A Wave-Analysis of Distributional Bias, Substantive Bias and Data Quality in a Mail Survey Among Dutch Municipalities

Goor, H. van; Stuiver, B.

Mudde, C.E. (1995), 'Right-Wing Extremism Analyzed. A Comparative Analysis of the Ideologies of Three Alleged Right-Wing Extremist Parties (NPD, NDP, CP'86)', *European Journal of Political Research* 27(2), pp. 203-24.

Mudde, C.E. (1998), *The Extreme Right Party Family: An Ideological Approach.* Dissertation Leiden: University of Leiden.

Mudde, C.E. and J.J.M. Van Holsteyn (1994), 'Over the Top: Dutch Right-Wing Extremist Parties in the Elections of 1994', *Politics* 14(3), pp. 127-134.

Pennings, P. and K. Brants (1985), *Rechts-Extreme Jongeren. Theorie en Praktijk van Politiek Gedrag.* Amsterdam: Vakgroep Collectief Politiek Gedrag.

Pfahl-Traughber, A. (1993), *Rechtsextremismus. Eine Kritische Bestandsaufnahme nach der Wiedervereinigung.* Bonn: Bouvier Verlag.

Stöss, R. (1994), 'Forschungs- und Erklärungsansätze – ein Überblick', in: W. Kowalsky and W. Schroeder (eds.), *Rechtsextremismus. Einführung und Forschungsbilanz.* Opladen: Westdeutscher Verlag.

Sumner, W.G. (1940), *Folkways. A Study of the Sociological Importance of Usages, Manners, Customs, Mores, and Morals.* Boston: Ginn and Co. [originally published in 1905].

Ueltzhöffer, J. (1983), 'Rechtsextremismus', in: D. Nohlen (ed.), *Pipers Wörterbuch zur Politik.* München: Piper, Teil 2.

Van der Eijk, C. and G. Irwin (1988), 'Survey. Het Nationaal Kiezersonderzoek', in: W. Derksen, A.F.A. Korsten and A.F.M. Bertrand (eds.), *De Praktijk van Onderzoek. Problemen bij Onderzoek van Politiek, Bestuur en Beleid.* Groningen: Wolters-Noordhoff.

Van Holsteyn, J. and C. Mudde (1992a), *Op Zoek naar Rechtsextremisme. Opmerkingen naar Aanleiding van een in en om Haarlem Gehouden Onderzoek.* Leiden: Vakgroep Politieke Wetenschappen.

Van Holsteyn, J.J.M. and C.E. Mudde (1992b), 'Voorzichtig Omspringen met Begrip Rechtsextremisme!', *Namens* 7(4), pp. 19-23.

Research Note

A Wave-Analysis of Distributional Bias, Substantive Bias and Data Quality in a Mail Survey among Dutch Municipalities

H. van Goor and B. Stuiver
University of Groningen

Abstract

This paper discusses a study of the consequences of enhancing the response rate for distributional bias, substantive bias and data quality in a mail survey among Dutch municipal officials on the local implementation of a national caravan sites policy. We found both distributions and relationships to be seriously biased at a response level of 46%. Higher response levels (62% and 74%) led to better results: distributional and substantive biases decrease as the response rate increases. However, even at a final response rate of 74%, distributions and relationships are still biased to some extent. This is due to the fact that the remaining non-respondents are composed more and more of specific, 'extreme' subgroups. There are indications that the quality of the data deteriorates somewhat at later waves. However, the effects are not so damaging that response-enhancement is unwarranted. Nevertheless, data quality remains an important topic of concern, especially when trying to enhance response rates even more to obtain almost complete response.

1 Introduction

Survey response has been declining for decades (Goyder 1987; Hox and De Leeuw 1994). Falling response is a cause for concern because results may be no longer representative or valid due to non-response bias. Non-response bias is a function of both response rate and differences between respondents and non-respondents. Most attention has been directed to response rates, because, almost by definition, little is known about non-respondents, especially with respect to their attitudes and behaviour.

In all likelihood, most survey researchers are of the opinion that the higher the response rate the better the survey: surveys with higher response rates are considered more representative and more valid than surveys with lower response rates (Groves 1987: S161; Dillman 1991: 229). This assumes that the higher the response rate the smaller the differences will be between respondents and non-respondents. Therefore, as the response rate increases, distributional and substantive biases will diminish.

Some researchers, however, maintain that the survey method is so robust that results are unbiased even when response rates are quite low. As Bradburn (1992: 393-394) observed, this opinion is particularly widespread in applied and commercial survey research. Additionally, some claim that response-enhancement may be harmful because reluctant respondents will give low-quality answers. If true, this means that in trying hard to reduce non-response errors, we run the risk of enhancing response errors (cf. Groves 1987: S168).

On the other hand, we also find survey researchers endorsing very high response rates, approximating 100 per cent. For example, Bebbington (1970) stated that, with increasing response, refusers will form more and more a deviant subcategory. Hence, valid outcomes can only be gained by very high response rates, approximating 100 per cent.

Because some loss of respondents is almost inevitable due to inaccessibility or refusal, the question survey researchers face in actual research is whether the response rates realized nowadays are still high enough to warrant unbiased outcomes. Or to broaden the question: are high response rates indeed necessary, and how high is high enough? Are efforts to increase response rates well spent, and does response enhancement not have other drawbacks, such as a deterioration of the quality of the data?

In this article, we will examine these questions in a systematic way by conducting a wave analysis. The data have been taken from a mail survey among Dutch municipalities on the local implementation of a national policy on caravan sites. A key informant in each municipality completed the questionnaire. Research on non-response is often confined to descriptive, mostly socio-demographic variables, but we are fortunate to have information from independent sources on the substantive or dependent variable central to our study as well. Therefore, we are not only able to study the differences between respondents and non-respondents across cumulative waves and their consequences for distributional bias, but also for the relationships between variables (substantive bias). In addition, we will examine whether response-enhancement affects data quality negatively.

2 Hypotheses

To test the consequences of non-response for *distributional bias*, we are going to proceed in two ways. First, we will compare the interim distributions after each successive, cumulative wave with the final distribution of the complete sample. Second, we will compare the differences between respondents and non-respondents for each successive, cumulative wave.

Most authors report initial distributional biases, when comparing the first wave with their final response rate. In addition, most claim that biases become smaller with successive, cumulated waves (e.g. Pavalko and Lutterman 1973; Hawkins 1975; Goudy 1976; Fitzgerald and Fuller 1982; Dolsen and Machlis 1991). Therefore, we expect to find clearly biased distributions after wave 1. When the response rate increases (waves 1 + 2, and 1 + 2 + 3), we expect the biases to diminish steadily and progressively.

H1: *Distributional bias: Comparing interim and final distributions*
At a low response level (in our study 46%), distributions will be biased, but the biases will decrease steadily when the response rate increases (in our study to 62% and then to 74%).

Two rival hypotheses can be formulated about the relationship between response-enhancement and differences between respondents and non-respondents. On the one hand, attempts to increase the response rate may result in persuading initially reluctant categories of respondents to participate as yet in the survey. As a consequence, response-enhancement leads to decreasing differences between respondents and non-respondents. On the other hand, Bebbington maintained that hard core refusers differ more and more from both early and late respondents. If that is true, response-enhancement only causes differences between respondents and non-respondents to increase.

H2: *Distributional bias: Comparing respondents and non-respondents across cumulative waves*
a. Respondents and non-respondents differ at low response rates, but these differences decrease steadily as the response rate increases.
b. Initial differences between respondents and non-respondents will increase or fluctuate randomly, as the response rate increases.

Little work has been done on *substantive bias*. However, available research suggests that substantive biases exist between successive cumulative waves (Hawkins 1975; Goudy 1976, 1978; Fitzgerald and Fuller 1982). Clear and stable patterns of unbiased outcomes seem to emerge only at high response levels.

We are going to test in three different but related ways whether response rates affect substantive results: (1) conventional null-hypothesis testing; (2) relative non-response bias in correlation coefficients; and (3) multiple regression.

H3: *Substantive bias*
At a low response rate, substantive outcomes will be seriously biased; when response levels increase, substantive outcomes will become increasingly less biased.

Different types of survey errors may be not only interrelated, but, more seriously, they may be at cross purposes. In particular, it is important to know whether response-enhancement leads to a deterioration of the quality of the data. Therefore, we will examine whether response-enhancement leads to an increase in item non-response and to a deterioration of our policy process data. Results to date are mixed. While Donald (1960) found that 'no answer' or 'don't know' are more frequent among later-wave respondents, others failed to do so (e.g. Green 1991). Unlike before, we are comparing the quality of the data of separate waves, not across cumulative waves.

H4 *Data quality*
Item non-response will increase and the quality of the policy process data will deteriorate with each successive wave of respondents.

3 Data

The data have been taken from a study of the local implementation of a national policy towards caravan dwellers in the Netherlands. In all, 286 'eligible' municipalities were included (see for more details on study design and data: Van Goor and Stuiver forthcoming).

On January 27, 1987, the questionnaire with a covering letter (on our department letterhead) and a prepaid return envelope was mailed to the municipalities. A handwritten identification number was clearly visible on the cover. On March 11, we sent a second letter, reminding those municipalities that had not yet responded about the questionnaire. New questionnaires were sent on request. From April 16, the remaining municipalities were contacted by phone, at least twice if necessary. The last questionnaires were returned in July.

After the data gathering had been completed, we realized that the postal services had delivered one of the questionnaires to the wrong municipality (with an almost identical name), and that the latter had returned the questionnaire. So, in effect, we failed to contact just one municipality. Because we had been able to contact key-persons in all other eligible municipalities, non-response in our study is unambiguously made up of refusals. Therefore, ambiguities resulting from the widespread inability in mail surveys to differentiate between access-failure and refusals do not play a role in our research.

Of the 285 eligible and contacted municipalities, 71 failed to return the questionnaire. Two questionnaires were returned unanswered. Four more were sent back, whereby only a few questions had been answered, and the key-person indicated that he was unable or unwilling to answer the remaining questions. These cases were also counted as refusals (in all, 77 or 27%). The

remaining cases, totalling 208 (73.0%), were classified as respondents: 127 (44.6%) replied to the first mailing, 47 (16.5%) to the second, and another 34 (11.9%) returned the questionnaire after one or more telephone calls.

Apart from the questionnaire, we gathered provincial documents on the original policy plans and provincial reports on the progress made by the municipalities in executing the policy. In this way, we could establish, independently of the survey, the exact assignments and the extent and speed of the implementation of the plan targets of almost all municipalities involved.[1] In the analyses presented here, we used the provincial data on number of caravan dwellers residing in a municipality in 1978, task assignment and policy performance for both responding and non-responding municipalities.

Data on municipal size (number of inhabitants in 1986), party politics (percentage of municipal council seats occupied by left-wing parties after the 1986 municipal elections), and urbanization were taken from statistical sources published by the Netherlands Central Statistics Office (Centraal Bureau voor de Statistiek). Data on centre-periphery (a composite index measuring differences in socio-economic development between regions) were taken from Van Engelsdorp Gastelaars et al. (1980).

4 Results

4.1 Distributional bias

Many researchers use means to compare waves of respondents, but as Bebbington (1970: 172) noted, the distribution of a variable may be truncated at both extremes without the mean of the distribution being seriously affected. It is not clear whether such a pattern can be encountered frequently, but to be certain we need to examine the frequency distribution in some detail when looking for distributional bias.

Table 1 gives the cumulative distribution of eight variables, six 'independent' variables (size, urbanization, party politics, centre-periphery, number of caravan dwellers in 1978, and task assigned), and two 'dependent' variables measuring the effectiveness of the policy implementation (policy performance, and success of realization of first caravan site).[2] Because we want to know to what extent each single category of a variable is *biased* (overrepresented or underrepresented) on successive cumulative waves (in comparison to the relative size of this category in the final distribution), we calculated the *percentage errors* between successive cumulative waves and the total distribution for all categories of our variables.[3]

Because we are going to use these same eight variables again in our analysis of relationships, we confined the distributional analysis to the 257 municipalities

Table 1 Distributions of characteristics of responding municipalities by successive cumulative waves

Characteristics	Wave 1	1 + 2	1 + 2 + 3	All municipalities	Percentage error W1 vs All	W1+2 vs All	W1+2+3 vs All
Municipal size (number of inhabitants x 1000)							
< 5	12.0	9.4	8.5	10.1	+19	-7	-16
5 - 10	29.9	27.7	28.0	25.3	+18	+10	+11
10 - 20	34.2	36.5	35.4	32.7	+5	+12	+8
20 - 50	19.7	20.8	22.8	23.7	-17	-12	-4
> 50	4.3	5.7	5.3	8.2	-48	-31	-35
(*N*)	(117)	(159)	(189)	(257)			
Urbanization							
Agrarian municipalities	26.5	27.7	28.6	27.6	-4	+0	+4
Suburbanized municipalities	35.0	34.0	21.2	30.4	+15	+12	+3
Commuter municipalities	20.5	22.0	23.3	21.4	-4	+3	+9
Small towns	14.5	12.6	13.2	13.2	+10	-5	0
Large towns	3.4	3.8	3.7	7.4	-54	-49	-50
(*N*)	(117)	(159)	(189)	(257)			
Party politics (percentage left-wing councillors)							
0	6.0	5.7	5.3	6.6	-9	-14	-20
6 - 24	35.9	34.0	33.3	331.1	+15	+9	+7
25 - 39	40.2	37.1	38.6	38.1	+6	-3	+1
40 - 62	17.9	23.3	22.8	24.1	-26	-3	-5
(*N*)	(117)	(159)	(189)	(257)			
Centre – periphery							
Central region	18.8	18.9	18.5	18.7	+1	+1	-1
Intermediate region	33.3	34.0	34.9	33.5	-1	+2	+4
Semi-periphery	21.4	20.8	22.2	25.3	-15	-18	-12
Periphery	26.5	26.4	24.3	22.6	+17	+17	+8
(*N*)	(117)	(159)	(189)	(257)			
Caravans in 1978							
0	79.5	78.0	75.7	72.8	+9	+7	+4
1 - 5	10.3	10.1	11.6	12.1	-15	-17	-4
6 - 20	4.3	6.9	7.9	8.6	-50	-20	-8
21 - 155	6.0	5.0	4.8	6.6	-9	-24	-27
(*N*)	(117)	(159)	(189)	(257)			
Number of caravans assigned for accommodation							
1 - 3	29.1	27.0	24.9	23.3	+25	+16	+7
4 - 6	29.9	30.8	31.7	29.6	+1	+4	+7
≥ 7	41.0	42.1	43.4	47.1	-13	-11	-8
(*N*)	(117)	(159)	(189)	(257)			
Policy performance							
Early completion	26.3	23.4	23.0	26.0	+1	-10	-12
Recently completed	28.9	29.2	28.4	24.8	+17	+18	+15
Partly completed	21.9	23.4	24.0	21.5	+2	+9	+12
Nothing completed	22.8	24.0	24.6	27.6	-17	-13	-11
(*N*)	(114)	(154)	(183)	(246)			
Success first caravan site							
Early success	33.6	30.1	30.8	32.5	+3	-7	-5
Recently successful	31.9	31.4	31.9	28.0	+14	+12	+14
Unsuccessful	34.5	38.6	37.4	39.4	-12	-2	-5
(*N*)	(113)	(153)	(182)	(246)			

for which policy performance was a relevant variable, that is, those which had to enlarge an existing site or to build one or more new sites. The remaining municipalities only had to preserve an already existing site (24 cases), or had an unknown assignment (4 cases; cf. note 1). The cumulative response rates for these 257 municipalities are 45.5%, 61.9%, and 73.5% respectively.

As Table 1 demonstrates, some distributional bias can be observed at the first wave for all variables. The percentage errors show that some categories are initially strongly overrepresented, while others are underrepresented. For example, small municipalities were quick in returning the questionnaire, while large municipalities, especially large towns, were laggards.

What happens when the response rate increases? When we compare the percentage errors at the first wave with those at the three cumulated waves (the total response), percentage errors diminished in 20 out of 32 categories. Thus, an increase in response rate from 46 to 74% results, in most cases, in a convergence on the final distribution. Nevertheless, over one fourth of the categories show an increase in percentage errors. Furthermore, twelve categories still have percentage errors of more then 10 per cent at the final response rate. We thus conclude that a higher response rate (74%) is better than a lower one

(46%), but also that biases have not disappeared completely. Our results support Hypothesis 1.

At the beginning of this section, we questioned the use of means in comparing frequency distributions across waves. When we look at our variables, a distribution truncated at both extremes can be observed at least from the second wave on for municipal size, party politics, policy performance, and success of first site, that is, with four of the eight variables. It is noteworthy that both of our outcome variables are biased in this way.

To test Hypothesis 2, that is whether differences between respondents and non-respondents increase or in fact decrease, we compared, first of all, the respondents to the first wave to all other municipalities (non-respondents to the first wave); then the respondents to the first two waves to the remaining municipalities; and last, the respondents to all three waves to the remaining, final non-responding municipalities. We compared the distributions of the eight variables analysed before, using χ^2 to test for differences between the distributions.

Table 2 Differences in distribution of characteristics between responding and non-responding municipalites by successive cumulative waves (χ^2 test)

Characteristics		Wave 1 respondents *vs* all other municipalities	Wave 1+2 respondents *vs* all other municipalities	Wave 1+2+3 respondents *vs* final non-responding municipalities
Municipal size	χ^2	7.41	3.95	9.36
	p	.12	.41	.05
Urbanization	χ^2	6.07	6.66	12.69
	p	.19	.16	.01
Party politics	χ^2	3.78	1.33	3.64
	p	.29	.72	.30
Centre – periphery	χ^2	3.49	.52	4.83
	p	.32	.04	.19
Caravans 1978	χ^2	8.73	7.60	4.26
	p	.03	.06	.24
New task	χ^2	5.27	4.93	3.98
	p	.15	.18	.26
Policy performance	χ^2	3.30	6.94	10.84
	p	.35	.07	.01
Success first site	χ^2	2.44	2.42	5.06
	p	.30	.30	.08

Note: for *N* see Table 1.

Table 2 presents the summary outcomes. Four variables (size, urbanization, centre-periphery, policy performance) show a pattern of increasing differences, reaching significance after three cumulative waves ($p \leq .10$). Initial differences disappear in one case (caravans 1978). The remaining three variables show no clear pattern of increasing or decreasing differences. Our results best fit Bebbington's contention (Hypothesis 2b): remaining non-respondents become more and more a deviant category. Moreover, more detailed analyses of the joint distribution of several combinations of an independent and a dependent variable reveal that several 'extreme' categories are particularly overrepresented among the final non-respondents, such as large, less successful municipalities and also, to a lesser degree, small, successful municipalities.

4.2 Substantive bias

The possibilities of examining substantive bias are limited in our study. In our substantive analyses, we are especially interested in the way size and party politics affect the way municipal officials handle the policy implementation, and consequently, affect policy performance. However, information on the strategies used by municipalities to carry out the tasks assigned to them, and the way higher authorities and the local population react to these strategies, is only available for the municipalities which responded. For this reason, we have restricted our analysis to the relations between the eight variables used before: six 'independent' and two 'dependent' variables.[4]

1. What happens if we test the bivariate relations between the independent and dependent variables for statistical significance on successive cumulative waves? Although conventional null hypothesis significance testing has been severely critized for many years (see for a recent critique Cohen 1994), its widespread use among researchers warrants paying attention to the consequences of different response rates for null hypothesis testing.

Taking $\alpha \leq .05$ (two-tailed) as our criterion, we draw the wrong conclusions in five of the twelve relationships between independent and dependent variables in wave 1 and wave 1 + 2, and in four cases in wave 1 + 2 + 3, compared to the results for all municipalities (Table 3). In addition, our results show that stability of the results of significance tests across cumulative waves is no sure guarantee that these results are correct.

2. To test the differences between the correlation coefficients across successive cumulative waves in a more systematic way, we devised a measure of *Relative Non-Response Error* (RNE), by transforming *r*-values into *z*-scores.

Table 3 Product-moment correlation coefficients by successive cumulative waves

Relationships	Wave 1	1 + 2	1 + 2 + 3	All municipalities
Size x policy performance	.015	.052	.068	.156*
Size x success first site	-.254*	-.225*	-.192*	-.094
Urbanization x policy performance	-.031	.000	.055	.146
Urbanization x success first site	-.244*	-.217*	-.120	-.035
Party politics x policy performance	.055	.103	.100	.157*
Party politics x success first site	-.048	-.017	.006	.157*
Centre-periphery x policy performance	.056	.087	.064	-.005
Centre-periphery x success first site	.015	.054	-.014	-.040
Caravans 1978 x policy performance	.077	.096	.089	.164*
Caravans 1978 x success first site	-.158	-.138	-.113	.005
New task x policy performance	.098	.119	.120	.139
New task x success first site	-.230*	-.215*	-.215*	-.162*

Note: * p < .05 (two-tailed).

$$\text{RNE} = \frac{|Z_o - Z_v|}{s}$$

Z_o = the *z*-score of an observed correlation coefficient on a particular cumulative wave;

Z_v = the *z*-score of the correlation coefficient for all municipalities, used for validating purposes;

$$s = \frac{1}{\sqrt{N-3}}$$

RNE represents the deviation from the 'true' or 'validating' correlation coefficient in standard-deviation units.[5] In theory, RNE may be positive or negative, but we ignore the direction of the bias here. This measure is attractive, because the outcomes can be interpreted in terms of confidence intervals (see Table 4). We use RNE here as a measure of *agreement* between the observed and the validating values of the correlation coefficients to test whether bias increases or rather decreases across cumulative waves. Part A of Table 5 presents RNE for all 28 relationships between our eight variables, while Part B presents RNE for the relationships between independent and dependent variables only.

Table 4 Relation between amount of agreement and relative non-response error (*RNE*) in standard-deviation units with corresponding confidence intervals

Amount of agreement	*RNE* (in *s*)	Confidence intervals (approximately)
High	< .500	< 38%
Substantial	1.000 .501	68 - 38%
Moderate	1.650 1.001	90 - 68%
Poor	1.960 1.651	95 - 90%
Very poor	> 1.960	> 95%

As Table 5, part A, shows, 9 out of 28 coefficients fall outside the 95%-confidence interval in the first wave, that is almost one third. Across waves, the number of coefficients outside the 95%-confidence interval diminish from nine (wave 1) via seven (wave 1 + 2) to one (wave 1 + 2 + 3). The number of coefficients falling within the 95-90% and 90-68%-intervals show some increase, while the number of coefficients in the 68-38%-interval and within the 38%-interval are more or less stable. Across cumulative waves, the most extreme deviations from the 'true' value (outside the 95% confidence interval) almost disappear. On the other hand we see no increase within the 38%

Table 5 Relative non-response error (RNE) by successive cumulative waves

Cumulative waves	Deviations from validating scores in standard deviation units*					χ^2	*p*
	> 1.960 (> 95%)	1.960-1.651 (95-90%)	1.650-1.001 (90-68%)	1.000-.500 (68-38%)	< .500 (< 38%)	(df = 3)**	
A. All relationships (*N* = 28)							
Wave 1	9	1	8	6	4	23.9	< .001
Wave 1+2	7	1	10	7	3	13.1	< .02
Wave 1+2+3	1	4	12	6	5	10.8	< .02
Expected (normal) distribution	1.4	1.4	6.2	8.4	10.6		
B. Only relationships between independent and dependent variables (*N* = 12)							
Wave 1	5	0	4	3	0		
Wave 1+2	4	0	4	3	1		
Wave 1+2+3	0	1	6	2	3		
Expected (normal) distribution	0.6	0.6	2.6	3.6	4.6		

Notes: * Corresponding confidence intervals given in parentheses.

** To test the observed distribution against the expected distribution in Part A of the table, we combined the > 1.960 and 1.960-1.651 categories because of the small expected frequencies. For the same reason, we did not test Part B of the table.

confidence interval. In the end, the bulk of the correlation coefficients turn up in the middle categories. Therefore, too many correlation coefficients are biased, even at a final response rate of 74%.

The figures presented in Table 5, part B, show that the outcomes for the correlation coefficients between independent and dependent variables are largely the same as the outcomes for all correlation coefficients.

The outcomes of the foregoing analysis can be tested in a more rigorous way by examining the complete distributions of deviations from the validating scores. Even when the successive cumulative waves are unbiased, we expect some random fluctuation around the 'true' (validating) value. By assuming that these random fluctuations are normally distributed, we can compute the goodness-of-fit between the observed distribution of deviations on a particular cumulative wave and the expected normal distribution by a one-sample χ^2 test. This test proves that the distributions for all successive cumulative waves are biased, but also that the bias decreases with an increasing response rate (Table 5, Part A).

3. So far, substantive biases have only been analysed in the bivariate relations between our eight variables. However, we also want to know whether the response rate affects the outcomes of multivariate analyses. Therefore, we performed multiple regression analyses using policy performance and success of realization of first site, respectively, as dependent variables, and using size, party politics, centre-periphery and new task as our independent variables.[6]

Table 6 Unstandardized, and in parentheses standardized regression coefficients by successive, cumulative waves, with success of realization of first site as dependent variable

	Wave 1		Wave 1 + 2		Wave 1 + 2 + 3		All municipalities	
	B	(β)	B	(β)	B	(β)	B	(β)
Size	-.483	(-.216)*	-.386	(-.171)*	-.338	(-.149)	-.120	(-.055)
Party politics	.006	(.103)	.006	(.098)	.007	(.122)	.008	(.123)*
Centre-Periphery	-.008	(-.010)	.030	(.039)	-.022	(-.027)	-.041	(-.050)
New Task	-.009	(-.144)	-.010	(-.155)	-.011	(-.165)*	-.011	(-.172)**
(constant)	3.902**		3.518**		3.402**		2.563**	
r^2 (adj.)	.049		.047		.044		.023	

* p ≤ .10; ** p ≤ .05.

The regression analysis for policy performance reveals that for all three cumulative waves and all municipalities alike, our municipal characteristics are not related in a linear way to policy performance (significance of $F > .05$ in all analyses). However, to explain the success of realization of first site, the response rate does indeed matter (Table 6). Overall, our municipal characteristics appear to play only a minor role. But, we also see that the variance explained decreases from almost five per cent to a little over two per cent when the response rate increases. Furthermore, our interpretation of what constitutes the most important variable or variables changes when the response rate increases. At low response rates, size is the only variable of any importance. After three cumulative waves, new task becomes more important. Finally, for all municipalities, party politics emerges as an additional factor, next to new task. Looking both at the unstandardized and the standardized coefficients, we see the outcomes of the multiple regression analysis becoming less and less biased when the response rate increases: size steadily declines in importance across cumulative waves, while new task continually grows in significance.

Thus, our study demonstrates that multivariate analyses are susceptible to non-response biases too, at least when the explained variance is low – which is not an uncommon situation in the social and political sciences. Our data do not enable us to say whether this conclusion also holds when r^2 is much higher. Because the answer to this question may have far-reaching consequences, more research is badly needed here.

On the whole, Hypothesis 3 is supported by the results of the foregoing three related tests: substantive biases diminish when the response rate increases. However, at our final response rate (74%), substantive biases have still not disappeared completely.

4.3 Data quality

What is the use of enhancing the response rate, if subsequently the additional data gathered is of inferior quality? The quality of the data is the point where non-response errors and response errors may clearly work at cross-purposes. To test whether response-enhancing techniques, in particular follow-up procedures, lead to a deterioration of the data gathered at later waves, we studied item non-response with two sets of questions: a set of six miscellaneous questions at the beginning of the questionnaire, and a set of eleven Likert-type attitudinal questions at the end of the questionnaire. In between, we asked questions about the policy implementation process. Because the number of questions relevant to a municipality varies with the task assigned and the progress made in implementing the task, a simple measure of item non-response cannot be used here. Instead, we assessed the quality of the data provided by the municipalities, using the information in the provincial reports as our standard. Unlike the analyses of distributional and substantive bias, the analysis of data quality uses answers from all 208 respondents. The results can be found in Table 7.

Table 7 Data quality by wave: item non-response and process data quality

	Wave 1	2	3	
1 *Six questions at beginning of questionnaire*				Analysis of variance:
Mean % no answer	1.2	1.1	3.9	$F = 3.797$; $p = .024$
2 *Eleven attitudinal questions at end of questionnaire*				Analysis of variance:
Mean % no answer	8.7	6.2	13.9	$F = 1.156$; n.s.
3 *Assessment quality process data*				
% questionnaires with incomplete data	5.5	6.4	5.9	$\chi^2 = 5.536$; n.s.
% questionnaires with only information on second provincial plan	6.3	17.0	5.9	
(N)	(127)	(47)	(34)	

Note: Response completeness (percentage of questions answered) was used in the analysis-of-variance.

To start with, respondents to the third wave answered both sets of questions less frequently than respondents to the first two waves. However, using response completeness (percentage of questions answered), we only found significant differences for the first group of six questions. All eleven attitudinal questions show consistent differences: respondents to the second wave answered most questions, and respondents to the third wave answered least, but these differences fail to reach significance.

In the analysis of the quality of the process data, two types of errors emerged. First, the information given by some respondents was incomplete, incoherent, or difficult to understand. Second, some questionnaires contained only information on the then current (second) provincial plan, while we asked for information about the first as well as the second plan. As can be observed in Table 7, no systematic deterioration of the quality of the process data can be detected across waves.

In all, the data support Hypothesis 4 to some extent. Item non-response is largest for wave 3. On the other hand, the quality of the process data does not deteriorate systematically across waves.

5 Discussion

In this article, we tested a number of hypotheses concerning the consequences of enhancing the response rate for distributional bias, substantive bias and data quality in a mail survey among municipal officials on the implementation of a national caravan sites policy in Dutch municipalities. The results of our study are primarily of interest for studies of organizations using key informants as their data source. However, just like the study by Tomaskovic-Devey et al. (1994), our research suggests that response behaviour of individuals and organizational key informants does not differ greatly.

We found both distributions and relationships to be seriously biased at a low response level (46%). Higher response levels (62% and 74%) indeed led progressively to better results, thereby confirming Hypotheses 1 and 3 which predict a decrease in the distributional bias and substantive bias respectively as the response rate increases. However, even at a final response rate of 74%, distributions and relationships are still biased to some extent. This is caused by the fact that the remaining non-respondents are composed more and more of deviant subgroups. Hence, Hypothesis 2a, which predicted decreasing differences between respondents and non-respondents was refuted, while Hypothesis 2b, which predicted increasing differences between them, was confirmed.

There are indications that the quality of the data deteriorates somewhat at later waves, lending some support to Hypothesis 4. However, the effects are not so damaging that response-enhancement is unwarranted. Nevertheless, data quality remains an important topic of concern, especially when trying to enhance response rates even more to reach almost complete response (cf. Dillman 1991: 240-241).

Our study demonstrates that response-enhancement is certainly worthwhile. In particular, the supplementary response gained at wave 3, from 62% to 74%, greatly reduces the bias in our results. Unfortunately, the outcomes do not enable us to say with certainty which response rate is high enough. Obviously, a final response rate of 74% still appears to be inadequate. Hence, our study underlines the necessity of high response rates to obtain valid outcomes. To err on the safe side, a response rate of considerably more than 80% seems desirable.

Our results also demonstrate that striving for a high response rate in itself may not be sufficient, because the remaining non-respondents are composed more and more of specific subcategories. We believe that this divergence between respondents and non-respondents as the response rate increases is of crucial importance. For, it explains why distributional and substantive biases cannot be ignored, not even at rather high response levels. In our opinion, this outcome is also of great significance for attempts to obtain not only response rates which are as high as can be attained within reason, but are also sufficiently unbiased. To reach this twofold goal, a dual field strategy is in order. Apart from *general methods* to enhance the response rate (e.g., follow-up procedures, layout of questionnaire, monetary incentives), we have to devise *specific* strategies which are tailored to the needs and characteristics of particular categories of the population studied.

Instead of further enhancing the response rate, some might prefer to control for bias by weighting for it afterwards. However, as we have shown elsewhere, weighting seldom turned out to be a satisfactory substitute to response-enhancement in our study (Van Goor and Stuiver forthcoming). Hence, we our of the opinion that response-enhancement is the best strategy to obtain unbiased results.

Notes

1. The municipal assignments were taken from formal policy documents adopted by the county council (Provinciale Staten) concerned. Data on the implementation were gathered from official documents prepared by the provincial administration for deliberation in the county council. To assess the progress made by the municipalities, we primarily relied upon formal decisions by the municipal councils (gemeenteraden) which are based on procedures laid down in laws and by-laws and which are subject to provincial review and assent (e.g. Law on Town and Country Planning [Wet op de Ruimtelijke Ordening]). These data make it possible to establish the progress validly in most cases.

In four cases we were unable to ascertain the task assigned to a municipality. Additionally, we were unable to ascertain the speed or extent of the implementation of the plan targets (i.e., the policy performance) in eleven cases. Also in eleven cases, information about the success of the implementation of the first site is missing.

2. Policy performance measures the speed and extent of implementation of the complete provincial plan targets over a period of approximately ten years. Success of realization of first caravan site measures an aspect of policy performance. As can be expected, both outcome variables are rather strongly related ($r = .683$; $N = 243$).

3. The percentage error is a measure of the extent to which the distribution of the respondents to the first wave, waves 1 + 2, and waves 1 + 2 + 3 differ from the distribution for all municipalities studied (cf. Hawkins 1975: 471, Table 1, note a).

In formula:

$$\frac{\%\ \textit{cumulative wave X} - \%\ \textit{all municipalities}}{\%\ \textit{all municipalities}} \times 100$$

4. Because the size of the municipalities varies greatly (from less than 1,000 to almost 600,000 inhabitants), with only a few very large municipalities, we use a logarithmic transformation of the size variable. Moreover, we use the original distributions for party politics, caravans in 1978, and number of caravans assigned instead of the categorized variables used in the analysis of distributional bias.

5. See for the formula of s Blalock (1979: 418-420). We use N of all municipalities in this formula, because r of all municipalities (r_p) is our point of departure for measuring the amount of agreement. Hence, we use the corresponding confidence intervals. Because of missing values, N varies between 243 and 257.

6. Because size and urbanization are rather strongly interrelated, and relate more or less in the same way with both dependent variables, we decided to run analyses using size or urbanization alternately. As the outcomes were roughly the same, only those analyses using size are presented here. Due to the strong relationship between caravans 1978 and new task, one had to be omitted. Hence, we decided to select new task as the most interesting variable.

Bibliography

Bebbington, A.C. (1970), 'The Effect of Non-Response in the Sample Survey with an Example', *Human Relations* 23, pp. 169-180.

Blalock, H.M. (1979), *Social Statistics* (revised 2nd edn.). New York: McGraw-Hill.

Bradburn, N.M. (1992), 'A Response to the Nonresponse Problem', *Public Opinion Quarterly* 56, pp. 391-397.

Cohen, J. (1994), 'The Earth is Round (p <.05)', *American Psychologist* 49, pp. 997-1003.

Dillman, D.A. (1991), 'The Design and Administration of Mail Surveys', *Annual Review of Sociology* 17, pp. 225-249.

Dolsen, D.E. and G.E. Machlis (1991), 'Response Rates and Mail Recreation Survey Results: How Much is Enough?', *Journal of Leisure Research* 23, pp. 272-277.

Donald, M.N. (1960), 'Implications of Nonresponse for the Interpretation of Mail Questionnaire Data', *Public Opinion Quarterly* 24, pp. 99-114.

Fitzgerald, R. and L. Fuller (1982), 'I Hear You Knocking But You Can't Come In: The Effects of Reluctant Respondents and Refusers on Sample Survey Estimates', *Sociological Methods & Research* 11, pp. 3-32.

Goudy, W.J. (1976), 'Nonresponse Effects on Relationships Between Variables', *Public Opinion Quarterly* 40, pp. 360-369.

Goudy, W.J. (1978), 'Interim Response to a Mail Questionnaire: Impacts on Variable Relationships', *Sociological Quarterly* 19, pp. 253-265.

Goyder, J. (1987), *The Silent Minority: Nonrespondents on Sample Surveys*. Cambridge: Polity Press.

Green, K.E. (1991), 'Reluctant Respondents: Differences between Early, Late, and Nonresponders to a Mail Survey', *Journal of Experimental Education* 59, pp. 268-276.

Groves, R.M. (1987), 'Research on Survey Data Quality', *Public Opinion Quarterly* 51, pp. S156- S172.

Hawkins, D.F. (1975), 'Estimation of Nonresponse Bias', *Sociological Methods & Research* 3, pp. 461-488.

Hox, J.J. and E.D. de Leeuw (1994), 'A Comparison of Nonresponse in Mail, Telephone, and Face-to-face Surveys: Applying Multilevel Modeling to Meta-analysis', *Quality and Quantity* 28, pp. 329-344.

Pavalko, R.M. and K.G. Lutterman (1973), 'Characteristics of Willing and Reluctant Respondents', *Pacific Sociological Review* 16, pp. 463-476.

Tomaskovic-Devey, D., J. Leiter and S. Thompson (1994), 'Organizational Survey Nonresponse', *Administrative Science Quarterly* 39, pp. 439-457.

Van Engelsdorp Gastelaars, R.E., W.J.M. Ostendorf and S. de Vos (1980), *Typologieën van Nederlandse gemeenten naar stedelijkheidsgraad: Milieu-differentiatie en maatschappelijke ongelijkheid binnen Nederland,* Monografieën Volkstelling 1971, Vol. 15B. The Hague: Staatsuitgeverij.

Van Goor, H. and B. Stuiver (Forthcoming), 'Can Weighting Compensate for Nonresponse Bias in a Dependent Variable? An Evaluation of Weighting Methods to Correct for Substantive Bias in a Mail Survey among Dutch Municipalities', *Social Science Research,* accepted for publication.

Janne Haaland Matláry, *Energy Policy in the European Union*, Macmillan, London, 1997, ISBN 0-333-64349-6, Dfl 50,80.

Energy Policy in the European Union explains why and how the EU was ble to become highly influential in European energy policy-making over the 1985-1992 period. Traditionally, energy was considered a matter of national interest and governments played a dominant role in the energy sector through regulation, public ownership, subsidization, etc. Therefore, the 1985 *White Paper on the Internal Market* did not cover the energy sector. However, by 1988 the Commission extended the internal market concept to include the energy sector with a series of proposals for the creation of the *Internal Energy Market* (IEM). Besides this, the Commission sought to achieve formal competence in energy matters and attempted to formulate a *Common Energy Policy* (CEP) beyond the deregulatory nature of the IEM. Since 1992, as is argued in this study, the EU has reverted to more modest ambitions regarding the further development of the IEM and the CEP.

On the back cover, Haaland Matláry's work is labelled as 'the first comprehensive book-length study of EU energy policy', which it is not. Regrettably, it lacks an evaluation of the actual implementation of EU energy policy at the national level and the consequences for the supply and use of energy. *Energy Policy in the European Union* is essentially an empirical case study on the *process* of energy policy-making within the EU framework, and the role therein of the member states, the various EU institutions and interest groups. In this sense, it is a welcome and relevant addition to the existing sector-based studies of EU policy-making. Moreover, as energy analysts follow the developments in 'Brussels' meticulously, the *process* of EU energy policy-*making* has not received much attention so far, and therefore the book is a valuable contribution to European energy studies.

The book also claims to contribute to European integration theory in general by asking to what extent an intergovernmental approach is appropriate to explain the evolution of EU energy policy-making. The hypotheses tested are, first: *the only significant actors are governments* and, secondly: *all policy-making outcomes in the EU process can be traced to prior government interests.* The alternative hypotheses are: *institutional EU actors matter independently and interests may be formed during the policy-making process itself.*