The Determinants of Panel Attrition in Telephone Survey:

An Experience from Taiwan's Election Study (draft)

Kah-yew Lim

Ph. D. Student at Department of Political Science & Research Assistant in Election
Study Center, National Chengchi University, Taiwan.

,

The Determinants of Panel Attrition in Telephone Survey:

An Experience from Taiwan's Election Study (draft)

Kah-yew Lim

Ph. D. Student at Department of Political Science & Research Assistant at Election
Study Center, National Chengchi University, Taiwan.

,

Abstract

In this paper, we investigate the effect from interaction between interviewer and respondent on the decision to refuse or to accept the interview in the second wave survey. We explore a two-stage multilevel model to analysis the interviewer characteristics and respondent characteristics in the same model, and include the possible interaction terms. We were able to expand our information on respondent from the first wave survey.

The data which used in this paper consists of two major parts. First, there are fifteen waves rolling survey have conducted in pre-election and a panel survey conducted after the presidential election. We collected data on 9818 individual in the first wave. In this research, the dependent variable is the result of the interview, a variable with three categories (cooperate, refuse and lost contact). The independent variables include respondent's characteristics (gender, age, education level) and interviewer characteristic (gender, age, level, job, etc.). We also include some control variables in the model include the length of interview in first wave and the time distance between first and second wave.

In conclusion, the panel design variables have significant influence on respondent's decision to participate the second wave interview. On the other hands, the characteristic of respondents, include their gender, age, and education level have same influence on the result.

Keywords: panel attrition, multilevel analysis, telephone survey

1. Introduction

There are many puzzles of politics and political behavior which cannot answer with cross-sectional data analysis. Indeed, for decade's political scientist have devoted vast resources to collect panel data. Panel studies, in which the same respondents are interviewed repeatedly at different points of time, are widely used in many areas in the U.S and European Countries. One of the key concerns on data quality in panel surveys is panel attrition. Even if a panel starts out with a good representation of the population of interest, attrition problem can quickly accumulate and render the sample unrepresentative.

The scholars have developed a conceptual framework where the influencing factors of result in surveys are separated in four groups (Groves and Cooper, 1998). These four broad groups include the social environment, the sampling unit, the sampling design, the interviewer, and the last two factors can be controlled by the researchers. Whether a respondent cooperate or refuses depends largely on the interaction between the respondent and the interviewer. In panel study, additional complication arises when respondents are being surveyed in few waves.

Therefore, by using the available panel data sets from telephone survey in Taiwan, this paper attempt to explore the causes of sample attrition in a this panel study.

2. Theoretical

Response rates to social and behavioral surveys have been declining for several decades(Groves and Cooper, 1998), increasing the likelihood that differences between respondents and non-respondents may be sufficient to bias survey estimates (Caetano, 2001). There are several methodologies that have been employed to investigate the potential effects of nonresponse on survey estimates. In those methodologies issue, panel attrition studies compare the baseline characteristics of respondents who are and who are not lost to follow-up waves of interview. Attrition also can be conceptualized as the combination of three cumulative factors: mortality, migration between waves, and survey non-response (Groves and Cooper, 1998).

Recent years have seen a surge in the challenges posed by panel survey (Ashenfelter et al. 1986). It has become increasingly difficult to maintain response rates, causing the costs of data collection to rise considerably. A comprehensive analysis of panel attrition in a household panel was presented for the Panel Study of Income Dynamics (PSID), a panel survey running since 1968. Fitzgerald et al. (1998) give evidence of non-random attrition, but the potential bias in the empirical analysis is considered low.

2.1 Interviewer

As a basic theoretical review of survey participation, Couper and Groves (1992) point out the importance of the interviewer-respondent interaction. They argue that much depends on interviewer's ability to persuade potential respondent. Interviewers might influence respondents' willingness to participate in surveys through their personal performance patterns when contacting household.

Empirical evidence shows that some interviewer characteristics have an effect on individual response rates. Generally, socio demographic variables, personal traits and socio skills of interviewer have been explored, either alone or in-combination with respondent characteristics. A number of studies have investigated the role of different interviewer characteristics on both individual response rates and the quality of interviewed data obtained.

2.2 Respondent

Evidence from panel attrition studies confirmed the differences of non-responders and responders in terms of socio-demographic and economic variables. With regards to socio-demographic characteristics of respondent, the findings relating to respondent characteristics are somewhat inconsistent. Several researchers have found age to be correlated with nonresponse, but, the impact of other respondent characteristic (such as race, gender, education, socio-economic status, etc.) is mixed. Analysis of the PSID suggest that the potential for attrition is higher amongst young adults, most likely a result of the increased mobility required of important transitions associated with young adulthood, such as schooling decision, family formation and employment search. Additionally, Zabel compared the labour-market behavior of attritors and non-attritors in both the PSID and the Survey of Income and Program Participation (SIPP). For both the PSID and the SIPP, he found that individuals who left the survey were more likely to reside in urban locations, live in South or West, be non-white, be unmarried, not own homes, and have fewer children than the individuals who remained in the survey (Zabel, 1998).

2.3 Panel Design

Panel attrition is also affected by factors in the design of the survey itself; interview length, consistency of interviewer across panel waves, and the frequency of waves can all affect the completion and retention rates. In some research, shorter interview length, interviewer consistency, and less frequent waves are all associated with higher completion rate. Interview length has a positive impact on attrition, likely due to the fatigue and loss of interest resulting from repeated long interviews (Zabel,

1998). As Dillman (1978) argue, why people respond to a social survey is a social exchange between survey interviewer and respondent. Dillman posits that the respondent participate because the act of participation is expected to bring rewards that exceed the cost of participation. Interview who takes a lot of time will increased the time cost of respondent. In addition, an interview which takes in the time when the respondent is not available, they will more intend to lost contact or refuse the interview.

2.4 Survey Environment

Household size has been found to have positive impact on response rates. Relative high refusal rates are often seen with single person households. For the area characteristics, several socio-environmental determinants for survey participation have been identified in the survey literature. Most of the factors correlating with survey participation are urbanicity, population density and crime rates. Furthermore, some scholar found that migration is linked to village level characteristics; in particular, the stability and resilience of social networks increase the likelihood of re-interview. For example, migrants from villages that were linguistically homogeneous were more likely to be found and re-interviewed, probably due to the greater interconnectedness experienced by linguistically homogeneous villages. Additionally, researchers found inverse relationship between the likelihood of survey response and income. In developed countries, the high opportunity cost of time amongst relatively affluent people may make many potential respondents reluctant to participate in a study (Jimenez-Martin and Peracchi, 2002). Contrastingly, migration-related attrition is less likely to be a serious issue, as communications and information networks in developed countries enable researcher to locate most respondents who have changed residence between waves. In developing countries, generally have the reverse problem (Defo, 1992). Experience has indicated that non-response is very low in developing countries, both in cross sectional and panel surveys. However, the issue of migration-related attrition is more problematic. In developed countries, survey respondents are often a phone call away, tracking movers in developing countries may require considerable ingenuity, time, and resources on the part of both researchers and fieldworkers.

3. Methodology and Data

3.1 Variables in the Analyses

3.1.1 Dependent Variables(Lost Contact, Refusal)

In this research, the dependent variable is the result of the interview. We segment the result into the following three subgroups in terms of their panel participation status: (1) co-operate sample, (2) refusal sample, and (3) lost contact sample. Refusal and noncontacts generally make up the two most important components of unit nonresponse. It is important to separate noncontacts from refusals when examining survey participation, a simultaneously analysis of both components seems relevant when assessing interviewer effect. In this paper, noncontact and nonresponse will be analyzed separately as figure 1 displayed.
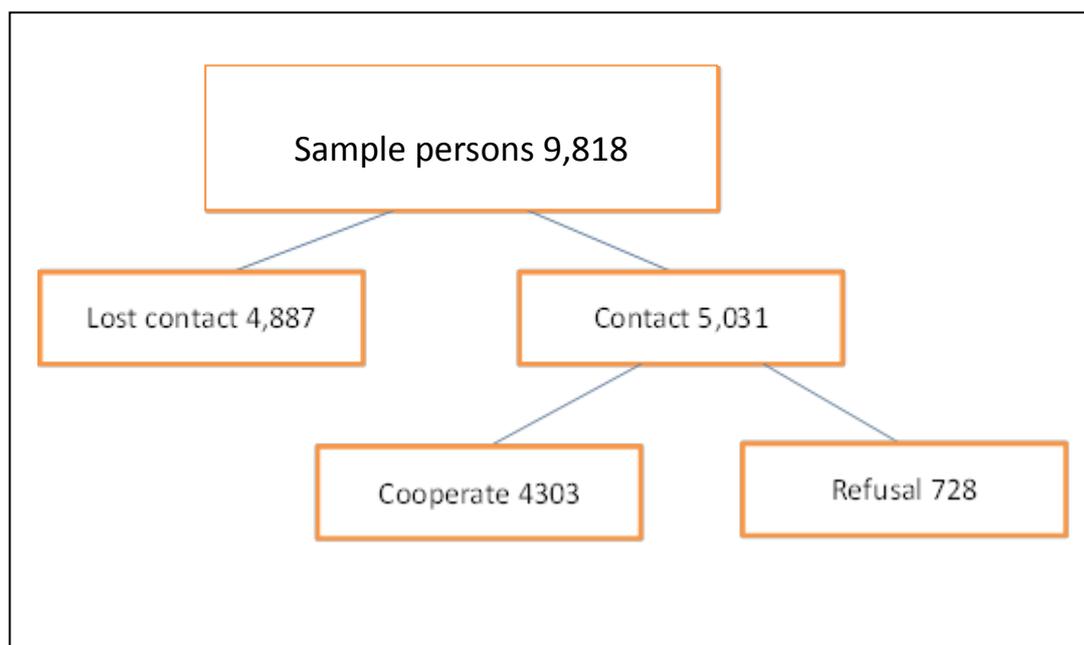

Figure 1 Subgroups of sample person

In terms of the theoretical framework for contacting household and survey cooperation, there are four main categories influencing whether interviewers are able to get re-contact or get cooperation:

3.1.2 Interviewer (Gender, Experience)

During the panel survey, the interview was conducted by 73 interviewers. Of those interviewers 54 were female (74 %) and 19 male (26 %). The interviewer average experience was 1609 hours. On average, each interviewer conducted 134 interviews.

3.1.3 Respondent (Gender, Age, Education Level)

In the panel survey, 9818 respondents were re-contacted. Of those respondent 5208 were female (53%) and 4610 male (47 %). The respondent average age was 46 years old. Education level of respondent has included into the analysis model, about 10% respondent is illiteracy and only primary school level education, 60% of them take secondary school, and 30% of them are undergraduate and above .

3.1.4 Survey Environment (Household size, Urbanization)

The average household size of respondent in the data was 3 people in their house. For theoretical reason, household size will be a dichotomous variable (1=1 person, 0= more than 1 persons) in the model. Besides, urbanization level of respondent residential area will be included. Urbanization level was measured by village, town and city.

3.1.5 Panel Design (Days Distance between Wave, Life Timetable Effect, Length of Last Interview)

In this category, three main variables include days distance, life timetable effect and length of 1st time interview. In terms of the theoretical framework for contacting household, days distance and life timetable were included. The average days distance were 48 days between 1st time and 2nd time interview. Life timetable was measure in minutes, and in average of 120 minutes. The average length of interview in 1st time interview was 11 minutes.

These factors have an impact on the result of survey, including whether successfully to contact sample person or the decision of respondent to cooperate or refuse to participate.

3.2 Using of Multilevel Model

Due to respondents are nested within interviewers, appropriates statistical model to use should consider the hierarchical structures as Figure 2 for better estimation. A series of multilevel models were conducted to estimate the potential associations between the interviewer and respondent. The key feature of this method is that the independent variables which included in the model can be defined on different levels of data structure. For the dependent variables in this paper which are dichotomous variables, we need to specify a non-linear analysis appropriate for binary outcomes. While the basic principle of standard multilevel model is a normal sampling model and identity link function, the binary outcome model uses a binomial sampling model and a logit link (Raudenbush and Bryk, 2002). This allows examining a multilevel

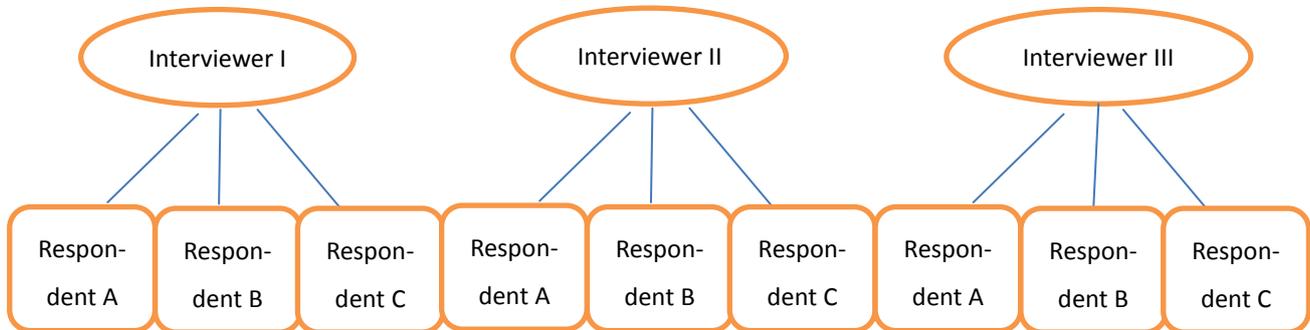model for dependent variable with binary outcomes.

Figure 2 Hierarchical Structures between Interviewer and Respondent

3.3 Data

The data which used in this paper consists of two major parts. First, there are fifteen waves rolling survey have conducted in pre-election and a panel survey conducted after the presidential election. We collected data on 9818 individual in the first wave. Due to the panel survey was conducted by two different organizations with systematic separate on sample; the analysis result will be separate into two parts. The panel survey for the sample from 1st to 10th waves rolling survey was conducted by Taiwan Real Survey Co., Ltd. The panel survey for the sample from 11st to 15th waves rolling survey was conducted by Election Study Center in National Chengchi University, Taiwan.

Data analyzed in this article were from Taiwan's Election and Democratization Studies, 2012: Presidential and Legislative Election (TEDS 2012) (NSC 100-2420-H-002-030). The coordinator of multi-year project TEDS is Professor Chi Huang (National Chengchi University). TEDS2012 is a yearly project on the president and legislator election in 2005. The principal investigator is Professor Yun Han Chu. More information is on TEDS website (http://www.tedsnet.org). The author(s) appreciate the assistance in providing data by the institute and individual(s) aforementioned. The author(s) are alone responsible for views expressed herein.

4 Result

In table 1, four models have been presented. Model A1 and A2 are the models conducted the multilevel analysis on panel survey which the baseline sample from 1st wave to 10th wave rolling survey. While the Model B1 and B2 are the models

conducted the multilevel analysis on panel survey which the baseline sample from 11st wave to 15th wave rolling survey. The main different between model A and model B is the date conduct the baseline survey. For these four models presented in table 1, two equations are presented, both equations including the respondent level predictors and interviewer level predictors, but the equations for model A2 and model B2 will included the interaction variables.

In Model A1 and A2, five factors were found to be associated with survey contact and noncontact: day distance with lastime survey, respondent life time table effect, household size, respondent age, and respondent gender. As the first column of table 1 indicates, when the day distance between 1st time survey and 2nd time survey is shorter, the odds to get contact with respondent is higher. If the time to re-contact the respondent in 2nd time survey near to the 1st time survey, the odds to get contact with respondent is higher. When the household size is more than one person who available to ask, the odds to get contact is higher. The older respondent has higher odds to get contact compare to younger respondent. The female respondent has higher odds to get contact compare to the male respondent. However, in the Model A2 which included interaction variables, showed nothing different with the result of Model A1. This means interaction between interviewer experience and panel design variables have no significance effect on dependent variable.

In Model B1 and B2, which the baseline survey conducted later than the sample in Model A1 and A2, showed some different result. There are three variables found to have significance effect on survey contact and noncontact: urbanization of respondent residential, respondent age, and respondent education level. Those variables include: day distance with lastime survey, respondent life time table effect, and household size have no more significance effect on dependent variables. The Model B1 result showed, when the level of urbanization becomes higher, the odds to contact respondent become higher. In those respondent variables, the respondent with higher education level and the older respondent have bigger odds to get contact compare to the respondent who have lower education level and the younger respondent. In the Model B2 which included interaction variables, showed the interaction terms between interviewer experience and respondent's life time-table have significance effect on dependent variable. The fourth column in Table 1 showed if the time to re-contact the respondent in 2nd time survey near to the 1st time survey, the odds to get contact with respondent is higher. Additionally, for an interviewer who has more experience, have higher odds to get contact with the respondent compare to the interviewer who less experience and the odds become higher when the time to re-contact the respondent is

far from the first time interview.

In table 2, another four models have been presented. For the same reason, the table included four models to analyze the reason respondent to co-operate or to reject the panel survey. For these four models presented in table 2, two equations are presented, both equations including the respondent level predictors and interviewer level predictors, but the equations for model A4 and model B4 will included the interaction variables.

In Model A3 and A4, two main factors were found to be associated with survey co-operate and refusal: the first time interview length and the experience level of interviewer. As the first column of table 2 indicates, when the length of first time interview is shorter, the odds to get co-operate from respondent is higher. The interviewer with more experience has higher odds to get co-operate from respondent compare to the interviewer with less experience. However, in the Model A4 which included interaction variables, showed a poor result, all the independent variable which included can't reject the hypothesis about coefficient equal to zero. The interaction variables which included have no significance effect at all. This means interaction between interviewer experience and panel design variables have no significance effect on dependent variable.

In Model B3 and B4, which the baseline survey conducted later than the sample in Model A3 and A4, showed some different result. There are three variables found to have significance effect on survey co-operate and refusal: respondent life time table effect, respondent education level, and interviewer experience level. The Model B3 result showed when the time to re-contact the respondent in 2nd time survey near to the 1st time survey, the odds to get co-operates from respondent is higher. In those respondent variables, the respondent with higher education level has bigger odds to get cooperation from respondent compare to the respondent who have lower education level. For the interviewer effect, the result showed the interviewer with more experience has higher odds to get co-operate from respondent. In the Model B4 which included interaction variables, showed the interaction terms between interviewer experience and first time interview length have significance effect on dependent variable. The fourth column in Table 2 showed the odds of higher education level respondent to co-operate is higher compare to the education level lower interviewer. Additionally, for an interviewer who has more experience, have higher odds to get co-operate from the respondent compare to the interviewer who less experience and the odds become higher when the length of the first time interview is longer.

Table 1: Multilevel Analysis of Contact and Noncontact in Panel Survey

	1st -10th waves Rolling		11th -15th wave Rolling	
	A1	A2	B1	B2
	B	B	B	B
	(s.e)	(s.e)	(s.e)	(s.e)
Fixed(1st-level)				
Panel Design				
Distance	-0.007 ***	-0.008 ***	-0.003	-0.001
	(0.001)	(0.002)	(0.003)	(0.004)
Life time-table	-0.001 *	0.000	0.000	-0.001 **
	(0.000)	(0.000)	(0.000)	(0.000)
Survey Area				
Urbanization	-0.039	-0.038	0.058 *	0.060 *
	(0.035)	(0.035)	(0.035)	(0.035)
Household Size	0.406 ***	0.406 ***	0.052	0.058
	(0.096)	(0.096)	(0.099)	(0.099)
Respondent's				
Male	-0.147 **	-0.147 **	-0.095	-0.094
	(0.061)	(0.061)	(0.062)	(0.062)
Age	0.023 ***	0.023 ***	0.022 ***	0.022 ***
	(0.003)	(0.003)	(0.003)	(0.003)
Education level	0.029	0.030	0.0871 ***	0.086 ***
	(0.024)	(0.024)	(0.025)	(0.025)
2nd level				
Male	-0.146	-0.146	0.026	0.030
	(0.346)	(0.346)	(0.121)	(0.121)
Experience	4.5E-05	4.79E-05	8.51E-05	0.0000263
	(4E-05)	(5.9E-05)	(0.00016)	(0.000287)
Interaction				
Distance*Experience		2.55E-07		-9.3E-06
		(5.7E-07)		(1.07E-05) *
Life *Experience		-1.6E-07		1.33E-06
		(1.2E-07)		(6.99E-07)
Constant	-0.700 **	-0.710 **	-1.465	-1.488
	(0.283)	(0.297)	(0.245)	(0.254)
N1	4863	4863	4478	4478
N2	52	52	34	34

Table 2: Multilevel Analysis of Cooperate and Refusal in Panel Survey

	1st -10th waves Rolling		11th -15th wave	
	A3	A4	B3	B4
	B	B	B	B
	(s.e)	(s.e)	(s.e)	(s.e)
Fixed(1st-level)				
Panel Design				
Life time-table	0.000	0.000	0.001 **	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Length	-0.031 *	-0.025	-0.036 *	-0.005
	(0.019)	(0.023)	(0.020)	(0.027)
Survey Area				
Urbanization	-0.079	-0.079	0.069	0.077
	(0.081)	(0.081)	(0.085)	(0.085)
Respondent's				
Male	0.227	0.227	-0.065	-0.056
	(0.142)	(0.143)	(0.152)	(0.152)
Age	-0.005	-0.005	0.003	0.003
	(0.006)	(0.006)	(0.007)	(0.007)
Education level	0.078	0.076	0.295 ***	0.290 ***
	(0.056)	(0.056)	(0.062)	(0.062)
2nd level				
Male	0.054	0.053	-0.011	-0.012
	(0.386)	(0.386)	(0.182)	(0.184)
Experience	0.0000836 **	0.000103	0.001 *	0.002 **
	(4.22E-05)	(9.7E-05)	(0.000)	(0.001)
Interaction				
Life		9.60E-08		1.57E-09
		(2.8E-07)		(2E-06)
Length*Experience		-2.9E-06		-0.00012 *
		(7.2E-06)		(6.8E-05)
Constant	2.151 ***	2.126 ***	0.565	0.158
	(0.519)	(0.542)	(0.575)	(0.631)
N1	1995	1995	1727	1727
N2	50	50	34	34

5   Conclusion and Discussion

This essay intends to explore reason respondent participation in telephone panel survey. This is done through combining both respondent level variables and interviewer level variables and build up a multilevel model to analyze. The result reveals several findings of significant importance.

First, the findings reported in table 1 and table 2 suggests that some panel design variables are associated with both contact/noncontact and co-operate/refuse decision. In the question of contact or noncontact, the date of first time interview seems important and makes the result different. From table 1, we may say for the baseline surveys who conduct earlier, those samples have higher probability to lost contact in panel survey. By the way, the sample from the latter baseline survey, have no problem on the day distance from the baseline survey and panel survey. Moreover, respondent life time table have significance effect on the contact result. This suggests the fieldworkers better find the respondent in the same time as they want to improve the contact rate.

Secondly, in the question of co-operate/refuse, respondent more intend to refuse the panel when the lastime survey take a lot of time. As Dillman (1978) argue, why people respond to a social survey is a social exchange between survey interviewer and respondent. When the time cost becomes higher, the respondent will intend to refuse the panel survey. That is better for researcher to reduce the time length or shortened the questionnaire for baseline survey. Equally, the life time table effect is exist in panel survey. In the panel survey, respondent that gets contact in the time as close as the time in baseline survey, they will more intend to accept the panel survey. This suggest the researchers better re-contact the respondent in the same time to increase the co-operate rates.

Finally, but not the least, this paper findings suggest that the chance for both noncontact and refusals are subject to respondent, interviewer, as well as to panel design. However, one has to keep in mind that other unmeasured characteristics of interviewers, for example their appearance and their voice may also play a role in the respondent's decision. Important implications for further empirical researches are to collect more detail information on interviewer and improve the extent their training to cope with the challenges of the issue of nonresponse in survey data collection.

Reference

Ashenfelter, Orley, Angus Deaton, and Gary Solon, 1986, "Collecting Panel Data in Developing Countries: Does it Make Sense?" *LSMS Working Paper 23*, Washington, DC: The World Bank.

Caetano, R. 2001, "Non-response in alcohol and drug surveys: A research topic in need of further attention," *Addiction*,   96:1541–1545.

Couper, Mick. P., Robert. M. Groves. 1992, "The role of the interviewer in survey participation," *Survey Methodology,* 18: 263-77.

Defo, Barthelemy Kuate. 1992, "Mortality and Attrition Processes in Longitudinal Studies in Africa: An Appraisal of the Ilford Surveys," *Population Studies*, 46: 327-348

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt, 1998, "An Analysis of Sample Attrition in Panel Data," *The Journal of Human Resources*, 33 (2): 251-99.

Groves, Robert. M. 1990, "Theories and Methods of Telephone Surveys," Annual Review of Sociology, 16: 221-240.

Groves, Robert. M. and Couper, Mick. P. 1998, *Nonresponse in Household Interview Surveys*, New York: John Wiley & Sons.

Jimenez-Martin, Sergi and Franco Peracchi. 2002, "Sample Attrition and Labour Force Dynamics: Evidence from the Spanish Labour Force Survey," *Spanish Economic Review*, 4:79-102

Raudenbush, Stephen W., and Anthony S. Bryk. 2002, *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.

Zabel, Jeffrey 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 (2): 479-506.