Survey Non-Response and Unemployment Duration

Gerard J. van den Berg
Maarten Lindeboom
Peter J. Dolton

September 2004

Survey Non-Response and Unemployment Duration

Gerard J. van den Berg
Free University Amsterdam, IFAU-Uppsala,
Tinbergen Institute, CEPR, INSEE-CREST and IZA Bonn

Maarten Lindeboom
Free University Amsterdam,
Tinbergen Institute and IZA Bonn

Peter J. Dolton
University of Newcastle-upon-Tyne,
CEP, London School of Economics and IZA Bonn

ABSTRACT

Survey Non-Response and Unemployment Duration[*]

Social surveys are often used to estimate unemployment duration distributions. Survey non-response may then cause a bias. We study this using a unique dataset that combines survey information of individual workers with administrative records of the same workers. The latter provide information on unemployment durations and personal characteristics of all survey respondents and non-respondents. We develop a method to empirically distinguish between two explanations for a bias in results based on only survey data: (1) selectivity due to related unobserved determinants of unemployment durations and non-response, and (2) a causal effect of a job exit on non-response. The latter may occur even in fully homogeneous populations. The methodology exploits variation in the timing of the duration outcome relative to the survey moment. The results show evidence for both explanations. We discuss implications for standard methods to deal with non-response bias.

JEL Classification: J64, C41, C81, C23, C14, C42

Keywords: non-response bias, unemployment measurement, hazard rate, sample selection, event history

Corresponding author:

Gerard J. van den Berg
Department of Economics
Free University of Amsterdam
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands
Email: gjvdberg@xs4all.nl

1 Introduction

Non-response is a commonly encountered phenomenon in social surveys. Whether non-response affects the statistical analysis of the survey data depends on the variables one is interested in. If one aims to use the survey to estimate the fraction in the population with a certain characteristic, then a systematically high or low non-response among those who have this characteristic biases the estimate. If one aims to estimate a model, and the only difference between the ultimate sample of respondents and the intended survey sample is in the distribution of explanatory variables on which one conditions in the analysis, then non-response does not affect the estimation results. This requires that non-response behaviour is unrelated to any unobserved determinants of the endogenous variable of interest (i.e. the variable whose values one aims to explain in the analysis). Indeed, non-response must be unrelated to measurement errors in the data on the latter variable.

If non-response is related to unobserved determinants of the variable of interest, and if this is ignored,[1] then in general the estimation results are inconsistent. Empirical studies based on social survey data do not pay much attention to such non-response, for the reason that it is felt that there is nothing one can do about it.[2] Most studies merely provide the non-response percentage. To obtain an indication of systematic non-response, studies sometimes compare the marginal distributions of explanatory or endogenous variables among respondents to those in census data. The differences between the marginal distributions can be used to construct non-response weights for the respondents, giving a higher weight to respondents who seem to be underrepresented. The underlying idea of this approach is that if the marginal distributions among respondents are similar to the corresponding population distributions then, hopefully, the conditional distributions of the endogenous variable given the explanatory variables are also similar.

This paper pursues the character of non-response in more detail, with an unusually informative sampling frame that combines survey information of individual workers with administrative records of the same workers. From administrative records, a random sample of unemployed workers in the UK was taken, and a survey was conducted among these workers six months after that, with follow-up surveys taking place at regular time intervals after the first survey interview. The survey focuses on labour market outcomes and personal characteristics. We study non-response at the first survey interview which for simplicity we often refer to as "the" survey. The survey response rate is about 56%. It should be stressed from the outset that we are only concerned with unit non-response in the survey,

---

[1] See Diggle and Kenward (1994) for clarification of "ignorable" and "non-ignorable" drop out.

[2] As Horowitz and Manski (1998) state, "[With nonresponse,] the only way to identify population statistics is to make assumptions that determine the distribution of the missing data. A fundamental problem of empirical analysis is that such assumptions are untestable." There is of course a strong analogy to empirical analysis based on a truncated sample in the presence of self-selection.

and not with item non-response. The administrative records provide unemployment durations of all survey respondents and non-respondents, and they also contain some personal characteristics of all these individuals.

We focus on the relation between the duration of unemployment and non-response, and on the effect of such a relation on the survey estimates of the unemployment duration distribution. Clearly, the bias in the estimated unemployment duration distribution due to non-response follows from a comparison of the estimate based on data from respondents to the estimate based on data from the complete sample.[3,4] In the present paper we go substantially further than this. Specifically, we distinguish between two reasons for a relation between non-response on the one hand, and unemployment duration on the other. First of all, job search behaviour and the behaviour towards survey participation may be affected by the same underlying unobserved individual-specific characteristics. An individual with a relative dislike for social contacts may refuse to cooperate with the survey interview and may also be reluctant to apply for a job and/or to be exposed to job search counselling by his case worker. An individual who spends a lot of his time searching for a job may not want to spend time with a survey interview. Badly motivated people may have difficulties finding a job and may be less inclined to participate in a survey, especially when this survey is about job search behaviour and labour market prospects. In sum, the unobserved determinants of unemployment durations and non-response may be related, and this gives rise to a selection effect. The second reason for a relation between non-response and unemployment duration is that the acceptance of a job makes it more difficult for the agency to contact the individual. Job acceptance may entail a movement of the individual to another geographical location - which could easily be out of the scope of the survey. Also, the individual may be away from home more often. These concern a causal effect of a job exit on non-response. The second relation is fundamentally different from the first relation, as the causal effect runs directly from job acceptance to non-response, and this effect does not depend on the presence of unobserved characteristics. In the presence of a causal effect, if one of two identical individuals purely by chance finds a job before the survey

---

[3]For both estimations we use the administrative or register data. In principle, the duration outcomes among respondents who are unemployed at the survey date can be observed from the subsequent survey interviews, but in practice this is hampered by survey attrition.

[4]Conceptually, this approach is remotely related to a number of other studies. Wang, Sedransk and Jinn (1992) use combined survey/administrative data to estimate the relation between firm characteristics and *item* non-response by firms concerning these characteristics. Potthoff, Manton and Woodbury (1993) use the number of callbacks in telephone surveys needed to interview a sample member, in order to correct for the bias due to non-availability, which is a particular type of non-response. Holmes and Schmitz (1996) use information obtained from past business partners of non-respondents in a study of business ownership. Kamionka and Lacroix (2003) estimate treatment effects from survey data where the non-respondents are basically replaced by administrative records from another sample of individuals, who did not participate in the experiment.

2

date, that individual has a higher probability of non-response, and the survey estimates will be biased.[5]

We develop a method to empirically distinguish between two explanations for a bias in survey estimates. The administrative data provide us with random drawings from the bivariate distribution of the duration of unemployment and the binary non-response indicator. Typically, to study the causal effect of unemployment duration on non-response in the presence of selectivity, one would resort to instrumental variable methods, i.e. methods that require a variable that causally affects unemployment duration but does not have a direct causal effect on non-response. Such a variable is typically hard to find. Instead, we exploit the variation in the timing of the duration outcome relative to the survey date. We demonstrate that the shape of the conditional distribution of the unemployment duration given the value of the non-response indicator, evaluated around the survey date, is informative on the presence of a causal effect. If many non-respondents exit unemployment before the survey date and few shortly after, compared to the respondents, then this indicates a causal effect. Similarly, the shape of the conditional duration distribution given non-response, after the survey date, is informative on the presence of a selection effect.[6]

The survey data also include a variable indicating the *type* of survey non-response, from 22 different behavioral categories, for the non-respondents. For example, it indicates whether the individual has moved his residence, or whether he refused to be interviewed. We exploit this information to further study the relation between non-response and unemployment duration. First, we provide a taxonomy of the type of non-response that is tailored for our focus on unemployment duration as the outcome of interest. Certain types of non-response may indicate that the individual has left unemployment (e.g. if the individual has moved). Other types may indicate a selection effect (e.g. if the individual refuses to cooperate with the survey interview).[7] The data allow us to study the effects of personal characteristics on the type of non-response, and, by implication, the difference between the distributions of these characteristics in the full dataset and among respondents. Note that social surveys often do not contain any information on the type of non-response or on characteristics of non-respondents, so that their relation cannot be studied.[8] We also

---

[5]Bring and Carling (1999) interview workers for whom the destination state upon leaving the Swedish employment office register is not reported to the office by the worker. It turns out that a substantial fraction left the register because they found a job. This suggests that exit out of unemployment may be an important explanation of not giving information (although of course attrition from a register is not the same as non-response at a survey).

[6]Studies of attrition in longitudinal survey data often model attrition as a selection problem. Van den Berg, Lindeboom and Ridder (1994) estimate an unemployment duration model where exit to work has a causal effect on the probability of attrition (but selection effects are absent).

[7]One could call these the "reasons of non-response" rather than the "types of non-response", but we use the latter to avoid confusion with the two aforementioned "reasons of non-response bias".

[8]Occasionally, "coversheet" information is collected for non-respondents (see e.g. Brehm, 1987). In

study the relation between the type of non-response and unemployment duration, and the relation between the type of non-response and the reason for non-response bias.

The results suggest a targeting of individuals whose non-response is likely to distort the empirical analysis of individual labour market behaviour with survey data. A particular novelty is that we address the causal effect from the outcome of interest on non-response. The results are of interest for agencies that run surveys as well as for researchers modelling the length of unemployment spells who are not so well endowed with data as in the present example. Moreover, the results in the paper enable an assessment of the usefulness of various popular methods to tackle non-response problems, notably instrumental variable methods and imputation methods. More in general, the paper provides a methodology to address the presence of a causal effect of a duration variable on a binary outcome in the presence of selectivity.

The data used in this paper were originally collected for the purpose of evaluating the effects of a specific treatment, called "Restart", of unemployed individuals. For this reason, the original sample is randomly divided into a treatment and a control group. In previous studies, these data have been used to estimate the treatment effects (see Dolton and O'Neill, 1995, 1996a, 1996b, and O'Neill and Dolton, 2002). In the literature on treatment effects, studies often only have survey data at their disposal (see e.g. Dolton, 1993, for a survey). The present paper therefore also has some relevance for the analysis of treatment effects, as it shows that conclusions based on survey data with non-response may be biased.

The remainder of the paper is organised as follows. The next section describes the data sources, the non-response types, and the way in which non-response varies with explanatory variables. Section 3 considers the non-response bias in the estimated unemployment duration distribution. Section 4 develops the method to distinguish between the two reasons for non-response bias. Section 5 assesses the evidence for both. Section 6 summarizes and concludes.

2 The data

2.1 Data sources

In 1989 the Policy Studies Institute in the UK was commissioned by the UK Employment Service to evaluate the impact of the "Restart" policy program for unemployed workers (see White and Lakey, 1992). At the time, the Restart program consisted of compulsory six-monthly meetings between the unemployed individual and a counsellor of the Employment

---

telephone surveys, the agency also collects the date and time of the interview attempt, and it usually records whether non-response is due to non-availability or due to refusal.

Office, for each unemployment benefits claimant in the UK.[9] During these meetings or "interviews", the counsellor offers advice on job search, and he may place workers in contact with employers or training agencies. If the individual does not attend a Restart interview or is deemed not to be available for work then their case is referred to an adjudication officer and they may be faced with the possibility of having their benefits reduced or suspended. Over the period of July to September 1989 over 270,000 such adjudication decisions were made and in 57% of cases the person's UB was stopped. The main aim of the program is to reduce the amount of time people spend unemployed, and to reduce their dependency on unemployment benefits.

To avoid confusion, it must be stressed from the outset that the Restart interviews are *not* survey interviews. For the purposes of the present paper, the main relevance of the Restart interviews is that the planned date of the first Restart interview (6 months after entry into unemployment) affects the sampling design. In particular, to evaluate the Restart program, a random sample of 8925 unemployed workers was selected around March/April 1989 who would approach their 6th month of unemployment around May/June 1989. Individuals were retained in the sample even if they subsequently did not attend a scheduled Restart interview. The median of the distribution of the Restart interview date is at the end of May 1989.

Every Employment Office throughout Britain was contacted while constructing the sample, in order to eliminate regional biases. Individuals were selected for the sample from the inflow lists, on the basis of their National Insurance (NI) numbers. Of this set, a control group of 582 people was randomly chosen again by means of previously specified NI digit sequences. Members of the control group, although eligible for a Restart interview, were deliberately not offered a Restart interview after the first 6 months of unemployment. The existence of a random control group allows for the evaluation of the impact of the program without having to deal with the issue of self-selection.

For the sample of 8925 individuals, administrative information on a few personal characteristics, such as sex, age, and travel-to-work area, was collected from the Employment Services. The information on an individual's travel-to-work area was linked to the National Online Manpower Information System (NOMIS) data, in order to obtain data on local labour market conditions. In addition, the data are linked to the Joint Unemployment and Vacancies Operating System (JUVOS) Cohort database collected by the Employment Service. The JUVOS data provide accurate administrative records on the claimant's unemployment history from 1982 up to January 1995. In the present study we focus on the unemployment spell that has led to the invitation to the Restart interview. Unfortunately, the administrative data do not record the destination state upon exit out of unemployment. This could be employment, a training programme or simply signing off the claiming of un-

---

[9]In 1989 the Restart programme was relatively large with over 2 million meetings taking place.

employment benefit (to obtain benefits, one needs to register at the Employment Service). However, by comparing the administrative data to the survey data for respondents, O'Neill and Dolton (2002) show that most exits out of unemployment amount to a transition into employment.

After excluding individuals who lacked JUVOS data or travel-to-work area information (180 and 736 individuals, respectively), or whose age was below 16 or above 65 (141 individuals), or whose unemployment duration was substantially longer or shorter than 6 months in May 1989 (37 individuals), we are left with a sample of 8012. Of these, 512 are in the experimental control group. Only 229 out of the 8012 unemployment spells (approximately 3%) are right-censored at the end of the observation window.

About 6 months after the identification of the full sample (i.e., in September/October 1989), a survey organisation (Social and Community Planning Research, or SCPR) conducted a survey of these individuals. It is the non-response to this survey that we are interested in. The survey is intended to supply additional information on background variables and job search behaviour of the individuals. The median of the distribution of the actual survey interview dates among respondents is at the beginning of October 1989. The median of the mixed distribution of actual survey interview dates among respondents and intended survey dates among non-respondents is at the end of September 1989.

At the individual level, the survey is carried out as follows. First, the Employment Office provides the information necessary to locate the sample member. The address is the address given by the sample member for official unemployment related business. Next, the interviewer attempts to establish contact with the sample member him- or herself, to make an appointment for the face-to-face interview. If the attempt does not result in a contact then the interviewer makes another attempt, up to at least four times in total. Different attempts are always made at different days of the week and at different times of the day. The interviewers' earnings depend on the number of actual interviews. There is anecdotal evidence that interviewers often continue to try to establish contact if all four attempts were unsuccessful. After the interview, the interviewer returns the completed response forms by mail to the survey agency. In Subsection 2.3 we relate these stages to the various types of non-response.

Of the original sample of 8925 individuals, 5200 individuals completed the survey. Of the sample of 8012 (see above), 4708 completed the survey. These non-response rates of around 40% are within the expected range. Follow-up surveys (i.e., additional waves of the longitudinal survey) took place at 6 months and at 18 months after the first survey. We do not use those data in this paper. The attrition rate from the first to the second survey amounts to about 50%.

2.2 Survey non-response

Table 1 presents means of variables for the individuals in our sample. The first column of the table gives means of the total sample of 8012. The second column of the table refers to information of the respondents to the survey. The unemployment duration variable is measured from the sample selection date. This date is not exactly known for all individuals in the sample. We approximate it by the first day of the month two months prior to the month in which the Restart interview date takes place. This probably ante-dates the true selection date for some individuals, but it prevents that we obtain negative durations as measured from the selection date to the date of exit out of unemployment. In the empirical analysis we take this into account.

The local unemployment change variable measures the decline in the unemployment rate in % between 1988 and 1990 in the individual's travel to work area (i.e. it equals the average local unemployment rate in % in 1988 minus the value in 1990). Between 1988 and 1990, local labor market conditions improved in each area in the sample, so this variable is always positive.

It is clear from the table that the means and standard deviations of the variables do not differ very much across the response status. The average age of the respondents slightly is higher than of the non-respondents. The average of the residual duration beyond the selection date is higher among respondents than in the total sample. Among non-respondents, its average is 285 days. So, on average, individuals with lower exit rates are over-represented among survey participants. This suggests that survey non-response may be selective with respect to unemployment duration.

Table 2 reports results on a logit analysis for non-response. We also estimate a logit that excludes all explanatory variables. This can be used to test whether non-response is random. The likelihood ratio test for joint significance of the included regressors yields a statistic of 151. From this it can be concluded that this hypothesis is strongly rejected. Due to the interaction terms in the model specification, the estimates in Table 2 are not straightforward to translate into effects of personal characteristics. It can be shown that non-response is higher if the individual is male, and if male has age around 30, lives in an inner city, and if the local unemployment rate is increasing. Therefore, samples over-represent individuals with characteristics opposite to these. For women, non-response is lowest around age 30. We also estimated model versions that include regional dummy variables. The regional effects are jointly significant. In particular, non-response is higher in London and the South East.

Table 1: Descriptive statistics of variables in the total sample and among the survey respondents

Variable	Total sample mean (st.dev.)	In the survey mean (st.dev.)
Age	32.6 (12)	33.6 (13)
Female	0.30	0.32
Local unemployment rate decline	0.35 (0.05)	0.35 (0.05)
Living in an inner city area	0.20	0.18
Member of control group	0.06	0.07
Unempl.duration (in days) beyond selection date	302 (407)	315 (412)
Right censored duration	0.03	0.03
# individuals	8012	4708

Note: st.dev. is the standard deviation of the variable.

Table 2: Non-response and explanatory variables: results from a logit for non-response

Variable		
Constant	–0.70	(2.2)
Age	0.06	(4.6)
Age^2	–0.001	(5.8)
Female	1.60	(2.7)
Female∗Age	–0.13	(4.5)
Female∗Age^2	0.002	(4.6)
Local unempl.decline	–1.41	(2.6)
Female∗Unempl.decline	0.88	(0.9)
Inner city area	0.27	(4.1)
Female∗Inner city	–0.10	(0.8)
Control group	–0.05	(0.5)
Female∗Control	–0.25	(1.2)
– log likelihood	5337.2	

Explanatory note: The outcome 1 denotes non-response. The t-values are in parentheses.

2.3 A taxonomy of survey non-response

Table 2 provides a list of the 22 non-response categories as coded by the survey agency. For reasons of transparency, we aggregate these into four *types*. Specifically, we distinguish between non-response due to refusal of the individual to cooperate (labelled as REFUSE), non-response because the individual had moved residence (labelled as MOVED), non-response due to inability of the survey agency to contact a respondent, for reasons not directly associated with mobility (e.g. the respondent is never at home; labelled as NO-CONTACT) and finally non-response for other reasons (e.g. ill health, respondent does not speak English; labelled as OTHER). The flow diagram titled "Types of non-response" relates the types of non-response to the various stages of the survey participation process. Table 4 gives the numbers of non-respondents in the 4 types of non-response.

NOCONTACT and MOVED cover categories associated with a lack or loss of contact between the agency running the survey and the individual respondent. Though the lack of contact could be initiated by the individual respondent, it is presumed that an agency, to a large extent, can affect such drop-out by providing more intensive tracing efforts. An example is "not contacted, never in". It seems that interviewers tried to contact the individual up to as many as three times, but it is still possible that a larger number of attempts increases the frequency in this category. In general, extra efforts of the agency may influence the composition of the sample directly.

Our grouping of the non-response categories is to a large extent driven by our interest in the unemployment duration variable. The types are constructed in such a way that it may be expected that the within-type categories are relatively homogenous in their effect (except of course for OTHER). This should avoid any offsetting effects of separate categories within a type. It is expected that those who left the sample for reasons that assign them to MOVED will have, on average, higher exit rates out of unemployment. The REFUSE type may cover less motivated individuals with poor labour market opportunities. Such individuals lead to non-response bias due to selectivity. We should emphasize that the "type of non-response" does not automatically lead to a specific "reason of non-response bias". For example, the fact that an individual moved to another residence can be related to the acceptance of a job in another town, but it can also be a reflection of the fact that an individual dislikes stability. In both cases the type of non-response is MOVED, but the first case, the reason for non-response bias is "causal" as defined in Section 1, whereas in the second case the non-response bias is due to selectivity. However, it is likely that the type MOVED is more often associated to the causal reason of non-response bias than the other types. Similarly, one may expect REFUSE to be more often associated to the selectivity reason of non-response bias.

It is not always obvious to which type a specific code should be assigned. For example, although "mover - follow up address given" refers to a change of address, it is not clear why

Table 3: Non-response categories

Variable	code	# individuals
No trace of address	1	68
Address vacant or derelict	2	213
Premises demolished	3	5
Business/industrial premises only	4	19
Remote address (not issued to interviewer)	5	56
Mover - follow-up address given	19	43
Mover - follow-up address not known	20	922
Respondent deceased	21	3
No contact at address (4+ calls)	22	398
Complete refusal of info about occupants	23	14
Address given is benefits office	24	36
Refusal to office	70	26
Not contacted (e.g never in)	71	350
Personally refused interview	72	578
Broke appointment; could not be recontacted	73	194
Ill (at home) during survey period	74	12
Away/in hospital during survey period	75	100
Incapacitated	76	10
Refusal on behalf of respondent	77	129
Respondent could not speak adequate English	78	15
Other type of non-response	79	98
Lost in Postal system	80	15
Interview obtained	51	4708
Total number of individuals	8012	

Flow diagram "Types of non-response".

Arabic numbers: non-response categories. Roman numbers: stages of survey participation process

i = Listed in sample
ii = Address known
iii = Valid address
iv = Contact made at address
v = Address confirmed as individual's
vi = Contact made with individual
vii = Appointment made
viii = Response achieved
ix = Response received

i

Y

NOCONTACT

ii N → 1

Y

iii → 5
N → 2, 3, 4, 24, 25

Y

MOVED

22 ← N **iv** N → 23

Y

19, 20 ← N **v**

Y

No direct contact
Direct contact possible

21, 74, 75, 76, 78, 79 ← N **vi** N → 71   77

Y

vii N → 70, 72

Y

viii N → 73

REFUSE

Y

80 ← N **ix**

OTHER

51

12
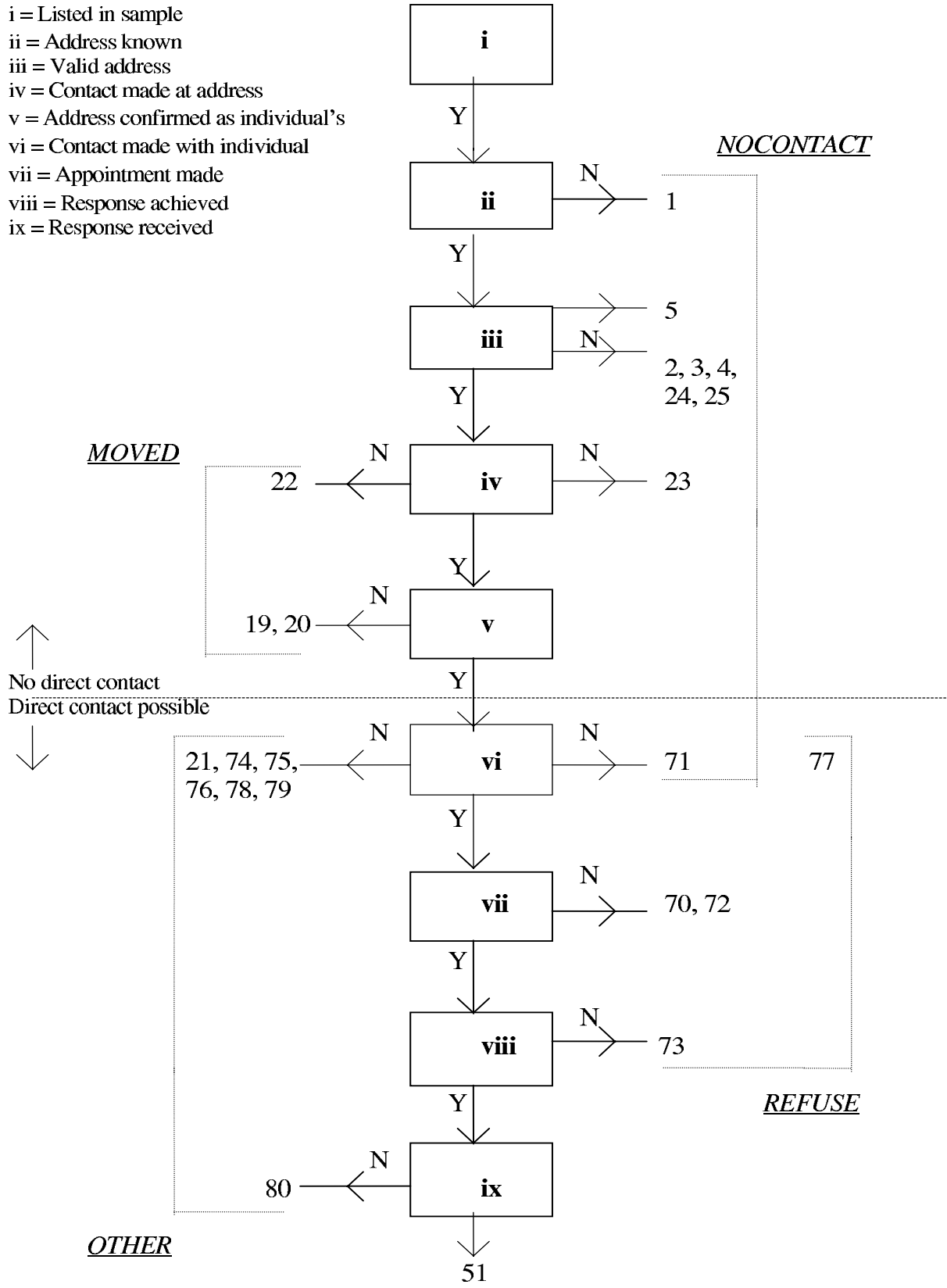
Table 4: Counts for aggregated non-response categories

Category	# individuals
MOVED (codes 19,20,22)	965
REFUSE (codes 23,70,72,77)	747
NOCONTACT (codes 1,2,3,4,5,24,71,73)	1354
OTHER (codes 21,74,75,76,78,79,80)	238
Total non-response	3304

the agency could not contact this individual and whether more efforts of the agency would have resulted in an interview. We choose to classify it as "MOVED" as we expect that the decision to move may signal something of the labour market prospects of the individual and that as such this category is very similar to the category "mover, address unknown". In sensitivity analyses we experimented with small modifications of the type definitions, and it turned out that the results are insensitive to this.

We perform a multinomial logit analysis to relate the type of non-response to a range of personal characteristics (see Table 5). The respondents serve as the reference group and hence the effects presented are relative to this group. The results are in line to those in Table 2 for the logit analysis for non-response in general, in the sense that the latter estimates are averages of the former. Again, the age pattern for women appears to be quite different from that of men. Note that living in an inner city area increases the probability that no contact can be established with the sample member. There are no effects for the variable Control.

Table 5: Non-response and explanatory variables: results from a multinomial logit

Variable	MOVED		NOCONTACT		REFUSE		OTHER	
Constant	−2.48	(4.7)	−1.88	(4.3)	−2.07	(3.7)	−3.59	(4.1)
Age	0.11	(4.1)	0.10	(4.6)	0.06	(2.4)	0.02	(0.5)
Age^2	−0.002	(5.5)	−0.002	(5.4)	−0.0008	(2.4)	−0.0003	(0.5)
Female	2.97	(2.9)	1.28	(1.5)	1.36	(1.4)	0.91	(0.5)
Female∗Age	−0.21	(3.9)	−0.15	(3.5)	−0.09	(2.0)	−0.05	(0.5)
Female∗Age^2	0.003	(3.7)	0.002	(3.6)	0.001	(2.0)	0.0007	(0.6)
Local unempl.decline	−0.59	(0.7)	−1.93	(2.6)	−2.38	(2.5)	0.60	(0.4)
Female∗Unempl.decline	0.16	(0.1)	2.15	(1.6)	0.74	(0.5)	−1.31	(0.5)
Inner city area	0.15	(1.5)	0.50	(6.0)	−0.10	(0.8)	0.30	(1.7)
Female∗Inner city	−0.12	(0.6)	−0.08	(0.5)	−0.008	(0.0)	−0.07	(0.2)
Control group	0.10	(0.6)	−0.11	(0.8)	−0.10	(0.5)	−0.21	(0.6)
Female∗Control	−0.50	(1.3)	−0.16	(0.5)	−0.04	(0.1)	1.15	(1.1)
− log likelihood	9366.7							

Explanatory note: The respondents are taken as reference group. The t-values are in parentheses.

3 Non-response bias in estimates of the unemployment duration distribution

3.1 Non-parametric estimates

In this section we examine the consequences of survey non-response for estimates obtained from basic duration analyses. We start with a comparison of the non-parametric estimates of the duration distribution for the full sample and for respondents. The relevance of this exercise is limited. First, obviously, differences in the distribution of observed explanatory variables are not taken into account. Secondly, the sample contains ongoing spells of unemployment, since individuals are selected when their elapsed duration equals about 4 months. This means that non-parametric estimators do not estimate the unemployment duration distribution but rather a conditional distribution that is truncated from below. Moreover, the elapsed durations in the data are somewhat dispersed around 4 months instead of being identically equal to 4 months. The standard deviation of the elapsed duration at the sample selection date is 21 days. This means that non-parametric estimators estimate a conditional (truncated) distribution with a dispersed truncation point. The Kaplan-Meier estimator can be adapted to this, by focusing on the unemployment duration beyond the selection date (the residual duration), conditional on the elapsed duration at the date of selection. In practice this adaption amounts to an appropriate redefinition of the risks sets at the observed times that a failure occurs.[10] Note however that the interpretation of the Kaplan-Meier estimate may be problematic in the interval of truncation points.

With our data, the first about 130 days of unemployment duration cover on average the elapsed duration on which we condition. At these durations, the shape of the unemployment duration distribution is not identified. As we have seen in the previous section, the sample selection date is imputed (rather than directly observed) in such a way that one would expect to observe few transitions out of unemployment in the first weeks after the imputed selection date. This is confirmed by the Kaplan-Meier estimates. In the about 45 days following the first 130 days, the survivor functions are flat because there are virtually no observed exits out of unemployment. This does not reflect the actual survivor functions in this interval. Rather, at durations between 130 and 175 days, the shape of the unemployment duration distribution is not identified. Note that this does not affect the estimates at higher durations. Because of all this, and because we want to restrict attention

---

[10]With flow data for instance, the risk set at a point t includes durations that are equal to or exceed t. So the risk set R_t satisfies $R_t = \sum_i I(t_i > t)$, where $I(.)$ is the indicator function and t_i the duration of individual/spell i. In case of the distribution of the residual duration r conditional on the elapsed duration p, a spell $(t = r + p)$ is included in the risk set if additionally t exceeds p. So the risk set R_t^* for the residual duration conditional on the elapsed duration equals $R_t^* = \sum_i I(t_i > t, p_i < t)$.
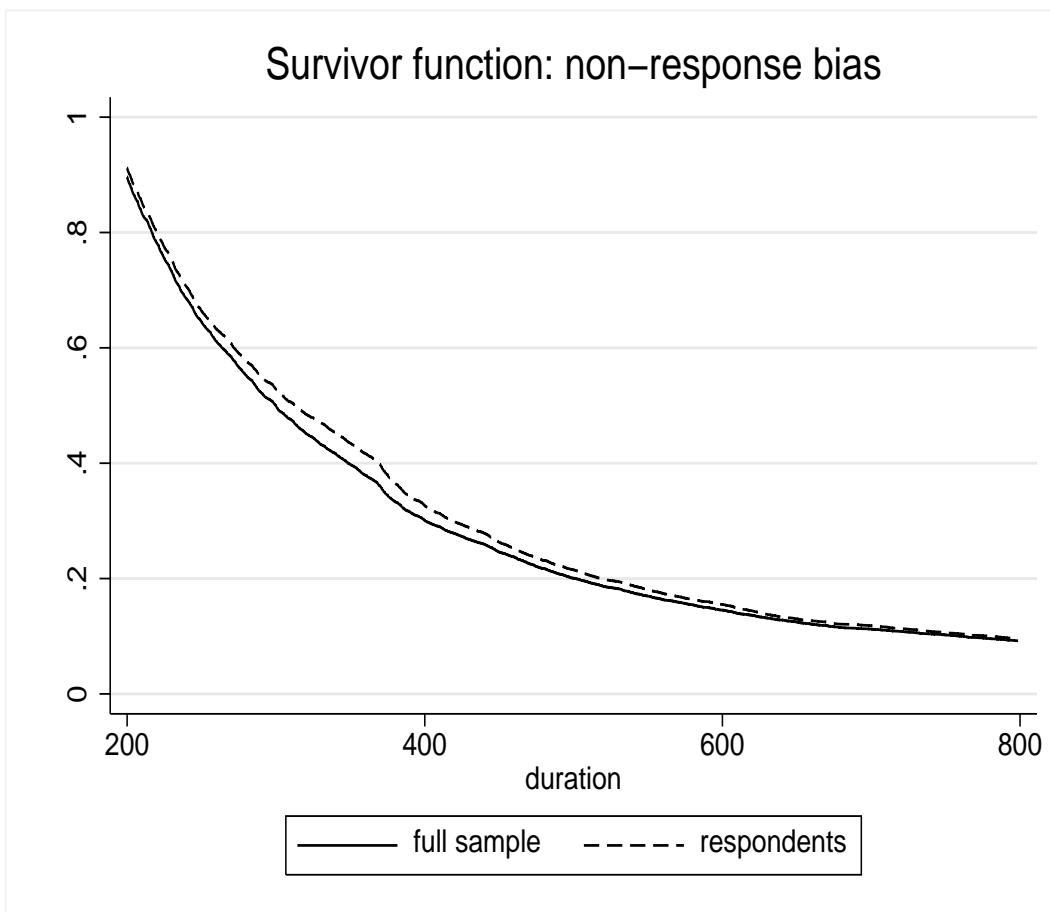
Figure 1: Kaplan-Meier estimates of the conditional unemployment duration distribution, for the full sample and for respondents.

to durations exceeding the largest truncation point, we only present non-parametric results for the durations exceeding 200 days.

Figure 1 plots the estimated survivor functions for the full sample and for the respondents. The survivor function for respondents is above the other, implying that the unemployment exit rates are lower for this respondents. There is a concentration of duration outcomes close to one year. The fact that the curves are relatively close suggests that the non-response bias is small. However, this may be because different types of non-respondents are lumped together, and moreover it is difficult to infer from this figure to what extent the exit rates out of unemployment differ. To formally test for equality, we focus on a comparison of the estimates for respondents and non-respondents. Figure 2 plots the estimated survivor functions for respondents and non-respondents. Clearly, the difference is more pronounced than in Figure 1. The log-rank test for equality of the survivor functions in Figure 2 yields a test statistic of 26, which by far exceeds the 95th percentile of
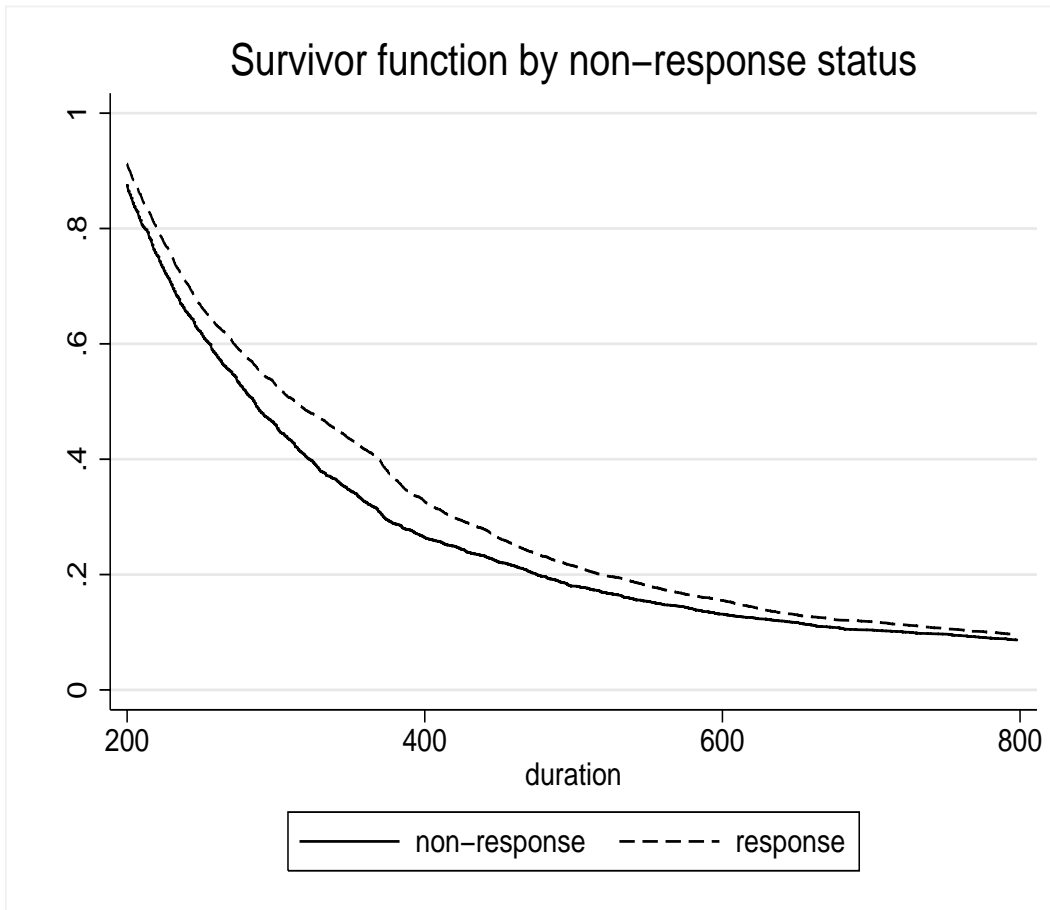
Figure 2: Kaplan-Meier estimates of the conditional unemployment duration distribution by (non-)response status.

the χ_1^2 distribution. Equality of the survivor functions is therefore strongly rejected. This is confirmed by plotting the confidence intervals for the Kaplan-Meier estimates (not shown here).

It is to be expected that the type of non-response is also informative with respect to the relation between non-response and unemployment duration. For example, MOVED suggests a direct causal effect, in which case the estimated survivor function may decline more steeply at low durations. Note that one should be careful with interpreting these estimates, because the type of non-response is not an explanatory variable of unemployment duration. In Figure 3 the estimated survivor functions for the different types of non-response are plotted. There are clear differences between the estimates. The survivor function for OTHER is uniformly below all other functions, implying that the unemployment exit rates are higher for this type. Individuals who refuse to participate in the survey also have substantially higher exit rates than survey participants. As expected, movers (MOVED) also

17

have higher exit rates than survey participants. The figure shows a ranking of the survivor functions: the function for OTHER lies below the function for MOVED, which in turn lies below the NOCONTACT survivor function. However, the survivor functions for respondents and NOCONTACT intersect. Initially, respondents have the lowest exit rate, but later this reverses.

The log-rank test for equality of the survivor functions in Figure 3 yields a test statistic of 73, which by far exceeds the 95th percentile of the χ_4^2 distribution. Equality of the survivor functions is therefore strongly rejected. This is confirmed by plotting the confidence intervals for the Kaplan-Meier estimates (not shown here).

As noted, the non-parametric estimates are affected by individual heterogeneity. We therefore proceed by estimating semi-parametric unemployment duration models that incorporate observed explanatory variables.

3.2 Semi-parametric estimates

We adopt the familiar Cox Proportional Hazard (PH) model for the duration of unemployment, with, in obvious notation, hazard rate $\theta(t|X) = \lambda(t) \exp(X'\beta)$. We employ partial likelihood estimation (see e.g. Lancaster, 1990). This does not impose any functional form restriction on the duration dependence term or baseline hazard $\lambda(t)$, so it allows for an unrestricted non-parametric baseline hazard.[11]

Because of the variation in the elapsed duration at the sample selection date, we modify the standard partial likelihood procedure along the lines of the previous subsection. That is, we apply partial likelihood estimation to the residual durations conditional on the elapsed durations, allowing for delayed entry of individuals, where the delay time depends on the elapsed duration.[12] An alternative is to use a stratified partial likelihood approach (see Ridder and Tunalı, 1999) in which the stratification is on the elapsed duration. This procedure allows for a non-parametric baseline hazard for each stratum. We also use this estimation procedure, and the main conclusions are not altered, so we do not report them here.

For all results below, we estimate the variance-covariance matrix by way of the inverse of the Hessian as well as by way of the sandwich estimator of Lin and Wei (1989).

---

[11]The PH model is incorrect if the controls have different duration dependence than the treated, which seems likely because the former do not have 6-monthly meetings with their case worker whereas the latter do. We ignore this issue here for reasons of simplicity. Note that the controls only constitute 6% of the sample.

[12]Somewhat loosely, contributions to the partial likelihood function are based on the conditional probability that a spell i ends, given the risk set R_i, defined as the set of spells having the same duration as spell/individual i or longer. Here, our risk set R_i^p is defined as the set containing all spells exceeding the length of spell i (t_i) of which the elapsed duration p_i is smaller than t_i.
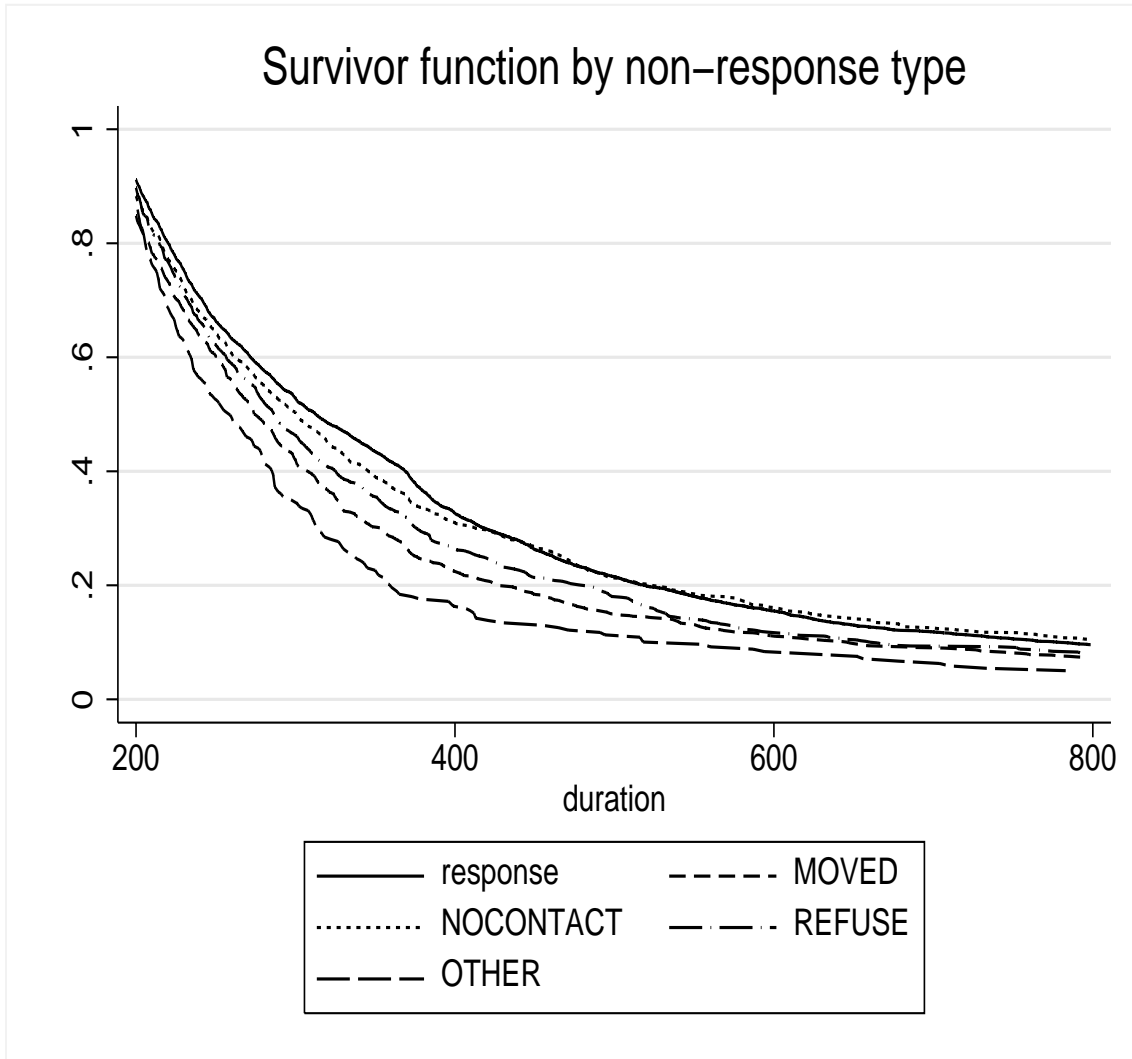
Figure 3: Kaplan-Meier estimates of the conditional unemployment duration distribution by (non-)response type.

This estimator is robust to misspecifications of the model. Consequently, a comparison of the conventionally estimated standard errors with those based on the sandwich estimator provides an informal specification test. It turns out that the differences are invariably extremely small, and therefore we do not report both.

Table 6 reports the partial likelihood estimation results of the PH model using different samples. For our purposes, the most important aspect of the table concerns the difference between the estimates in the second and third columns. If these are the same then non-response does not affect the estimates of the covariate effects on the duration distribution. It turns out that the estimates of the covariate effects are very close, and certainly not as different as might be expected from the high non-response rate. Indeed, for all means and purposes, they are qualitatively the same. So our first general observation is that estimates of covariate effects are not seriously biased by non-response. We formally test this using a Stratified Partial Likelihood Ratio test inspired by Andersen et al. (1993). This test does not impose equal baseline hazards by response status. The test accepts the null hypothesis of equal covariate effects.

One may argue that this is because the omission of non-respondents primarily leads to changes in the distribution of the unobserved determinants of the duration of unemployment, and that these changes are absorbed by the estimated baseline hazard function. The theoretical results of Ridder (1987) (see also the survey in Van den Berg, 2001) may support this explanation. In general, estimates of duration models with flexible baseline hazards (like in our case) are less sensitive to misspecification than more restrictive models, in particular if there is virtually no right-censoring of the duration variable, as in our data. The theoretical literature also predicts that estimated covariate effects are typically smaller and not larger, if unobserved heterogeneity is not taken into account. This prediction is confirmed by our empirical results that show that to the effect that the estimated covariate effects differ, they are smaller if the full sample (containing more heterogeneity) is used. Against this explanation one may argue that selective non-response involves a more serious selection problem than a misspecification of, say, the functional form of an unobserved heterogeneity distribution. To investigate this further, we estimate models where the baseline hazard is restricted to be constant (i.e. the duration has an exponential distribution given the observed explanatory variables) and models where the baseline hazard has the Weibull functional form. Table 7 presents the results for the former. It turns out that the differences between the estimates of the full sample and the sample of respondents are larger than in Table 6. The same applies to the case of the Weibull specification. Therefore, a flexible specification of the duration model is to be preferred over more restrictive specifications, in order to reduce biases due to sample non-response.

We now briefly discuss the covariate effects themselves, in columns 2–4 of Table 6. The coefficients on the variables Age and Age^2 suggest that exit rates increase after age 23, but

Table 6: Non-response and unemployment duration: results from partial likelihood analyses

Variable	full sample		full sample		respondents		non-respondents	
Personal characteristics								
Age	−0.023	(6.0)	−0.023	(6.1)	−0.031	(6.1)	−0.016	(2.6)
Age^2	0.0005	(5.9)	0.0005	(5.8)	0.0007	(6.0)	0.0004	(2.4)
Female	0.397	(2.4)	0.408	(2.4)	0.486	(2.2)	0.313	(1.1)
Female∗Age	0.011	(1.6)	0.010	(1.4)	0.020	(2.2)	−0.0005	(0.0)
Female∗Age^2	−0.0004	(2.4)	−0.0004	(2.2)	−0.0006	(2.5)	−0.0003	(0.8)
Local unempl.decline	1.587	(6.1)	1.598	(6.2)	1.908	(5.6)	1.304	(3.3)
Female∗Unempl.decline	−0.468	(1.0)	−0.517	(1.1)	−0.842	(1.4)	−0.083	(0.1)
Inner city area	−0.172	(5.1)	−0.166	(5.0)	−0.155	(3.4)	−0.200	(4.1)
Female∗Inner city	−0.095	(1.5)	−0.099	(1.6)	−0.125	(1.5)	−0.052	(0.5)
Control group	−0.228	(4.3)	−0.222	(4.2)	−0.273	(3.9)	−0.151	(1.8)
Female∗Control	0.003	(0.0)	−0.018	(0.2)	0.035	(0.3)	−0.084	(0.5)
Non-response types								
MOVED	0.238	(6.6)	–		–		–	
NOCONTACT	0.048	(1.5)	–		–		–	
REFUSE	0.136	(3.4)	–		–		–	
OTHER	0.415	(6.2)	–		–		–	
– log likelihood	62804.7		62842.0		34490.6		23069.2	
# individuals	8011		8011		4708		3303	

Explanatory note: The partial likelihood is based on the distribution of the residual duration conditional on the elapsed duration. In parentheses we report t-values.

Table 7: Non-response and unemployment duration: results for an exponential model

Variable	full sample		full sample		respondents	
Personal characteristics						
Constant	−6.202	(65.0)	−6.158	(65.0)	−6.344	(50.7)
Age	−0.034	(8.5)	−0.035	(8.6)	−0.042	(8.2)
Age^2	0.0008	(8.0)	0.0008	(8.0)	0.0009	(7.7)
Female	0.550	(3.2)	0.582	(3.4)	0.800	(3.6)
Female∗Age	0.018	(2.4)	0.017	(2.3)	0.028	(3.0)
Female∗Age^2	−0.0007	(3.4)	−0.0006	(3.2)	−0.0008	(3.4)
Local unempl.decline	1.763	(6.8)	1.820	(7.0)	2.325	(6.9)
Female∗Unempl.decline	−0.658	(1.4)	−0.764	(1.6)	−1.570	(2.7)
Inner city area	−0.199	(5.9)	−0.192	(5.7)	−0.166	(3.6)
Female∗Inner city	−0.154	(2.4)	−0.161	(2.5)	−0.173	(2.0)
Control group	−0.287	(5.2)	−0.283	(5.1)	−0.352	(4.8)
Female∗Control	0.033	(0.3)	0.013	(0.1)	0.088	(0.7)
Non-response types						
MOVED	0.25	(7.0)	–		–	
NOCONTACT	0.011	(0.3)	–		–	
REFUSE	0.126	(3.2)	–		–	
OTHER	0.410	(6.1)	–		–	
– log likelihood	7013.4		7053.4		4164.8	
# individuals	8011		8011		4708	

Explanatory note: t-values are in parentheses.

22

in fact the quantitative change is very small. The coefficient on Female is significantly positive, but to evaluate the gender effect one has to take the interaction effects into account as well. The over-all average effect is negligible. The local unemployment decline variable has a large and significant positive effect on the exit rate, so better labour market conditions increase the exit rates of the unemployed. The inner city variable indicates lower exit rates for inner city inhabitants. Finally note that the control group experiences significantly lower exit rates. Therefore, the results indicate that the Restart treatment is effective in increasing labour market prospects of the unemployed. From column 4 on the non-respondents, it is clear that female non-respondents have a lower exit rate out of unemployment than female respondents. However, the difference is not statistically significant.

Next, we examine the non-response bias in the estimated duration dependence (i.e. the estimated baseline hazard $\lambda(t)$) as well as in the mean level of the hazard rate. The baseline hazard estimates for the full sample and for the respondents are derived from the above partial likelihood estimations (the second and third column in Table 6). Figure 4 plots the smoothed[13] baseline hazards. As in the previous subsection, we restrict attention to durations exceeding 200 days. Moreover, at durations above 600 there are not many observations and these are widely scattered, so we do not consider durations exceeding 600.

Clearly, the curves in Figure 4 are rather different from each other, especially at lower durations. The baseline hazard for the full sample is larger at first, but at higher durations it drops to a level that is slightly smaller than for the sub-sample of respondents. However, the comparison is somewhat hampered by the fact that the estimates are based on separate partial likelihood estimations, with different estimated covariate effects. This is even more relevant for a comparison of the estimated mean hazard level. We solve this by estimating the model with stratified partial likelihood estimation, stratifying on the response status. This assumes identical covariate effects but allows the mean level and duration dependence to differ between respondents and non-respondents. Recall that the covariate effects were already found to be equal. It turns out that the shapes of the baseline hazard estimates are very similar to those from separate partial likelihood estimations for respondents and non-respondents. Baseline hazard estimates by response status are also important for the analysis in the upcoming sections.

Figure 5 plots the smoothed[14] baseline hazard by response status. Clearly, the curves are very different from each other, reinforcing the conclusions of the previous paragraph.

---

[13]We use median splines with 10 sub-intervals.

[14]Here we use median splines with only 5 sub-intervals, because of the smaller sample size of the non-respondents. Notice that the dashed curve in Figure 4 corresponds to the estimated baseline hazard for the respondents in Figure 5, the only differences being the degree of smoothing and the scale on the vertical axis.
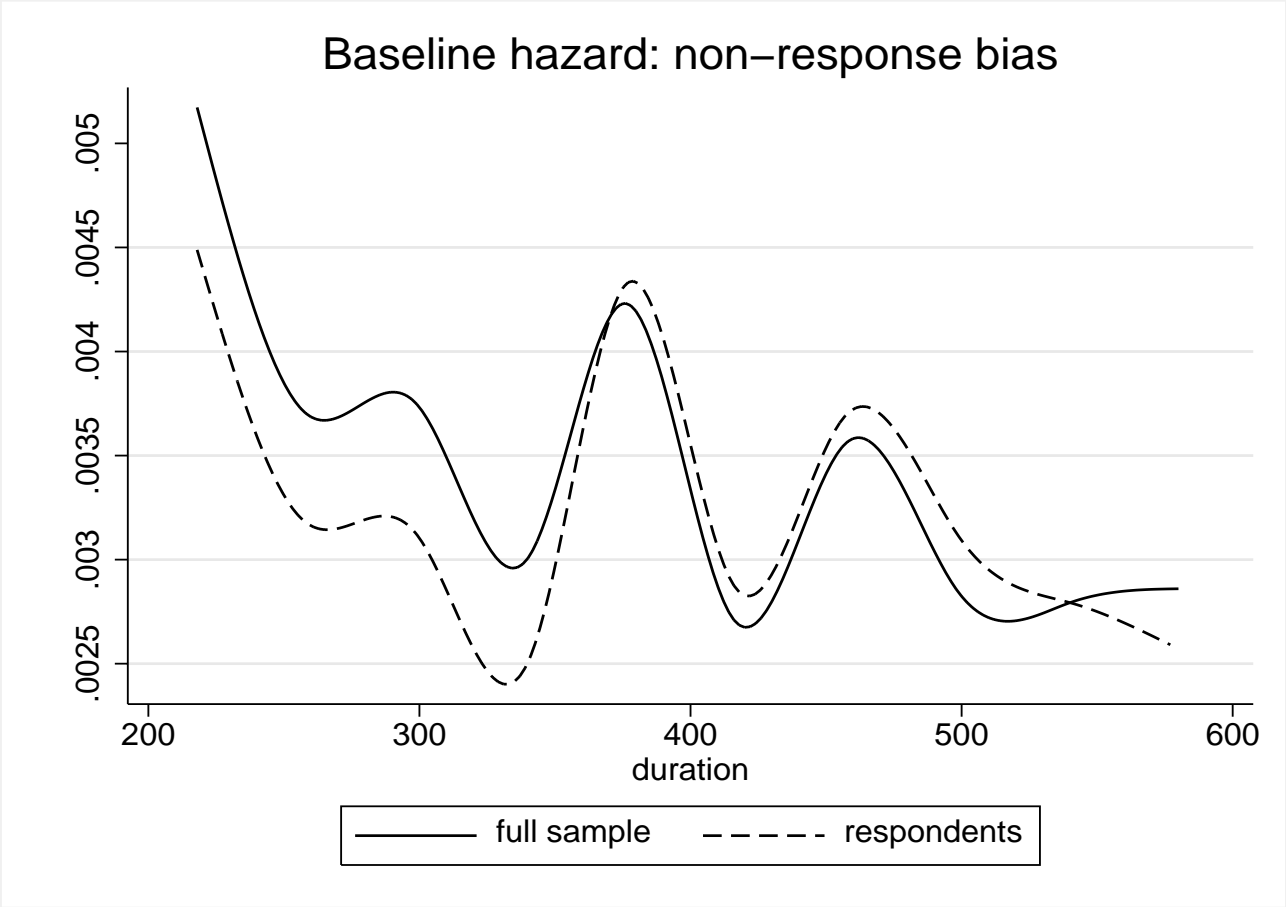
Figure 4: Estimates of the baseline hazard, for the full sample and for respondents, based on separate partial likelihood estimation.
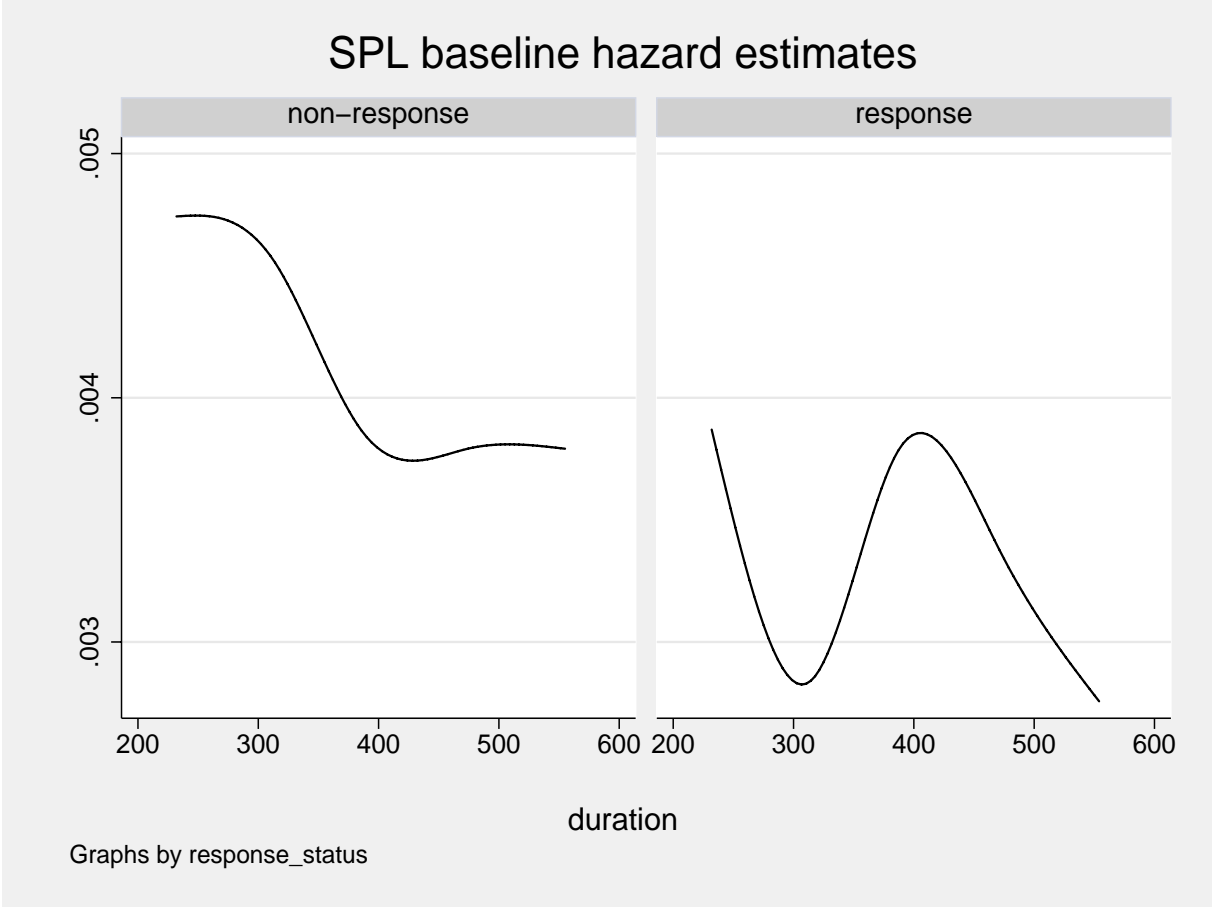
Figure 5: Estimates of the baseline hazard by response status, based on stratified partial likelihood estimation.

The baseline hazard for non-respondents is much larger at first, but at higher durations it drops to a level that is somewhat closer to that for the respondents. The mean level is higher for non-respondents. This is in accordance to the non-parametric results. Not accounting for survey non-response therefore gives a downward bias of the exit rate out of unemployment, while the duration dependence is estimated to be less negative than in reality. The latter is obviously relevant from a policy point of view, as it suggests that the scarring effect of long-term unemployment is larger than what is inferred from survey estimates. Moreover, survey estimates lead to erroneous conclusions concerning the prevalence of long-term unemployment and the associated welfare problems.

We repeat the above analyses using the Ramlau-Hansen estimator for the baseline hazards (see e.g. Andersen et al., 1993). This is based on kernel smoothing instead of spline smoothing. One advantage is that it enables the estimation of confidence intervals. The results reinforce those in the previous paragraphs, for kernel bandwidths ranging from 10 to 60 days.

We end this subsection by briefly examining the extent to which the mean level of the hazard varies across the four different types of non-response. The first column of Table 6 estimates the model on the complete sample (i.e. respondents and non-respondents), including the 4 non-response type indicators as additional explanatory variables. There are two caveats. First, the parameters associated with those indicators cannot be given a causal interpretation, because from a behavioural point of view there is no causal effect from (the type of) non-response on unemployment exit. Significant coefficients merely signal differences in the distribution of unemployment durations across non-response types. At most, the non-response types may be interpreted as indicators of unobserved characteristics that affect unemployment durations. A significant effect may then signal the presence of relevant (normally unobserved) heterogeneity. The second caveat is that the estimation assumes identical baseline hazards across response status, which we already know is incorrect. We therefore also use stratified partial likelihood estimation, where stratification is taken with respect to response status and non-response type. It turns out that the estimated coefficients on the non-response types are virtually identical to those reported here.

The coefficients of all non-response types are positive. A likelihood ratio test leads to rejection of the null hypothesis that all 4 coefficients are zero. This again confirms that survey non-respondents have a higher mean hazard level. The rank order of coefficients across types is in agreement to Figure 3. The exit rate for NOCONTACT is not significantly higher than for the sample respondents. This is consistent with the notion that NOCONTACT is to a larger extent determined by the efforts of the survey agency than the other non-response types. On the other hand, one may expect that NOCONTACT is informative on unobserved characteristics of the individual. Also, recall from the non-parametric

analysis that the ranking of the hazard rates of NOCONTACT and the respondents is non-monotonic. NOCONTACT seems to be associated with a particularly low exit rate at low durations.

Finally, note that the covariate effects are virtually identical across the first two columns. If the non-response type merely represents unobserved heterogeneity, then, according to the line of reasoning earlier in this subsection, one would expect the coefficients in the first column to be larger.

4 The methodology for assessing the causality explanation and the selectivity explanation of the non-response bias

4.1 Definition of the causal and the selection effect

For ease of exposition, we assume that unemployment duration T is a continuous random variable, and that the time interval between the moment of inflow into unemployment and the moment of the survey date is exactly c months. Also, for simplicity, we do not use notation in terms of counterfactuals.

We assume that the individual exit rate out of unemployment depends on the elapsed unemployment duration t and the observed and unobserved explanatory variables X and V, and we express this rate as $\theta(t|X,V)$. The individual probability of being non-respondent depends on whether $T < c$ and on the observed and unobserved explanatory variables X and ε, and we express this rate as $\pi(X, \varepsilon, \mathrm{I}(T < c))$, where $\mathrm{I}(.)$ is the indicator function which equals 1 if its argument is true and 0 otherwise. The individual outcome variables are T and the binary non-response indicator Z, where we take $Z = 1$ to denote non-response. Note that we specify a deterministic model for θ and π at the individual level, and that T and Z given X, V, ε are stochastic, so there is randomness in the outcomes at the individual level. For example, we may specify a latent variable model for $Z|T$, as follows: $Z = 1$ iff $Z^* > 0$, with $Z^* = \gamma_0 + \gamma_1 X + \gamma_2 \mathrm{I}(T < c) + \varepsilon + \eta$. In that case, $\pi(X, \varepsilon, \mathrm{I}(T < c)) = 1 - H(-\gamma_0 - \gamma_1 X - \gamma_2 \mathrm{I}(T < c) - \varepsilon)$, with H being the distribution function of η. In general, we implicitly make some regularity assumptions on θ and π, for example that θ is continuous in t and positive.

We observe a random sample from the joint distribution of $T, Z|X$. The random variables T and Z given X can be dependent for two reasons: because of a causal effect of Z on T, and because the unobserved determinants V and ε of T and Z are dependent. We do not aim to specify a model that is fully identifiable. The aim is to demonstrate that it is nevertheless possible to identify whether a causal or selection effect is present when it can

not be ruled out that the other effect is present as well. The causal effect is represented by the effect of T on $\pi(X, \varepsilon, \mathrm{I}(T < c))$. Of course, this can not be directly observed. The selection effect is represented by a stochastic relation between V and ε. If non-respondents often have $T < c$ then this may be due to a causal effect or to a selection effect. In case of a selection effect, individuals with values of V that increase θ also have values of ε that increase π.

4.2 Assessing the presence of a selection effect

As noted in the introduction, although the causal effect runs from Z to T, we find it more useful to examine the conditional distribution of $T|Z, X$. This should not be confused with the distribution of T in the population, which is directly observable. We focus on the conditional distribution of $T|Z, X$ because, as we shall see, it is informative on the types of dependence between T and Z given X. Note that even in case of a causal effect, Z is not a deterministic function of T, so that typically the conditional distribution of $T|Z, X$ is not degenerate.

Below we use θ and f as generic symbols for a hazard rate and a density, respectively. Bayes' theorem implies that

$$f(t|Z, X) = f(t|X) \frac{\Pr(Z = z|T = t, X)}{\Pr(Z = z|X)} \tag{1}$$

This implies the following equation which is particularly useful for our purposes,

$$\theta(t|Z, X) = \theta(t|X) \frac{\Pr(Z = z|T = t, X)}{\Pr(Z = z|T \geq t, X)} \tag{2}$$

where $\theta(t|X) \equiv \mathbb{E}_V[\theta(t|X, V) \mid T \geq t]$ denotes the hazard rate in the population aggregated over V among the survivors at t. All terms in (2) are in principle observable. Now consider the issue of whether a selection effect is present or not, when a causal effect can not be ruled out. We consider the hazard rate $\theta(t|Z, X)$ at durations $t > c$. First, suppose there is no selection effect. Then $V \perp\!\!\!\perp \varepsilon$, and

$$\Pr(Z = 1|T = t, X) = \mathbb{E}_{\varepsilon, V}[\Pr(Z = 1|T = t, X, \varepsilon, V) \mid T = t, X] = \mathbb{E}_\varepsilon[\pi(X, \varepsilon, 0) \mid X] \tag{3}$$

where we use that $[\varepsilon \perp\!\!\!\perp V]|T = t, X$. Similarly,

$$\Pr(Z = 1|T \geq t, X) = \mathbb{E}_{\varepsilon, V}[\Pr(Z = 1|T \geq t, X, \varepsilon, V) \mid T \geq t, X] = \mathbb{E}_\varepsilon[\pi(X, \varepsilon, 0) \mid X] \tag{4}$$

By substituting these equations into equation (2), it follows that in the absence of a selection effect,

$$\theta(t|Z=1,X) = \theta(t|X) = \theta(t|Z=0,X) \qquad \text{(for all } t>c\text{)}.$$

In the presence of a selection effect, i.e. $V \not\!\perp \varepsilon$, equation (3) is replaced by

$$\Pr(Z=1|T=t,X) = \mathbb{E}_V\left[\mathbb{E}_\varepsilon(\pi(X,\varepsilon,0)\mid V,X)\mid T=t,X\right] \tag{5}$$

with a similar result to replace (4). Consequently,

$$\theta(t|Z=1,X) = \theta(t|X)\,\frac{\mathbb{E}_V\left[\mathbb{E}_\varepsilon(\pi(X,\varepsilon,0)\mid V,X)\mid T=t,X\right]}{\mathbb{E}_V\left[\mathbb{E}_\varepsilon(\pi(X,\varepsilon,0)\mid V,X)\mid T\geq t,X\right]}$$

In general, this does not equal $\theta(t|X)$ or $\theta(t|Z=0,X)$. To obtain an equality one needs to consider pathological cases, like a model where the dynamic weeding out due to selection is finished at t, or a model where the hazard rates are identically zero at t.

In words, because of the selection effect, conditioning on $Z = 1$ means examining a subset of survivors at t among which individuals with certain values of V are over-represented, compared to conditioning on $Z = 0$. A causal effect does not play a role here because we condition on survival up to $t > c$.

4.3 Assessing the presence of a causal effect

Now consider the issue of whether a causal effect is present or not, when a selection effect can not be ruled out. We want to examine the hazard rates $\theta(t|Z,X)$ around $t = c$. First, consider equation (1) for the density of $T|Z,X$. At the right-hand side, the density $f(t|X)$ is the density of a duration variable with a continuous hazard rate, averaged over the unobserved heterogeneity distribution. Such a density is continuous (see e.g. Lancaster, 1990). Next, the probabilities $\Pr(Z=z|T=t,X)$ can be expressed as

$$\Pr(Z=1|T=t,X) = \mathbb{E}_V\left[\mathbb{E}_\varepsilon(\pi(X,\varepsilon,\mathrm{I}(t<c))\mid V,X)\mid T=t,X\right] = 1-\Pr(Z=0|T=t,X) \tag{6}$$

The probabilities $\Pr(Z=z|T=t,X)$ are discontinuous in t at $t=c$ if and only if there is a causal effect. The expectation over the unobserved heterogeneity distribution of $V|T=t,X$ does not affect this result. As a result, the density $f(t|Z,X)$ is discontinuous at $t=c$ if and only if there is a causal effect. Note that the integral of a function that is continuous except at, at most, one point, is always continuous, so the survivor function of $T|Z,X$ is continuous. By dividing both sides of equation (1) by this survivor function, we obtain the result that the hazard rate $\theta(t|Z,X)$ is discontinuous at $t=c$ if and only if there is a causal effect.

In practice, a discontinuity is hard to distinguish from a decrease of a continuous function, in particular if the duration data contain small measurement errors or are time-aggregated. It is easier to examine whether one function decreases stronger than another. In the presence of a causal effect, the probability $\Pr(Z = 1|T = t, X)$ jumps downward at $t = c$ whereas the probability $\Pr(Z = 0|T = t, X)$ jumps upward at $t = c$. In the absence of a causal effect, both probabilities are continuous at $t = c$. We therefore examine

$$\frac{\theta(c^+|Z = 1, X)/\theta(c^-|Z = 1, X)}{\theta(c^+|Z = 0, X)/\theta(c^-|Z = 0, X)}$$

where $\theta(c^-|Z, X) := \lim_{t \uparrow c} \theta(t|Z, X)$ and $\theta(c^+|Z, X) := \lim_{t \downarrow c} \theta(t|Z, X)$. It can be shown that the above expression equals $(\Pr(Z = 1|t = c^+, X)/\Pr(Z = 1|t = c^-, X))/(\Pr(Z = 0|t = c^+, X)/\Pr(Z = 0|t = c^-, X))$. In the presence of a causal effect the above expression is smaller than 1, whereas in the absence of a causal effect it is equal to 1. In words, with a causal effect, non-respondents are more likely to have a duration outcome just before the survey date than just after, compared to respondents. A selection effect can not give rise to such a strong local stochastic dependence of the duration outcome on the response status. Note that the causal reason for non-response bias can be present even if the population is fully homogeneous. In that case there will always be individuals who by chance find a job before the survey date, and this increases their probability of becoming non-respondent.

The above expression neatly corrects for any other reasons for why the hazard may be discontinuous or sharply increasing or decreasing around $t = c$. From equation (2) it is obvious that the hazards around c also reflect the duration dependence of the hazard rate $\theta(t|X, V)$, which is a model primitive. For example, seasonal conditions in the labour market may be improving around $t = c$. However, this duration dependence affects the observed duration dependence of $\theta(t|Z, X)$ around c alike for respondents and non-respondents. So, by examining the difference of the observed duration dependence of respondents and non-respondents, we omit its effect. There is an analogy to the econometric method of difference-in-differences to estimate a parameter of interest (see e.g. Heckman, LaLonde and Smith, 1999, for an overview). By comparing respondents and non-respondents, we correct for the part of the observed duration dependence that is due to factors that we are not interested in.

We summarize the results in the following proposition,

Proposition 1. *There is a selection effect if and only if $\theta(t|Z = 1, X) \neq \theta(t|Z = 0, X)$ for $t > c$, regardless of whether there is a causal effect or not. There is a causal effect if and only if $\theta(c^+|Z = 1, X)/\theta(c^-|Z = 1, X) < \theta(c^+|Z = 0, X)/\theta(c^-|Z = 0, X)$, regardless of whether there is a selection effect or not.*

The above results may be generalized under some additional regularity conditions. For example, we may allow the individual non-response probability to depend on the realization

of T if it occurs before $T = c$. This captures the idea that the closer the exit out of unemployment to the survey date, the less difficult it is for the survey agency to find the individual. Also, we may the unobserved heterogeneity terms V and ε to vary stochastically over time.

5 The empirical evidence concerning the causality and selectivity explanations

The test on selection effects is straightforwardly applied to the estimation results of the previous section on the baseline hazard. Recall that we estimate the baseline hazard for respondents and for non-respondents separately, without imposing functional forms. The latter is important because if a functional form is misspecified then the fit to the functional form on the duration interval $(0, c)$ may affect the fit at $t > c$. With partial likelihood estimation, the baseline hazard functions are estimated non-parametrically.

We focus on durations exceeding the survey date. That date corresponds to approximately 340 days of unemployment duration (on average, 130 days of elapsed duration at the sample selection date plus some 7 months from there until the survey date). From Figure 5 we observe that the duration dependence is clearly different across respondents and non-respondents at durations exceeding 340. Therefore, the estimation results unambiguously show the presence of a selection effect.

Figure 5 also strongly suggests the presence of a causal effect. The baseline hazard for non-respondents has a smaller gradient than the baseline hazard for respondents, around 340 days.

We now proceed towards a more formal analysis of the presence of a causal effect, by estimating a duration model and subsequently examining the estimated duration dependence by response status around the survey data. However, recall that, unfortunately, the survey date is not always at exactly the same elapsed duration for all individuals. We deal with this by estimating models that allow for two different types of duration dependence: the dependence of the current hazard rate on the elapsed unemployment duration (i.e., the usual duration dependence), and the dependence on the time relative to the survey date. We are only interested in the latter, and in particular whether, just before the survey date, the effect of the time relative to the survey date decreases more among non-respondents than among respondents.

The model that we estimate allows for piecewise constant duration dependence as well as piecewise constant dependence on the time relative to the survey date. The two types of duration dependence and the covariates are assumed to act proportionally on the hazard rate. We estimate the model separately for respondents, non-respondents, as well as

for each of the four types of non-response.[15] Note that, as in Section 3, unemployment duration models that condition on survey response outcomes should not be interpreted as behavioural models of unemployment duration, but rather as descriptions of the distribution of unemployment duration conditional on survey response outcomes. For the effect of the time relative to the survey date we use 24 monthly dummy variables, and for convenience we label them as denoting the time since the selection of the sample, which approximately equals the time relative to the survey date plus 7 months.[16] Because of the small number of observations for the "OTHER" category, the model for that category only takes 18 of these monthly dummy variables. The likelihood function conditions on the elapsed duration at the sample selection date, and we estimate the model with maximum likelihood.

Table 8 presents the estimates.[17] To some extent, the duration dependence effect of "Month since selection of the sample" seems to picks up misspecification of the ordinary duration dependence, as it slowly decreases over time. This it not surprising, given that (i) both effects start early on in the spell and move in line as time proceeds, and (ii) the ordinary duration dependence is only captured by 4 coefficients. We therefore do not focus in great detail on these estimates.

To test for a causal effect we examine the hazard rates when going from month 7 to month 8 after the sample selection date (note that this point in time does not generally coincide with a jump of the duration dependence function). The hazard rate for respondents increases by 0.32 (standard error 0.09) whereas the hazard rate for non-respondents decreases by 0.09 (standard error 0.10). (The parameter estimates capturing "Month since selection of the sample" have relatively large standard errors, but the standard errors of the difference between parameters of adjacent months are much smaller for a given subsample.) Clearly, the difference in direction between respondents and non-respondents is in accordance to a causal effect. Moreover, the magnitude of the difference of the change is 0.41, with an estimated standard error of 0.13. The difference is significantly positive, so the comparison between respondents and non-respondents confirms the presence of a causal effect.

The results in Table 8 show that, among the non-response types, the hazard rate de-

---

[15]We omit a further 7 individuals from the sample of 8011 because of their large elapsed unemployment durations.

[16]Recall that for non-respondents we only observe the intended survey date.

[17]The first column gives estimates for the full sample where we condition on the type of non-response (i.e., we include 4 binary non-response indicators, and the baseline category is being respondent). We also estimate the model on the full sample without these four variables (estimates not shown). This results in a drop of the log likelihood of 34 points, so the LR test on the null hypothesis that unemployment durations are independent of non-response status is strongly rejected (test statistic 68; χ^2_4 distribution under the null), confirming our previous results.

crease when going from month 7 to month 8 is largest for the "MOVED" type of non-respondents. This is precisely the type for which a causal effect is a priori most likely. For "MOVED" we find a decrease of –0.40 (standard error 0.20). The difference of 0.72 in comparison to the respondents has a standard error of 0.21. The difference is therefore strongly significant. For the NOCONTACT type of non-response, the difference in comparison to the respondents is also significant (difference 0.40 with standard error 0.20). For REFUSE and OTHER we do not find a significant difference. This is to be expected: REFUSE concerns individuals who were actually contacted, so this type of non-response does not seem to reflect causal effects. OTHER only includes a small number of non-respondents.

In conclusion, the relation between non-response and unemployment duration is only partly due to the fact that both are affected by unobserved individual-specific characteristics. We also find strong evidence of a causal effect of exit out of unemployment on non-response.

To identify the magnitudes of both types of non-response bias one would need to specify an identified model of unemployment duration and non-response, incorporating a causal effect at the individual level, as well as unobserved heterogeneity to explain selection effects. We feel that this is beyond the scope of the paper. Moreover, it is not clear to what extent the results would be generalisable to other situations.

6   Conclusions and recommendations

With a unique dataset that merges survey data and register data for the same individuals, we assess the survey non-response bias in estimates based on survey data. We focus on unemployment duration as the variable of interest. For each non-respondent, the register data provide observations of the unemployment duration and covariates.

We find, first of all, that non-response indeed biases the estimates of the unemployment duration distribution. In particular, the estimated mean level and duration dependence of the hazard rate are biased. With survey data, the hazard rate is under-estimated, and the estimated hazard rate does not decrease as fast as in the population. However, in our data, the covariate effects are found to be correctly estimated from the survey data, provided that a flexible semi-parametric empirical approach is used.

Next, we develop and apply a method to inquire the reasons for the non-response bias. We distinguish between a bias due to a selection effect (i.e. due to related unobserved determinants) and a bias due to a causal effect (i.e. because the individual left unemployment before the survey date). The method does not require functional form assumptions or a semi-parametric model structure, and indeed it does not require a fully identified model. It exploits the data variation in the timing of the moment of exit out of unemployment relative to the survey date.

Table 8: Estimates for the exit rate out of unemployment, with a piecewise constant duration dependence as well as dependence on the time relative to the survey date, and conditional on non-response status.

Variable	full sample		respondents		non-respondents		MOVED=1	
Personal characteristics:								
Age/10	-0.388	(0.06)	-0.464	(0.08)	-0.291	(0.11)	0.162	(0.22)
Age^2/100	0.045	(0.01)	0.055	(0.01)	0.033	(0.02)	-0.025	(0.03)
Female	0.225	(0.02)	0.242	(0.03)	0.188	(0.04)	0.299	(0.07)
Inner city area	-0.202	(0.03)	-0.196	(0.04)	-0.218	(0.04)	-0.229	(0.08)
Local unempl.decline	1.410	(0.21)	1.615	(0.28)	1.224	(0.35)	0.938	(0.62)
Control group	-0.228	(0.04)	-0.250	(0.06)	-0.186	(0.08)	-0.292	(0.13)
Duration dependence since entry into unemployment:								
< 6 months	-3.38	(0.23)	-3.58	(0.29)	-2.99	(0.40)	-3.92	(0.69)
between 6 and 12 months	-2.91	(0.22)	-2.98	(0.28)	-2.64	(0.39)	-3.84	(0.67)
between 12 and 18 months	-3.00	(0.20)	-2.85	(0.26)	-3.12	(0.36)	-4.03	(0.62)
> 18 months	-3.36	(0.13)	-3.25	(0.17)	-3.41	(0.23)	-3.85	(0.43)
Months since selection of sample (> 24 has value=0):								
1	-0.67	(0.20)	-0.51	(0.27)	-0.89	(0.34)	-0.52	(0.60)
2	0.64	(0.18)	0.84	(0.24)	0.34	(0.32)	0.83	(0.56)
3	1.32	(0.18)	1.42	(0.23)	1.12	(0.31)	1.59	(0.54)
4	1.48	(0.17)	1.60	(0.23)	1.26	(0.31)	1.86	(0.54)
5	1.37	(0.18)	1.43	(0.23)	1.24	(0.31)	1.97	(0.54)
6	1.10	(0.18)	1.16	(0.23)	0.96	(0.31)	1.42	(0.55)
7	1.13	(0.18)	1.08	(0.23)	1.11	(0.31)	1.91	(0.55)
8	1.28	(0.17)	1.42	(0.22)	1.02	(0.31)	1.51	(0.54)
9	1.30	(0.16)	1.32	(0.21)	1.24	(0.29)	1.74	(0.50)
10	1.10	(0.16)	1.07	(0.21)	1.14	(0.29)	1.35	(0.51)
11	1.08	(0.16)	1.06	(0.21)	1.08	(0.30)	1.18	(0.53)
12	1.16	(0.16)	1.13	(0.21)	1.21	(0.30)	1.23	(0.53)
13	0.91	(0.17)	0.80	(0.22)	1.08	(0.30)	1.18	(0.54)
14	0.98	(0.15)	0.93	(0.20)	1.04	(0.27)	1.52	(0.42)
15	1.20	(0.10)	1.17	(0.13)	1.25	(0.16)	0.57	(0.40)
16	1.20	(0.10)	1.28	(0.13)	1.01	(0.18)	1.01	(0.35)
17	1.24	(0.11)	1.41	(0.13)	0.87	(0.20)	-0.12	(0.59)
18	1.12	(0.12)	0.97	(0.16)	1.31	(0.17)	1.15	(0.34)
19	0.50	(0.16)	0.47	(0.20)	0.55	(0.25)	0.01	(0.59)
20	0.69	(0.15)	0.70	(0.19)	0.65	(0.25)	0.89	(0.40)
21	0.75	(0.15)	0.83	(0.18)	0.58	(0.26)	0.12	(0.59)
22	0.74	(0.15)	0.64	(0.20)	0.88	(0.23)	0.45	(0.52)
23	0.80	(0.15)	0.92	(0.18)	0.56	(0.28)	0.94	(0.43)
24	0.63	(0.17)	0.64	(0.22)	0.61	(0.28)	-0.11	(0.72)
Non-response types:								
MOVED	0.24	(0.03)	–		–		–	
REFUSE	0.11	(0.03)	–		–		–	
NOCONTACT	0.05	(0.03)	–		–		–	
OTHER	0.38	(0.06)	–		–		–	
# individuals	8004		4706		3298		963	

Explanatory note: standard errors are in parentheses.

Variable	REFUSE=1		NOCONTACT=1		OTHER=1	
Personal characteristics:						
Age/10	-0.285	(0.19)	-0.499	(0.18)	-0.264	(0.34)
Age^2/100	0.034	(0.02)	0.055	(0.02)	0.029	(0.04)
Female	0.224	(0.07)	0.094	(0.07)	0.211	(0.15)
Inner city area	-0.241	(0.08)	-0.197	(0.06)	-0.122	(0.15)
Local unempl.decline	1.590	(0.64)	1.452	(0.59)	-0.853	(1.25)
Control group	-0.150	(0.14)	-0.058	(0.12)	-0.645	(0.31)
Duration dependence since entry into unemployment:						
< 6 months	-2.89	(0.73)	-2.68	(0.61)	-1.42	(1.89)
between 6 and 12 months	-2.36	(0.70)	-2.31	(0.59)	-0.70	(1.86)
between 12 and 18 months	-3.06	(0.63)	-2.84	(0.54)	-1.44	(1.74)
> 18 months	-3.64	(0.41)	-3.13	(0.37)	-2.66	(0.80)
Months since selection of sample (> 24 has value=0):						
1	-1.60	(0.67)	-0.67	(0.52)	-1.92	(1.81)
2	-0.19	(0.60)	0.27	(0.49)	-0.31	(1.74)
3	0.62	(0.58)	1.10	(0.47)	0.18	(1.72)
4	0.80	(0.57)	1.11	(0.47)	0.32	(1.72)
5	0.78	(0.58)	1.00	(0.47)	0.20	(1.73)
6	0.52	(0.58)	0.92	(0.47)	0.11	(1.73)
7	0.33	(0.58)	1.00	(0.47)	0.40	(1.73)
8	0.57	(0.57)	0.92	(0.46)	0.40	(1.71)
9	0.94	(0.52)	1.15	(0.43)	-0.18	(1.66)
10	1.07	(0.52)	1.05	(0.43)	0.27	(1.64)
11	0.89	(0.53)	1.22	(0.43)	-0.50	(1.73)
12	0.91	(0.54)	1.42	(0.43)	-0.03	(1.68)
13	1.04	(0.54)	1.00	(0.45)	0.43	(1.63)
14	1.00	(0.51)	0.62	(0.42)	-0.03	(1.59)
15	1.60	(0.29)	1.35	(0.26)	0.14	(1.12)
16	0.85	(0.40)	1.01	(0.30)	1.11	(0.76)
17	1.06	(0.38)	1.16	(0.29)	0.47	(1.04)
18	1.64	(0.31)	0.91	(0.33)	2.04	(0.57)
19	0.26	(0.59)	0.86	(0.35)	–	
20	-0.11	(0.72)	0.67	(0.39)	–	
21	1.02	(0.43)	0.56	(0.42)	–	
22	0.69	(0.52)	1.03	(0.35)	–	
23	1.16	(0.43)	-1.12	(1.00)	–	
24	0.85	(0.52)	0.70	(0.42)	–	
# individuals	926		1157		252	

Table 8 (continued).

Application of the methodology leads to the conclusion that there is both a selection effect and a causal effect. Thus, survey estimates are biased for two reasons: (1) non-respondents are inherently different, and (2) for a given individual, finding a job before the survey date increases the probability of non-response. Note that the second reason for non-response bias is present even if the population is fully homogeneous. In such a population, there will always be individuals who by chance find a job before the survey date, and this increases their probability of becoming non-respondent.

The causal effect gives rise to an under-estimate of the hazard rate before the survey date, and an under-estimate of its decrease. The selection effect gives rise to a bias everywhere. In our sample, the latter results in an under-estimate of the hazard rate. The causal effect is present for individuals who, according to the survey agency, could not be interviewed because they had moved residence and could not be contacted anymore.

What are the implications of our results for application of methods to correct for non-response bias if one only has access to survey data? First, weighting or imputation of missing values using the survey data does not necessarily remove the bias due to the selection effect and does not remove the bias due to the causal effect. Secondly, Instrumental Variable methods do not remove the bias to due to a causal effect. They do help to remove bias due to selection. However, it is not clear at all what would constitute a valid instrumental variable, i.e. a variable that affects non-response but does not have a direct causal effect on unemployment duration, and is not allowed to depend on unobservables that affect unemployment duration. A candidate is the identity of the interviewer, in particular if this is randomized across sample members (see O'Muircheartaigh and Campanelli, 1999).

To deal with non-response bias due to the causal effect, it may be useful if the survey agency puts maximum effort into tracing individuals who moved residence between the sample selection date and the survey date. For this, it is necessary to record the type of non-response. Finally, one may try to obtain administrative records to inquire the unemployment status of those who moved.

References

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

Brehm, J. (1987), "Who's missing? An analysis of non-response and under coverage in the 1986 national election studies post-election survey", Working paper, National Election Studies.

Bring, J. and K. Carling (1999), "Attrition and misclassification of drop-outs in the analysis of unemployment duration", *Journal of Official Statistics*.

Diggle, P. and M.G. Kenward (1994) "Informative drop-out in longitudinal data analysis", *Applied Statistics*, 43, 49–93.

Dolton P. (1993), "The econometric assessment of training schemes: a critical review", Working paper, University of Newcastle-upon-Tyne.

Dolton, P. and D. O'Neill (1995), "The impact of Restart on reservation wages and long-term unemployment", *Oxford Bulletin of Economics and Statistics*, 57, 451–70.

Dolton, P. and D. O'Neill (1996a), "Unemployment duration and the Restart effect: some experimental evidence", *Economic Journal*, 106, 387–400.

Dolton, P. and D. O'Neill (1996b), "The Restart effect and the return to full-time stable employment", *Journal of the Royal Statistical Society, Series A*, 159, 275–88.

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.

Holmes, T.J. and J.A. Schmitz (1996), "Nonresponse bias and business turnover rates: the case of the Characteristics of Business Owners Survey", *Journal of Business and Economic Statistics*, 14, 231–241.

Horowitz, J.L. and Manski, C.F. (1998), "Censoring of outcomes and regressors due to survey nonresponse: identification and estimation using weights and imputations", *Journal of Econometrics*, 84, 37–58.

Kamionka, T. and G. Lacroix (2003), "Assessing the impact of non-response on the treatment effect in the Canadian self-sufficiency experiment", Working paper, INSEE-CREST, Paris.

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

Lin D.Y. and L.J. Wei (1989), "The robust inference for the Cox proportional hazards model", *Journal of the American Statistical Association*, 84, 1074–1078.

O'Muircheartaigh, C. and P. Campanelli (1999), "A multilevel exploration of the role of interviewers in survey non-response", *Journal of the Royal Statistical Society, Series A*, 162, 437–446.

O'Neill, D. and P. Dolton (2002), "The long-run effects of unemployment monitoring and work search programmes", *Journal of Labor Economics*, 20, 381–403.

Potthoff, R.F., K.G. Manton and M.A. Woodbury (1993), "Correcting for nonavailability bias in surveys by weighting based on number of callbacks", *Journal of the American Statistical Association*, 88, 1197–1207.

Ridder, G. (1987), "The sensitivity of duration models to misspecified unobserved heterogeneity and duration dependence", Working paper, Groningen University.

Ridder, G. and I. Tunalı (1999), "Stratified partial likelihood estimation", *Journal of Econometrics*, 92, 193–232.

Van den Berg, G.J., M. Lindeboom, and G. Ridder (1994), "Attrition in longitudinal panel data, and the empirical analysis of dynamic labour market behaviour", *Journal of Applied Econometrics*, 9, 421–435.

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.

Wang, R., J. Sedransk and J.H. Jinn (1992), "Secondary data analysis when there are missing observations", *Journal of the American Statistical Association*, 87, 952–961.

White, M. and J. Lakey (1992), "The Restart effect: Does active labour market policy reduce unemployment?", Working paper, Policy Studies Institute, London.