

Data-collection and Fieldwork Procedures

JOHANNES H. SMIT, MAARTEN Z. DE VRIES,
JOHANNES L. POPPELAARS

Introduction

This chapter provides information on the steps taken in the data-collection for the second cycle of the Longitudinal Aging Study Amsterdam (LASA). Between September 1995 and December 1996 data were collected according to various methods and procedures. For the sake of clarity, these various methods and procedures will be referred to in this chapter as 'main', 'medical' and 'self-administered'. Because of the longitudinal design of LASA, only minor changes were made in the questions and procedures described by Smit and De Vries (1994). Additional data-collection methods and procedures were incorporated in the LASA study (medical part) on the topic of 'risk factors for falls and osteoporotic fractures'. For each procedure, the respondents, data-collection procedures and (non-)response will subsequently be described. Procedures of the second LASA cycle which differed from procedures of the first cycle (Smit and De Vries 1994), will be explicitly described. The last part of this chapter gives an overview of the LASA data available per December 1996, and some remarks will be made on the handling and consequences of non-response and attrition.

Attrition between cycles

At the end of the first LASA cycle, data on 3107 respondents had been collected. In the years between the end of the first cycle (October 1993) and the beginning of the second cycle (September 1995) much effort was directed towards keeping the address database of respondents up to date with regard to the last known addresses, moves and other mutations. Each year a newsletter reporting some findings of

the study was mailed to the participants, the purpose of which was twofold. In the first place the intention was to keep the respondents interested and involved in LASA, and hopefully motivate them to participate in the following data-collection cycle. A second important spin-off of sending newsletters is the fact that a number of them will be returned as undelivered, indicating that these particular respondents have moved or are deceased. This information gives the fieldwork team the opportunity to investigate the whereabouts of these particular respondents, for instance by contacting a proxy.

In addition, each year the municipalities in which the respondents live are contacted and asked whether each particular respondent is still living at the same address and, if not, they are asked to provide the new address. If the municipality reports that the respondent is deceased, the date of death and registration number of the death certificate are requested. If a respondent has moved to another municipality, the new municipality is asked the same questions. To give an indication of the effort involved in keeping the database up to date: at the beginning of the first cycle respondents lived in 11 different municipalities, at the end of the second cycle there were respondents living in 114 different municipalities.

At the start of the second cycle 417 (13.4%) respondents had died and 13 (0.4%) respondents could not be located for reasons of (temporarily) moving abroad or an unknown change of address. Attrition due to death was related to age and sex ($p < 0.01$), more male respondents had died than female respondents, regardless of their age. No interactive effect was found of age and sex on attrition due to death, indicating that in all age-groups the precedence of male over female deaths was similar.

At the beginning of the second LASA cycle, 2677 respondents were still available for interviewing.

The main interview

The main interview included a broad selection of topics: demography, (mental) health-related questions and activities (e.g. depression, self-efficacy, chronic diseases, physical performance), and questions related to the social activities of respondents (e.g. social network contacts and social participation). To interview the 2677 respondents available for the main interview, 40 interviewers were selected. In the south and north-east of the Netherlands, local interviewers were recruited, mainly from the pool of interviewers who had conducted interviews for LASA in the first cycle. In the region of Amsterdam, new interviewers were recruited through advertisements in local newspapers and the bulletin boards of universities. Depending on their prior LASA interviewing experience, the interviewers received a three or five-

day interview training, in which they were taught standard survey interviewing techniques. Parts of the interview were practised by video-taped role-playing. The interviewers received a fieldwork manual which contained (1) guidelines on how to answer questions which were frequently asked by respondents, (2) basic instructions on how to conduct an interview on emotional or difficult topics and (3) instructions on how to react and who to contact when difficulties arise. During the fieldwork, personal feedback was given, based on listening to tapes that were recorded during an interview, with the respondents' consent. Every two months additional group training sessions were organized for interviewers.

All respondents received a personal letter, which included the name and telephone number of the interviewer. In the letter participants were informed that they would receive a small present in gratitude of their participation (a silver spoon with the logo of the Vrije Universiteit). All letters were personally signed by the Scientific Director of LASA. Approximately one week after the letter was sent an interviewer contacted the respondent to make an appointment for an interview. From the first LASA cycle it was known that especially frail elderly people might find the interview tiring. If so, they were given the opportunity to stop and continue it on another day. This happened in 130 interviews. Completeness of data-collection was given priority over extra cost factors, such as extended interview time and additional travelling costs of the interviewer. Only in a few cases ($n=17$), when the interview was discontinued did the interviewer not succeed in making another appointment.

For respondents who indicated that they did not want to, or could not participate in the study anymore, a special procedure was initiated. First, the refusal was accepted, but permission was asked to contact the respondent again at a later date. After a few months the fieldwork co-ordinator and some especially selected interviewers (those with the highest response rates) contacted the respondents again and tried to persuade them to participate. Participation was requested successively for a full interview, a short face-to-face version of the full interview, a short telephone interview, and finally a short telephone interview with a proxy. The last option was always tried if there were indications that the respondent was not longer able (too frail) to participate in the study.

During the fieldwork two especially assigned fieldwork co-ordinators supervised the activities. These various procedures resulted in data from 2545 respondents. An overview of the enrollment attempts can be found in Table 2.1, and a further specification is given in Table 2.2.

Table 2.1
Results of enrollment attempts for the main interview

	N	% of total	% of eligible
Deceased before approach	417	13.4	–
Refusals	90	2.9	3.4
Ineligible	38	1.2	–
Not contacted	4	0.1	0.2
Moved abroad, etc.	13	0.4	–
Data main interview present	2545	81.9	96.4
Total	3107	100.0	(2639) 100.0%

Table 2.2
Specification of the attrition

Deceased before approach	417
Refusals	90
– not interested/motivated	86
– bad experience previous LASA interview	4
Ineligible	38
– not able: cognitive indication	4
– not able: physical indication	34
Not contacted	4
Moved (not contacted)	13
– moved elsewhere in the Netherlands: address unknown	4
– moved abroad temporarily	3
– moved abroad permanently	6

The results of the enrollment attempts were satisfactory, but these figures should be seen in the context of the diversity of data-collection. Although data were collected from 2545 respondents, a full interview was only obtained from 2204 (71% of the original 3107 respondents and 83.5% of the 2639 eligible respondents). The distribution percentages of data-collection for the main interview are presented in Table 2.3.

Table 2.3
Sources of data collected

	N	% of total
Face-to-face interview completed	2204	86.6
Face-to-face interview not completed (short/terminated)	98	3.8
Telephone interview	165	6.6
Telephone interview with proxy	78	3.1
Total	2545	100.0

The mode of the interview was associated with the living arrangements of the respondents. Respondents who participated in the short version of the interview, or for whom a proxy answered the questions, lived less often independently than other respondents (Table 2.4).

Table 2.4
Distribution of independent vs institutionalized over the second cycle enrollment categories

	Complete	Short	Telephone	Proxy
Total	2204	98	165	78
Independent	96.1	72.4	84.0	66.7
Institutionalized	3.9	27.6	16.0	33.3

The medical interview

The medical interview consisted of health-related questions including smoking and drinking habits, diet and medication use. A number of cognitive tests, and measurements of blood pressure, body composition, lung capacity and vision were also carried out. A sub-sample of respondents participated in an additional, diagnostic interview on depression. A new section in the medical interview focused on risk factors for falls and osteoporotic fractures. This gave rise to additional questions on falling and additional antropometric measurements. Participants in the medical interview were also asked to visit a local hospital where blood samples, urine samples and

ultrasound measurements of the heels were taken. In a sub-sample of respondents (i.c. those living in the west of the Netherlands) extra Dual X-ray Absorptiometry (DXA) measurements and X-rays of the spine were taken by staff of the Academic Hospital of the Vrije Universiteit in Amsterdam.

Only those respondents who participated in a face-to-face interview, and were born before 1931, were asked to participate in the medical interview, which took place approximately two weeks after the main interview. For the medical interview, a total of eight interviewers were selected. Six interviewed the majority of the respondents, and the other two interviewers were on stand-by to replace interviewers who were on holiday or ill. The recruited interviewers had a medical background, and three had already been involved in the medical interviews of the previous cycle.

The interviewers received a training of approximately one week, provided by various experts, and received further specific training on body measurements during the fieldwork. They received a special fieldwork manual which contained instructions on all measurements. A special fieldwork co-ordinator supervised the fieldwork activities for the medical interviews. Another fieldwork co-ordinator supervised the hospital procedures. Of the 3107 respondents interviewed in the first LASA cycle, a total of 2429 respondents were born before 1931. Of these respondents, 397 (16.3%) were deceased and 11 (0.5%) respondents could not be contacted. This resulted in 2021 potential respondents for the medical interview. Only those respondent who took part in the face-to-face mode of the main interview (1720) were asked to participate in the medical interview. An overview of the enrollment attempts is given in Table 2.5.

Table 2.5
Results of enrollment attempts for the medical interview

	N	% of total	% of eligible
Deceased before approach	5	2.9	–
Refusals	160	9.3	9.6
Ineligible	43	2.5	–
Not contacted	3	1.7	1.8
Data medical interview present	1509	87.7	90.4
Total	1720	100.0	(1669) 100.0%

The reason most frequently mentioned by participants for refusing to participate in the medical interview was that they were tired of medical procedures. They indicated that they had seen more than enough nurses, physicians and hospitals, and did not want to see any more. Almost all participants who refused the medical interview were willing to remain active participants in other parts of the LASA study.

The self-administered questionnaire

Questions on topics such as personality traits, self-perceived health status, metamemory and social involvement were included in the self-administered questionnaire. All respondents who participated in the face-to-face mode of the main interview (2302 respondents), received this questionnaire, which they were asked to complete in the week after the personal interview. Respondents who participated in a medical interview were asked to hand in the completed questionnaire to the medical interviewer. The medical interviewer checked the questionnaire for completeness (item non-response) and offered help (additional explanations) if specific questions had not been answered. Those respondents who did not participate in the medical interview (refused or born in 1931 or later) received a pre-stamped addressed envelope, and were asked to return the self-administered questionnaire by mail.

To improve response rates compared with 1992–1993, a new procedure was implemented. A target date was calculated for all respondents. On this target date the self-administered questionnaire was expected to be mailed or handed in. If this was not the case a reminder letter, including a new questionnaire, was sent to the respondent with a repeated request to fill in the questionnaire. After three weeks, those respondents who had not reacted to the reminder were contacted by telephone. In the telephone contact, reasons for not reacting were asked, and assistance in completing the questionnaire was offered. A total of 2051 respondents (89.1% of those asked, 77.7% of the 2639 respondent available at the start) returned the questionnaire, which is an improvement in response rate of 15% compared with 1992/1993. Thus, although the new procedure was quite costly (sending reminders with new questionnaires) and time-consuming for the fieldwork staff, it proved to be worthwhile with regard to response rates. The results of the procedures are shown in Table 2.6.

Table 2.6
Results of enrollment attempts for the self-administered questionnaire

	N	% of total
Refused to complete questionnaire	251	10.9
Returned questionnaire	2051	89.1
Total	2302	100.0

Overview of LASA data 1992–1996

Table 2.7 gives an overview of the availability of LASA data. No correction has been made for partial non-response or item non-response, and it should be noted that some patterns are a consequence of the design i.c. no medical interview in the second cycle for those respondents born in 1931 or later.

For example: pattern D represents data on the respondents who participated in the main interview, the medical interview and the self-administered questionnaire in the first cycle, but were deceased or lost to follow-up in the second cycle (deceased, refused, moved). Pattern T represents the respondents for whom complete data are available (1127 respondents) after the second cycle.

Handling non-response

Almost every researcher involved in a longitudinal study is confronted with the phenomenon of missing data. From Table 2.7 it can be seen that LASA is no exception. Missing values due to unit or item non-response on variables in longitudinal designs result in a reduction in the size of the database, and therefore in less accurate estimates. Moreover a number of standard methods of analysis (e.g. multiple regression, (M)ANOVA) can only be applied to cases with complete data. Furthermore, bias in the results and the generalisation of those results to specific populations may exist if there are systematic differences (sex, age, health) between respondents and non-respondents with regard to certain variables.

Researchers try to solve the problem of missing data in different ways, e.g. by means of the imputation method, weighting procedures, or model-based procedures. For a comprehensive discussion on various approaches see Little and Rubin (1987). In the following paragraphs a number of different imputation methods will

Table 2.7
Overview of LASA data

Pattern	Mainc1	Medc1	Selfc1	Mainc2	Medc2	Selfc2	Number of cases	Percentage of case
A	1	-	-	-	-	-	166	5.3
B	1	1	-	-	-	-	151	4.9
C	1	-	1	-	-	-	37	1.2
D	1	1	1	-	-	-	451	14.5
E	1	-	-	1	-	-	37	1.2
F	1	1	-	1	-	-	26	0.8
G	1	-	1	1	-	-	11	0.4
H	1	1	1	1	-	-	77	2.5
I	1	-	-	1	1	-	14	0.5
J	1	1	-	1	1	-	26	0.8
K	1	-	1	1	1	-	1	0.0
L	1	1	1	1	1	-	61	2.0
M	1	-	-	1	-	1	42	1.4
N	1	1	-	1	-	1	81	2.6
O	1	-	1	1	-	1	40	1.3
P	1	1	1	1	-	1	479	15.4
Q	1	-	-	1	1	1	69	2.2
R	1	1	-	1	1	1	192	6.2
S	1	-	1	1	1	1	19	0.6
T	1	1	1	1	1	1	1127	36.3
Total							3107	100.0

Legend:

1: data; -: no data; Mainc1: Main interview cycle 1; Medc1: Medical interview cycle 1; Selfc1: Self-administered questionnaire cycle 1; Mainc2: Main interview cycle 2; Medc2: Medical interview cycle 2; Selfc2: Self-administered questionnaire cycle 2.

be discussed. It should be emphasized that there is no single solution for the problem of missing data. The approach chosen should always be placed in the context of the research question to be answered.

Imputation methods

Group mean

In the method of imputation of the 'group mean', various approaches are followed. They all have in common that the missing value is replaced by some sort of 'average', with continuous variables as the mean (group or overall), with nominal variables as the category with the highest frequency and, if the variable has an ordinal level of measurement, the middle category is often chosen for imputation. All kinds of nuances are found in this approach.

When the number of missing data is small, as is often the case in constructed variables (scales), this can be a very useful and simple way to replace missing values. It should be realised, however, that by imputing in such a way the resulting variability of the imputed distribution is an under-estimation of the true variability, which increases the chances of finding non-existing 'significant results' from data imputed with this method.

Hot deck

In the 'hot deck' approach missing values are replaced with values from respondents who are similar to nonrespondents on variables observed for both (matching variables). An example: the scores of a respondent on sex, age, education and marital status are known, but the score on a specific variable, e.g. evaluation of health, is missing. The value of the score on the evaluation of health from a sampled matched respondent (matched on sex, age, education and marital status) with a score on evaluation of health is imputed for the person with the missing value.

Regression estimates

In this approach, the missing value of the variable is predicted by means of a regression model. Various approaches are possible dependent on e.g. the measurement level of the variables. The basic approach is to estimate the missing variables for a respondent on the basis of predicted values from the regression on the known variables for that respondent. For example, in a study on the effect of education and age on depression, for a number of respondents the depression score is missing, but the scores on education and age are known. It is then possible to calculate – for the complete cases – a regression equation in which the score for depression is estimated. For the incomplete cases the depression score can be calculated and imputed. As a consequence of this approach, the predicted and imputed values are located on the estimated regression line, and therefore under-estimate the true variability of the depression scores for incomplete cases.

Last Value Carried Forward

The Last Value Carried Forward (LVCF) approach can be used when there are multiple observations per subject, for example in longitudinal data. The last value observed is used to fill in missing values at a later stage in the study. The assumption of LVCF is that the response remains constant after the last observed value. If the timing and withdrawal of the non-respondent is related to the phenomena under study, LVCF imputations can be biased. For example, in the case of degenerative processes such as cognitive decline or chronic diseases, the assumption that responses remain constant is clearly violated and the resulting variability of the imputed distribution is an under-estimation of the true variability and changes.

Multiple imputations

The idea behind multiple imputations is that for each missing value in a dataset several values are imputed. This method proposed by Rubin (see among others, Rubin 1987), corrects the problems associated with the single imputation methods mentioned above. The most important disadvantage of single imputation is that the single value imputed cannot reflect the sampling variability, and therefore systematically under-estimates uncertainty. By imputing several copies of missing data, different sets of 'complete data' are formed for which summary statistics and standard errors can be calculated and the uncertainty in the statistic can be assessed by an analysis of variance 'within and between' imputations.

New software is under development, in which the imputation methods described above will be easily available, and it will be possible to evaluate the consequences of different imputation methods for the research question to be answered (e.g. Solas 1998).

Conclusions

After the second cycle of LASA two remarks on data-collection and the database should be made. First, the investments made in keeping the respondents involved in the LASA study (newsletters, tracking respondents through municipal records, offering different options for data-collection) seem to have paid off. Apart from attrition due to natural causes (death, ineligibility), attrition due to refusal rates are low. Therefore, the LASA database has become a large and rather unique, high quality source for multi-disciplinary research.

Unfortunately, due to systematic attrition, specific inclusion criteria for certain sections of the LASA design (e.g. born in or before 1930 for the medical interview in

cycle 2) and participation in satellite projects, the LASA database has also become rather complex. Keeping track of which data are available for which respondent, is becoming a time-consuming activity. Furthermore, the rather complex sampling procedures, in combination with selective attrition or non-response, imply that researchers should be very cautious in generalizing their results to the general population. Many tests of significance, for example, use standard errors with the assumption that these standard errors are the result of simple random sampling techniques, where non-response or attrition is at random for the variables used in the study. The question of whether this assumption is still appropriate, or that more complex methods are required, will probably need to be addressed in many future research projects which make use of the LASA data.

Finally, the complexity of the LASA database requires documentation of extremely high quality in the database in order to prevent under-use or improper use of data collected in the context of LASA.

References

- Little RJA, Rubin DB (1987). *Statistical Analysis with Missing Data*. New York: Wiley & Sons.
Rubin DB (1987). *Multiple Imputations for Nonresponse in Surveys*. New York: Wiley.
Smit JH, Vries MZ de (1994). Procedures and result of the fieldwork. In Deeg DJH, Westendorp-de Serière M (eds), *Autonomy and Well-being in the Aging Population I: Report from the Longitudinal Aging Study Amsterdam 1992–1993*. Amsterdam: VU University Press.
Solas (1998). *Solas for Missing Data Analysis*. Cork: Statistical Solutions Ltd.