On selectivity of nonresponse in discrete-time, multi-wave panel studies.

Taris, T.W.

published in
Quality and Quantity
1997

document version
Publisher's PDF, also known as Version of record

Quality & Quantity **31**: 79–94, 1997.

On selectivity of nonresponse in discrete-time multi-wave panel studies*

TOON W. TARIS
Kurt Lewin Institute/Free University of Amsterdam, Department of Social Psychology, Van der Boechorststr. 1, 1081 BT Amsterdam, The Netherlands (E-Mail: AW.TARIS@PSY.VU.NL)

Abstract. The current paper deals with the issue of the detection of selective nonresponse in discrete-time, multi-wave panel studies. If groups in a population differ with respect to the chances that they will be (and remain) in a longitudinal sample, we speak of selective nonresponse. Ultimately, selective nonresponse may lead to a sample that is very different from the target population. We discuss ways to detect and quantify the amount of selectiveness by means of discrete-time Markov models. Then we proceed by addressing how a researcher may gain understanding of how to solve the problems caused by selective nonresponse, and the degree to which these solutions will be effective, by means of data on the nonresponse during a three-wave panel study involving 2800 young Dutch adults.

Key words: (selective) nonresponse, representativeness, longitudinal research, discrete-time panel analysis, Markov models

Introduction

In almost all field surveys it is the case that data are obtained from less respondents than was hoped for. For example, adresses and phone numbers from 2000 respondents may be available; however, usually a considerably smaller portion of the respondents will actually cooperate in the study. This phenomenon–not being able to obtain data from all respondents who initially were meant to be in the sample, either due to explicit refusal to participate in the study, or to other reasons is called *nonresponse*.

Goyder's (1987) review has shown that a nonresponse rate of about 30 to 40 per cent is not exceptionally high in many panel studies. Apart from the fact that a high nonresponse rate implies that many potential respondents have to be approached to obtain an acceptably large sample (hence, a sufficiently high statistical power), nonresponse can also be a major threat to the *validity* of a study. It is by now canon that a high nonresponse rates endangers the representativity of the study, because chances are that nonresponse is

*This research was supported by a grant of the Netherlands Organization for Scientific Research NWO, grant number PGS 56-381.

selective or *non-random*, i.e., some groups in the target population may have a larger or smaller chance to be in the sample than other groups – but see Taris (1995, 1996), who shows that a high nonresponse rate can coincide with a completely representative sample and vice versa; therefore, this rule of thumb is not generally valid.

Van de Pol's (1989) review demonstrated that inhabitants of large cities, women, youngsters, elderly, low SES, and low education respondents are often underrepresented in samples. The generalisability of conclusions that are based on such a biased sample can, of course, not be warranted. Restriction of range-effects can be expected, and if variables of interest are related to variables that 'cause' nonresponse, the estimates of their effects on other variables will be biased. Hence, non-random nonresponse can seriously jeopardize the generalisability of the conclusions of a study. This applies even more to longitudinal research, as nonresponse tends to *accumulate* across waves: all too often, respondents who have dropped out once are not contacted for further waves of the study.

Below we first discuss how selective nonresponse can be detected. First we briefly treat traditional approaches. Then we show how Markov modeling can be applied in examining and quantifying nonresponse. Finally, by means of an example we show how to test for the degree of selectiveness of the sample, and whether the mechanism that generated the nonresponse has been identified. The latter is of importance in subsequent data analysis, because many ways of correcting for nonresponse (most importantly, weighting the sample) either implicitly or explicitly assume that this mechanism is known.

Approaches to identifying selectiveness of nonresponse

Below we first discuss three approaches that are commonly used in assessing the representativeness of longitudinal samples, (1) comparison of the sample to figures on the composition of the target population; (2) comparison of the sample to figures obtained in other, comparable, studies; and (3) relating nonresponse to variables in the study. Then we proceed by treating an alternative to these methods, namely (4) inspection of the nonresponse pattern across time.

Figures on the composition of the target population

A simple way to obtain some insight in the degree to which the sample is representative for the target population is to compare figures on the composi-

Table 1. Marginals cannot yield conclusive evidence about representativeness

Table 1a
Strong association between variables

	Employed	Not employed	
Male	10	40	50
Female	40	10	50
	50	50	100

Table 1b
Independence of variables

	Employed	Not employed	
Male	25	25	50
Female	25	25	50
	50	50	100

tion of the target population with the distributions of the corresponding variables in the sample. For instance, for almost all (Western) countries information about the distributions of the population concerning age, gender, socio-economic status, and level of education is publicly available. If our sample is to be representative of the inhabitants of such a country, the distributions of the sample on these variables should be very similar to these population figures. Usually at least some of such figures can be obtained with respect to these biographic variables, occasionally even in cross-classified form, which is rather more useful. For instance, let in a particular population half of the respondents be male, and half of the respondents be employed. It could be the case that one's sample is not an adequate reflection of the composition of the population, even if 50 per cent of the respondents in the sample would indeed be male, and 50 per cent of the respondents would be employed; there are many compositions of the sample that could comply with these prerequisites. This is illustrated in Table 1, presenting two of the many possible compositions of a 100-person sample that comply with the figures given above (i.e., half of the respondents employed; half of the respondents male). In the first case (Table 1a), there is a strong association between gender and employment status, Pearson chi-square with 1 $df = 36.0$; $p < 0.001$; in the second case (1b) these variables are completely independent (the proportions in the cells can be computed as the product of the proportions in the table margins, which is impossible if the variables are correlated to some degree).

Clearly, the composition of the sample can be quite different, even if the *univariate* distributions regarding some variables are identical. The important point to note here is that information about the distributions of particular variables in the population can *only* be informative concerning the representativeness of the sample, if it can be assumed that these variables are uncorrelated. If they are not, comparing the sample distributions to the population distributions can be highly misleading.

A second major problem with this approach of detecting selective nonresponse is that its logical status is questionable. The implicit assumption seems to be that if the sample resembles the target population with respect to *these* variables, it will be representative in other respects as well, both with regard to the univariate distributions of the variables and their multivariate distributions – a conclusion that is simply one bridge too far. Of course, one could maintain that a sample is representative unless the opposite is shown – but I guess that anyone taking this argument seriously would ultimately end up refraining from checking one's data for selective nonresponse: a scientifically extremely undesirable situation. Hence, this approach can never warrant the conclusion that the data are representative – it can only help one in identifying possible biases in the data.

Comparison to other studies

Another possible source to draw data from concerning the distribution of variables in the target population is to compare one's own sample to studies that have examined a similar, or even the same, population. Many studies routinely report means and standard deviations of the researcher's sample. If one uses the same or a similar instrument in one's own study, one can compare the distributions of this instrument in one's own study with these other studies. The advantage of such a comparison is that information on other variables than biographic variables only can be compared. Of course, there is a catch here; if no difference is found, it could simply be the case that one's own study is just as bad as the previous study with regard to (selective) nonresponse, while differences could both mean that one's study is better or worse than the other study. Additionally, the same difficulties as mentioned above in the discussion of relating the sample to figures on the target population apply here as well.

Relating nonresponse to variables in the study

A third way to test for selective nonresponse is to relate variables in the study that were measured at time one, to a dichotomous variable that indicates whether one is still in the study at a follow-up wave. For example, one could find out whether gender is significantly associated with nonresponse, and so on. Ideally, we would find that none of the variables chosen is related to later nonresponse, which could be taken to mean that the sample is at least not getting less representative than it was at the first wave concerning the variables of interest. Of course, the weakness of this approach is that one still has to test for representativeness of the sample at the first occasion,

and that the sample may be getting less representative with respect to variables that were not measured during the first wave. In that sense, this approach is often not very informative.

Inspection of nonresponse patterns across time

In two earlier publications I have proposed to take the nonresponse pattern across time as an indicator of the degree to which selective nonresponse is at work (Taris, 1995, 1996). Many longitudinal studies show a decreasing nonresponse rate across waves. But, contrary to what one would intuitively assume, this is *not* a good sign. This is illustrated by the following hypothetical example. Let a population consist of two equal-sized groups (A and B), that have different response probabilities R_a and R_b. Let their time-invariant response probabilities be 0.6 and 0.9, respectively. Finally, assume that the members of these groups have an equal chance to be asked to participate in the study. If we would like to have a first-wave sample of 1,000 persons, we would have to contact a total number of 1,334 persons (667 of each group).[1] The total nonresponse for the first wave will therefore be slightly more than 25 per cent. Of the 667 members of group B, 600 respondents will agree to participate; however, only 400 members of group A will be in the sample. The time-one A-to-B ratio will, hence, be 1.5: for each member of A, there will be 1.5 member of group B in the sample at the first wave.

It is easy to see that the overrepresentation of group B will increase with every successive wave. This implies that the nonresponse of each wave, compared to its predecessor, will *decrease*, because group B – whose members will constitute a larger part of the sample with every wave – has a lower nonresponse rate than group A. For four waves and $R_a = 0.6$ and $R_b = 0.9$, the percentages nonresponse will be 25.0, 22.0, 19.2, and 16.7, respectively. Note that these very modest differences in nonresponse rates already imply that at the fourth wave there will be five members of group B for every member of group A. Thus, even a slightly decreasing nonresponse pattern informs us that selective nonresponse has taken place, and that serious bias may have occurred as a result.

The advantages of the latter procedure in comparison to the preceding three approaches to assessing selectivity of nonresponse are threefold. Firstly, it is *simple*; just eyeballing the nonresponse pattern already informs us whether there is reason to suspect selective nonresponse. Secondly, it is *general*; while more traditional approaches can only show whether a sample is representative with respect to particular variables that happen to be in the study, the curent approach will detect any type of selectivity, irrespective of whether

the source of this selectiveness is in the study or not. For instance, assume that gender is the only important predictor of selective drop out. Using any of the three traditional approaches we can only detect selective nonresponse if we have a measurement of gender in the study. However, the nonresponse pattern will reveal that selective nonresponse has taken place, *whether or not* we actually have a measure of gender.

Thirdly, if we accept that the nonresponse pattern across time gives an indication of the degree to which nonresponse was selective (Taris, 1995, 1996), we have a handle to quantify the amount of selectiveness, which in turn (1) suggests a method to identify the mechanism that generated the nonresponse by means of discrete-time Markov models, and (2) leads to a method to examine to which degree controlling particular variables reduces the amount of selectiveness. These two points are elaborated in the this paper. Below we first discuss the essentials of the discrete-time Markov model. Successively we proceed by presenting a simple application of our methodology.

Nonresponse as a Markov process

Following Taris (1996), we propose to construe nonresponse as the result of the operation of a *discrete-time Markov process* (e.g., Bartholomew, 1981; Coleman, 1968). In its simplest form, a Markov-process is a probabilistic process occurring in discrete time, concerning the score of a respondent on a discrete (qualitative) variable X. The score on X should be available for all waves of the panel study. With regard to nonresponse, X is a dichotomous variable with categories x_1 = 'still in the study' and x_2 = 'nonresponse'. For simplicity we assume in the following that respondents who do not participate in a particular wave are not contacted for any of the follow-up waves, i.e., once one is in state 'nonresponse' one remains in that state (i.e., 'nonresponse' is an *absorbing state*), but this assumption can easily be relaxed.

A *first-order* Markov process is described by the vector $\mathbf{P}$ of probabilities that the respondent belongs to state i ($i = 1, \ldots, p$) and the $p \times p$ matrix of w to $w + 1$ transition probabilities $\mathbf{O}$ (Bartholomew, 1981). Assuming that the waves are equally spaced in time, we can compute $\mathbf{P}$ for every wave w by means of

$$\mathbf{P}_w = \mathbf{P}_0 \times \mathbf{O}^w, \tag{1}$$

with $\mathbf{P}_0$ representing the fraction of the sample that belongs to x_1 and x_2 at the first wave. Take, for example, group A in Table 1, with response proba-

bility $R_a = 0.6$. The transition matrix **O** that corresponds with this probability is

$$\begin{bmatrix} 0.6 & 0.4 \\ 0 & 1 \end{bmatrix},$$

i.e., the probability to remain in state x_1 ('still in the study') at wave w, given that one was in the study at $w - 1$, is 0.6; the likelihood to experience a transition to x_2 at time $w + 1$ ('dropped out'), given that one is in x_1 at w, is 0.4; the probability to experience transition to x_1 given x_2 is zero; and the probability to remain in x_2, once one belongs to x_2, is 1. This example shows that the operation of a constant response probability R can be described as a time-invariant first-order Markov process, i.e., the simple first-order w to $w + 1$ transition matrix which is constant over waves and the vector $\mathbf{P}_0$ are sufficient to describe the attrition process at any given point in (discrete) time: at any wave w it is precisely known how many respondents will be in the study, and how many will already have dropped out.

It is also possible that the observed nonresponse pattern does not fit the hypothesis that it is the result of the operation of a first-order Markov process. For instance, a second- or higher-order Markov chain could be responsible for the nonresponse pattern (i.e., $\mathbf{P}_w$ cannot be predicted from **O** and $\mathbf{P}_{w-1}$ only; we need more information to do this, for example, also information about $\mathbf{P}_{w-2}$). For simplicity such higher-order Markov processes are not considered here however.[2]

Maximum likelihood-estimates of the parameters of these simple (as well as many other, much more complicated) Markov models can be obtained using a version of the EM-algorithm (Van de Pol & Langeheine, 1989), while the fit of the model can be assessed by comparing observed frequencies with the frequencies as expected on the basis of parameter estimates (using a chi-square distributed likelihood ratio test). The necessary arithmetic can be found in for example Anderson (1954)[3]

Now, if it is the case that random nonresponse can be conceptualized as the result of the operation of such a simple time-invariant first-order Markov process, the difference between the observed nonresponse and the nonresponse as predicted on the basis of a first-order Markov chain can be seen as an indication of the *amount of selectivity* of the nonresponse. Stated differently, the degree to which a simple first-order Markov chain fits the observed data (as reflected in the test statistic), indicates whether the observed nonresponse pattern differs significantly from what one would expect if nonresponse was random. If the data are acceptably well accounted for by a time-invariant first-order Markov chain, there is no reason to assume

selective nonresponse. If, however, such a simple model is not good enough, we can assume that selectiveness is a problem.

In such circumstances we can use this methodology to examine whether particular background variables (gender, age, etc) are related to nonresponse. If they are, one can *stratify* the panel on the basis of this variable, fit a first-order Markov model to all subgroups separately, and see to which degree 'controlling' this variable reduces the amount of selectivity as reflected in the value of the test statistic. That is, by fitting time-invariant first-order transition matrices to all groups separately, we allow the transition matrices to be different across groups, but not across time. This corresponds with a situation in which nonresponse rates are different *across* groups, while nonresponse is random *within* groups.

If the value of the test statistic is still significant, we have not yet fully accounted for the selectivity of the nonresponse. In that case one has to continue the quest for other variables that possibly account for the bias in the sample. If the test statistic is not significant (i.e., it is plausible that the nonresponse is random, once one controls the effects of particular background variables), one has discovered the mechanism behind the nonresponse, and one may proceed by applying procedures to adjust for nonresponse (i.e., weighting particular classes of respondents).

How important is a particular variable in accounting for selective nonresponse?

The amount of selectiveness accounted for by splitting the sample with regard to a particular variable can be computed by relating the decrease in the test statistic when fitting a stratified model where the subgroups are allowed to have different transition matrices, relative to the fit of a model in which these transition matrices are constrained to be equal across groups. This yields what we will call a pseudo-R^2 value:

$$\text{pseudo-}R^2 = \frac{\chi_n^2 - \chi_t^2}{\chi_n^2}, \tag{2}$$

where subscript n refers to the model where the groups are restricted to have equal transition probabilities (the null model), and t to a target model (the model where the groups were allowed to have different transition probabilities). Pseudo-R^2 ranges from 0 to 1, with 1 indicating that selectiveness has completely been accounted for (cf. Tuma & Hannan, 1984, who propose a similar measure in assessing the fit of event history models).

In short, above we propose to consider the degree to which nonresponse

rates can be considered due to a time-invariant first-order Markov chain as an indication of the amount of selectiveness in a particular discrete-time panel study. The amount of selectiveness can be expressed in terms of the value of the test statistic, comparing the observed data to the data as reproduced on the basis of the estimated first-order Markov chain. Stratification of the sample on the basis of variables that are related to the nonresponse will usually lead to a lower value of the test statistic; the importance of such a stratifying variable can be assessed by computing a *Pseudo-R^2 value*. *Below* we present an application of this methodology to real data, showing how our procedure may be carried out in everyday research.

Example: The socialization process of young adults in the Netherlands

Sample. A longitudinal panel study was conducted in the Netherlands to examine the development of the socialization process of young adults aged 18 to 30 years. During the observed period three waves with two years in between were conducted, in fall/winter 1987/88, fall/winter 1989/90, and fall/winter 1991/92, respectively. For the first wave, 2,800 subjects in total were to be asked to participate in the study. Of these, 1,775 subjects participated in the first wave (a nonresponse percentage of 36.6). Two years later, 1,419 of these subjects also took part in the second wave (a nonresponse 20.1 per cent). Finally, 1,257 respondents participated in all three waves (nonresponse is 11.4 percent). Respondents who dropped out were not asked to participate in the follow-up waves. A detailed account of the response and nonresponse per wave, cross-tabulated by birth cohort membership (1961, 1965 and 1969) and gender, is presented in the Appendix.

Nonresponse. As mentioned above, the nonresponse rates per wave were 36.6, 20.1, and 11.4 per cent, respectively. Hence, on the basis of this decreasing nonresponse rate one must suspect that nonresponse was not random. A simple first-order Markov chain could account for the pattern in the data: L^2 with 4 degrees of freedom degrees of freedom was 45.5; $p < 0.001$. Hence, the null hypothesis that the observed nonresponse pattern was the result of a simple first-order Markov chain had to be rejected.

Therefore, we examined the nonresponse pattern more closely. We were interested in explaining this nonresponse by examining nonresponse patterns in different sub-groups of the sample. Two divisions could readily be made: one on the basis of gender, the other on the basis of birth cohort membership. These variables – age and gender – have often been shown to be related to nonresponse. In two earlier publications (Dijkstra & Smit, 1989, and Taris

et al., 1993) it was shown for the current sample that nonresponse was indeed related to these variables, though not very strongly. Of course, other variables could also be important in determining nonresponse, but the scores of the subjects on these other variables were not available for the subjects who did not participate in the first wave; only information on their sex and age was present. We suspect that this will not be very different in many other panel studies. Therefore, this example will provide us with a fairly realistic impression of the merits and disadvantages of our approach. Below we first discuss the results of a stratification on the basis of gender and birth cohort; successively we examine the joint effects of these variables.

Gender. A two-group Markov analysis was conducted, taking males and females as separate groups. The transition probabilities were taken as *group-invariant*, i.e., the transition matrix **O** had to be equal across both groups. This led to a L^2-value of 57.72 with 10 *df*; $p < 0.001$, which we take as our null model here. The question is, to which degree does allowing the transition matrix **O** to be different across groups lower the L^2-value? Estimating the model for both groups separately yielded a L^2-value of 53.01 with 8 *df*; $p < 0.001$. Therefore, we gain 4.71 L^2-points with a loss of 2 *df*; $p < 0.10$. Looking at the L^2-contributions of each group separately reveals that the male group contributes 44.52 L^2-points with 4 *df*, while the female group contributes only 8.49 L^2-points, also with 4 *df* ($p > 0.05$). Hence, it seems that for females a simple first order Markov chain applies (i.e., nonresponse is *not* selective within this group), while for males nonresponse is still selective.

Applying equation (2) shows that controlling gender reduces the amount of selectiveness of the nonresponse with 8.2 per cent. Though this figure is not impressive, our analyses revealed that the selectiveness of the nonresponse was concentrated within the male group of the sample, and in this sense our results are helpful. It appears that if analyses were restricted to women only, there would be much less reason to fear the harmful consequences of selective nonresponse (at least not after weighting the sample with the appropriate figures).

Birth cohort membership. In a similar vein, we conducted a three-group first-order Markov analysis with birth cohort as the stratification variable. First the transition matrix was constrained to be equal across all three groups. This yielded a L^2-value of 54.09 with 16 *df*; $p < 0.001$. Hence, it appears that there is reason to suspect that nonresponse was selective; a simple first-order Markov chain cannot account for the observed nonresponse pattern.

Could it be that birth cohort membership is a factor that influences nonresponse? After all, one often observes that age is related to nonresponse.

Therefore it was decided to allow the transition matrix to be estimated separately for each birth cohort. This results in a L^2-value of 50.24 with 12 *df*; $p < 0.001$: i.e., a decrease of 3.85 L^2-points (with a loss of 4 *df*). Pseudo-R^2 is 0.07 in this case, which shows that the amount of selectiveness that is accounted for by birth cohort membership is actually quite small (though previous analyses, which are not reported here, had shown that nonresponse was significantly affected by birth cohort membership). The result obtained here should be interpreted in such a way, that controlling birth cohort membership will *not substantially* reduce the consequences of selective nonresponse (namely, with a mere 7 per cent).

However, as we have seen above, marginals cannot yield conclusive evidence with regard to the possible selectiveness of a sample. Therefore we decided to split the sample on the basis of gender and birth cohort membership, resulting in six birth cohort-gender groups.

Gender and birth cohort: Joint effects? Finally we examined the joint effects of gender and birth cohort membership. Six gender-birth cohort groups were formed, and a simple first-order Markov chain with constant transition probabilities across groups was fitted to the data. The results indicated that this initial model was not tenable (L^2 with 34 *df* was 71.10; $p < 0.001$; χ^2 is 69.60). The less restricted model where the transition matrix was constrained to be equal across time, but not across groups, yielded a χ^2 of 59.53 with 24 *df*; $p < 0.001$. Hence, again the improvement in fit is not very large: Pseudo-R^2 was only 0.13 (i.e., 13 percent of the selective nonresponse is accounted for).

Figure 1 shows the observed nonresponse rate for each wave for each group separately (curved lines), as well as the estimated nonresponse rates (the horizontal line, indicating constancy across waves). It is evident that in the case of men, the observed nonresponse rates decrease strongly; this is for all three groups an almost straight line. Clearly, the simple Markov model does not fit the data for men. However, as was already suggested by the male/female analysis, this picture is rather different for women. Especially for the oldest and youngest birth cohorts, the observed curves are much closer to the expected nonresponse rate than in the case of men. This is also observed from the contributions of each of the groups to the χ^2-value, given in Table 2. This table shows that for the women, a first-order Markov chain is tenable, even if the possible confounding effect of birth cohort membership is taken into account; the considerably larger χ^2-contribution of women born in 1965 to the overall-x^2 value is still not large enough to indicate a significant

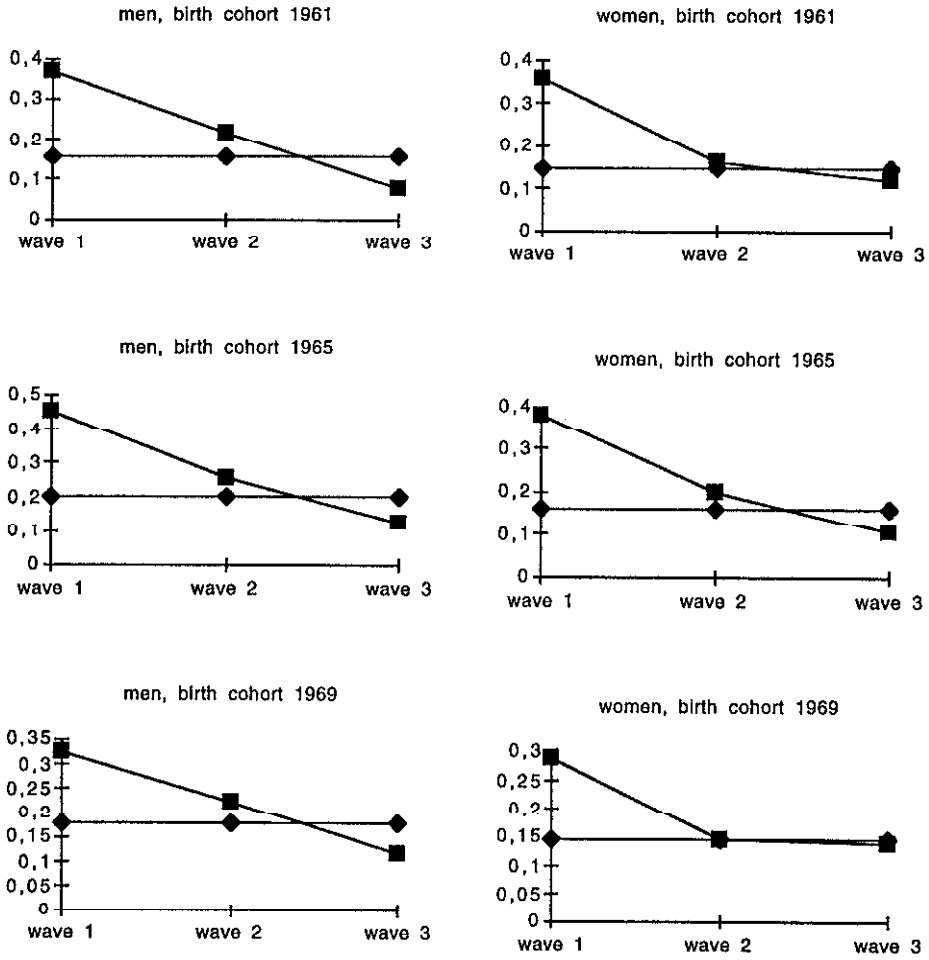

Figure 1. Expected (straight line) and observed nonresponse rates for six groups (observed rates calculated from the data in the Appendix, expected rates as expected on the basis of a first-order Markov chain).

Table 2. Contributions of the six groups to χ^2-value, transition matrices constrained to be equal across time but not across groups

Group	χ^2-contribution	$p(df)$
Men, birth cohort 1961	19.58	< 0.05 (4)
Men, birth cohort 1965	14.11	< 0.05 (4)
Men, birth cohort 1969	10.66	< 0.05 (4)
Women, birth cohort 1961	3.85	> 0.05 (4)
Women, birth cohort 1965	9.79	> 0.05 (4)
Women, birth cohort 1969	1.54	> 0.05 (4)

deviation from randon nonresponse. Summing over cohorts, the χ^2-contribution of the women to the overall chi-square value is 15.18 with 12 *df*; $p > 0.10$. The men, therefore, contribute 44.35 χ^2-units, also with 12 degrees of freedom ($p < 0.001$).

On the basis of the previous analyses we may conclude that controlling gender and birth cohort will result in an only modest reduction of the bias in the sample. Selective nonresponse will certainly continue to be a threat to the validity of the results, even after, for example, weighting the sample on the basis of the response probabilities of each gender/birth cohort combination: a somewhat disappointing conclusion.

On the other hand, our analyses indicated that the selectiveness was largely concentrated within the male stratum of the sample, while the bias was the greatest for birth cohort 1965 (cf. Table 2). In trying to reveal the mechanism behind the nonresponse one should therefore focus on these groups. Additionally, we should note that although the rather low Pseudo-R^2 values indicate that we cannot reduce the effects of selective nonresponse substantially, this does not in itself mean that the response bias will be large. If the bias in the sample is not severe to begin with, there is not much to account for, and low pseudo-R^2 values are only to be expected

Discussion

Above we have presented a new approach to studying selective nonresponse in discrete-time panel studies. An inspection of the pattern of nonresponse rates across time can reveal whether there is reason to expect bias in the sample or not, while discrete-time Markov models can be used to quantify the degree of selectiveness of the nonresponse. One may even stratify the sample on the basis of variables that are suspected to be linked to the nonresponse, and see whether controlling these variables will cure the bias in the sample in any substantial degree.

In a real-data example we have shown how this procedure works in practice. The results of this example was rather disappointing, as we have shown that controlling two factors that are commonly related to nonresponse – gender and age – would only result in a very modest decrease of bias. Insofar this is typical for other longitudinal studies, it points at the fact that it is better to be safe than to be sorry: Because the negative effects of selective dropout can probably not always completely be cured by statistical means (rather an understatement!), a practical recommendation should be that one should always try to *minimize dropout* as much as possible. While this does not warrant that selective nonresponse will be absent (cf. Taris, 1996), it will

at least reduce its impact: the lower the nonresponse, the less drastic the effects of possible selectiveness can be.

The approach to analyzing nonresponse that was proposed in the current research is not intended to *replace* earlier approaches. On the contrary, we feel that these older approaches are indispensable when it comes to analyzing nonresponse. In practice one could think of an integrated type of analysis in which one first visually inspects the nonresponse pattern, then tries to find variables that are linked to the nonresponse, and then examines the degree to which these variables can account for the observed nonresponse. In this sense, we consider the approach proposed here as a useful complement of these earlier approaches.

Appendix. Description of the data used in this paper (adapted from Dijkstra & Smit, 1989, and Taris et al., 1993).

Men					Women				
cohort 1961	wave 1	wave 2	wave 3	*N*	cohort 1961	wave 1	wave 2	wave 3	*N*
	0	0	0	173		0	0	0	163
	1	0	0	64		1	0	0	48
	1	1	0	18		1	1	0	29
	1	1	1	216		1	1	1	216
cohort 1965	wave 1	wave 2	wave 3	*N*	cohort 1965	wave 1	wave 2	wave 3	*N*
	0	0	0	241		0	0	0	181
	1	0	0	75		1	0	0	59
	1	1	0	27		1	1	0	25
	1	1	1	191		1	1	1	209
cohort 1969	wave 1	wave 2	wave 3	*N*	cohort 1969	wave 1	wave 2	wave 3	*N*
	0	0	0	144		0	0	0	123
	1	0	0	66		1	0	0	44
	1	1	0	27		1	1	0	36
	1	1	1	203		1	1	1	222

N.B. Score 0 indicates 'nonresponse' in this wave, 1 indicates 'participation'.

In all analyses conducted in this paper, 0.5 was added to the observed cell frequencies to prevent empty-cell problems. As respondents who did not cooperate once were not asked to participate in any of the follow-up waves, the four other logically possible patterns ([0, 0, 1]; [0, 1, 0]; [0, 1, 1]; and [1, 0, 1], respectively) did not occur.

Notes

1. The necessary arithmetic is elementary. We must solve the equation $R_a A + R_b B = 1000$, constraining A to be equal to B (Taris, 1996).
2. We limit our discussion to first-order Markov chains for two reasons. Firstly, a first-order Markov process is easier to discuss than a higher-order process, while the extension to higher-order chains is easy (e.g., Van de Pol & Langeheine, 1989). Secondly, testing higher-order Markov models requires more rounds of data collection, and longitudinal studies involving more than two or three waves are rare. For example, testing a first-order Markov chain one needs at least three waves of data, while testing a second-order Markov chain requires at least four waves.
3. A convenient and easy-to-use computer program that estimates the parameters of many different types of Markov models (including multiple group, latent class, and mixed Markov models), is the PANMARK 2.2 program by Van de Pol *et al.* (1990). PANMARK was used for all analyses reported in this paper.

Reference

Anderson, T. W. (1954). Probability models for analyzing time changes in attitudes. In P. F. Lazarsfeld (ed.), *Mathematical Thinking in the Social Sciences*. Glencoe, IL: The Free Press.

Bartholomew, D. J. (1981). *Mathematical Models in Social Science*. New York: Wiley.

Coleman, J. S. (1968). The mathematical study of change. In H. M. Blalock & A.B. Blalock (eds.), *Methodology in Social Research* (pp. 428–478). New York: McGraw-Hill.

Dijkstra, W. & Smit, J. (1989). Nonrespons (Nonresponse). In W. Dijkstra (ed.), *Het proces van sociale integratie van jong-volwassenen: De gegevensverzameling voor de eerste hoofdmeting* (pp. 59–70). Amsterdam: VU-Uitgeverij.

Goyder, J. (1987). *The Silent Minority*. Cambridge: Blackwell University Press.

Taris, T. W. (1995). Nonrespons en representativiteit in panel studies: Een korte notitie (Nonresponse and representativity in panel studies: A short note). *Kwantitatieve Methoden* 16, 41–44.

Taris, T. W. (1996). Modeling nonresponse in multiwave panel studies using discrete-time Markov models. *Quality & Quantity* 30: 189–203.

Taris, T. W., Van der Vaart, W. & Dijkstra, W. D. (1993). Nonrespons en de representativiteit van de follow-up metingen (Nonresponse and representativity of the follow-up waves). In W. D. Dijkstra (ed.), *Het proces van sociale integratie van jong-volwassenen; De gegevensverzameling voor de tweede hoofdmeting* (pp. 65–76). Amsterdam: VU-Uitgeverij.

Tuma, N. B. & Hannan, M. T. (1984). *Social Dynamics; Models and Methods*. New York: Lawrence Eribaum.

Van de Pol, F. J. R. (1989). *Issues of Design and Analysis of Panels*. Amsterdam: Sociometric Research Foundation.

Van de Pol, F. J. R., De Jong, W. & Langeheine, R. (1990). *PANMARK 2.2 User Manual: Panel Analysis Using Markov Chains*. Groningen: IEC ProGAMMA.

Van de Pol, F. J. R. & Langeheine, R. (1989). Mixed Markov latent class models, from description toward explanation. In C. C. Clogg (ed.), *Sociological Methodology* 1990.