Attrition in Panel Data: the Effectiveness of Weighting

Leen Vandecasteele and Annelies Debels, K.U. Leuven

This is a pre-copyedited, author-produced version of an article accepted for publication in the European Sociological Review by Oxford University Press following peer review. The version of record Vandecasteele, L., & Debels, A. (2006). Attrition in panel data: the effectiveness of weighting. European Sociological Review, 23(1), 81-97. is available online at: https://doi.org/10.1093/esr/jcl021.

Abstract

Although weighting seems to be a popular strategy for coping with panel attrition in sociological research it is not clear how effective it is in reducing attrition bias. This article aims to fill this gap by giving an assessment of the effectiveness of weighting in the European Community Household Panel. Estimates of the distribution of social class and education appear to be dropout biased in the ECHP. Moreover, the direction of the bias differs across countries. Weighting with the ECHP longitudinal weight tends to reduce this bias to some extent; yet, unexpectedly, it also sometimes increases it. The latter problem is largely avoided, however, by replacing the available ECHP-weight with a new longitudinal weight that includes better predictors of dropout, such as covariates related to the interview process, to respondents' previous interview experiences and education.

Key Words: **attrition / dropout / weighting / ECHP / social class / education**

Introduction

The unique design of panel studies offers a wide and appealing range of opportunities to social scientists. However, most panel surveys are plagued with dropout or attrition of observation units throughout the subsequent waves. Attrition may not only decrease the sample size and thereby diminish the efficiency of estimates. It may also lead to biased estimates whenever cases are not dropping out randomly from the original sample (Engel and Reinecke, 1996; Rose, 2000).

There are different ways of dealing with attrition, based on varying assumptions about the dropout process. The aim of this article is to assess the effectiveness of one of these strategies, namely weighting. The dataset used to perform this evaluation is the European Community Household Panel (ECHP) provided by Eurostat. Although this dataset is widely used in the social sciences, most researchers pay little attention to the problem of dropout: a majority of articles based on the ECHP ignore it altogether. If possible dropout bias is acknowledged, then weighting is by far the most commonly applied strategy.[1] Several researchers have shown that attrition in the ECHP is substantial (Behr *et al.*, Bellgardt and Rendtel, 2005; Nicoletti and Peracchi, 2005; Watson, 2003). Yet, a thorough analysis of the extent to which weighting helps to reduce dropout bias is currently lacking in the literature.

After a brief overview of common strategies of dealing with attrition, this article devotes a section to how panel weights that take into account dropout probabilities are usually developed. It is argued that the predicting variables used to construct these weights should be strongly related to dropout. We identify the most influential predictors of dropout in the literature and discuss to what extent European panel surveys use these in the weights they recommend.

In the empirical part of the article, the implications of this theoretical discussion are assessed by means of illustrations based on the ECHP. The question in this respect is double: 1. to what extent does dropout influence estimations of the relative frequencies of social class and education? and 2. to what extent does weighting reduce this bias? The examples of social class and education are chosen because of their relevance in the sociological literature.

Three methods are used to investigate these questions. 1. First, we examine the effect of social class on the time to dropout from the ECHP by means of a discrete-time logistic hazard model. Next, we add as covariates to this model the variables used to construct the ECHP-weights, in order to see whether weighting would reduce dropout bias in estimates with social class as the outcome variable. 2. Secondly, we compare the distribution of social class in the first wave for respondents who never drop out with that for all initial sample respondents, including the ones who subsequently attrite from the panel. In this way we can assess both whether dropout would lead to misrepresentations of the distribution of social class and to what extent weighting accounts for this. 3. Thirdly, for Denmark only, the distribution of education in the ECHP is compared with external population data. Again, the bias of the ECHP estimates is assessed for both weighted and unweighted distributions.

Along with this threefold assessment, a new longitudinal weight is proposed that takes into account most of the weighting variables identified as important predictors of dropout in the first part of the article. The performance of this new longitudinal weight is assessed and compared to the ECHP-weight.

Attrition in panel surveys: mechanisms and coping strategies

Assumptions about dropout mechanisms

Dropout occurs when people who participate at least once in the panel, do not continue their participation until the end of the study. In the missing data literature a number of assumptions about missing data mechanisms are distinguished, each involving different consequences for analysis (Little and Rubin, 1987). Suppose dropout yields missing data on an outcome variable Y in year t (Y_t). The missing data on Y_t due to dropout are *missing completely at random* (MCAR) if the response probability in year t (R_t) is unrelated both to the outcome variable Y_t and to a set of i variables X_i that are not subject to nonresponse. These variables can consist of both time-constant person-specific characteristics (e.g. gender, nationality), or of time-varying variables with non-missing information on previous measurements with the person (e.g. household income in year t-1). In the MCAR case, the units with missing values on Y_t form a random subsample of the complete sample. The dropout mechanism is considered *missing at random* (MAR) when R_t depends only on observed variables X_i and not on Y_t (Little, 1995). In this case, the subsample of units with missing values is not random with respect to Y_t, but within the subclasses of X_i it is. The worst scenario occurs when the response probability R_t is related not only to the observed variables X_i, but also to the unobserved variable of interest Y_t. This missing value-mechanism is considered *missing not at random* (MNAR) or non-ignorable.

How to cope with attrition?

As explained in the introduction, inferences made from an incomplete dataset will influence the outcome. Therefore, it is essential to consider an appropriate method for taking dropout into account. In what follows an overview is given of the most commonly adopted approaches in the handling of nonresponse.

The first and most frequently applied strategy is to ignore the nonresponse altogether and pursue an analysis with the available cases, so-called complete case analysis or listwise deletion (Allison, 2002; Graham and Hofer, 2000). The appeal of this approach is its simplicity and general usability (Little and Rubin, 1989). In addition, the standard errors of the obtained estimates are correct, because the sample size is known (Graham and Hofer, 2000). In the context of longitudinal panel data, one could opt to pursue the analysis with a balanced subpanel of individuals with complete participation in all waves (Verbeek, 2004). However, the longer the panel study, the smaller this subpanel becomes, and with a mature panel the loss of power often becomes problematic. An alternative is to include all individuals even if they are not observed at all time points. However, both alternatives may suffer from a more severe drawback: the parameter estimates will be biased when the data are not MCAR (Diggle *et al.*, 2002).

When the data are not MCAR, one could opt to pursue the analysis with only the available cases, but assign a weight to these cases to compensate for dropouts. Although weighting is popular and rather straightforward, it is based on the assumption that data are MAR. In particular it is assumed that dropout occurs randomly after weighting with those observed variables in the dataset that are associated with dropout. A drawback of weighting is that it tends to decrease the precision of the estimates because the standard errors are overestimated (Raab *et al.*, 2005).

An alternative approach is to impute missing data, a method usually applied in the case of item nonresponse (Kalton and Kasprzyk, 1986). Imputation methods replace missing values by a reasonable estimate and consequently the data are analysed as if they were complete. For example, in the context of longitudinal research, the last observation could be carried forward or a summary measure of the general trend could be imputed (Diggle *et al.*, 2002). A fundamental problem arises however when imputed data are analysed in standard procedures as if they were real complete data, because the precision of test statistics is overestimated and standard errors are underestimated (Allison, 2002; Little and Rubin, 1989). This problem can be avoided by using multiple imputation (Graham and Hofer, 2000; Kalton and Kasprzyk, 1986).

For some types of analysis it is possible to use maximum likelihood estimation. This method yields unbiased estimates on the condition that the missing data pattern is MAR. Diggle and Kenward (1994) have applied the maximum likelihood estimation method to the case of dropout in longitudinal data analysis.

An important assumption with the methods presented up until now is that the data should be either MCAR (with available case analysis) or MAR (with the other methods). Often however, data will be MNAR. In this case, one could consider applying one of the models that take into account this non-ignorable missing data problem by simultaneously modelling the nonresponse and the substantial analysis. Several models have been proposed, and they have proven usable in a longitudinal panel data framework. Examples are selection models (such as the Heckman model) and pattern-mixture models (Allison, 2002; Diggle *et al.*, 2002; Little, 1995; Verbeek, 2004). Yet, it must be noted that these models are not a panacea for all problems, because dissimilar models often yield different results and because they are based on rather strict assumptions (Allison, 2002; Little, 1995).

Weighting in panel surveys

Weighting strategies

This section explains how weighting for nonresponse is accomplished in panel surveys and how it differs from weighting for other purposes. Generally, weighting is applied in order to make survey estimates a better representation of population estimates. The reasons why survey estimates may not be representative of the population are double: this may be due either to unequal sample selection probabilities in a stratified design, or to nonresponse of sample units. Weights developed to correct for the former are called design weights and are calculated as the inverse of the sample selection probability (Raab *et al.*, 2005). Nonresponse weights aim to compensate for unit nonresponse, both initial sample nonresponse and subsequent attrition.

As opposed to design weights, which are known exactly, nonreponse weights need to be estimated on the basis of available information about the nonrespondents. One way to develop nonresponse weights is to rely on external sources (e.g. censuses) that provide population information regarding a number of control variables, like age, sex and region. Weights are then constructed so that the weighted sample distribution over these control variables mirrors their population distribution. This is called post-stratification (Kalton and Kasprzyk, 1986), and may be applied in order to reduce nonresponse bias, but also other sampling inadequacies.

Another way to proceed is to use information that has been collected on nonrespondents, such as basic data on geographical location, neighbourhood or reason for nonresponse. Additionally, in case of attrition, there is often a considerable amount of information available in the form of variables collected in the waves prior to attrition. For the construction of panel nonresponse weights, it is recommended to utilise this information to the best possible extent. Usually, it is advised to search for those weighting variables that are related most closely to the response propensity, because including these variables will generally reduce nonresponse bias (Kalton and Brick, 2000).

Selective attrition in panel studies: previous findings

In the construction of weights aimed at reducing dropout bias, it is important to select those weighting variables that are significant predictors of dropout patterns. Therefore, this section gives an overview of previous research into the predictors of panel nonresponse and panel dropout in socio-economic household panels.

Lepkowksi and Couper (2002) explain that panel nonresponse depends on the successful completion of three steps, namely: *locating* respondents, *contacting* them and obtaining their *cooperation*. Research by Nicoletti and Peracchi (2005) has shown that good predictors of the probability of contact in the ECHP are number of children, home ownership, length of residence at current address, an index of item nonresponse to household income, length of the fieldwork and duration of the interview and number of visits by the interviewer; while the probability of cooperation conditional on contact is closely related to labour force status, living in a couple, infrequent interactions with neighbours and mode of the interview. Moreover, Behr *et al.* (2005) conclude that the strongest indicators of low response probabilities are interviewer change and moving address. Additionally, individuals not completing their interview in the first wave and persons with missing values on crucial variables in previous waves are more likely to drop out (Watson, 2003). These findings suggest that including variables relating to the data collection process in the weighting procedure might be useful. Other factors with a positive impact on staying in the ECHP or in other household panels such as the American Panel Study of Income Dynamics (PSID) are: not moving, being married, having children, being middle-aged and highly educated (Behr *et al.*, 2005; Fitzgerald *et al.*, 1998; Lillard and Panis, 1998; Nicoletti and Peracchi, 2002; Watson, 2003; Zabel, 1998).

Not only does attrition appear to be selective with respect to these variables, but it also turns out that the direction and significance of these effects varies substantially across countries and waves (Behr *et al.*, 2005). For example, Watson (2003) points out that in Northern European countries, higher educated people are less likely to drop out, but that this effect is reversed in Southern European countries, where higher educated people are more likely to be lost. A similar interaction effect has been established with respect to income and country (Watson, 2003).

Although evidence abounds that panel attrition is selective along important social characteristics, some studies conclude that this is not necessarily problematic for a correct estimation of statistics (Behr *et al.*, 2003; Lillard and Panis, 1998; Watson, 2003).

Weighting variables in household panel surveys

In the previous section, we presented a number of important predictors of panel attrition found in earlier research. In this section these determinants of dropout are compared with variables used for weighting in the largest European panel studies.

Although some differences can be discerned concerning the choice of weighting variables across surveys, Table 1 shows that major similarities prevail. This is not surprising given that most of these covariates have proven to be influential determinants of attrition. However, it must be noted that a number of relevant variables are omitted in the construction of the ECHP-weight in particular. In contrast with the German Socio-Economic Panel (GSOEP) recommendations, the ECHP-weight does not include any information on panel design features or panel experiences, which nonetheless have been shown successful in predicting attrition. Another important variable that is included in the GSOEP and British Household Panel Survey (BHPS) weighting procedures but not in the ECHP-weight is whether the household has moved. Finally, the omission of education level in the ECHP-weight is rather surprising given the substantial importance of this variable for most sociological research.

Table 1 Comparative overview of determinants of dropout found in the literature and variables used for longitudinal weighting in panel studies.

Determinants of dropout found in literature		**Variables used for longitudinal weighting**		
Panel Study of Income Dynamics	**European Community Household Panel**	**European Community Household Panel longitudinal weight**[1]	**British Household Panel Study longitudinal weight**[2]	**German Socio-Economic Panel determinants of dropout**[3]
age	age	age	age	age of household head
		sex	sex	sex of household head
			race	
		region	region	large city
				resident of East Berlin
				migration from East to West Germany
marital status	marital status	whether split-off household		splitt-off household
	living in a couple			separation/divorce
				single household
number of children	children	household size	household size	
		arrivals to and departures from the household		
			organisational membership	
		nr of economically active persons in the household	employment status	expected loss of job
household income	household income	household income	income	household income
		main source of income	income composition	
			number of cars	
			ownership of durables	
owner of dwelling	tenure status	tenure status	tenure status	
education level	education level		educational qualifications	
	labour force status			
	duration of interview			
	mode of interview			
	number of visits			
	interviewer change			interviewer change
				number of interviews with same interviewer
	length of fieldwork			
	item nonresponse on crucial variables			item nonresponse on income
	uncompleted interview in first wave			
	interaction with neighbours			
	length of stay at current address			
	moving		moving	household moved

[1] *Source: Eurostat (2003a)*
[2] *Source: Buck (2003)*
[3] *Determinants of dropout rates, referred to in longitudinal weighting procedure (Haisken-DeNew and Frick, 2003)*

Data and operationalisation

Data and main concepts

The ECHP is a standardised socio-economic survey that has been submitted annually to a panel of individuals and households in different EU member states (Eurostat, 2003b). The panel study started in 1994, and ran for eight waves until 2001. In the first wave of the ECHP, a sample of approximately 60.500 households and 130.000 individuals aged 16 and over were interviewed, with the aim of being representative of all individuals living in private households within the European Union at that time. The countries selected for this article have complete coverage of the eight waves of the ECHP and their weights have been constructed according to the standardised procedure described by Eurostat (Eurostat, 2003a). Included are Belgium, Denmark, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal and Spain.[2]

Dropout can take place both at the individual and the household level. In this article the focus is on individual dropout, in order to be able to take into account the differences in participation within households. Attrition may occur for different reasons. First, it is possible that a person is no longer a member of the population under observation. This happens when a person dies, in the case of institutionalisation, migration to a country outside the European Union, or movement of a non-sample person to a household without sample persons.[3] In these cases, the respondent is no longer approached for participation in the panel study. Alternatively, dropout can be due to the non-participation of persons who are selected for the personal interview. In this article, only this last kind of attrition is considered.

The variable social class is based on the occupational position of the main breadwinner in the household. Social class can be considered a household characteristic as it is shared between household members. We apply a reduced (6 categories) version of the Goldthorpe (12 categories) social class typology. In addition, a separate class is reserved for persons living in a household where the main breadwinner is not employed. Unemployment and inactivity are widespread phenomena, but are too often ignored in social class analysis. In sum, we distinguish the following classes: 1. unemployed or inactive, 2. semi-unskilled manual, 3. supervisory & skilled manual, 4. self-employed, 5. routine nonmanual, 6. lower professionals and 7. higher professionals.

Education has three categories: high education, i.e. recognized third level education (ISCED 5-7), middle education, i.e. second stage of secondary education (ISCED 3) and low education, i.e. less than second stage of secondary education (ISCED 0-2).

Construction of personal weights in the ECHP

For a detailed discussion of cross-sectional and longitudinal weights in the ECHP we refer to the user documentation (Eurostat, 2003a). Here, the most important steps are summarised. The construction of the ECHP-weights departs from a starting weight that is assigned to every sample person (Eurostat, 2003a). In the first wave this starting weight consists of the design weight, while in the subsequent waves it consists of the finally obtained base weight from the previous wave. The starting weight is multiplied by a factor taking into account response probabilities, calculated on the basis of logistic regressions.[4] These models are estimated on the basis of the covariates enumerated in Table 1. The obtained weight is poststratified in order to reflect the distributions of age, sex and household size[5] in the population. This results in a longitudinal (=base) weight for each sample person. Non-sample persons receive a zero base weight. The base weight is suitable for longitudinal analyses that use the waves up till the wave of the weight (Eurostat, 2003b). The ECHP also provides cross-sectional weights, calculated as the average of the base weights of all sample persons in the households and assigned to both sample and non-sample persons in the household.

Selective attrition and the effectiveness of the ECHP-weight: a hazard model approach

In this section we address the question how social class influences the hazard of dropping out of the ECHP, both without and with controlling for the variables of the ECHP-weights. With respect to the relationship between social class and attrition, we expect to find the same North/South divide in Europe as has been found for education level and income (Behr *et al.*, 2005; Watson, 2003). It can indeed be argued that there are links between someone's occupation, attained education level and income (Covello and Bollen, 1979; Schooler and Schoenbach, 1994; Teichler, 2001). The expectation is that there will be more attrition in the manual classes in Northern European countries, compared to the nonmanual classes. The reverse effect is expected in Southern European countries. For the category of inactive/unemployed, we expect to find more dropout because questions about professional status, income and other economic characteristics may be threatening and/or irrelevant for this group.

In Table 2 the results of two discrete-time logistic hazard models are presented. The dependent variable is the hazard of dropout, i.e. the probability of dropping out, given that one has not dropped out before; the independent variable is social class. Only respondents to the first wave are included in both models. Both models are weighted with the ECHP base weight of the first wave in order to distinguish the effect of weighting for attrition from the effect of weighting for design issues and poststratification of the initial sample. Model 1 displays univariate regression outcomes. There is a highly significant effect of social class on dropout in all countries, but the direction of the effects is not consistent. As expected, in the Southern countries -and in Ireland- manual classes, but also the routine nonmanual and self-employed classes tend to remain longer in the panel than higher (and lower) professionals. However, Spain is an exception. This country does not fit the Southern pattern, in the sense that manual classes do not have a significantly lower dropout hazard while the self-employed and routine nonmanual class do have a higher dropout hazard than the reference category of higher professionals. In the latter aspect it resembles more closely the Northern countries, where the routine nonmanuals are often more likely to drop out quickly than the higher professionals. However, the hypothesis that lower social classes attrite sooner in Northern countries is not consistently confirmed. When comparing the dropout hazard of the manual classes with that of the higher professionals, the expected positive effect is significant only in Denmark and in Belgium. Furthermore, in line with the hypothesis, we find that the unemployed/inactive have higher dropout hazards than the professional classes in all countries. Consequently, we could say that the effects of social class only partially confirm our hypotheses based on previous findings regarding income and education.

Model 2 in Table 2 provides estimates of social class, but now under control of the covariates used for weighting in the ECHP (see Table 1).[6] The covariates are measured in the wave prior to dropout. If these covariates are effective weighting variables, we would expect the effect of social class to diminish as compared to model 1.

Indeed, some odds ratios become insignificant. In Ireland and Portugal, after controlling for the ECHP weighting covariates, the dropout hazard of the routine nonmanual, self-employed and manual classes is no longer lower compared to the reference category of higher professional. In other cases, the odds ratios remain significant, or ratios that were insignificant under model 1 are now significant. The Wald Chi Squares of the effect of social class also show that it remains highly significant in all countries. All these findings suggest that weighting with the ECHP-weights may not always reduce dropout bias in estimates with the social class variable. Moreover, the adjusted R^2's are only slightly higher in the second models, suggesting that stronger predictors of attrition could be used for weighting.

Table 2 Discrete-time logistic hazard models, modelling the effect of social class$_{t-1}$ on dropout probability (Model 1), under control of ECHP weighting variables (Model 2).

		Odds Ratios			Model information	
	Ref. Higher professionals	**Model 1**	**Model 2**		**Model 1**	**Model 2**
Belgium	Lower professionals	1.07 ns	1.18 ns	*Pooled N*	*30245*	*26902*
	Routine nonmanual	1.29 **	1.30 *	*-2 Log L*	*19264.09*	*16941.52*
	Self-employed	1.93 ***	2.38 ***	*Adj.R²*	*0.0131*	*0.0394*
	Supervisory & skilled manual	1.32 **	1.19 ns	*Beta Wald X²*	*182.65*	*439.47*
	Semi-unskilled manual	1.09 ns	1.04 ns	*(df=6)*	***	***
	Unemployed	1.88 ***	5.49 ***			
Denmark	Lower professionals	1.29 **	1.22 *	*Pooled N*	*24914*	*24580*
	Routine nonmanual	1.94 ***	1.67 ***	*-2 Log L*	*17339.04*	*16758.67*
	Self-employed	2.71 ***	2.32 ***	*Adj.R²*	*0.0196*	*0.0349*
	Supervisory & skilled manual	1.66 ***	1.51 ***	*Beta Wald X²*	*230.87*	*117.42*
	Semi-unskilled manual	1.98 ***	1.65 ***	*(df=6)*	***	***
	Unemployed	2.40 ***	2.69 ***			
Germany	Lower professionals	0.90 ns	0.94 ns	*Pooled N*	*66404*	*64863*
	Routine nonmanual	1.14 ns	1.19 *	*-2 Log L*	*28559.62*	*26969.39*
	Self-employed	0.96 ns	0.87 ns	*Adj.R²*	*0.0104*	*0.031*
	Supervisory & skilled manual	0.77 ***	0.76 ns	*Beta Wald X²*	*241.42*	*448.00*
	Semi-unskilled manual	1.03 ns	1.08 ***	*(df=6)*	***	***
	Unemployed	1.57 ***	3.40 ***			
France	Lower professionals	0.85 **	0.82 **	*Pooled N*	*72214*	*65735*
	Routine nonmanual	1.05 ns	0.96 ns	*-2 Log L*	*43686.31*	*38010.95*
	Self-employed	1.09 ns	1.14 ns	*Adj.R²*	*0.0158*	*0.0362*
	Supervisory & skilled manual	0.97 ns	0.97 ns	*Beta Wald X²*	*523.23*	*1143.66*
	Semi-unskilled manual	1.06 ns	1.03 ns	*(df=6)*	***	***
	Unemployed	1.76 ***	3.52 ***			
Greece	Lower professionals	0.86 *	0.94 ns	*Pooled N*	*62535*	*61026*
	Routine nonmanual	0.71 ***	0.88 ns	*-2 Log L*	*35159.20*	*31810.61*
	Self-employed	0.49 ***	0.78 **	*Adj.R²*	*0.0269*	*0.1148*
	Supervisory & skilled manual	0.60 ***	0.78 **	*Beta Wald X²*	*708.10*	*1184.88*
	Semi-unskilled manual	0.69 ***	0.97 ns	*(df=6)*	***	***
	Unemployed	1.30 ***	3.81 ***			
Italy	Lower professionals	0.68 ***	0.69 ***	*Pooled N*	*94131*	*88767*
	Routine nonmanual	0.68 ***	0.71 ***	*-2 Log L*	*49008.10*	*44282.37*
	Self-employed	1.02 ns	1.32 ***	*Adj.R²*	*0.0205*	*0.0776*
	Supervisory & skilled manual	0.71 ***	0.72 ***	*Beta Wald X²*	*775.29*	*2065.23*
	Semi-unskilled manual	0.66 ***	0.75 ***	*(df=6)*	***	***
	Unemployed	1.52 ***	4.19 ***			
Ireland	Lower professionals	0.83 **	0.81 **	*Pooled N*	*41175*	*34515*
	Routine nonmanual	1.03 ns	1.28 ***	*-2 Log L*	*35212.51*	*28938.56*
	Self-employed	1.00 ns	1.44 ***	*Adj.R²*	*0.019*	*0.0694*
	Supervisory & skilled manual	0.74 ***	0.86 ns	*Beta Wald X²*	*451.54*	*1286.58*
	Semi-unskilled manual	0.75 ***	0.95 ns	*(df=6)*	***	***
	Unemployed	1.57 ***	5.56 ***			
Portugal	Lower professionals	1.10 ns	1.19 ns	*Pooled N*	*63274*	*60415*
	Routine nonmanual	0.73 ***	1.10 ns	*-2 Log L*	*28406.69*	*25277.71*
	Self-employed	0.85 *	1.84 ***	*Adj.R²*	*0.024*	*0.0984*
	Supervisory & skilled manual	0.59 ***	1.07 ns	*Beta Wald X²*	*576.43*	*1072.12*
	Semi-unskilled manual	0.65 ***	1.00 ns	*(df=6)*	***	***
	Unemployed	1.67 ***	6.01 ***			
Spain	Lower professionals	1.10 ns	1.16 *	*Pooled N*	*84001*	*82754*
	Routine nonmanual	1.22 ***	1.41 ***	*-2 Log L*	*56931.40*	*54265.12*
	Self-employed	1.18 **	1.90 ***	*Adj.R²*	*0.0233*	*0.0633*
	Supervisory & skilled manual	1.01 ns	1.25 ***	*Beta Wald X²*	*974.77*	*2355.12*
	Semi-unskilled manual	0.95 ns	1.22 **	*(df=6)*	***	***
	Unemployed	2.15 ***	5.67 ***			
The Netherlands	Lower professionals	1.12 *	1.10 ns	*Pooled N*	*50093*	*49774*
	Routine nonmanual	1.29 ***	1.29 ***	*-2 Log L*	*31513.66*	*30953.27*
	Self-employed	1.61 ***	1.90 ***	*Adj.R²*	*0.006*	*0.0207*
	Supervisory & skilled manual	0.92 ns	0.91 ns	*Beta Wald X²*	*154.04*	*283.32*
	Semi-unskilled manual	1.02 ns	1.03 ns	*(df=6)*	***	***
	Unemployed	1.51 ***	2.80 ***			

** p<0.05, ** p<0.01, *** p<0.001*

Source: own calculations on ECHP UDB - version December 2003 (Eurostat)

Towards a solution? The construction of a new longitudinal weight

The conclusion of the previous section states that weighting with the ECHP-weights does not always reduce the effects of selective dropout. This could be due to the omission of some influential predictors of dropout from the models used to construct the ECHP-weight. As discussed in the section about previous findings on selective attrition, it concerns in particular variables related to the interview process, the respondent's previous interview experience and education. In this section a new longitudinal weight is proposed that also includes these omitted variables.

The new longitudinal weight aims to correct for attrition of eligible sample persons.[7] Hence, it is constructed for participating sample persons from the second wave onwards. For every wave the construction of the new longitudinal weight requires a starting weight. In the second wave, this starting weight consists of the ECHP base weight of the first wave. This should enable us to weight also for design issues and to take into account the poststratification correction performed by Eurostat (Eurostat, 2003a). In all subsequent waves the starting weight consists of the obtained new longitudinal weight from the previous waves.

For all waves, the starting weights are corrected for dropout by multiplying these with the inverse probability of participating in the current wave, given that one is eligible for a personal interview and participated in the previous wave. These probabilities are estimated through logistic regression models that not only take into account the variables already used to estimate the ECHP-weight (see Table 1, column 3), but also variables that are thought to be additionally important in the dropout process (see Table 1, column 2).[8]

These are variables related to the interview process: duration of the personal interview in the previous wave and duration squared, mode of interview in the previous wave (PAPI or not), number of visits to the household in the previous wave and whether or not the interviewer has changed since the previous wave. Also, some variables related with previous wave response were included, namely item-nonresponse on the income variable in the previous wave, and whether the person participated in the first wave. Other predictors of dropout included are frequency of contact with neighbours (previous wave), length of stay at the current address (previous wave), education (previous wave) and labour force status (previous wave). Where covariates show a substantial number of missings, a category indicating a missing answer is included in the estimations.[9] For persons who did not participate in the previous wave, or in the waves where the number of missings on a covariate was too small to be modelled, the ECHP longitudinal weight is imputed.[10] In order to avoid extreme weights, the estimated probabilities are trimmed at one percent before their inverse is multiplied with the starting weight. Weights are rescaled so that their mean over the population of participating sample persons in a given wave equals one.

Two illustrations of the effect of weighting

Comparison of two ECHP subsamples

The purpose of this section is to provide a concrete example of how attrition may influence estimates of the distribution of the population over the social class variable. In order to enable comparison, the sample is divided into two partially overlapping subsamples: one comprising all sample persons participating in the first wave, another comprising only sample persons participating in all eight subsequent waves.[11] The latter subsample, which will be called the longitudinal sample, is affected by dropout. The aim is to compare the relative frequency distribution of social class in the first wave between both subsamples. The findings are presented in Table 3 for Denmark and Greece, because these are the countries for which we found highly significant effects of all social classes on dropout in the discrete-time logistic regressions of the previous section. Also, Denmark represents the Northern countries, while Greece is representative of the Southern rim, where the effect of social class on dropout tends to go in the opposite direction.

Table 3 Comparison of four percentage distributions of social class for Denmark and Greece

	Respon-dents Wave 1	Respondents to all waves			Ratio of respondents to all waves relative to respondents to wave 1		
	A.1.	**B.1.**	**C.1.**	**D.1.**	**B.2.**	**C.2.**	**D.2.**
Social class	**Wave 1 base weight Eurostat**	**Wave 1 base weight Eurostat**	**Longi-tudinal weight Eurostat**	**Longi-tudinal weight New**	**= B.1./A.1. *100**	**= C.1./A.1. *100**	**= D.1./A.1. *100**
			Denmark				
Unemployed or inactive	29.34	23.42	22.81	27.00	79.82	77.74	92.02
	[28.3-30.42]	[22-24.9]	[21.06-24.67]	[25.32-28.75]			
Higher professionals	16.44	20.67	19.08	18.15	125.73	116.06	110.40
	[15.61-17.31]	[19.34-22.06]	[17.55-20.72]	[16.88-19.49]			
Lower professionals	15.78	18.3	16.55	16.91	115.97	104.88	107.16
	[14.97-16.63]	[17.03-19.63]	[15.12-18.1]	[15.65-18.26]			
Routine non-manual	14.82	14.73	15.52	14.96	99.39	104.72	100.94
	[14-15.68]	[13.55-15.99]	[13.61-17.65]	[13.69-16.33]			
Self-employed	3.8	3.63	4.77	3.85	95.53	125.53	101.32
	[3.4-4.26]	[3.06-4.31]	[3.78-6.01]	[3.21-4.61]			
Supervisory & skilled manual	9.39	9.48	9.82	9.13	100.96	104.58	97.23
	[8.74-10.08]	[8.55-10.5]	[8.68-11.1]	[8.19-10.17]			
Semi-unskilled manual	10.42	9.77	11.43	10	93.76	109.69	95.97
	[9.71-11.17]	[8.78-10.86]	[10-13.03]	[8.92-11.19]			
N	*5472*		*2537*				
Nmiss	*38*		*19*				
			Greece				
Unemployed or inactive	31.08	27.63	28.21	29.25	88.90	90.77	94.11
	[30.34-31.84]	[26.68-28.61]	[27.17-29.27]	[28.15-30.38]			
Higher professionals	7.23	5.77	7.47	7.09	79.81	103.32	98.06
	[6.81-7.67]	[5.28-6.32]	[6.78-8.22]	[6.4-7.84]			
Lower professionals	6.77	5.63	6.13	6.27	83.16	90.55	92.61
	[6.36-7.21]	[5.14-6.17]	[5.56-6.74]	[5.7-6.9]			
Routine non-manual	10.77	10.08	10.16	11.32	93.59	94.34	105.11
	[10.26-11.31]	[9.42-10.78]	[9.46-10.9]	[10.43-12.26]			
Self-employed	27.64	33.26	30.14	28.73	120.33	109.04	103.94
	[26.9-28.39]	[32.22-34.31]	[29.1-31.2]	[27.71-29.78]			
Supervisory & skilled manual	10.18	11.27	11.9	11.43	110.71	116.90	112.28
	[9.69-10.69]	[10.59-11.99]	[11.12-12.73]	[10.62-12.3]			
Semi-unskilled manual	6.34	6.36	5.99	5.9	100.32	94.48	93.06
	[5.94-6.76]	[5.83-6.93]	[5.47-6.56]	[5.35-6.52]			
N	*11757*		*6089*				
Nmiss	*120*		*53*				

Source: own calculations on ECHP-UDB - version December 2003 (Eurostat).
Note: Numbers between square brackets indicate 90% confidence intervals

When comparing the percentages in column B.1. with those in column A.1. in Table 3, it is possible to see how dropout influences the estimation of the size of the different social classes. The under- or overrepresentation of a certain social class in the group of respondents to all waves relative to the group of wave one respondents, is also summarised in column B.2. There, a number below 100 indicates underrepresentation of the social class in the longitudinal sample, while a number higher than 100 points to overrepresentation. So, the closer to 100, the better the estimate of the longitudinal sample reflects one without dropout. Note that the percentages in columns A.1. and B.1. have been weighted with the Eurostat wave one base weight in order to account for design and initial nonresponse effects.

In the columns C.1. and D.1. of each table, the longitudinally weighted relative frequency distribution of social class is presented for the longitudinal subsample. The percentages under C.1. are weighted with the Eurostat longitudinal weight of wave eight, while the percentages under D.1. are weighted with the newly constructed longitudinal weight of wave eight (see previous section). The relative over- or underrepresentation of certain social classes can again be read from the ratios in columns C.2. and D.2.

The tables show that for a majority of the classes, dropout does not seem to affect the percentage estimations a lot. This is true especially for the lower social classes, i.e. for the semi-unskilled manual, supervisory and skilled manual and routine nonmanual. In these cases, weighting would not be necessary, but doing so may sometimes worsen the estimates.

More interestingly, in both countries, attrition seems to influence the estimate of the percentage in the unemployed/inactive category. Apparently, the unemployed/inactive are underrepresented in the longitudinal sample because they drop out more frequently. In Greece the relative underrepresentation is around 11%, while in Denmark it amounts to almost 20%. What is the effect of weighting on the biases in the longitudinal estimates? Weighting with the longitudinal Eurostat-weight slightly deteriorates the estimate of the unemployed/inactive in Denmark and slightly improves it in Greece. In contrast, the new longitudinal weight reduces the underrepresentation of the unemployed/inactive substantially in both countries. The confidence intervals confirm that the estimates based on the new longitudinal weight come statistically close to the estimates based on all wave one respondents.

The classes of higher and lower professionals form another example where percentages based on the wave one respondents and those based on respondents to all waves differ to some degree. Yet, for these classes the effect of dropout is different across countries. Because in Denmark, higher and lower professionals tend to drop out less, they are overrepresented in the longitudinal sample with respectively 26% and 16%. In contrast, in Greece -the representative of the Southern rim- higher and lower professionals appear to drop out more frequently, causing a downward bias of their share in the longitudinal sample. Irrespective of the direction of the bias, weighting both with the Eurostat and the new weight tends to reduce the bias. For the higher professional class the new longitudinal weight reduces a little more bias than the Eurostat-weight in both countries. For the lower professional class, the new longitudinal weight reduces somewaht more bias in Greece, but for Denmark the bias reduction of this new weight is slightly smaller than that of the Eurostat weight. Overall however, differences are modest.

For the class of the self-employed, there only seems to be a significant dropout problem in Greece. There, the self-employed are more likely to continue their participation in the panel and are therefore overrepresented in the longitudinal sample (with 20% compared to wave one respondents). This overrepresentation is reduced to 9% when weighting with the Eurostat longitudinal weight and it shrinks further to less than 4% when weighting with the new longitudinal weight. The information on confidence intervals confirms this trend.

Comparison with external population data

The aim of this section is to investigate how cross-sectional descriptive statistics based on the ECHP are nonresponse biased and how well the cross-sectional weight corrects for this bias. To accomplish this, ECHP data on education are compared with an external data source containing reliable population data.

Dropout or attrition does not only affect the longitudinal representativeness of the initial sample. Also cross-sectional analyses with household panel data are likely to be biased due to selective dropout patterns. The longer a panel study runs, the less representative the sample becomes of the respective cross-sectional populations. Not only selective attrition leads to misrepresentation of certain categories in the population. Also, fresh persons in the study like children or new household members of initial sample persons, are added on a selective basis. This is because only the non-random group of initial sample persons who are still in the study will bring in new household members in the household panel.[12] To increase the cross-sectional representativeness of the panel Eurostat provides a cross-sectional weight.[13]

In what follows we will compare weighted and unweighted yearly cross-sectional distributions of education level based on ECHP estimates with administrative education data in the Danish population. The Danish case has been chosen for this comparison because it is the only country for which administrative education level data could be found that are compatible with the ECHP education data.

Table 4 Comparison of percentage distribution of education in the ECHP sample with administrative data, Denmark

Higest attained education level		**Administrative data**	**ECHP (unweighted)**		**ECHP (cross-sectionnally weighted)**	
		%	**%**	**CI**	**%**	**CI**
1998	Low	38.5	27.94	[26.82-29.1]	29.95	[28.55-31.39]
	Middle	42.8	48.71	[47.44-49.98]	48.89	[47.33-50.44]
	High	18.7	23.35	[22.29-24.44]	21.16	[19.96-22.42]
	N	*3678779*		*4180*		*4180*
	X² test		*0.912*	*	*0.787*	
1999	Low	37.4	26.52	[25.38-27.69]	27.8	[26.37-29.28]
	Middle	43.2	49.15	[47.84-50.45]	49.73	[48.09-51.37]
	High	19.3	24.33	[23.23-25.47]	22.47	[21.17-23.83]
	N	*3685592*		*3982*		*3982*
	X² test		*0.929*	*	*0.863*	
2000	Low	36.7	25.45	[24.31-26.63]	27.49	[25.98-29.05]
	Middle	43.4	49.10	[47.77-50.43]	49.6	[48.87-51.34]
	High	19.9	25.45	[24.31-26.63]	22.91	[21.57-24.3]
	N	*3679976*		*3831*		*3831*
	X² test		*0.943*	*	*0.839*	
2001	Low	35.9	24.17	[23.05-25.34]	26.9	[25.25-28.62]
	Middle	43.6	49.78	[48.44-51.11]	49.36	[47.47-51.25]
	High	20.5	26.05	[24.89-27.24]	23.74	[22.21-25.35]
	N	*3679976*		*3785*		*3785*
	X² test		*0.955*	**	*0.829*	
2001/1998	Low	0.93	0.87		0.90	
	Middle	1.02	1.02		1.01	
	High	1.10	1.12		1.12	

Sources: Statistics Denmark - statbank.dk / Own calculations based on ECHP-UDB (2003)
** p<0.10, ** p<0.05*
Note: Numbers between square brackets indicate 90% confidence intervals

In Table 4 the yearly cross-sectional distributions of attained education level are given from 1998 to 2001. These waves of the ECHP have been chosen for comparability reasons.[14] The unweighted distributions based on the ECHP clearly underrepresent the lowest education level in all years between 1998 and 2001, while the highest education level is overrepresented. The distribution based on cross-sectionally weighted ECHP data still underrepresents the lowest education group and overrepresents the highest education level, though to a smaller degree. The X^2 tests show that the distribution of education level in the ECHP data differs from the distribution in the administrative data at the 0.1 significance level. The difference between the weighted ECHP data and the administrative data is insignificant.

The biased representation of education level in the ECHP dataset is probably influenced by both non-random initial response and selective attrition. How can the effect of dropout on the estimation of education be separated out from the effect of initial nonresponse? A way to proceed is to investigate whether misrepresentation of the education variable becomes larger between 1998 and 2001. In the bottom rows of Table 4, the ratio of the proportions in each educational group is given for 2001 over 1998. As an example, we discuss the ratios for the lowest educated. Their 2001/1998 ratio based on the administrative data is 0.93; indicating that the share of the lowest education level has decreased between 1998 and 2001. The 2001/1998 ratio based on the unweighted ECHP data (0.87) shows a larger decrease in the share of the lowest education level than is observed in the population data. Hence, social researchers using unweighted ECHP data might overestimate the decrease in the share of the lowest education group. It also shows that this overestimation must be caused by dropout bias and not by initial nonresponse bias. The 2001/1998 ratio of the weighted data (0.90) is in between that of the unweighted data and the population data. This would suggest that weighting corrects the dropout bias to a certain degree. However, the differences between the ratios are small and no hard conclusions should be drawn.

Conclusion

As weighting is an often applied strategy to reduce bias resulting from non-random attrition, this article has examined its effectiveness in the European Community Household Panel. Both the selectiveness of attrition and the bias reduction obtained by weighting have been assessed using three methods: 1. discrete-time hazard models, 2. a comparison of two subsamples in the ECHP: the sample of respondents to the first wave and the smaller subsample of 'survivors' who participated in all ECHP waves, and 3. a comparison with external population data from Statistics Denmark.

The main outcome is that there is considerable bias in estimates of social class and education resulting from selective attrition. The direction of this bias was different across countries, largely following a North-South divide. In particular, lower social classes drop out more quickly in Northern Europe, while a reversed effect is found in Southern Europe and Ireland. Furthermore, the unemployed/inactive have a higher attrition rate in all countries.

Weighting with the ECHP-UDB provided weight in most cases reduces this bias, but often only slightly and in other cases it deteriorates the estimate. However, this weight is mainly based on socio-demographic characteristics. Yet, an overview of the literature suggests that also survey-related variables are important predictors of dropout. Therefore, a new longitudinal weight was constructed including these omitted variables.

Generally speaking, the bias reduction of the new longitudinal weight was a bit higher than that obtained by the original ECHP-weight.

We conclude by discussing the implications of these findings for applied social researchers. First of all, could dropout be ignored? While some researchers find no substantial dropout bias for income-related variables in the ECHP (Behr *et al.*, 2003; Watson, 2003), our findings suggest that there is a problem for social class and education. Secondly, is weighting an appropriate strategy to cope with selective attrition? Although our findings suggest that weighting may occasionally worsen dropout bias, we also find that the effectiveness of weights improve if they are constructed on the basis of better predictors for dropout. Good candidates for inclusion are survey-related variables. However, constructing new weights is not always practicable and may be time-consuming for a multi-country panel like the ECHP. Before entering into this endeavour, it might be more convenient to first apply one of the methods presented in this article for checking the effectiveness of available weights.

Acknowledgments

Both authors contributed equally to this article. The research by Leen Vandecasteele was funded on the basis of research project G0398.02 of the Research Foundation-Flanders. The work of Annelies Debels on this research was possible thanks to an ‘Aspirant’ scholarship of the Research Foundation-Flanders. Special thanks also go to Prof. dr. J. Berghman, dr. Ludi Simpson, Prof. dr. J. Billiet, Bart Meuleman and three anonymous reviewers for their useful comments.

Notes

[1] We conducted a search on ISI Web of Science with keywords 'European Community Household Panel' or 'ECHP'.

[2] For Germany the GSOEP-dataset is used, as Eurostat integrated this national panel into the ECHP.

[3] For a description of sample and non-sample persons in the ECHP, we refer to Peracchi (2000).

[4] Thus, the starting weight is multiplied by the probability for an individual being resident (=a member of an interviewed household in wave i), if it was resident in the previous wave and divided by the probability for an individual for having been resident in the last wave i-1, if it is resident in the current wave. The probability of being interviewed when eligible, is also calculated. The probabilities were modelled in PROC CATMOD of the SAS system and trimmed to avoid extreme weights (Eurostat, 2003a). Unfortunately, it was impossible to obtain more estimation details from Eurostat.

[5] Calibration by household size only took place in seven countries: Germany, Ireland, Luxembourg, the Netherlands, Austria, Finland and Spain.

[6] However, the covariate 'arrivals to or departures from the household' is nowhere included since it overlaps with the variable split-off household. In the Netherlands, there is no information available concerning split-off status, so the effect could not be estimated. This also applies to the effect of region in the Netherlands and in Denmark.

[7] Eligible for the personal interview are these persons who live in a private household in the EU and have reached the age of 16.

[8] Due to space constraints, estimation details from these logistic regressions are not presented, but they can be obtained from the authors upon request.

[9] Further in this article, we will focus on the functioning of the weights in Denmark and Greece. Because of a substantial number of missings, a category indicating a missing answer was

included in the estimations for Denmark for the variable length of stay at current address and in Greece for contact with neighbours. In Greece, interviewer change was not included because of a too high percentage of missings that could not be modelled.

[10] For the countries further elaborated in this article, the percentage of imputed cases in the different waves lies between 6,96% and 9,3% in Greece and between 6,68% and 13,19% in Denmark.

[11] Persons who either die, move to an institution or move outside the EU during the panel study, are dropped from the subsamples. As such, we represent the population under scope of the ECHP.

[12] To be complete, we should mention that not only selective drop-out and drop-in patterns lead to a worsening of the cross-sectional representation of household panel data over time. Also, changes in the population structure caused by immigration flows are not accounted for if no new households are added to represent the new immigrant groups. In this case, the household sample would not perfectly reflect the population structure in later time points, even if there were complete participation. For this exercise, changes in the population structure due to immigration are assumed to be negligible because of the short time-span under study.

[13] To improve the cross-sectional representation of a household panel, one could also opt to extend the initial sample with selectively sampled fresh households. In the case of the ECHP, no fresh sample has been added to the initial sample.

[14] Attained education level in the ECHP dataset has been measured in 1994 and then in all years between 1998 and 2001, but the categorisation of the 1994-measurement deviates from the one of the other years.

References

Allison, P. D. (2002). *Missing data*, Thousand Oaks, CA: Sage.

Behr, A., Bellgardt, E. and Rendtel, U. (2003). *Comparing poverty, income inequality and mobility under panel attrition. A cross country comparison based on the European Community Household Panel. CHINTEX Working Paper no. 12*, URL: http://www.destatis.de/chintex/download/paper12.pdf

Behr, A., Bellgardt, E. and Rendtel, U. (2005). Extent and Determinants of Panel Attrition in the European Community Household Panel. *European Sociological Review*, **21**, 489-512.

Buck, N. (2003). *BHPS User Documentation,* Institute for Social and Economic Research, URL: http://iserwww.essex.ac.uk/ulsc/bhps/doc/index.html.

Covello, V. T. and Bollen, K. A. (1979). Status Consistency in Comparative Perspective: An Examination of Educational, Occupational, and Income Data in Nine Societies. *Social Forces*, **58**, 528-39.

Diggle, P. J., Heagerty, P., Liang, K. and Zeger, S. L. (2002). *Analysis of Longitudinal Data,* Cornwall: T.J. International Ltd.

Diggle, P. J. and Kenward, M. G. (1994). Informative Drop-Out in Longitudinal Data Analysis. *Applied Statistics*, **43**, 49-93.

Engel, U. and Reinecke, J. (1996). Introduction. In Engel, U. and Reinecke, J. (Eds), *Analysis of change. Advanced techniques in panel data analysis.* New York: Walter de Gruyter, pp. 1-14.

Eurostat (2003a). *Construction of weights in the ECHP. Doc PAN 165/2003-06*, Luxembourg: Eurostat.

Eurostat (2003b). *ECHP UDB Manual. European Community Household Panel Longitudinal Users' Database. Waves 1 to 8, Survey years 1994 to 2001. Doc PAN 168/2003-06*, Luxembourg: Eurostat.

Fitzgerald, J., Gottschalk, P. and Moffit, R. (1998). An analysis of sample attrition in panel data. The Michigan Panel Study of Income Dynamics. *The Journal of Human Resources,* **33**, 251-99.

Graham, J. W. and Hofer, S. M. (2000). Multiple Imputation in Multivariate Research. In Little, T. D., Schnabel, K. U. and Baumert, J. (Eds), *Modeling Longitudinal and Multilevel Data*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers, pp. 201-18.

Haisken-DeNew, J. P. and Frick, J. R. (2003). *Desktop Companion to the German Socio-Economic Panel Study (SOEP),* Berlin: DIW.

Kalton, G. and Brick, M. (2000). Weighting in household panel surveys. In David, R. (Ed), *Researching social and economic change. The uses of household panel studies*. London/New York: Routledge, pp. 96-112.

Kalton, G. and Kasprzyk, D. (1986). The treatment of missing survey data. *Survey methodology*, **12**, 1-16.

Lepkowski, J. M. and Couper, M. P. (2002). Non-response in the second wave of longitudinal household surveys. In Groves, R. M., Dillman, D. A., Eltinge, J. L. and Little, R. J. A. (Eds), *Survey Nonresponse*. New York: John Wiley & Sons, pp. 259-72.

Lillard, L. A. and Panis, C. W. A. (1998). Panel attriton from the Panel Study of Income Dynamics. Household income, marital status, and mortality. *The Journal of Human Resources,* **33**, 437-57.

Little, J. A. (1995). Modeling the Drop-Out Mechanism in Repeated-Measures Studies. *Journal of the American Statistical Association,* **90**, 1112-121.

Little, J. A. and Rubin, D. B. (1987). *Statistical analysis with missing data*, New York: John Wiley & Sons.

Little, J. A. and Rubin, D. B. (1989). The Analysis of Social Science Data with Missing Values. *Sociological Methods and Research*, **18**, 292-326.

Nicoletti, C. and Peracchi, F. (2002). *A cross-country comparison of survey participation in the ECHP*. *ISER Working Papers, 2002-32,* Colchester: University of Essex.

Nicoletti, C. and Peracchi, F. (2005). Survey response and survey characteristics: microlevel evidence from the European Community Household Panel. *Journal of the Royal Statistical Society A,* **168**, 763-81.

Peracchi, F. (2000). The European Community Household Panel: A review. *Empirical Economics*, **27**, 63-90.

Raab, G., Purdon, S. and Buckner, K. (2005). *Practical Exemplars and Survey Analysis,* URL: http://www.napier.ac.uk/depts/fhls/peas/

Rose, D. (2000). Household panel studies: an overview. In Rose, D. (Ed), *Researching social and economic change. The uses of household panel studies*. London: Routledge, pp. 3-35.

Schooler, C. and Schoenbach, C. (1994). Social Class, Occupational Status, Occupational Self-Direction and Job Income: A Cross-National Examination. *Sociological Forum,* **9**, 431-58.

Teichler, U. (2001). Education and employment. In Smelser, N. J. and Baltes, P. B. (Eds), *International Encyclopedia of the Social and Behavioral Sciences*. Oxford: Elsevier Science Ltd, pp. 4178-182.

Verbeek, M. (2004). *A guide to modern econometrics*, Chichester: John Wiley and Sons, Ltd.

Watson, D. (2003). Sample attrition between waves 1 and 5 in the ECHP. *European Sociological Review*, **19**, 361-78.

Zabel, J. E. (1998). An analysis of attrition in the panel study of income dynamics and the survey of income and program participation with an application to a model of labor market behavior. *The Journal of Human Resources,* **33**, 479-506.

Authors' Addresses

Leen Vandecasteele (corresponding author), Katholieke Universiteit Leuven, Centre for Sociological Research (CESO), E. Van Evenstraat 2B, 3000 Leuven, Belgium. Tel: + 32-16-323176. Fax: +32-16-323365. E-mail: leen.vandecasteele@soc.kuleuven.be

Annelies Debels, Katholieke Universiteit Leuven, Centre for Sociological Research (CESO), E. Van Evenstraat 2B, 3000 Leuven, Belgium. Tel: + 32-16-323176. Fax: +32-16-323365. E-mail: annelies.debels@soc.kuleuven.be

Biographical notes

Leen Vandecasteele is a PhD student at the Centre for Sociological Research of the KULeuven (Belgium). Her PhD study investigates the impact of social stratification determinants on poverty and income dynamics in European welfare regimes. Other research interests include panel data research, individualisation and values studies.

Annelies Debels is a PhD student at the Centre for Sociological Research of the KULeuven (Belgium). She is currently working on a PhD about the effects of labour market flexibility on income and poverty dynamics. Other research interests include panel data research, social security and welfare states.