Social Experiments and Instrumental Variables with Duration Outcomes

Jaap H. Abbring *
Gerard J. van den Berg †

September 8, 2004

Abstract

This paper examines the empirical analysis of treatment effects on duration outcomes from data that contain instrumental variation. We focus on social experiments in which an intention to treat is randomized and compliance may be imperfect. We distinguish between cases where the treatment starts at the moment of randomization and cases where it starts at a later point in time. We derive exclusion restrictions under various informational and behavioral assumptions and we analyze identifiability under these restrictions. It turns out that randomization (and by implication, instrumental variation) by itself is often insufficient for inference on interesting effects, and needs to be augmented by a semiparametric structure. We develop corresponding non- or semiparametric tests and estimation methods.

*Department of Economics, Free University Amsterdam, and Tinbergen Institute. Email: jabbring@econ.vu.nl.

†Department of Economics, Free University Amsterdam, Tinbergen Institute, IFAU-Uppsala, IZA, and CEPR. Email: gberg@feweb.vu.nl.
Address for correspondence: Department of Economics, Free University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.

Keywords: event-history analysis, intention to treat, non-compliance, selection.
JEL codes: C14, C31, C41, J6.

Thanks to participants at the EC^2 Conference (London, December 2003), the CEPR conference at Bristol (April 2004), and seminars at IFAU-Uppsala, Tinbergen Institute Amsterdam, CEU Budapest, Princeton, Yale, Brown, and USC for useful comments. Jaap Abbring acknowledges financial support by the Royal Netherlands Academy of Arts and Sciences. Previous title: Social Experiments in Real Time.

1 Introduction

Social experiments have become important tools for policy evaluation in general, and the evaluation of active labor market policies in particular (see Heckman, LaLonde and Smith, 1999, for a survey). Until recently, it was generally thought that policy evaluation using social experiments merely involved trivial comparisons of outcomes between the various treatment statuses assigned in the experiment. The studies of Ham and LaLonde (1996), Meyer (1996), and Eberwein, Ham and LaLonde (1997) show that complications arise if the outcome variable of interest is a duration variable (e.g. unemployment duration) or depends on the realization of a duration variable (e.g. the post-unemployment wage or the subsequent employment duration). The evaluation based on outcomes among those who survive up to a certain point is confounded with dynamic selection effects even if initial treatment assignment is randomized. Further problems arise if compliance to the treatment status assigned in the experiment (the intention to treat, or ITT) is imperfect and if actual program enrollment is a time-consuming process. In this paper, we explore the use of social experiments in duration and event-history analysis. We consider the randomized ITT as an instrumental variable (IV), and more generally consider the use of instrumental variables in duration analysis.

To motivate the paper and outline its contributions, it is useful to give (in the next paragraphs) a taxonomy of different cases that may arise in practice. This corresponds to the way the paper is organized. In all cases we allow the population under study to be heterogeneous. Throughout the paper, we use the terms "randomization" and "random assignment" to denote situations in which an ITT is assigned independently of agents' individual characteristics. An agent is said to "comply" with the ITT outcome if the actual treatment status coincides with it.[1] In every case, the policy setting determines what is meant with "treatment".

In Section 2, we start with the benchmark case where (*i*) randomization of ITT occurs at time 0 (which is usually the moment of inflow into the state of which the subsequent time spent in it is the outcome of interest), (*ii*) agents are immediately subject to the treatment, and (*iii*) there is perfect compliance. We show that dynamic selection effects may arise, and care has to be taken how to define and estimate treatment effects. We re-appraise the insights from Ham and LaLonde (1996). We also briefly discuss what happens if assumption (*ii*) is

[1] In the literature, compliance is often only used for agents who are assigned to be treated instead of being in the control group, but here we also use it for those who are assigned to the control group, unless stated otherwise.

violated in the sense that the actual treatment occurs at a later point in time. This is less relevant from an empirical point of view, because, with a positive time interval between ITT assignment and actual treatment, perfect compliance becomes a very strong assumption.

Subsequently, as the second main case, we relax in Section 3 the perfect compliance assumption (*iii*) but maintain assumptions (*i*) and (*ii*). This case arises e.g. if there is randomization of program offers (ITT) at time 0 and agents choose immediately whether to accept the offer and enroll. In such cases, ITT can be used as an IV for, but does not coincide with, actual treatment. We first develop and analyze a nonparametric IV estimator that, unlike existing estimators, allows for censoring. We then provide novel semiparametric identification results, and a corresponding semiparametric IV estimator.

To avoid problems with imperfect compliance, researchers often analyze the effect of ITT instead of actual treatment in reduced-form models. Methodologically, such an analysis fits our first main case. Moreover, under the appropriate IV conditions, the null of no treatment effects and the null of no ITT effects are equivalent (e.g. Robins and Tsiatis, 1991, and Abadie, 2002). Two advantages of an ITT-analysis are often cited (see e.g. Robins and Tsiatis, 1991). First, randomization can be ensured even if compliance is imperfect and selective. Second, ITT rather than actual treatment is the relevant public-policy instrument if the policy maker cannot control compliance any better than the analyst.[2] On the other hand, an IV analysis allows to some extent to disentangle the effects of compliance (participation) and the effects of actual treatment. If enough (parametric) structure is imposed, the results of an IV analysis can be extrapolated beyond the scope of the experiment. Therefore, and because an ITT-analysis is already covered by our first case, we consider the effects of actual treatment using ITT as an IV.

The third main case, which we examine in Section 4, is the case where both assumptions (*ii*) and (*iii*) are violated, i.e. randomization of ITT occurs at time 0, the actual treatment occurs later, and compliance is imperfect. Again, ITT can be used as an instrument (see e.g. Eberwein, Ham and LaLonde, 1997). However, as argued by Rosenzweig and Wolpin (2000) and Abbring and Van den Berg (2003b), it is often hard to justify exclusion restrictions in a dynamic setting with forward-looking agents. In particular, if ITT affects the treatment, as it should if it is to be a valid instrument, then it is likely to affect the outcome hazard up to the moment of treatment enrollment as well. Therefore, we introduce a

[2]For example, the threat of a punishment treatment may be considered as a treatment itself (Abbring, Van den Berg and Van Ours, 1997, and Black et al., 2003).

weaker-than-usual IV assumption that only requires "ex post exclusion", that is exclusion of ITT from the outcome hazard after treatment enrollment. We discuss informational and behavioral assumptions under which this weaker IV assumption holds, and discuss its implications for identifiability. We show that, even though ITT is randomly assigned, "ex ante" effects of ITT on treatment assignment and on outcome hazards before treatment enrollment cannot be identified, and can typically not even be signed. Moreover, the effects of actual treatment are poorly identified in this case. More constructively, we argue that data on ex post ITT effects are informative on selection effects. After all, under ex post exclusion these effects cannot be causal. We demonstrate that a selectivity test based on this idea bears a remote analogy to tests on cohort effects in mortality rates (Vaupel and Yashin, 1985) and on unobserved heterogeneity in duration models with time-varying explanatory variables (Van den Berg and Van Ours, 1997).

All results apply to the case where there is no deliberate randomization at time 0 but the data contain appropriate instrumental variation (e.g., because of a natural experiment). Also, our results carry over to more complex, dynamic experimental designs than the one considered here (Section 5 provides some discussion). In all cases, we follow nonparametric and semiparametric approaches.[3] We do not restrict attention to effects on survival probabilities but, in line with duration analysis in general, we focus on the effects on the hazard rate of the outcome duration variable, because of the intimate link between the hazard rate and economic behavioral models (Van den Berg, 2001). Knowing the effect on individual behavior as reflected in the hazard rate and the way this changes over time enables one to learn something about the reasons for the effectiveness or ineffectiveness of a policy, and this allows one to extrapolate the experimental results to slightly different policies and policies in slightly different environments. As an example, consider the third case above, where the empirical distinction between "ex ante" and "ex post" effects calls for a hazard rate analysis. We return to this below.

Various papers consider methods for IV analysis of distributional treatment effects that apply to some extent to the problems studied in this paper. The nonparametric approach for effects on survival probabilities in our second main case is closely related to Imbens and Rubin (1997) and Abadie (2002), who dis-

[3]We also generalize results to cases where the treatment is not binary but varies continuously. Further generalizations concern cases where treatment and instrument may change over time at the same discrete (deterministic or random) points of times. This avoids the substantive problems discussed in Abbring and Van den Berg (2003b), but unlikely to have many applications. The methodology for such cases has been discussed in the context of transformation models by Bijwaard (2003). We do not consider such cases.

cuss identification, estimation, and testing of distributional treatment effects in a nonparametric setting.[4] Their methods, however, do not handle censoring. In contrast, our focus is on methods that allow for right-censoring. Duration-model IV estimators that allow for censoring have been developed by Robins and Tsiatis (1991), Bijwaard and Ridder (2003), Bijwaard (2003), and Chesher (2003). These estimators apply to our second and/or third main cases. They require (semi-)parametric structure, and additional substantive assumptions like perfect compliance in the control group (Bijwaard and Ridder, 2003). With some exceptions, they do not focus on effects on the individual hazard rate. We will discuss these estimators where appropriate, notably when we evaluate exclusion restrictions in our third main case.

Throughout the paper, we consider experiments at face value, and we do not address generalizability issues. Specifically, we do not address endogenous selection into the experimental population, differences between the experiment and the permanent imposition of a policy, and equilibrium effects in general. See Heckman, LaLonde and Smith (1999) for a detailed discussion, Ferrall (2002) for a comprehensive dynamic economic framework, and Van den Berg and Van der Klaauw (2001) for an empirical illustration in a reduced-form duration-analysis framework.

The remainder of the paper is organized as follows. Sections 2, 3, and 4 discuss our three main cases. Section 5 concludes.

2 Randomization with perfect compliance

2.1 Potential-outcome framework, treatment effects, and available data

We consider the population of agents or individuals flowing into a state of interest, and the durations these individuals subsequently spend in that state. Upon inflow, an individual is assigned to a treatment from a set $\mathcal{S}$. In this section, we assume that the individual complies with the assigned treatment, so that ITT and actual treatment coincide. We are interested in the causal effect of the treatment on the

[4] Also, Abadie, Angrist and Imbens (2002) develop and apply a semiparametric IV estimator of quantile treatment effects. This estimator does not allow for censoring. It is moreover based on quantile models that are designed for outcomes like earnings and that do not apply naturally to duration and event-history outcomes.

duration spent in the state of interest (the "outcome").[5]

We model this effect using the potential-outcome framework pioneered by Neyman (1923). To each treatment $s \in \mathcal{S}$ corresponds a random variable $T(s)$, the potential outcome duration that would prevail if we would intervene and assign treatment s. We assume that $\{T(s)\} := \{T(s); s \in \mathcal{S}\}$ is a measurable stochastic process.[6] Moreover, for ease of exposition we assume that each $T(s)$ is continuous, and we denote the hazard rate of $T(s)$ by $\theta_{T(s)}$. We are interested in contrasts between the distributions of $T(s)$ and $T(s')$ corresponding to treatments $s, s' \in \mathcal{S}$. These contrasts are summarized in so-called treatment effects. We focus on two of these,

$$\frac{\theta_{T(s')}(t)}{\theta_{T(s)}(t)} \qquad \text{and} \qquad \Pr(T(s') > t) - \Pr(T(s) > t), \tag{1}$$

which are, respectively, the relative effect on the hazard rate at t and the additive effect on the survival probability at t of replacing one treatment s by another treatment s', as functions of t. The former captures the effect on the most interesting feature of the duration distribution. The latter, as we shall see, can easily be related to the standard literature on treatment evaluation. For the time being we consider a single subject, or, equivalently, a homogeneous group of subjects. We may of course consider a whole range of alternative treatment effects, like

$$\Pr(T(s') > t)/\Pr(T(s) > t) \qquad \text{and} \qquad \mathbb{E}[T(s')]/\mathbb{E}[T(s)].$$

In fact, the relative effect on the survival probability may often be more interesting than the additive effect. However, as we shall see in the next subsection, nonparametric inference simultaneously provides estimates of all of these. Moreover, as we shall see in Subsection 2.4, heterogeneity across subjects causes related methodological difficulties for all treatment effects except for the additive effect on the survival probability (and the effects derived from it). The two special effects in (1) therefore cover the whole range of treatment effects from a methodological point of view.

The treatment is assigned according to a $\mathcal{S}$-valued random variable S. Throughout this section, we assume that treatment assignment is randomized, i.e.[7]

Assumption 1 (Randomization). $S \perp\!\!\!\perp \{T(s)\}$.

[5] For expositional clarity, we restrict attention to a single outcome duration. Most methods in this paper are easy to extend to more general event-history outcomes. For examples, the Nelson-Aalen estimator used in this section has been developed for general event-history processes.

[6] This process, viewed as a random function $s \mapsto T_s$, can alternatively be interpreted as a nonparametric structural equation for the determination of the outcome by the treatment.

[7] More generally, randomization could be conditional on observed covariates, or even observed

The actual outcome duration is $T := T(S)$; all other potential outcomes are counterfactual.

We allow for random right-censoring.[8] To this end, define a random censoring time C that is independent of (T, S). Our data are derived from the full-information random sample $((T_1, S_1, C_1), \ldots, (T_n, S_n, C_n))$ from the distribution of (T, S, C). Suppose that observation i is censored if $T_i > C_i$, and complete if $T_i \leq C_i$. Then our data are the limited-information random sample $((\widetilde{T}_1, S_1, D_1), \ldots, (\widetilde{T}_n, S_n, D_n))$, where D_i is defined as $D_i := I(T_i \leq C_i)$, so $D_i = 1$ indicates a complete observation, and

$$\widetilde{T} = \begin{cases} T_i & \text{if } T_i \leq C_i \\ C_i & \text{if } T_i > C_i \end{cases}$$

is the possibly censored outcome duration. In the sequel we do not consider treatment effects on moments of the outcome duration (like the mean) because typically the observation window is finite (i.e., observed durations are always right-censored if they exceed some finite time) and the moments are not nonparametrically identified from such data.

2.2 Nonparametric estimators and tests

With randomization, as in Assumption 1, standard hazard regressions (e.g. Andersen et al., 1993, and Fleming and Harrington, 1991) are directly informative on $\{T(s)\}$. We briefly discuss nonparametric estimation and testing methods from the perspective of the treatment-evaluation problem.

Let the integrated hazard of $T(s)$ be denoted by $\Theta_{T(s)}$, so

$$\Pr\left(T(s) > t\right) = \exp\left(-\Theta_{T(s)}(t)\right).$$

If S is discrete, then $\{\Theta_{T(s)}\}$ can be estimated by repeated application of the Nelson-Aalen estimator. Without loss of generality, denote $\mathcal{S} = \{0, 1, \ldots, k\}$, where possibly $k = \infty$. Then, the Nelson-Aalen estimator of $\Theta_{T(s)}$ is given by

external covariate processes. Throughout most of this paper, we ignore observed covariates. If appropriate, it is implicitly understood that results hold conditional on covariates. In the case of discrete covariates, all empirical methods can be directly applied to strata defined by the covariates. In practice, continuous covariates are harder to handle, because of the usual computational problems.

[8] The censoring mechanism specified here is usually referred to as "simple random right-censoring". Extensions to more general forms of independent censoring and filtering are straightforward (see Andersen et al., 1993, and Fleming and Harrington, 1991).

(see e.g. Andersen et al., 1993, p.178)

$$\widehat{\Theta}_{T(s)}(t) = \sum_{\{i:D_i=1,\widetilde{T}_i\leq t,S_i=s\}} \left[R_s(\widetilde{T}_i)\right]^{-1},$$

with $R_s(t) := \sum_{j=1}^n I(\widetilde{T}_j \leq t, S_j = s)$ the number of spells in the subsample with $S_i = s$ that are at risk (not completed nor censored) at time t. The asymptotic behavior of $\widehat{\Theta}_{T(s)}(t)$ $(s \in \mathcal{S})$ follows from standard results for the Nelson-Aalen estimator (e.g., Andersen et al., 1993, Example IV.1.6):

$$\sqrt{np_s}\left(\widehat{\Theta}_{T(s)} - \Theta_{T(s)}\right) \xrightarrow{\mathcal{D}} \mathbb{G}_s \quad (s \in \mathcal{S})$$

jointly, where $p_s := \Pr(S = s)$. Here, $\mathbb{G}_s$ $(s \in \mathcal{S})$ are mutually independent Gaussian martingales such that $\mathbb{G}(0) = 0$ and, for $t < t'$,

$$\operatorname{cov}\left(\mathbb{G}_s(t), \mathbb{G}_s(t')\right) = \int_0^t \frac{\theta_{T(s)}(\tau)}{\overline{F}_{T(s)}(\tau)\,\overline{F}_C(\tau^-)} d\tau =: \sigma_s^2(t),$$

where $\overline{F}_{T(s)}$ and $\overline{F}_C$ are the survival functions of $T(s)$ and C, respectively. The variances $\sigma_s^2(t)$ can be consistently estimated by

$$\widehat{\sigma_s^2}(t) = \sum_{\{i:D_i=1,\widetilde{T}_i\leq t,S_i=s\}} \left[R_s(\widetilde{T}_i)\right]^{-2},$$

and np_s by $N_s := \sum_{i=1}^n I(S_i = s)$.

The potential duration distributions and survival probabilities are one-to-one related to the potential integrated hazards by

$$\overline{F}_{T(s)}(t) = \mathcal{P}_0^t\left(1 - d\Theta_{T(s)}(\tau)\right) = \exp\left(-\Theta_{T(s)}(t)\right).$$

Here, $\mathcal{P}$ is the product integral. The properties of the well-known Kaplan-Meier estimator of $\overline{F}_{T(s)}$,

$$\widehat{\overline{F}}_{T(s)}(t) = \mathcal{P}_0^t\left(1 - d\widehat{\Theta}_{T(s)}(\tau)\right) = \prod_{\{i:D_i=1,\widetilde{T}_i\leq t,S_i=s\}} \left(1 - [R_s(\widetilde{T}_i)]^{-1}\right)$$

follow from the properties of the Nelson-Aalen estimator. In particular, according to the delta method, $\sqrt{np_s}\left(\widehat{\overline{F}}_{T(s)} - \overline{F}_{T(s)}\right)$ is asymptotically equivalent to $-\sqrt{np_s}\overline{F}_{T(s)}\left(\widehat{\Theta}_{T(s)} - \Theta_{T(s)}\right)$, so that jointly

$$\sqrt{np_s}\left(\widehat{\overline{F}}_{T(s)} - \overline{F}_{T(s)}\right) \xrightarrow{\mathcal{D}} -\overline{F}_{T(s)}\mathbb{G}_s \quad (s \in \mathcal{S}). \tag{2}$$

Without censoring ($\overline{F}_C = 1$), $1 - \widehat{\overline{F}}_{T(s)}$ reduces to the empirical distribution function,

$$\operatorname{cov}(-\overline{F}_{T(s)}(t)\mathbb{G}_s(t), -\overline{F}_{T(s)}(t')\mathbb{G}_s(t')) = \overline{F}_{T(s)}(t)\overline{F}_{T(s)}(t')\sigma_s^2(t) = F_{T(s)}(t)\overline{F}_{T(s)}(t')$$

for $t \leq t'$, and $-\overline{F}_{T(s)}\mathbb{G}_s$ is an F-Brownian bridge.

These results are the basis for estimating potential-outcome contrasts and for various tests. One can immediately derive (i) uniform confidence bounds on the potential duration distributions, (ii) point-wise results for isolated survival probabilities, and (iii) results for treatment effects. See Andersen et al. (1993) for methods of inference for the hazard rates $\theta_{T(s)}$ that are derived in the above framework (these methods typically impose some smoothness of the hazard rates as functions of the duration).

As an example, consider the additive effect of replacing treatment s by s' on the survival function at t, defined as $\overline{F}_{T(s')}(t) - \overline{F}_{T(s)}(t)$. For given t, $\widehat{\overline{F}}_{T(s)}(t)$ consistently estimates the probability of surviving for t periods under treatment s, with asymptotic standard error $(np_s)^{-1/2}\overline{F}_{T(s)}(t)^2\sigma_s(t)$. The latter can be consistently estimated by $N_s^{-1/2}\widehat{\overline{F}}_{T(s)}(t)^2\widehat{\sigma}_s(t)$. The estimators $\widehat{\overline{F}}_{T(s)}(t)$ and $\widehat{\overline{F}}_{T(s')}(t)$ are asymptotically independent. In the evaluation of a training program, we could use the estimator to assess the effect of training on the probabilities of staying unemployed for 6 or 12 months. Because of right-censoring, discrete-choice models like probit models are not a good alternative. Also, in discrete-choice models it is hard to deal with time-varying explanatory variables. The dynamically assigned treatment of Section 4 of this paper provides an example of such an explanatory variable.

The Nelson-Aalen estimator is also a basis for tests of null hypotheses like $\overline{F}_{T(s)} = \overline{F}_{T(s')}$ for $s, s' \in \mathcal{S}$. Andersen et al. (1993), Chapter V, provides a unified treatment. One popular test is the log rank test, which behaves particularly well against proportional treatment effects

$$\overline{F}_{T(s)} = \exp(-\Theta_{T(s)}) = \exp(-c\Theta_{T(s')}) = \left(\overline{F}_{T(s')}\right)^c$$

for $0 < c \neq 1$. Other tests are Kolmogorov-Smirnov and Cramér-Von Mises tests for censored data.

In the case of continuous S, additional smoothness is required. For example, Cox's (1972) regression can be used to estimate the scalar parameter $\gamma > 0$ in the proportional hazards (PH) model $\theta_{T(s)}(t) = \gamma^s\lambda(t)$. The "baseline hazard" function λ can be estimated nonparametrically by the Breslow estimator. Dabrowska (1987) develops an alternative for Cox's regression based on kernel smoothing.

As in the nonparametric approach, discrete-choice models are not a good alternative if outcomes are censored or if there are time-varying explanatory variables. Moreover, discrete-choice analysis is typically less efficient than semiparametric duration analysis based on e.g. the PH model.

2.3 Treatment plans

In the model of Subsection 2.1 the ITT is (randomly) assigned at time 0 and is always observed by the analyst. The assumed perfect compliance ensures that the actual treatment equals the ITT. By implication, the treatment is also assigned at time 0, and is always observed.

The most straightforward application of this model takes $\mathcal{S}$ to be a set of programs in which agents either enroll at time 0 or not enroll at all. However, the model is versatile in dealing with alternative timing patterns. In particular, the treatment space itself may include one or more time dimensions. For example, a treatment in $\mathcal{S}$ may not only specify a particular program but also the time of enrolling into this program. Perfect compliance requires that agents comply with the full ITT, including the assigned timing of the program.[9] Moreover, because the ITT is always observed, so is the actual treatment. This is true even if program enrollment is intended to occur after the agent has left the state of interest.

If the treatment is discrete, for example if there is a finite set of programs and enrollment takes place at discrete times, then we can straightforwardly apply the nonparametric methods of the previous subsection. With continuous enrollment, we again need additional smoothness. Here, we should be careful in choosing appropriate semiparametric structure. In the typical case that agents observe the ITT at time 0, they will perfectly anticipate future enrollment and respond to the assigned treatment from time 0 onwards. It is well known that Cox's (1972) PH specification is typically inconsistent with dynamic economic theory in this case.[10] Therefore, the methods based on the PH model suggested in the previous subsection can typically not be used.

We do not elaborate on this. As noted in Section 1, perfect compliance is a strong assumption in this case. We postpone further discussion to Section 4, which addresses the more relevant case with imperfect compliance.

[9]Note that this may be applied to the case in which program enrollment at an intended time is contingent on not having left the state of interest at that time. Also, recall that the model may be interpreted as being conditional on some external observed covariate process. This allows for the possibility that the ITT is a plan contingent on such a covariate process.

[10]See Van den Berg (2001) for a discussion of the PH model from an job-search perspective.

2.4 Unobserved heterogeneity and treatment effects

2.4.1 Selectivity despite randomization

If the population is homogeneous, the nonparametric treatment effect estimators of Subsection 2.2 estimate the corresponding individual treatment effects. However, as noted in Section 1, we allow individuals to be heterogeneous, and in this subsection we examine the implications of using the nonparametric tools in a heterogeneous population.

We only consider ex ante individual heterogeneity in outcomes that can be captured by time-invariant unobserved characteristics V, because this is sufficient to obtain the main insights. Randomization (Assumption 1) implies that $S \perp\!\!\!\perp V$. For ease of exposition, we take V to be a continuous random variable and S to be a binary indicator with realizations 1 ("treatment") and 0 ("control"), and we assume that the hazard rates $\theta_{T(s)}(t|V)$ satisfy the usual regularity conditions that guarantee existence of the expressions below.

Individual treatment effects are now defined in terms of the distributions of $T(0)|V$ and $T(1)|V$, whereas average treatment effects now concern averages over the relevant population, i.e. over the distribution of V in the relevant population. Thus, the individual treatment effects on the hazard rate and the survival probability at t are defined as

$$\frac{\theta_{T(1)}(t|V)}{\theta_{T(0)}(t|V)} \qquad \text{and} \qquad \Pr(T(s') > t|V) - \Pr(T(s) > t|V)$$

respectively. The average additive treatment effect on the survival probability at t is naturally defined as $\mathbb{E}[\Pr(T(s') > t|V) - \Pr(T(s) > t|V)]$. This equals $\Pr(T(s') > t) - \Pr(T(s) > t)$, so the nonparametric tools straightforwardly apply to inference of this effect in the presence of heterogeneity. This result extends to much more general types of heterogeneity than considered here (e.g. V may vary over time).

Unfortunately, the above line of reasoning cannot be applied to any other treatment effect of interest. Consider the effect on the hazard rate. We demonstrate that the nonparametric estimators cannot be used to obtain consistent estimates of average treatment effects, and, related to that, that the definition of average treatment effects on the hazard rate is nontrivial.

The main difficulties follow from the fact that if the treatment has a causal effect on the duration, then, typically, the distribution of V among the survivors at points in time $t > 0$ depends on the treatment, so $V \not\perp\!\!\!\perp S|T > t$. In other words, there is no randomization at $t > 0$ despite the randomization at 0. To see this,

let f, F, and $\overline{F}$ be generic symbols for a density, a distribution function, and a survival function, with subscripts denoting the corresponding random variable (note that $\overline{F} := 1 - F$). Further, denote the hazard of T by θ_T, and its integrated hazard by Θ_T. Note that because of randomization at 0, $\theta_{T(s)}(t|V) = \theta_T(t|S = s, V)$ and $\Theta_{T(s)}(t|V) = \Theta_T(t|S = s, V)$. From e.g. Lancaster (1990),

$$f_V(v|T > t, S) = \frac{\overline{F}_T(t|S, V) f_V(v)}{\int_0^\infty \overline{F}_T(t|S, V) dF_V(v)}, \tag{3}$$

and $\overline{F}_T(t|S, V) = \exp\left[-\Theta_T(t|S, V)\right]$.[11]

It is not difficult to construct examples in which the distribution of V among the treated survivors at t is first-order stochastically dominated by the distribution of V among the non-treated survivors at t, in particular if there is a strong positive interaction between S and V in the hazard rate of T, and this hazard rate increases in V and S. Then the individual hazard rate at t is very large if both $S = 1$ and V is large, and as a result the survivors at t may contain relatively few treated individuals with a high V. As a specific example, let F_V be a gamma distribution with unit mean and variance, and let

$$\theta_{T(s)}(t|V) = \theta^*_{T(s)}(t) V \tag{4}$$

for all t, s, with $\Theta^*_{T(s)}(t)$ defined analogously. Then, the distribution of $V \mid T > t, S = 1$ equals the distribution of

$$\frac{1 + \Theta^*_{T(0)}(t)}{1 + \Theta^*_{T(1)}(t)} V \;\Bigg|\; T > t, S = 0,$$

so[12]

$$\frac{\mathbb{E}\left[V|T > t, S = 1\right]}{\mathbb{E}\left[V|T > t, S = 0\right]} = \frac{1 + \Theta^*_{T(0)}(t)}{1 + \Theta^*_{T(1)}(t)}.$$

What does this imply for inference on average treatment effects on the hazard rate? First, we have to define average effects. Consider the average treatment effect defined by

$$\mathbb{E}[\theta_{T(1)}(t|V)/\theta_{T(0)}(t|V)]$$

[11] Of course, the distributions of V among the survivors at t ($F_V(v|T > t, S = s)$ with $s = 0, 1$) differ from the distribution of V in the inflow into unemployment ($F_V(v)$), but this is not the point here.

[12] One can construct nongeneric examples where the distribution of V among survivors does not depend on S, notably if $\theta_T(t|S, V)$ is additive in a term depending on S and t on the one hand and a term depending on V and t on the other.

This involves aggregation over V in the population. However, a hazard concerns a subpopulation of survivors at t, which is systematically different from the population, so instead one would like to take the average over V among survivors. But because the selectivity of survivors at $t > 0$ depends on the treatment status, one has to be specific about this status as well. We propose the following average treatment effect on the individual hazard rate,

$$\mathbb{E}\left[\frac{\theta_{T(1)}(t|V)}{\theta_{T(0)}(t|V)} \;\middle|\; T(1) \geq t\right]$$

which can be called the *average treatment effect on the treated survivors at t*. It averages over the distribution of V among the survivors at t if the agents are assigned to the treatment. Under randomization, this is equivalent to averaging over the distribution of V among the treated survivors at t (so with $T \geq t, S = 1$). If equation (4) is satisfied then the average treatment effect on the treated survivors at t equals both the average and the individual treatment effect on the hazard at t (which then equal $\theta^*_{T(1)}(t)/\theta^*_{T(0)}(t)$). Note that in this case the individual treatment effect is homogeneous across individuals but not necessarily over time. Note that in general the average treatment effect on the treated survivors at t is a property of a subpopulation whose composition depends on the treatment effect on $[0, t)$.

Now consider nonparametric inference. Without unobserved heterogeneity, $\theta_T(t|S = 1)/\theta_T(t|S = 0)$ equals the average population treatment effect on the hazard rate at t. With unobserved heterogeneity

$$\frac{\theta_T(t|S=1)}{\theta_T(t|S=0)} = \frac{\mathbb{E}[\theta_T(t|S=1,V) \mid T \geq t, S=1]}{\mathbb{E}[\theta_T(t|S=0,V) \mid T \geq t, S=0]},$$

so this ratio reflects (i) the treatment effect and (ii) the selection effect that at $T = t$, among the survivors at t, the treated and controls have systematically different unobserved characteristics despite the randomization at $t = 0$. The nonparametric estimator of $\theta_T(t|S = 1)/\theta_T(t|S = 0)$ (which is basically its sample equivalent) therefore does not capture the treatment effect. In the earlier example with the gamma distribution for V, the nonparametric estimator of $\theta_T(t|S = 1)/\theta_T(t|S = 0)$ underestimates the average effect for the population.

In fact, one can construct examples where

$$\theta_T(t|S=1) < \theta_T(t|S=0)$$

even if

$$\theta_{T(1)}(t|V) > \theta_{T(0)}(t|V) \quad \text{almost surely for all } t. \tag{5}$$

For example, let V have a discrete distribution with $\Pr(V = 0.2) = \Pr(V = 2.5) = 0.5$, and let equation (4) be satisfied, with $\Theta^*_{T(1)}(t) > \Theta^*_{T(0)}(t)$ for all $t > 0$ (note that this is weaker than inequality (5) for $\theta_{T(s)}(t|V)$ in this example). Then $\theta_T(t|S = 1) < \theta_T(t|S = 0)$ for values of $\Theta^*_{T(s)}(t)$ in an interval around 1. In such cases the dynamic selection effect on the observed hazard rate dominates the treatment effect, in certain time intervals. Obviously, this may lead to invalid nonparametric inference on the sign of the treatment effect.

Similar results can be derived for e.g. the use of the sample equivalent of $F_T(t + a|T > t, S = 1)/F_T(t + a|T > t, S = 0)$ for $a > 0$ (e.g. from a probit analysis of whether $T \in (t, t + a]$ given $T > t, S$) to estimate the corresponding average effect on the individual conditional survival probability.[13]

2.4.2 Semiparametric approaches

To proceed towards more constructive results, we need to impose some semiparametric structure on the distribution of $T(s)|V$. Subsequently, two approaches can be taken. First, one may abandon point identification and estimation, and focus on bounding the parameters of interest (e.g. Manski, 1997). Second, we may impose a structure that is sufficiently specific to enable point identification and estimation of relevant treatment effects.

Following the first approach, we may assume that equation (4) applies and that the individual-level potential-outcome distributions can be ranked in terms of first-order stochastic dominance, i.e. $\Theta^*_{T(1)}(t) > \Theta^*_{T(0)}(t)$ for all $t > 0$, $\Theta^*_{T(1)}(t) = \Theta^*_{T(0)}(t)$ for all $t > 0$, or $\Theta^*_{T(1)}(t) < \Theta^*_{T(0)}(t)$ for all $t > 0$. In Appendix 1 we prove that, with unobserved heterogeneity,

$$\frac{\theta_T(t|S = 1)}{\theta_T(t|S = 0)} < \mathbb{E}\left[\frac{\theta_{T(1)}(t|V)}{\theta_{T(0)}(t|V)}\right] = \mathbb{E}\left[\frac{\theta_{T(1)}(t|V)}{\theta_{T(0)}(t|V)} \,\middle|\, T(1) \geq t\right] \tag{6}$$

if $\Theta^*_{T(1)}(t) > \Theta^*_{T(0)}(t)$, for all $t > 0$. The results for the other two cases follow as straightforward modifications. The ranking of the individual-level potential-outcome distributions can be inferred from the ranking of $F_{T;S=1}(t)$ and $F_{T;S=0}(t)$

[13] Average non-additive treatment effects on quantities that are defined for the whole population, such as the average of the relative effect on the survival probability defined in Subsection 2.1, can also not be inferred nonparametrically, but this is only because of the non-additivity. Note that if $\Theta_{T(1)}(t|V) > \Theta_{T(0)}(t|V)$ for all $t > 0$ and at all V, then the observable quantity $\overline{F}_T(t|S = 1)/\overline{F}_T(t|S = 0) - 1$ is always negative (this follows from $\overline{F}_T(t|S = s) = \mathbb{E}[\exp(-\Theta_{T(s)}(t|V))]$) and vice versa, so then this observable quantity always has the same sign as the average treatment effect $\mathbb{E}[\exp(-\Theta_{T(1)}(t|V))/\exp(-\Theta_{T(0)}(t|V))] - 1$. A similar result holds for $\theta_T(0|S = 0) - \theta_T(0|S = 1)$. All these quantities are based on samples drawn at $t = 0$, at which there has been no dynamic selection yet.

near $t = 0$, so the average treatment effects on the hazard rate can be bounded by the observable left-hand side of equation (6) (see also Meyer, 1996).

Now let us turn to the second semiparametric approach, imposing a structure that is sufficiently specific to enable point identification and estimation. An obvious choice is to impose that the hazard rate of $T(s)|V$ is multiplicative in t, s and V, so, in equation (4),

$$\theta^*_{T(s)}(t) = \gamma^s \, \lambda(t)$$

for some scalar parameter $\gamma > 0$ and some "baseline hazard" function λ. This results in the so-called two-sample MPH model for $T|S, V$, with

$$\theta_T(t|S, V) = \lambda(t) \, \gamma^S \, V. \tag{7}$$

Here, "two-sample" refers to the subsamples with $S = 0$ and $S = 1$. We are primarily interested in identification of the individual (and average) treatment effect parameter γ. Under the additional assumption that $\mathbb{E}[V] < \infty$ this parameter is identified from (see Elbers and Ridder, 1982, and Kortram et al., 1995)

$$\gamma = \lim_{t \downarrow 0} \frac{F_T(t|S = 1)}{F_T(t|S = 0)}. \tag{8}$$

In applications of MPH models, the "random effects" assumption that observed and unobserved explanatory variables are independent is often controversial. However, here this assumption follows from the randomization of S.

The analysis by Lenstra en Van Rooij (1998) suggests that the sample analog of the right-hand side of (8), $\left[1 - \widehat{\overline{F}}_{T(1)}(t_n)\right] / \left[1 - \widehat{\overline{F}}_{T(0)}(t_n)\right]$, is a consistent estimator of γ if $t_n \downarrow 0$ at an appropriate rate as $n \to \infty$.[14] Other semiparametric structures of the hazard rate of $T(s)|V$ may also lead to point identification and estimation of treatment effects (see Heckman and Taber, 1994, and Van den Berg, 2001, for surveys).

2.4.3 Post-duration outcomes

We end this subsection by re-addressing the results in Ham and LaLonde (1996) on problems with inference of treatment effects on post-spell outcomes if randomization occurs at $t = 0$. Consider an outcome W that is realized immediately after T. Let $W(s)$ be the random outcome that would prevail if we would hypothetically assign the treatment s to the agent. The process $\{W(s)\}$, which is assumed

[14] Horowitz (1999) provides an asymptotically normal estimator of γ for the case that S is continuous and the model satisfies additional smoothness conditions.

measurable, can again be given a structural interpretation. In general, $\{W(s)\}$ may depend on V. The observation window is denoted by $[0, \mathcal{T})$ with $\mathcal{T} < \infty$. We observe W iff $T < \mathcal{T}$. The distribution of $V|S = s$ among the agents for whom W is observed equals the distribution of $V|T < \mathcal{T}, S = s$. It follows that, among the observed W, the distribution of V among the treated in general differs from the distribution of V among the controls. So, despite randomization at $t = 0$, we cannot simply compare the observed mean outcomes of W among treated and controls in order to uncover e.g. $\mathbb{E}[W(1) - W(0)]$. However, let $\mathcal{T} \to \infty$. Then each sample member provides an observation of W. As $V \perp\!\!\!\perp S|T < \infty$ because of randomization, it follows that the distribution of V is the same among the treated and the controls for whom W is observed, and inference can be based on simple comparisons of observed outcomes of W.

This line of reasoning also applies if W is realized with a certain delay after T, and/or if the analysis is only based on realizations of W in a fixed time interval. One may think of T as the unemployment duration and W as the post-unemployment wage, the post-unemployment job duration, or the earnings of employed workers say two years after the experiment.

We conclude that problem with causal inference on post-duration outcomes arises because of a bounded observational window (i.e., right-censoring on a bounded interval). Without the latter, the former does not arise. With a bounded observational window, a solution is to estimate a model that takes unobserved heterogeneity into account.[15]

3 Randomization and immediate treatment with imperfect compliance

3.1 Model and data

We now relax Section 2's assumption of perfect compliance, but retain the assumption that treatment is immediate. Instead of being randomly assigned to a treatment as in Assumption 1, agents are randomly assigned a label $Z \in \mathcal{Z}$ at time 0. In the typical experimental setup, Z takes values in the set $\mathcal{Z} = \mathcal{S}$ of possible treatments and is interpreted as the ITT with the treatment Z. More generally, we allow Z to be an IV with support $\mathcal{Z} \neq \mathcal{S}$. We will nevertheless refer to Z as the ITT throughout.

[15] See Visser (1996) for a related empirical analysis of the duration of the second (AIDS) stage of HIV/AIDS if the observational window is fixed.

Let $S(z)$ be the random actual treatment that would prevail if we hypothetically assign the agent the label $z \in \mathcal{Z}$. The process $\{S(z)\}$, which is assumed measurable, can again be given a structural interpretation. The actual treatment is $S = S(Z)$. In the ITT case with perfect compliance, we would have that $S(z) = z$ for all $z \in \mathcal{Z} = \mathcal{S}$, and that $S = S(Z) = Z$. In the present section, however, we allow for imperfect compliance (so $S = S(Z) \neq Z$) by allowing $S(z)$ to be non-degenerate. If Z is discrete and an effective ITT, it should be the case that $\Pr(S(z) = z) > \Pr(S(z) = z')$ for all $z, z' \in \mathcal{S}$ such that $z' \neq z$.

Throughout this section we maintain

Assumption 2 (IV). (i) $z \mapsto S(z)$ is nontrivial (in a way that will be further specified in special cases below), and (ii) $Z \perp\!\!\!\perp (\{T(s)\}, \{S(z)\})$.

Underlying Assumption 2(ii) is the notion that (a) the ITT does not causally affect outcomes directly, so that potential outcomes $T(s)$ need not be indexed by z, (b) outcomes do not causally affect treatment, so that $S(z)$ need not be indexed by possible values t of T, and (c) the ITT is not causally affected by either outcomes or treatment (see Abbring, 2003, for discussion).

With imperfect compliance, the actual treatment $S = S(Z)$ and the potential outcomes $\{T(s)\}$ are typically dependent, because agents non-experimentally self-select or are selected in actual treatment. This sets the present analysis apart from that of Section 2. Note, however, that a reduced-form analysis of the effect of ITT on outcomes fits Section 2's framework. Formally, the outcome equation can be reduced to $\{T(S(z))\}$. Under Assumption 2, $Z \perp\!\!\!\perp \{T(S(z))\}$, so that Assumption 1 holds for the reduced-form model with Z replacing S.

The data of Subsection 2.1 are accordingly enriched with instrumental variation. We now have a random sample $((\widetilde{T}_1, S_1, D_1, Z_1), \ldots, (\widetilde{T}_n, S_n, D_n, Z_n))$ from the joint distribution of $(\widetilde{T}, S, D, Z)$.

3.2 Nonparametric estimation and testing with IV

For now, consider the binary treatment-binary instrument case that $\mathcal{S} = \{0, 1\}$ and $\mathcal{Z} = \{0, 1\}$. Again, it is convenient to think of the treatment statuses as "treatment" $(s = 1)$ and "control" $(s = 0)$ in this case. Define $p(z) := \Pr(S(z) = 1)$. Note that $p(z) = \Pr(S = 1|Z = z)$ under Assumption 2. Assume that Imbens and Angrist's (1994) monotonicity property holds, that is $S(0) \leq S(1)$ or $S(0) \geq S(1)$. Without further loss of generality, we take $S(0) \leq S(1)$. Then, the subpopulation that switches treatment status between propensities $p(0)$ and $p(1)$ all switch from treatment 0 to treatment 1. This subpopulation, called "compliers" by Imbens and Rubin (1997), is therefore $Q := \{S(0) = 0, S(1) = 1\}$,

and has probability mass $\Pr(Q) = p(1) - p(0)$. We formalize Subsection 3.1's Assumption 2(i) by assuming that $\Pr(Q) > 0$.

Standard methods from the treatment-effects literature can be adapted to learn about various average contrasts between $T(1)$ and $T(0)$ on Q, and the marginal distributions $F_{0;Q}$ and $F_{1;Q}$ of, respectively, $T(0)$ and $T(1)$ on Q. In this subsection we only consider the average additive treatment effect on the survival probabilities, as from Section 2 we know that nonparametric inference of other effects is not well possible if the population is heterogeneous. Of particular interest here are adaptations of existing methods that allow for censoring.

Identification of the marginal potential-outcome distributions $F_{0;Q}$ and $F_{1;Q}$ on Q is straightforward (Imbens and Rubin, 1997). In particular, in Appendix 2 we show that

$$\overline{F}_{0;Q}(t) = \frac{\Pr\left(T > t, S = 0 | Z = 0\right) - \Pr\left(T > t, S = 0 | Z = 1\right)}{p(1) - p(0)} \quad \text{and}$$
$$\overline{F}_{1;Q}(t) = \frac{\Pr\left(T > t, S = 1 | Z = 1\right) - \Pr\left(T > t, S = 1 | Z = 0\right)}{p(1) - p(0)}.$$

By implication, the mean survival probability contrast on Q

$$\Delta_Q(t) := \overline{F}_{1;Q}(t) - \overline{F}_{0;Q}(t),$$

is identified (Imbens and Angrist, 1994). This is a local average treatment effect on survival for at least t periods.

In applied work, researchers often choose only a few values of t and apply e.g. Imbens and Angrist's (1994) IV estimator to assess the effect of a program on survival through the corresponding intervals of time. For example, in the evaluation of a training program, one could define the outcomes to be survival in unemployment for 6 months. The appeal of such an approach lies in its computational and presentational simplicity. However, if richer survival data are available, it ignores potentially useful information on the effect of the program. Furthermore, standard methods cannot handle the fact that survival data are often censored, nor can they handle time-varying covariates. We therefore focus on econometric methods to learn about the *functions* $\overline{F}_{0;Q}$, $\overline{F}_{1;Q}$, and Δ_Q in the possible presence of right-censoring.

To this end, first note that

$$\Delta_Q = \frac{\overline{F}_{T;Z=1} - \overline{F}_{T;Z=0}}{p(1) - p(0)},$$

where $\overline{F}_{T;Z=z}(t) := \Pr(T > t | Z = z)$. Thus, the causal null that $\overline{F}_0 = \overline{F}_1$ is equivalent to the reduced-form null that $\overline{F}_{T;Z=0} = \overline{F}_{T;Z=1}$ (Robins and Tsiatis,

1991, and Abadie, 2002). Thus, under the IV assumptions we can test for distributional treatment effects using any of the nonparametric tests of Section 2.

Next, consider estimating $\overline{F}_{0;Q}$, $\overline{F}_{1;Q}$, and Δ_Q. First note that $\overline{F}_{T;Z=z}(t)$ can be estimated by the Kaplan-Meier estimator

$$\widehat{\overline{F}}_{T;Z=z}(t) = \prod_{\{j:D_j=1,\widetilde{T}_j\leq t,Z_j=z\}} \left(1-[R_z(\widetilde{T}_j)]^{-1}\right),$$

with $R_z(t) := \sum_{i=1}^n I(\widetilde{T}_i \leq t, Z_i = z)$ the appropriate risk set. The conditional survival probability $\overline{F}_{T;S=s,Z=z}(t) := \Pr(T > t|S = s, Z = z)$ can be estimated by a similar Kaplan-Meier estimator $\widehat{\overline{F}}_{T;S=s,Z=z}(t)$. Also, $p(z)$ can be estimated by

$$\widehat{p}(z) = \frac{\sum_{i=1}^n S_i I(Z_i = z)}{\sum_{i=1}^n I(Z_i = z)}.$$

Thus, $\overline{F}_0$ and $\overline{F}_1$ can be estimated by

$$\widehat{\overline{F}}_{0;Q} = \frac{[1-\widehat{p}(0)]\,\widehat{\overline{F}}_{T;S=0,Z=0} - [1-\widehat{p}(1)]\,\widehat{\overline{F}}_{T;S=0,Z=1}}{\widehat{p}(1)-\widehat{p}(0)} \quad \text{and} \tag{9}$$

$$\widehat{\overline{F}}_{1;Q} = \frac{\widehat{p}(1)\widehat{\overline{F}}_{T;S=1,Z=1} - \widehat{p}(0)\widehat{\overline{F}}_{T;S=1,Z=0}}{\widehat{p}(1)-\widehat{p}(0)}, \tag{10}$$

respectively, and Δ_Q can be estimated by

$$\widehat{\Delta}_Q = \widehat{\overline{F}}_{1;Q} - \widehat{\overline{F}}_{0;Q} = \frac{\widehat{\overline{F}}_{T;Z=1} - \widehat{\overline{F}}_{T;Z=0}}{\widehat{p}(1)-\widehat{p}(0)}.$$

Asymptotic behavior of these estimators follows from standard results for the Kaplan-Meier estimator. Let $q_z := \Pr(Z = z)$. We have

Proposition 1. *Under Assumption 2,* $\sqrt{n}\left(\widehat{\overline{F}}_{1;Q} - \overline{F}_{1;Q}, \widehat{\overline{F}}_{0;Q} - \overline{F}_{0;Q}\right) \xrightarrow{\mathcal{D}}$

$$\frac{1}{p(1)-p(0)}\Bigg(\left(\overline{F}_{T;S=1,Z=1} - \overline{F}_{1;Q}\right)\mathcal{N}_1 - \left(\overline{F}_{T;S=1,Z=0} - \overline{F}_{1;Q}\right)\mathcal{N}_0$$
$$- \sqrt{\frac{p(1)}{q_1}}\overline{F}_{T;S=1,Z=1}\mathbb{G}_{11} + \sqrt{\frac{p(0)}{q_0}}\overline{F}_{T;S=1,Z=0}\mathbb{G}_{10},$$
$$\left(\overline{F}_{T;S=0,Z=1} - \overline{F}_{0;Q}\right)\mathcal{N}_1 - \left(\overline{F}_{T;S=0,Z=0} - \overline{F}_{0;Q}\right)\mathcal{N}_0$$
$$+ \sqrt{\frac{1-p(1)}{q_1}}\overline{F}_{T;S=0,Z=1}\mathbb{G}_{01} - \sqrt{\frac{1-p(0)}{q_0}}\overline{F}_{T;S=0,Z=0}\mathbb{G}_{00}\Bigg),$$

with $\mathcal{N}_z$ *(*$z = 0, 1$*) a normal random variable with zero mean and variance* $p(z)\left[1 - p(z)\right]/q_z$*,* $\mathbb{G}_{sz}$ *(*$s, z = 0, 1$*) a Gaussian martingale such that* $G_{sz}(0) = 0$ *and, for* $t < t'$*,*

$$\operatorname{cov}\left(\mathbb{G}_{sz}(t), \mathbb{G}_{sz}(t')\right) = \int_0^t \frac{\theta_{T;S=s,Z=z}(\tau)}{\overline{F}_{T;S=s,Z=z}(\tau)\,\overline{F}_C(\tau^-)} d\tau =: \sigma^2_{sz}(t),$$

and $\mathcal{N}_1$*,* $\mathcal{N}_0$*,* $\mathbb{G}_{11}$*,* $\mathbb{G}_{01}$*,* $\mathbb{G}_{10}$*, and* $\mathbb{G}_{00}$ *mutually independent.*

Proof. See Appendix 2. □

It is instructive to note that the mathematical expressions in Proposition 1 differ in a number of ways from the corresponding expressions in Subsection 2.2 (see in particular equation (2)). First, each estimator $\widehat{\overline{F}}_{s;Q}$ now depends on two Kaplan-Meier estimators and on estimators of $p(z)$ (see equations (9) and (10)). For each estimator $\widehat{\overline{F}}_{s;Q}$, the limiting stochastic process is a sum of four independent terms reflecting the variation in the estimators of $p(1), p(0)$, and the two Kaplan-Meier estimators, respectively. The two estimators $\widehat{\overline{F}}_{s;Q}$ $(s = 0, 1)$ are asymptotically dependent because they both depend on the estimators of $p(0)$ and $p(1)$.

Proposition 1 can be used to compute asymptotic standard errors of $\widehat{F}_{0,Q}(t)$, $\widehat{F}_{1;Q}(t)$, and $\widehat{\Delta}_Q(t)$, and (point-wise and uniform) confidence bounds on $F_{0,Q}$, $F_{1;Q}$, and Δ_Q. In particular, for $\widehat{\Delta}_Q$ we have

Corollary 1. *For* $t \leq t'$*, the asymptotic covariance of* $\widehat{\Delta}_Q(t)$ *and* $\widehat{\Delta}_Q(t')$ *equals*

$$\begin{aligned}
\frac{1}{\left[p(1) - p(0)\right]^2}\Big\{ & \frac{p(1)\left[1 - p(1)\right]}{nq_1}\left[\overline{F}_{T;S=1,Z=1}(t) - \overline{F}_{T;S=0,Z=1}(t) - \Delta_Q(t)\right] \\
& \qquad \times \left[\overline{F}_{T;S=1,Z=1}(t') - \overline{F}_{T;S=0,Z=1}(t') - \Delta_Q(t')\right] \\
& + \frac{p(0)\left[1 - p(0)\right]}{nq_0}\left[\overline{F}_{T;S=1,Z=0}(t) - \overline{F}_{T;S=0,Z=0}(t) - \Delta_Q(t)\right] \\
& \qquad \times \left[\overline{F}_{T;S=1,Z=0}(t') - \overline{F}_{T;S=0,Z=0}(t') - \Delta_Q(t')\right] \\
& + \frac{p(1)\overline{F}_{T;S=1,Z=1}(t)\overline{F}_{T;S=1,Z=1}(t')\sigma^2_{11}(t)}{nq_1} \\
& + \frac{\left[1 - p(1)\right]\overline{F}_{T;S=0,Z=1}(t)\overline{F}_{T;S=0,Z=1}(t')\sigma^2_{01}(t)}{nq_1} \\
& + \frac{p(0)\overline{F}_{T;S=1,Z=0}(t)\overline{F}_{T;S=1,Z=0}(t')\sigma^2_{10}(t)}{nq_0} \\
& + \frac{\left[1 - p(0)\right]\overline{F}_{T;S=0,Z=0}(t)\overline{F}_{T;S=0,Z=0}(t')\sigma^2_{00}(t)}{nq_0}\Big\}.
\end{aligned} \tag{11}$$

A consistent estimator of this asymptotic covariance follows by plugging the estimators $\widehat{p}(z)$ of $p(z)$, $M_z := \sum_{i=1}^n I(Z_i = z)$ of nq_z, $\widehat{\overline{F}}_{T;S=s,Z=z}$ of $\overline{F}_{T;S=s,Z=z}$, $\widehat{\Delta}_{s;Q}$ of $\Delta_{s;Q}$, and consistent estimators[16]

$$\widehat{\sigma}^2_{sz}(t) = \sum_{\{j:D_j=1,\widetilde{T}_j\leq t,S_j=s,Z_j=z\}} \left[R_{sz}(\widetilde{T}_j)\right]^{-2}$$

of $\sigma_{sz}(t)$ into equation (11). Note that Corollary 1 provides the asymptotic variance of $\widehat{\Delta}_Q(t)$, and therefore its asymptotic standard error, for $t = t'$.

One special case deserves some attention. In the case that there is no censoring, $\widehat{\Delta}_Q(t)$ reduces to Imbens and Angrist's (1994) IV estimator of the local average treatment effect on the binary outcome $I(T > t)$, the Wald estimator[17]

$$\widehat{\Delta}_Q(t) = \frac{M_1^{-1}\sum_{i=1}^n Z_i I(T_i > t) - M_0^{-1}\sum_{i=1}^n (1-Z_i) I(T_i > t)}{M_1^{-1}\sum_{i=1}^n Z_i S_i - M_0^{-1}\sum_{i=1}^n (1-Z_i) S_i}.$$

In this case, Proposition 1 implies

Corollary 2. *If $\overline{F}_C = 1$ (no censoring), then, for $t \leq t'$, the asymptotic covariance of $\widehat{\Delta}_Q(t)$ and $\widehat{\Delta}_Q(t')$ equals*

$$\begin{aligned}
\frac{1}{[p(1)-p(0)]^2}\Big\{ & \frac{p(1)\,[1-p(1)]}{nq_1}\left[\overline{F}_{T;S=1,Z=1}(t) - \overline{F}_{T;S=0,Z=1}(t) - \Delta_Q(t)\right] \\
& \qquad \times \left[\overline{F}_{T;S=1,Z=1}(t') - \overline{F}_{T;S=0,Z=1}(t') - \Delta_Q(t')\right] \\
& + \frac{p(0)\,[1-p(0)]}{nq_0}\left[\overline{F}_{T;S=1,Z=0}(t) - \overline{F}_{T;S=0,Z=0}(t) - \Delta_Q(t)\right] \\
& \qquad \times \left[\overline{F}_{T;S=1,Z=0}(t') - \overline{F}_{T;S=0,Z=0}(t') - \Delta_Q(t')\right] \\
& + \frac{p(1) F_{T;S=1,Z=1}(t)\overline{F}_{T;S=1,Z=1}(t')}{nq_1} \\
& + \frac{[1-p(1)]\, F_{T;S=0,Z=1}(t)\overline{F}_{T;S=0,Z=1}(t')}{nq_1} \\
& + \frac{p(0) F_{T;S=1,Z=0}(t)\overline{F}_{T;S=1,Z=0}(t')}{nq_0} \\
& + \frac{[1-p(0)]\, F_{T;S=0,Z=0}(t)\overline{F}_{T;S=0,Z=0}(t')}{nq_0}\Big\}.
\end{aligned}$$

[16] See Section 2.

[17] See e.g. Angrist and Krueger (1999) and Heckman, LaLonde and Smith (1999) for general discussions of the Wald estimator in the treatment evaluation context.

For $t = t'$, this reduces to the asymptotic variance given by Imbens and Angrist (1994),

$$\frac{\mathbb{E}\left[\left(I(T>t) - \overline{F}_T(t) - \Delta_Q(t)\left(S - \mathbb{E}[S]\right)\right)^2 \left(Z - \mathbb{E}[Z]\right)^2\right]}{\left[\operatorname{cov}\left(S, Z\right)\right]^2}.$$

In general, the functions $\overline{F}_{0;Q}$, $\overline{F}_{1;Q}$, and Δ_Q inherit the disadvantages of the local average treatment effect parameter in Imbens and Angrist's (1994) work. Unless $p(0) = 0$ and $p(1) = 1$, in which case Q is (almost surely) the entire population, the set Q, and therefore these functions, are instrument-dependent for given propensity scores $p(0)$ and $p(1)$ (see Heckman, 1997, Heckman, LaLonde and Smith, 1999, and Abbring, 2003, for discussion). Policy-evaluation problems usually require information on "parameters of interest" other than the identified local average treatment effects. It is often more interesting to know average treatment effects on individual hazard rates than on survival functions. As elsewhere in the paper, two approaches can then be taken. First, one may focus on bounding the parameters of interest. In this section we do not pursue this approach. Instead, in the next subsection, we follow the second approach, which amounts to the imposition of some semiparametric structure and the investigation of point identification and estimation of treatment effects in the ensuing models.

3.3 Semiparametric IV in a proportional hazards framework

In this subsection we adopt structures for the hazard rate of $T(s)$ that are related to the familiar mixed proportional hazards (MPH) model, allowing for unobserved heterogeneity V across individuals. As in Subsection 2.4, this enables us to focus on individual treatment effects. We only allow $\{T(s)\}$ and S to be dependent by way of a common dependence on the individual V, so $\{T(s)\} \perp\!\!\!\perp S|V$. This means that in the case where S differs from the randomized assignment due to selective compliance, this selection mechanism is captured by V.[18]

We start by adopting the multiplicative structure for $\theta_{T(s)}(t|V)$ from Subsection 2.4.2, resulting in the two-sample MPH model for $T|S, V$,

$$\theta_T(t|S, V) = \lambda(t)\, \gamma^S\, V. \tag{7}$$

[18]Recall that in Subsection 2.4 we examined general models for $T(s)|V$ with $V \perp\!\!\!\perp S$. Compared to that, we now impose some structure on the distributions of $T(s)|V$ and we drop the assumption that $V \perp\!\!\!\perp S$.

We again assume that $\mathbb{E}[V] < \infty$, but we now replace the assumption that $S \perp\!\!\!\perp V$ by the assumption that there is an instrument Z that satisfies Assumption 2. We again take the "worst-case scenario" that Z is binary (again, one may think of Z as an ITT indicator). Among other things, Assumption 2(ii) in this case implies that $Z \perp\!\!\!\perp V$. We now formalize Assumption 2(i) by assuming that $p(1)\mathbb{E}[V|S=1,Z=1] \neq p(0)\mathbb{E}[V|S=1,Z=0]$. Note that $p(0) \neq p(1)$ is necessary for this condition to hold.

Just as in (8), it is useful to focus on limits as $t \downarrow 0$, because at 0 the dynamic selection that we examined in Subsection 2.4 has not yet taken place. There holds that[19]

$$\gamma = \lim_{t \downarrow 0} \frac{p(1)F_{T;S=1,Z=1}(t) - p(0)F_{T;S=1,Z=0}(t)}{[1-p(0)]\,F_{T;S=0,Z=0}(t) - [1-p(1)]\,F_{T;S=0,Z=1}(t)} \tag{12}$$

The right-hand side of this only depends on observable quantities. Thus, γ is identified. We summarize this result in

Proposition 2. *With a valid binary instrument (i.e. that satisfies Assumption 2), and under the assumption that $\mathbb{E}[V] < \infty$, the treatment effect parameter γ in an MPH model with an endogenous binary treatment is identified.*

Note that we do not require exogenous explanatory variables. We also do not require parametric assumptions (like a parametric latent-variable selection equation) on the treatment selection process $\{S(z)\}$. If, in violation of Assumption 2, Z is not informative on S (i.e., if $p(1) = p(0)$) then equation (12) does not have a solution for γ.

By analogy to Lenstra and Van Rooij (1998), it may be possible to demonstrate that γ can be consistently estimated by the sample equivalent of the right-

[19] To see this, note that the right-hand side of (12) equals

$$\lim_{t \downarrow 0} \frac{p(1)f_{T;S=1,Z=1}(t) - p(0)f_{T;S=1,Z=0}(t)}{[1-p(0)]\,f_{T;S=0,Z=0}(t) - [1-p(1)]\,f_{T;S=0,Z=1}(t)}$$

by De l'Hospital's rule, and that

$$f_{T;S=s,Z=z}(t) = \lambda(t)\gamma^s \mathbb{E}\left[V \exp\left(-\gamma^s V \int_0^t \lambda(\tau)d\tau\right) | S=s, Z=z, T \geq t\right].$$

The result follows from $\mathbb{E}[V|Z=1] = \mathbb{E}[V|Z=0]$ which implies that

$$p(1)\mathbb{E}[V|S=1,Z=1] - p(0)\mathbb{E}[V|S=1,Z=0] = [1-p(0)]\mathbb{E}[V|S=0,Z=0] - [1-p(1)]\mathbb{E}[V|S=0,Z=1].$$

hand side of (12),

$$\widehat{\gamma} = \frac{\widehat{p}(1)\left[1-\widehat{\overline{F}}_{T;S=1,Z=1}(t_n)\right] - \widehat{p}(0)\left[1-\widehat{\overline{F}}_{T;S=1,Z=0}(t_n)\right]}{[1-\widehat{p}(0)]\left[1-\widehat{\overline{F}}_{T;S=0,Z=0}(t_n)\right] - [1-\widehat{p}(1)]\left[1-\widehat{\overline{F}}_{T;S=0,Z=1}(t_n)\right]},$$

where $t_n \downarrow 0$ at an appropriate rate as $n \to \infty$. This IV estimator $\widehat{\gamma}$ can be seen as a version for our non-linear model of the Wald IV estimator of a treatment effect in the linear regression model.[20] To see this, note that at $t \downarrow 0$, the specification (7) resembles a non-linear regression model with an endogenous regressor and a constant treatment effect parameter, and that equation (12) can be re-expressed as follows,[21]

$$\gamma - 1 = \lim_{t\downarrow 0} \frac{\theta_T(t|Z=1) - \theta_T(t|Z=0)}{(1-p(0))\theta_T(t|S=0,Z=0) - (1-p(1))\theta_T(t|S=0,Z=1)} \quad (14)$$

Although the proposition does not concern identification of the full model, it is not difficult to achieve the latter e.g. by way of including exogenous explanatory variables X. One may use these identification results as a justification to estimate full models that consist of two model equations: (i) an MPH model equation

[20] See e.g. Angrist and Krueger (1999) and Heckman, LaLonde and Smith (1999) for discussions in a regression model context.

[21] One may elaborate on this intuition by noting that equation (12) can be re-expressed as (see also Footnote 19)

$$\gamma - 1 = \lim_{t\downarrow 0} \frac{f_{T;Z=1}(t) - f_{T;Z=0}(t)}{[1-p(0)]f_{T;S=0,Z=0}(t) - [1-p(1)]f_{T;S=0,Z=1}(t)}. \quad (13)$$

Heuristically, the numerator in the right-hand side of (13) satisfies $[f_{T;Z=1}(t) - f_{T;Z=0}(t)]\,dt = \mathbb{E}[I(t \le T < t+dt)|Z=1] - \mathbb{E}[I(t \le T < t+dt)|Z=0]$. Like the numerator of the population version of the Wald estimator, this is a difference in mean outcomes between the subpopulations with $Z=1$ and $Z=0$. Moreover, near $t=0$ this difference in mean outcomes equals the parameter of interest, $\gamma - 1$, times the denominator in the right-hand side of (13) (see also Footnote 19; we divide by $\lambda(t)$ to cover the cases in which λ is not bounded away from 0 and ∞ at 0):

$$\begin{aligned}\lim_{t\downarrow 0} \frac{f_{T;Z=1}(t) - f_{T;Z=0}(t)}{\lambda(t)} &= (\gamma-1)\left\{p(1)\mathbb{E}[V|S=1,Z=1] - p(0)\mathbb{E}[V|S=1,Z=0]\right\} \\ &= (\gamma-1)\lim_{t\downarrow 0} \frac{[1-p(0)]f_{T;S=0,Z=0}(t) - [1-p(1)]f_{T;S=0,Z=1}(t)}{\lambda(t)}.\end{aligned}$$

The denominator's factor $p(1)\mathbb{E}[V|S=1,Z=1] - p(0)\mathbb{E}[V|S=1,Z=0]$ is nonzero under Assumption 2(i), and reduces to $[p(1)-p(0)]\,\mathbb{E}[V] = (\mathbb{E}[S|Z=1] - \mathbb{E}[S|Z=0])\,\mathbb{E}[V]$ in the special case that $V \perp\!\!\!\perp (S,Z)$. More generally, it reflects not only the difference in mean treatments between the subpopulations with $Z=1$ and $Z=0$, but also the interaction with the unobservables V.

for $T|S,X,V$, say $\theta_T(t|S,X,V) = \lambda(t)\gamma^S \exp(\alpha X)V$, and (ii) a latent-variable selection equation, say $S^* = \beta_0 + \beta_1 Z + \beta_2 X + \varepsilon$, with $S := \mathrm{I}(S^* > 0)$, and where ε and V are possibly dependent. The identification results imply that the estimation results are not fully driven by functional-form assumptions.

It is clear that a continuous instrument Z enables identification of more general models. Chesher (2003) considers an MPH-type model with an endogenous continuous treatment indicator S as well as exogenous variables X, a continuous instrument Z, and a latent variable equation relating S and Z. He demonstrates local identification of ratios of the derivatives of the individual hazard rate with respect to S and X.

We now proceed to the case where Z represents an ITT and non-compliance is asymmetric in the sense that agents always comply if assigned to the control group ($z = 0$), i.e. $S(0) = 0$ and $p(0) = 0$. Under this restriction, Bijwaard and Ridder (2003) develop an estimator of a treatment effect in an MPH model with a parametric baseline hazard. They exploit that, because of randomization, the subpopulation of agents with $Z = 0$ is representative for the population. This ensures that all parameters except the treatment effect are identified from the data on this subpopulation. The treatment effect is subsequently identified from the outcomes of the agents with $Z = 1$. Here we follow the same approach. This requires an MPH model that is fully identified in the absence of treatments.[22] This is usually achieved by including exogenous X variables, so we augment equation (7) with such variables. In addition, we now allow the treatment effect γ to depend on the elapsed time t since treatment and on X,

$$\theta_T(t|S,V,X) \;=\; \lambda(t)\;\gamma(t,X)^S\;\phi(X)\;V, \tag{15}$$

and we make standard assumptions that ensure identification of λ, ϕ, and the distribution of V in the population in the absence of treatments (notably, this requires $X \perp\!\!\!\perp V$). The observed outcomes for $Z = 0$ then identify these quantities. Now consider the outcomes of non-compliers among those who are assigned to be treated. We demonstrate in Appendix 2 that these identify the distribution of $V|S = 0, Z = 1$. Together, this then also identifies the distribution of $V|S = 1, Z = 1$. The outcomes of agents with $S = 1$ and $Z = 1$ subsequently identify the treatment effect function γ. In sum, we have

Proposition 3. *Consider a standard MPH model that is augmented by an endogenous binary treatment and that is identified in the absence of this treatment. Assume perfect compliance among the controls. With a valid binary instrument*

[22] See Heckman and Taber (1994) and Van den Berg (2001) for surveys.

(i.e. that satisfies Assumption 2), the treatment effect as a function of the elapsed duration and observed covariates is identified.

Proof. See Appendix 2. □

From a policy point of view it is obviously important to be able to identify the way in which the individual treatment effect changes over time and across individuals.[23]

4 Randomization with later treatment and imperfect compliance

4.1 Model and data

We further extend the model of Subsection 3.1 by not only allowing for imperfect compliance, but also for positive amounts of time between treatment assignment and treatment enrollment. Again, a $\mathcal{Z}$-valued label Z is randomly assigned at time 0. Then, the agents engage in a time-consuming process of enrolling in a program. We focus on the case of a binary program, in which the agents either enroll at some time in $[0, \infty)$ or not enroll at all. Following Abbring and Van den Berg (2003b), we can formalize this by taking $\mathcal{S} = [0, \infty]$. Then, S simply denotes the random time at which an agent enrolls in the program. The point ∞ corresponds to never enrolling at all.

As before, denote the model for the treatment as a function of the instrument by $\{S(z)\}$. In a social experiment, perfect compliance would again arise if $\mathcal{Z} = \mathcal{S}$ and $S(z) = z$. An interpretation is that a full treatment plan Z, stipulating the timing of future program participation, is randomly assigned at time 0 and is adhered to in all states of the world. If Z is observed by the agent, which we typically assume it is, this is the perfect-foresight case alluded to in Subsection 2.3. In this section, we allow for imperfect compliance, i.e. non-degenerate $\{S(z)\}$. This case is more relevant than the case of perfect compliance and delayed treatment.

For expositional convenience, we take the instrument (ITT) to be binary ($\mathcal{Z} = \{0, 1\}$). Because $\mathcal{S}$ is larger than $\mathcal{Z}$, $\mathcal{Z}$ cannot contain a treatment plan for each possible treatment. Therefore, $S(0)$ and $S(1)$ cannot both be non-degenerate, except in the trivial case that S has binary support. Thus, there is imperfect compliance.

[23] One might want to consider inference in models where the individual treatment effect γ is allowed to depend in a general way on V, but this seems too ambitious.

In the next subsection we argue that the ITT is likely to affect outcomes directly in this case. To accommodate this, we augment the outcomes model by indexing potential outcomes not only by treatments in $\mathcal{S}$, but also by the labels in $\mathcal{Z}$. Thus, the potential-outcomes process is now $\{T(s,z)\} := \{T(s,z); (s,z) \in \mathcal{S} \times \mathcal{Z}\}$. We again assume that each $T(s,z)$ is continuously distributed, with hazard rate $\theta_{T(s,z)}$ and integrated hazard $\Theta_{T(s,z)}$. For expositional convenience, we restrict the joint distributions of $\{T(s,z)\}$ for fixed z as in

Assumption 3. For all $z \in \mathcal{Z}$, there exists a unit exponential random variable E_z such that $T(s,z) = \Theta^{-1}_{T(s,z)}(E_z)$ for all $s \in [0,\infty]$.

Because we never observe two potential outcomes jointly, Assumption 3 is empirically innocuous.

Following Abbring and Van den Berg (2003b), we assume that there is no anticipation of future treatment. This means that current potential integrated hazards do not depend on future treatment enrollment, i.e.

Assumption 4. For all $s \in [0,\infty)$, $z \in \mathcal{Z}$, $\Theta_{T(s,z)}(t) = \Theta_{T(\infty,z)}(t)$ for all $t \leq s$.

Note that Assumptions 3 and 4 imply that $T(s,z) = T(\infty,z)$ on $\{T(\infty,z) \leq s\}$.

We also assume that treatments are only observed if enrollment has taken place before the outcome spell is completed. This is natural in combination with Assumption 4, and natural in many applications.[24] Thus, we now have a random sample $((\widetilde{T}_1, S_1 I(S_1 < \widetilde{T}_1), D_1, I(S_1 < \widetilde{T}_1), Z_1), \ldots, (\widetilde{T}_n, S_n I(S_n < \widetilde{T}_n), D_n, I(S_n < \widetilde{T}_n), Z_n))$ from the joint distribution of $(\widetilde{T}, SI(S < \widetilde{T}), D, I(S < \widetilde{T}), Z)$.

4.2 Which exclusion restrictions can be derived from social experiments?

An equivalent of Assumption 2 for this section would require that (i) $z \mapsto S(z)$ is nontrivial, and (ii) $Z \perp\!\!\!\perp (\{T(s,Z)\}, \{S(z)\})$. A sufficient condition for (ii) is that ITT does not causally affect outcomes directly, i.e. $T(s,z) = T(s,z')$ for all $s \in \mathcal{S}$, $z, z' \in \mathcal{Z}$. With imperfect compliance and dynamic enrollment in

[24] The analysis can be straightforwardly extended to the case that treatments are always observed. A natural symmetric extension of the present model allows outcomes to affect future treatment and imposes that neither future outcomes nor future treatments are anticipated. Alternatively, under the assumption that treatments are not causally affected by outcomes at all, we can allow for for anticipation of future treatment. See Abbring and Van den Berg (2003b) and Abbring (2003) for details.

treatment, randomization is unlikely to ensure such exclusion. To see this, consider a social experiment in which the treatment is a public training program for the unemployed, and the outcome is the unemployment duration. Suppose that agents operate in a continuous-time dynamic environment in which they may affect treatment and outcomes by (i) investing in some (human) capital, and (ii) searching for job and training opportunities. Agents are informed about their ITT status and possible some other predetermined variables (called V below). Otherwise, information accumulates in the obvious way. In this framework, dynamic selection issues may arise because of the agents' private information. Moreover, at the level of an individual agent ITT will affect not only treatment but also outcomes directly (i) before enrollment ("ex ante") if ITT is informative on the arrival of future training (or other) opportunities and this is relevant to search behavior (e.g. the extent of discouraged worker effects), and (ii) after enrollment ("ex post") if ITT affects investments in financial or human capital.

If the ex ante effects of ITT on outcomes are absent, we say that an ex ante exclusion restriction is satisfied. Similarly, we refer to the absence of ex post effects of ITT as ex post exclusion. A violation of ex ante exclusion is very likely, because ITT has to affect actual treatment in order to be a useful instrument in the first place (as in Assumption 2(i)). An ex post exclusion restriction, on the other hand, is not per se inconsistent with ITT being an instrument and may be reasonable in some applications.

A formal statement of such an assumption requires notation that allows us to explicitly control for dynamic selection. To this end, again suppose that all ex ante heterogeneity is captured by a random variable V such that $\{T(s,z)\} \perp\!\!\!\perp \{S(z)\}|V$. Then, a weak version of Assumption 2 that only imposes ex post exclusion is

Assumption 5 (IV with ex post exclusion). (i) $z \mapsto S(z)$ is nontrivial, (ii) $Z \perp\!\!\!\perp (\{T(s,z)\}, \{S(z)\}, V)$, and (iii) for all $s \in [0,\infty)$, and $z, z' \in \mathcal{Z}$,

$$\theta_{T(s,z)}(t|V) = \theta_{T(s,z')}(t|V) \quad \text{almost surely,} \quad \text{for all } t > s. \tag{16}$$

Note that *ex ante* exclusion would require equation (16) to hold for all $t \leq s$.

4.3 Non-parametric approach

4.3.1 Identifiability

Suppose that $\Pr(S = T) = 0$. Following Abbring and Van den Berg (2003b) note that a large data set would provide[25]

$$\begin{aligned} &Q_{S;Z=z}(t,s) := \Pr(T > t, S > s, T > S | Z = z) \text{ and} \\ &Q_{T;Z=z}(t) := \Pr(T > t, T < S | Z = z) \end{aligned} \tag{17}$$

for all $(t,s) \in \mathbb{R}^2_+$ and $z = 0, 1$. These are the sub-survival-functions of (T,S) and T on $Z = z$ for the subpopulations with respectively $T > S$ and $T < S$.

We are interested in inference on (i) causal ITT effects, and (ii) causal effects of actual treatment on outcomes. By Assumption 5, the former are the ex ante effects of ITT on the treatment process, as embodied in $z \mapsto S(z)$, and the ex ante effects of ITT on outcomes, as in $z \mapsto T(\infty, z)$. The latter are the treatment effects embodied in $s \mapsto T(s,z)$. In general, these effects will depend on the ITT status (z).[26]

First consider identifiability of the ITT effects in (i) . One may tend to believe that, because of randomization of Z, ITT effects can be directly inferred from comparing outcomes and treatments between the subpopulations with $Z = 0$ and $Z = 1$. However, we will now argue that surprisingly little can be learned about ITT effects.

Note that the effects of ITT on the treatment process are only observed (and only relevant) on $[0,T]$. Moreover, by Assumption 5, ITT only possibly causally affects the outcome process on $[0,S]$. Intuitively, we should therefore learn about ITT effects from data on the "identified minimum" of (T,S), i.e. the smallest of T and S joint with the identity of this smallest duration, on $Z = 0$ and $Z = 1$. The distribution of this identified minimum on $Z = z$ is fully characterized by $(Q^0_{S;Z=z}, Q_{T;Z=z})$, with $Q^0_{S;Z=z}(\cdot) := Q_{S;Z=z}(-\infty, \cdot)$ (Tsiatis, 1975). We can think of such data as being generated by a competing risks model that is embedded in our model, and in which one risk is enrollment in the treatment (which terminates the ex ante effects of ITT on outcomes) and the other risk is realized by the outcome transition (which terminates observability of the effects of ITT on the treatment process).

[25] Note that simple random censoring does not matter for identification, provided that some straightforward support conditions hold.

[26] We can avoid such a dependence by focusing on $\theta_{T(s,z)}(\cdot|V)$ on (s,∞) (where $\theta_{T(s,0)}(\cdot|V) = \theta_{T(s,1)}(\cdot|V)$ by Assumption 5).

If $\{S(z)\} \perp\!\!\!\perp \{T(s,z)\}$ (randomized assignment) then $(Q^0_{S;Z=z}, Q_{T;Z=z})$ is consistent with an independent risks model in which the outcome risk has survival function $\overline{F}_{T(\infty,z)}$ and the treatment risk has survival function $\overline{F}_{S(z)}$. Because independent risks models are identified (e.g. Cox, 1962), $\overline{F}_{T(\infty,0)}$ and $\overline{F}_{S(0)}$ are identified from $(Q^0_{S;Z=0}, Q_{T;Z=0})$, and $\overline{F}_{T(\infty,1)}$ and $\overline{F}_{S(1)}$ are identified from $(Q^0_{S;Z=1}, Q_{T;Z=1})$.

However, here we allow for general dependence of $\{S(z)\}$ and $\{T(s,z)\}$, through a common dependence on the unobservable V. Then we know that $(Q^0_{S;Z=z}, Q_{T;Z=z})$ is consistent with a particular *dependent* risks model in which the outcome risk has marginal survival function $\overline{F}_{T(\infty,z)}$ and the treatment risk has marginal survival function $\overline{F}_{S(z)}$. However, dependent competing risks models are not nonparametrically identified (Cox, 1959, 1962; Tsiatis, 1975). We could find an independent competing risks model that fits the data, but the marginal survival functions of its risks would now typically not be $\overline{F}_{T(\infty,z)}$ and $\overline{F}_{S(z)}$.

Indeed, Peterson's (1976) results imply that $\overline{F}_{T(\infty,z)}$ and $\overline{F}_{S(z)}$ can at best be bounded, as in

Proposition 4. *For given data* $(Q^0_{S;Z=z}, Q_{T;Z=z})$, $\overline{F}_{T(\infty,z)}$ *and* $\overline{F}_{S(z)}$ *satisfy*

$$Q^0_{S;Z=z} + Q_{T;Z=z} \leq \overline{F}_{T(\infty,z)} \leq Q_{T;Z=z} + Q^0_{S;Z=z}(-\infty) \quad \textit{and}$$
$$Q^0_{S;Z=z} + Q_{T;Z=z} \leq \overline{F}_{S(z)} \leq Q^0_{S;Z=z} + Q_{T;Z=z}(-\infty),$$

for $z = 0, 1$. *The bounds are sharp.*

The bounds in Proposition 4 are typically wide, and may overlap across ITT groups even if there are ITT effects. Then, ITT effects cannot even be signed. This would imply that, contrary to typical IV analyses, one cannot infer empirically whether the ITT variable Z has a causal effect on S. Moreover, a priori information on one of the risks is not informative on the marginal distributions of the other risk. More formally, the bounds on either marginal distribution can be attained even if we arbitrarily fix the other marginal distribution.[27]

For example, suppose that $T(\infty, z)$ is exponential with parameter μ_z, $S(z)$ is exponential with parameter ν_z, and that $\{S(z)\} \perp\!\!\!\perp \{T(s,z)\}$. Then, if we do

[27] Peterson's bounds on, for example, $\overline{F}_{T(\infty,z)}$ follow from $\overline{F}_{T(\infty,z)} = Q_{T;Z=z} + Q^*_{T;Z=z}$, where $Q^*_{T;Z=z}(t) := \Pr(T > t, S < T | Z = z)$. We know $Q_{T;Z=z}$. Non-tight bounds on $\overline{F}_{T(\infty,z)}$ arise because we only know that $Q^0_{S;Z=z} \leq Q^*_{T;Z=z} \leq Q^0_{S;Z=z}(\infty)$. Now, suppose that we do not only know $Q^0_{S;Z=z}$ and $Q_{T;Z=z}$, but also $\overline{F}_{S(z)}$. This is equivalent to also knowing the marginal sub-survival-functions $Q^*_{S;Z=z}$ of S on $\{S > T\}$. Given $Q^0_{S;Z=z}$, $Q^*_{S;Z=z}$ is clearly not informative on the sub-survival-functions $Q^*_{T;Z=z}$ of T on $\{S < T\}$. Therefore, this additional information cannot be used to tighten the bounds on $\overline{F}_{T(\infty,z)}$.

not assume or know that $\{S(z)\} \perp\!\!\!\perp \{T(s,z)\}$, we have to settle for Proposition 4's bounds,

$$\exp\left[-(\mu_z+\nu_z)t\right] \leq \overline{F}_{T(\infty,z)}(t) \leq \frac{\mu_z}{\mu_z+\nu_z}\exp\left[-(\mu_z+\nu_z)t\right] + \frac{\nu_z}{\mu_z+\nu_z} \quad \text{and}$$
$$\exp\left[-(\mu_z+\nu_z)t\right] \leq \overline{F}_{S(z)}(t) \leq \frac{\nu_z}{\mu_z+\nu_z}\exp\left[-(\mu_z+\nu_z)t\right] + \frac{\mu_z}{\mu_z+\nu_z},$$

for $z = 0, 1$. Now let $\mu_0 = \nu_0 = 1$ and $\mu_1 = \nu_1 = 2$. This case could e.g. arise if T is an unemployment duration, S is the (unforeseeable) duration at which unemployment benefits are reduced, and the ITT concerns assignment of different rates of benefits reduction (e.g. Abbring, Van den Berg and Van Ours, 1997). Then, agents in the group with the higher rate of benefits reduction typically respond by increasing their unemployment exit rate. In this numerical example, the bounds reduce to

$$l_0(t) := \exp(-2t) \leq \overline{F}_{T(\infty,0)}(t), \overline{F}_{S(0)}(t) \leq \frac{1}{2}\exp(-2t) + \frac{1}{2} := u_0(t) \quad \text{and}$$
$$l_1(t) := \exp(-4t) \leq \overline{F}_{T(\infty,1)}(t), \overline{F}_{S(1)}(t) \leq \frac{1}{2}\exp(-4t) + \frac{1}{2} := u_1(t).$$

Now note that $l_0(0) = u_0(0) = l_1(0) = u_1(0) = 1$ and, more importantly, that $l_1 < l_0 < u_1 < u_0$ on $(0, \infty)$. Thus, even though the ITT effects are substantial, the bounds for the two ITT groups overlap and ITT effects cannot be signed.

To gain some intuition for this result, consider the following heuristic approach to inferring ITT effects on outcomes.[28] Consider the "crude" ex ante outcome hazard on $Z = z$,

$$\theta_T(t|S \geq t, Z = z) = \frac{-Q'_{T;Z=z}}{Q_{T;Z=z} + Q_{S;Z=z}}.$$

The fact that Z has been randomly assigned may lead one to believe that contrasts of this crude hazard between $Z = 0$ and $Z = 1$ are informative on (ex ante) ITT effects on outcomes. However, note that

$$\theta_T(t|S \geq t, Z = z) = f_T(t|S(z) \geq t, T(\infty, z) \geq t, Z = z),$$

so that for $t > 0$ a comparison of the crude hazard between the subpopulations with $Z = 0$ and $Z = 1$ does not only reveal causal ITT effects on outcomes, but also differences in unobservable characteristics between the subpopulations

$$\{S(0) \geq t, T(\infty, 0) \geq t, Z = 0\} \quad \text{and} \quad \{S(1) \geq t, T(\infty, 1) \geq t, Z = 1\}.$$

[28] Note that the argument symmetrically applies to the effects on treatment.

These differences will in particular exist because $z \mapsto S(z)$ is nontrivial by Assumption 5, so that $S(0) \geq t$ and $S(1) \geq t$ will in general select different subpopulations.

These selection effects disappear in the limit $t \downarrow 0$, but in the fully nonparametric case this cannot be exploited. If we impose additional smoothness on the embedded competing risks model, then ITT effects (or, at least, their sign) can be identified from behavior of the crude hazard near 0. For example, if we impose an multivariate MPH structure, then individual ITT effects can be identified (Abbring and Van den Berg, 2003a). Such semiparametric approaches are discussed in Subsection 4.4.

So far, we have focused on identification of ITT effects from the embedded competing risks data. The data on the residual outcome durations after enrollment in treatment are also informative on ITT effects, but in a very limited way. By Assumption 5, there are no direct causal effects of ITT on outcome hazards after treatment enrollment. Thus, any effects of ITT in the data should be due to indirect selection effects (that is, effects on the distribution of V among survivors). Such effects can only arise if there are ITT effects on outcomes, or, because $z \mapsto S(z)$ is nontrivial by Assumption 5, if $\{S(z)\}$ and $\{T(s,z)\}$ are dependent through joint dependence on the unobservable V. Note that this information cannot be used to further disentangle ITT effects on treatment and outcomes and selection effects. Information on ex post ITT effects can however be used to test for selection effects in general. In the next subsection, we explore this further.

Finally, note that the scope for nonparametric identifiability of ex post effects of treatment enrollment on outcomes is very limited. The case for identification is worse than in the similar single-spell case studied by Abbring and Van den Berg (2003b). This suggests that substantial semiparametric structure is needed to ensure identifiability of the ex post effects. Again, this is discussed in Subsection 4.4.

In conclusion, even though ITT is randomly assigned, not much can be learned about either (*i*) ex ante ITT effects or (*ii*) ex post treatment effects without imposing semiparametric structure. This means that results based on actual social (or, for that sake, laboratory) experiments depend on the chosen semiparametric structure.

4.3.2 Testing

The previous subsection suggests that Assumption 5 does not permit a nonparametric structural analysis in which ex ante ITT effects and ex post treatment

effects are identified or meaningfully bounded. Nevertheless, a reduced-form ITT analysis would still be of interest if the ITT itself is a public-policy instrument. However, the equivalence of the null hypothesis of no treatment effects and the null hypothesis of no ITT effects breaks down in this case. Thus, unlike in the second main case, a reduced-form analysis of ITT effects is not very informative on the effects of actual training.

We have also argued that, under Assumption 5, information on ex post ITT effects can be used to test for selection effects.[29] This suggests two approaches to testing. First, we may be able to test for the sign of overall causal ITT effects. Second, we can nonparametrically test for ex post ITT effects. This is informative on the presence of heterogeneity that leads to nonrandom inflow into the "ex post" state. If the null hypothesis of no ex post ITT effects is accepted then one may proceed by estimating models that ignore such heterogeneity.

4.4 Semiparametric approach

4.4.1 Bounds

Recent results by Bond and Shaw (2003) can to some extent be used to sign ITT effects in a fairly general class of models. Suppose that there exist increasing functions ξ_S and ξ_T such that $(S(0), T(\infty, 0))$ equals $(\xi_S(S(1)), \xi_T(T(\infty, 1)))$ in distribution. For example, in the bivariate MPH model

$$\Pr\left(S(z) > s, T(\infty, z) > t\right) = \mathcal{L}\left(\Lambda_{S(z)}(s), \Lambda_{T(\infty,z)}(t)\right) \qquad (z = 0, 1),$$

[29] The line of reasoning underlying this claim can be related to tests for selection effects in the event-history literature. First, consider the test by Van den Berg and Van Ours (1997) on selection effects due to unobserved heterogeneity in duration analysis. If the value of an exogenous time-varying explanatory variable in the first period is related to the hazard rate in the second period then this indicates such selection effects. Second, consider tests on selection effects in demographic mortality analysis where cohort effects may be present (Vaupel and Yashin, 1985, Lindeboom, Portrait and Van den Berg, 2003). The mortality rate at advanced ages may be related to indicators of childhood conditions early in life. This can be due to a causal effect (e.g., bad childhood conditions result in damage to organs, leading to higher mortality at advanced ages). It can also be due to a selection effect: under adverse childhood conditions, only the children with the "best" unobserved characteristics survive. Under additional semiparametric structure, one may distinguish between these hypotheses by examining the sign of the observed correlation. Third, in Abbring, Chiappori and Pinquet (2003)'s model of car-insurance claims under moral hazard and experience rating, individual claim propensities are only causally affected by the past occurrence of claims, but not their timing. Any variation of observed claim intensities with the timing of past claims conditional on their occurrence should therefore be due to selection effects. This is used to identify the sign of the occurrence-dependence effects under additional semiparametric structure on the model.

with $\mathcal{L}$ the bivariate Laplace transform of the mixing distribution, this holds for $\xi_S = \Lambda_{S(0)}^{-1} \circ \Lambda_{S(1)}$ and $\xi_T = \Lambda_{T(\infty,0)}^{-1} \circ \Lambda_{T(\infty,1)}$. Thus, this structure includes the MPH competing risks model of Abbring and Van den Berg (2003a). It also includes the more general model studied by Heckman and Honoré (1990).

Bond and Shaw call (ξ_S, ξ_T) a "covariate-time transformation".[30] This covariate-time transformation is all we need to know to rank the (ex ante) marginal potential-outcome and potential-treatment distributions in terms of first-order stochastic dominance. For example, if $\xi_S(t) = t$ for all t then $\overline{F}_{S(0)} = \overline{F}_{S(1)}$. If $\xi_S(t) > t$ for all t, on the other hand, then $\overline{F}_{S(1)} < \overline{F}_{S(0)}$, etcetera. Define $\overline{Q}_{S;Z=z}(t) := \Pr(S \leq t, T > S)$, $\overline{Q}_{T;Z=z}(t) := \Pr(T \leq t, T < S)$, and $F_{Z=z}(t) := \Pr(S \leq t, T \leq t)$, for $t \geq 0$.

Bond and Shaw's results immediately imply

Proposition 5. *Suppose that either $\xi_S \leq \xi_T$ or $\xi_S \geq \xi_T$. Then, either*

$$\begin{aligned} F_{Z=0}^{-1} \circ F_{Z=1} &\leq \xi_S \leq \overline{Q}_{S;Z=0}^{-1} \circ \overline{Q}_{S;Z=1} \quad \textit{and} \\ \overline{Q}_{T;Z=0}^{-1} \circ \overline{Q}_{T;Z=1} &\leq \xi_T \leq F_{Z=0}^{-1} \circ F_{Z=1} \end{aligned}$$

or

$$\begin{aligned} F_{Z=0}^{-1} \circ F_{Z=1} &\geq \xi_S \geq \overline{Q}_{S;Z=0}^{-1} \circ \overline{Q}_{S;Z=1} \quad \textit{and} \\ \overline{Q}_{T;Z=0}^{-1} \circ \overline{Q}_{T;Z=1} &\geq \xi_T \geq F_{Z=0}^{-1} \circ F_{Z=1}. \end{aligned}$$

These bounds are tight if $\xi_S = \xi_T$.

Proof. See Appendix 3. □

Note that either the lower bound on ξ_S coincides with the upper bound on ξ_T or the upper bound on ξ_S equals the lower bound on ξ_T. As a consequence, either the potential-treatment distributions or the potential-outcome distributions (or both) can be ranked.

4.4.2 Point-identification

Richer point-identification results can be derived if we are willing to impose further structure. In line with the semiparametric models in previous sections, which were based on the MPH model, we consider the following model framework,

$$\theta_{T(s,z)}(t|X,V) \;=\; \lambda_1(t)\, V_1\, \gamma^{\mathrm{I}(t>s)}\, e^{\beta_1 X + \eta_1 z\, \mathrm{I}(t \leq s)}$$

[30] More precisely, they call $t \mapsto ((t, \xi_S), (t, \xi_T))$ a covariate-time transformation. For both risks, $t \mapsto t$ is the (normalized) time-transformation for the first covariate value ($z = 0$).

$$\theta_{S(z)}(t|X,V) \;=\; \lambda_2(t)\; V_2\; e^{\beta_2 X+\eta_2 z},$$

where $V := (V_1, V_2)$. This is the Timing-of-Events model (Abbring and Van den Berg, 2003b) augmented with a (binary) ITT variable. The ex ante effect on outcomes is represented by η_1. An ex ante exclusion restriction would impose $\eta_1 = 0$. The parameter η_2 captures the causal effect of z on S. If Z is degenerate at say $Z = 0$ then η_1 and η_2 are not identified, and the model reduces to the Timing-of-Events model. The parameter γ represents the ex post effect.

The identification proofs of Abbring and Van den Berg (2003b) can be straighforwardly adapted to prove identification of the augmented Timing-of-Events model.[31]

Proposition 6. *Consider a Timing-of-Events model that is augmented by a binary treatment assignment indicator Z and where we do not impose ex ante exclusion restrictions. If this indicator is exogenous (e.g. randomized) then the model is identified.*

Some comments are in order. First, we can identify substantially more than Abbring and Van den Berg (2003b). In particular, we can now identify the ex ante effect on the outcome hazard rate, and we can now identify the ex post effect in deviation from the hazard rate for those who are assigned to a particular z, e.g. to the control group ($Z = 0$).

Second, the model is heavily over-identified.

Third, notice that for the first time in this paper we require a specification for the distribution of S as a function of its determinants (here Z, X and V).

5 Conclusion

Social experiments in which the outcome of interest is a duration variable are more difficult to analyze than social experiments with time-independent outcomes. First, the outcome may be censored. Second, the randomization occurs at time zero but the composition of survivors changes over time in different ways in the treatment and control groups. The paper studies the three most important benchmark cases, distinguished by whether treatment is immediate or not and whether compliance is perfect or not.

In the intermediate case of imperfect compliance and immediate treatment, one can make nonparametric inferences on local average treatment effects on

[31] In particular, the ITT is simply treated as another regressor, independent of V by randomization, in the embedded competing-risks model.

survival probabilities, and we derive the relevant asymptotic results. To infer average effects on hazard rates, one has to resort to semiparametric models. As a by-product, the paper establishes identification of MPH models with a binary endogenous regressor and a binary instrumental variable. In the most complicated case (imperfect compliance, later treatment), nonparametric analyses are not very informative on (ex ante) effects on survival before actual treatment (and their signs), nor on ex post effects. Again, semiparametric models provide identification, and the paper establishes identification of Timing-of-Events models augmented with ex ante effects.

The results of this paper lead to the conclusions that (1) while it is possible to make nonparametric inferences on additive effects on survival probabilities, the degree to which these are informative decreases with the complexity of the case at hand, and (2) to study the more interesting average effects on individual hazard rates and conditional exit probabilities one needs a semiparametric structure, despite the randomization at time zero. In sum, despite the randomization not much can be learned without a semiparametric structure. The underlying intuition is that hazard rates condition on survival up to a positive duration, so that a nonparametric comparison of hazard rates is affected by selection, whereas survival probabilities concern the population at time zero.[32]

One approach to these problems is to use more complex dynamic experimental designs than the simple design considered in this paper. However, the main arguments and results carry over to such designs in which randomization takes place repeatedly at discrete (possibly random) times. If actual treatment enrollment takes place at the same times, an extension of our first two cases applies. If agents enroll in treatment more frequently, say continuously, then an extension of our third case applies. With continuous outcomes and treatment enrollment, repeated randomization can at best reduce the inference problems, e.g. by narrowing bounds on some parameters of interests, but not solve these problems altogether.

The results have some implications for the design of social experiments and laboratory experiments. First, experimental inference requires semiparametric structure, so results depend on the chosen semiparametric structure. Second, it is useful to collect as many explanatory variables on the subjects as possible, for two reasons: it serves to reduce the magnitude of unobserved heterogeneity,

[32]Intuitively, the nonparametric hazard rates at zero are not yet affected by selection, and semiparametric identification involves an extrapolation of the treatment effect on the hazard at zero to positive durations. This does of course not mean that semiparametric estimation results are completely driven by extremely short durations.

and it facilitates the semiparametric inference. Third, it is advisable to minimize imperfect compliance.

Appendix

Appendix 1 Proof of equation (6) in Subsection 2.4

Proof. Note that the middle and right-hand sides of equation (6) equal $\theta^*_{T(1)}/\theta^*_{T(0)}$. We now turn to the left-hand side. Let $\mathcal{L}_V$ be the Laplace transform of F_V. There holds that

$$\overline{F}_T(t|S=s) = \mathcal{L}_V\left(\Theta^*_{T(s)}(t)\right) \tag{18}$$

Therefore,

$$\frac{\theta_T(t|S=1)/\theta_T(t|S=0)}{\theta^*_{T(1)}/\theta^*_{T(0)}} = \frac{\mathcal{L}'_V(\Theta^*_{T(1)}(t))/\mathcal{L}_V(\Theta^*_{T(1)}(t))}{\mathcal{L}'_V(\Theta^*_{T(0)}(t))/\mathcal{L}_V(\Theta^*_{T(0)}(t))} \tag{19}$$

Note that the right-hand side of (19) equals 1 if F_V is degenerate. Now take any $t > 0$. We need to prove that the right-hand side of (19) is smaller than 1 if $\Theta^*_{T(1)}(t) > \Theta^*_{T(0)}(t)$ and F_V is non-degenerate. Clearly, this is established if $-\mathcal{L}'_V/\mathcal{L}_V$ is decreasing. The latter follows because the derivative of $-\mathcal{L}'_V/\mathcal{L}_V$ at $\Theta^*_{T(s)}(t)$ equals minus the variance of $V|T \geq t, S = s$. □

Appendix 2 Proofs of statements in Section 3

Proof of identification of $\overline{F}_{0;Q}$, $\overline{F}_{1;Q}$, and Δ_Q. Consider the observed events

$$A_{sz} := \{T > t, S = s, Z = z\}$$

for some $t \geq 0$, and $s = 0, 1$ and $z = 0, 1$. Note that

$$\begin{aligned}
A_{00} &= \{T(0) > t, S(0) = 0, S(1) = 1, Z = 0\} \cup \{T(0) > t, S(1) = 0, Z = 0\} \\
A_{01} &= \{T(0) > t, S(1) = 0, Z = 1\} \\
A_{11} &= \{T(1) > t, S(0) = 0, S(1) = 1, Z = 1\} \cup \{T(1) > t, S(0) = 1, Z = 1\} \\
A_{10} &= \{T(1) > t, S(0) = 1, Z = 0\},
\end{aligned}$$

that the sets in the right-hand sides are disjoint, and that the mass of the subpopulation Q considered is $\Pr(Q) = p(1) - p(0)$. It immediately follows that the marginal potential-outcome distributions

$$\begin{aligned}
\overline{F}_{0;Q}(t) &= \frac{\Pr(T > t, S = 0|Z = 0) - \Pr(T > t, S = 0|Z = 1)}{p(1) - p(0)} \quad \text{and} \\
\overline{F}_{1;Q}(t) &= \frac{\Pr(T > t, S = 1|Z = 1) - \Pr(T > t, S = 1|Z = 0)}{p(1) - p(0)}
\end{aligned}$$

on Q are identified (Imbens and Rubin, 1997). □

Proof of Proposition 1. First write

$$\begin{aligned}
\sqrt{n}\left(\widehat{\overline{F}}_{0;Q}-\overline{F}_{0;Q}\right) =& \frac{1-\widehat{p}(0)}{\widehat{p}(1)-\widehat{p}(0)}\sqrt{n}\left(\widehat{\overline{F}}_{T;S=0,Z=0}-\overline{F}_{T;S=0,Z=0}\right) \\
&- \frac{1-\widehat{p}(1)}{\widehat{p}(1)-\widehat{p}(0)}\sqrt{n}\left(\widehat{\overline{F}}_{T;S=0,Z=1}-\overline{F}_{T;S=0,Z=1}\right) \\
&+ \overline{F}_{T;S=0,Z=0}\sqrt{n}\left(\frac{1-\widehat{p}(0)}{\widehat{p}(1)-\widehat{p}(0)}-\frac{1-p(0)}{p(1)-p(0)}\right) \\
&- \overline{F}_{T;S=0,Z=1}\sqrt{n}\left(\frac{1-\widehat{p}(1)}{\widehat{p}(1)-\widehat{p}(0)}-\frac{1-p(1)}{p(1)-p(0)}\right) \quad \text{and} \\
\sqrt{n}\left(\widehat{\overline{F}}_{1;Q}-\overline{F}_{1;Q}\right) =& \frac{\widehat{p}(1)}{\widehat{p}(1)-\widehat{p}(0)}\sqrt{n}\left(\widehat{\overline{F}}_{T;S=1,Z=1}-\overline{F}_{T;S=1,Z=1}\right) \\
&- \frac{\widehat{p}(0)}{\widehat{p}(1)-\widehat{p}(0)}\sqrt{n}\left(\widehat{\overline{F}}_{T;S=1,Z=0}-\overline{F}_{T;S=1,Z=0}\right) \\
&+ \overline{F}_{T;S=1,Z=1}\sqrt{n}\left(\frac{\widehat{p}(1)}{\widehat{p}(1)-\widehat{p}(0)}-\frac{p(1)}{p(1)-p(0)}\right) \\
&- \overline{F}_{T;S=1,Z=0}\sqrt{n}\left(\frac{\widehat{p}(0)}{\widehat{p}(1)-\widehat{p}(0)}-\frac{p(0)}{p(1)-p(0)}\right).
\end{aligned}$$

Next, note that $\sqrt{n}\left(\widehat{\overline{F}}_{T;S=0,Z=z}-\overline{F}_{T;S=0,Z=z}\right)$ $(z=0,1)$, $\sqrt{n}\left(\widehat{\overline{F}}_{T;S=1,Z=z}-\overline{F}_{T;S=1,Z=z}\right)$ $(z=0,1)$, and $\sqrt{n}\left(\widehat{p}(z)-p(z)\right)$ $(z=0,1)$ converge jointly in distribution to $-\overline{F}_{T;S=0,Z=z}\mathbb{G}_{0z}/\sqrt{(1-p(z))\,q_z}$ $(z=0,1)$, $-\overline{F}_{T;S=1,Z=z}\mathbb{G}_{1z}/\sqrt{p(z)q_z}$ $(z=0,1)$, and $\mathcal{N}_z$ $(z=0,1)$, with $\mathcal{N}_1$, $\mathcal{N}_0$, $\mathbb{G}_{11}$, $\mathbb{G}_{01}$, $\mathbb{G}_{10}$, and $\mathbb{G}_{00}$ mutually independent with distributions as given in Proposition 1. Furthermore, $\widehat{p}(z)\xrightarrow{\mathcal{D}}p(z)$ by the law of large numbers. The claimed result follows from consecutively applying Slutsky's lemma, the delta method, and the continuous-mapping theorem (see e.g. Van der Vaart and Wellner, 1996). □

Proof of Proposition 3. The X variables only play a role in the identification of λ, ϕ and F_V from the outcomes for $Z=0$. We therefore proceed conditional on X, subsume $\phi(X)$ into λ, and suppress X in the notation.

Among agents with $S=0, Z=1$ there holds that

$$\theta_{T;S=0,Z=1}(t|V)=\lambda(t)V \quad \text{and}$$

$$\overline{F}_{T;S=0,Z=1}(t)=\mathcal{L}_{V;S=0,Z=1}\left(\int_0^t\lambda(\tau)d\tau\right),$$

with $\mathcal{L}_{V;S=0,Z=1}$ being the Laplace transform of $[V|S=0,Z=1]$. Its argument $\int_0^t\lambda(\tau)d\tau$ is an already identified function. Thus, $\mathcal{L}_{V;S=0,Z=1}$ is identified, and therefore $F_{V;S=0,Z=1}$.

From

$$F_V(v)=F_{V;Z=1}(v)=p(1)F_{V;S=1,Z=1}(v)+[1-p(1)]F_{V;S=0,Z=1}(v)$$

we can now also identify $F_{V;S=1,Z=1}$, since all other quantities in this equation are observed or identified.

Among agents with $S=1, Z=1$ there holds that

$$\theta_{T;S=1,Z=1}(t|V)=\lambda(t)\gamma(t)V \quad \text{and}$$

$$\overline{F}_{T;S=1,Z=1}(t) = \mathcal{L}_{V;S=1,Z=1}\left(\int_0^t \lambda(\tau)\gamma(\tau)d\tau\right),$$

with $\mathcal{L}_{V;S=1,Z=1}$ being the Laplace transform of $[V|S=1,Z=1]$. Because this is already identified, and the left-hand side is observed, we can back out the argument of this Laplace transform. Since λ is also already (almost everywhere) identified, it follows that γ is (almost everywhere) identified. □

Appendix 3 Proofs of statements in Section 4

Proof of Proposition 5. This proof follows as a special case of Bond and Shaw (2003). First, suppose that $\xi_S \geq \xi_T$. Then,

$$\begin{aligned}
\overline{Q}_{S;Z=1}(t) &= \Pr\left(S(1) \leq t, T(\infty,1) > S(1)\right) \\
&= \Pr\left(S(0) \leq \xi_S(t), \xi_T^{-1}(T(\infty,0)) > \xi_S^{-1}(S(0))\right) \\
&\geq \Pr\left(S(0) \leq \xi_S(t), T(\infty,0) > S(0)\right) = \overline{Q}_{S;Z=0}(\xi_S(t)),
\end{aligned}$$

$$\begin{aligned}
\overline{Q}_{T;Z=1}(t) &= \Pr\left(T(\infty,1) \leq t, T(\infty,1) < S(1)\right) \\
&= \Pr\left(T(\infty,0) \leq \xi_T(t), \xi_T^{-1}(T(\infty,0)) < \xi_S^{-1}(S(0))\right) \\
&\leq \Pr\left(T(\infty,0) \leq \xi_T(t), T(\infty,0) < S(0)\right) = \overline{Q}_{T;Z=0}(\xi_T(t)),
\end{aligned}$$

and

$$\begin{aligned}
F_{Z=0}(\xi_T(t)) \leq F_{Z=1}(t) &= \Pr(T(\infty,1) \leq t, S(1) \leq t) \\
&= \Pr\left(T(\infty,0) \leq \xi_T(t), S(0) \leq \xi_S(t)\right) \leq F_{Z=0}(\xi_S(t)).
\end{aligned}$$

Taken together, these inequalities imply that

$$\begin{aligned}
F_{Z=0}^{-1} \circ F_{Z=1} \quad &\leq \xi_S \leq \quad \overline{Q}_{S;Z=0}^{-1} \circ \overline{Q}_{S;Z=1} \quad \text{and} \\
\overline{Q}_{T;Z=0}^{-1} \circ \overline{Q}_{T;Z=1} \quad &\leq \xi_T \leq \quad F_{Z=0}^{-1} \circ F_{Z=1}.
\end{aligned}$$

Similarly, if $\xi_S \leq \xi_T$ we have that

$$\begin{aligned}
F_{Z=0}^{-1} \circ F_{Z=1} \quad &\geq \xi_S \geq \quad \overline{Q}_{S;Z=0}^{-1} \circ \overline{Q}_{S;Z=1} \quad \text{and} \\
\overline{Q}_{T;Z=0}^{-1} \circ \overline{Q}_{T;Z=1} \quad &\geq \xi_T \geq \quad F_{Z=0}^{-1} \circ F_{Z=1}.
\end{aligned}$$

□

References

Abadie, A. (2002), "Bootstrap tests for distributional treatment effects in instrumental variable models", *Journal of the American Statistical Association*, 97, 284–292.

Abadie, A., J. Angrist, and G. Imbens (2002), "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings", *Econometrica*, 70, 91–117.

Abbring, J.H. (2003), "Dynamic econometric program evaluation", Working paper, Free University, Amsterdam [Paper prepared for the H. Theil Memorial Conference, Amsterdam, 16-18 August 2002].

Abbring, J.H., P.A. Chiappori, and J. Pinquet (2003), "Moral hazard and dynamic insurance data", *Journal of the European Economic Association*, 1, 767–820.

Abbring, J.H. and G.J. van den Berg (2003a), "The identifiability of the mixed proportional hazards competing risks model", *Journal of the Royal Statistical Society Series B*, 65, 701–710.

Abbring, J.H. and G.J. van den Berg (2003b), "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 (1997), "The effect of unemployment insurance sanctions on the transition rate from unemployment to employment", Working paper, Tinbergen Institute, Amsterdam.

Andersen, P.K., Ø. Borgan, R.D. Gill, and N. Keiding (1993), *Statistical Models Based on Counting Processes*, Springer-Verlag, New York.

Bijwaard, G. (2003), "Instrumental variable estimation for duration data", Working paper, Erasmus University, Rotterdam.

Bijwaard, G. and G. Ridder (2003), "Correcting for selective compliance in a re-employment bonus experiment", *Journal of Econometrics*, forthcoming.

Black, D.A., J.A. Smith, M.C. Berger, and B.J. Noel (2003), "Is the threat of reemployment services more effective than the services themselves? Evidence from random assignment in the UI system", *American Economic Review*, 93, 1313–1327.

Bond, S.J. and J.E.H. Shaw (2003), "Bounds on the covariate-time transformation for competing-risks survival analysis", Mimeo, Department of Statistics, University of Warwick, Coventry, UK.

Chesher, A. (2003), "Semiparametric identification in duration models", Working paper, UCL, London.

Cox, D.R. (1959), "The analysis of exponentially distributed life-times with two types of failure", *Journal of the Royal Statistical Society Series B*, 21, 411–421.

Cox, D.R. (1962), *Renewal Theory*, Methuen, London.

Cox, D.R. (1972), "Regression models and life tables (with discussion)", *Journal of the Royal Statistical Society Series B*, 34, 187–220.

Dabrowska, D.M. (1987), "Non-parametric regression with censored survival time data", *Scandinavian Journal of Statistics*, 14, 181–197.

Eberwein, C., J.C. Ham, and R.J. 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.

Elbers, C. and G. Ridder (1982), "True and spurious duration dependence: The identifiability of the proportional hazard model", *Review of Economic Studies*, 49, 403–410.

Ferrall, C. (2002), "Estimation and inference in social experiments", Working paper, Queen's University, Kingston.

Fleming, T.R. and D.P. Harrington (1991), *Counting Processes and Survival Analysis*, Wiley, New York.

Ham, J.C. and R.J. LaLonde (1996), "The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training", *Econometrica*, 64, 175–205.

Heckman, J.J. (1997), "Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations", *Journal of Human Resources*, 32, 441–462.

Heckman, J.J. and B.E. Honoré (1989), "The identifiability of the competing risks model", *Biometrika*, 76, 325–330.

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, editors, *Handbook of Labor Economics, Volume III*, North-Holland, Amsterdam.

Heckman, J.J. and C.R. Taber (1994), "Econometric mixture models and more general models for unobservables in duration analysis", *Statistical Methods in Medical Research*, 3, 279–302.

Horowitz, J.L. (1999), "Semiparametric estimation of a proportional hazard model with unobserved heterogeneity", *Econometrica*, 67, 1001–1029.

Imbens, G.W. and J.D. Angrist (1994), "Identification and estimation of local average treatment effects", *Econometrica*, 62, 467–475.

Imbens, G.W. and D.B. Rubin (1997), "Estimating outcome distributions for compliers in instrumental variables models", *Review of Economic Studies*, 64, 555–574.

Kortram, R.A., A.J. Lenstra, G. Ridder, and A.C.M. van Rooij (1995), "Constructive identification of the mixed proportional hazards model", *Statistica Neerlandica*, 49, 269–281.

Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.

Lenstra, A.J. and A.C.M. van Rooij (1998), "Nonparametric estimation of the mixed proportional hazards model", Working paper, Free University, Amsterdam.

Lindeboom, M., F. Portrait, and G.J. van den Berg (2003), "Individual mortality and macro-economic conditions from birth to death", Working paper, Tinbergen Institute, Amsterdam.

Manski, C.F. (1997), "Monotone treatment response", *Econometrica*, 65, 1311–1334.

Meyer, B.D. (1996), "What have we learned from the Illinois reemployment bonus experiment?", *Journal of Labor Economics*, 14, 26–51.

Neyman, J. (1923), "On the application of probability theory to agricultural experiments. Essay on principles", *Roczniki Nauk Rolniczych*, 10, 1–51 [in Polish; edited and translated version of Section 9 by D.M. Dabrowska and T.P Speed (1990), *Statistical Science*, **5**, 465–472].

Peterson, A.V. (1976), "Bounds for a joint distribution function with fixed subdistribution functions: Application to competing risks", *Proceedings of the National Academy of Sciences*, 73, 11–13.

Ridder, G. (1990), "The non-parametric identification of generalized accelerated failure-time models", *Review of Economic Studies*, 57, 167–182.

Robins, J.M. and A.A. Tsiatis (1991), "Correcting for non-compliance in randomized trials using rank preserving structural failure time models", *Communications in Statistics–Theory and Methods*, 20, 2609–2631.

Rosenzweig, M.R. and K.I. Wolpin (2000), "Natural "natural experiments" in economics", *Journal of Economic Literature*, 38, 827–874.

Van den Berg, G.J. (2001), "Duration models: Specification, identification, and multiple durations", in J.J. Heckman and E. Leamer, editors, *Handbook of Econometrics, Volume V*, North Holland, Amsterdam.

Van den Berg, G.J. and B. van der Klaauw (2001), "Counseling and monitoring of unemployed workers: Theory and evidence from a controlled social experiment", Working paper, Free University and Tinbergen Institute, Amsterdam.

Van den Berg, G.J. and J.C. van Ours (1997), "Eyeballing state dependence and unobserved heterogeneity in aggregate unemployment duration data", *Research in Labor Economics*, 16, 369–394.

Van der Vaart, A.W. and J.A. Wellner (1996), *Weak Convergence and Empirical Processes*, Springer, New York.

Vaupel, J.W. and A.I. Yashin (1985), "The deviant dynamics of death in heterogeneous populations", in N.B. Tuma, editor, *Sociological Methodology 1985*, Jossey-Bass, San Francisco.

Visser, M. (1996), "Nonparametric estimation of the bivariate survival function with an application to vertically transmitted AIDS", *Biometrika*, 83, 507–518.