = t_s] dP [T_s \leq t_s] \\
&= \Pr [T_c \in (p, q)] \Pr [T_s \in (a, b)] \\
&+ \int_a^b (\Pr [T_c \in (p, q] | T_s = t_s] - \Pr [T_c \in (p, q)]) dP [T_s \leq t_s] \\
&= \Pr [T_c \in (p, q)] \Pr [T_s \in (a, b)] \\
&+ \int_a^b E \{ \Pr [T_c \in (p, q] | T_s = t_s, X] - \Pr [T_c \in (p, q) | X] \} dP [T_s \leq t_s].
\end{aligned}$$

It is possible that the latter integral is zero even if

$$\Pr [T_c \leq t|T_s, X] \neq \Pr [T_c \leq t|X]. \quad (1)$$

Moreover, if this integral is small then the events $T_c \in (p, q]$ and $T_s \in (a, b]$ are approximately independent. Thus, if the dependence of $\Pr [T_c \leq t|T_s, X]$ on T_s is substantially reduced after X is integrated out, then the inequality (1) may no longer be detectable by comparing $\Pr(T_c \in (p, q], T_s \in (a, b])$ with $\Pr(T_c \in (p, q]) \cdot \Pr(T_s \in (a, b])$. Therefore, despite the results of the above preliminary data analysis a conditional treatment effect may still be possible.

A slightly different way to look for possible treatment effects is to compare the conditional probabilities $\Pr[T_c \in (p, q]|T_s \leq p]$ for the treatment group ($G = 1$) and the control group ($G = 0$) because again the individuals involved in the treatment group have received the actual treatment. Thus, the differences between these conditional probabilities may indicate possible treatment effects. The conditional probabilities involved can be straightforwardly estimated from the results in Table 1, as reported in Table 3. The standard errors of the differences have been calculated under the null hypothesis that the conditional probabilities involved are equal. Under this null hypothesis the ratio of the difference and its standard error is approximately standard normally distributed, hence none of these differences are significantly different from zero at any conventional significance level.

Table 3. Estimated probabilities

$\Pr[T_c \in (p, q]|T_s \leq p] \times 100\%$

$(p, q]$	$(1, 6]$	$(6, 12]$	$(12, \infty]$
Treatment group	21.0	18.5	56.0
Control group	21.8	15.6	59.7
Differences	-0.8	2.9	-3.7
Standard errors	5.6	3.8	5.0

To check whether the marginal distribution of job search depends on the group assignment G , divide the last two column entries in Table 1 by their column totals to get nonparametric estimates of $\Pr(T_s \in (a, b])$, and compare the differences, similar to Table 3. See Table 4. Again, each of these differences are not significantly different from zero at any conventional significance level.

Table 4. Estimated probabilities $\Pr(T_s \in (a, b]) \times 100\%$

$(a, b]$	$(0, 1]$	$(1, 6]$	$(6, 12]$
Treatment group	50.1	40.9	7.6
Control group	58.2	37.3	4.5
Differences	-8.1	3.6	3.1
Standard errors	5.0	4.9	2.2

4 Modeling strategy and empirical results

4.1 Average and conditional treatment effects

The problem with determining the effect of treatment is that the outcomes of interest, say Y_1 for the treated ($T = 1$) and Y_0 for the non-treated ($T = 0$), are never observed for the same individual, hence the actual treatment effect $Y_1 - Y_0$ is unobserved. What is observed is $Y = T.Y_1 + (1 - T)Y_0$ and T itself. Nevertheless, given that Y_1 and Y_0 are independent of T , denoted by

$$(Y_1, Y_0) \perp T, \quad (2)$$

and $p = \Pr[T = 1]$ is known (by random assignment in the treatment group), the average treatment effect $ATE = E[Y_1 - Y_0]$ can be calculated as

$$\begin{aligned} ATE &= E[T.Y_1/p] - E[(1 - T)Y_0/(1 - p)] \\ &= E((T/p)E[Y_1|T]) - E((1 - T)/(1 - p)E[Y_0|T]) \\ &= E[Y_1].E[T/p] - E[Y_0]E[(1 - T)/(1 - p)] \\ &= E[Y_1 - Y_0]. \end{aligned} \quad (3)$$

The independence condition (2) may be too strong, but it is often not unreasonable to assume that conditional on a vector X of sufficiently many covariates, Y_1 and Y_0 are independent of T , denoted by

$$(Y_1, Y_0) \perp T \mid X. \quad (4)$$

This is called the unconfoundedness hypothesis in the treatment literature. See for example Imbens (2004) and the references therein. Denoting

$$p(X) = \Pr[T = 1|X],$$

which is known as the propensity score [see for example Rosenbaum and Rubin (1983)], and assuming that $\Pr[0 < p(X) < 1] = 1$, it follows then similar to (3) that

$$E \left[\frac{T \cdot Y}{p(X)} - \frac{(1-T)Y}{1-p(X)} \middle| X \right] = E[Y_1 - Y_0 | X]. \quad (5)$$

This is the conditional treatment effect. Taking the expectation of (5) yields the *ATE*.

In our case the variable of interest is the recidivism duration T_c , and the propensity score is

$$p(X) = \Pr[T_c > T_s, G = 1 | X].$$

However, the unconfoundedness hypothesis (4) does not hold in our case because the actual treatment depends on T_c . Moreover, T_c and T_s are interval censored, so that in order to determine the conditional treatment effect we need to determine first the joint conditional distribution of T_c and T_s given the covariates, with treatment build in. To solve these problems, we will use the following approach.

4.2 The Abbring-Van den Berg model

Abbring and Van den Berg (2003) consider the problem of identification of treatment effects in a bivariate mixed proportional hazard model, where one duration, S , is the timing of an intervention on another duration Y . In their Model 1a they specify the hazard functions of these duration as

$$\theta_S(t|X, V) = \lambda_S(t)\varphi_S(X)V_S \quad (6)$$

for the duration S and

$$\theta_Y(t|S, X, V) = \begin{cases} \lambda_Y(t)\varphi_Y(X)V_Y & \text{if } t \leq S \\ \lambda_Y(t)\varphi_Y(X)\delta(t|S, X)V_Y & \text{if } t > S \end{cases} \quad (7)$$

for the duration Y , where X is a vector of covariates with support \mathbb{X} , $V = (V_S, V_Y)' \in (0, \infty) \times (0, \infty)$ is a vector of possibly mutually dependent unobserved heterogeneity variables which are independent of X , the $\lambda_i(t)$ and $\varphi_i(X)$, $i = S, Y$, are the baseline and systematic hazards, respectively,

and $\delta(t|S, X)$ represents the conditional treatment effect. Implicit in this specification is the "no anticipation" condition³

$$\theta_S(t|s_1, X, V) = \theta_S(t|s_2, X, V) \text{ if } t \leq \min(s_1, s_2)$$

The focus in Abbring and Van den Berg (2003) is on nonparametric identification of the baseline and systematic hazards $\lambda_i(t)$ and $\varphi_i(X)$, $i = S, Y$, the treatment function $\delta(\cdot)$ and the joint distribution $G(v)$ of V , rather than on estimation. In particular, they show that this model is non-parametrically identified if

$$\{(\varphi_S(x), \varphi_Y(x))'; x \in \mathbb{X}\} \text{ contains an open set in } \mathbb{R}^2, \quad (8)$$

and $E[V_S] < \infty$, $E[V_Y] < \infty$.

In the case of the ESEO program, S is the job search duration T_s and Y is the recidivism duration T_c . Since T_s and T_c are interval censored, and the support \mathbb{X} of the covariates is countable, which violates condition (8), we cannot take semi-nonparametric unobserved heterogeneity into account. See Bierens (2008, Section 9). It may be possible that the model is identified for certain parametric specifications of $G(v)$, for example, let $\ln(V_S)$ and $\ln(V_Y)$ be jointly normally distributed, but due to the interval censoring of T_c and T_s that will make our model unduly complicated and possibly more prone to misspecification. Therefore, we will set $V_S = V_Y = 1$ in (6) and (7).

We specify the proportional hazard of job search T_s similar to (6), with $V_S = 1$, and we specify the conditional hazard of the recidivism duration T_c for the control group ($G = 0$) as well as for the treatment group ($G = 1$) in the case $t \leq T_s$ similar to (7) for the case $V_Y = 1$, $t < S$. Moreover, we specify the systematic hazards parametrically in the usual way, as the $\exp(\cdot)$ of linear combinations of the covariates. Thus, in our notation,

$$\theta_s(t|X) = \exp(\beta'_s X) \lambda_s(t) \quad (9)$$

is the conditional hazard of the job search duration T_s , and

$$\theta_c(t|T_s, X, G) = \begin{cases} \exp(\beta'_c X) \lambda_c(t) & \text{if } t \leq T_s \\ \exp(\beta'_c X) \delta(t|T_s, X, G) \lambda_c(t) & \text{if } t > T_s \end{cases} \quad (10)$$

is the conditional hazard of the recidivism duration T_c , where $\delta(t|T_s, X, G) = 1$ for the control group $G = 0$, and $\delta(t|T_s, X, G)$ is to be determined for the treatment group $G = 1$.

³Abbring and Van den Berg (2003, Assumption 1).

Implicit in this specification is that for the control group, T_s and T_c are independent conditional on X , due to the absence of unobserved heterogeneity. Thus, in the case $G = 0$ the joint conditional survival function of T_c and T_s given X is assumed to be

$$\begin{aligned} S(t_c, t_s|X) &= \Pr [T_c > t_c, T_s > t_s|X] \\ &= \exp((\beta_c + \beta_s)'X) \int_0^{\max(0, t_c)} \lambda_c(\tau_c) d\tau_c \int_0^{\max(0, t_s)} \lambda_s(\tau_s) d\tau_s. \end{aligned} \quad (11)$$

Note that the preliminary data analysis suggests that $\Pr[T_c > t_c, T_s > t_s] = \Pr[T_c > t_c] \cdot \Pr[T_s > t_s]$, which is the case if

$$E[\exp((\beta_c + \beta_s)'X)] = E[\exp(\beta_c'X)] \cdot E[\exp(\beta_s'X)].$$

A sufficient (but not necessary) condition for the latter is that one of the parameter vectors β_s and β_c is a zero vector. As we will see in the next two subsections, the null hypothesis $\beta_c = 0$ is rejected but the hypothesis $\beta_s = 0$ can not be rejected.

Given the proportional hazard structure of the model, the systematic and baseline hazards will be specified in a data-driven way. Since the duration T_s and T_c are interval-censored, there is in first instance no need to specify the baseline hazards parametrically. This enables us to let the data determine how the baseline hazards look like, and which components of β_s and β_c can be set to zero. Only the treatment effect factor $\delta(t|T_s, X, G)$ has to be specified parametrically.

4.3 Job search

Given the hazard (9) the conditional survival function of T_s is

$$S_s(t|X) = \exp(-\exp(\beta_s'X_s) \Lambda_s(t)),$$

where $\Lambda_s(t) = \int_0^t \lambda_s(\tau) d\tau$ is the integrated hazard and X_s is a subvector of covariates relevant for job search. In first instance we have selected for X_s all available covariates: $X_s = X$. Then

$$\begin{aligned} \Pr [T_s \in (a, b]|X] &= S_s(a|X) - S_s(b|X) \\ &= \exp(-\exp(\beta_s'X) \Lambda_s(a)) - \exp(-\exp(\beta_s'X) \Lambda_s(b)). \end{aligned}$$

Since we can only estimate $\Lambda_s(t)$ for $t \in \{1, 6, 12\}$, we may without loss of generality assume that $\Lambda_s(t)$ is piecewise linear:

$$\begin{aligned}\Lambda_s(t|\alpha_s) &= \sum_{k=1}^{i-1} \alpha_k (b_k - b_{k-1}) + \alpha_i (t - b_{i-1}) \text{ for } t \in (b_{i-1}, b_i], \quad (12) \\ b_0 &= 0, b_1 = 1, b_2 = 6, b_3 = 12 \\ \alpha_i &> 0 \text{ for } i = 1, 2, 3, \alpha_s = (\alpha_1, \alpha_2, \alpha_3)'\end{aligned}$$

Note that $\Lambda_s(t|\alpha_s)$ is homogenous of degree one in α_s : $\Lambda_s(t|c.\alpha_s) = c.\Lambda_s(t|\alpha_s)$. Therefore, we cannot allow a constant 1 in X .

The parameter vectors β_s and α_s have been estimated by maximum likelihood. Then we conduct a series of Wald and likelihood ratio tests to determine the subvector X_s of covariates that are relevant for the job search duration. Moreover, on the basis of the estimation results for the piecewise linear integrated baseline hazard (12) we deduct the functional form of the underlying smooth baseline hazard $\lambda_s(t)$. The details of this specification analysis can be found in the Appendix.

We find that only the location dummy variables matter for job search, so that

$$X_s = (CHICAGO, SANDIEGO)'$$

Moreover, we cannot reject the null hypothesis that the baseline hazard is of the Weibull type: $\lambda_s(t) = \alpha_{1,s}\alpha_{2,s}t^{\alpha_{2,s}-1}$, where $\alpha_{1,s}$ plays the role of scale parameter. Thus, the survival function now takes the form

$$\begin{aligned}S_s(t|X) &= \exp(-\exp(\beta'_s X_s) \alpha_{1,s} t^{\alpha_{2,s}}) \\ &= \exp(-\exp(\beta'_s X_s + \ln(\alpha_{1,s})) t^{\alpha_{2,s}}).\end{aligned}\quad (13)$$

The estimation results for this model are presented in Table 5.

Table 5. Job search

Covariates	Estimates	t-val.
<i>CHICAGO</i>	-1.228695	-7.988
<i>SANDIEGO</i>	-0.379488	-2.745
Parameters		
$\alpha_{1,s}$	1.197083	10.880
$\alpha_{2,s}$	0.884122	15.957
LogL, n	-422.063	503

where here and in the sequel "LogL" denotes the log-likelihood value and n is the sample size.

Note that the p-value of the two-sided test of the null hypothesis $\alpha_{2,s} = 1$ is 0.03649, hence the Weibull baseline hazard involved is decreasing in t at the 5% significance level.

The results in Table 5 are only final with respect to the model specification. The coefficients involved will be re-estimated jointly with those of the recidivism model specified below. At that point we will interpret the results.

4.4 Recidivism of the control group

Similar to the job search case, the conditional survival function of T_c for the control group will be modeled as a proportional hazard model:

$$S_c(t|X) = \exp(-\exp(\beta'_c X) \Lambda_c(t)),$$

where $\Lambda_c(t)$ is the integrated baseline hazard. Again, we may without loss of generality assume that $\Lambda_c(t)$ is piecewise linear, as in (12).

We have followed the same specification strategy as for job search. The details can be found in the Appendix. Surprisingly, we find that none of the covariates matter for recidivism, and that the baseline hazard is constant. Thus, the distribution of T_c for the control group is exponential, so that the survival function involved takes the form

$$S_c(t|X) = \exp(-\alpha_c t). \tag{14}$$

The maximum likelihood estimation result for α_c is presented in Table 6.

Table 6. Recidivism ($G = 0$)

Parameter	Estimate	t-val.
α_c	0.043681	7.396
LogL, n	-139.357	134

Note that this result implies that $E[T_c|G = 0] = 1/\alpha_c \approx 23$ months, so that approximately,

$$\begin{aligned} \Pr(T_c \in (0, 1]|G = 0) &\approx 0.04274 \\ \Pr(T_c \in (1, 6]|G = 0) &\approx 0.18781 \\ \Pr(T_c \in (6, 12]|G = 0) &\approx 0.17740 \\ \Pr(T_c > 12|G = 0) &\approx 0.59205 \end{aligned}$$

4.5 Incorporating conditional treatment

For the treatment group ($G = 1$), treatment is only received if $T_c > T_s$. Therefore we will assume that if $T_c \leq T_s$ the distribution of T_c is the same as for the control group:

$$\Pr [T_c \leq t | T_s, X, G = 1] = 1 - \exp(-\alpha_c \cdot t) \text{ if } t \leq T_s.$$

See (14). This is the "no anticipation" condition in Abbring and Van den Berg (2003, Assumption 1). Admittedly, in view of the effect of group assignment on attrition this may be a strong assumption. However, it follows from Table 2 that the number of individuals in the treatment group for which $T_c \leq T_s$ is relatively small, hence if this assumption is incorrect its impact on the results will be minor.

Recall from the results of the preliminary data analysis that if there is a treatment effect then it will likely work via the covariates. Therefore, let

$$\begin{aligned} \Pr [T_c \leq t | T_s, X_c, G = 1] \\ = 1 - \exp(-\alpha_c \cdot T_s) \exp(-\alpha_c \cdot \exp(\beta'_c X_c) \cdot (t - T_s)) \text{ if } t > T_s, \end{aligned}$$

where X_c is the vector of covariates involved, which now **also includes** 1 for the constant term. Thus, the conditional survival function of T_c given T_s , X_c , and G is specified as

$$\begin{aligned} S_c(t | T_s, X_c, G) = \Pr [T_c > t | T_s, X_c, G] = I(t \leq T_s) \exp(-\alpha_c \cdot t) \quad (15) \\ + I(t > T_s) \exp(-\alpha_c \cdot T_s) \cdot \exp(-\alpha_c \cdot \exp(G \cdot \beta'_c X_c) \cdot (t - T_s)), \end{aligned}$$

where $I(\cdot)$ is the indicator function. The reason for this particular specification is to preserve the continuity of $S_c(t | T_s, X_c, G)$ in a parsimonious way.

Note that the corresponding conditional hazard takes the form (10):

$$\begin{aligned} \theta_c(t | T_s, X, G) &= \frac{-\partial S_c(t | T_s, X_c, G) / \partial t}{S_c(t | T_s, X_c, G)} \\ &= \alpha_c (1 + I(t > T_s) (\exp(G \cdot \beta'_c X_c) - 1)), \end{aligned}$$

where the systematic hazard $\exp(\beta'_c X_c)$ is now equal to 1, the baseline hazard $\lambda_c(t)$ is equal to the constant α_c , and

$$\delta(t | T_s, X, G) = 1 + I(t > T_s) (\exp(G \cdot \beta'_c X_c) - 1).$$

Next, rewrite the survival function of T_s as

$$S_s(t|X) = \exp(-\exp(\beta'_s X_s) t^{\alpha_s}), \quad (16)$$

where now $X_s = (\text{CHICAGO}, \text{SAN DIEGO}, 1)'$ and $\alpha_s = \alpha_{2,s}$. See Table 5. Then it follows from (15) and (16) that for $0 \leq a < b \leq p < q$,

$$\begin{aligned} \Pr [T_c \in (p, q], T_s \in (a, b]|X, G] &= \int_a^b S_c(t|t_s, X_c, G) d(-S_s(t_s|X_s)) \\ &= - \int_a^b \exp(-\alpha_c \cdot (1 - \exp(G \cdot \beta'_c X_c)) t_s) d(\exp(-\exp(\beta'_s X_s) t_s^{\alpha_s})) \\ &\quad \times (\exp(-\alpha_c \cdot \exp(G \cdot \beta'_c X_c) \cdot p) - \exp(-\alpha_c \cdot \exp(G \cdot \beta'_c X_c) \cdot q)) \\ &= \int_{S_s(b|X_s)}^{S_s(a|X_s)} \exp[-\alpha_c \cdot \exp(-\alpha_s^{-1} \beta'_s X_s) \\ &\quad \times (1 - \exp(G \cdot \beta'_c X_c)) (\ln(1/u))^{1/\alpha_s}] du \\ &\quad \times (\exp(-\alpha_c \cdot \exp(G \cdot \beta'_c X_c) \cdot p) - \exp(-\alpha_c \cdot \exp(G \cdot \beta'_c X_c) \cdot q)). \end{aligned}$$

The parameters involved can now be (re-)estimated by maximum likelihood. In particular, denoting

$$\begin{aligned} Y = 1 &\text{ if } T_s \in (0, 1] \quad \text{and } T_c \in (1, 6] \\ Y = 2 &\text{ if } T_s \in (0, 1] \quad \text{and } T_c \in (6, 12] \\ Y = 3 &\text{ if } T_s \in (0, 1] \quad \text{and } T_c > 12 \\ Y = 4 &\text{ if } T_s \in (1, 6] \quad \text{and } T_c \in (6, 12] \\ Y = 5 &\text{ if } T_s \in (1, 6] \quad \text{and } T_c > 12 \\ Y = 6 &\text{ if } T_s \in (6, 12] \quad \text{and } T_c > 12 \\ Y = 0 &\text{ if otherwise} \end{aligned}$$

and

$$\begin{aligned} \rho(i, X, G|\theta_0) &= \Pr [Y = i|X, G] \text{ for } i = 1, \dots, 6, \\ \rho(0, X, G|\theta_0) &= 1 - \sum_{i=1}^6 \rho(i, X, G|\theta_0) = \Pr [Y = 0|X, G], \end{aligned}$$

with $\theta_0 = (\alpha_s, \beta'_s, \alpha_c, \beta'_c)'$ the true parameter vector, the log-likelihood involved takes the form $\sum_{j=1}^n \ln \rho(Y_j, X_j, G_j|\theta)$, where $\{(Y_j, X_j, G_j)\}_{j=1}^n$ is the

sample. In this way there is no need to compute $\Pr [T_c \in (p, q], T_s \in (a, b]|X, G]$ for $p = a, q = b$ and $q < a$.

Note that if there is a treatment effect then the effect is positive, in the sense that treatment reduces the risk of recidivism, if for $t > T_s$,

$$\Pr[T_c > t|T_s, X_c, G = 1] > \Pr[T_c > t|T_s, X_c, G = 0],$$

which is the case if $\beta'_c X_c < 0$.

4.6 Joint maximum likelihood results

In first instance we have chosen for X_c in (15) the vector of all available covariates, including 1 for the constant term. Again, we have conducted a series of Wald and likelihood ratio test to remove insignificant covariates. See the Appendix for the details. The result is that only two covariates matter for treatment: Age and the location dummy Boston:

Table 7. Job search, recidivism and treatment effects

Job search			Recidivism		
Parameter	Estimate	t-val.	Parameter	Estimate	t-val.
α_s	0.875049	12.320	α_c	0.041905	7.664
Covariates	Estimates	t-val.	Covariates	Estimates	t-val.
<i>CHICAGO</i>	-1.225155	-7.440	<i>BOSTON</i>	0.425046	2.304
<i>SANDIEGO</i>	-0.324369	-2.141	<i>AGE</i>	-0.045503	-2.720
1	0.184070	1.868	1	1.222241	2.536
LogL, n	-847.358	503			

Note that the estimation results for job search are very close to the corresponding estimates in Table 5, as expected. Moreover, observe that the estimate of α_c in Table 7 is close to the estimate of α_c in Table 6. These comparisons indicate that the model is not too far from to the truth, as the Hausman (1978) model specification test is based on similar comparisons. However, conducting a formal Hausman (1978) test is beyond our scope because the preliminary estimation results are based on a partially specified model, which violates one of the key conditions in Hausman (1978).⁴

⁴Consequently, in order to conduct a formal Hausman test in our case one has to extend Hausman's (1978) paper to the case of partially specified models, which is possible but to the best of our knowledge has not yet been done.

The significant negative signs of the location dummies in the first panel of Table 7 indicate that the job search takes longer in Chicago and San Diego than in Boston, and in Chicago longer than in San Diego. Thus, it seems that the ex-offenders in Boston receive more or better help with the job search than in the other two cities. Other possible reasons for these effects are differences in labor market conditions, efficiency of programs, attitudes of employers regarding ex-convicts, and policies regarding the release of criminal records information, to mention a few. However, we do not have enough data information to pinpoint the reasons for the differences in job search durations in these three locations.

Recall that for the control group the recidivism duration has an exponential distribution which does not depend on covariates. The two covariates in the second panel of Table 7 are therefore related to the effect of treatment for the treatment group only. How to interpret these results will be discussed in the next subsection.

4.7 Treatment effects

It is straightforward to verify from (15) that

$$\begin{aligned} & \Pr[T_c \leq t | T_c > T_s, T_s, X_c, G] \\ &= 1 - \exp(-\alpha_c \cdot \exp(G \cdot \beta'_c X_c) \cdot (t - T_s)) I(t > T_s) \end{aligned}$$

hence

$$E[T_c | T_c > T_s, T_s, X_c, G] = T_s + \alpha_c^{-1} \exp(-G \cdot \beta'_c X_c)$$

and thus

$$\begin{aligned} & E[T_c | T_c > T_s, T_s, X_c, G = 1] - E[T_c | T_c > T_s, T_s, X_c, G = 0] \quad (17) \\ &= \alpha_c^{-1} (\exp(-\beta'_c X_c) - 1). \end{aligned}$$

This expression may be interpreted as (a version of) the conditional treatment effect. Thus, the treatment has a positive effect, in the sense that treatment increases the expected time between release and rearrest, if $\beta'_c X_c < 0$, regardless the job search duration.

It follows from the results in Table 7 that

$$\widehat{\beta}'_c X_c = 1.222241 + 0.425046.BOSTON - 0.045503.AGE. \quad (18)$$

As to the "Boston" effect, (18) is larger for Boston than for the other two locations, so that *ceteris paribus* the conditional treatment effect on recidivism in Boston is less than in Chicago and San Diego. This difference increases with age. Moreover, it follows from (18) that the treatment reduces the risk of recidivism in Chicago and San Diego if

$$AGE > 1.222241/0.045503 \approx 27, \tag{19}$$

and in Boston if

$$AGE > 2.180753/0.045503 \approx 36.$$

Note that Uggen (2000), using a different data set and methodology, concludes that work programs reduce recidivism only for ex-convicts over the age of 26, which corresponds to our finding (19). Thus, in general, treatment only reduces the risk of recidivism for older ex-inmates, and increases the risk of recidivism for younger ex-inmates! With how much is illustrated in Figure 1.

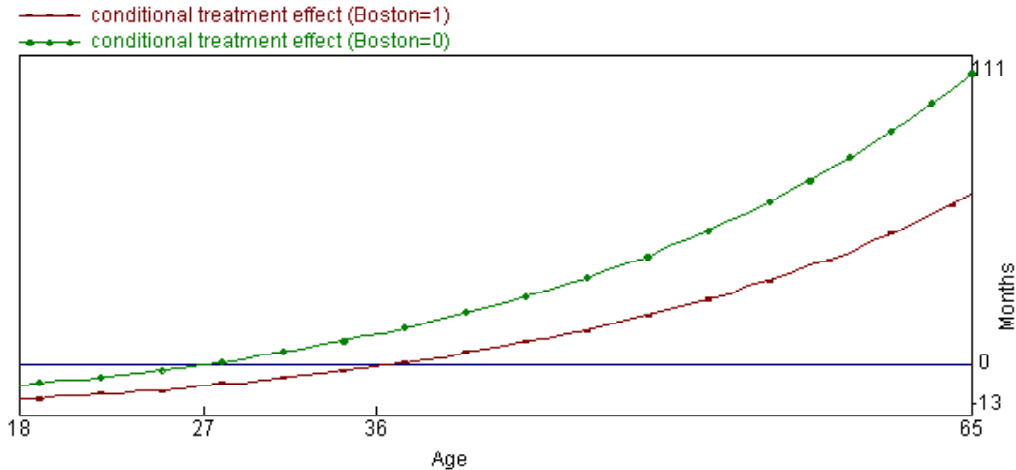


Figure 1. Conditional treatment effect on recidivism

4.8 Comparison with the preliminary data analysis

In the preliminary data analysis we have argued that if the dependence of $\Pr [T_c \leq t | T_s, X_c]$ on T_s is substantially reduced after X_c is integrated out, then the inequality (1) is no longer detectable. To verify this conjecture, we have estimated $\Pr [T_c \in (p, q) | T_s = t_s, X_c, G = 1]$ for $t_s < p$ on the basis of

the results in Table 7 and then averaged these estimates over the treatment group, which yield the results in Table 8.

Table 8. Estimated $\Pr [T_c \in (p, q] | T_s = t_s, G = 1]$

p	q	Range of t_s	Mean	Minimum	Maximum
1	6	$0 \rightarrow 1$	0.20697	0.20567	0.20829
6	12	$0 \rightarrow 1$	0.18507	0.18405	0.18610
6	12	$1 \rightarrow 6$	0.19164	0.18610	0.19753
12	∞	$0 \rightarrow 1$	0.56324	0.56194	0.56457
12	∞	$1 \rightarrow 6$	0.57202	0.56457	0.58015
12	∞	$6 \rightarrow 12$	0.59190	0.58015	0.60480

Indeed, the dependence of $\Pr [T_c \in (p, q] | T_s = t_s, G = 1]$ on $t_s < p$ is weak, which explains why we could not find any dependence.

5 Conclusions

In this paper we have investigated whether the ESEO program has an effect on recidivism, using a version of the bivariate mixed proportional hazard model proposed by Abbring and van den Berg (2003). In first instance we have left the baseline hazards involved unspecified, and included all available covariates in the systematic hazards. Subsequently, we have determined the relevant covariates and baseline hazard specifications via a sequence of Wald and likelihood ratio tests. Thus, the model specification is to a large extent done in a data driven way.

In contrast with previous studies we find that the ESEO program has an effect on recidivism, but this effect depends on age and location: the ESEO program reduces the risk of recidivism only for ex-inmates over the age of 27 in San Diego and Chicago, and over the age of 36 in Boston, but increases the risk of recidivism for the other ex-inmates in the treatment group. The effect of age on recidivism is not surprising, as this has been documented in the criminology literature before. See for example Uggen (1999, 2000) and Hanson and Bussiere (1998). The latter study focuses on sex offenders, however. In view of Figure 1 it seems that the positive effect of the treatment for the older ex-convicts outweighs the negative effect for the younger ex-convicts, in terms of the expected number of months with

which the rearrest will be postponed. Hence, heterogeneity of impact is an important point to consider when evaluating re-entry programs.

One of the agreements in the literature on program evaluation is that given the specificities of the groups of people who usually make use of those services, some programs that work well for a given group may not work so well for others. In other words, the effects of programs may be heterogenous. See, for example, Heckman et al. (1999). That is exactly what we find.

Our results provide evidence that employment programs for ex-offenders can reduce recidivism, provided that these programs take the heterogeneity of the population of ex-offenders into account. A program that is uniform for all ex-offenders may not yield the expected results. This paper has therefore made a positive contribution to the debate in the criminological literature about the likely effects of ex-offenders employment programs. See Visher et al. (2005). However, the eligibility criteria for the ESEO program and the fact that we had to condition on absence of attrition make it difficult to extrapolate our findings to other recidivism reducing intervention programs.

In unemployment duration studies, age and education are usually important factors for the length of the unemployment spell. In the case under review, however, the job search duration does not depend on any individual-specific covariates. This may be due to the fact that all individuals in the sample have one dominant characteristic in common, namely being ex-convicts.

Recall from Table 7 that the job search duration in San Diego and Chicago are significant higher than in Boston. On the other hand, in view of Table 7 and Figure 1 it seems that the post-placement service rendered to the ex-convicts in the treatment group in Boston is less effective in reducing recidivism than in the other locations. However, the two Boston effects may be related. Suppose that the job search assistance in Boston is aimed at lesser quality jobs than in Chicago and San Diego, which takes less time than looking for better jobs. Since it is plausible that ex-offenders placed in low quality jobs are more prone to recidivism than those placed in better quality jobs (see Uggen 1999), this hypothesis would explain the Boston effects. Of course, this is a speculative explanation because we do not have data on the quality of the job placements.

It is important for the scientific evaluation of future employment programs for ex-offenders that more and better data are made available. For instance, it would have been helpful for our analysis if we have had details about the teams that help each ex-inmate in finding a job, such as their case

loads, and more details about the ex-offenders themselves, such their criminal and employment history, the status of their release (parole, probation, or unconditional release), time served versus sentence time, family characteristics, level of participation in the job search, the types of jobs searched for, and past and presents employer’s evaluations. Moreover, we fail to understand the reason why in the ESEO data the job search and recidivism durations were interval censored. If we had observed uncensored durations, we would have been able to conduct a more sophisticated econometric analysis, for example by including unobserved heterogeneity in our model. Furthermore, future program evaluations should pay more attention to the attrition problem, in particular by trying to trace down the drop-outs and gathering information about the reasons for dropping out.

Although our methodology is specific for the ESEO program, it may be of general interest as an example of how to extract information from bad data by careful econometric modeling. A model gives structure to the data information, and the less of the latter, the more model assumptions have to be made in order to compensate for the lack of data quality. Admittedly, such a model (and our model in particular) may not be a perfect representation of the data-generating process, hence the numerical estimation result should not be taken at face value. However, the qualitative results, in particular which variables have an impact and the directions of their impact, are usually robust under model misspecification. In other words, misspecified models may generate no results but will unlikely generate spurious results.

6 Appendix

6.1 Preliminary data analysis

To test whether

$$\Pr [T_c \in (1, 6]] = \Pr [T_c \in (1, 6]|T_s \in (0, 1]], \quad (20)$$

$$\Pr [T_c \in (6, 12]] = \Pr [T_c \in (6, 12]|T_s \in (0, 1], T_s \in (1, 6]] \quad (21)$$

and

$$\begin{aligned} & \Pr [T_c \in (12, \infty)] \\ &= \Pr [T_c \in (12, \infty)|T_s \in (0, 1], T_s \in (1, 6], T_s \in (6, 12]] \end{aligned} \quad (22)$$

we have estimated Logit models for each of these conditional probabilities, and for each group separately:

$$\Pr [T_c \in (1, 6] | T_s \in (0, 1]] = F(\beta_{1,0} + \beta_{1,1}I(T_s \in (0, 1]) + \beta_{1,2}) \quad (23)$$

$$\begin{aligned} & \Pr [T_c \in (6, 12] | T_s \in (0, 1], T_s \in (1, 6]] \\ & = F(\beta_{2,0} + \beta_{2,1}I(T_s \in (0, 1]) + \beta_{2,2}I(T_s \in (1, 6])) \end{aligned} \quad (24)$$

$$\begin{aligned} & \Pr [T_c \in (12, \infty) | T_s \in (0, 1], T_s \in (1, 6], T_s \in (6, 12]] \\ & = F(\beta_{3,0} + \beta_{3,1}I(T_s \in (0, 1]) + \beta_{3,2}I(T_s \in (1, 6]) + \beta_{3,3}I(T_s \in (6, 12])) \end{aligned} \quad (25)$$

where $F(x)$ is the logistic distribution function. However, in the case $G = 0$ we have $I(T_s \in (0, 1]) + I(T_s \in (1, 6]) + I(T_s \in (6, 12]) = 1$, so that in the case (25) we can only estimate

$$\begin{aligned} & \Pr [T_c \in (12, \infty) | T_s \in (0, 1], T_s \in (1, 6]] \\ & = F(\beta_{3,0} + \beta_{3,1}I(T_s \in (0, 1]) + \beta_{3,2}I(T_s \in (1, 6])) \end{aligned} \quad (26)$$

Note that we do not need to worry about misspecification of these Logit models, because the explanatory variables involved are mutually exclusive dummy variables. Therefore, the hypothesis (20) is equivalent to the hypothesis $\beta_{1,1} = 0$ in (23), the hypothesis (21) is equivalent to the hypothesis $\beta_{2,1} = \beta_{2,2} = 0$ in (24), and the hypothesis (22) is equivalent to the hypothesis $\beta_{3,1} = \beta_{3,2} = \beta_{3,3} = 0$ in (25) if $G = 1$, and to the hypothesis $\beta_{3,1} = \beta_{3,2} = 0$ in (26) if $G = 0$.

The estimation and test results are presented in Tables A.1 and A.2. In all cases the null hypothesis is not rejected by the Wald test (or squared t test in the case $i = 1$) at any conventional significance level!

Table A.1. Logit results for (23)-(25)/(26)

Treatment group ($G = 1$)

i	$\beta_{i,0}$	$\beta_{i,1}$	$\beta_{i,2}$	$\beta_{i,3}$	Wald test
	(t-value)	(t-value)	(t-value)	(t-value)	(p-value)
1	-1.2809338	-0.0391111			0.0225
	(-7.17)	(-0.15)			(0.88076)
2	-1.5040774	-0.0606246	0.1094842		0.88076
	(-3.33)	(-0.12)	(0.22)		(0.83261)
3	0.4054651	-0.1335314	-0.2595112	0.1823216	1.19
	(0.44)	(-0.14)	(-0.28)	(0.18)	(0.75601)

**Table A.2 . Logit results for (23)-(25)/(26)
Control group ($G = 0$)**

i	$\beta_{i,0}$	$\beta_{i,1}$	$\beta_{i,2}$	Wald test
	(t-value)	(t-value)	(t-value)	(p-value)
1	-1.2992830	0.0216225		0.0025
	(-3.99)	(0.05)		(0.96012)
2	-1.6094379	-0.1973594	0.0930904	0.35
	(-1.47)	(-0.17)	(0.08)	(0.83801)
3	-0.6931472	1.1093076	1.1826954	1.69
	(-0.80)	(1.24)	(1.29)	(0.42906)

6.2 Job search model specification

The initial maximum likelihood results for job search are presented in Table A.3. Recall that the α 's are the parameters of the integrated baseline hazard (12).

Table A.3. Job search: Model 1

Covariates	Estimates	t-val.	
<i>AGE</i>	0.001374	0.141	*
<i>AGE1ARR</i>	0.011526	0.964	*
<i>DRUGS</i>	-0.060380	-0.526	*
<i>WHITE</i>	0.128538	1.051	*
<i>MALE</i>	-0.206205	-1.258	*
<i>CHICAGO</i>	-1.188883	-7.057	
<i>SANDIEGO</i>	-0.353766	-2.454	
<i>EDUC_L</i>	-0.261975	-1.393	*
<i>EDUC_H</i>	-0.094193	-0.592	*
Parameters			
α_1	1.139577	2.788	
α_2	0.806235	2.577	
α_3	1.082720	2.186	
LogL, n	-417.622	503	

The Wald test of the hypothesis that the coefficients of the covariates in Table A.3 indicated by an asterix (*) are jointly zero has p-value 0.45137.

The Wald test of the null hypothesis $\alpha_1 = \alpha_2 = \alpha_3$ has p-value 0.10356, so that the null hypothesis involved cannot be rejected at the 10% significance level. Recall that the latter hypothesis implies that the baseline hazard is constant. However, for the time being we will not implement the restriction $\alpha_1 = \alpha_2 = \alpha_3$. First, we will get rid of the insignificant covariates. The results are presented in Table A.4.

Table A.4. Job search: Model 2

Covariates	Estimates	t-val.
<i>CHICAGO</i>	-1.211832	-7.739
<i>SANDIEGO</i>	-0.326150	-2.303
Parameters		
α_1	1.185860	10.711
α_2	0.847022	6.014
α_3	1.061613	3.719
LogL, n	-420.811	503

The null hypothesis $\alpha_1 = \alpha_2 = \alpha_3$ is now rejected: the Wald test involved has p-value 0.01131. On the other hand, the null hypothesis $\alpha_2 = \alpha_3$ is accepted: The Wald test involved has p-value 0.39498. The latter result indicates that the baseline hazard is non-increasing. This suggests to specify a Weibull baseline hazard. The estimation results involved are presented in Table 5.

As a double-check whether the model can be reduced from the initial model in Table A.3 to the model in Table 5 we have conducted the likelihood-ratio test: The LR test involved has p-value 0.35230, hence the hypothesis cannot be rejected at any conventional significance level. Moreover, the t-test statistic of the null hypothesis $\alpha_{2,s} = 1$ has value -2.091 , which is borderline significant at the 5% level for the two-sided t-test, and significant for the left-sided t-test (the corresponding left-sided p-value is 0.01826). Since it is implausible that the "hazard" of finding a job increases with the job search duration, the left-sided result prevails.

It is conceivable⁵ that the job search duration T_s also depends on the group assignment G . To check this, we have re-estimated the model in Table A.3 with G and the products of G with the covariates in Table A.3 included as additional covariates. The Wald test that this extended model can be

⁵As noted by a referee.

reduced to the model in Table A.4 has p-value 0.54600. Consequently, the hypothesis involved can not be rejected at any conventional significance level.

6.3 Recidivism of the control group

The initial maximum likelihood results are presented in Table A.5.

Table A.5. Recidivism of the control group: Initial model

Covariates	Estimates	t-val.
<i>AGE</i>	-0.057819	-2.098
<i>AGE1ARR</i>	-0.061837	-1.716
<i>DRUGS</i>	0.232762	0.733
<i>WHITE</i>	-0.012282	-0.036
<i>MALE</i>	-0.290917	-0.721
<i>CHICAGO</i>	0.287844	0.725
<i>SANDIEGO</i>	0.069299	0.165
<i>EDUC_L</i>	-0.103330	-0.206
<i>EDUC_H</i>	0.748567	1.741
Parameters		
α_1	0.366003	0.906
α_2	0.609465	0.957
α_3	0.493359	1.010
LogL, n	-134.728	134

The covariate *AGE* is borderline significant at the 5% level; all the other covariates are insignificant at any conventional significance level. The Wald test that all the coefficients of the covariates (including the one for *AGE*) are zero has p-value 0.32881, hence the null hypothesis that T_c does not depend on covariates cannot be rejected at any conventional significance level. Moreover, the Wald test of the null hypothesis $\alpha_1 = \alpha_2 = \alpha_3$ has p-value 0.80083, and therefore cannot be rejected at any conventional significance level. Recall that this hypothesis is equivalent to the hypothesis that $\Lambda_c(t) = \alpha_1 t$. Thus, the distribution of T_c for the control group is exponential, without covariates!

The LR test that the model in Table A.5 can be reduced to the exponential model in Table 6 has p-value 0.59809, hence the latter cannot be

rejected at any conventional significance level. Therefore, we will adopt the exponential model (14) for the recidivism of the control group.

Recall that our model assumes that conditional on the covariates in X , T_c and T_s are independent. As a check on this assumption, we have also estimated and tested a version of the model in Table A.5 with the dummy variables $I(T_s \in (0, 1])$, $I(T_s \in (1, 6])$ and $I(T_s \in (6, 12])$ included as covariates. The Wald test that the coefficients of these three dummy variables are jointly zero has p-value 0.73451, and the Wald test that all the covariates, including these three dummy variables, have zero coefficients has p-value 0.83536. Thus, our conclusion that the recidivism duration T_c for the control group is distributed exponential without covariates stands.

6.4 Joint maximum likelihood results

The initial maximum likelihood estimation results are presented in Table A.6, for recidivism only.

Table A.6. Recidivism: Model 1

Covariates	Estimates	t-val.	
<i>DRUGS</i>	-0.217296	-1.027	*
<i>CHICAGO</i>	-0.523601	-2.073	
<i>SANDIEGO</i>	-0.461512	-1.926	
<i>WHITE</i>	-0.278290	-1.332	*
<i>MALE</i>	-0.050078	-0.154	*
<i>EDUC_L</i>	0.576838	1.874	
<i>EDUC_H</i>	-0.330336	-0.939	*
<i>AGE</i>	-0.040221	-2.200	
<i>AGE1ARR</i>	-0.030686	-1.292	*
1	2.245037	3.226	
Parameter			
α_c	0.041529	7.672	
LogL, n	-842.896	503	

The parameters indicated by an asterix (*) are jointly insignificant: The p-value of the Wald test involved is 0.42136. Therefore, in the next estimation round these covariates have been removed. See Table A.7.

As to the results in Table A.7, the Wald test that only *AGE* matters for recidivism yields p-value 0.03464, hence this hypothesis is rejected at

the 5% significance level. Moreover, the Wald test that the coefficients of the two location dummy variables are equal has p-value 0.53307, which indicates that we may replace these dummy variables by the dummy variable $BOSTON = 1 - CHICAGO - SANDIEGO$. Furthermore, the test of the same hypothesis, jointly with the hypothesis that the coefficient of $EDUC_L$ is zero, has p-value 0.20804.

Table A.7. Recidivism: Model 2

Covariates	Estimates	t-val.
<i>CHICAGO</i>	-0.301132	-1.350
<i>SANDIEGO</i>	-0.460476	-2.030
<i>EDUC_L</i>	0.487441	1.639
<i>AGE</i>	-0.047794	-2.870
1	1.707907	3.581
Parameter		
α_c	0.040407	7.550
LogL, n	-845.992	503

Therefore, we have re-estimated the model without the education dummy, and with the location dummies replaced by the dummy variable $BOSTON$, yielding the results in Table 7.

As a double-check we have conducted the LR test that the initial model in Table A.6 can be reduced to the model in Table 7. The p-value of the test is 0.25816.

References

- Abbring, J.H., and G.J. van den Berg (2003): The non-parametric identification of treatment effects in duration models, *Econometrica*, 71, 1491-1517.
- Abbring, J.H., G.J. van den Berg and J.C. van Ours (2005): The effect of unemployment insurance sanctions on the transition rate from unemployment to employment, *Economic Journal*, 115, 602-630.
- Bierens, H.J. (2008): Semi-nonparametric interval-censored mixed proportional hazard models: Identification and consistency results, *Econometric Theory*, 24, 749-794.
- Bierens, H.J., and J.R. Carvalho (2010): Attrition in the employment services for ex-offenders program. http://econ.la.psu.edu/~hbierens/ESEO_ATTRITION.PDF.

- Bloom, H.S. (1984): Accounting for no-shows in experimental evaluation design, *Evaluation Review*, 8, 225-246.
- Bloom, D. (2006): Employment-focused programs for ex-prisoners. What have we learned, what are we learning, and where should we go from here?, Michigan Disability Rights Coalition. <http://www.mdrc.org/publications/435/full.pdf>.
- Bureau of Justice Statistics (1994): Recidivism of prisoners released in 1994.
- Beck, A.J., and B.E. Shipley (1989): Recidivism of prisoners released in 1983, Special report, Bureau of Justice Statistics.
- Carvalho, J.R., and H.J. Bierens (2007): Conditional treatment and its effect on recidivism, *Brazilian Review of Econometrics*, 27, 53-84.
- Devine, T., and N. Kiefer (1991): *Empirical Labor Economics: The Search Approach*, New York, USA: Oxford University Press.
- Eberwein, C., J. Ham, and R. Lalonde (1997): The impact of being offered and receiving classroom training on the employment histories of disadvantaged women: Evidence from experimental data, *Review of Economic Studies*, 64, 655-682.
- Fagan, J., and R. Freeman (1999): Crime and work, *Crime and Justice*, 25, 225-290.
- Farrington, D.P., and B.C. Welsh BC (2005): Randomized experiments in criminology: What have we learned in the last two decades?, *Journal of Experimental Criminology*, 1, 9-38.
- Freeman, R. (1999): Economics of crime, Chapter 52 in Ashenfelter, A.C., and D. Card (Eds), *The Handbook of Labor Economics, Vol 3C*, Amsterdam, the Netherlands: Elsevier, 3529-3571.
- Freeman, R. (2003): Can we close the revolving door? Recidivism versus employment of ex-offenders in the US., *Urban Institute Reentry Round Table*, New York University Law School.
- Gendreau, P., T. Little, and C. Goggin (1996): A meta-analysis of the predictors of adult offender recidivism: What works!, *Criminology*, 34, 575-607.
- Hanson, R.K., and M.T. Bussiere (1998): Predicting relapse: A meta-analysis of sexual offender recidivism studies, *Journal of Consulting and Clinical Psychology*, 66, 348-362.
- Harer, M.D. (1994): Recidivism among federal prisoners released in 1987, *Journal of Correctional Education*, 46, 98-127.
- Hausman, J.A. (1978): Specification tests in econometrics, *Econometrica*, 46, 1251-1271.

- Heckman, J.J., R.J. Lalonde, and J.A. Smith (1999): The economics and econometrics of active labor market programs. Chapter 12 in Ashenfelter, A.C., and D. Card (Eds), *Handbook of Labor Economics, Vol. 3A*, Amsterdam, the Netherlands: Elsevier, 1865-2097.
- Heckman, J.J., J. Smith, and C. Taber (1998): Accounting for dropouts in evaluations of social programs, *The Review of Economics and Statistics*, 80, 1-14.
- Imbens, G.W. (2004): Nonparametric estimation of average treatment effects under exogeneity: A review, *Review of Economics and Statistics*, 86, 4-30.
- Maltz, M.D. (1984): *Recidivism: Quantitative Studies in Social Sciences*, Orlando, FL, USA: Academic Press.
- Milkman, R. (2001): Employment services for ex-offenders, 1981-1984: Boston, Chicago, and San Diego, Discussion Paper 8619, ICPSR.
- Milkman, R., A. Timrots, A. Peyser, M. Toborg, B.G.A. Yezer, L. Carpenter, and N. Landson (1985): Employment services for ex-offenders field test, Discussion paper, The Lazar Institute.
- Phipps, P., K. Korinek, S. Aos, and R. Lieb (1999), Research findings on adult corrections programs: A review, Washington State Institute for Public Policy.
- Pissarides, C.A. (2000): *Equilibrium Unemployment Theory*, Second edition, Cambridge, MA, USA: MIT Press.
- 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.
- Sampson, R.J., and J.H. Laub (1997): A life-course theory of cumulative disadvantage and the stability of delinquency, *Advances in Criminological Theory*, 7, 133-161.
- Seiter, R.P., and K.R. Kadela (2003): Prisoner reentry: What works, what does not, and what is promising, *Crime and Delinquency*, 49,360-388.
- Schmidt, P., and A.D. Witte (1988): *Predicting Recidivism Using Survival Models: Research in Criminology*, Berlin, Germany: Springer-Verlag.
- Timrots, A. (1985): An evaluation of employment services programs for ex-offenders, Masters thesis, University of Maryland, College Park.
- Uggen, G. (1999): Ex-offenders and the conformist alternative: A job quality model of work and crime, *Social Problems*, 46, 127-151.
- Uggen, C. (2000): Work as a turning point in the life course of criminals: A duration model of age, employment and recidivism, *American Sociological Review*, 65, 529-546.

- Van den Berg, G.J. (1999): Empirical inference with equilibrium search models of the labor market, *The Economic Journal*, 109, 283-306.
- Van den Berg, G.J. (2000): Duration models: Specification, identification, and multiple durations, Chapter 52 in Heckman, J.J., and E.E. Leamer (Eds), *Handbook of Econometrics, Vol. V*, Amsterdam, the Netherlands: Elsevier, 3381-3460.
- Van den Berg, G.J., B. van der Klaauw, and J.C. van Ours (2004): Punitive sanctions and the transition rate from welfare to work, *Journal of Labor Economics*, 22, 211-241.
- Visher, C.A., L. Winterfield, and M.B. Coggeshall (2005): Ex-offender employment programs and recidivism: A meta-analysis, *Journal of Experimental Criminology*, 1, 295-316.
- Western, B. (2002): The impact of incarceration on wage mobility and inequality, *American Sociological Review*, 67, 526-546.