Representativity of the Time Use Survey

08

Adelaide Ariel and Barry Schouten

Discussion paper (08013)

Statistics Netherlands

Voorburg/Heerlen, 2008

Explanation of symbols

.	= data not available
*	= provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2005-2006	= 2005 to 2006 inclusive
2005/2006	= average of 2005 up to and including 2006
2005/'06	= crop year, financial year, school year etc. beginning in 2005 and ending in 2006
2003/'04–2005/'06	= crop year, financial year, etc. 2003/'04 to 2005/'06 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Prinses Beatrixlaan 428
2273 XZ Voorburg

end of Augustus 2008:
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands - Facility Services

Cover
TelDesign, Rotterdam

Information
Telephone .. +31 88 570 70 70
Telefax .. +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order
E-mail: verkoop@cbs.nl
Telefax .. +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1572-0314

6008308013 X-10

REPRESENTATIVITY OF THE TIME USE SURVEY

Summary: In 2006 the selection of respondents for the Time Use Survey ('Tijdbestedingsonderzoek' or 'TBO') consisted of three stages. First, a sample of persons was drawn from the households that responded to the 2005 and 2006 Mobility Surveys (Mobiliteitsonderzoek Nederland or MON). Second, the selected MON respondents be asked to participate in a TBO contact interview. Third, a sample of respondents to this contact interview was invited to complete a one-week TBO diary..

Both MON and contact interview questionnaires contain items that relate strongly to the TBO diary. From the perspective of non-response reduction and adjustment, it is therefore informative to investigate the representativity of the final TBO respondents with respect to the original sample. This analysis may aid future data collection for the TBO as well as a more effective two-step adjustment procedure.

In this paper we employ R-indicators to investigate the representativity of the TBO response. R-indicators are proposed by Schouten and Cobben (2007) as measures to evaluate the similarity between the response and population of interest. We make three comparisons: between TBO diary and TBO contact interview response, between TBO contact interview response and MON sample, and directly between TBO diary and MON sample. We find that persons who are more active according to MON or contact interview are more likely to be respondents to the TBO full interview.

This research was sponsored by the Netherlands Institute for Social Research (SCP). The 2005 and 2006 MON datasets were provided by DANS[1].

Keywords: Non-response, R-indicator, non-response adjustment, TBO, MON

1. Introduction

The Dutch Time Use Survey (in Dutch: 'Tijdbestedingsonderzoek' or 'TBO') is conducted to get insight into the type and duration of activities spend on various daily activities such as paid work, commuting, doing household tasks, and socializing. Survey participants are asked to detail their 24-hour activities for a period of one week by means of a time use diary. This survey is carried out by The Netherlands Institute for Social Research (SCP) and partners once every five years. In 2006, the survey was conducted in accordance with Eurostat guidelines (Harmonized European Time Use Survey, HETUS) in order to enhance comparability within the EU. One of the major design changes was the fieldwork period; the 2005 TBO was held in October only while the 2006 TBO was

[1] DANS data sets (MON 2005, ID P1733, 01/01-12/31 2005, and MON 2006, ID P1734, 01/01-12/31 2006), produced by Ministry of Transport, Public Works and Water Management.

split in 12 monthly portions. In order to be able to compare the 2006 TBO to the 2005 TBO the October sample size was taken larger than those of the other months. The fieldwork for the 2006 TBO was carried out by SocialData.

One of the primary concerns of survey researchers is non-response, since the bias resulting from non-response may hamper strong conclusions about the population of interest. The impact of non-response is often measured through the response rate; the response rate of the 2006 TBO was 62% (1875 out of 3041 persons returned the TBO diary, Socialdata, 2007). However, it is recognized in the literature, e.g. Schouten and Cobben (2007) that response rates by themselves may be poor indicators of the quality of the response to a survey. It is imperative that the composition of the response is compared with that of the population of interest with respect to important background characteristics. Survey researchers often calibrate the response distribution to a number of population statistics like age, household type, gender and degree of urbanization. Statistics Netherlands has the favourable position that it can link several administrative data sets and registers to its surveys. Background characteristics may, however, not relate directly to the variables of interest in the survey. Therefore, sometimes follow-up studies are conducted in order to investigate possible relationships between survey participation and survey topic. Examples are Kersten and Bethlehem (1984), Stoop (2005) and Schouten (2007). This way, survey researchers can verify whether survey participation indeed correlates with survey topic.

In the case of the 2006 TBO, no follow-up study was done, but a considerable amount of information about respondents and non-respondents is available from the Mobility Survey (in Dutch MON, or Mobiliteits Onderzoek Nederland). The 2006 TBO sample is drawn from respondents of the 2005 and 2006 MON. First, random persons in the selected MON households are invited to participate in a TBO contact interview. If a selected person responds to this request, then he or she is considered to be eligible to fill in the detailed one-week TBO diary. From the set of eligible persons a random sample is drawn which is asked to keep a one-week TBO diary. The 2006 TBO design enables us to investigate the composition of the response to the TBO diary with respect to MON and contact interview characteristics. These characteristics relate strongly to the items in the TBO diary.

It is our main objective in this paper to analyse the extent to which TBO respondents are representative for the original response to the TBO contact interview and to the MON. We do so by applying the R-indicator defined by Schouten and Cobben (2007). The rationale behind this indicator is that the response to a survey can be considered to be representative when all individual response probabilities are the same. The proposed R-indicator measures a deviation from this ideal situation by computing the variation of estimated response probabilities. Response probabilities are estimated using the available background characteristics; in case of the TBO the contact interview and MON variables.

This leads us to the following research questions:

1. What is the representativity of the response in the Contact Interview and the response in TBO, with regard to survey variables of MON and Contact Interview?

2. Which survey variables of MON and Contact Interview are different in distinguishing the definitive response of the TBO?
3. Which variables of MON and Contact Interview are candidates for non-response adjustment?

In Section 2, we elaborate on the TBO sample and data collection design. In Section 3, we give some background to the methods used in the analyses. In Section 4, we answer the three research questions. We discuss our findings in Section 5.

2. The data collection design of the Time Use Survey

In 2006 the selection of respondents for the Time Use Survey (TBO) consisted of three stages:

1. *Stage 1 - Response to the 2005 and 2006 MON*. A random sample was drawn from responding households to the 2005 and 2006 Mobility Surveys (MON). One person was randomly selected from this household using the Kish-selection grid method (Kish 1949) and asked to participate in a contact interview.
2. *Stage 2 – Response to TBO Contact Interview*. A random sample was drawn from the respondents to the TBO contact interview. This sample was invited to respond to a detailed TBO questionnaire.
3. *Stage 3 –Response to TBO diary*. Part of the sample responds to the one-week TBO diary.

In 2005 and 2006 MON was fielded for a period of one year and split in monthly portions. MON was a mail survey where all members of the selected households were asked to specify their mobility activities in detail. The Transport Research Centre of the Ministry of Transport, Public Works, and Water Management (AVV) has commissioned Socialdata to conduct this household survey among the Dutch population; the number of respondents is about 53,000 (AVV 2007).

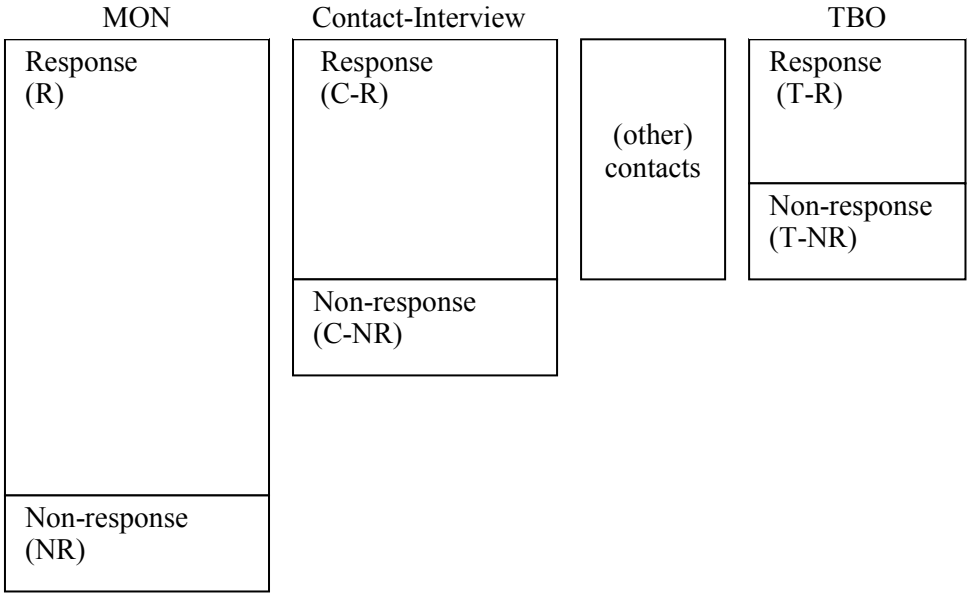
The selection probabilities of the MON households were taken proportional to their sizes The Kish-selection grid method was used to ensure a random selection of persons within these households. Once the TBO sample for the diary was obtained, a series of telephone contacts was made to guide and motivate the selected persons. A complete description of the selection procedure is given in Socialdata (2007).

In Figure 2.1 we illustrate the three-stage selection procedure for TBO. Response rates for the three stages were: 71% for MON, 92% for the Contact Interview, and 62% for TBO.

The MON respondents provided the details of their commuting activities for one day (24 hours): departure and arrival time, mode of transportation, main purpose of the commuting, and so on. In the Contact Interview, they were asked to specify all their activities (sleeping, working, doing chores, commuting, socializing, etc) from 4 a.m on the same day they filled in the MON questionnaire to 4 a.m. the following day. This

interview was done by telephone and took about 15 minutes on average. The one-week TBO diary was similar to the Contact Interview; the main differences were that the TBO respondents had to detail their activities in a 7-day diary chronologically for 10-minute intervals. They had to do so without the assistance of the interviewer.

Figure 2.1. Overview of the TBO sample procedure



For the three stages in figure 2.1 sample and response sizes are as follows:

- Stage 1: MON
  Response: R (53,000 people), Non-response: NR

- Stage 2: Contact Interview
  Response: C-R (2811 people), Non-response: C-NR (230 people)

- Stage 3: TBO
  Response: T-R (1875 people), Non-response: T-NR + C-NR (1166 people)

Clearly, there is a strong overlap between the survey variables in each of the three stages. Therefore, the survey variables of the first two stages may be strong predictors for response in the detailed TBO, the third stage. We may expect that if these variables give only a low explanation of response behaviour in the TBO, the non-response is missing-at-random with respect to survey topics. Also, they may be used as weighting variables in a separate non-response adjustment step. We are especially interested to identify, if any, which MON and Contact Interview variables could predict non-response in the TBO survey. For the analysis, the MON and Contact Interview variables will be used as predictors for response probabilities estimate. R-indicators are then computed based on these estimates. We detail our approach in the following section.

3. A measure for the selectivity of non-response: the R-indicator

It is known that response rates alone cannot be used as indicators of survey quality (see e.g., Groves 1989 and Biemer and Lyberg 2003). The R-indicator (Schouten and Cobben 2007, Cobben and Schouten 2008) has been developed as a tool to measure the representativity of response or analogously to measure the selectivity of non-response.

Representative response is defined as equal response probabilities for all individuals in the population of interest. The R-indicator computes variation in estimated individual response probabilities; the smaller the variation of these probabilities, the stronger the representativity. The R-indicator takes values between 0 and 1. A value close to 1 implies that we cannot distinguish the non-response from the response, and thus the response can be considered representative with respect to predictive variables. The lower the value of the R-indicator the stronger the predictive variables separate response from non-response. Hence, the R-indicator helps us not only to identify which variables are predictive but also to measure the extent to which these variables differentiate response and non-response.

Clearly, the true values of response probabilities are unknown to the survey researcher, and, thus, need to be estimated. Individual response probabilities can be estimated using models that incorporate available background characteristics. In our case, the MON and Contact Interview survey variables are known for both the TBO respondents and non-respondents. Hence, these survey variables can be employed to estimate response probabilities of the TBO. In Section 4 we construct explanatory models for response and compute corresponding R-indicators.

The TBO diary variables themselves are known only for respondents and can, therefore, not be used as explanatory variables in the estimation of response probabilities. However, since the MON and Contact Interview variables are closely related to the detailed TBO variables, we believe they are good predictors of the detailed diary items.

In addition to the R-indicator values, Schouten and Cobben (2007) recommend to compute upper bounds for the maximal absolute bias of key items in the survey that can be attributed to non-response. The R-indicator can be translated to a maximal absolute non-response bias by assuming that there is an optimal correlation between response to the survey and answers to the survey questions. When assuming such a worst case scenario, one gets an upper bound for the maximal absolute bias. In Section 4 we compute this upper bound in order to show the maximal impact of the non-response.

We remark that the R-indicator itself is a random variable with a standard error that depends on the sample size of the survey. The larger the sample size, the more precise the R-indicator. For this reason, in Section 4, we also show 95% confidence intervals for the R-indicator.

4. R-indicators for the response to the Time Use Survey

We analyse the response in Stages 2 and 3 of Figure 2.1. We restrict analysis to these stages for two reasons. First, we do not have access to the original MON samples. However, we may analyse MON response behaviour by comparing respondents to population statistics. Second, we assume that non-response adjustment for MON is well investigated and elaborated. Hence, if TBO respondents are representative of MON respondents, then we can apply the MON non-response adjustment procedures.

In Section 4.1 we derive R-indicators for Stages 2 and 3 based on MON variables. Next, in Section 4.2 we look at R-indicators for Stage 3 based on Contact Interview variables.

4.1 Stage 2 and 3: R-indicators related to MON survey variables

MON mobility activities portray how people spend their time. Therefore, we have chosen to use the following MON variables:

- public transportation user (yes, no),
- total commuting time
- total commuting distance
- number of trips
- commuting category (did not commute at all, only work-related trips, only non-work-related trips, both work and non-work related trips)
- income
- occupation
- working category (less than 12-hour per week, 12-hour till 30-hour per week, more than 30-hour per week)
- month and weekday of MON survey participation.

Some of the variables were obtained by recoding the original variables in MON. *Total commuting time* and *total commuting distance* were obtained by summing travel time of each trip, and trip distance, respectively. Interval values were used for these variables. *Commuting category* was a recode of seven main motives of commuting in MON.

Furthermore, we also included standard auxiliary variables such as age (eight intervals), gender, household size (five categories), degree of urbanization (five categories), and province (12 provinces) into the model.

The details of the selected MON variables are given in the Appendix 1.

We used logistic regression models to predict response behaviour. Model selection procedures were based on forward, backward, and enter strategies in SPSS in order to obtain a robust set of variables. The forward and backward strategies were based on 5% significance levels for entering or removing variables.

Table 4.1.1 contains the MON variables that were selected in the logistic regression models for Stages 2 and 3. The variables are ordered with respect to their predictive power. In Stage 2 age turns out to be the strongest variables, while in Stage 3 it is the month of the MON interview.

Table 4.1.1: MON variables selected in logistics regression models for response in Stages 2 and 3.

Stage	*Selected MON variables*
2	Age, Number of trips, Month, and Income
3	Month, Age, Number of trips, and Commuting category

The R-indicators and the maximum absolute bias were calculated based on the variables in Table 4.1.1. We used naive non-parametric bootstrapping with 500 replications to calculate corresponding confidence intervals. Table 4.1.2 specifies the R-indicators and the maximum bias, whereas Table 4.1.3 and 4.1.4 specify the corresponding logistic regression parameters.

Table 4.1.2. Estimated R-indicators and maximum absolute bias using MON variables

Significant Variables	$R\ (CI_{0.05})$	$B\ (CI_{0.05})$
Stage 2 (Contact Interview)		
Age	88.7% (86.0, 91.4)	6.1% (4.6, 7.5)
Age + Number of trips	85.6% (82.8, 88.6)	7.7% (6.1, 9.2)
Age + Number of trips + Month	85.0% (83.0, 86.7)	8.1% (7.1, 9.2)
Age + Number of trips + Month + Income	80.9% (77.8, 83.8)	10.3% (8.6, 11.9)
Stage 3 (TBO)		
Month	70.5% (69.1, 72.0)	24.1% (22.7, 25.4)
Month + Age	66.8% (65.5, 68.1)	27.4% (26.1, 28.8)
Month + Age + Number of trips	64.6% (63.1, 66.1)	28.8% (27.3, 30.3)
Month + Age + Number of trips + Commuting category	62.5% (60.9, 63.9)	30.6% (29.1, 32.1)

From Table 4.1.2 we can conclude that the R-indicator values are much higher for the Contact Interview response (Stage 2) than for the TBO response (Stage 3). This implies that, with respect to MON survey variables, the representativity of the Contact Interview response is higher than the TBO response. In other words, the Contact Interview respondents resemble MON respondents better than do the TBO respondents with respect to MON characteristics.

Table 4.1.3: Stage 2: Response to Contact Interview. The parameters in the logistic regression model with standard errors and p-values. Bold denotes significance at 5% level. Each last category is used as a reference category.

Category	*Parameter*	*Standard error*	*p-value*
Age			
65 years and older	0		
less than 10 years	**-3.815**	.815	.000
10 - 14 years	-.054	.491	.912
15 - 17 years	.231	.517	.655
18 - 29 years	-.335	.293	.253
30 - 39 years	.404	.307	.187
40 - 49 years	-.101	.265	.701
50 - 64 years	-.146	.238	.539
Number of trips			
9 and more	0		
0	**-.995**	.461	.031
1	-1.165	.668	.081
2	-.294	.449	.512
3	.174	.547	.751
4	.095	.460	.837
5	1.153	.689	.094
6	.354	.510	.487
7	.681	.727	.349
8	.973	.758	.199
MON interview month			
December	0		
January	-.291	.361	.419
February	-.023	.371	.950
March	.595	.396	.133
April	-.025	.395	.949
May	.434	.418	.300
June	.359	.422	.395
July	-.081	.368	.825
August	.140	.407	.730
September	-.140	.399	.725
October	**1.133**	.414	.006
November	-.238	.418	.568
Income (€)			
Unknown	0		
Younger than 12 year	.470	.679	.489
No income	.566	.318	.075
Less than 7 500	**1.253**	.381	.001
7 500 - 15 000	**.692**	.264	.009
15 000 - 22 500	**.763**	.278	.006
22 500 - 30 000	**.907**	.300	.002
More than 30 000	**.783**	.286	.006
Constant	**2.055**	.577	.000

Table 4.1.4: Stage 3: Response to detailed TBO interview. The parameters in the logistic regression with standard errors and p-values. Bold denotes significance at 5% level. Each last category is used as a reference category.

Category	*Parameter*	*Standard error*	*p-value*
Age			
65 years and older	0		
less than 10 years	**-1.697**	.484	.000
10 - 14 years	**.846**	.198	.000
15 - 17 years	**.952**	.244	.000
18 - 29 years	**.614**	.167	.000
30 - 39 years	.072	.141	.609
40 - 49 years	**.288**	.140	.040
50 - 64 years	.137	.127	.280
Number of trips			
9 and more	0		
0	**-.814**	.267	.002
1	**-1.422**	.440	.001
2	**-.659**	.234	.005
3	-.446	.271	.100
4	**-.462**	.231	.046
5	-.021	.278	.941
6	-.131	.249	.597
7	.123	.316	.698
8	.206	.310	.507
MON interview month			
December	0		
January	**2.211**	.241	.000
February	**2.308**	.243	.000
March	**2.515**	.237	.000
April	**1.689**	.249	.000
May	**2.283**	.248	.000
June	**2.136**	.247	.000
July	**2.441**	.245	.000
August	**2.173**	.253	.000
September	**2.042**	.249	.000
October	**2.997**	.234	.000
November	**1.228**	.266	.000
Commuting category			
Mixed	0		
Did not commute	-.**814**	.267	.002
Non-work only	**.264**	.120	.028
Work only	.058	.167	.728
Constant	**-1.594**	.332	.000

In the Contact Interview, the number of trips is the only MON mobility variable that separates response from non-response: the more respondents travel the more likely they respond to the Contact Interview, as opposed to those who did not commute at all (see

Table 4.1.3). This finding may be counterintuitive as people who are busier may be expected to be less willing to participate in a survey. However, at the same time, people who commute frequently may have a stronger interest in a survey about daily activities. Research on survey nonresponse provides a discussion about who would be survey participants; see for example Abraham, Maitland, and Bianchi (2006) and Stoop (2007).

The relation with the number of trips is also found for the TBO. From Table 4.1.4 we observe that those with the largest number of trips tend to respond better. However, in the TBO, the response does not only depend on the number of trips but also on the purposes of commuting (commuting category): as long as a non-work related reason is involved, it is more likely for the commuter to respond.

For both Contact Interview and TBO, age and the MON interview month play a role. Their significance is more pronounced for the TBO respondents. For the TBO, the respondents are more likely to be young. The month of interview in MON turns out to be a very strong predictor of response in TBO. December MON respondents are almost all TBO non-respondents, while October MON respondents have a relatively high response rate. In Appendix 2 we provide the distribution of the respondents for the MON survey months. We suspect that the strong dependence on the MON fieldwork month is a fieldwork effect, i.e. the effort in obtaining response depends strongly on the original month of response in MON. This dependence can partially be explained by the objective to compare the 2006 TBO to the 2005 TBO which was fielded only in October. For this reason it was imperative to have a sufficiently large number of respondents in October for the 2006 TBO. Furthermore, as the fieldwork stopped at the end of December 2006, there was little time to get response from selected November and December MON respondents.

It may be conjectured, however, that the MON interview month does not relate strongly to the TBO diary items. Therefore, it is interesting to also model response behaviour and compute R-indicators without taking into account the month of interview in MON. We use the same forward, backward and enter strategies to construct a logistic regression model.

Table 4.1.5: MON variables selected in logistics regression models for response in Stages 2 and 3 excluding the MON interview month.

Stage	*Selected MON variables*
2	Age, Number of trips, and Income
3	Age, Number of trips, Occupation and Commuting category

In Table 4.1.5 we give the selected MON variables for Stages 2 and 3 leaving out MON interview month as an explanatory variable. It turns out that the same variables are selected and in the same order for Stage 2. However, for Stage 3, removing the survey months results in the selection of *occupation* as a significant variable.

R-indicators and maximal absolute bias for Stages 2 and 3 are shown in Table 4.1.6. Tables 4.1.7 and 8 specify the logistic regression parameters for the selected variables for Stages 2 and 3, respectively.

Table 4.1.6. R-indicators and maximum bias using MON variables, without MON-survey months.

Significant Variables	$R\ (CI_{0.05})$	$B\ (CI_{0.05})$
Stage 2 (Contact Interview)		
Age	86.8% (83.8, 89.8)	7.0% (5.4, 8.6)
Age + Number of trips	84.1% (81.6, 86.5)	8.5% (7.2, 9.9)
Age + Number of trips + Income	81.3% (78.4, 84.3)	10.1% (8.4, 11.7)
Stage 3 (TBO)		
Age	84.7% (83.0, 86.2)	11.9% (10.7, 13.2)
Age + Number of trips	79.2% (77.8, 80.6)	16.8% (15.6, 18.0)
Age + Number of trips + occupation	77.3% (76.4, 78.2)	18.3% (17.6, 19.0)
Age + Number of trips + occupation + commuting category	77.2% (76.2, 78.1)	18.0% (17.2, 18.8)

Removing the MON interview month from the analysis did not alter the conclusions about response behaviour for Stage 2. For Stage 3, however, the R-indicators increased considerably from 62.5% to 77.2%, which implies higher response representativity. Furthermore, we obtain additional information about the selectiveness of response to the one-week TBO diary. People who work between 12-30 hours per week as well as students were more likely to respond to the TBO.

4.2 Stage 3: R-indicators related to the Contact Interview survey variables

In this section we investigate the response to Stage 3 with respect to Contact Interview variables. The response consists of 1875 persons out of 2811 selected Contact Interview respondents; leading to a response rate of 67%.

In the analysis we selected the following Contact Interview variables:

- number of activities (range = 10 to 62, mean = 24)
- total commuting time
- number of trips (range = 0 to 18, mean = 3.5)
- commuting category (did not commute at all, only work-related trips, only non-work-related trips, both work and non-work related trips)
- month and weekday of Contact Interview participation.

Number of trips and commuting category were obtained by recoding the kind of activity recorded in the Contact Interview. Furthermore we recoded the values of the number of activities into three categories: below average, average, and above average. The number

of trips was recoded into ten intervals. Age, gender, household size, degree of urbanization degree, and province were again also recorded in the Contact Interview and were included as explanatory variables. We applied the same variable selection strategies as in section 4.1.

Table 4.1.7: Stage 2: Response to the Contact Interview, without MON interview month. The parameters are given together with standard errors and p-values. Bold denotes significance at 5% level. Each last category is used as a reference category.

Category	*Parameter*	*Standard error*	*p-value*
Age			
65 years and older	**0**		.000
less than 10 years	**-3.705**	.802	.000
10 - 14 years	-.022	.485	.965
15 - 17 years	.131	.511	.798
18 - 29 years	-.402	.290	.165
30 - 39 years	.385	.302	.202
40 - 49 years	-.130	.262	.618
50 - 64 years	-.170	.235	.469
Number of trips			
9 and more	**0**		.000
0	**-.999**	.457	.029
1	-1.167	.659	.077
2	-.305	.446	.493
3	.210	.546	.701
4	.081	.457	.858
5	1.181	.684	.084
6	.330	.507	.516
7	.690	.724	.341
8	.918	.745	.218
Income (€)			
Unknown	**0**		.012
Younger than 12 year	.471	.674	.485
No income	**.617**	.313	.049
Less than 7 500	**1.282**	.379	.001
7 500 - 15 000	**.674**	.261	.010
15 000 - 22 500	**.778**	.276	.005
22 500 - 30 000	**.908**	.297	.002
More than 30 000	**.770**	.284	.007
Constant	**2.247**	.494	.000

Table 4.1.8: Stage 3: Response to the TBO, without MON months. The parameters are given with standard errors and p-values. Bold denotes significance at 5% level. Each last category is used as a reference category.

Category	*Parameter*	*Standard error*	*p-value*
Age			
65 years and older	0		
less than 10 years	**-2.152**	.517	.000
10 - 14 years	.261	.345	.450
15 - 17 years	.279	.364	.444
18 - 29 years	.283	.202	.161
30 - 39 years	-.094	.165	.569
40 - 49 years	.044	.164	.787
50 - 64 years	-.041	.137	.767
Number of trips			
9 and more	0		
0	**-.758**	.259	.003
1	**-1.499**	.434	.001
2	**-.703**	.227	.002
3	-.462	.262	.078
4	**-.554**	.225	.014
5	-.132	.268	.622
6	-.209	.241	.386
7	.098	.306	.749
8	.071	.297	.811
Occupation			
Other incl. Retired/unemployed	0		
12-30 hour/week	**.379**	.158	.017
More than 30 hour/week	.123	.145	.398
Homemaker	.013	.143	.925
Students	**.593**	.293	.043
Commuting category			
Mixed	0		
Did not commute	-.**758**	.259	.003
Non-work only	**.288**	.120	.017
Work only	.090	.159	.574
Constant	**.622**	.256	.015

Table 4.2.1 shows the selected Contact Interview variables as well as the corresponding values of the R-indicator and maximal absolute bias. Again the TBO response does depend strongly on the month in which the Contact Interview was held. This variable is the strongest predictor of response. Age and again the number of activities give additional explanation to TBO response.

Table 4.2.1. R-indicators and maximum bias using (significant) Contact Interview variables.

Significant Variables	$R\ (CI_{0.05})$	$B\ (CI_{0.05})$
Month	71.2% (69.7, 73.0)	21.1% (19.7, 22.4)
Month + Age	67.1% (65.5, 68.5)	25.4% (24.1, 26.9)
Month + Age + Number of activities	64.3% (62.7, 65.9)	26.4% (25.0, 28.0)

Table 4.2.2: The parameters in the logistic regression for Contact Interview variables with standard errors and p-values. Bold denotes significance at 5% level. Each last category is used as a reference category.

Category	*Parameter*	*Standard error*	*p-value*
Age			
65 years and older	**0**		.000
less than 10 years	1.613	.981	.100
10 - 14 years	**1.226**	.221	.000
15 - 17 years	**1.171**	.268	.000
18 - 29 years	**.810**	.178	.000
30 - 39 years	.074	.140	.596
40 - 49 years	**.372**	.141	.008
50 - 64 years	.201	.129	.119
Number of activities			
More than 24	**0**		.000
Less than 20	**-.638**	.109	.000
Between 20 and 24	**-.325**	.109	.003
Contact Interview month			
December	**0**		.000
January	**2.451**	.250	.000
February	**2.509**	.250	.000
March	**2.589**	.242	.000
April	**1.917**	.254	.000
May	**2.472**	.258	.000
June	**2.373**	.253	.000
July	**2.740**	.256	.000
August	**2.304**	.260	.000
September	**2.225**	.256	.000
October	**3.138**	.240	.000
November	**1.330**	.271	.000
Constant	**-1.604**	.232	.000

In Table 4.2.2 the details of the logistics regression model are given. The logistic parameters show that significantly fewer people participated in the TBO if the Contact Interview was conducted in December. Again, October shows the opposite effect. This is not surprising as we saw in the previous section that those who participated in MON during December would not be likely to be TBO respondents. Young people were more likely to participate in TBO. The table also shows that people reporting more activities

than average were more likely to respond to TBO, as opposed to those with smaller numbers of activities. This finding is consistent with the conclusion that persons who report more trips in MON are better respondents in the TBO.

Again, we conjecture that the Contact Interview month does not strongly influence the TBO diary items. Based on this assumption, we recalculated the R-indicators without the Contact Interview survey month in Table 4.2.3. After excluding the Contact Interview month, the variable selection resulted in selecting again age and number of trips. However, the values of the R-indicator increased considerably, from 64.3% to 84.6%. This result is similar to the analysis based on MON variables in Section 4.1.

Table 4.2.4 gives the parameters for the logistic regression model without the Contact Interview month.

Table 4.2.3. R-indicators and maximum bias for response to TBO using Contact Interview variables, excluding the Contact Interview month.

Significant Variables	$R\ (CI_{0.05})$	$B\ (CI_{0.05})$
Age	85.3% (84.8, 85.8)	11.1% (10.8, 11.4)
Age + The number of activities	84.6% (84.2, 85.0)	11.4% (11.1, 11.7)

Table 4.2.4: Stage 3: Response to TBO, excluding the Contact Interview month. The parameters are given with standard errors and p-values. Bold denotes significance at 5% level. Each last category is used as a reference category.

Category	*Parameter*	*Standard error*	*p-value*
Age			
65+ years	0		
less than 10 years	.708	.823	.390
10 - 14 years	**1.178**	.212	.000
15 - 17 years	**1.001**	.253	.000
18 - 29 years	**.700**	.168	.000
30 - 39 years	.051	.131	.696
40 - 49 years	.250	.132	.058
50 - 64 years	.106	.122	.384
Number of activities			
More than 24	0		
Less than 20	**-.432**	.102	.000
Between 20 and 24	**-.222**	.102	.029
Constant	**.669**	.112	.000

5. Discussion

In this report we demonstrate the application of R-indicators to measure response representativity of the detailed Time Use Survey ('Tijdsbestedingsonderzoek' or 'TBO') on employing variables in the Mobility survey (MON or Mobiliteits Onderzoek Nederland) and the TBO Contact Interview. Using the MON variables, the R-indicator values are much higher for the Contact Interview than for TBO. This difference can to a large extent be explained by the strong dependence of response on the month in which MON and Contact Interview were conducted. This is apparent from the large increase of the R-indicator values when the survey month was excluded. We suspect that this effect is largely due to the effort spent in obtaining response by Socialdata. October was an important interview month as this month was used for comparison of the 2005 and 2006 TBO's. Furthermore, as the fieldwork stopped in December 2006, there was little time to conduct interviews for selected households in November and December.

Based on their response to the MON survey variables, the TBO respondents can be characterized as those who commute more frequently, whose reasons of commuting involve at least a non-work purpose. Based on their response to the Contact Interview variables, it appears that commuting variables did not significantly separate the respondent from the non-respondent. Instead, it suggests that the TBO respondents were more active than the non-respondents, as those with higher than average number of activity were more likely to respond to the TBO. It is perhaps necessary to take these variables into account for weighting the TBO results. It is also interesting to note that age is the only auxiliary variable that significantly distinguishes response from non-response. Young people, except those who are between 30-39 years old, would tend to respond to TBO. In this analysis we do not investigate interaction effects that may occur.

The overall results indicate that the TBO respondents are more active people than non-respondents. Results from the American Time Use Survey (Abraham, Maitland, and Bianchi, 2006) provide little support for the hypothesis that busy people were less likely to respond, but considerable support for the hypothesis that people who were less socially engaged were less likely to be respondents. Marital status, presence of children, home ownership, being out of the labour force, and areas of living were characteristics used to determine the level of social engagement. In our case, being out of the labour force was one of the significant variables for non-response, and any other variables like household size and degree of urbanization did not account significantly for non-response.

Although we may conclude that more active people are overrepresented in the TBO, we need to be cautious, as we cannot rule out the possibility of response bias. It may be appealing to check for consistency in the response of MON and Contact Interview, however, there is a risk that their results may not be easily compared as the methods are different: MON is dominantly a mail survey where Contact Interview is a telephone survey.

Acknowledgements

The authors would like to thank Carlijn Kamphuis, Jan Spit and Ineke Stoop (all SCP), Socialdata, and Jelke Bethlehem (CBS) for their assistance, comments, and suggestions.

6. References

Abraham, K.G., Maitland, A., and Bianchi, S.M. (2006), Non-response in the American Time Use Survey: Who is missing from the data and how much does it matter? Public Opinion Quarterly, 70, 676-703.

Biemer, P.P., and Lyberg, L.E. (2003), Introduction to survey quality, John Wiley & Sons, New York, NY. USA.

Cobben, F. and Schouten, S. (2007), An empirical validation of R-indicators, Discussion paper 08006, CBS, Voorburg.

Groves, R.M. (1989), Survey errors and survey costs, John Wiley & Sons, New York, NY. USA.

Kersten, H.M.P., and Bethlehem, J.G. (1984), Exploring and reducing the non-response bias by asking the basic question, *Statistical Journal of the United Nations*, ECE, 2, 369-380.

Kish, L. (1949), A procedure for objective respondent selection within the household, Journal of the American Statistical Association, 44, 247, 380 – 87

Ministerie van Verkeer en Waterstaat, Rijkswaterstaat, Adviesdienst Verkeer en Vervoer (2007), Mobiliteitsonderzoek Nederland 2006, Het onderzoek.

Schouten, B. (2007), A follow-up of non-response in the Dutch Labour Force Survey, Discussion paper 07004, CBS, Voorburg.

Schouten, B. and Cobben, F. (2007), R-indexes for the comparison of different fieldwork strategies and data collection modes, Discussion paper 07002, CBS, Voorburg.

Socialdata (2007), Tijdbestedingsonderzoek 2006 Nieuwe stijl, Veldwerkverslag, Socialdata BV, Heerlen.

Stoop, I. (2007), No time, too busy. Time strain and survey cooperation. In: Geert Loosveldt, Marc Swyngedouw and Bart Cambré (eds.) *Measuring Meaningful Data in Social Research*. Leuven: Acco, pp. 301-314.

Stoop, I. (2005), Surveying nonrespondents, Field Methods 16, 23 – 54.

Appendix 1: Recoding of MON and Contact Interview variables

Label	*MON-variables*	*Comment*
Total commuting time (rsdduur_sum_int)	rsdduur, ritnr	Numeric Summing up ***rsdduur*** (in minutes) with ***ritnr*** = 1. Divide its values into intervals: [0], [1,15], [16,30] [31, 60] [61,120], [121,240], [241,∞)
Total commuting distance (afstvsum_int)	afstv, ritnr	Numeric Summing up ***afstv*** (in hectometer) with ***ritnr*** = 1 Divide its values into intervals: [0], [1, 49] [50, 99], [100, 199], [200, 299], [300, 399], [400, 499], [500, 599], [600, 699], [700, 799], [800, 899], [900, 999], [1000, 1999], [2000, 2999], [3000,∞)
Number of trips (aantvpl_int)	aantvpl	Numeric Divide its values into intervals: 0,1,2,3,…,8, 9++ (more than 9)
Commuting category (cat_verp)	kmotief	Category Divide ***kmotief*** into intervals: 0 (no commute) if aantvpl = 0 1 (nonwork) if kmotief for each rsdduur does not contain kmotief=1 or kmotief=2 2 (only work) if kmotief for each rsdduur only contain kmotief=1 or kmotief=2 3 (mixed) if kmotief for each rsdduur contain kmotief=1 or kmotief=2, and any other value of kmotief.
MON weekday (weekdag)	weekdag	Category Weekdays (7 days)
MON month (maand)	maand	Category Months (12 months)

Label (cont.)	*MON-variables*	*Comment*
Working category (betaaldw)	betaaldw	Category Unknown, less than 12 hour per week, between 12-30 hour per week, more than 30 hour per week.
Occupation (maatparticipatie)	maatpart	Category (some of its values are combined) Younger than 6 is combined together with retired, unemployed, disablement social support (WAO), and unkown; work 12-30 hour per week; work more than 30 hour per week; home maker; students.
Income (inkomen)	inkomen	Category Less than 12 years, no income, less than 7500, between 7500-15000, 15000-22500, 22500-30000, more than 30000, unknown.
Public transportation user (persopv)	persopv	Category yes, no
Gender (geslacht)	geslacht	Category Man, woman
Age (ageclass)	leeftijd	Numeric Divide its values into intervals 0-9, 10-14, 15-17, 18-29, 30-39, 40-49, 50-64, 65++
Household size (phh)	aantpers	Numeric Divide its values according to the number of people in the house (aantpers) 1,2,3,4, and 5+
Degree of urbanisation (sted)	sted	Category Very high, high, average, low, very low.
Province (prov)	prov	Category 12 provinces in the Netherland

Label	*Contact-Interview*	*Comment*
Number of activity (actantal_int)	actantal	Numeric Divide its values into intervals Less than average [1,20], average [21, 25], more than average [26, ∞)
Total commuting time (rsdduur_sum_int_Con)	Startuur, startmm, enduur, endmm, act	Numeric, category For each (commuting) act, time duration is calculated using: ***enduur-startuur***)*60+(***endmm-startmm***). Commuting is denoted by ***act*** = 900, 901, 902, 9041, 9042, 905, 906, 909, 910, 911, 912, 913, 914, 915, 917, 918, 9201, 9202, 9211, 9212, 922, 923, 928. Total commuting time (in minute) is divided into intervals: [0], [1,15], [16,30] [31, 60] [61,120], [121,240], [241,∞)
Number of trips (aantvpl_int_TBO)	act	Category. Count the total number of commuting, using act denoted above. Divide its values into intervals: 0,1,2,3,…,8, 9++ (more than 9)
Commuting category (cat_verp_TBO)	act	Category. Divide act into intervals: 0 (no commute) if aantvpl_int_TBO equals to 0, 1 (non work) if ***act*** does not contain 9041 or 9042 2 (only work) if act only contains 9041 or 9042 3 (mixed) if act contains 9041 or 9042 and any other act values specified for commuting.
Contact-Interview day (invuldag)	invuldag	Category Weekdays (7 days)
Contact-Interview month (inv_mm)	inv_mm	Category Months (12 months)

Appendix 2: Response for the MON and Contact Interview survey months.

MON	*Stage 2 (Contact Interview)*		*Total*
	Non-response	*Response*	
January	26	241	267
February	22	242	264
March	16	323	339
April	17	188	205
May	13	217	230
June	12	217	229
July	23	243	266
August	18	182	200
September	16	202	218
October	13	452	465
November	14	134	148
December	16	168	184
Total	206	2809	3015

MON	*Stage 3 (TBO)*		*Total*
	Non-response	*Response*	
January	98	169	267
February	92	172	264
March	102	237	339
April	97	108	205
May	80	150	230
June	84	145	229
July	84	182	266
August	79	121	200
September	85	133	218
October	98	367	465
November	87	61	148
December	154	30	184
Total	1140	1875	3015

Contact Interview	*Stage 3 (TBO)*		*Total*
	Non-response	*Response*	
January	72	168	240
February	72	179	251
March	87	230	317
April	82	113	195
May	64	145	209
June	71	148	219
July	62	180	242
August	60	121	181
September	69	133	202
October	85	367	452
November	73	61	134
December	139	30	169
Total	936	1875	2811