Survey Research Methods (2009)
Vol.3, No.3, pp. 139-155
ISSN 1864-3361
http://www.surveymethods.org

Measurement Errors in Retrospective Reports of Event Histories A Validation Study with Finnish Register Data

Marjo Pyy-Martikainen
Åbo Akademi University and Statistics Finland

Ulrich Rendtel
Freie Universitat Berlin

It is well known that retrospective survey reports of event histories are affected by measurement errors. Yet little is known about the determinants of measurement errors in event history data or their effects on event history analysis. Making use of longitudinal register data linked at person-level with longitudinal survey data, we provide novel evidence about 1) type and magnitude of measurement errors in survey reports of event histories, 2) validity of classical assumptions about measurement errors, 3) measurement error bias and 4) effect of measurement accuracy in event history analysis. The classical assumptions about measurement errors are not supported by our measurement error models. Measurement error in both spell durations and spell outcomes are shown to be important causes of bias in an event history analysis. The effects of education and earnings-related unemployment benefit are estimated with sizeable bias. The magnitude of bias in estimated covariate effects does not depend on model type whereas the Cox model produces clearly less biased estimates of baseline hazard compared to the Weibull model. The large bias in the Weibull baseline hazard is shown to be almost entirely due to low measurement accuracy in survey data.

Keywords: measurement error bias, validation study, event history data, unemployment spells

1 Introduction

Event history data are frequently used to analyze person-specific processes such as fertility, poverty and labour market transitions. Event history data typically consist of information about durations of spells in a state of interest (such as poverty, unemployment, having no children), the outcome or terminal event of the spell (transition to non-poverty, to employment or out of labour force, birth of first child), as well as a set of covariates explaining the durations and outcomes.

Event history data can be collected retrospectively by using either a multi-state or an event occurrence framework (see Lawless 2003). In the multi-state framework the reference period of interest is split into shorter time intervals and for each interval, the state occupied by the person is determined. The event occurrence framework asks for dates of specific events such as transitions between the states of interest. The Survey of Labour and Income Dynamics (SLID) uses the event occurrence framework for information on job and jobless spells during the year preceding the interview. The Survey of Income and Program Participation (SIPP) collects information about spells on food stamps program and spells without health insurance by using a multi-state framework where the 4-month reference period is split into time intervals of one month. The European Community Statistics on Income and Living Conditions (EU-SILC) uses a multi-state framework very similar to that used in the European Community Household Panel (ECHP) to collect month-level labour market state information for the year preceding the interview.

It is well-known that retrospective survey reports of event histories are affected by measurement errors (Eisenhower, Mathiowetz and Morganstein 1991; Bound, Brown and Mathiowetz 2001). A measurement error is the discrepancy between the observed value of a variable provided by the survey respondent and its underlying true value. Measurement errors in event histories are manifested as failure to report a spell (omission), reporting a spell that did not occur (overreporting) and misreporting the duration of a spell (misdating) (Mathiowetz 1986; Holt, McDonald and Skinner 1991).[1] In longitudinal surveys, misdating is typically manifested as the heaping of spell starts and ends at the seam between two reference periods, a phenomenon called the seam effect.[2] Even though spell outcomes may also be misreported (e.g. misclassification of a transition out of labour force as a transition to employment), this topic has received little attention in the literature.

Bound, Brown and Mathiowetz (2001) discuss the causes of measurement errors in survey reports. The respondents' ability to report accurately is believed to depend on the cognitive processes related to answering a survey ques-

Contact information: Marjo Pyy-Martikainen, Department of Economics and Statistics, Åbo Akademi University and Statistics Finland, FIN-00022 Statistics Finland, e-mail: marjo.pyy-martikainen@stat.fi

[1] These definitions of measurement error types are somewhat different from those used in a recent study by Jäckle (2008a). She uses definitions that are based on single events and not, as in our case, on spells which consist of two events (initial and terminal) and the time in between.

[2] As pointed out by Jäckle (2008a), seam effects can also arise as a consequence of chopping of long spells spanning three or more reference periods. Chopping may occur e.g. due to misclassification of the state at the middle waves.

tion, the social desirability of the event being reported and on various features of the survey design. The longer the recall period, the more difficult the reporting task and the less salient the event, the more difficult it is to retrieve the information requested. Socially undesirable events tend to go unreported while the opposite is true for socially desirable events. Survey design features, such as mode and method of data collection, interviewer characteristics, frequency and time interval between interviews of a longitudinal survey are likely to affect survey data quality (Groves 1989). However, as noted by Bound, Brown and Mathiowetz (2001), there are no decisive results with respect to the direction and magnitude of measurement errors attributable to these survey design features.

Despite the recognition of the existence of measurement errors in survey-based data on event histories, little is known about their effects on an event history analysis. Skinner and Humphreys (1999) studied spells generated from a Weibull distribution under the assumption of no censoring. They showed both analytically and by a simulation study that the standard estimators of regression coefficients of a Weibull model are approximately unbiased when measurement errors in spells are independent of each other, spell durations and covariates. The estimator of the shape parameter that determines the duration dependence of the hazard is, however, biased. Empirical evidence of measurement error bias in event history analysis concerns residence histories (Courgeau 1992), occupational spells (Hill 1994), time to benefit receipt or to nonemployment (Pierret 2001) and spells of benefit receipt (Jäckle 2008b). The findings from these studies are mixed: both attenuation and strengthening of covariate effects as well as both weakening and strengthening of duration dependence of baseline hazard were detected. Moreover, the studies by Hill (1994) and Pierret (2001) are not able to provide precise information about measurement error bias as they are based on the comparison of two survey data sets having different data collection methods or recall periods. Both data sets are thus subject to measurement errors as well as possibly different non-response patterns.

The studies by Skinner and Humphreys (1999) and Augustin (1999) are the only studies we are aware of that propose methods to adjust for measurement errors in spells. A common feature of the methods proposed is that they rely on rather restrictive assumptions: that spells are generated from certain parametric duration models, there is no censoring and measurement errors are independent of each other, spell durations and covariates.

Our study provides novel evidence of measurement errors in event history data by using longitudinal register data linked at person-level with longitudinal survey data. The combined survey-register longitudinal data enables us to 1) provide information on the type and magnitude of measurement errors in survey reports of event histories, 2) test the plausibility of common assumptions about measurement errors and 3) study measurement error bias in event history analysis. The survey data used in our study is collected by a multi-state framework with a reference period of one year split into one-month intervals. Comparisons of the survey data with register data measured at day level are affected by differences in the measurement accuracy. A fourth aim of our study is to evaluate the separate biasing effects of measurement accuracy and measurement error. This is done by discretizing the day-level register data into month-level data and by comparing results from the three data sets.

The next section discusses the details of the data and the research design. Section 3 studies the magnitude and type of measurement errors in survey reports of event histories. Section 4 specifies models for the process of reporting event histories in order to assess the validity of common assumptions about measurement errors. Section 5 shows how measurement errors affect standard event history analyses. Section 6 evaluates the separate biasing effects of measurement accuracy and measurement error. The findings and implications of our study are discussed in Section 7.

2 The data

Unemployment spells were used as the study variables of interest. We conducted a complete record-check validation study of reports of unemployment spells in the Finnish subset of European Community Household Panel (FI ECHP) data by making use of longitudinal register data linked at person-level with FI ECHP survey data. The register data were assumed to contain true, error-free information about unemployment spells. This is, of course, a simplifying assumption. However, as unemployed persons need to register into the records of an employment office in order to receive unemployment benefits, the register data can be claimed to be more accurate than the survey data.

The ECHP is an input-harmonised sample survey conducted in 15 EU member states between 1994 and 2001 and co-ordinated by Eurostat. The ECHP covers a wide range of topics concerning living conditions, the core topics being income and employment, see Peracchi (2002) for a review of the ECHP. The Finnish ECHP started in 1996. The FI ECHP is documented in Pyy-Martikainen et al. (2004). We used the first five waves of FI ECHP covering the years 1996-2000.

In the ECHP, retrospective labour market state data were collected by a multi-state framework in the form of a month-by-month main activity state calendar obtained for the year preceding the interview. The respondent was first asked whether there were changes in his/her main activity state during the preceding year. If not, the respondent was asked to choose a main activity state from a showcard with 10 options. If there were changes, the respondent was asked to choose a main activity state from the showcard for each month of the year beginning from January:

> "Were there any changes in your main activity in <year>?" [yes/no]
>
> if no:
>
> "What was your main activity state in <year> according to this list?"
>
> if yes:
>
> "What was your main activity state in <month>?"

Interviewers were given the following instructions: if a person's weekly working hours are 15 or more, an option related to employment should be chosen. If a person has had various activity states during a month, employment should be preferred over other states. Thus, in principle, having worked for 15 hours during one week in a specific month is enough to be defined as having been employed in that month. In FI ECHP, a person is defined as unemployed if he/she is without a job, available for work and looking for work through the employment office or newspaper advertisements or some other way. Persons dismissed temporarily are also regarded as unemployed.[3]

Our analysis was based on the FI ECHP sample persons aged 16 or over and thus eligible for a personal interview at the beginning of 1996 (11,641 persons altogether). The sample persons were defined as all members of the initial sample of households. Initial non-respondents (3,146 persons, 27.0%) were excluded because no survey information was available for them (for missingness patterns in the FI ECHP, see Pyy-Martikainen and Rendtel 2008). Temporary dropouts (921 persons, 7.9%) were also excluded because their inclusion would have posed the problem of left-censored spells. Left-censored spells are not only a source of bias in an event history analysis but they would have also artificially increased the heaping of spell starts in January. These restrictions left us with 7,574 (65.1%) sample persons, of whom 4,364 responded in each of the five interviews and 3210 attrited during years 1997 to 2000. For the total respondents, information about unemployment spells was obtained for the five-year period covering the years 1995-1999. For the attriters, information was obtained up to the end of the year that precedes the last interview. Unemployment spells ongoing at the end of the relevant reference period were right-censored. Spells ongoing in January 1995 were dropped because their starting date was unknown. The resulting survey data contain 2719 unemployment spells of 1,482 persons.

Validation data were obtained from the Ministry of Labour's Job-seekers Register. The register contains day-level information about unemployment spell starts and ends. For each spell, the outcome is also registered. In the register, an unemployed job seeker is defined as being without a job and seeking a new job. Registering with the employment office is considered as evidence of seeking a job. Persons dismissed temporarily are regarded as unemployed. Register spells ongoing between 1 January 1995 and 31 December 1999 were linked at person-level to the survey data by personal identification codes.[4] This time period corresponds to the main activity state reference periods of the first five years of the FI ECHP. We constructed register spell data covering, for each person, the same time span as his/her follow-up time in the survey data. For the total respondents, this means using register spells ongoing between 1 January 1995 and 31 December 1999. For the attriters, register spells ongoing between 1 January 1995 and the end of the year preceding the last interview were used. Spells ongoing at the end of the relevant reference period were right-censored. Left-censored spells (ongoing at 1 January 1995) were dropped. Spells lasting at most two days were also dropped as they were not re-

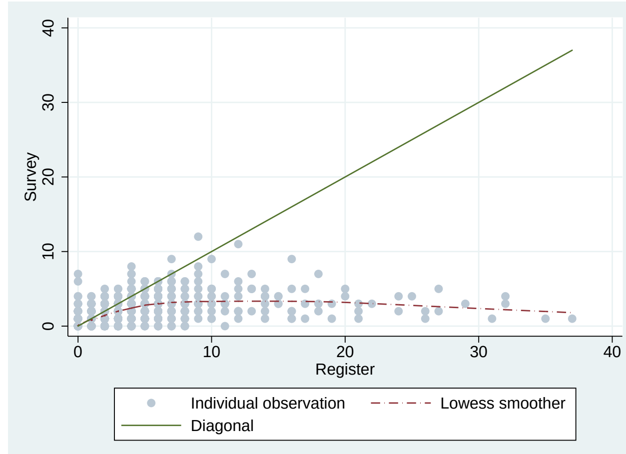

Figure 1. Number of unemployment spells in register and survey over the 5-year follow-up period

garded as true unemployment spells but registrations into the records of the employment office for some legislative reason. The register data contain 6,050 spells of 1,854 persons. Apart from covariates related to the fieldwork, covariates used in subsequent analyses were also taken from various administrative registers.

The magnitude and type of measurement errors were evaluated by person-level comparisons of survey reports and register data. The effects of measurement errors on event history analysis were assessed by comparing estimates based on the two data sources. No survey weights were used in the analysis. Likewise, no attempts were made to correct for the non-response bias. Although estimates based on both survey and register data are affected by non-response, the differences in the estimates cannot be attributed to non-response bias as both the survey and the register data contain the same persons. This was also the main reason why we neglected the use of survey weights in this study.

3 Magnitude and type of measurement errors

Figure 1 shows for each person the number of unemployment spells calculated both from the register and survey data over the 5-year follow-up period. For clarity, the x-axis

[3] The implementation of FI ECHP differs here from Eurostat recommendations, according to which main activity states apart from those related to employment be determined according to self-declaration on the basis of most time spent.

[4] All Finnish citizens are registered in the Finnish Population Information System (FPIS), which is a national register that contains basic information such as name, date of birth and address. As part of the registration process, citizens are issued with a personal identity code (PIC) that is used as a means of identification of persons. The FPIS is used throughout Finnish society's information services and management, including the production of statistics and research.

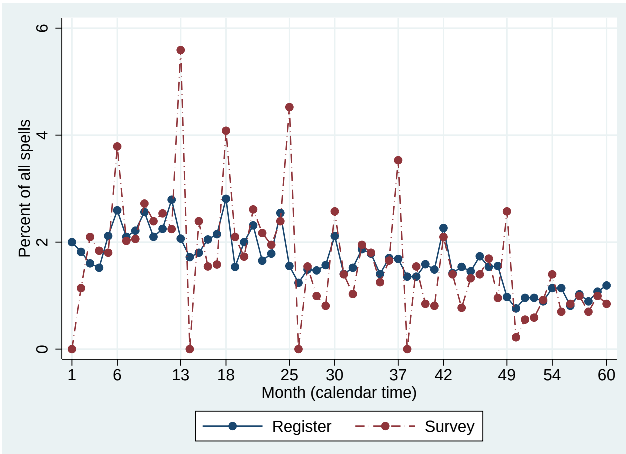

Figure 2. Spell starts in register and survey data

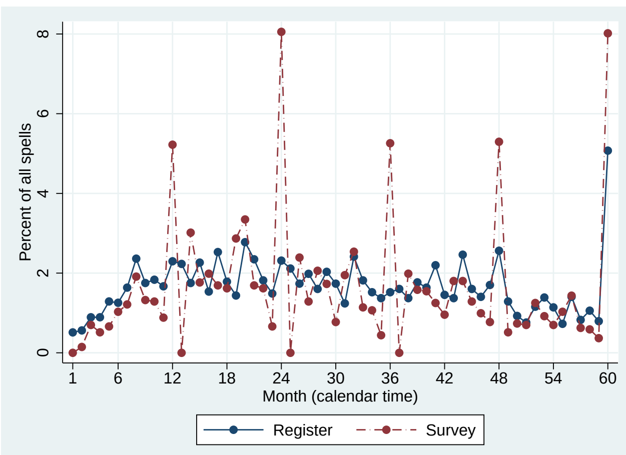

Figure 3. Spell ends in register and survey data

is truncated at 40.[5] A Lowess scatterplot smoother and a diagonal line are also shown.[6] If the number of survey spells and register spells were approximately equal, the points in Figure 1 would lie in the vicinity of the diagonal line. This is not the case, instead, the Lowess line is almost flat implying there is no association between the number of survey and register spells. There is both omitting and overreporting of unemployment spells, omitting being much more important. The omitting of unemployment spells is largely due to the differences in measurement accuracy in survey and register data.

There is a strong heaping effect of unemployment spell starts and ends at the seams between the reference periods of consecutive panel waves (Figures 2 and 3). Unemployment spells tend to start in January and end in December. There is also heaping of spell starts in June. Moreover, there is evidence of backward telescoping of spell starts: following the peaking of spell starts in January there is a lack of spells starting in February. This is likely a consequence of memory decay: events occurring early in the reference period are more difficult to recall.

Table 1: Spell outcomes in register and survey data

	Register		Survey	
Outcome	spells	%	spells	%
Employment	3,238	53.5	1,638	60.2
Subsidised work	720	11.9	58	2.1
OLF[a], Other	1,544	25.5	592	21.8
Attrition	274	4.5	213	7.8
End of follow-up	274	4.5	218	8.0
All	6,050	100.0	2,719	100.0

[a]OLF Out of Labour Force

An often ignored issue is that there may be measurement error in reported spell outcomes as well. In the analysis of unemployment duration, the outcome of interest is often becoming employed. In the survey data, 60.2% of spells ended in becoming employed, whereas in the register data only 53.5% of spells ended for this reason (Table 1). A person-level comparison of register and survey data shows that getting subsidised work is often misclassified by survey respondents as normal employment (Table 2). The higher percentage of survey spells that end because of attrition or end of follow-up reflects the fact that the survey spells are, on average, longer than register spells. The comparison in Table 2 was restricted to persons having one unemployment spell according to both survey and register data during the entire follow-up period. This restriction was done in order to make sure that the spells being compared are the same. The linking of multiple spells per person would have been too unreliable for measurement accuracy and measurement error reasons.

4 Determinants of measurement errors

Because of measurement errors, the true durations T^* are not observed in the survey. The reported durations T can be thought of as consisting of the true duration and a measurement error: $T = T^* + \epsilon$.[7] According to the classical assumptions (see e.g. Bound, Brown and Mathiowetz 2001, Skrondal and Rabe-Hesketh 2004) the measurement errors ϵ have zero mean and are independent of each other, true durations T^* and any covariates explaining T^*. We aimed at testing the validity of these assumptions by modelling $\epsilon = T - T^*$ as a

[5] Only three persons had more than 40 register spells during the follow-up period.

[6] For an introduction to the Lowess procedure see, for example, Fan and Gijbels (1996).

[7] An alternative for the additive measurement error model is the multiplicative model $T = T^* \times \epsilon$, see e.g. Skinner and Humphreys (1999) and Augustin (1999). According to the multiplicative model, the longer the spell lasts the larger the measurement error tends to be. Because of the way unemployment data was collected in the ECHP, there is substantial error in the measurement of short spells also -the reason why we chose to work with the additive model.

Table 2: Misclassification of spell outcomes (sample n: 351)

	Outcome in survey						
Outcome in register	(a)	(b)	(c)	(d)	(e)	(f)	All
(a) Employed	93.2	0.0	2.9	1.0	0.0	2.9	100.0
(b) Subsidised work	85.0	2.5	10.0	0.0	2.5	0.0	100.0
(c) OLF[a]	13.5	1.1	80.9	0.0	2.3	2.3	100.0
(d) Other	50.0	0.0	36.4	4.6	0.0	9.1	100.0
(e) End of follow-up	0.0	0.0	7.9	0.0	92.1	0.0	100.0
(f) Attrition	1.7	0.0	6.8	0.0	0.0	91.5	100.0

[a]OLF Out of Labour Force

function of the true duration and covariates x. We included in our models also some fieldwork-related covariates that are believed to affect measurement errors. Because the survey and register data can be reliably linked only at person-level (and not at spell-level), we defined our measurement error variable as the difference between the sum of unemployment durations from the survey and the sum of unemployment durations from the register, calculated separately for each person $i = 1, \ldots, n$ and for each panel wave $j = 1, \ldots, K_i$ in which the person was unemployed according to both survey and register:

$$\epsilon_{ij} = \sum_{s=1}^{S_{ij}} T_{sij} - \sum_{r=1}^{R_{ij}} T^*_{rij}.$$

S_{ij} and R_{ij} are the numbers of survey and register spells for person i and wave j. ϵ_{ij}'s can be thought of as estimates of cumulated measurement errors in the unemployment spells reported by person i in the wave j interview. To calculate ϵ_{ij}'s, unemployment spells extending over two or more waves were cut at the seams between the waves. We modelled measurement errors in two phases: in the first phase, we modelled the probability of reporting no unemployment spells in a specific wave, given that at least one unemployment spell was found in the register.[8] In the second phase, we modelled the magnitude of cumulated measurement error in the reported unemployment spells, given that at least one unemployment spell was both reported and found in the register.

For the first phase model, assume there are latent variables y^*_{ij} describing the propensity of person i to omit reporting unemployment spells occurring in wave j. The latent variables are assumed to follow the model

$$y^*_{ij} = x_{ij}\beta + \zeta_i + \varepsilon_{ij},$$

where x_{ij} is a $(1 \times p)$ vector of covariates (including a constant) possibly varying with time and person, β is a $(p \times 1)$ vector of the parameters to be estimated and $\zeta_i \sim N(0, \sigma^2_\zeta)$ are person-specific random effects. The random effects ζ_i were incorporated in the model in order to allow for the possibility of correlation of responses by the same person. Error terms ε_{ij} are assumed to be independent and to follow a logistic distribution with mean zero and variance $\sigma^2_\varepsilon = \pi^2/3$. [9] It is assumed that ε_{ij} and ζ_i are uncorrelated. The model can be alternatively expressed as

$$\text{logit}[P(y_{ij} = 1 \mid x_{ij}, \zeta_i)] = x_{ij}\beta + \zeta_i,$$

where

$$y_{ij} = \begin{cases} 1 & \text{if } y^*_{ij} > 0 \\ 0 & \text{if } y^*_{ij} \leq 0. \end{cases}$$

Variables y_{ij} are thus binary variables telling whether person i omits reporting unemployment spells occurring in wave j or not. The intracluster correlation i.e. the correlation among the latent responses by the same person is $\rho = \sigma^2_\zeta/(\sigma^2_\zeta + \frac{\pi^2}{3})$. The model is estimated by maximum likelihood, using a Gauss-Hermite quadrature to approximate the integral over the random terms ζ_i in the log-likelihood function (see e.g. Skrondal and Rabe-Hesketh 2004). In the empirical application, a 12-point quadrature was used.

The second phase model was specified as a random effects linear model:

$$\epsilon_{ij} = x_{ij}\gamma + \nu_i + \delta_{ij},$$

where ϵ_{ij} are the estimates of cumulated measurement errors defined earlier, x_{ij} is a $(1 \times p)$ vector of covariates (including a constant) possibly varying with time and person and γ is a $(p \times 1)$ vector of parameters to be estimated. The assumptions about the random terms ν_i and δ_{ij} are: $\nu_i \sim N(0, \sigma^2_\nu), \delta_{ij} \sim N(0, \sigma^2_\delta)$ and $\text{cov}(\nu_i, \delta_{ij}) = 0$. The intracluster correlation is $\rho = \sigma^2_\nu/(\sigma^2_\nu + \sigma^2_\delta)$. The model was estimated by maximum likelihood.

The distribution of the ϵ_{ij}'s is shown in Figure 4. Compared to a normal distribution (solid line), the empirical distribution (kernel density estimate shown by dashed line) has more mass in the vicinity of zero.

The model estimates are reported in Table 3. The covariates were arranged into three groups: 1) covariates related to

[8] We did not model the probability of overreporting spells given that the register data show none since such a reporting error was found in less than 1 % of person-years.

[9] Variance $\sigma^2_\varepsilon = \pi^2/3$ results from setting the scale parameter of logistic distribution equal to one.

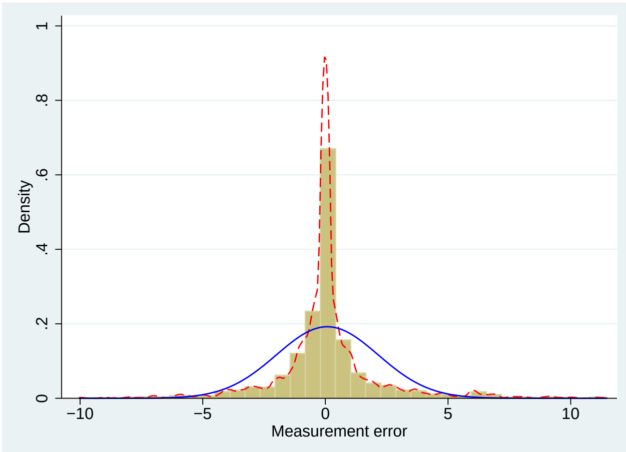

Figure 4. Distribution of cumulated measurement errors

the study variable of interest; 2) covariates used in the event history model (whose estimation is assumed to be the main target of analysis) and 3) covariates related to fieldwork. Covariates in groups 1) and 2) were used to test the classical assumptions about measurement errors. All the covariates are measured at the same year as the dependent variables. The covariates of the event history model are described in section 5. The covariates related to fieldwork include mode of interview (face-to-face vs. telephone), nature of the respondent (self vs. proxy) and the year of interview. Even though the mode of interview and the nature of the respondent are likely to influence the quality of survey reports, it is not clear from theory or empirical evidence how these survey design features affect the direction and magnitude of reporting errors (Bound, Brown and Mathiowetz 2001). Studies which do not control the assignment of respondents to self/proxy or face-to-face/telephone groups are subject to potential self-selection bias (Moore 1988). For example, it may well be that persons with more complex unemployment histories (and, therefore, more prone to reporting errors) are more difficult to reach and, therefore, less likely to give a personal face-to-face interview. However, this problem should be alleviated by the use of covariates related to unemployment history in the measurement error model. During 1996-1997, the fieldwork of the FI ECHP was conducted during February-May, whereas from 1998 onwards the fieldwork period was shifted to autumn. This caused a lengthening of the recall period by several months. Because of memory decay, this was expected to lead to a higher probability of omission and increased magnitude of measurement errors.

Having less than one month of cumulated unemployment time increases the odds of omission by a factor of almost five[10], a consequence of the lower accuracy of measurement and the preference given to activities related to employment on the survey questionnaire (Table 3, Model 1). Each additional month of unemployment decreases the odds of omission by 23.7%. [11] Being a female increases the odds of omission by 27.6%. Age has a u-shaped effect on the probability of omission. The probability decreases until the age of 37 and starts to increase thereafter. A higher probability of omission among the older is likely a consequence of decreasing cognitive ability along with age whereas the young tend to have shorter spells which are both more difficult to recall and more likely too short to be reported in the monthly main activity state scheme. Persons living in Eastern Finland and receiving earnings-related unemployment benefit are more likely than other persons to report unemployment spells. Conducting a proxy interview instead of an interview with the person of interest increases the odds of omission by 72.8%. During the years 1998-2000, the odds of omission are more than double compared to the year 1995. The estimated correlation between the latent responses by the same person is 0.281 and highly significant according to likelihood ratio test.

Both the amount of cumulated unemployment time and the number of unemployment spells affect the magnitude of cumulated measurement errors (Table 3, Model 2). Respondents with cumulative unemployment time less than one month are more likely to overreport which is expected since the reported unemployment time cannot be less than one month. Respondents with longer cumulative unemployment time and more unemployment spells are more likely to underreport. Females are more likely to overreport while persons with an upper secondary or higher education, living outside the capital region and receiving earnings-related unemployment benefit tend to underreport. The estimated correlation between the cumulated measurement errors by the same person is 0.123, again highly significant according to the likelihood ratio test.

5 Effects of measurement errors in event history analysis

Previous sections showed that measurement errors in spell durations are not only of nonnegligible magnitude but also do not conform to the classical independence assumptions. The spell outcomes were also shown to be misclassified. What is the impact of measurement errors in event history analysis based on survey data? This was evaluated by comparing Kaplan-Meier estimates of survival function and estimates from Cox and Weibull proportional hazards models based on register and survey data. The estimates based on register data were used as benchmarks against which the bias due to measurement errors in the survey-based estimates was evaluated.

The study design is described in Table 4. In the first phase, we assessed the impact of measurement errors in spell durations only. Measurement errors in spell durations include not only the effect of misdating of spells but also the effect of omissions and overreporting. Phase 1 analyses ignore spell outcome i.e. study the rate of exit from unemployment regardless of the reason for the exit. The Phase 1 survey data consist of survey spell durations and register

[10] Calculated as exp(1.604)

[11] Calculated as $1 - \exp(-0.270)$

Table 3: Determinants of measurement errors. Model 1: model for the probability of omission. Model 2: model for the magnitude of measurement error

	Model 1		Model 2	
	coef.	se	coef.	se
Constant	**1.551**	(0.528)	*0.724*	(0.420)
Covariates related to the study variable				
Sum of reg UE[a] months	**-0.270**	(0.016)	**-0.125**	(0.011)
Sum of reg UE[a] months lt 1	**1.604**	(0.154)	**1.138**	(0.193)
Number of reg UE[a] spells	0.019	(0.024)	**-0.075**	(0.022)
Covariates of the EH[b] model				
Female	**0.244**	(0.101)	**0.164**	(0.075)
Age	**-0.119**	(0.028)	0.025	(0.022)
Age squared	**0.002**	(0.000)	-0.000	(0.000)
Upper secondary education	0.051	(0.118)	**-0.183**	(0.085)
Higher education	0.227	(0.176)	**-0.524**	(0.135)
Semi urban municipality	0.074	(0.139)	0.000	(0.105)
Rural municipality	-0.050	(0.123)	0.131	(0.091)
Southern Finland	-0.210	(0.142)	**-0.331**	(0.109)
Eastern Finland	**-0.488**	(0.173)	**-0.273**	(0.128)
Central Finland	-0.105	(0.179)	**-0.295**	(0.138)
Northern Finland	-0.109	(0.194)	**-0.374**	(0.146)
Earnings-rel. UE[a] benefit	**-0.381**	(0.105)	**-0.268**	(0.079)
Covariates related to fieldwork				
Telephone interview	*0.199*	(0.115)	0.045	(0.092)
Proxy interview	**0.547**	(0.166)	0.139	(0.136)
Interview in 1997	0.112	(0.123)	-0.115	(0.084)
Interview in 1998	**0.926**	(0.129)	0.096	(0.097)
Interview in 1999	**0.939**	(0.139)	**-0.327**	(0.104)
Interview in 2000	**0.748**	(0.159)	-0.151	(0.121)
Intracluster correlation	**0.281**	(0.034)	**0.123**	(0.020)
-2 log likelihood	4,591		15,456	
number of persons	2,028		1,626	
number of person-years	5,103		3,673	
number of person-years with no reported spells	1,430		-	

Estimates significant at 5% (10%) risk level are displayed in **boldface** (*italics*).

[a]UE unemployment

[b]EH Event History

Table 4: Effects of measurement errors in event history analysis: study design

Phase	Measurement error in	Type of data	Benchmark data from	Survey data from
1	spell duration	spell duration	Register	Survey
		covariates	Register	Register
2	spell duration	spell duration	Register	Survey
	spell outcome	spell outcome	Register	Survey
		covariates	Register	Register

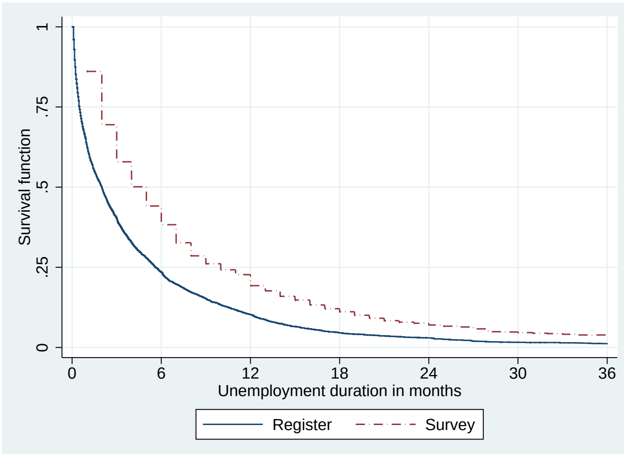

Figure 5. Phase 1. Kaplan-Meier survival function estimates for register and survey data.

covariates. By using the same source of covariates in the two data sets, the differences in estimates could only be attributed to differences in register and survey spells.

In the second phase, measurement errors in survey spell outcomes were taken into account by conducting a cause-specific analysis. In this analysis, the outcome of interest was becoming employed. Phase 2 survey data analyses were conducted using survey spell durations and outcomes, and register covariates.

Results from Phase 1 analyses are shown in the following whereas results from Phase 2 analyses are shown in the Appendix. Figure 5 shows Phase 1 Kaplan-Meier estimates for register and survey data. The Kaplan-Meier estimator is defined as $\hat{S}(t_l) = \prod_{j=1}^{l}(1 - d_j/r_j)$, where t_l is the duration of the lth ordered spell, r_j is the size of the risk set and d_j is the number of spells ending at time t_l. $\hat{S}(t)$ is an estimator of the survival function $S(t) = P(T \geq t)$ that describes the probability of a spell ending later than at time t.[12] In Figure 5 and in all subsequent figures describing the distribution of unemployment spells, the x-axis is truncated at 36 months because very few spells were longer than this. Survey spells end at a lower rate than register spells at all durations. The median duration of a spell is 2 months in the register and 5 months in the survey data. According to the cause-specific Kaplan-Meier estimates (Figure A.6 in Appendix), survey spells end in employment at a lower rate than register spells at durations less than 14 months. Thereafter, the situation is reversed. The crossing of the curves is due to the misclassification of subsidised work as normal employment by survey respondents. If in register data subsidised work is classified as normal employment, the register-based Kaplan-Meier curve lies below the survey-based curve at all durations (results not shown here).

We estimated both Cox and Weibull proportional hazards models in order to assess the measurement error bias in the estimates of the covariate effects and the baseline hazard. A proportional hazards model specifies the hazard function as a product of two terms: $\lambda(t \mid x) = \lambda_0(t)g(x)$. The hazard function $\lambda(t \mid x)$ describes the conditional probability of exit from unemployment, given the covariates and given that the spell has not ended before time t. Function $\lambda_0(t)$ is a baseline hazard specifying the dependency of the hazard function on the duration of interest. The covariates have a multiplicative effect on the hazard function via $g(x)$. Usually $g(x) = \exp(x\beta)$, where x is a $(1 \times p)$ vector of (possibly time-varying) covariates and β is a $(p \times 1)$ vector of parameters.[13] For the Weibull model, the baseline hazard is specified as $\lambda_0(t) = pt^{p-1}$. The shape parameter p determines whether the hazard function is monotonically decreasing ($p < 1$), increasing ($p > 1$) or constant ($p = 1$). The Cox model is estimated by a partial likelihood function that does not involve the $\lambda_0(t)$ terms. The shape of the hazard function is therefore completely unrestricted, which makes the model flexible when compared to fully parameterized models. Both belonging to the class of proportional hazards models, the parameter estimates of Cox and Weibull models are directly comparable. The parameter estimates of proportional hazards models are reported as hazard ratios. The hazard ratio of the i^{th} coefficient is calculated as $\exp(\beta_i)$ and it is interpreted as the ratio of the hazards for a 1-unit increase in the i^{th} covariate.

We hypothesize that estimates of the covariate effects of duration models with a flexible baseline hazard, such as the Cox proportional hazards model, are less biased by measurement errors than estimates from fully parameterized models. For example, it may well be that the effect of heaping of spell starts and ends is absorbed by a flexible baseline hazard. Van den Berg et al. (2004) found that covariate estimates of a Cox proportional hazards model were less biased by non-response than estimates of an exponential or a Weibull model. In order to assess our hypothesis, we compared the size of bias of the survey estimates of the Cox and Weibull proportional hazards models.

Sometimes dummies for heaping months are included as covariates in an attempt to correct for the heaping effect (e.g. Hujer and Schneider 1989, Hunt 1995, Kraus and Steiner 1998). We estimated models both with and without dummies for January and December in order to see whether such heaping dummies protect against measurement error bias in covariate effects or in the baseline hazard.

A set of covariates similar to those used in econometric analyses of unemployment duration was used (see e.g. Meyer 1990, Carling et al. 1996, Abbring et al. 2005). The covariates are spell-specific and they are usually measured at the end of the year preceding the start of the unemployment spell. Age is measured in years. Level of education divides persons into three classes. Basic education corresponds to the completion of comprehensive school. Upper secondary education comprises matriculation examina-

[12] The cause-specific Kaplan-Meier estimator $\hat{S}_c(t)$ describes the probability of an event of type c occurring later than at time t.

[13] A cause-specific proportional hazards model describes the conditional probability of exit due to the event of interest at time t, given that the spell has not ended before t.

tion and upper secondary vocational education. Higher education comprises, for example, tertiary vocational college education and university education. The possible state dependency in unemployment durations is measured by the proportion of time (since 1 January 1995) spent in unemployment before the spell in question. Variation in local labor market conditions is taken into account by information on residential area and statistical grouping of municipalities. The residential area dummies are based on the NUTS2 classification of regions. The statistical grouping of municipalities divides municipalities into urban, semi-urban and rural ones by the proportion of the population living in urban settlements and by the population of the largest urban settlement. Earnings-related unemployment benefit indicates whether a person has received this kind of benefit at the starting year of the unemployment spell. This variable, or variants of it, is often the variable of main interest in an unemployment duration analysis. Other covariates were directly determined by the spell itself and were therefore always taken from the same data source as the spell information. Indicators for the starting year of the unemployment spell aim at capturing the effect of economic fluctuations over time. The January dummy indicates whether the spell started in January (January 1995 excluded). The December dummy is specified as a time-varying indicator variable that gets value 1 in December and zero otherwise.[14]

The estimates from Phase 1 regression analyses are shown in Table 5.[15] The estimates from Phase 2 analyses are shown in Table A.1 in the Appendix. For each model, covariate hazard ratios and their standard errors are reported. Robust estimates of standard errors were calculated in order to take into account the clustering of spells within persons (Lin 1994).

Except for the year dummies, the magnitude and direction of measurement error bias in estimated covariate effects are similar in all estimated models (Table 5). The survey estimates of the year dummies are very much affected by the inclusion of heaping dummies, see footnote 16. The estimated effects of sex, level of education and the dummy for living in Northern Finland have all large biases, the absolute values exceeding 10 percentage points. The effect of education is larger, i.e. further from 1, in the survey-based models, whereas the opposite is true for the effects of sex and living in Northern Finland. Having high education has a markedly stronger effect in the survey-based models: the bias ranging from 18 to 30 percentage points. The shape parameters of the Weibull models are badly biased, which is clearly illustrated in Figure A.2. Both the Cox and the Weibull models show similar effects of January and December dummies. The register spells are less likely to end in December than in other months. This seasonal variation effect in spell ends is masked in the survey estimate by the heaping of spell ends in December. Survey spells beginning in January have a lower hazard of exit, implying longer spell durations while the January dummy has no effect in the register data. This is an indication of backward telescoping of survey spell starts. The effect of January and December dummies in other estimated covariate effects is negligible except for the year dummies of the survey models.[16] The results in Table 5 do not give support to our hypothesis about the Cox model coefficient estimates having smaller bias.

The competing risks analysis with becoming employed as the outcome of interest (Table A.1) shows similar biases in the effects of sex and level of education as before (Table 5). The survey-based models underestimate the effect of receiving earnings-related unemployment benefit by over 28 percentage points. [17] Compared to the analysis that ignores the outcome of interest, the biases in the year dummies and in the shape parameters of the Weibull models have become more pronounced. Moreover, most of the area dummies have now large biases. Introducing an additional source of measurement error, error in spell outcome, has apparently increased the measurement error bias. The effect of the heaping dummies as well as their effect on other estimated covariate effects is similar to before. Again, there is no indication of the Cox model coefficient estimates being more robust with respect to measurement error bias.

Figures A.1 and A.2 show the estimated baseline hazard functions for the Cox model and for the Weibull model without the heaping dummies.[18] For the estimated baseline hazard contributions of the Cox model (see Kalbfleisch and Prentice 2002), a kernel smoother with the Epanechnikov kernel function and a bandwidth of two months was applied (see e.g. Klein and Moeschberger 2003). The hazard func-

[14] Note that the December dummy is defined in a different time scale than the analysis time. The analysis time is specified as time from the beginning of each unemployment spell, whereas the December dummy is specified in calendar time.

[15] For the survey data, we estimated also complementary log-log (cloglog) models corresponding to Cox proportional hazard and Weibull models. The cloglog model is suitable for survival times that are grouped into discrete intervals of time but that are intrinsically continuous. Estimates from the cloglog models were very close to the results from ordinary continuous time Cox and Weibull models.

[16] The effect of the year dummies is weaker in the survey models without the January dummy. This is because the effect of a spell beginning in January is confounded with the effect of the starting year of the spell. Compared to the year 1995, spells beginning during the years 1996-1999 have a higher hazard of exit. The fact that spells beginning in January 1995 are excluded (because they are left-censored) attenuates this effect as spells beginning in January have also a lower hazard of exit.

[17] In the register-based models, the effect of receiving earnings-related unemployment benefit instead of basic unemployment allowance is to increase the exit rate into employment, which is contrary to expectations. A similar effect was found by Hujer and Schneider (1989) and, as noted by Hunt (1995), is likely a result of positive unobserved qualities of receivers of earnings-related unemployment benefit. In a study making use of the same data set, Pyy-Martikainen and Rendtel (2008) estimated a shared frailty Cox hazard model that controls for person-specific unobserved heterogeneity. The effect of receiving earnings-related unemployment benefit was to lower the hazard of exit, which is in accordance with the results from search theory.

[18] The estimated baseline hazards from the models including heaping dummies are almost identical and, therefore, not reported.

Table 5: Phase 1. Proportional hazards models. Measurement error in spell duration only

	1. Cox			2. Cox, heaping dummies			3. Weibull			4. Weibull, heaping dummies		
	Register	Survey		Register	Survey		Register	Survey		Register	Survey	
Variable	hr[a] (se)	hr (se)	bias[b]	hr (se)	hr (se)	bias	hr (se)	hr (se)	bias	hr (se)	hr (se)	bias
Female	**1.128** (0.067)	0.925 (0.045)	-20.3	*1.121* (0.667)	0.944 (0.047)	-17.7	**1.129** (0.069)	0.921 (0.052)	-20.8	*1.124* (0.069)	0.936 (0.053)	-18.8
Age	1.014 (0.017)	*1.024* (0.014)	1.0	1.015 (0.017)	*1.026* (0.015)	1.1	1.014 (0.018)	1.026 (0.016)	1.2	1.015 (0.018)	*1.029* (0.017)	1.4
Age squared	*1.000* (0.000)	**0.999** (0.000)	-0.1	*1.000* (0.000)	**0.999** (0.000)	-0.1	*1.000* (0.000)	**0.999** (0.000)	-0.1	*1.000* (0.000)	**0.999** (0.000)	-0.1
Upper secondary educ.	1.052 (0.074)	**1.128** (0.065)	7.6	1.052 (0.074)	**1.120** (0.065)	6.8	1.051 (0.076)	**1.153** (0.076)	10.2	1.051 (0.076)	**1.145** (0.076)	9.4
Higher education	**1.373** (0.134)	**1.555** (0.124)	18.2	**1.365** (0.134)	**1.584** (0.127)	21.9	**1.390** (0.140)	**1.662** (0.152)	27.2	**1.383** (0.139)	**1.684** (0.155)	30.1
Proportion of UE												
tabfnm[c] time	1.001 (0.001)	*0.998* (0.001)	-0.3	1.001 (0.001)	**0.997** (0.001)	-0.4	1.001 (0.001)	**0.998** (0.001)	-0.3	1.001 (0.001)	**0.997** (0.001)	-0.4
Semi urban municipality	1.135 (0.106)	1.071 (0.075)	-6.4	1.137 (0.107)	1.066 (0.077)	-7.1	1.140 (0.107)	1.067 (0.088)	-7.3	1.141 (0.108)	1.054 (0.089)	-8.7
Rural municipality	1.029 (0.070)	0.970 (0.058)	-5.9	1.028 (0.070)	0.979 (0.060)	-4.9	1.032 (0.072)	0.961 (0.066)	-7.1	1.031 (0.072)	0.972 (0.067)	-5.9
Southern Finland	1.189 (0.164)	**1.213** (0.085)	2.4	1.194 (0.165)	**1.162** (0.081)	-3.2	1.200 (0.165)	**1.224** (0.097)	2.4	1.204 (0.166)	*1.158* (0.091)	-4.6
Eastern Finland	1.151 (0.174)	1.134 (0.095)	-1.7	1.157 (0.175)	1.091 (0.092)	-6.6	1.160 (0.175)	1.161 (0.109)	0.1	1.164 (0.176)	1.110 (0.104)	-5.4
Central Finland	1.069 (0.146)	1.145 (0.106)	7.6	1.074 (0.146)	1.105 (0.104)	3.1	1.079 (0.148)	1.127 (0.120)	4.8	1.082 (0.148)	1.078 (0.116)	-0.4
Northern Finland	*1.331* (0.200)	1.156 (0.110)	-17.5	*1.334* (0.201)	1.136 (0.108)	-19.8	**1.356** (0.205)	1.181 (0.125)	-17.5	**1.358** (0.205)	1.159 (0.121)	-19.9
Earnings-rel. UE[a] benefit	1.068 (0.067)	1.010 (0.053)	-5.8	1.063 (0.067)	0.996 (0.053)	-6.7	1.070 (0.069)	1.022 (0.062)	-4.8	1.067 (0.068)	1.003 (0.062)	-6.4
Year 1996	1.073 (0.049)	**1.137** (0.063)	6.4	1.069 (0.050)	**1.228** (0.071)	15.9	1.079 (0.052)	**1.163** (0.079)	8.4	1.076 (0.052)	**1.268** (0.090)	19.2
Year 1997	**1.151** (0.062)	**1.156** (0.072)	0.5	**1.145** (0.063)	**1.275** (0.084)	13	**1.168** (0.065)	*1.207* (0.088)	3.9	**1.164** (0.066)	**1.361** (0.104)	19.7
Year 1998	**1.326** (0.088)	**1.342** (0.090)	1.6	**1.314** (0.089)	**1.421** (0.102)	10.7	**1.365** (0.093)	**1.486** (0.111)	12.1	**1.357** (0.094)	**1.639** (0.129)	28.2
Year 1999	**1.363** (0.118)	1.003 (0.096)	-36.0	**1.297** (0.115)	0.943 (0.111)	-35.4	**1.427** (0.125)	**1.225** (0.122)	-20.2	**1.363** (0.122)	**1.320** (0.148)	-4.3
Begin in January				1.003 (0.057)	**0.716** (0.041)	-28.7				0.994 (0.059)	**0.678** (0.044)	-31.6
December				**0.532** (0.037)	**1.341** (0.103)	80.9				**0.538** (0.037)	1.119 (0.079)	58.1
Weibull shape							**0.760** (0.020)	**1.081** (0.017)	32.1	**0.759** (0.020)	**1.087** (0.018)	32.8
-2 log pseudolikelihood	85,037	31,867		84,922	31,804		21,407	7,493		21,296	7,442	

number of register spells: 6,050 of which number of events: 5,502
number of survey spells: 2,717 of which number of events: 2,287
Estimates significant at 5% (10 %) risk level are displayed in **boldface** (*italics*).
Standard errors adjusted for clustering of spells within persons.

[a] hr hazard ratio. $H_0 : hr = 1$

[b] bias= $100 \times$ (Survey − Register)

[c] UE Unemployment

tion estimates were calculated setting the continuous variables at their mean values and the dummy variables to zero.[19] The survey baseline hazard from the Cox model is close to the register baseline hazard, although it displays a tendency towards underestimation. Due to the lower accuracy of measurement of spells in the survey, the survey baseline hazard is not able to reach the spike displayed by the register at the shortest durations. The survey baseline hazard from the Weibull model is nearly constant while the register baseline hazard shows negative duration dependence. The survey-based Weibull baseline hazard thus leads to erroneous conclusions about the duration dependence while the Cox baseline hazards from survey and register both display negative duration dependence. With respect to the estimation of the baseline hazard, the flexibility of the Cox model is clearly an advantage. As will be shown in section 6, the shape of the Weibull hazard is completely determined by spells shorter than one month.

Taking spell outcome into account markedly increases the measurement error bias in the estimated baseline hazard. The survey-based cause-specific hazard from the Cox model severely overestimates the true baseline hazard (Figure A.7 in Appendix). Moreover, the survey-based hazard is more kinked than the corresponding register-based hazard. If in the register data subsidised work is classified as employment, the register baseline hazard shifts somewhat upwards and exhibits similar kinks (results not shown here). The cause-specific Weibull baseline hazards from survey and register data lead again to different conclusions about the duration dependence (Figure A.8 in Appendix).

6 Effect of measurement accuracy

The previous section showed that the survey-based estimates of both the distribution of spells and of covariate effects were biased. This is a consequence of not only measurement errors but also of the way event history data were collected in the survey. In ECHP, information on main activity state is collected at the accuracy of one month. Moreover, as employment is preferred over unemployment, it is difficult to obtain information on unemployment spells shorter than one month. We aimed at separating the biases due to measurement error and measurement accuracy by discretizing the register spells and repeating the analyses with discretized data. Discrepancies between estimates based on survey data and discretized register data could then be taken as estimates of bias due to measurement error. Respectively, bias due to measurement accuracy could be evaluated by comparing results from original and discretized register data.

Register data were discretized in the following way: for each month, the number of unemployment days was calculated. If the number of days was at least 28, the register-based state of that month was defined as unemployed. The unemployment spell duration was then calculated by using these monthly indicators of unemployment state. Obtaining spell outcome information was not possible as this would have necessitated register information about other spells than unemployment. This information was not available in our data. Spells ongoing at December were censored if the person in question attrited from the survey the following year or if the spell was ongoing at the end of the reference period (December 1999).

Figure A.3 shows that the upward bias in the survey-based survival curve is to a large extent due to the lack of short spells. The survey and register curves are now for all practical purposes equal at durations less than approximately 8 months. The median spell duration in the discretized register data is 4 months, which is only one month shorter than in the survey data.

The estimates from the proportional hazard models based on the discretized register data, as well as the estimated biases due to measurement error and measurement accuracy, are shown in Table 6. Coarsening the measurement accuracy in register data diminishes the effect of being a female. This is due to the fact that females have a shorter median unemployment duration and thus, dropping out short spells affects more females than males. Measurement error operates in the same direction as measurement accuracy, attenuating the effect of being a female. Both measurement accuracy and measurement error cause a positive bias in the effect of higher education, the bias due to measurement error being markedly larger. This suggests that persons with higher education tend to underreport spell durations, a result supported by the model for the magnitude of measurement error (see Table 3). The area dummies have large biases due to both measurement accuracy and measurement error, but the biases tend to work in opposite directions. As for the time dummies, the biases due to measurement error and measurement accuracy are largest for year 1999, but they mostly work in opposite directions.[20] The biases in January and December dummies show that the heaping of spell starts and ends really is a measurement error and not a measurement accuracy problem. By contrast, the bias in the shape parameters of the Weibull models is for the most part due to measurement accuracy and, more specifically, the lack of short spells.

The estimated baseline hazard functions from the Cox proportional hazard models without time dummies are shown in Figure A.4. The lack of short spells in discretized register data and in survey data leads to underestimation of the baseline hazard for durations shorter than six months. For longer durations, the biases due to measurement accuracy and mea-

[19] This corresponds to a 36-year-old male with a basic level of education living in an urban municipality in the capital region, receiving basic unemployment allowance and having been unemployed 34 percent of the follow-up time before the spell in question. His unemployment spell started in 1995 the models including time dummies, the spell did not start in January and did not include December.

[20] The effect of the dummy for year 1999 increases markedly when the heaping dummies are included (models 2 and 4). In the discretized register data, an unemployment spell is ongoing in December 1999 is always censored because it is the last month of the follow-up period. this attenuates the effect of year dummies and, especially the effect of year 1999, in models not containing heaping dummies.

Table 6: Phase 1. Proportional hazards models. Measurement error in spell duration only

	1. Cox			2. Cox, heaping dummies			3. Weibull			4. Weibull, heaping dummies		
	Register2	MA	ME	Register2	MA	ME	Register2	MA	ME	Register2	MA	ME
Variable	hr[a] (se)	bias[b]	bias[c]	hr (se)	bias	bias	hr (se)	bias	bias	hr (se)	bias	bias
Female	0.995 (0.043)	-13.3	-7.0	0.985 (0.042)	-13.6	-4.1	1.001 (0.048)	-12.8	-8.0	0.989 (0.047)	-13.5	-5.3
Age	1.017 (0.013)	0.3	0.7	1.017 (0.129)	0.2	0.9	1.023 (0.015)	0.9	0.3	1.020 (0.014)	0.5	0.9
Age squared	**0.999** (0.000)	-0.1	0.0	**0.999** (0.000)	-0.1	0.0	**0.999** (0.000)	-0.1	0.0	**0.999** (0.000)	-0.1	0.0
Upper secondary educ.	1.061 (0.054)	0.9	6.7	1.076 (0.054)	2.4	4.4	1.074 (0.061)	2.3	7.9	1.087 (0.061)	3.6	5.8
Higher education	**1.420** (0.109)	4.7	13.5	**1.402** (0.106)	3.7	18.2	**1.461** (0.125)	7.1	20.1	**1.443** (0.122)	6.0	24.1
Proportion of UE[d] time	**0.997** (0.001)	-0.4	0.1	**0.998** (0.001)	-0.3	-0.1	**0.997** (0.001)	-0.4	0.1	**0.997** (0.001)	-0.4	0.0
Semi urban municipality	**1.200** (0.067)	6.5	-12.9	**1.190** (0.065)	5.3	-12.4	**1.248** (0.076)	10.8	-18.1	**1.229** (0.073)	8.8	-17.5
Rural municipality	**1.119** (0.060)	9.0	-14.0	*1.106* (0.058)	7.8	-12.7	**1.137** (0.068)	10.5	-17.6	*1.113* (0.065)	8.2	-14.1
Southern Finland	1.075 (0.065)	-11.4	13.8	1.081 (0.064)	-11.3	8.1	1.076 (0.072)	-12.4	14.8	1.083 (0.070)	-12.1	7.5
Eastern Finland	1.052 (0.076)	-9.9	8.2	1.069 (0.075)	-8.8	2.2	1.047 (0.084)	-11.3	11.4	1.068 (0.083)	-9.6	4.2
Central Finland	1.034 (0.080)	-3.5	11.1	1.045 (0.080)	-2.9	6.0	1.019 (0.088)	-6.0	10.8	1.035 (0.087)	-4.7	4.3
Northern Finland	1.134 (0.095)	-19.7	2.2	1.130 (0.094)	-20.4	0.6	1.139 (0.105)	-21.7	4.2	1.135 (0.103)	-22.3	2.4
Earnings-rel. UE[f] benefit	1.039 (0.046)	-2.9	-2.9	1.026 (0.045)	-3.7	-3.0	1.046 (0.053)	-2.4	-2.4	1.036 (0.051)	-3.1	-3.3
Year 1996	**1.157** (0.056)	8.4	-2.0	**1.168** (0.058)	9.9	6.0	**1.196** (0.067)	11.7	-3.3	**1.214** (0.070)	13.8	5.4
Year 1997	**1.259** (0.067)	10.8	-10.3	**1.288** (0.070)	14.3	-1.3	**1.326** (0.080)	15.8	-11.9	**1.362** (0.083)	19.8	-0.1
Year 1998	**1.366** (0.079)	4.0	-2.4	**1.465** (0.083)	15.1	-4.4	**1.493** (0.096)	12.8	-0.7	**1.624** (0.102)	26.7	1.5
Year 1999	**1.238** (0.100)	-12.5	-23.5	**1.857** (0.141)	56.0	-91.4	**1.492** (0.126)	6.5	-26.7	**2.209** (0.171)	84.6	-88.9
Begin in January				0.929 (0.059)	-7.4	-21.3				0.900 (0.064)	-9.4	-22.2
December				**0.364** (0.033)	-16.8	97.7				**0.358** (0.031)	-18.0	76.1
Weibull shape							**1.139** (0.015)	37.9	-5.8	**1.163** (0.015)	40.4	-7.6
-2 log pseudolikelihood	42,824			42,610			9,261			9,031		

number of discretized register spells: 3,448 of which number of events: 2,998
Estimates significant at 5% (10 %) risk level are displayed in **boldface** (*italics*).
Standard errors adjusted for clustering of spells into persons.

[a]hr hazard ratio. $H_0 : hr = 1$

[b]bias due to measurement accuracy,

calculated as = $100 \times$ (Register2 − Register)

[c]bias due to measurement error,

calculated as = $100 \times$ (Survey − Register2)

[d]UE unemployment

surement error work in opposite directions. Measurement accuracy creates a small positive bias leading to overestimation of the baseline hazard. The hazard spikes are however correctly placed in time. As measurement error creates a large negative bias, the joint effect of these two sources of bias leads to the underestimation of the baseline hazard. The effect of measurement error is, moreover, to flatten the shape of the baseline hazard. Figure A.5 shows the estimated Weibull hazard functions. Measurement accuracy has a dominating effect here: the exclusion of short spells leads to a badly biased shape of the baseline hazard, while measurement error only leads to slight underestimation of the level of the hazard.

7 Conclusion

Our study provided novel evidence on the existence, determinants and effects of measurement errors in event history analysis. Using longitudinal register data linked at person-level with longitudinal survey data, we were able to 1) provide information on the type and magnitude of measurement errors in retrospective survey reports of event histories, 2) assess the plausibility of classical assumptions about measurement errors, 3) study measurement error bias and 4) study the effect of measurement accuracy on event history analysis based on survey data.

Unemployment spells obtained from the FI ECHP data were used as the study variables of interest. Register data on unemployed jobseekers were used as the validation data. Available for all sample persons, having a definition of unemployment similar to that in the survey and giving precise information not only about the beginning and ending dates of each spell but also about spell outcomes, the validation data used in this study can be considered as being of outstanding quality.

According to our analysis, unemployment spells were subject to both omissions and, to a lesser extent, overreporting. Spell starts and ends were strongly heaped at the seams between the reference periods of consecutive panel waves. These findings are consistent with earlier studies on measurement errors in unemployment spells (Mathiowetz 1986, Mathiowetz and Duncan 1988, Kraus and Steiner 1998). A usually unnoticed issue is the classification error in reported spell outcomes. There was an excess of exits into employment in the survey data due to the fact that exits into subsidised work were often misclassified by respondents as becoming employed.

The model for the magnitude of measurement errors showed that the classical assumptions about measurement errors are not valid: cumulated measurement errors were correlated across survey waves, with variables related to true spells and with covariates used to explain the duration of spells. The model for the probability of omission of spells exhibited similar dependencies. Conducting a proxy interview instead of an interview with the person of interest and the lengthening of the recall period increased sharply the probability of omission while these survey design features had no effect on the magnitude of the cumulated measurement error.

The measurement error bias in an event history analysis was shown to result from both erroneously measured spell durations and misclassified spell outcomes. The survey data overestimated both the median duration of unemployment and the median time to becoming employed. There was no evidence of an overall attenuation effect of measurement errors on the estimated covariate effects, a result consistent with earlier empirical studies. The effect of education and in the competing risks analysis also the effect of receiving earnings-related unemployment benefit were estimated with sizeable bias. As for the estimated covariate effects, neither dummies for the heaping months nor the more flexible Cox model did protect against measurement error bias whereas the baseline hazard was much more accurately estimated by the Cox model. The survey-based estimates of the Weibull baseline hazard led to erroneous conclusions about the duration dependency of the hazard. The misclassification of spell outcomes was shown to be an important cause of bias in the estimate of cause-specific baseline hazard from the Cox model.

The survey data used in our study is collected by a multi-state framework with an accuracy of one month. Comparisons of the survey data with register data measured at day level are affected by differences in the measurement accuracy. Our attempts to separate the bias due to measurement accuracy and measurement error showed that measurement accuracy is an important source of bias in both the estimates of the distribution of spells and of covariate effects. Most notably, the bias in the Weibull baseline hazard was shown to be almost entirely due to lower measurement accuracy.

It is well-known that retrospective survey reports of event histories are affected by measurement errors. A few recent studies – including ours – suggest that measurement errors in survey spells have a non-negligible effect on an event history analysis. This has implications both for the survey organization collecting event history data and for the data analyst. In the light of our study, the use of proxy interviews should be kept to a minimum as they tend to lead to spell omissions. For the same reason, the time interval between the survey interview and the end of the reference period of the event history questions should be kept as short as possible. Paying attention to a careful definition of states in a multi-state data collection framework is important in order to avoid misclassification errors. Information about the spell distributions should be taken into account already in the questionnaire design phase in order to find an appropriate level of measurement accuracy. Our results suggest that attempts to control for heaping effects in the analysis phase by the inclusion of dummies for the heaping months are not helpful. As for the estimated covariate effects, the Cox model did not turn out to be more robust with respect to measurement error bias than the Weibull model. This contradicts earlier empirical findings concerning the robustness of Cox model with respect to non-response bias (van den Berg et al. 2004). However, the flexibility of the Cox model was clearly advantageous in the estimation of the baseline hazard. There have been only few attempts to develop methods to adjust for bias due to measurement error in spells in event history analysis. Moreover, the proposed methods assume that measurement errors are independent of each other, of the true durations and of the covariates used to explain the durations. Our study suggests that methods making more realistic assumptions need to be developed in order to effectively adjust for measurement errors.

References

Abbring, J., van den Berg, G., & van Ours, J. (2005). The Effect of unemployment insurance sanctions on the transition rate from unemployment to employment. *The Economic Journal*, *115*, 602-630.

Augustin, T. (1999). *Correcting for measurement error in parametric duration models by quasi-likelihood.* Collaborative Research Center 386, Discussion Paper 157.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In J. Heckman & E. Leamer (Eds.), *Handbook of econometrics* (Vol. 5, p. 3705-3833). Amsterdam: Elsevier.

Carling, K., Edin, P., Harkman, A., & Holmlund, B. (1996). Unemployment duration, unemployment benefits and labor market programs in Sweden. *Journal of Public Economics*, *59*, 313-334.

Courgeau, D. (1992). Impact of response errors on event history analysis. *Population: an English Selection*, *4*, 97-110.

Eisenhower, D., Mathiowetz, N., & Morganstein, D. (1991). Recall error: Sources and bias reduction techniques. In P. Biemer, R. G. L. Lyberg, N. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (p. 127-144). New York: Wiley.

Fan, J., & Gijbels, I. (1996). *Local polynomial modeling and its applications*. London: Chapman and Hall.

Groves, R. (1989). *Survey errors and survey costs*. New York: Wiley.

Hill, D. (1994). The relative empirical validity of dependent and independent data collection in a panel survey. *Journal of Official Statistics*, *10*, 359-380.

Holt, D., McDonald, J., & Skinner, C. (1991). The Effect of measurement errors on event history analysis. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (p. 127-144). New York: Wiley.

Hujer, R., & Schneider, H. (1989). The Analysis of labor market mobility using panel data. *European Economic Review*, *33*, 530-536.

Hunt, J. (1995). The Effect of unemployment compensation on unemployment duration in Germany. *Journal of Labor Economics*, *13*, 88-120.

Jäckle, A. (2008a). *Measurement error and data collection methods: Effects on estimates from event history data.* ISER Working Paper Series 13.

Jäckle, A. (2008b). *The causes of seam effects in panel surveys.* ISER Working Paper Series 14.

Kalbfleisch, J., & Prentice, R. (2002). *The statistical analysis of failure time data* (2nd ed.). New York: Wiley.

Klein, J., & Moeschberger, M. (2003). *Survival analysis. Techniques for censored and truncated data* (2nd ed.). New York: Springer.

Kraus, F., & Steiner, V. (1998). *Modelling heaping effects in unemployment duration models – With an application to retrospective event data in the German Socio-Economic Panel*. Stuttgart: In Jahrbücher für Nationalökonomie und Statistik. Lucius & Lucius.

Lawless, J. (2003). Event History Analysis and Longitudinal Surveys. In R. Chambers & C. Skinner (Eds.), *Analysis of Survey Data.* Chichester: Wiley.

Lin, D. (1994). Cox regression analysis of multivariate failure time data: The marginal approach. *Statistics in Medicine*, *13*, 2233-2247.

Mathiowetz, N. (1986). *The problem of omissions and telescoping error: New evidence from a study of unemployment.* Proceedings of the section on Survey Research Methods, American Statistical Association.

Mathiowetz, N., & Duncan, G. (1988). Out of work, out of mind: Response errors in retrospective reports of unemployment. *Journal of Business and Economic Statistics*, *6*, 221-229.

Meyer, B. (1990). Unemployment insurance and unemployment spells. *Econometrica*, *58*, 757-782.

Moore, J. (1988). Self/proxy response status and survey response quality. *Journal of Official Statistics*, *4*, 155-172.

Peracchi, F. (2002). The European Community Household Panel: A review. *Empirical Economics*, *27*, 63-90.

Pierret, C. (2001). Event history data and survey recall. *Journal of Human Resources*, *36*, 439-466.

Pyy-Martikainen, M., & Rendtel, U. (2008). Assessing the impact of initial nonresponse and attrition in the analysis of unemployment duration with panel surveys. *Advances in Statistical Analysis*, *92*, 297-318.

Pyy-Martikainen, M., Sisto, J., & Reijo, M. (2004). *The ECHP study in Finland. Quality report*. Helsinki: Statistics Finland.

Skinner, C., & Humphreys, K. (1999). Weibull regression for lifetimes measured with error. *Lifetime Data Analysis*, *5*, 23-37.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling*. Boca Raton, FL: Chapman and Hall.

van den Berg, G., Lindeboom, M., & Dolton, P. (2004). *Survey non-response and unemployment duration.* IFAU working paper 2004:12.

Appendix

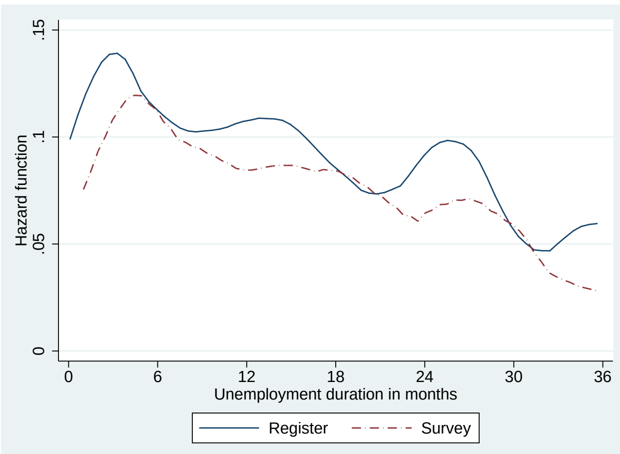

Figure A.1. Phase 1. Estimated baseline hazard function from the Cox model. No heaping dummies.

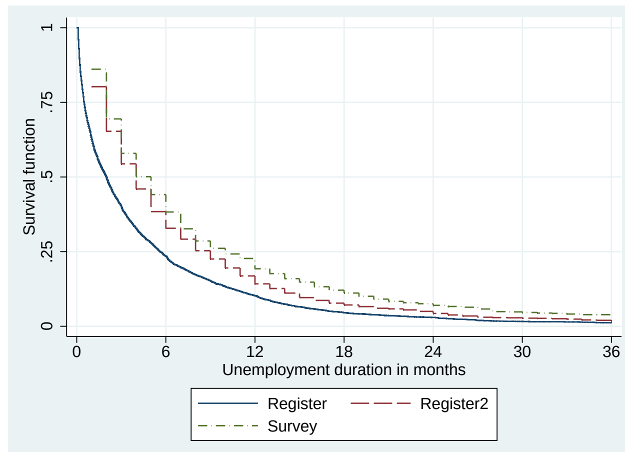

Figure A.3. Phase 1. Kaplan-Meier survival function estimates. Register 2: discretized register data.

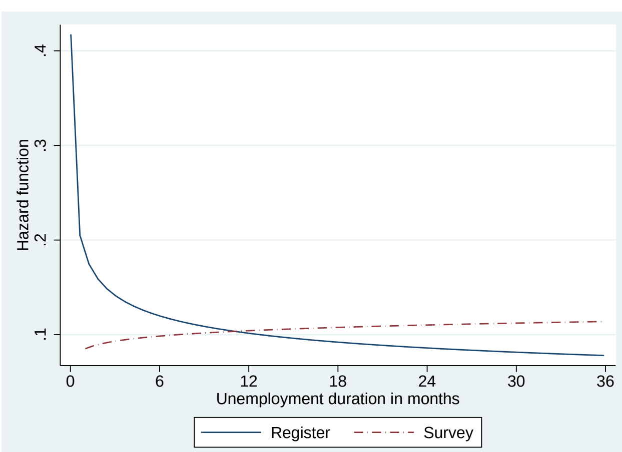

Figure A.2. Phase 1. Estimated baseline hazard function from the Weibull model. No heaping dummies.

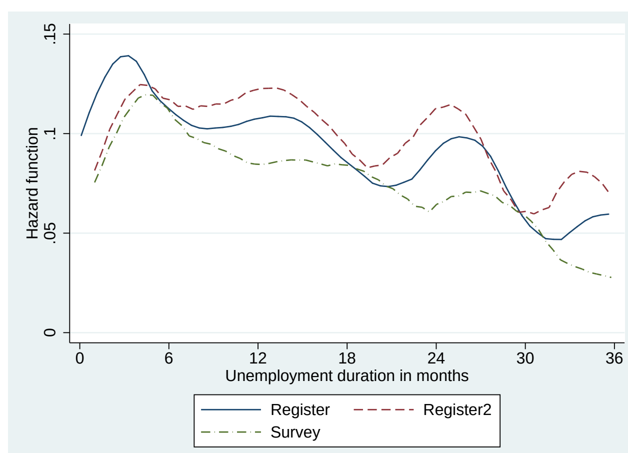

Figure A.4. Phase 1. Estimated baseline hazard function from the Cox model. No time dummies. Register 2: discretized register data.

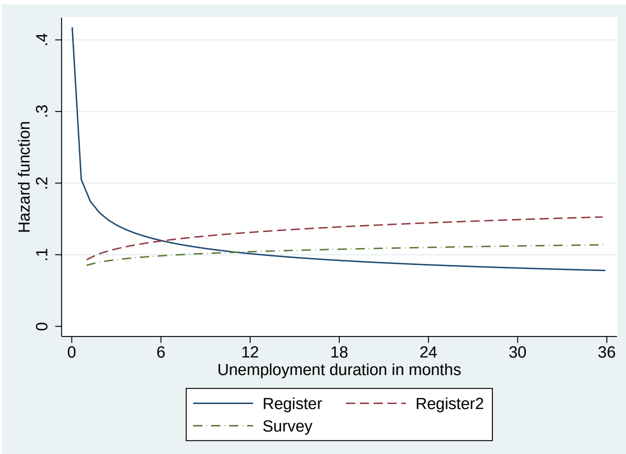

Figure A.5. Phase 1. Estimated baseline hazard function from the Weibull model. No time dummies. Register 2: discretized register data.

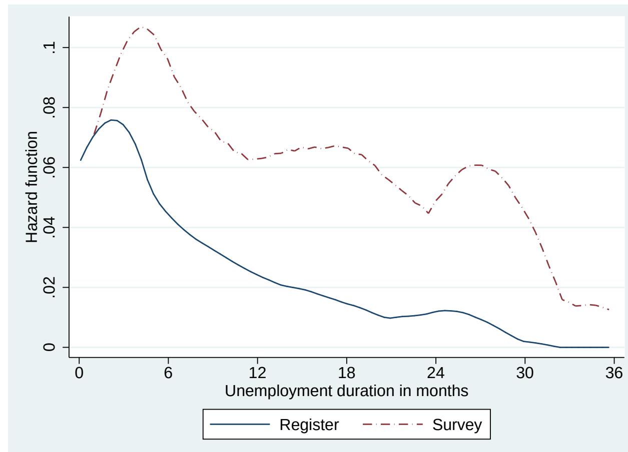

Figure A.7. Phase 2. Estimated baseline hazard function from the Cox proportional hazards model. No time dummies. Outcome of interest: becoming employed.

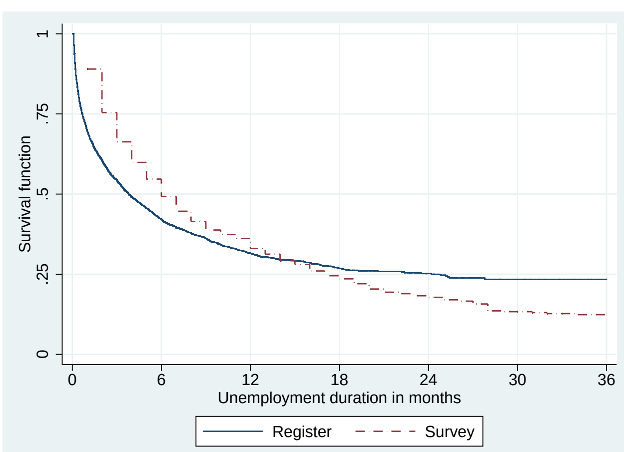

Figure A.6. Phase 2. Kaplan-Meier survival function estimates for the register and survey data. Outcome of interest: becoming employed.

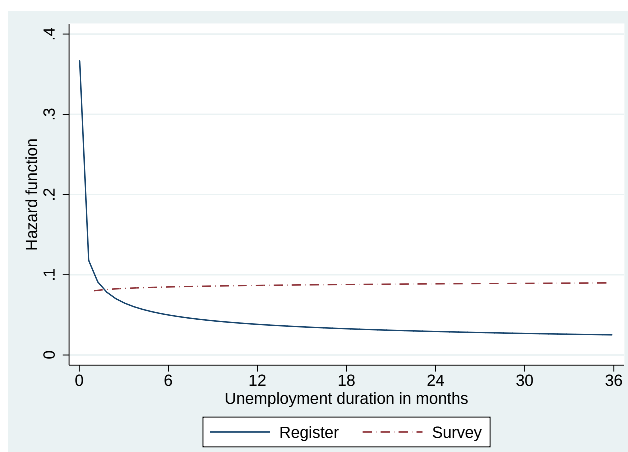

Figure A.8. Phase 2. Estimated baseline hazard function from the Weibull model. No time dummies. Outcome of interest: becoming employed.

Table A.1: Phase 2. Proportional hazards models. Outcome of interest: becoming employed

	1. Cox			2. Cox, heaping dummies			3. Weibull			4. Weibull, heaping dummies		
	Register	Survey		Register	Survey		Register	Survey		Register	Survey	
Variable	hr[a] (se)	hr (se)	bias[b]	hr (se)	hr (se)	bias	hr (se)	hr (se)	bias	hr (se)	hr (se)	bias
Female	1.039 (0.099)	**0.798** (0.047)	-24.2	1.033 (0.099)	**0.807** (0.048)	-22.6	1.033 (0.102)	**0.794** (0.052)	-23.9	1.028 (0.101)	**0.803** (0.053)	-22.4
Age	**1.098** (0.032)	**1.129** (0.020)	3.2	**1.099** (0.032)	**1.134** (0.020)	3.5	**1.105** (0.033)	**1.131** (0.022)	2.6	**1.105** (0.033)	**1.135** (0.023)	3.0
Age squared	**0.999** (0.000)	**0.998** (0.000)	-0.1	**0.999** (0.000)	**0.998** (0.000)	-0.1	**0.999** (0.000)	**0.998** (0.000)	-0.1	**0.999** (0.000)	**0.998** (0.000)	-0.1
Upper secondary educ.	1.034 (0.118)	1.106 (0.077)	7.2	1.033 (0.119)	1.097 (0.077)	6.4	1.047 (0.124)	1.137 (0.089)	9.0	1.046 (0.124)	1.126 (0.088)	7.9
Higher education	**1.377** (0.204)	**1.558** (0.151)	18.1	**1.366** (0.203)	**1.565** (0.152)	19.9	**1.433** (0.219)	**1.690** (0.179)	25.8	**1.423** (0.217)	**1.687** (0.179)	26.4
Proportion of UE[c] time	1.002 (0.002)	**0.997** (0.001)	-0.5	*1.003* (0.002)	**0.997** (0.001)	-0.6	1.002 (0.002)	**0.997** (0.001)	-0.5	1.002 (0.002)	**0.997** (0.001)	-0.5
Semi urban municipality	1.133 (0.171)	1.114 (0.092)	-1.9	1.135 (0.172)	1.107 (0.093)	-2.8	1.158 (0.175)	1.114 (0.104)	-4.5	1.160 (0.175)	1.098 (0.104)	-6.2
Rural municipality	1.016 (0.110)	0.972 (0.074)	-4.4	1.013 (0.110)	0.981 (0.075)	-3.2	1.031 (0.116)	0.957 (0.080)	-7.4	1.030 (0.116)	0.967 (0.081)	-6.3
Southern Finland	1.345 (0.339)	1.141 (0.099)	-20.4	1.354 (0.341)	1.091 (0.094)	-26.3	1.380 (0.348)	*1.173* (0.112)	-20.7	1.387 (0.349)	1.104 (0.104)	-28.2
Eastern Finland	1.283 (0.348)	1.135 (0.118)	-14.8	1.296 (0.351)	1.095 (0.113)	-20.0	1.307 (0.354)	1.178 (0.134)	-12.9	1.315 (0.356)	1.124 (0.126)	-19.2
Central Finland	1.171 (0.291)	1.186 (0.133)	1.5	1.178 (0.292)	1.147 (0.130)	-3.0	1.193 (0.297)	1.179 (0.148)	-1.4	1.198 (0.298)	1.132 (0.141)	-6.7
Northern Finland	*1.568* (0.414)	1.175 (0.137)	-39.3	*1.574* (0.415)	1.154 (0.134)	-42.0	*1.628* (0.430)	1.212 (0.154)	-41.6	*1.633* (0.430)	1.186 (0.148)	-44.7
Earnings-rel. UE[c] benefit	**1.319** (0.140)	1.035 (0.065)	-28.4	**1.310** (0.139)	1.016 (0.063)	-29.4	**1.331** (0.143)	1.046 (0.073)	-28.5	**1.325** (0.143)	1.018 (0.070)	-30.7
Year 1996	1.087 (0.077)	**1.151** (0.077)	6.4	1.078 (0.076)	**1.250** (0.088)	17.2	1.118 (0.083)	**1.192** (0.093)	7.4	1.110 (0.083)	**1.315** (0.108)	20.5
Year 1997	1.091 (0.091)	*1.158* (0.089)	6.7	1.081 (0.091)	**1.292** (0.105)	21.1	1.141 (0.101)	**1.219** (0.105)	7.8	1.133 (0.101)	**1.399** (0.128)	26.6
Year 1998	**1.323** (0.135)	**1.383** (0.111)	6.0	**1.304** (0.136)	**1.526** (0.129)	22.2	**1.440** (0.152)	**1.558** (0.137)	11.8	**1.423** (0.154)	**1.796** (0.164)	37.3
Year 1999	**1.321** (0.168)	1.004 (0.110)	-31.8	*1.250* (0.161)	1.105 (0.147)	-14.5	**1.508** (0.193)	*1.246* (0.142)	-26.2	**1.434** (0.187)	**1.539** (0.199)	10.5
Begin in January				1.050 (0.079)	**0.693** (0.047)	-35.7				1.036 (0.083)	**0.646** (0.049)	-39.0
December				**0.536** (0.050)	1.025 (0.095)	48.9				**0.545** (0.049)	0.873 (0.077)	32.8
Weibull shape							**0.616** (0.019)	*1.032* (0.018)	41.6	**0.615** (0.018)	**1.050** (0.019)	43.5
-2 log pseudo likelihood	51,838	23,084		51,770	23,055		17,973	6,679		17,910	6,637	

#(register spells): 6050 of which #(events): 3238
#(survey spells): 2717 of which #(events): 1637
Estimates significant at 5% (10%) risk level are displayed in **boldface** (*italics*).
Standard errors adjusted for clustering of spells within persons.

[a] hr hazard ratio. $H_0 : hr = 1$

[b] bias= $100 \times (\text{Survey} - \text{Register})$

[c] UE Unemployment