Journal of Official Statistics, Vol. 23, No. 3, 2007, pp. 387–408

Nonresponse Among Ethnic Minorities: A Multivariate Analysis

Remco Feskens, Joop Hox, Gerty Lensvelt-Mulders, and Hans Schmeets[1]

This article examines the nonresponse among ethnic minorities in the Netherlands. We have constructed a structural equation model with various response outcomes, controlling the effect of ethnicity on the response outcomes for various socio-economic and socio-demographic variables. The effect of ethnicity on response is almost entirely mediated by the degree of urbanization. We have also performed multiple group analyses to examine differences between ethnic groups in the response outcome predictors. Here again, we note that urbanization has a negative effect on the response probabilities in all the ethnic groups and in particular on the contact probabilities. This negative effect is somewhat larger, however, among sampled units with a non-Western background.

Key words: Nonresponse; ethnic minorities; structural equation models.

1. Introduction

Nonresponse rates in survey research have increased in recent years in almost all the Western countries (De Heer and De Leeuw 2002). For several reasons, this is a problem. Firstly, nonresponse reduces the number of respondents and consequently the precision of estimates. Secondly, nonresponse can increase the costs of survey research since greater efforts are needed to reach the desired sample size. Thirdly, if nonresponse is selective, the survey estimates may be biased and not accurately reflect the true values of the target population (Groves and Couper 1998; Thornberry and Massey 1988). Nonresponse is selective when nonrespondents differ systematically from respondents in terms which matter to the survey objectives. (Groves and Couper 1998). The most disturbing consequence of nonresponse is the bias in point estimators (Groves 1989). Biased estimates are more likely to occur if specific groups exhibit below-average response rates. This makes it more likely that the nonrespondents differ systematically from the respondents, since the nonresponse is not random. So in order to speculate about nonresponse bias, it is important to look at response rates among various subgroups (Thornberry and Massey 1988).

Due to their above-average nonresponse rates Statistics Netherlands has difficulties in surveying the ethnic minority or immigrant population. Ethnic minorities constitute about

[1]R.C.W. Feskens, J. Hox, G. Lensvelt-Mulders (Utrecht University) and J.J.G. Schmeets (Statistics Netherlands), Heidelberglaan 1, 3584 CS Utrecht, The Netherlands (Email: r.c.w.feskens@uu.nl). The views expressed in this article are those of the authors and do not necessarily reflect the policies of Statistics Netherlands.
Acknowledgment: The authors wish to thank the anonymous reviewers for their helpful suggestions and comments.

20% of the Dutch population (http://statline.cbs.nl). The ethnic minority or immigrant population is defined in the Netherlands as "Everyone residing in the Netherlands with one or both parents who were born abroad" (Reep 2003). A further distinction is usually drawn between people with one or both parents born in Europe, North America, Australia, Japan or Indonesia and people from non-Western countries (mainly Turks, Moroccans, Surinamese and Antilleans).[2] The two groups are of approximately the same size. The response problems among ethnic minorities are not restricted to Statistics Netherlands. Ethnic minorities have lower response rates in almost all the Western countries (Eisner and Ribeaud 2007; Feskens et al. 2006). Nonresponse among ethnic minorities is becoming politically relevant. Ethnic minority interest groups in the United States have organized to avoid alleged under-counts in survey research. Comparing groups and cultures is an essential feature of survey research (Harkness et al. 2003). However, differences in the response rates of various ethnic groups may bias overall survey estimates. Couper and de Leeuw (2003) indicate that the under-representation of ethnic groups may threaten studies on values and norms, e.g., with regard to Sunday observance or commercial activities. Schmeets (2005a) presents other results where below-average response rates among ethnic minorities may bias survey results. Even after correction for age, ethnic minorities are found to be less happy, less healthy and less active in club life and have a greater sense of insecurity. Above that, better ethnic minority response rates are needed for good estimates of subpopulations.

One way to find out more about possible selectivity would be to use background information available for all the sampled units. Recent changes in Dutch legislation have enabled Statistics Netherlands to link administrative records, resulting in a unique database (Houbiers 2004). It provides rich background information about the nonrespondents, enabling us to more thoroughly examine the nonresponse patterns.

A previous study by Schmeets and Michiels demonstrates that the high ethnic minority nonresponse rates can be attributed to socio-economic status and urbanization (Schmeets and Michiels 2003). In particular, non-Western foreigners tend to have lower response rates than the native population. At the same time they live predominantly in urban areas, and they are more often unemployed and have lower education levels than the native population. These characteristics correlate negatively with response rates (Goyder et al. 1992; Lavrakas 1993; Groves and Couper 1998; Stoop 2004; van Goor et al. 2005).

The study by Schmeets and Michiels uses a logistic regression model. In this study we examine whether their conclusion also holds true if structural equation techniques are used. Structural equation modelling is more suitable than log linear modelling in several ways. Firstly, structural equation modelling makes it possible to include indirect effects. A variable can be independent as well as dependent simultaneously. Secondly, structural equation models can incorporate latent variables. In this analysis, we examine the concepts of social economic status and urbanization in a more detailed way than would otherwise be possible. Using latent variables makes it possible to measure these constructs more precisely. Thirdly, structural equation models provide more model fit statistics than simple logistic models do. We use the method proposed by Schneekloth and Leeven (2003) to assess the nonresponse bias introduced by below-average ethnic minority response rates.

[2] For reasons of simplicity we use "Western foreigners" and "non-Western foreigners" in this article.

They use logistic regression analysis to evaluate the degree to which the sample nonresponse can be traced back to population characteristics. Pseudo R square values are used to evaluate the explanatory power of the total model. In addition, the model is elaborated by including more variables and multiple group analyses.

Lastly, the effects on noncontacts and refusals are illustrated. These considerations have resulted in three research questions:

(1) Do ethnic minorities in the Netherlands have lower response rates, contact rates and cooperation rates?
(2) What is the effect of ethnicity on the various response outcomes if controlled for other socio-economic and socio-demographic variables?
(3) Do response models differ between various ethnic groups?

By addressing these questions, we hope to gain a better understanding of the nonresponse problem among ethnic minorities in a multivariate environment. The available data are described in the second section of this article, the methods are described in the third and the results in the fourth, which is divided into three parts. In the first part we show the response rates among ethnic minorities, the second part describes the construction of the structural equation model used to examine the effect of ethnicity on the response controlled for other variables, and the third part presents the results of our multiple group analyses. Lastly, our conclusions are given in Section 5.

2. Data

We have performed our analyses on the survey files of the *Continuous Survey on Living Conditions* (POLS) 1998 conducted by Statistics Netherlands. About 40,000 interviews are conducted every year. POLS is an integrated survey on living conditions of the Dutch population in private households. The POLS design is based on a modular structure consisting of a joint sample frame and a joint questionnaire. The observation units are individuals. The sample frame is the Population Register from all Dutch municipal basic administrations. POLS is a cluster sample. First communities are drawn, and then people. Large cities are automatically included (Schouten 2003). Communities and persons are drawn in such a way that the first order inclusion probabilities are equal across all sampled units with the exception of age, since the target population in some modules has age restrictions. We concentrate our analyses on the joint questionnaire, with the total Dutch population (except residents of nonprivate households) as the target population. Participation is voluntary in POLS and the survey is solely administered in Dutch. Every month a sample of about 3,500 people is drawn. In 1998, there is a two-month fieldwork period for the twelve consecutive samples. In the first month, Statistics Netherlands collects data with a CAPI mode and nonrespondents with a known telephone line are re-approached with a CATI technique. The nonrespondents without a known telephone line and the sampled units who cannot cooperate due to illness are reapproached with CAPI (Schouten 2003).

The POLS survey is supplemented by administrative data from the Population Register and information about employment and social benefits (Schmeets and Michiels 2003; Schouten 2003). Linking the administrative records makes socio-demographic and

socio-economic information available on the nonrespondents at the individual and postal code level. This information has been gathered at Statistics Netherlands in the Social Statistical Database, in which several registers are linked to each other as well as to data from sample surveys (Houbiers 2004). For an extensive summary of the construction of this database, see Houbiers (2004). In the POLS 1998 Survey, 39,431 sampled units were drawn and the number of respondents according to AAPOR response definition two was 23,993 (60.8%) (AAPOR 2006), which is not unusually low in the Netherlands (see e.g., De Heer 1999).

The additional information provided by the link to administrative data makes it possible to study the nonrespondents. However, no extra information is available on 1,143 of the units sampled (2.9% of the total sample). Since no systematic missing data pattern is to be found with respect to important background variables for these 1,143 cases, they are considered to be missing completely at random and deleted from the data file. Because these numbers are quite small, they can be dropped from the sample without a significant loss of information. Since we only analyse sampled units aged 15–65,[3] 28,542 sampled units are left for analysis. The response in this subsample is somewhat lower: 60.4% or 17,123 sampled units respond. To avoid capitalization on chance in the analysis, we randomly split this new file into an exploration file consisting of odd case numbers (14,271 cases) and a validation file consisting of the even case numbers (14,271 cases).

As is noted above, in the Netherlands ethnic minorities are defined as everyone residing in the Netherlands with one or both parents born abroad. Table 1 shows the ethnicity distribution in the Netherlands in 1998, the year the POLS survey was conducted.

Table 1. Population in the Netherlands according to ethnicity 1998 (http://statline.cbs.nl)

	Population	Percentage of total population
Total population	15,654,192	
Native population	13,033,792	83.3
Ethnic minorities	2,620,400	16.7
Western foreigners	1,341,947	8.6
Indonesia	407,885	2.6
Germany	405,911	2.6
Belgium	111,537	0.7
Non-Western foreigners	1,278,453	8.2
Suriname	290,467	1.8
Turkey	289,777	1.8
Morocco	241,982	1.5
Netherlands Antilles & Aruba	92,105	0.6

3. Methods

First of all, we looked at the bivariate relationships between ethnic groups and several response categories, and this provides information for answering the first research question. To address the second research question, we construct a structural equation model.

[3] The absolute number of ethnic minorities (especially those of non-Western descent) above 65 is very limited. In the year that the survey was held this number was 22,675 or less than 0.15% of the total population.

Structural equation modelling allows us to combine latent variables and structural relationships between them and other observed variables (Kline 1998). The advantages of path models (the so-called structural component) and factor models (the measurement component) are combined in structural equation models. Using latent variables reduces the effect of measurement errors. Structural equation modelling also makes it possible to analyse models in which variables are both exogenous and endogenous, and hence, the use of indirect effects. Indirect effects are useful for evaluating the combined effect of ethnicity, urbanization and socio-economic status on response probabilities. We first construct a structural equation model in the exploration file. This model is validated in the validation file. As Groves and Couper (1998) note, "dissecting the nonresponse phenomenon into one of noncontacts, refusals and other causes sensitizes us to considering alternative causes of each outcome." We thus not only analyse the response outcome, we also address noncontacts and refusals. Other causes of nonresponse only have a minor effect on the response rate, as is shown in Table 2, and are therefore not further analysed.

We also want to know which variables the two groups (native/ Western foreigners vs non-Western foreigners) differ on as regards the response phenomenon. We address this point by conducting a multiple group analysis and statistically comparing the path coefficients of the groups. The multiple group analyses are conducted on the total sample aged 15–65. Here again we dissect the response outcomes, which makes it possible to examine the various response predispositions in the subgroups. Since the sample size is large, the assessment of model fits is based on two goodness-of-fit indices that are less sensitive to sample size, namely the Bentler comparative fit index (CFI; see Bentler 1990) and the root mean squared error of approximation (RMSEA) value (Browne and Cudeck 1993). The CFI value indicates the degree of improvement of the overall fit of the specified model relative to an independence model in which the variables are assumed to be uncorrelated (Kline 2005).

$$CFI = 1 - \frac{\chi^2_m - df_m}{\chi^2_b - df_b} \tag{1}$$

The RMSEA fit index is an exact fit in which the null hypothesis states that the model corresponds to the data (RMSEA = .00). This value is calculated as follows:

$$RMSEA = \sqrt{\frac{\chi^2_m - df_m}{ndf_m}} \tag{2}$$

4. Results

4.1. Do Ethnic Minorities in the Netherlands Have Lower Response Rates, Contact Rates and Cooperation Rates?

The overall response rate for sampled units aged 15–65 in the POLS 1998 survey is 60.4% (AAPOR response definition number two). Further dissecting the nonresponse in alternative outcomes is as follows for the three ethnic groups:

Table 2. Response outcomes among ethnic groups in POLS 1998 in percentages

	Native population	Western foreigners	Non-Western foreigners
Response	62.0	57.2	39.7
Noncontacts	12.0	15.9	26.4
Refusals	24.7	23.0	20.1
Language problems	0.0	2.5	13.0
Other	1.3	1.8	1.0
N	24,005	2,511	2,026

Figures may not add up to 100% due to rounding.

The distribution of nonresponse outcomes is more or less similar among Western foreigners and the native population. The relative numbers of noncontacted sampled units and nonresponse due to difficulties with the survey language are higher among non-Western foreigners than among the native population, a result also found in other surveys (Feskens et al. 2006). These are the primary reasons for the lower response rates among non-Western foreigners.

We also address the bivariate relationships between ethnicity and various socio-demographic and socio-economic variables. Here again, the observed characteristics of Western foreigners are very similar to those of the native population, whereas those of non-Western foreigners differ substantially (Schmeets 2005a). We consequently decided to focus our analysis on non-Western foreigners. Ethnicity is a dichotomous variable in the following analyses with non-Western foreigners as the first and Western foreigners and the native population as the second group.

4.2. What Is the Effect of Ethnicity on the Various Response Outcomes if Controlled for Other Socio-economic and Socio-demographic Variables?

4.2.1. Structural Equation Model

We wanted to see whether the nonresponse is still affected by ethnicity if controlled for socio-demographic and socio-economic variables. First we considered the bivariate relationships between each of the socio-demographic and socio-economic variables and the response for all the ethnic groups separately. Then we looked at the bivariate relationships between the ethnic groups and the socio-economic and socio-demographic variables. With this information and the theoretical consideration that urbanization and socio economic status (SES) have been related to survey nonresponse for many years (Goyder et al. 1992; Lavrakas 1993; Groves and Couper 1998; Stoop 2004; Van Goor et al. 2005), we constructed a structural equation model enabling us to control the effect of ethnicity on the various response outcomes for the other variables (see Figure 1).

The measurement part of the model consists of the latent variables urbanization and SES. The latent variable urbanization is measured by the observed variables urbanization at postal code level, degree of urbanization of the city, and city size. These variables are sufficiently correlated with each other, but not correlated to the extent that they measure the same. The latent variable SES is measured by home values and an indicator for receiving social benefits. For identification purposes, we fixed the factor loadings of the indicators urbanization of the city and home value at one.

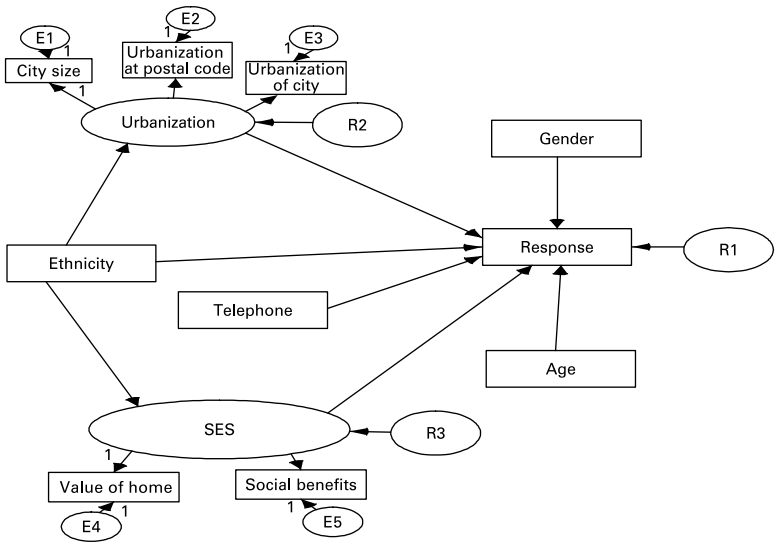

Fig. 1. Structural equation model

Probit regressions are estimated for the categorical factor indicators, and simple linear regressions are estimated for the continuous factor indicators (Muthén 1998-2004). We construct the two latent variables because using the extra information of all the indicators makes it possible to identify urbanization and SES more precisely. Moreover, some indicators are subject to subjective classification. Using latent variables with more indicators reduces this form of measurement error. The relationships between the other observed variables and latent variables on response constitute the structural part of the model. These observed variables are ethnicity, an indicator for having a known telephone land-line (telephone), gender and age. These observed variables are regressed on the binary outcome variable response.

Since we wanted to see whether the relationship between ethnicity and response is mediated by SES and urbanization, we also regressed these latent variables on ethnicity. The relationships are assumed to be unidirectional, and the latent variables SES and urbanization and also SES and having a known telephone land-line are assumed to covary (these relationships are not drawn in the figure). We also explored including interaction terms in the model, but this did not improve the model fit significantly. The regression coefficients are estimated with the unweighted least square estimator and are interpreted as probit regression coefficients. The estimates between parentheses are standardized coefficients using the variances of the continuous latent variables as well as the variances of the background and outcome variables (Muthén and Muthén 1998-2004). The sign of the regression coefficients in Tables 3–8 shows the extent to which this characteristic changes the probability in the nonresponse (-) or response direction (without sign). We assumed that the measurement errors are uncorrelated.

4.2.2. Response

The results of application of the structural equation model with the dependent variable response defined as the AAPOR response definition number two are presented in Table 3.

Table 3. Structural equation model with dependent variable response

	Explained variables					
	Estimates on response	Standard errors	Estimates on urbanization	Standard errors	Estimates on SES	Standard errors
Predictors						
Size of city			1.000 (0.874) fixed	0.000		
Urbanization of city			0.825 (0.914)**	0.011		
Urbanization of postal code			0.870 (0.887)**	0.015		
Value of home					1.000 (0.648) fixed	0.000
Social benefits					−0.370 (−0.491)**	0.020
Gender	0.075 (0.037)**	0.021				
Age	−0.002 (−0.033)**	0.001				
Urbanization	−0.097 (−0.142)**	0.011				
SES	0.047 (0.063)**	0.018				
Telephone	0.341 (0.139)**	0.031				
Ethnicity	−0.285 (−0.072)**	0.048	1.453 (0.250)**	0.054	−1.805 (−0.340)**	0.073

Note: dependent variable coded 1 = response, 0 = nonresponse; gender coded 1 = female, 0 = male; ethnicity coded 1 = non − Western foreigners, 0 = native population and Western foreigners.

**$p < 0.01$.

Chi square = 125.432 (df = 4); RMSEA = 0.046; CFI = 0.972, *R* square = 0.072.

Validation file: Chi square = 91.611 (df = 3); RMSEA = 0.045; CFI = 0.980, *R* square = 0.082.

Table 4. Structural equation model with dependent variable contact

	Explained variables					
	Estimates on contact	Standard errors	Estimates on urbanization	Standard Errors	Estimates on SES	Standard errors
Predictors						
Size of city			1.000 (0.874) fixed	0.000		
Urbanization of city			0.825 (0.914)**	0.011		
Urbanization of postal code			0.870 (0.887)**	0.015		
Value of home					1.000 (0.649) fixed	0.000
Social benefits					−0.368 (−0.489)**	0.020
Gender	0.060 (0.030)**	0.027				
Age	0.004 (0.056)**	0.001				
Urbanization	−0.183 (−0.267)**	0.012				
SES	−0.014 (−0.019) ns	0.020				
Telephone	0.386 (0.156)**	0.035				
Ethnicity	−0.215 (−0.054)**	0.052	1.453 (0.250)**	0.054	−1.806 (−0.340)**	0.073

Note: dependent variable coded 1 = contact, 0 = no contact; gender coded 1 = female, 0 = male; ethnicity coded 1 = non − Western foreigners, 0 = native population and Western foreigners.

**$p < 0.01$, ns = not significant.

Chi square = 123.290 (df = 4); RMSEA = 0.046; CFI = 0.972. *R* square = 0.104.

Validation file: Chi square = 90.715 (df = 3); RMSEA = 0.045; CFI = 0.980, *R* square = 0.113.

Table 5. Structural equation model with dependent variable co-operation

	Explained variables					
	Estimates on cooperation	Standard errors	Estimates on urbanization	Standard Errors	Estimates on SES	Standard errors
Predictors						
Size of city			1.000 (0.860) fixed	0.000		
Urbanization of city			0.856 (0.911)**	0.012		
Urbanization of postal code			0.896 (0.882)**	0.017		
Value of home					1.000 (0.623) fixed	0.000
Social Benefits					−0.400 (−0.501)**	0.025
Gender	0.070 (0.035) ns	0.024				
Age	−0.003 (−0.046)**	0.001				
Urbanization	−0.041 (−0.057)**	0.013				
SES	0.042 (0.054)*	0.021				
Telephone	0.257 (0.102)**	0.034				
Ethnicity	0.188 (0.045)**	0.058	1.324 (0.226)**	0.057	−1.715 (−0.321)**	0.082

Note: dependent variable coded 1 = co − operation, 0 = refusal; gender coded 1 = female, 0 = male; ethnicity coded 1 = non − Western foreigners, 0 = native population and Western foreigners.

*$p < 0.05$, **$p < 0.01$, ns = not significant.

Chi square = 177.247 (df = 6); RMSEA = 0.048; CFI = 0.949, *R* square = 0.025.

Validation file: Chi square = 142.455 (df = 5); RMSEA = 0.047; CFI = 0.956, *R* square = 0.033.

Table 6. Multiple group analysis on response

Predictors on response	Explained variables			
	Native population & Western foreigners		Non-Western foreigners	
	Estimates	Standard Errors	Estimates	Standard errors
Gender	0.063 (0.032)**	0.016	0.153 (0.076)**	0.058
Age**	−0.002 (−0.022)*	0.001	−0.008 (−0.092)*	0.002
Urbanization**	−0.085 (−0.127)**	0.006	−0.180 (−0.244)**	0.032
SES	0.021 (0.036)**	0.007	0.016 (0.027)ns	0.032
Telephone*	0.318 (0.129)**	0.020	0.173 (0.080)*	0.074

Note: gender coded 1 = female, 0 = male.

$*p - \text{value} < 0.05$,$**p - \text{value} < 0.01$, ns = not significant.

*after variable name reflects significant difference between groups for this variables at $p < 0.05$ level, $**p - \text{value} < 0.01$.

Chi square = 14.512 (df = 5) $p - \text{value} = 0.0126$; RMSEA = 0.012, CFI = 0.698; R square1 = 0.047, R square2 = 0.097.

Table 7. Multiple group analysis on contact

Predictors on contact	Explained variables			
	Native population & Western foreigners		Non-Western foreigners	
	Estimates	Standard Errors	Estimates	Standard Errors
Gender	0.076 (0.038)**	0.020	0.152 (0.075)*	0.085
Age*	0.006 (0.082)**	0.001	−0.002 (−0.019) ns	0.003
Urbanization**	−0.138 (−0.206)**	0.007	−0.247 (−0.336)**	0.036
SES	0.002 (0.004) ns	0.007	0.002 (0.004) ns	0.035
Telephone**	0.291 (0.118)**	0.022	0.198 (0.092)*	0.096

Note: gender coded 1 = female, 0 = male.
p − value < 0.05, **p − value < 0.01, ns = not significant.
*after variable name reflects significant difference between groups for this variables at $p < 0.05$ level, **p − value < 0.01.
Chi square = 13.198 (df = 5) p − value = 0.0215; RMSEA = 0.011, CFI = 0.739; R square1 = 0.075, R square2 = 0.143.

Table 8. Multiple group analysis on co-operation

Predictors on refusals	Explained variables			
	Native population & Western foreigners		Non-Western foreigners	
	Estimates	Standard Errors	Estimates	Standard Errors
Gender	0.044 (0.022)*	0.018	0.142 (0.071)*	0.072
Age*	−0.005 (−0.062)**	0.001	0.003 (0.043) ns	0.003
Urbanization*	−0.004 (−0.059)**	0.007	−0.136 (−0.179)**	0.050
SES	0.015 (0.028)*	0.006	−0.022 (−0.039) ns	0.048
Telephone	0.264 (0.104)**	0.022	0.185 (0.088) ns	0.096

Note: gender coded 1 = female, 0 = male.

*p − value < 0.05,**p − value < 0.01, ns = not significant.

*after variable name reflects significant difference between groups for this variables at $p < 0.05$ level, **p − value < 0.01.

Chi square = 2.804 (df = 2) p value = 0.2429; RMSEA = 0.006, CFI = 0.911; R square1 = 0.023, R square2 = 0.047.

The relation between ethnicity and response is almost entirely mediated by urbanization and the socio-economic status of the sampled units. A large amount of the negative effect of ethnicity on response is mediated by urbanization. The standardized probit regression of urbanization on response is -0.142 and the standardized effect of ethnicity on urbanization is 0.250. In particular, urbanization and telephone have a large effect on the response probability. The latent variable SES does not strongly affect the probability of responding. The total standardized effect of ethnicity on response is -0.129, which is the sum of the direct effect and all the indirect effects. The sum of all the unstandardized indirect effects is -0.227, which is the sum of the product of the unstandardized coefficients for the paths from ethnicity via SES to response and ethnicity via urbanization to response. The standard error for all the indirect effects of ethnicity on response is 0.025.

The sum of the direct and indirect effects of ethnicity ($-.129$) is still less than the direct effects of urbanization ($-.142$) and owning a landline telephone (.139). Ethnic minorities have lower response rates than the native population, but ethnic minorities also disproportionally live in urban areas. These results suggest that in particular, it is this urbanization effect that "causes" lower response rates among ethnic minorities, and not ethnicity itself. If controlled for other variables, ethnicity only has a small effect on the response probability and a fairly small standardized coefficient of ethnicity on response (-0.072) remains in this multivariate environment.

The socio-economic status of the sampled units barely affects the response probabilities, suggesting that the negative effect of ethnicity on response is mainly mediated by socio-demographic rather than socio-economic characteristics. This also holds true of the native population. Regardless of their ethnicity and socio-economic status, people who live in large cities have lower response probabilities than those who do not. The availability of a land-line proves to be a strong indicator for response. This not only holds true of the second part of the fieldwork period when the reapproaching strategies differ between a CATI mode for nonrespondents with a known telephone line, or otherwise CAPI, it also holds true of the first part of the fieldwork period when no telephone calls are made. Age and gender only have a small influence on response; older sampled units have slightly smaller response probabilities.

We not only inspect path coefficients to examine the effect of ethnicity on response, we also use the Schneekloth and Leeven (2003) method to evaluate nonresponse bias introduced in this case by ethnicity. This is done by examining the pseudo R square values of the total model. This value can serve as an indicator of the amount of nonresponse bias introduced while including background variables. If the model does not predict, or only poorly predicts, whether the sampled units will or will not respond, the nonresponse pattern can be seen as random and thus as following the basic logic of probability sampling. According to Andress et al. (1997), values below 0.05 indicate low and negligible correlation, and values above 0.2 indicate a strong correlation. Of course, the explanatory power is heavily dependent on the availability of information for the respondents and nonrespondents. However, as is noted above, the construction of the Social Statistical Databases at Statistics Netherlands provides detailed socio-demographic and socio-economic information.

The effect of ethnicity decreases substantially if the model controls for other variables. Not only does the path coefficient decrease if the indicator for ethnicity is included in the

analysis, the pseudo R square only increases by 1.0% (from 6.2% to 7.2%). Small path coefficients and a low pseudo R square suggest that predicting response is still fairly difficult, even if rich background information is available. This suggests that the effect of ethnicity on response and the response bias is not as high as some bivariate relationships seem to indicate. However, there may be a certain amount of selective nonresponse in urban areas. Urban residents, regardless of the ethnic group they belong to, are somewhat under-represented in this survey.

4.2.3. Contact

Fieldwork strategies that can successfully increase contact rates have been described in greater detail in recent years (Groves and Couper 1998; Bates 2004). The number of contact efforts and the time required to contact sampled units are among the factors noted to explain contactability. Unfortunately, this kind of fieldwork information is not available for this survey and thus cannot be included in the model examining the pattern of contacting the sampled units. For this reason, and for the comparability between the models, we use the same model to explain the contact process and in the following section the cooperation process. The results of the structural equation model on dependent variable contact defined as AAPOR contact definition number three are presented in Table 4.

In line with the argument formulated by Groves and Couper (1998), the role of urbanization is even more pronounced in the contact process. The standardized coefficient of urbanization on the contact probability is -0.267. The effect of ethnicity on contact is less than on the response rate, indicating that in this model, the effect of ethnicity is mediated even more strongly by urbanization. The unstandardized estimate for all the indirect effects from ethnicity to contact is -0.240 here, with a standard error of 0.029. Groves and Couper (1998) note that it is more difficult to establish contact with urban sampled units than nonurban sampled units.

These results show that the same is true of ethnic minorities. Nonresponse among ethnic minorities is heavily determined by low contact rates, which are not unique to ethnic minorities since they are largely mediated by urbanization. This also helps explain why ethnic minorities have lower response rates. Contact difficulties are mainly concentrated in urban areas. Nonrespondents with a known landline are reapproached in the second month with a CATI mode. Other nonrespondents are reapproached with a CAPI mode. However, due to a shortage of the interview staff, not all the nonrespondents without a known land-line are reapproached, or fewer contact efforts are made than in the case of nonrespondents in the CATI mode. Not surprisingly, the regression coefficient from telephone on the contact probability is high. Nonetheless, this positive effect of having a known land telephone on the contact probability is also found in the first month of the interview process when no telephone calls are made. Furthermore, the results show that women and the elderly are somewhat easier to contact, albeit with small probabilities.

4.2.4. Cooperation

To examine the process of cooperation (defined as AAPOR cooperation definition number two) we include only those sampled units who are contacted in the first place so

that 12,202 sampled units remain for the exploration file and 12,366 for the validation file. The results are presented in Table 5.

Somewhat surprisingly, ethnicity has a positive effect on the cooperation probability. Non-Western foreigners tend to refuse a request to participate in a survey less often than the other sampled units. Because of the low contact rates and the high nonresponse due to language problems among ethnic minorities, the results should be interpreted with caution. Sampled units who are not contacted do not have an opportunity to refuse a request to take part in a survey (Hox and de Leeuw 1998). Moreover, sampled units can use language problems as a friendly way to refuse to participate. The indirect effect of ethnicity on cooperation via SES and urbanization is negative, -0.127 (standard error 0.028), which also suggests that the positive effect of ethnicity on the cooperation rate should not be overinterpreted. The effect of urbanization, which is large in the response and contact process, is also lower, suggesting that there is a contact problem and not a participation problem in urban areas. The effect of SES is very small. Again, having a known land telephone proves to be a strong predictor. Older sampled units have a somewhat higher probability of refusing to participate. As indicated by a low pseudo R square value, the cooperation process is even more difficult to predict than the response and contact processes.

4.3. Do Response Models Differ Between Various Ethnic Groups?

In order to examine whether predictors for various response outcomes have the same influence for various ethnic groups, we perform a multiple group analysis with the same two groups. In multiple group analyses, the significance of the effect of group differences on model parameters can be tested by imposing cross-group equality constraints (Kline 1998). More general references on multiple group analysis can be found in Bollen (1989) and the Mplus technical appendix (www.statmodel.com). The native population and Western foreigners constitute the first group, and the non-Western foreigners the second. The results of the multiple group analysis with outcome variable response are presented in Table 6.

We perform our analysis on 26,479 native sampled units and Western foreigners (first group) and 1,893 non-Western foreigners (second group). Group membership moderates the relationship between having a known land telephone and response. The negative effect of urbanization on response probability is only slightly larger for the non-Western foreigners. Age has a somewhat larger negative effect on the response rate of the second group than on that of the first group. This reflects the higher nonresponse due to language problems, which is almost entirely found among older non-Western foreigners. In sum, the two groups do differ in their response process on the path coefficients age, urbanization, and having a known land-line.

Table 7 shows the results for the two groups on the dependent variable contact, which demonstrate that the response probability of the first group is more influenced than that of the second by having a known land telephone.

Again, the parameter indicator for having a known land telephone varies across groups. Contact probabilities of the first group are heavily influenced by this predictor. In the second group, the negative effect of urbanization is larger than in the first group. Nonetheless, urbanization has also a relatively large negative effect on the contact rate

among the native population and Western foreigners. Table 8 shows the results of the multiple group analysis with the outcome variable refusal.

For this multiple group analysis, where we examine the predictors for the sampled units who refuse to participate across the groups, we again only include the sampled units who are contacted. The selection of contacted units results in 23,210 remaining sampled units for the first group, and 1,359 non-Western foreigners with whom contact has been established remain for analysis. For the non-Western foreigners, urbanization has a somewhat more negative effect on the response outcome, in this case the refusal rate. Having a known land telephone again proves to be a strong positive predictor for the first group but does not have much effect on the second group.

5. Conclusions

Ethnic minorities are a growing part of Western societies, and are increasingly relevant for policy-makers. According to the Statistics Netherlands definition, almost 20% of the Dutch population have a foreign background and are called "allochtonous." Predictions are that the percentage of ethnic minorities in the Netherlands will increase to 35% by 2050 (De Jong and Hilderink 2004). With an increasing demand for data about ethnic minorities and decreasing response rates among them, more attention is devoted to the quality of the data about ethnic minorities. Nonresponse itself does not automatically imply bias in point estimates. However, nonresponse rates can serve as an indicator for potential bias problems. Nevertheless, reducing nonresponse should focus on reducing nonresponse error. Simply trying to increase response rates can actually increase the survey error (Merkle and Edelman 2002). Therefore it is important to know which societal groups have high nonresponse rates, so that tailored strategies can be developed to reduce nonresponse in these under-represented subgroups.

The analyses in this article are based on the results of the survey on living conditions in the Netherlands in 1998. Bivariate tables of response and ethnicity show large differences in the response rates between various ethnic groups. One of the most interesting findings is the high noncontact rate among ethnic minorities and more specifically among non-Western foreigners. Surprisingly, the cooperation rate among ethnic minorities is higher than among the native population. If sampled units are not contacted, of course it is impossible for them to refuse a request to participate. An increase in the minimum number of contact efforts in cases of earlier noncontact at Statistics Netherlands in March 2004 nonetheless shows a substantial increase in the contact and response rate among non-Western foreigners, but not in the refusal rate (Schmeets 2005b).

The results of the structural equation models show that the negative effect of ethnicity on response partially disappears if we control for other variables. In particular, urbanization has a strong effect on nonresponse. However, ethnicity still affects the response probability. Knowing that nonresponse among ethnic minorities is especially found in highly urbanized areas nevertheless enables the researcher to focus on this very specific group. Of course there is a high correlation between urban areas and ethnic minorities, but in the Netherlands about 70% of the population of foreign descent and 70% of the non-Western foreigners do *not* live in one of the four major cities. Urbanization has been related to survey nonresponse and more specifically nonresponse due to noncontact

for many years (Groves and Couper 1998; Steeh et al. 2001). Indeed, one of the reasons for higher nonresponse rates in the Netherlands than most other countries might be the urban nature of the western part of the country.

Survey response remains a process greatly influenced by chance. Some groups, like ethnic minorities, do have lower response rates. But, fortunately, even with the inclusion of background information, it is extremely difficult to predict whether a specific person will respond or not. Although a great deal of administrative information is available, a low pseudo R square value (0.072) suggests that the nonresponse is not very systematic. In particular, the refusal process seems to be greatly influenced by chance, as indicated by even lower pseudo R square values. Although regression coefficients and R square are small, the focus should be on enlarging the response in urban areas, where there is a serious contact problem. Tailoring data collection strategies for higher contact rates will lead to increased response rates among ethnic minorities.

To sum up, urban sampled units are more difficult to contact and more contact efforts are needed. This holds particularly true of non-Western urban sampled units. However, this extra negative effect is limited. Steeh et al. (2001) observe a trend of increasing noncontact rates in some American metropolitan areas, and urge survey methodologists to plan for this eventuality. To keep fieldwork procedures at a manageable level, a tailored fieldwork strategy might be conceivable, involving all the urban sampled units receiving more contact efforts. In the Netherlands, the minimum number of contact efforts by earlier noncontact was changed from three to six in March 2004[4] (Snijkers and Kockelkoren 2004). This has had a very positive effect on the contact and response rates, in particular among ethnic minorities (see Feskens et al. 2006; Schmeets 2005b). This supports our findings on the contact problem among ethnic minorities. Increasing contact efforts will decrease the nonresponse among ethnic minorities. However, the question still remains as to the kind of nonrespondents there will be after a change in the fieldwork procedure. Future research could focus on this.

The multiple group analysis results show that ethnic groups are not homogeneous in their response processes. Age and urbanization have a more negative effect on the response and contact probability of non-Western foreigners than of the native population and Western foreigners. However, having a known land telephone proves to be a very strong indicator for responding among the native population and Western foreigners as well as non-Western foreigners. This positive effect is much smaller among non-Western foreigners.

This study has a number of limitations. Although the Social Statistical Database contains a rich amount of data on respondents and nonrespondents, not all the theoretical considerations can be taken into account. And since non-Western foreigners are only about 8% of the population in the 15–65 age-group they might not have much effect on the overall estimates. Although they may not have much effect on the overall estimates, better response rates among ethnic minorities may still be needed to obtain better estimates on the subpopulations. In addition, since these results are based on a Dutch survey, they can

[4] It is also the maximum number, except for appointments in the sixth attempt. In that case a seventh attempt is allowed.

only be partially generalized to other countries. Future research could focus on these limitations and use our findings as hypotheses to study in further detail. Nevertheless, our analysis outcomes suggest that although ethnic minorities have lower response rates, the focus should be on enlarging the response in urban areas. Ethnic minorities do not respond as well as the native population, but the explanations and hence the solutions have less to do with divergent response behaviour among ethnic minorities, and more to do with living conditions. Additional efforts should be made to increase the contact rate in urban areas. A possible solution can be to approach the sampled units with specially tailored strategies, e.g., a higher minimal number of contact efforts than for nonurban sampled units.

Appendix

American Association for Public Opinion Research (2006), *Standard Definitions*[5].

Response rate 2 $= \frac{(\mathbf{I}+\mathbf{P})}{(\mathbf{I}+\mathbf{P})+(\mathbf{R}+\mathbf{NC}+\mathbf{O})+(\mathbf{UH}+\mathbf{UO})}$

Contact rate 3 $= \frac{(\mathbf{I}+\mathbf{P})+\mathbf{R}+\mathbf{O}}{(\mathbf{I}+\mathbf{P})+\mathbf{R}+\mathbf{O}+\mathbf{NC}}$

Cooperation rate 2 $= \frac{(\mathbf{I}+\mathbf{P})}{(\mathbf{I}+\mathbf{P})+\mathbf{R}+\mathbf{O}}$

Variables

Sex; sex of the sampled unit
0 = male
1 = female

Age; age of the sampled unit
15–65 years

Size of city; size of the community where the sampled unit is registered
1 = small
8 = large

Urbanization of community; urbanization of community where the sampled unit is registered
1 none
2 weak
3 moderate
4 strong
5 very strong

Urbanization at postal code level; urbanization at postal code level of the address where the sampled unit is registered
1 $<$ 500 addresses per square km
2 500– $<$ 1,000 addresses per square km
3 1,000– $<$ 1,500 addresses per square km
4 1,500– $<$ 2,500 addresses per square km
5 $>$ 2,500 addresses per square km

[5] **I** = Complete interview **P** = Partial interview **R** = Refusal and break-off **NC** = Non − contact **O** = Other **UH** = Unknown if household/occupied **HU UO** = Unknown, other.

Indicator for telephone; Does the sampled unit have a known registered land telephone?
0 = no known registered land telephone
1 = known registered land telephone
Response; Did the sampled unit respond (partially)?
0 = nonresponse
1 = response
Value of home; value of the home in Dutch guilders where the sampled unit is registered
1 < 50 thousand
2 50–75 thousand
3 75–100 thousand
4 100–125 thousand
5 125–150 thousand
6 150–200 thousand
7 200–250 thousand
8 250–300 thousand
9 300–350 thousand
10 350–400 thousand
11 400–500 thousand
12 > 500 thousand
Indicator for social benefits; Does the sampled unit receive some form of social benefit?
0 = no
1 = yes
Ethnicity; Is the sampled unit a non-Western foreigner?
0 = no
1 = yes
Contact; Was contact established with the sampled unit during the fieldwork period?
0 = no, no contact
1 = yes, contact
Refusals; Did the sampled unit refuse to participate in the survey?
0 = sampled unit refused to provide requested information
1 = sampled unit cooperated

6. References

American Association for Public Opinion Research (2006). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 4th Edition. Lenexa, KS: AAPOR.

Andress, H.J., Hagenaars, J.A., and Kühnel, S. (1997). Analyse von Tabellen und kategorialen Daten. Berlin: Springer. [In German]

Bates, N.A. (2004). Contact Histories: A Tool for Understanding Attrition in Panel Surveys. Paper presented at the 59th Annual AAPOR Conference, May 11–13.

Bentler, P.M. (1990). Comparative Fit Indexes in Structural Models. Psychological Bulletin, 107, 238–246.

Bollen, K.A. (1989). Structural Equations with Latent Variables. New York: Wiley.

Browne, M.W. and Cudeck, R. (1993). Alternative Ways of Assessing Model Fit. In J.S. Long (ed.). Testing Structural Equation Models, 136–162. Newbury Park, CA: Sage.

Couper, M.P. and de Leeuw, E.D. (2003). Nonresponse in Cross-Cultural and Cross-National Surveys. In J.A. Harkness, F.J.R. v.d. Vijver and P.P. Mohler (eds). Cross-Cultural Survey Methods, 157–178. New York: Wiley.

De Heer, W. (1999). International Response Trends: Results of an International Survey. Journal of Official Statistics, 15, 129–142.

De Heer, W. and de Leeuw, E.D. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In Survey Nonresponse, R. Groves et al. (eds). New York: John Wiley.

De Jong, A. and Hilderink, H. (2004). Lange Termijn Allochtonenscenario's voor Nederland, Bevolkingstrends, 52, 77–82. [In Dutch]

Eisner, M. and Ribeaud, D. (2007). Conducting a Criminological Survey in a Culturally Diverse Context: Lessons from the Zurich Project on the Social Development of Children. European Journal of Criminology, 4, 271–298.

Feskens, R.C.W., Hox, J., Lensvelt-Mulders, G., and Schmeets, H. (2006). Collecting Data among Ethnic Minorities in an International Perspective. Field Methods. 18, 284–304.

Goyder, J., Lock, J., and McNair, T. (1992). Urbanization Effects on Survey Nonresponse: A Test within and across Cities. Quality and Quantity, 26, 39–48.

Groves, R.M. (1989). Survey Errors and Survey Costs. New York: Wiley.

Groves, R.M. and Couper, M.P. (1998). Nonresponse in Household Interview Surveys. New York: Wiley.

Harkness, J.A., Vijver, F.J.R.V.d., and Mohler, P.P. (eds) (2003). Cross-Cultural Survey Methods. New York: John Wiley.

Houbiers, M. (2004). Towards a Social Statistical Database and Unified Estimates at Statistics Netherlands, Journal of Official Statistics, 20, 55–75.

Hox, J.J. and de Leeuw, E.D. (1998). Nonrespons in Surveys: Een Overzicht. Kwantitatieve Methoden, 19, 31–53.

Kline, R.B. (1998). Principles and Practice of Structural Equation Modeling. New York: Guilford Press.

Kline, R.B. (2005). Principles and Practice of Structural Equation Modeling. New York: Guilford Press.

Lavrakas, P.J. (1993). Telephone Survey Methods: Sampling, Selection and Supervision (2nd edition). Newbury Park, CA: Sage.

Merkle, D.M. and Edelman, M. (2002). Nonresponse in Exit Poll: A Comprehensive Analysis. In Survey Nonresponse, R. Groves et al. (eds). New York: John Wiley.

Muthén, B.O. (1998-2004). Mplus Technical Appendices. Los Angeles, CA: Muthén and Muthén.

Muthén, L.K. and Muthén, B.O. (1998-2004). Mplus Users's Guide (3rd edition). Los Angeles, CA: Muthén and Muthén.

Reep, C. (2003). Moeilijk Waarneembare Groepen. Een Inventarisatie. CBS-report H1568-03-SOO, Heerlen: Statistics Netherlands. [In Dutch]

Schmeets, H. (2005a). De Leefsituatie van Allochtonen. In Over het Omgaan met Allochtonen in Enquêteonderzoek, H. Schmeets and R. van der Bie (eds). Heerlen: Statistics Netherlands. [In Dutch]

Schmeets, H. (2005b). De Invloed van het Aantal Contactpogingen op de Respons van Allochtonen. In Over het Omgaan met allochtonen in enquêteonderzoek, H. Schmeets and R. van der Bie (eds). Heerlen: Statistics Netherlands. [In Dutch]

Schmeets, H. and Michiels, J. (2003). Het Effect van Non-respons onder Allochtonen, Bevolkingstrends, 51, 52–57. [In Dutch]

Schneekloth, U. and Leeven, I. (2003). Woran bemisst sich eine "gute" allgemeine Bevölkerungsumfrage? Analysen zu Ausmaß, Bedeutung und zu den Hintergründen von Nonresponse in zufallsbasierten Stichprobenerhebungen am Beispiel des ALLBUS. ZUMA-Nachrichten, 53, 27, 16–57. [In German]

Schouten, B. (2003). De verschillende Typen Non-respons in POLS 1998. CBS-report 2319-03-TMO. Heerlen: Statistics Netherlands. [In Dutch]

Snijkers, G. and Kockelkoren, S. (2004). De gestandaardiseerde Benaderingsstrategie 2003: Evaluatie en Aanbevelingen. CBS-report H415-04-SOO, Heerlen: Statistics Netherlands. [In Dutch]

Steeh, C., Kirgis, N., Cannon, B., and DeWitt, J. (2001). Are They Really as Bad as They Seem? Nonresponse Rates at the End of the Twentieth Century. Journal of Official Statistics, 17, 227–247.

Stoop, I.A.L. (2004). Surveying Nonrespondents. Field Methods, 16, 23–54.

Thornberry, O.T.Jr. and Massey, J.T. (1988). Trends in United States Telephone Coverage across Time and Subgroups. In Telephone Survey Methodology, R. Groves et al. (eds). New York: John Wiley and Sons.

van Goor, H., Jansma, F., and Veenstra, R. (2005). Differences in Undercoverage and Nonresponse between City Neighbourhoods in a Telephone Survey. Psychological Reports, 96, 867–878.

Received June 2005
Revised March 2007