Tilastokeskus
Statistikcentralen
Statistics Finland

Tutkimuksia
Forskningsrapporter
Research Reports
260

Statistical Analysis of Survey-based Event History Data

with Application to Modeling of Unemployment Duration

Marjo Pyy-Martikainen

Tilastokeskus
Statistikcentralen
Statistics Finland

Tutkimuksia
Forskningsrapporter
Research Reports
260

Statistical Analysis of Survey-based Event History Data

with Application to Modeling of Unemployment Duration

Marjo Pyy-Martikainen

Päätoimittaja – Chefredaktör – Principal editor:
Timo Alanko

The publications in the Research Reports series do not represent the views of Statistics Finland. The writers of the publications are responsible for the contents of the publications.

Taitto – Ombrytning – Layout:
Hilkka Lehtonen

ISSN 1798–8578
= Research Reports
ISBN 978–952–244–456–1 (pdf)
ISBN 978–952–244–455–4 (print)

Acknowledgements

I am most indebted to my primary supervisor Leif Nordberg, Professor (emeritus) of Statistics and Econometrics, Åbo Akademi University, for his support and patience during the many years that the completion of this dissertation took. His experience in and insight into conducting scientific work has guided me along the way. I am also deeply grateful to my second supervisor Ulrich Rendtel, Professor of Statistics, Free University Berlin. I admire his enthusiasm for panel surveys. Collaboration with him was both fruitful and fun.

The reviewers of this thesis, Peter Lynn, Professor of Survey Methodology, University of Essex, and Risto Lehtonen, Professor of Statistics, University of Helsinki, are warmly thanked for their comments and constructive criticism.

I thank Statistics Finland for the possibility to construct and use longitudinal micro-merged survey-register data, for giving me leaves of absence and for publishing the dissertation. Markku Koivula kindly helped me with the construction of data. The help of Ari Toikka and Seppo Suomalainen with SAS programming was essential for the progress of my work. Mia Kokkinen, Pia Maljanen and Jari Tolonen revised the language of the thesis. The layout of the publication was skilfully done by Hilkka Lehtonen. Hannele Sauli, Veli-Matti Törmälehto and Tuula Kuula explained me various features of ECHP and EU-SILC. To crown it all, Hannele delighted my days by lending me excellent literature from her bookshelves.

I thank my colleagues at the Statistical Methods unit for creating a working atmosphere that is both inspiring and relaxed. This gave me the energy necessary to finish the thesis alongside work.

I am grateful to my parents Kaija and Seppo Pyy for their love and unconditional support. I thank my sister Paula Pallasaho for leading by example and showing me that it is possible to combine successfully family, work and PhD studies. To my children Antti and Emilia I owe my warmest thanks. By bringing so much love and joy into my life they have helped me clarify the priorities of life.

My dear friends: how much more mirthless life would be without you! I wish we will have many hilarious moments ahead.

I thank the School of Statistical Information, Inference and Data Analysis and the Finnish Graduate School in Stochastics and Statistics for the financial support. I am also grateful to the Finnish Work Environment Fund for a grant that allowed me to complete this dissertation.

Helsinki, September 2013

Marjo Pyy-Martikainen

Abstract

Longitudinal surveys are increasingly used to collect event history data on person-specific processes such as transitions between labour market states. Survey-based event history data pose a number of challenges for statistical analysis. These challenges include survey errors due to sampling, non-response, attrition and measurement.

This study deals with non-response, attrition and measurement errors in event history data and the bias caused by them in event history analysis. The study also discusses some choices faced by a researcher using longitudinal survey data for event history analysis and demonstrates their effects. These choices include, whether a design-based or a model-based approach is taken, which subset of data to use and, if a design-based approach is taken, which weights to use.

The study takes advantage of the possibility to use combined longitudinal survey register data. The Finnish subset of European Community Household Panel (FI ECHP) survey for waves 1–5 were linked at person-level with longitudinal register data. Unemployment spells were used as study variables of interest.

Lastly, a simulation study was conducted in order to assess the statistical properties of the Inverse Probability of Censoring Weighting (IPCW) method in a survey data context.

The study shows how combined longitudinal survey register data can be used to analyse and compare the non-response and attrition processes, test the missingness mechanism type and estimate the size of bias due to non-response and attrition. In our empirical analysis, initial non-response turned out to be a more important source of bias than attrition. Reported unemployment spells were subject to seam effects, omissions, and, to a lesser extent, overreporting. The use of proxy interviews tended to cause spell omissions. An often-ignored phenomenon, classification error in reported spell outcomes, was also found in the data. Neither the Missing At Random (MAR) assumption about non-response and attrition mechanisms, nor the classical assumptions about measurement errors, turned out to be valid. Both measurement errors in spell durations and spell outcomes were found to cause bias in estimates from event history models. Low measurement accuracy affected the estimates of baseline hazard most. The design-based estimates based on data from respondents to all waves of interest and weighted by the last wave weights displayed the largest bias. Using all the available data, including the spells by attriters until the time of attrition, helped to reduce attrition bias. Lastly, the simulation study showed that the IPCW correction to design weights reduces bias due to dependent censoring in design-based Kaplan-Meier and Cox proportional hazard model estimators.

The study discusses implications of the results for survey organisations collecting event history data, researchers using surveys for event history analysis, and researchers who develop methods to correct for non-sampling biases in event history data.

Key words: longitudinal surveys, survey errors, event history analysis

Tiivistelmä

Pitkittäisillä surveytutkimuksilla kerätään yhä useammin yksilöitä koskevia tapahtumahistoriatietoja, kuten esimerkiksi tietoja siirtymistä eri työmarkkinatilojen välillä. Tällaisten tietojen tilastollisessa analyysissa tulee ottaa huomioon surveytutkimuksen virhelähteet, joita ovat muun muassa otannasta, kadosta ja attritiosta sekä mittaamisesta johtuvat virheet.

Tässä tutkimuksessa tarkastellaan kadosta, attritiosta ja mittaamisesta johtuvia virheitä surveyaineistoon perustuvissa tapahtumahistoriatiedoissa sekä niiden aiheuttamaa harhaa tapahtumahistoria-analyysissa. Tutkimus käsittelee surveyaineistoon perustuvien tapahtumahistoriatietojen tilastollisessa analyysissa vastaantulevia valintoja ja niiden vaikutusta tuloksiin. Tällaisia valintoja ovat: valitaanko asetelma- vai malliperusteinen lähestymistapa; mitä osaa aineistosta hyödynnetään sekä valittaessa asetelmaperusteinen lähestymistapa, mitä painoja käytetään.

Tutkimuksessa hyödynnetään yhdistettyä pitkittäistä survey-rekisteriaineistoa. Eurooppalaisen elinolotutkimuksen (ECHP, European Community Household Panel) Suomea koskeva, tutkimuskerrat 1–5 kattava aineisto yhdistettiin henkilötasolla rekisteripaneeliaineistoon. Tutkimusmuuttujina olivat työttömyysjaksojen kestot.

Tutkimuksessa tarkastellaan lisäksi simulointimenetelmin käänteisen sensurointitodennäköisyyden painotusmenetelmän (IPCW, Inverse Probability of Censoring Weighting method) tilastollisia ominaisuuksia surveyaineistoon perustuvassa elinaika-analyysissa.

Tutkimuksessa näytettiin, kuinka yhdistettyä pitkittäistä survey-rekisteriaineistoa voidaan hyödyntää kadon ja attrition analysoinnissa, puuttuneisuuden mekanismin testaamisessa sekä kadosta ja attritiosta johtuvan harhan estimoinnissa. Ensimmäisen tutkimuskerran kato osoittautui tutkimuksen empiirisissä analyyseissa attritiota merkittävämmäksi harhan lähteeksi. Surveyvastauksiin perustuvissa työttömyysjaksoissa esiintyi jaksojen alkujen ja loppujen kasautumista viiteajankohtien ääripäihin, jaksojen raportoimatta jättämistä ja jossain määrin myös yliraportointia. Raportoimatta jättämisen todennäköisyys oli yhteydessä sijaisvastaajan käyttöön. Työttömyysjaksojen päättymissyissä esiintyi luokitteluvirheitä. Empiiristen analyysien perusteella klassiset oletukset mittausvirheistä tai oletukset puuttuneisuuden satunnaisuudesta (MAR, Missing At Random) eivät pitäneet paikkaansa. Sekä työttömyysjaksojen kestoon että päättymissyihin liittyvät mittausvirheet aiheuttivat harhaa tapahtumahistoria-analyysin tuloksiin. Työttömyysjaksojen alhainen mittaustarkkuus aiheutti eniten harhaa perushasardifunktion estimointiin. Empiiristen analyysien perusteella harhaisimpia olivat kaikkiin tutkimusaaltoihin vastanneiden henkilöiden osa-aineistoon perustuvat, viimeisen tutkimusaallon painoilla painotetut asetelmaperusteiset estimaatit. Aineiston laajentaminen kattamaan kaikki vähintään ensimmäiseen tutkimusaaltoon vastanneet henkilöt pienensi harhaa. Simulointitutkimuksen tulosten perusteella asetelmapainojen IPCW-korjaus pienentää asetelmaperusteisten Kap-

lan-Meier- ja Coxin verrannollisten hasardien mallin kovariaattivaikutusten estimaattorien informatiivisesta sensuroinnista aiheutuvaa harhaa.

Tutkimuksen tulosten merkitystä arvioidaan tapahtumahistoriatietoa surveytutkimuksilla keräävien organisaatioiden ja tietoja käyttävien tutkijoiden sekä menetelmäkehittäjien näkökulmasta.

Avainsanat: Pitkittäiset surveytutkimukset, surveytutkimuksen virhelähteet, tapahtumahistoria-analyysi

Sammanfattning

Longitudinella surveyundersökningar används mer och mer för att samla in händelsehistorik för individer, såsom t.ex. uppgifter om förändringar i arbetsmarknadsstatus. Vid statistisk analys av sådana uppgifter bör man beakta felkällorna i surveyundersökningar, dvs. fel som beror på användningen av urval, förekomsten av bortfall och attrition samt dålig mätprecision.

I denna undersökning granskas förekomsten av fel i händelsehistorikdata som beror på bortfall, attrition och mätfel samt den bias som orsakas av dessa i analysen av datat. I undersökningen analyseras vilka val man ställs inför vid statistisk analys av survey-baserat händelsehistorikdata och vilken inverkan olika lösningar har på resultaten. Bland de frågor som forskaren ställs inför kan nämnas huruvida man ska välja ett designbaserat eller ett modellbaserat betraktelsesätt, vilken del av materialet som skall användas samt, vid val av ett designbaserat betraktelsesätt, vilka vikter som skall användas.

I undersökningen utnyttjas ett kombinerat longitudinellt survey-registermaterial. Det finländska materialet i den europeiska undersökningen om levnadsförhållanden (ECHP, European Community Household Panel) i omgångarna 1–5 kombinerades på individnivå med longitudinellt registermaterial. Undersökningsvariabler var längden på arbetslöshetsperioderna.

I undersökningen granskas även effekterna av att utnyttja vikter som bygger på censureringssannolikheterna i olika faser av datainsamlingsperioden (IPCW, Inverse Probability of Censoring Weighting).

I undersökningen visas hur ett kombinerat longitudinellt survey-registermaterial kan utnyttjas vid analys av bortfall och attrition, testning av olika antaganden om typen av bortfall samt estimering av bias på grund av bortfall och attrition. Empirisk analys visade att bortfallet vid första undersökningsomgången var en mer betydande källa till bias än attritionen. Respondenternas svar angående början och slutet på arbetslöshetsperioder tenderade att i viss mån koncentera sig till början och slutet av referensperioderna. Vissa perioder rapporterades inte alls, men å andra sidan noterades även överrapportering i någon mån. Sannolikheten för att en period blev orapporterad var större när man intervjuade en annan person istället för intervjupersonen. Klassificeringsfel förekom ifråga om orsakerna till avslutade arbetslöshetsperioder. Varken de klassiska antagandena om mätningsfelens egenskaper eller bortfallets slumpmässighet (MAR, Missing At Random) visade sig vara valida. Mätningsfelen i såväl längden av arbetslöshetsperioderna som orsakerna till att de tog slut gav upphov till bias i analysresultaten. Den låga precisionen i mätningen arbetslöshetsperiodernas längd orsakade särskilt mycket bias vid estimeringen av baslinjehasarden. Designbaserade estimat baserade på det delmaterial som omfattade endast de personer som besvarat alla undersökningsomgångar och viktade med den sista undersökningsomgångens vikter uppvisade mest bias. En utvidgning av materialet till att omfatta alla personer som besvarat minst den första undersökningsomgången minskade biasen. På basis av resultaten från simuleringsundersökningen minskar en IPCW-korrigering av designvikterna den bias som orsakas av informativ censurering vid

design-baserad estimering av parametrarna i Kaplan-Meiers modell och i Cox proportionella hasardmodell.

I undersökningen redogörs även för hur organisationer som samlar in survey-baserat händelsehistorikdata, forskare som utnyttjar uppgifterna och metodutvecklare kan dra nytta av resultaten i denna studie.

Nyckelord: longitudinella surveyundersökningar, felkällorna i surveyundersökningar, analys av händelseförlopp

Contents

List of original publications

1 Introduction

Longitudinal surveys are increasingly used to collect event history data on person-specific processes such as transitions between labour market states. Event history data collected by longitudinal surveys pose a number of challenges for statistical analysis. These challenges include, survey errors due to sampling, non-response, attrition and measurement.

Survey errors are problematic because they diminish the accuracy of estimates. The concept of total survey error [5] provides a theoretical framework for survey errors. For empirical analysis of errors in survey data, combined survey register data are considered a valuable tool [6, 7]. However, combining survey data with register data may be time-consuming and costly. Also, legal and ethical problems may be involved. Therefore, only few studies use this method to assess errors in event history data.

This study takes advantage of the possibility to use combined longitudinal survey register data. Finnish subset of European Community Household Panel (FI ECHP) data for waves 1–5 were linked at person-level with longitudinal register data. Unemployment spells were used as study variables of interest.

The study deals with non-response and measurement errors and the bias caused by them. The study shows how longitudinal combined survey register data can be used to conduct an analysis of non-response and attrition in longitudinal survey data. The study makes a contribution to the pool of evidence on the existence, determinants and effects of non-response, attrition and measurement errors in event history data based on longitudinal surveys. The study also assesses statistical properties of the Inverse Probability of Censoring Weighting (IPCW) method in design-based survival analysis in the presence of dependent censoring.

A researcher using longitudinal survey data for event history analysis has to make several choices that affect the results of the analysis. These choices include the following: whether a design-based or a model-based approach is taken, which subset of data to use and, if a design-based approach is taken, which weights to use. These choices are discussed in [8, 9, 10]. However, the effect of these choices in event history analysis have not been assessed yet with combined survey register data. This study makes a contribution by providing empirical evidence on the effect of these choices.

2 Background

2.1 Examples of longitudinal surveys

The first longitudinal social surveys were launched during the late 1960's and early 1980's in the UK, USA and Germany. Canada launched a number of panel surveys in the early 1990's. The first EU level household panel was launched in 1994. During the 2000's, new panel surveys have been launched in Australia (2001), New Zealand (2002) and South Africa (2008) [11].

The Panel Study of Income Dynamics (PSID, www.psidonline.isr.umich.edu) run by the University of Michigan is the pioneer of household panel surveys. Launched in 1968 and still running, it is the longest running household panel survey in the world. Its original focus was on income and poverty dynamics but its study topics have been extended to cover areas such as labour force and residential dynamics.

Survey of Income and Program Participation (SIPP, www.census.gov/sipp) run by the US Census Bureau since 1984 is another long-running panel survey in the USA. It has a rotating design with panels ranging from 2.5 to 4 years. It was mainly designed to measure the effectiveness and future costs of government transfer programs such as the food stamps program.

German Socio-Economic Panel (SOEP, www.diw.de/en/soep) launched in 1984 and run by the German Institute for Economic research is the European pioneer of household panel surveys. A special feature of SOEP is that it follows all persons ever interviewed, regardless of their relationship to the original sample persons [12].

The British Household Panel Survey (BHPS, www.iser.essex.ac.uk/bhps) was run by the Institute for Social and Economic Research during 1991–2008, with 18 yearly data collection waves. The BHPS sample was incorporated in 2010 in the second round of a new household panel survey, *Understanding Society* (http://www.understandingsociety.org.uk).

In the early 1990's, Statistics Canada launched several longitudinal surveys, including the *Survey of Labour and Income Dynamics* (SLID, see www.statcan.gc.ca/imdb-bmdi/3889-eng.htm). SLID is a rotating panel survey with new six-year panels beginning every three years [8]. One of the main aims of SLID is to support analyses of income mobility and labour market dynamics.

An ambitious multicountry panel survey, *the European Community Household Panel* (ECHP, epp.eurostat.ec.europa.eu/portal/page/portal/microdata/echp), was launched in 1994. The key features of ECHP are its comparability across countries, achieved by input-harmonisation, as well as the wide range of topics covered. The ECHP was carried out by national data collection units, mostly national statistical institutes, with the Statistical Office of the European Communities (Eurostat) providing centralised support and coordination [13]. Finland started compiling ECHP data in 1996, a year after becoming a member of the EU. The ECHP was designed for the analysis of individual change over time and in this respect, it can be claimed to be the first real Finnish longitudinal social survey. The panel was run until 2001, resulting in 6 annual waves. The Finnish ECHP survey is described in [14].

The European Union Statistics on Income and Living Conditions (EU-SILC, epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_ silc) was launched in 2003. Like ECHP, it is a multicountry household panel survey designed and coordinated by Eurostat. The design is however output harmonised, giving national data collection units more freedom with respect to implementing the survey. Most EU-SILC countries implement a rotating panel design with a new four-year panel beginning each year, reflecting the fact that cross-sectional estimates are considered as of primary importance. EU-SILC has been compiled in Finland since 2003. Finland has recently adopted the recommended 4-year rotating panel design.

The Millennium Cohort Study (MCS, www.cls.ioe.ac.uk) run by the Centre for Longitudinal Studies is the most recent of UK's four ongoing national longitudinal birth cohort studies. The study has been tracking the Millennium children born in UK through their early childhood years and plans to follow them into adulthood.

2.2 Collection of event history data by longitudinal surveys

Many longitudinal surveys collect event history data related to person-specific processes such as fertility, income and labour market dynamics. Event history data consists of information about durations of spells in a state of interest (such as poverty, unemployment, having no children), the outcome 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 [15]. In the multi-state framework the time 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.

PSID uses an event occurrence framework to collect information on residence and labour force status histories. The timing of transitions between different states are recorded at the accuracy of one third of a month. These data are converted into month level information in the public release dataset. [16]. SLID uses the event occurrence framework for information on job and jobless spells during the year preceding the interview. 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. EU-SILC uses a multi-state framework very similar to that used in ECHP to collect month-level labour market state information for the year preceding the interview.

2.3 Errors in longitudinal surveys

A major objective in the design of any survey is to maximise the accuracy of key estimates, given cost and time constraints [17]. The concept of total survey error [5] provides a theoretical framework for assessing accuracy of survey estimates. The following discussion on total survey error bases on [17]. Total survey error refers to the accumulation of all errors that may arise in the design, collection, processing and analysis of data. Survey errors are problematic because they diminish the accuracy of estimates. The objective of maximising accuracy is equivalent to minimising total survey error.

Total survey errors can be decomposed into sampling errors and non-sampling errors. Sampling errors arise because a survey measures only a subset of the population of interest. Even if the total population was measured, the estimates would contain errors due to survey non-response and deficiencies in the specification of survey questions, frame, measurement or data processing. These errors are called non-sampling errors. Non-sampling errors can be viewed as mistakes or unintentional errors that can be made at any stage of the survey process whereas sampling errors are intentional in the sense that their magnitude can be controlled [18]. Each of the error sources may contribute a variable error, a systematic error, or both. Variable errors are reflected in the variance and systematic errors in the bias of an estimate.

Total survey error is usually measured in terms of mean squared error (MSE). Each estimate has a corresponding MSE reflecting the effects of all error sources [18]. The mean squared error of an estimate $\hat{\beta}$ is defined as the expected squared difference between the estimate and the value of the target parameter β, the expectation being taken over all possible realisations of the survey process:

$$\mathrm{MSE}(\hat{\beta}) = \mathrm{E}(\hat{\beta} - \beta)^2 .$$

Mean squared error can be decomposed into squared bias and variance:

$$\mathrm{MSE}(\hat{\beta}) = \left[\mathrm{E}(\hat{\beta}) - \beta\right]^2 + \mathrm{E}\left[(\hat{\beta} - \mathrm{E}(\hat{\beta}))^2\right] = \left[Bias(\hat{\beta})\right]^2 + Var(\hat{\beta}).$$

Both the bias and the variance components can be further decomposed according to the error source. Biemer and Lyberg [18] use the following decomposition reflecting the most important sources of bias and variance:

$$\mathrm{MSE}(\hat{\beta}) = \left[Bias_{spec}(\hat{\beta}) + Bias_{nr}(\hat{\beta}) + Bias_{fr}(\hat{\beta}) + Bias_{meas}(\hat{\beta}) + Bias_{dp}(\hat{\beta})\right]^2$$
$$+ Var_{samp}(\hat{\beta}) + Var_{meas}(\hat{\beta}) + Var_{dp}(\hat{\beta}),$$

the subscripts *spec, nr, fr, meas, dp* and *samp* referring to errors due to specification, non-response, frame, measurement, data processing and sampling.

Estimation of MSE is a complex and costly process. Therefore, usually only a few of the most important components are estimated [18]. This study deals with non-response and measurements errors and the bias caused by them.

2.3.1 Non-response errors

A non-response error is caused by unsuccessful attempts to obtain the desired information from eligible units. The failure to obtain any information at all from an eligible unit results in unit non-response whereas item non-response refers to a situation where a responding unit fails to answer some questions. Our focus is on unit non-response. Hereafter, unit non-response is called simply non-response.

In longitudinal surveys, non-response may occur in three different patterns. *Total non-respondents* provide data for none of the survey waves. *Attrition non-respondents* drop permanently out of the survey at some wave after the first, while *temporary non-respondents* return to the survey after missing one or more waves [19].

Non-response errors in event history data are manifested in three different ways: due to total non-response, attrition and temporary non-response, spells may *not* be *observed* at all. Attrition and temporary non-response may cause *right-censoring* of spells. In this case the follow-up ends before the end of the spell, leaving the ending date of the spell and its outcome unknown. Temporary non-response may also cause *left-truncation* of spells. A spell is left-truncated if it has begun before the start of the follow-up period. Longitudinal surveys usually follow individuals over a fixed follow-up time with pre-specified start and end dates. Right-censoring and left-truncation may also occur because of the fixed follow-up time, a reason not related to non-response.

Figure 1 demonstrates the different non-response errors in event history data created by the different non-response patterns. The follow-up time is the time period [0,3]. Interviews are conducted at time points t=1, t=2 and t=3. At the time t interview, information about spells and covariates are collected for the time period (t–1,t]. Person **a** responds in all three interviews. He has three spells, the first being left-truncated by the start of the follow-up period and the third being right-censored by the end of follow-up period. His second spell is completely observed. The spells by person **b** are not observed due to total non-response. Person **c** attrits at wave 2 and therefore, only his first spell, the spell ongoing during (0,1] is observed. The spell is right-censored due to attrition at time t=1. His second spell is

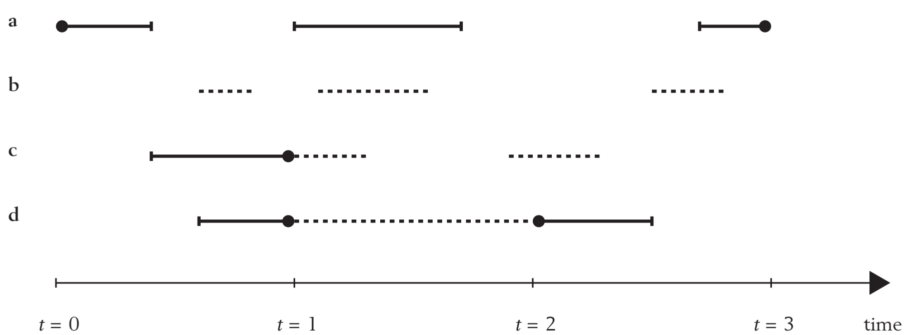

Figure 1 Non-response errors in event history data.
Solid line: part of spell observed in survey. Dashed line: Part of spell not observed by non-response. Observed starting and ending dates of spells are marked by ticks. Left-truncation and right-censoring are marked by circles.

not observed due to attrition. Person **d** misses the wave 2 interview and is, therefore, a temporary non-respondent. The spell by person **d** is observed as two spells, the first spell being right-censored and the second spell being left-truncated.

It is usually assumed that data are Missing At Random (MAR) [20]. The MAR assumption in the context of event history data is discussed in [1]. In this context, the MAR assumption means that non-response is independent of current and future events, given past events and covariates. The assumption of an independent right-censoring mechanism discussed in [3] is equivalent to the MAR assumption. If the MAR assumption does not hold, i.e. if data are Missing Not At Random (MNAR), one has to model the missing data mechanism in order to get unbiased estimates.

An example of a MNAR mechanism is a situation where persons with long unemployment spells drop out from the survey more frequently than otherwise similar persons with shorter spells. In this case, falsely assuming a MAR missingness mechanism leads to biased estimates of the distribution of unemployment duration. If, in addition, the covariate effects differ among persons with long and short spells, there will be bias in the estimated covariate effects, too. The validity of the MAR assumption is usually impossible to test because the values of study variables are unobserved from the time of non-response. [1].

Either weighting or imputation may be used to correct for non-response. Both of these methods rely on the MAR assumption. As it is very difficult to impute all items in a missing wave without distorting associations between survey variables, weighting is usually the preferred method to correct for unit non-response [19]. Survey data sets are usually equipped with weights aiming to correct for non-response. These weights are to be used with all variables included in the survey data set. It is not always clear, however, how to use weights in event history analysis, see [9, 4].

In longitudinal surveys collecting event history data, specific weights may need to be developed to account for dependent censoring that violates the independency asssumption, see e.g. [21]. Dependent censoring means that the probability of censoring is related to the length of the spells of interest. Dependent censoring may cause a bias in estimates from event history analysis. Robins [22] proposed an Inverse Probability of Censoring Weighting (IPCW) method to adjust for bias in survival analysis due to dependent censoring. Lawless [23] discussed the use of the IPCW method in a complex survey data context.

Sample selection models aim to correct for non-response that is MNAR. These models are mainly used in the analysis phase and not in the production phase of survey data. The studies by van den Berg, Lindeboom and Ridder [24] and van den Berg and Lindeboom [25] are early examples of sample selection modeling of labour market transition data.

2.3.2 Measurement errors

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 history data are manifested as a failure to report a spell (omission), reporting a spell that did not occur (overreporting) and misreporting the duration of a spell (misdating) [26, 27]. In event history data, 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. Even though spell outcomes may also be misreported, this topic has received little attention in the literature. [2].

Because of measurement errors, the true spell 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 ϵ: $T = T^* + \epsilon$.[1] Referring to Figure 1, Figure 2 demonstrates measurement errors in reported spell durations. Thick lines show the true durations T^* and thin lines the measurement error ϵ in the spell duration. At the first interview at time t=1, person **a** reported a spell that did not occur (overreporting). At the second interview at time t=2, he misdated the start of the spell. Person **b** is a total nonrespondent and, therefore, reports no spells. Person **c** correctly reports his first spell. The first spell is right-censored and the second spell not reported due to attrition at wave 2. Person **d** omitted his first spell. The true duration T^* and measurement error ϵ cancel each other out so that T=0.

According to the classical assumptions [28, 7], measurement errors ϵ have zero mean and are independent of each other, true durations T^* and any covariates explaining T^*. Under a linear regression model, classical measurement errors in the dependent variable do not cause bias in the estimates of regression coefficients [7]. If the model specified is nonlinear or measurement errors are not classical, bias may result. The validity of the classical assumptions is usually impossible to test because the true durations T^* are not observed.

Skinner and Humphreys [29] and Augustin [30] proposed methods to correct 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 satisfy the classical assumptions.

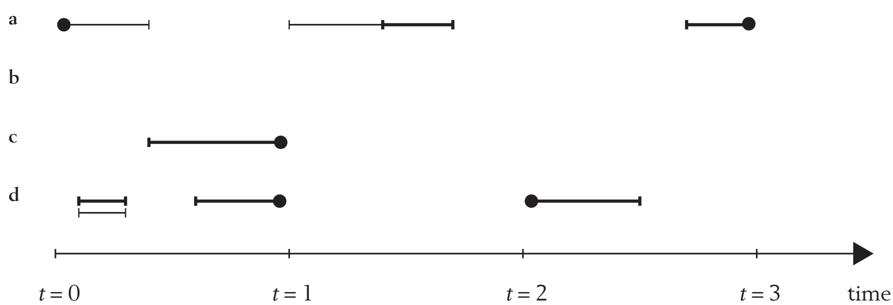

Figure 2: Measurement errors in event history data.
Thick line: true duration. Thin line: measurement error.

2.3.3 Estimating non-response and measurement error biases

In practice, only one realisation of the survey process is observed. Therefore, $Bias(\hat{\beta})$ is unknown. Bias may, however, be estimated using the deviation of the value of $\hat{\beta}$ obtained from the survey and the true value β. The true value is un-

1 Subscripts indicating individual and spell ignored.

known but sometimes additional, gold standard data are collected, which for evaluation purposes are considered to be the truth [18].

A reinterview study revisits respondents from the original survey sample and asks some of the questions that were asked in the original survey. Reinterview questions are designed to reference the same time period as in the original interview. The goal is to obtain higly accurate responses that can be used to estimate the true value of the parameter. Then, bias due to measurement error can be estimated as a difference between estimates from the original survey and the reinterview survey. This approach is, however, not without problems. The longer the recall period, the more erroneous the responses tend to be [7]. Also, it is likely that not all respondents from the original survey respond to the reinterview survey. The estimate from the reinterview survey may thus be plagued by non-response bias. As a consequence, reinterview data may be as erroneous as the data that is being evaluated [17].

An external validation study compares survey estimates with external estimates that are considered to be more accurate. External estimates may be obtained from administrative records or from a survey that is considered to be a gold standard for the estimate being evaluated [18]. External validation studies may suffer from differences in target populations or definitions of variables of interest in the two data sources. Moreover, external validation studies do not allow the decomposition of bias due to different sources.

Record check studies link administrative register data to survey data at individual-level. Record check studies may be classified into *prospective record check studies*, *reverse record check studies* and *complete record check studies*. Prospective record check studies link administrative records to survey respondents in order to confirm the reported behaviors. Reverse record check studies sample units from administrative records with desired characteristics and then attemp to interview them. Prospective record check studies may be used for measuring overreporting of events while reverse record check studies may be used for measuring underreporting of events. Complete record check studies with validation data for all sampled persons allow both the estimation of overreporting and underreporting. Moreover, they allow the estimation of bias due to non-response. As in external validation studies, the comparisons may be hampered by differences in the definitions of variables from the two data sources.

Even though no gold standard data are error-free, they can be very useful if the errors are small relative to errors in data being evaluated. As Biemer and Lyberg [18] point out, gold standard data provides a silver rather than a gold standard.

2.4 Analysing event history data based on a complex longitudinal survey

Surveys often use complex sampling designs involving stratification, clustering and unequal selection probabilities of units. Longitudinal surveys have an additional stage of clustering arising from the repeated observations by the same sample units. Non-response, attrition and measurement errors bring additional challenges to the analysis of longitudinal survey data. How should these complexities be taken into account in event history analysis?

Pfefferman and Sverchkov [31] and Pfefferman[32] discuss different approaches to the modelling of survey data. Kovačević and Roberts [10] discuss model-based and design-based approaches to the modelling of event history data. In the model-based approach, the target parameters of interest are parameters β of a superpopulation model that is assumed to have generated the variable values in the finite population. The standard model-based approach ignores the probability distribution $P(S)$ induced by the sampling design. The only source of random variation in the superpopulation model parameter estimator $\hat{\beta}$ is due to the random component in the model. Accordingly, the model-based standard errors of parameter estimates reflect the uncertainty due to the model. Sample design variables or sample weights might be incorporated as covariates of the model in order to protect against nonignorable sample design.

The design-based approach is traditionally used for descriptive inference. However, the ideas of design-based inference can be applied to analytic inference as well. In this approach the target parameter of interest is defined as a finite population parameter B that would be obtained from the model estimation procedure if all data values in the finite population were available. In the design-based approach, the only source of random variation in the estimation procedure is the sampling distribution of the estimator $\hat{B}$. Inference about B could in principle be carried out with certainty if all elements of the population were measured [33]. In practice, there would be uncertainty in the estimates even in this case due to non-sampling errors. An analyst taking the design-based approach would conduct a weighted analysis. The design information would be used to calculate the standard errors of parameter estimates. These standard errors reflect the uncertainty due to making inferences on the basis of a sample only instead of the whole population.

The test for ignorability of sample design suggested by Pfeffermann [34] may be used to choose between the design-based and the model-based approaches for event history analysis. The test compares design-based and model-based estimates of parameters of interest and rejects the null hypothesis of ignorability of sample design if the model-based estimates are "too far" from the design-based estimates. In this case, a design-based approach for the analysis should be taken.

Longitudinal analyses often use only respondents to each wave of interest [35, 19]. Even though the available data until the time of attrition could be used, attriters are often discarded from the analysis. In an analysis using weights this can be motivated by the fact that weights are usually adjusted for non-response and attrition. However, the general purpose weights included in a survey data set may not fully correct for non-response and attrition that is selective with respect to the particular response variable of interest. The inclusion of the available data from the attriters might in this case help to reduce the bias due to attrition. It is not clear, however, which weights should be used in an analysis including attriters, see [9].

3 Aims of the study

Article [1].
To show how register data combined at person-level with survey data can be used to conduct a non-response and attrition analysis that enables to 1) study the determinants of non-response and attrition; 2) test the validity of the MAR assumption; and 3) estimate the size of bias due to non-response and attrition. To apply this analysis to unemployment spell data from FI ECHP survey in order to provide novel information on relative importance and determinants of non-response and attrition in event history data and their effects on event history analysis.

Article [2].
To conduct a complete record check validation study of retrospective reports of unemployment spells from FI ECHP survey data in order to provide novel evidence about 1) the type, magnitude and determinants of measurement errors in survey reports of event histories, 2) the validity of classical assumptions about measurement errors, 3) the size of bias due to measurement errors and low measurement accuracy in event history analysis of survey data.

Article [3].
To study statistical properties of the Inverse Probability of Censoring Weighting (IPCW) method in design-based survival analysis based on complex survey data.

Article [4].
To discuss the following choices involved in event history analysis of survey data: 1) whether to take a design-based or a model-based approach for modelling; 2) which subset of data to use; and 3) if a design-based approach is chosen, which weights to use. To demonstrate the effect of these choices by using unemployment spell data from FI ECHP survey combined at person-level with register data.

4 Data and Methods

4.1 Combined longitudinal survey register data

The Finnish subset of the European Community Household Panel (FI ECHP) survey data were combined at person-level with longitudinal register data. Empirical analyses were based on data from FI ECHP sample persons aged 16 or over at the beginning of the panel. Sample persons are defined in the ECHP as all members of the initial sample of households. The first five waves of the FI ECHP data covering the years 1996–2000 were used in the analyses. Temporary non-respondents were excluded, leaving 10,720 persons for the analysis. Unemployment spells were used as study variables of interest.

FI ECHP target population and sample design. The target population of FI ECHP consists of members of private households permanently resident in Finland. As most household panel surveys, FI ECHP aims to remain cross-sectionally representative of the household population over time. This is strived for using certain follow-up rules of the sample persons, see [14]. The FI ECHP sample is a two-phase stratified network sample. The population information system of the Population Register Centre was used as a frame. The frame population consisted of persons permanently living in Finland aged 15 and over. In the first phase, a master sample of target persons was drawn from the frame. Dwelling units were constructed by adding all the persons sharing the same domicile code as the target persons to the master sample. The master sample was merged with the most recent taxation records and their information was used to form a socio-economic group for each target person. The second phase consisted of drawing the final sample from the master sample using stratification according to socio-economic groups.

Collection of event history data on labour market states in FI 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. 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.

Construction of combined survey-register panel data. FI ECHP survey data were merged with administrative data on unemployment spells retrieved from the Ministry of Labour's register of jobseekers. The register contains day-level information on the starts and ends of unemployment spells, as well as on spell outcomes. All register spells ongoing between 1 January 1995 and 31 December 1999 were used in the analysis. This time period corresponds with the main activity state reference periods of the first five waves of the FI ECHP. The register of jobseekers and other administrative registers such as Statistics Finland's register of completed education and degrees, the population information system of the population register centre and registers of the tax administration were also used to retrieve the background variables used in the analyses. Personal identification numbers were used in order to merge the data files at person-level across time, across various administrative registers and across survey and register data.

4.1.1 Article [1]

Register data were used as a source of information on unemployment spells and covariates. This information is available for all sample persons irrespective of the response status. Survey data were used only to obtain, for each wave, the sample person's participation status in the FI ECHP. This way we obtained directly comparable information for respondents and non-respondents and were able to detect a pure non-response effect, free from measurement errors. The statistical analyses were conducted in a model-based framework.

Assessing determinants of non-response and attrition. Separate models were estimated for the non-response and attrition processes. The initial non-response analysis was conducted by estimating logit models for the probability of being a non-respondent at the first wave of the panel. The analysis was restricted to sample persons having at least one spell of unemployment during the follow-up period (2,956 persons). The attrition process was modeled by a discrete-time hazard model where the conditional probability of attrition at a specific year, given that the person has remained in the survey until the year in question, is explained by a set of time-varying covariates. Initial non-respondents were excluded, leaving 2,085 persons for the attrition analysis.

Testing the validity of the MAR assumption. Covariates describing number of days spent in unemployment and number of unemployment spells were used to test the MAR assumption. For the initial non-response analysis, the number of unemployment days and the number of spells were calculated both before and after the time of the interview (or time of contact, if an interview was not obtained) in wave 1. In the attrition analysis, the number of unemployment days and the number of spells were calculated for each wave before and after the last obtained interview. If the initial non-response mechanism is MAR, none of these covariates should explain probability of non-response. In the attrition model, a MAR non-response mechanism implies that covariates measured after the last obtained interview should not affect probability of non-response. The validity of the MAR assumption was tested by looking at the statistical significance of these covariates.

Estimating non-response bias. The participation behaviour in the survey is known for each sample person having one or more spells during the observation period. It was assumed that unemployment spells are observed until the time of the last interview or until the end of the observation period, whichever comes first. This creates a number of different cases:

a Spells that end before the last interview (or before 31 December 1999, whichever comes first) are *fully observed*.
b Spells ongoing at the time of the last interview, which is followed by attrition, are *right censored by attrition* at the time of the last interview.
c Spells that start after the last interview, which is followed by attrition, are *not observed by attrition*.
d Spells by persons without any interviews are *not observed by initial non-response*.

On the basis of this taxonomy, three different sets of unemployment spells were constructed:

The *full information* set of spells uses the entire register information without restrictions by initial non-response or attrition. Cases **a, b, c, d** (10,734 spells).

The *partial information* set of spells is a subset of the full information set of spells, obtained by excluding spells unobserved by initial non-response. Cases **a, b, c** (7,712 spells).

The *observed information* set of spells is a subset of the partial information set of spells, obtained by excluding spells unobserved by attrition and the remaining length of the spells censored by attrition. Cases **a, b** (6,496 spells).

The size of bias due to non-response and attrition was estimated by comparing Kaplan-Meier estimates of survival function and estimates of regression coefficients from a Cox shared frailty model based on the three sets of unemployment spells. The analyses were conducted in a cause-specific setting, the outcome of interest defined as transition from unemployment to employment. The bias due to non-response was estimated as

$$\hat{Bias}_{nr} = \hat{\beta}_{partial} - \hat{\beta}_{full},$$

the bias due to attrition was estimated as

$$\hat{Bias}_{attr} = \hat{\beta}_{obs} - \hat{\beta}_{partial},$$

and the joint effect due to non-response and attrition as

$$\hat{Bias}_{nr+attr} = \hat{\beta}_{obs} - \hat{\beta}_{full}.$$

The Hausman test [36] was used to test the statistical significance of bias.

4.1.2 Article [4]

The full information and observed information sets of spells described in the previous section were used together with a *total respondents* set of spells. The total respondents set of spells uses data from respondents who provide data on all waves of interest (4,066 spells).

Design-based estimates based on the full information data set were used as benchmark estimates $\hat{B}_{bm}$ against which estimates based on the observed information and the total respondents sets of spells were evaluated. These benchmark estimates are free from the effects of non-response and attrition. The benchmark estimates were taken to be the best available estimates of $\boldsymbol{B}$, the finite population regression parameters, which, if the model postulated is correct, in turn estimate the model parameters β. Model-based and design-based estimates of Cox proportional hazard models for the total respondents and observed information data sets were calculated. The outcome of interest was defined as transition from unemployment to employment.

The design-based total respondents analyses were weighted by the last wave base weights (described in [35]). The design-based observed information estimates were calculated using both first wave base weights and base weights from the starting year of the spell.

The test proposed by Pfeffermann [34] was used to test ignorability of sample design. A Mahalanobis type of distance measure was used to assess the closeness of estimated coefficients to the benchmark estimates.

4.1.3 Article [2]

A complete record check validation study of retrospective reports of unemployment spells from the FI ECHP survey data was conducted. The survey data consists of all unemployment spells reported by FI ECHP sample persons (2,710 spells). For each person, the validation data cover the same time span as his/her follow-up time in the survey. The validation data contains 6,050 register spells. The statistical analyses were conducted in a model-based framework.

Assessing determinants of measurement errors. To study determinants of measurement errors and test validity of the classical assumptions, measurement error variables were constructed for each person. The survey and register data can be reliably linked only at person-level and not at spell-level. Therefore, the measurement error variable was calculated as the difference between the sums of spell durations from the survey and the register. Measurement error variables were calculated separately for each person and for each panel wave in which the person was unemployed according to both survey and register. Measurement errors were modeled in two phases. In the first phase, a random effects logit model was specified for the probability of reporting no unemployment spells in a specific wave, given that at least one unemployment spell was found in the register. In the second phase, a random effects linear model was specified for the magnitude of measurement errors in the reported unemployment spells, given that at least one unemployment spell was both reported and found in the register.

Testing the validity of the classical assumptions. Covariates related to length and number of unemployment spells and covariates of the model explaining unemployment duration were used to test the classical assumptions about measurement errors. Statistically significant effects of these covariates were taken as evidence of violation of the classical assumptions.

Estimating bias due to measurement errors and low measurement accuracy. Kaplan-Meier estimates of survival function and estimates from Cox and Weibull proportional hazards models based on survey data were compared with estimates from the validation data. The estimates based on validation data were used as benchmarks against which the bias due to measurement errors in survey-based estimates was evaluated. Both analyses ignoring spell outcome and cause-specific analyses were conducted. In the cause-specific analyses, the outcome of interest was defined as transition from unemployment to employment.

The analyses were conducted in two phases. The phase 1 analyses were concerned with measurement errors in spell durations only. Therefore, spell outcomes were ignored. The phase 1 survey data analyses were conducted using survey spells and register covariates. By using the same source of covariates as in the validation data, the differences in estimates could only be attributed to differences in spell durations. The Phase 2 analyses took measurement errors in spell outcomes into account by conducting cause-specific analyses. Phase 2 survey data analyses were conducted using survey spell durations and outcomes, and register covariates.

Differences in estimates based on survey and validation data result not only from measurement errors but also from low measurement accuracy in survey data. Survey reports on main activity state were collected at the accuracy of one month. Moreover, if a person has had various activity states during a month, employment was preferred over other states. Therefore, it is difficult to obtain information on unemployment spells shorter than one month. We aimed at separating the effects of measurement accuracy and measurement error by discretizing the register spells at the accuracy of one month and repeating the analyses with discretized data. Differences between estimates based on survey data and discretized register data (reg2) could then be taken as estimates of bias due to measurement error:

$$\hat{B}ias_{me} = \hat{\beta}_{survey} - \hat{\beta}_{reg2}.$$

Bias due to measurement accuracy could be estimated by calculating differences of estimates from original (reg) and discretised register data:

$$\hat{B}ias_{ma} = \hat{\beta}_{reg2} - \hat{\beta}_{reg}.$$

4.2 Simulation study (Article [3])

Statistical properties of the IPCW method in design-based survival analysis in the presence of dependent censoring were assessed by simulation methods. The parameters of interest were defined as the values of the finite population survival function $S(t)$ at certain time points and the finite population regression coefficient B from a Cox proportional hazards model.

Generation of the populations. Four different populations of persons, each of size $N = 10{,}000$ and corresponding to the following scenarios were generated:

1 The variable determining the censoring mechanism is known,
2 A variable that is either a) strongly or b) weakly associated with the variable determining the censoring mechanism is observed,
3 The variable determining the censoring mechanism is unknown.

The population characteristics consist of three binary variables: social exclusion, sex and level of education and a variable describing the length of the unemployment spell. Social exclusion determines the probability of censoring but is unobserved. Sex was used both as an auxiliary variable in the censoring model and as a stratification variable in the sampling stage. Level of education is the covariate in the survival model whose effect on the length of unemployment spells is of interest. The four populations differ by the degree of association between variables sex and social exclusion, see Table 2 in [3]. Perfect (No) association between sex and social exclusion corresponds to scenario 1 (3) above.

Unemployment spells were generated from the Weibull distribution using a value of 0.8 for the shape parameter (a decreasing hazard rate) and scale parameters depending on the level of education and social exclusion. The median duration of the unemployment spells, as well as the effect of education on the hazard of spell completion, are different among the excluded and the non-excluded. Censoring that depends on social exclusion thus biases both the estimates of survival function and the estimate of the regression coefficient.

Sampling design and estimation. From each population, 500 stratified simple random samples of size $\boldsymbol{n}$=600 were drawn without replacement and using sex as a stratification variable. Inclusion probabilities of 0.07 for men and 0.05 for women were used. For each sample, an artificial 2-wave panel survey was conducted. It was assumed that there is no non-response at wave 1. Selective survey attrition at wave 2 was generated by stratifying the samples according to exclusion status and drawing 80% samples of respondents among the non-excluded and 20% samples of respondents among the excluded. For each sample $\boldsymbol{s}_j$, the IPC corrected design weights (see equation 7 in [3]) were constructed using sex as an auxiliary variable in the censoring model. Estimates $\hat{B}_j$ and $\hat{S}_j(t)$ were calculated using these weights. The empirical distribution of these estimates was used as an approximation of the sampling distribution of $\hat{B}$ and $\hat{S}(t)$.

5 Results

5.1 Article [1]

Determinants of non-response and attrition. Initial non-response and attrition turned out to be different processes driven by different background variables, Tables 3 and 4 in [1]. Low level of education, high household disposable income, small family size as well as being middle-aged, living in an urban municipality or in the capital region, not being married, being unemployed or outside the labour force were associated with a high probability of initial non-response. There were far fewer strong predictors of attrition which suggests that attrition was less selective than initial non-response. Young age, low level of education, low household disposable income and living in Northern Finland were associated with high probability of attrition. The difficulties in fieldwork in 2000 due to uncertainty about the continuation of the panel, showed as a peak in the attrition hazard, Table 4 in [1].

Validity of the MAR assumption. Both the initial non-response and attrition processes were non-ignorable with respect to analysis of unemployment duration. Being in the uppermost decile with respect to the number of unemployment days after the time of the first interview, raised the odds of initial non-response by 30.5% in a model including covariates of the unemployment spell model. An increase of 100 days of unemployment after the last obtained interview increased the odds of attrition hazard by 3%.

Size of bias due to non-response and attrition. Initial non-response caused downwards bias in the estimated survival function, whereas attrition did not have a biasing effect. The Hausman tests showed that both initial non-response and attrition caused bias in the coefficient estimates of a Cox shared frailty model, Table 6, [1]. The bias due to initial non-response tended to be larger than the bias due to attrition, Table 1. The largest biases were caused to the effect of receiving earnings-related unemployment benefit.

Table 1:
Analysis of unemployment duration. Non-response and attrition bias in estimates of Cox shared frailty models.

Variable	Non-response bias %	Attrition bias %
Female	–46.9	–5.4
Age	20.9	15.8
Age squared	15.4	13.4
Upper secondary education	–56.2	71.0
Higher education	–26.1	21.3
Prop. of UE[1] time	10.5	11.7
Semi urban municipality	–6.6	–31.5
Rural municipality	–18.7	–57.5
Southern Finland	–9.1	15.7
Eastern Finland	–42.0	33.7
Central Finland	–13.3	2.2
Northern Finland	–13.0	17.5
Earnings-related UE benefit	–682.8	226.6
Year 1996	20.2	17.5
Year 1997	–657.7	67.9
Year 1998	–24.9	–26.7
Year 1999	–14.5	–3.5

1 UE Unemployment

5.2 Article [2]

Type and magnitude of measurement errors in reported unemployment spells. The retrospective reports of unemployment spells showed both omitting and overreporting of spells, omitting being much more important, Figure 1 in [2]. The starts and ends of survey spells were strongly heaped at the seams between the reference periods of consecutive panel waves, Figures 2 and 3 in [2]. Of register spells ending in subsidised work, 85% were misclassified as ending because of normal employment in the survey, Table 2 in [2].

Determinants of measurement errors in reported unemployment spells. Conducting a proxy interview instead of an interview with the person of interest increased the odds of omitting unemployment spells by 72.8%, Table 3 in [2]. During the years 1998-2000, the odds of omission were more than double compared to the year 1995. This is likely a consequence of the shifting of the fieldwork period from spring to autumn from 1998 onwards, and of the resulting prolongation of the recall period by more than six months. The fieldwork covariates did not have a clear effect on the magnitude of measurement errors.

Validity of classical assumptions. Both the probability of omission and the magnitude of measurement errors depended on variables related to unemployment spells and covariates used in the event history model, Table 3 in [2]. Moreover, both the propensity to omit reporting unemployment spells and the measurement errors were correlated across survey waves. The classical assumptions about measurement errors were thus not valid.

The size of bias due to measurement errors, effect of measurement accuracy. The survey data overestimated both the median duration of unemployment (5 months vs. 2 months) and the median time to become employed (6 months vs. 3.8 months), Figures 5 and A.6 in [2]. The effect of education and in the competing risks model also the effect of receiving earnings-related unemployment benefit were estimated with sizeable bias (biases ranging from 18 to 30 percentage points and 28 to 30 percentage points, respectively), Table 6 in [2]. The bias in the effect of education was mainly due to measurement errors. Neither dummies for the heaping months, nor a more flexible model specification, protected against bias in coefficient estimates, Table 5 in [2]. The biases in January and December dummies showed that the heaping of spell starts and ends was a measurement error and not a measurement accuracy problem, Table 6 in [2]. The lack of short spells in survey data and in discretised register data led to underestimation of the baseline hazard function from the Cox proportional hazard models for durations shorter than six months, Figure A.4 in [2]. For longer durations, the biases due to measurement accuracy and measurement error worked in opposite directions. Measurement accuracy created a small positive bias leading to overestimation of the baseline hazard. The hazard spikes were however correctly placed in time. Measurement error created a large negative bias and flattened the shape of the baseline hazard. The joint effect of measurement accuracy and measurement errors was underestimation of the baseline hazard. The low measurement accuracy and the resulting lack of short spells in survey data led to badly biased shape of the baseline hazard from the Weibull model, while measurement errors only led to slight underestimation of the level of the baseline hazard (Figure A.5 in [2]).

5.3 Article [3]

The IPC corrected design weighted estimators of ***S(t)*** and ***B*** had the smallest bias in Scenario 1, see Table 3 in [3]. Scenario 1 corresponds to a situation where the censoring mechanism is known. This is an ideal situation for the IPCW method. The bias of IPC corrected design weighted estimators growed as information on the censoring mechanism lessened but was always smaller than the bias of design weighted estimators (Scenarios 2a and 2b). When the censoring mechanism was unknown (Scenario 3), the bias of IPC corrected design weighted Kaplan-Meier estimators was equal to that of design weighted estimators. In that case, there was no gain from using IPC corrected design weights in survival curve estimation. By contrast, the IPC corrected design weighted estimators of the hazard ratio performed quite well even in this case.

5.4 Article [4]

The observed information estimates of covariate effects of Cox proportional hazard models were closer to the benchmark estimates than the total respondents estimates, Table 1 in [4]. Thus using all the available data in the analysis, including the spells by attriters until the time of attrition, helped to reduce attrition bias. Comparison of the model-based and the design-based estimates revealed that the weighting correction for attrition is not very helpful in our analysis. The weights from the last wave analysed and the weights from the starting wave of the spell produced estimates that were further from the benchmark than the corresponding unweighted estimates.

The design-based estimates with total respondents data and the last wave weights were furthest from the benchmark estimates. The design-based estimates from the observed information data and weighted by the first-wave weights were closest to the benchmark estimates. However, the tests indicated nonignorability of the sample design (Table 2 in [4]) and a model-based analysis would be valid in this case. Contrary to expectations, the inclusion of design variables moved estimates *farther* from the benchmark estimates.

6 Discussion

6.1 Discussion of methods

6.1.1 Combined longitudinal survey register data

Combining longitudinal survey data with administrative register data is time-consuming and costly. In many countries, linking of various data sources is difficult because of a lack of a variable that uniquely identifies persons. Also, as noted by Calderwood and Lessof [6], legal and ethical problems may be involved. As a consequence, there are only a few studies available on non-response bias or measurement error bias in event history analysis based on combined longitudinal survey register data. Van Den Berg, Lindeboom and Dolton [37] studied initial non-response bias in the analysis of unemployment spells. Pyy-Martikainen and Rendtel [38] tested the validity of the assumption of independent censoring in event history analysis with the same data set as in this study. Mathiowetz and Duncan [39] studied the type, magnitude and determinants of measurement errors in retrospective reports of unemployment. Jäckle [40] studied measurement error bias in analysis of benefit receipt spells. I am unaware of previous studies using combined longitudinal survey register data to assess the effects of different approaches to event history analysis.

Even though combined longitudinal survey register data are considered a valuable tool for assessing errors in survey data, there are potential problems related to the use of such data. Next I discuss the relevance of four potential problems raised by Bound, Brown and Mathiowetz [7]and Biemer and Lyberg [18] to the study:

1. *The time periods for the administrative data and the survey data may not coincide*

The time periods in the survey data and the register data used in this study have a complete overlap.

2. *The definitions of the characteristic of interest may differ in administrative data and in survey data*

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. In the register, an unemployed job seeker is defined as being without a job and seeking a new job. Registering at the employment office is considered as evidence of seeking a job. Persons dismissed temporarily are regarded as unemployed. The definitions of unemployment in survey and register data are thus close to each other.

3. *The micro-merging of register and survey data is often restricted to a very specific population which makes generalisation of results problematic*

Register data were merged to all FI ECHP sample persons eligible for interview. The results obtained are thus generalisable to the population aged 16 and over and residing in Finland. This is an advantage compared to the studies by Mathiowetz and Duncan [39] and by Jäckle [40], who use samples restricted to very specific populations.

4. *Administrative data can be prone to errors as well*

As unemployed persons need to register at employment office in order to utilise their services and to receive unemployment benefits, register information is likely to cover most unemployed persons. Moreover, the duration of unemployment is likely to be precisely measured as register information on unemployment is used in order to pay unemployment benefits. However, persons who get a new job do not always inform the employment office about the job. Thus, an unemployment spell in the data base may erroneously continue for some time after the true ending date of the spell.

Lastly, linking of register and survey data is virtually error-free due to personal identity codes. All Finnish citizens are registered in the Finnish Population Information System, 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 a personal identity code that is used as a means of identifying persons. Data from the Finnish Population Information System is used throughout Finnish society's information services and management, including the production of statistics and research.

6.1.2 Simulation study

The IPC corrected design weights are time-dependent and change each time a censoring occurs in the data. To incorporate time varying weights in the analysis, the data had to be transformed into a counting process form. Each unemployment spell was split into several intervals, the splitting points being defined by the times at which censorings occurred in the sample. Time was defined as time from the beginning of the unemployment spell. The estimations were conducted by R software, which supports estimation of Kaplan-Meier survival function and Cox proportional hazard model based on counting process form data. Due to problems with computing capacity in R, the number of replicate samples had to be restricted to 500. For the same reason, the artificial data had to be generated so that all censorings occurred during first 30 days (so that there was a maximum of thirty weights per person). For applications of this method to real data, it might be useful to model the censoring process as a discrete time process where the probability of censoring changes only at the time points defined by survey interviews. The discrete-time hazard model used in [1] is one option.

6.2 Discussion of main results

The results of our study have implications for 1) survey organisations collecting event history data by longitudinal surveys; 2) researchers using longitudinal surveys for event history analysis; and 3) researchers who develop methods to correct for non-sampling errors in event history data.

6.2.1 Implications for survey organisations

Our study demonstrated a novel way to conduct a non-response analysis of longitudinal survey data. The linking of register data at person-level to survey data enables to analyse and compare the non-response and attrition processes, test the type of the missingness mechanism and estimate the size of bias due to non-response and attrition. Our study also contributed to the pool of evidence on the existence, determinants and effects of non-response, attrition and measurement errors in event history data based on longitudinal surveys. This pool may be used to provide both collectors and users of data with information on data quality, in adjusting survey estimates for non-sampling bias, and to optimise future collection of event history data by longitudinal surveys.

Our results suggest that initial non-response may be a more important source of bias than attrition in event history analysis. Other studies with different variables and different analyses have reached similar conclusions. The studies by Fitzgerald, Gottschalk and Moffitt [41] and Sisto [42] even suggested that the bias in cross-sectional estimates of income distribution and socioeconomic status caused by initial non-response may fade away over the life of the panel. These results challenge the common view of attrition being the main threat to the value of panel data [41, 43, 44], and argue in favor of conducting panel surveys in order to provide not only longitudinal but also cross-sectional data. Moreover, the existence of a fade away effect would imply that long-term panels should be preferred over short-term panels. However, a recent study with Finnish subsample of EU-SILC survey finds a clear biasing effect of panel attrition on estimates of transition probabilities between household income quintiles [45]. More research with different variables, panel surveys and countries are needed in this important issue.

According to our analysis, reported 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 by Mathiowetz [26], Mathiowetz and Duncan [39] and Kraus and Steiner [46]. The use of proxy interviews tended to cause spell omissions and should, therefore, be avoided in the collection of event history data.

A previously unnoticed finding was the classification error in reported spell outcomes. There was an excess of exits into employment in survey data due to the fact that exits into subsidised work were often misclassified by respondents as becoming employed. Attention needs to be paid to the definition of states in a multi-state framework in order to minimize misclassification errors.

Almost 40% of the register spells were shorter than one month. A measurement accuracy of one month used in ECHP main activity state calendar and currently in EU-SILC is clearly too coarse and leads to biased estimates. Register information about the distribution of the spells of interest should be taken into account in the questionnaire design phase in order to find an appropriate level of measurement accuracy.

6.2.2 Implications for researchers using longitudinal surveys for event history analysis

An unsettling result for the researchers using longitudinal surveys for the analysis of labour market transitions is that some of the key covariates such as type of unemployment benefit and level of education, had large biases due to non-response and measurement errors. Compared to the Weibull model, the more flexible Cox model did not turn out to be more robust with respect to measurement errors in estimated covariate effects. This contradicts an earlier empirical finding concerning the robustness of the Cox model with respect to initial non-response bias [37]. However, the flexibility of the Cox model was clearly advantageous in the estimation of the baseline hazard. In the light of our results, including dummies for the heaping months is not helpful in correcting measurement error bias in estimated covariate effects or distribution of spells.

As discussed in Boudreau [8], the choice of approach for analytical inference of survey data is a controversial topic. Kovačević and Roberts [10] discuss and demonstrate model-based and design-based approaches for event history analysis. The test for ignorability of sample design suggested by Pfeffermann [34] may be used to choose between these two approaches. The test compares design-based and model-based estimates of parameters of interest and rejects the null hypothesis of ignorability of sample design if the model-based estimates are "too far" from the design-based estimates. In this case, a design-based approach for the analysis should be taken. However, the use of this test may be problematic in some cases. It is not always clear in longitudinal analyses which set of weights should be used. The choice of weights may affect the result of the test. Also, our results showed that the design-based estimates may be even more biased than model-based estimates.

Longitudinal analyses often use only respondents to each wave of interest, thus discarding attriters from the analysis. In a design-based analysis using weights this can be motivated by the fact that weights are usually adjusted for non-response and attrition. However, the general purpose weights included in a survey data set may not fully correct for non-response and attrition that is selective with respect to the response variable of interest. The inclusion of the available data from the attriters might, in this case, help reduce the bias due to attrition. This is a topic shortly discussed in [9]. Our results point towards the importance of using all the available data in the analysis. The often recommended way to use survey data for longitudinal analyses; total respondents with last wave weights [19, 35] is not a modeling strategy to recommend in the light of our results.

Results from the simulation study suggest that combined design IPC weights may be useful in event history analyses based on survey data with dependent censoring. These weights were effective in reducing bias due to dependent censoring even when there was little information available about the censoring mechanism. Results from recent simulation studies by Lawless and Hajducek [47, 48] are in line with our results. However, due to the very specific purpose of the design IPC weights and the fact that the weights are time-dependent, their calculation is not easily integrated in the routine production of survey data. Instead, analysts of event history data may benefit from constructing them for their own research purposes.

6.2.3 Implications for the development of methods to correct for non-sampling errors in survey data

The number of days unemployed after the last interview had a statistically significant effect on the probabilities of non-response and attrition in our study. Moreover, measurement errors in reported unemployment spells were shown to be correlated across survey waves, with variables related to true spells and with covariates used to explain the duration of spells. Thus, neither the MAR assumption about non-response and attrition mechanisms, nor the classical assumptions about measurement errors, were valid in our study. Our results suggest that methods that make more realistic assumptions about the mechanisms generating non-sampling errors need to be developed.

6.3 Areas for future research

The performance of the IPCW method has not yet been studied with real survey data. Lawless and Hajducek [47, 48] illustrated the use of the method using jobless spell durations from Statistics Canada's Survey of Labour and Income Dynamics. However, they lacked gold standard data and were thus not able to assess neither the size of bias due to censoring nor the effectiveness of the IPCW method in reducing bias. Pyy-Martikainen and Rendtel [38] showed that censoring is independent with respect to analysis of unemployment spells in FI ECHP data. There is thus no scope for the IPCW method unless dependent censoring is generated in the data. Studies with other combined longitudinal survey register data sets might shed light on the usefulness of this method for event history analysis based on survey data.

More studies with different combined longitudinal survey register data sets and different event history variables are needed to increase our understanding of the existence, determinants and effects of non-sampling errors in event history data. Also, the effects of different modelling approaches for event history analysis, the choice of the subset of data used in analysis, and the choice of weights to use in event history analysis are areas where more research is needed.

References

[1] Marjo Pyy-Martikainen and Ulrich Rendtel. Assessing the impact of initial nonresponse and attrition in the analysis of unemployment duration with panel surveys. *Advances in Statistical Analysis*, 92:297–318, 2008.

[2] Marjo Pyy-Martikainen and Ulrich Rendtel. Measurement errors in retrospective reports of event histories. A validation study with Finnish register data. *Survey Research Methods*, 3(3):139–155, 2009.

[3] Marjo Pyy-Martikainen and Leif Nordberg. Inverse probability of censoring weighting method in survival analysis based on survey data. *Statistics in Transition*, 8(3):487–501, 2007.

[4] Marjo Pyy-Martikainen. Approaches for event history analysis based on complex longitudinal survey data. *Advances in Statistical Analysis*, 97:297–315, 2013.

[5] Robert Groves. Survey Errors and Survey Costs: *An Introduction to Survey Errors. Wiley*, 1989.

[6] Lisa Calderwood and Carli Lessof. Enhancing longitudinal surveys by linking to administrative data. In Peter Lynn, editor, *Methodology of Longitudinal Surveys*, pages 55–72.Wiley, 2009.

[7] John Bound, Charles Brown, and Nancy Mathiowetz. Measurement error in survey data. In James Heckman and Edward Leamer, editors, *Handbook of econometrics*, volume 5, pages 3705–3833. Elsevier, 2001.

[8] Christian Boudreau. *Duration Data Analysis in Longitudinal Surveys*. PhD thesis, University of Waterloo, 2003.

[9] Georgia Roberts and Milorad Kovačević. New research problems in analysis of duration data arising from complexities of longitudinal sur- veys. In *Proceedings of the Survey Methods Section*, pages 111–116. SSC Annual Meeting, 2001.

[10] Milorad Kovačević and Georgia Roberts. Modelling durations of multi- ple spells from longitudinal survey data. *Survey Methodology*, 33(1):13–22, 2007.

[11] Peter Lynn. Methods for longitudinal surveys. In Peter Lynn, editor, *Methodology of Longitudinal Surveys*, pages 1–19. Wiley, 2009.

[12] Matthias Schonlau, Nicole Watson, and Martin Kroh. Household survey panels: how much do following rules affect sample size? SOEPpaper 347 on Multidisciplinary Panel Data Research, 2010.

[13] Franco Peracchi. The European Community Household Panel: A review. *Empirical Economics*, 27(1):63–90, 2002.

[14] Marjo Pyy-Martikainen, Johanna Sisto, and Marie Reijo. The ECHP study in Finland. Quality report. Living conditions 2004:1, Statistics Finland, 2004.

[15] Jerald Lawless. Event history analysis and longitudinal surveys. In Ray Chambers and Chris Skinner, editors, *Analysis of Survey Data*, pages 221–243. Wiley, 2003.

[16] Mario Callegaro. *Seam Effects Changes due to Modifications in Question Wording and Data Collection Strategies. A Comparison of Conventional Questionnaire and Event History Calendar Seam Effects in the PSID*. PhD thesis, University of Nebraska, 2007.

[17] Paul Biemer. Total survey error: design, implementation and evaluation. *Public Opinion Quarterly*, 74(5):817–848, 2010.

[18] Paul Biemer and Lars Lyberg. *Introduction to Survey Quality*. Wiley, 2003.

[19] Grahan Kalton and Michael Brick. Weighting in household panel surveys. In David Rose, editor, *Researching Social and Economic Change: the Uses of Household Panel Studies*. Routledge, 2000.

[20] Donald Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

[21] John Kalbfleisch and Ross Prentice. *The Statistical Analysis of Failure Time Data*. Wiley, 1980.

[22] James Robins. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proceedings of the Biopharmaceutical Section*, pages 24–33. American Statistical Association, 1993.

[23] Jerald Lawless. Censoring and weighting in survival estimation from survey data. In *Proceedings of the Survey Methods Section*, pages 31–36. SSC Annual Meeting, 2003.

[24] Gerard van den Berg, Maarten Lindeboom, and Geert Ridder. Attrition in longitudinal data and the empirical analysis of dynamic labour market behaviour. *Journal of Applied Econometrics*, 9(4):421–435, 1994.

[25] Gerard van den Berg and Maarten Lindeboom. Attrition in panel data and the estimation of dynamic labor market models. Leiden university research memorandum, 1994.

[26] Nancy Mathiowetz. The problem of omissions and telescoping error: New evidence from a study of unemployment. In *Proceedings of the section on Survey research Methods*, pages 482–487, 1986.

[27] Daniel Holt, J McDonald, and Chris Skinner. The effect of measurement errors on event history analysis. In Paul Biemer, Rober Groves, Lars Lyberg, Nancy Mathiowetz, and Seymour Sudman, editors, *Measurement Errors in Surveys*, pages 665–685. Wiley, 1991.

[28] Anders Skrondal and Sophia Rabe-Hesketh. Genera*lized Latent Variable Modeling. Chapman and Hall*, 2004.

[29] Chris Skinner and K Humphreys. Weibull regression for lifetimes mea- sured with error. *Lifetime Data Analysis*, 5:23–27, 1999.

[30] Thomas Augustin. Correcting for measurement error in parametric duration models by quasi-likelihood. Discussion paper 157, Collaborative Research Center 386, 1999.

[31] Danny Pfeffermann and Michail Sverchkov. Inference under informative sampling. In Danny Pfeffermann and C Rao, editors, *Sample Surveys: Inference and Analysis*, volume 29B, pages 455–487. Elsevier, 2009.

[32] Danny Pfeffermann. Modelling of complex survey data: Why model? Why is it a problem? How can we approach it? *Survey Methodology*, 37(2):115–136, 2011.

[33] Ray Chambers and Chris Skinner, editors. *Analysis of Survey Data*, chapter 1. Wiley, 2003.

[34] Danny Pfeffermann. The role of sampling weights when modeling survey data. *International Statistical Review*, 61(2):317–337, 1993.

[35] Eurostat. ECHP UDB manual: *European Community Household Panel Longitudinal User's Database, waves 1 to 7, survey years 1994 to 2000*. Doc. Pan 168/2003-6, 2003.

[36] Jerry Hausman. Specification tests in econometrics. *Econometrica*, 46(6):1251–1271, 1978.

[37] Gerard van den Berg, Maarten Lindeboom, and Peter Dolton. Survey non-response and the duration of unemployment. *Journal of the Royal Statistical Society A*, 169(3):585–604, 2006.

[38] Marjo Pyy-Martikainen and Ulrich Rendtel. The effects of panel attri- tion on the analysis of unemployment spells. Chintex working paper 10, 2003.

[39] Nancy Mathiowetz and Greg Duncan. Out of work, out of mind: Re- sponse errors in retrosepctive reports of unemployment. *Journal of Business and Economic Statistics*, 6(2):221–229, 1988.

[40] Annette Jäckle. Measurement error and data collection methods: Effects on estimates from event history data. Iser working paper series 13, 2008.

[41] John Fitzgerald, Peter Gottschalk, and Robert Moffitt. An analysis of sample attrition in panel data. The Michigan panel study of income dynamics. *Journal of Human Resources*, 33(2):251–299, 1998.

[42] Johanna Sisto. Attrition effects on the design-based estimates of dispos- able household income. Chintex working paper 9, 2003.

[43] Ugo Trivellato. Issues in the design and analysis of panel studies: a cursory review. *Quality and Quantity*, 33:339–352, 1999.

[44] Greg Duncan. Using panel studies to understand household behaviour and well-being. In David Rose, editor, *Researching Social and Economic Change. The Uses of Household Panel Studies*. Routledge, 2000.

[45] Tara Junes. Initial wave nonresponse and panel attrition in the Finnish subsample of EU-SILC. Master's thesis, University of Helsinki, 2012.

[46] Florian Kraus and Viktor Steiner. Modelling heaping effects in unem- ployment duration models -with an application to retrospective event data in the German socio-economic panel. *In Jahrbücher für Nationalökonomie und Statistik*. Lucius & Lucius, 1998.

[47] Jerald Lawless and Dagmar Hajducek. Duration analysis in longitudinal studies with ascertainment of information at intermittent followup times and losses to followup. *Canadian Journal of Statistics*, 40(1):1–21, 2012.

[48] Jerald Lawless and Dagmar Hajducek. Estimation of finite population duration distributions from longitudinal survey panels with intermittent followup. *Lifetime Data Analysis*, 19:371–392, 2013.

Tutkimuksia-sarja Research Reports Series

Tilastokeskus on julkaissut Tutkimuksia v. 1966 alkaen, v. 1990 lähtien ovat ilmestyneet seuraavat:

164. **Henry Takala,** Kunnat ja kuntainliitot kansantalouden tilinpidossa. Tammikuu 1990. 60 s.

165. **Jarmo Hyrkkö,** Palkansaajien ansiotasoindeksi 1985=100. Tammikuu 1990. 66 s.

166. **Pekka Rytkönen,** Siivouspalvelu, ympäristöhuolto ja pesulapalvelu 1980-luvulla. Tammikuu 1990. 70 s.

167. **Jukka Muukkonen,** Luonnonvaratilinpito kestävän kehityksen kuvaajana. 1990. 119 s.

168. **Juha-Pekka Ollila,** Tieliikenteen tavarankuljetus 1980-luvulla. Helmikuu 1990. 45 s.

169. **Tuovi Allén – Seppo Laaksonen – Päivi Keinänen – Seija Ilmakunnas,** Palkkaa työstä ja sukupuolesta. Huhtikuu 1990. 90 s.

170. **Ari Tyrkkö,** Asuinolotiedot väestölaskennassa ja kotitaloustiedustelussa. Huhtikuu 1990. 63 s.

171. **Hannu Isoaho – Osmo Kivinen – Risto Rinne,** Nuorten koulutus ja kotitausta. Toukokuu 1990. 115 s.

171b. **Hannu Isoaho – Osmo Kivinen – Risto Rinne,** Education and the family background of the young in Finland. 1990. 115 pp.

172. **Tapani Valkonen – Tuija Martelin – Arja Rimpelä,** Eriarvoisuus kuoleman edessä. Sosioekonomiset kuolleisuuserot Suomessa 1971–85. Kesäkuu 1990. 145 s.

173. **Jukka Muukkonen,** Sustainable development and natural resource accounting. August 1990. 96 pp.

174. **Iiris Niemi – Hannu Pääkkönen,** Time use changes in Finland in the 1980s. August 1990. 118 pp.

175. **Väinö Kannisto,** Mortality of the elderly in late 19th and early 20th century Finland. August 1990. 50 pp.

176. **Tapani Valkonen – Tuija Martelin – Arja Rimpelä,** Socio-economic mortality differences in Finland 1971–85. December 1990. 108 pp.

177. **Jaana Lähteenmaa – Lasse Siurala,** Nuoret ja muutos. Tammikuu 1991. 211 s.

178. **Tuomo Martikainen – Risto Yrjönen,** Vaalit, puolueet ja yhteiskunnan muutos. Maaliskuu 1991. 120 s.

179. **Seppo Laaksonen,** Comparative Adjustments for Missingness in Short-term Panels. April 1991. 74 pp.

180. **Ágnes Babarczy – István Harcsa – Hannu Pääkkönen,** Time use trends in Finland and in Hungary. April 1991. 72 pp.

181. **Timo Matala,** Asumisen tuki 1988. Kesäkuu 1991. 64 s.

182. **Iiris Niemi – Parsla Eglite – Algimantas Mitrikas – V.D. Patrushev – Hannu Pääkkönen,** Time Use in Finland, Latvia, Lithuania and Russia. July 1991. 80 pp.

183. **Iiris Niemi – Hannu Pääkkönen,** Vuotuinen ajankäyttö. Joulukuu 1992. 83 s.

183b. **Iiris Niemi – Hannu Pääkkönen – Veli Rajaniemi – Seppo Laaksonen – Jarmo Lauri,** Vuotuinen ajankäyttö. Ajankäyttötutkimuksen 1987–88 taulukot. Elokuu 1991. 116 s.

184. **Ari Leppälahti – Mikael Åkerblom,** Industrial Innovation in Finland. August 1991. 82 pp.

185. **Maarit Säynevirta,** Indeksiteoria ja ansiotasoindeksi. Lokakuu 1991. 95 s.

186. **Ari Tyrkkö,** Ahtaasti asuvat. Syyskuu 1991. 134 s.

187. **Tuomo Martikainen – Risto Yrjönen,** Voting, parties and social change in Finland. October 1991. 108 pp.

188. **Timo Kolu,** Työelämän laatu 1977–1990. Työn ja hyvinvoinnin koettuja muutoksia. Tammikuu 1992. 194 s.

189. **Anna-Maija Lehto,** Työelämän laatu ja tasa-arvo. Tammikuu 1992. 196 s.

190. **Tuovi Allén – Päivi Keinänen – Seppo Laaksonen – Seija Ilmakunnas,** Wage from Work and Gender. A Study on Wage Differentials in Finland in 1985. 88 pp.

191. **Kirsti Ahlqvist,** Kodinomistajaksi velalla. Maaliskuu 1992. 98 s.

192. **Matti Simpanen – Irja Blomqvist,** Aikuiskoulutukseen osallistuminen. Aikuiskoulutustutkimus 1990. Toukokuu 1992. 135 s.

193. **Leena M. Kirjavainen – Bistra Anachkova – Seppo Laaksonen – Iiris Niemi – Hannu Pääkkönen – Zahari Staikov,** Housework Time in Bulgaria and Finland. June 1992. 131 pp.

194. **Pekka Haapala – Seppo Kouvonen,** Kuntasektorin työvoimakustannukset. Kesäkuu 1992. 70 s.

195. **Pirkko Aulin-Ahmavaara,** The Productivity of a Nation. November 1992. 72 pp.

196. **Tuula Melkas,** Valtion ja markkinoiden tuolla puolen. Kanssaihmisten apu Suomessa 1980-luvun lopulla. Joulukuu 1992. 150 s.

197. **Fjalar Finnäs,** Formation of unions and families in Finnish cohorts born 1938–67. April 1993. 58 pp.

198. **Antti Siikanen – Ari Tyrkkö,** Koti – Talous – Asuntomarkkinat. Kesäkuu 1993. 167 s.

199. **Timo Matala,** Asumisen tuki ja aravavuokralaiset. Kesäkuu 1993. 84 s.

200. **Arja Kinnunen,** Kuluttajahintaindeksi 1990=100. Menetelmät ja käytäntö. Elokuu 1993. 89 s.

201. **Matti Simpanen,** Aikuiskoulutus ja työelämä. Aikuiskoulutustutkimus 1990. Syyskuu 1993. 150 s.

202. **Martti Puohiniemi,** Suomalaisten arvot ja tulevaisuus. Lokakuu 1993. 100 s.

203. **Juha Kivinen – Ari Mäkinen,** Suomen elintarvike- ja metallituoteteollisuuden rakenteen, kannattavuuden ja suhdannevaihteluiden yhteys; ekonometrinen analyysi vuosilta 1974 – 1990. Marraskuu 1993. 92 s.

204. **Juha Nurmela,** Kotitalouksien energian kokonaiskulutus 1990. Marraskuu 1993. 108 s.

205a. **Georg Luther,** Suomen tilastotoimen historia vuoteen 1970. Joulukuu 1993. 382 s.

205b. **Georg Luther,** Statistikens historia i Finland till 1970. December 1993. 380 s.

206. **Riitta Harala – Eva Hänninen-Salmelin – Kaisa Kauppinen-Toropainen – Päivi Keinänen – Tuulikki Petäjäniemi – Sinikka Vanhala,** Naiset huipulla. Huhtikuu 1994. 64 s.

207. **Wangqiu Song,** Hedoninen regressioanalyysi kuluttajahintaindeksissä. Huhtikuu 1994. 100 s.

208. **Anne Koponen,** Työolot ja ammatillinen aikuiskoulutus 1990. Toukokuu 1994. 118 s.

209. **Fjalar Finnäs,** Language Shifts and Migration. May 1994. 37 pp.

210. **Erkki Pahkinen – Veijo Ritola,** Suhdannekäänne ja taloudelliset aikasarjat. Kesäkuu 1994. 200 s.

211. **Riitta Harala – Eva Hänninen-Salmelin – Kaisa Kauppinen-Toropainen – Päivi Keinänen – Tuulikki Petäjäniemi – Sinikka Vanhala,** Women at the Top. July 1994. 66 pp.

212. **Olavi Lehtoranta,** Teollisuuden tuottavuuskehityksen mittaminen toimialatasolla. Tammikuu 1995. 73 s.

213. **Kristiina Manderbacka,** Terveydentilan mittarit. Syyskuu 1995. 121 s.

214. **Andres Vikat,** Perheellistyminen Virossa ja Suomessa. Joulukuu 1995. 52 s.

215. **Mika Maliranta,** Suomen tehdasteollisuuden tuottavuus. Helmikuu 1996. 189 s.

216. **Juha Nurmela,** Kotitaloudet ja energia vuonna 2015. Huhtikuu 1996. 285 s.

217. **Rauno Sairinen,** Suomalaiset ja ympäristöpolitiikka. Elokuu 1996. 179 s.

218. **Johanna Moisander,** Attitudes and Ecologically Responsible Consumption. August 1996. 159 pp.

219. **Seppo Laaksonen** (ed.), International Perspectives on Nonresponse. Proceedings of the Sixth International Workshop on Household Survey Nonresponse. December 1996. 240 pp.

220. **Jukka Hoffrén,** Metsien ekologisen laadun mittaaminen. Elokuu 1996. 79 s.

221. **Jarmo Rusanen – Arvo Naukkarinen – Alfred Colpaert – Toivo Muilu,** Differences in the Spatial Structure of the Population Between Finland and Sweden in 1995 – a GIS viewpoint. March 1997. 46 pp.

222. **Anna-Maija Lehto,** Työolot tutkimuskohteena. Marraskuu 1996. 289 s.

223. **Seppo Laaksonen** (ed.), The Evolution of Firms and Industries. June 1997. 505 pp.

224. **Jukka Hoffrén,** Finnish Forest Resource Accounting and Ecological Sustainability. June 1997. 132 pp.

225. **Eero Tanskanen,** Suomalaiset ja ympäristö kansainvälisestä näkökulmasta. Elokuu 1997. 153 s.

226. **Jukka Hoffrén,** Talous hyvinvoinnin ja ympäristöhaittojen tuottajana – Suomen ekotehokkuuden mittaaminen. Toukokuu 1999. 154 s.

227. **Sirpa Kolehmainen,** Naisten ja miesten työt. Työmarkkinoiden segregoituminen Suomessa 1970–1990. Lokakuu 1999. 321 s.

228. **Seppo Paananen,** Suomalaisuuden armoilla. Ulkomaalaisten työnhakijoiden luokittelu. Lokakuu 1999. 152 s.

229. **Jukka Hoffrén,** Measuring the Eco-efficiency of the Finnish Economy. October 1999. 80 pp.

230. **Anna-Maija Lehto – Noora Järnefelt** (toim.), Jaksaen ja joustaen. Artikkeleita työolotutkimuksesta. Joulukuu 2000. 264 s.

231. **Kari Djerf,** Properties of some estimators under unit nonresponse. January 2001. 76 pp.

232. **Ismo Teikari,** Poisson mixture sampling in controlling the distribution of response burden in longitudinal and cross section business surveys. March 2001. 120 pp.

233. **Jukka Hoffrén,** Measuring the Eco-efficiency of Welfare Generation in a National Economy. The Case of Finland. November 2001. 199 pp.

234. **Pia Pulkkinen,** ”Vähän enemmän arvoinen” Tutkimus tasa-arvokokemuksista työpaikoilla. Tammikuu 2002. 154 s.

235. **Noora Järnefelt – Anna-Maija Lehto,** Työhulluja vai hulluja töitä? Tutkimus kiirekokemuksista työpaikoilla. Huhtikuu 2002. 130 s.

236. **Markku Heiskanen,** Väkivalta, pelko, turvattomuus. Surveytutkimusten näkökulmia suomalaisten turvallisuuteen. Huhtikuu 2002. 323 s.

237. **Tuula Melkas,** Sosiaalisesta muodosta toiseen. Suomalaisten yksityiselämän sosiaalisuuden tarkastelua vuosilta 1986 ja 1994. Huhtikuu 2003. 195 s.

238. **Rune Höglund – Markus Jäntti – Gunnar Rosenqvist** (eds.), Statistics, econometrics and society: Essays in honour of Leif Nordberg. April 2003. 260 pp.

239. **Johanna Laiho – Tarja Nieminen** (toim.), Terveys 2000 -tutkimus. Aikuisväestön haastatteluaineiston tilastollinen laatu. Otanta-asetelma, tiedonkeruu, vastauskato ja estimointi- ja analyysiasetelma. Maaliskuu 2004. 95 s.

240. **Pauli Ollila,** A Theoretical Overview for Variance Estimation in Sampling Theory with Some New Techniques for Complex Estimators. September 2004. 151 pp.

241. **Minna Piispa.** Väkivalta ja parisuhde. Nuorten naisten kokeman parisuhdeväkivallan määrittely surveytutkimuksessa. Syyskuu 2004. 216 s.

242. **Eugen Koev.** Combining Classification and Hedonic Quality Adjustment in Constructing a House Price Index. Tammikuu 2013. 67 pp.

243. **Henna Isoniemi – Irmeli Penttilä** (toim.), Perheiden muuttuvat elinolot. Artikkeleita lapsiperheiden elämänmuutoksista. Syyskuu 2005. 168 s.

244. **Anna-Maija Lehto – Hanna Sutela – Arto Miettinen** (toim.), Kaikilla mausteilla. Artikkeleita työolotutkimuksesta. Toukokuu 2006. 385 s.

245. **Jukka Jalava – Jari Eloranta – Jari Ojala** (toim.) Muutoksen merkit – Kvantitatiivisia perspektiivejä Suomen taloushistoriaan. Tammikuu 2007. 373 s.

246. **Jari Kauppila.** The Structure and Short-Term Development of Finnish Industries in the 1920s and 1930s. An Input-output Approach. Elokuu 2007. 274 s.

247. **Mikko Myrskylä.** Generalised Regression Estimation for Domain Class Frequencies. Elokuu 2007. 137 pp.

248. **Jukka Jalava.** Essays on Finnish Economic Growth and Productivity, 1860–2005. Joulukuu 2007. 154 s.

249. **Yrjö Tala.** Kirkon vai valtion kirjat? Uskontokuntasidonnaisuuden ongelma Suomen väestökirjanpidossa 1839–1904. 317 s.

250. **Hanna-Kaisa Rättö.** Hyvinvointi ja hyvinvoinnin mittaamisen kehittäminen. Huhtikuu 2009. 82 s.

251. **Pertti Koistinen** (toim.), Työn hiipuvat rajat. Tutkielmia palkkatyön, hoivan ja vapaaehtoistyön muuttuvista suhteista. Huhtikuu 2009. 159 s.

252. **Kirsti Ahlqvist.** Kulutus, tieto, hallinta. Kulutuksen tilastoinnin muutokset 1900-luvun Suomessa. Maaliskuu 2010. 314 s.

253. **Jukka Hoffrén (editor) – Eeva-Lotta Apajalahti – Hanna Rättö.** Economy-wide MFA with Hidden Flows for Finland: 1945–2008. 89 pp.

254. **Hannu Pääkkönen.** Perheiden aika ja ajankäyttö. Tutkimuksia kokonaistyöajasta, vapaaehtoistyöstä, lapsista ja kiireestä. Toukokuu 2010. 260 s.

255. **Anna Pärnänen.** Organisaatioiden ikäpolitiikat:strategiat, instituutiot ja moraali. Helmikuu 2011. 291 s.

256. **Ilja Kristian Kavonius.** Kädestä suuhun – Makro- ja mikrotaloudellinen tarkastelu suomalaisten kotitalouksien säästämisestä ja sen mittaamisesta 1950-luvulla. Keskäkuu 2011. 193 s.

257. **Vesa Kuusela.** Paradigms in Statistical Inference for Finite Populations. Up to the 1950s. Elokuu 2011. 236 pp.

258. **Arto Kokkinen.** On Finland's Economic Growth and Convergence with Sweden and the EU15 in the 20th Century. Maaliskuu 2012. 253 pp.

259. **Hanna Sutela.** Määräaikainen työ ja perheellistyminen Suomessa 1984–2008. Joulukuu 2012. 208 s.

260. **Marjo Pyy-Martikainen.** Statistical Analysis of Survey-based Event History Data with Application to Modeling of Unemployment Duration. Lokakuu 2013. 114 pp.

The Research Reports series describes Finnish society in the light of up-to-date research results. Scientific studies that are carried out at Statistics Finland or are based on the datasets of Statistics Finland are published in the series.

Longitudinal surveys are increasingly used to collect event history data on person-specific processes such as transitions between labour market states. Survey-based event history data pose a number of challenges for statistical analysis. These challenges include survey errors due to sampling, non-response, attrition and measurement.

This study deals with non-response, attrition and measurement errors in event history data and the bias caused by them in event history analysis. The study also discusses some choices faced by a researcher using longitudinal survey data for event history analysis and demonstrates their effects. These choices include, whether a design-based or a model-based approach is taken, which subset of data to use and, if a design-based approach is taken, which weights to use.

The study takes advantage of the possibility to use combined longitudinal survey register data. The Finnish subset of European Community Household Panel (FI ECHP) survey for waves 1–5 were linked at person-level with longitudinal register data.

ISSN 1798–8578
= Research Reports
ISBN 978–952–244–456–1 (pdf)
ISBN 978–952–244–455–4 (print)

Tietopalvelu, Tilastokeskus
puh. 09 1734 2220
www.tilastokeskus.fi

Julkaisutilaukset, Edita Publishing Oy
puh. 020 450 05
asiakaspalvelu.publishing@edita.fi
www.editapublishing.fi

Informationstjänst, Statistikcentralen
tel. +358 9 1734 2220
www.stat.fi

Beställning av publikationer, Edita Publishing Oy
tel. +358 20 450 05
www.editapublishing.fi

Information Service, Statistics Finland
tel. +358 9 1734 2220
www.stat.fi

Publication orders, Edita Publishing Oy
tel. +358 20 450 05
www.editapublishing.fi