*Quality & Quantity* **32:** 77–91, 1998.



# Handling Missing Data by Re-approaching Non-respondents

MARK HUISMAN[1], BOUDIEN KROL[2] & ERIC VAN SONDEREN[2]
[1]*Department of Statistics, Measurement Theory & Information Technology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS, Groningen, The Netherlands;* [2]*Northern Center for Healthcare Research, University of Groningen, The Netherlands*

**Abstract.** When handling missing data, a researcher should be aware of the mechanism underlying the missingness. In the presence of non-randomly missing data, a model of the missing data mechanism should be included in the analyses to prevent the analyses based on the data from becoming biased. Modeling the missing data mechanism, however, is a difficult task. One way in which knowledge about the missing data mechanism may be obtained is by collecting additional data from non-respondents. In this paper the method of re-approaching respondents who did not answer all questions of a questionnaire is described. New answers were obtained from a sample of these non-respondents and the reason(s) for skipping questions was (were) probed for. The additional data resulted in a larger sample and was used to investigate the differences between respondents and non-respondents, whereas probing for the causes of missingness resulted in more knowledge about the nature of the missing data patterns.

**Key words:** missing data, follow-up, cause of missingness, scale data.

## 1. Introduction

Missing data is a problem a researcher is often confronted with. There are frequently persons not answering all questions in a questionnaire and the resulting item non-response can cause serious problems (Little & Schenker, 1995). The procedures to treat the missing data can be grouped into three categories (Little & Rubin, 1987): (1) weighing procedures, (2) imputation procedures, and (3) direct analysis of the incomplete data. An important feature determining the success of a treatment procedure is the mechanism underlying the missing data. When data are non-randomly missing, analyses may be seriously biased due to differences between respondents and non-respondents, and the missing data mechanism should be modeled in the analysis.

Insight into the missing data mechanism requires knowledge about which factors contribute to not answering questions. When the cause of missingness is known the missing data mechanism is called accessible, and when included properly in the analysis it will cause no bias even if the data is non-randomly missing (Graham & Donaldson, 1993). Inaccessible, non-ignorable missing data mechanisms, however,

can only be analyzed by either making reasonable guesses about the mechanism based on what one considers the true mechanism to be (Little & Rubin, 1987; Rubin, 1987), or by collecting additional data to make the mechanism accessible (Graham & Donaldson, 1993).

In this paper we describe the method of re-approaching non-respondents who participated in a study among patients on a waiting list of orthopaedic practices (Krol, 1996). These follow-ups are used to obtain answers to previously skipped questions and to find the cause of missingness. Follow-ups are usually applied in situations when there are complete cases missing (unit non-response) (see e.g., Rao, 1983). In this study, however, the sample of follow-ups consists of respondents who only skipped certain questions.

Part of the questionnaire used in the waiting list study consists of scales measuring latent properties of the respondents. Therefore, some of the outcome variables of interest are not observed, but are latent traits computed from several observed items. These latent outcome variables are considered missing if one or more observable items with which they are determined are missing. The new answers obtained by re-approaching the non-respondents will be used to compute the latent outcome variables, and the sample of follow-ups will be used to investigate if there are systematic differences between respondents and non-respondents with respect to these latent traits (see also Rubin, 1987).

In Section 2 the data derived from the study among the patients of orthopaedic practices are described followed by the method of re-approaching the persons who did not answer all questions. The results of the re-approachment are presented in Section 3. These results include a description of the extent of the missing data and the success of the re-approachment in terms of response rates, obtained reasons for missingness, and additional data.

## 2. Methods

### 2.1. THE DATA SET

The study was conducted at the request of a health insurance company in order to obtain more insight in the waiting list problem of orthopaedic practices. By means of a questionnaire, the entire referral traject, from the general practioner to the ultimate treatment by the specialist, was highlighted. From all possible orthopaedic conditions the seven most frequently presented were chosen. To this end the patients treated by an orthopaedic surgeon from the end of 1994 until the beginning of 1995 were asked to cooperate.

The first part of the questionnaire used in this study was directed towards questions concerning socioeconomic status and detailed questions as to the entire traject from the beginning of the orthopaedic complaints until the ultimate treatment by the orthopaedic surgeon. The second part of the questionnaire contained three standardized scales with two or more subscales each. The *Groningen Activity Restriction Scale* (GARS) evaluates the level of activity restrictions and consist of

two subscales: Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) (Kempen, Doeglas & Suurmeijer, 1993). Two subscales of the *Multidimensional Pain Inventory – Dutch Language Version* (MPI-DLV) are included, measuring pain-relevant psychosocial aspects (Lousberg, 1994). Finally, five subscales of the *Nottingham Health Profile* (NHP) (Hunt, McKenna & McEwan, 1993) are included in the second part of the questionnaire, measuring the amount of energy, emotional well-being, sleep, physical mobility, and social isolation of the respondents. After the three standardized questionnaires, two sets of eight questions evaluating the change in the concepts measured by the subscales mentioned above were added. The first set dealt with the change during the referral traject and the second with the change after the treatment by an orthopaedic surgeon.

According to the files of the health insurance company a total of 1891 patients were treated by an orthopaedic surgeon for one of the seven selected conditions. The questionnaire was sent to all these patients and a reminder sent three weeks later to those individuals who did not return their questionnaire. Ultimately, 1330 insurants responded (71%). Of those 56% were females and 44% males. The mean age of the sample was 48.9 years (range: 7–96; sd: 19.7); 69% were married, the educational level of 30% of the respondents was low (only primary school) and 4% were highly educated (university). Of the insurants 42% had a paid job, 19% were housewives, 17% received an elderly pension and 13% received income from social security for the disabled. For 13% of the patients the questionnaire was filled in by a relative or friend, and the most frequently mentioned orthopaedic condition were knee complaints (61%). In Table I the demographic data of the sample are represented.

### 2.2. RE-APPROACHING NON-RESPONDENTS

In order to handle the missing data properly, one needs information about the mechanism underlying the missingness, and therefore information about the non-respondents. The most direct method to obtain information on non-respondents is to re-approach some of them. Other ways to obtain information is by using theory, logic, or prior data, and make reasonable guesses about the mechanism. Graham & Donaldson (1993) describe these and other strategies for modeling the missing data mechanism.

When re-approaching non-respondents, three important points have to be considered: (1) the sample of non-respondents, (2) the method of re-approaching, and (3) the moment of re-approachment.

In this paper the *group of non-respondents* is determined using the second part of the questionnaire, consisting of the scales and the questions concerning change, i.e., the respondents with missing values in the scales are taken into account. However, not all respondents with missing scale items were followed-up. If the reason for the missing data was known (explicitly stated by the respondent or determined

*Table I.* Demographic variables of the waiting list study ($N$=1330).

| | | |
|---|---|---|
| **Sex:** | Female | 56.4% |
| ($N$ = 1329) | Male | 43.6% |
| **Age:** | 48.9 years | range: 7–96 |
| ($N$ = 1205) | | sd: 19.7 |
| **Civil status:** | Married | 68.7% |
| ($N$ = 1296) | Not married | 16.7% |
| | Widowed | 12.4% |
| | Divorced | 2.2% |
| **Educational level:** | Primary school | 29.9% |
| ($N$ = 1222) | LBO | 28.6% |
| | MULO/MAVO | 11.9% |
| | MBO | 19.6% |
| | HAVO/VWO | 5.8% |
| | HBO/University | 4.3% |
| **Occupational status:** | Retired | 16.6% |
| ($N$ = 1177) | Employed | 42.1% |
| | Housewife/man | 19.0% |
| | Disabled income | 12.6% |
| | Unemployed | 3.7% |
| | Student | 6.0% |
| **Help with filling in questionnaire:** | No | 86.8% |
| ($N$ = 1330) | Yes | 13.2% |
| **Orthopaedic condition:** | Knee | 60.6% |
| ($N$ = 1328) | Hip | 17.5% |
| | Shoulder/elbow | 7.5% |
| | Ankle/foot | 10.6% |
| | Wrist/hand | 3.8% |
| | Back | 0.2% |

from earlier answers) and valid, the respondent was not re-approached. Therefore, the group of non-respondents which were candidates to be followed-up, consisted of respondents with missing data in part two of the questionnaire with an unknown reason. This method is depicted in Figure 1.

A sample of the above-mentioned non-respondents was taken and these individuals were followed-up. This sample will help to obtain the desired information on the non-respondents and the missing data mechanism. Little & Rubin (1987) and Graham & Donaldson (1993) argue that even if only a small sample is followed-up,

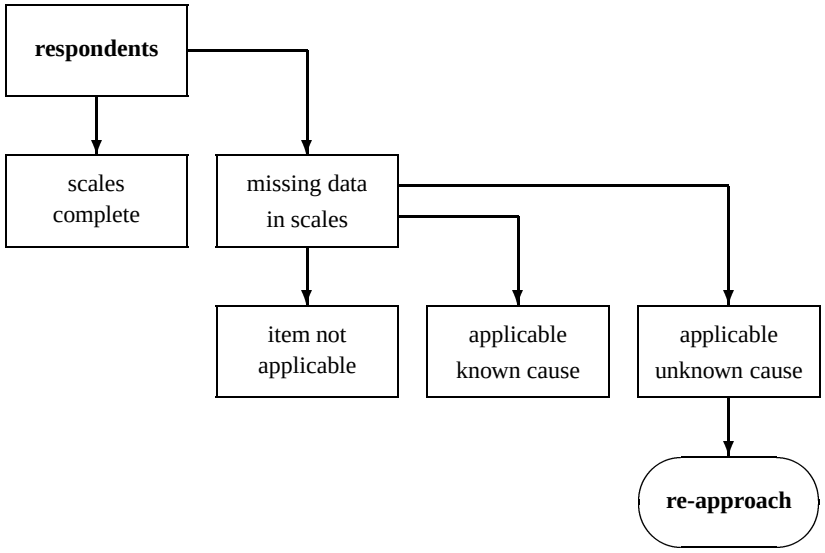


*Figure 1.* The re-approaching of respondents with missing data in the scales.

this can be very helpful in reducing sensitivity of inference caused by the missing data.

The *method of re-approachment* was largely dependent on the amount of missing data, and to a smaller extent, on the (expected) nature of the missingness. Telephone interviews were conducted with the non-respondents who had a small or moderate amount of missing data, regardless of the nature of the missingness. With a standard introduction, the non-respondent was told that some obscurities emerged while processing the data and (s)he was asked whether some questions could be repeated. When the respondent was unwilling to respond a second time, the interviewer was instructed to probe for the reason for missing data. Respondents with a large amount of missing data, especially with missing values for (groups of) consecutive questions (e.g., skipping one or more pages), were sent a copy of the questionnaire with the request to answer the missing questions or, when the respondent was unwilling to give an answer, to state the reason for the non-response.

The last point to consider is the *moment of re-approachment.* The time between the returning of the questionnaire and the follow-up had to be as short as possible. Especially when probing for reasons for missing data, the respondent should be able to remember the situation of the first (non)-response. Therefore, the non-respondents were contacted within two weeks after the return of the questionnaire.

### 2.3. SCENARIO

Two pilot studies were conducted with 17 and 15 respondents, respectively. Based on the results of the pilot (with respect to missing data), the following scenario was followed after the returning of the questionnaires:

1. A check whether or not there were any missing data in the questionnaire and an evaluation if the respondent belonged to the group of non-respondents.
2. Determination whether the non-respondent should be re-approached (only a sample of the non-respondents was followed-up), and whether the re-approach would be held by mail or telephone.
3. In case the telephone number of the second group of non-respondents was not known, they would be re-approached by mail.
4. Pages with missing data were copied and send to the non-respondents, with the request to fill in the unanswered questions or to state the reason of missingness.
5. Telephone interviews were to be held within two weeks after receiving the questionnaire to obtain answers on skipped items or to probe the causes for missing data.

## 3. Results

First, the response rates of the waiting list study and of the re-approaching will be discussed. Also the occurrence of missing data is investigated in both samples. Second, the results of re-approaching non-respondents in terms of causes of non-response and obtaining additional data will be shown.

### 3.1. RESPONSE

Of 1891 patients treated by an orthopaedic surgeon, 71% responded ($N$ = 1330). In this sample up to 1237 (93%) had one or more missing values in the entire questionnaire. Following the scenario in the previous section, 435 (33%) non-respondents were selected to be re-approached. The results of the re-approachment procedure are shown in Table II. The number of non-respondents, the method of re-approaching, and the success of the re-approachment (in terms of response rates) are shown for every week in which questionnaires were returned.

Of the 435 non-respondents, 95 were not re-approached because either the item was not applicable or the cause of missingness was known. Of the remaining 340 persons, 279 (82%) were followed-up by telephone and 61 (18%) by mail. 256 non-respondents gave new answers and/or stated the reason for not answering the question(s).

Of the sample of 340 respondents to follow-up, 61% were females and 39% males. The mean age was 56.8 years, which is higher than the mean age of the original sample. The educational level of 46% of the non-respondents was low and 2% was highly educated.

*Table II.* Results of the re-approaching of non-respondents – method of re-approaching and response. For the re-approach by telephone a difference is made between not reached (no resp.) and unwilling to cooperate (unwil.).

| Week | Number of non-respondents | Re-approaching | | | Telephone | | | Mail | |
|---|---|---|---|---|---|---|---|---|---|
| | | None | Tel. | Mail | Resp. | Unwil. | No resp. | Resp. | No resp. |
| 1[a] | 30 | 1 | 23 | 6 | 22 | 1 | 0 | 5 | 1 |
| 2 | 249 | 60 | 154 | 35 | 114 | 17 | 23 | 20 | 15 |
| 3 | 61 | 13 | 40 | 8 | 31 | 4 | 5 | 4 | 4 |
| 4[b] | 16 | 3 | 11 | 2 | 8 | 2 | 1 | 1 | 1 |
| 5 | 56 | 14 | 32 | 10 | 27 | 1 | 4 | 7 | 3 |
| 6 | 23 | 4 | 19 | 0 | 17 | 0 | 2 | 0 | 0 |
| Total | 435 | 95 | 279 | 61 | 219 | 25 | 35 | 37 | 24 |
| Percentage | 100 | 22 | 64 | 14 | 50 | 6 | 8 | 9 | 6 |

[a] There was only one day on which the questionnaires returned.

[b] The reminder was sent this week to respondents who had not returned their questionnaire.

*Table III.* Occurrence of item non-response.

| | | | | Units | Mean | Item non-response | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $N$ | $k$ | Cat | INR | INR | $Q_0$ | Min | Max | d$Q_0$ | $nMD$ |
| GARS | 1309 | 18 | 4 | 11.8 | 0.44 | 2.5 | 0.5 | 6.3 | 12 | 0 |
| MPI-DLV | 1275 | 13 | 7 | 26.1 | 0.86 | 6.6 | 2.6 | 14.5 | 7 | 0 |
| NHP | 1288 | 30 | 2 | 20.3 | 1.37 | 4.6 | 2.5 | 7.9 | 6 | 0 |
| Change | 1285 | 16 | 5 | 13.5 | 0.72 | 4.5 | 2.8 | 6.7 | 2 | 0 |
| GARS[a] | 418 | 18 | 4 | 31.6 | 1.27 | 7.1 | 1.4 | 18.4 | 12 | 0 |
| MPI-DLV[a] | 388 | 13 | 7 | 62.6 | 2.43 | 18.7 | 8.0 | 31.7 | 8 | 0 |
| NHP[a] | 401 | 30 | 2 | 50.1 | 4.02 | 13.4 | 6.7 | 21.2 | 6 | 0 |
| Change[a] | 397 | 16 | 5 | 33.0 | 2.05 | 12.8 | 7.6 | 18.9 | 3 | 0 |

$N$: number of respondents; $k$: number of items in the scale; Cat: number of categories; units INR: percentage respondents having at least one item missing; mean INR: mean of the number of items missing across all respondents; $Q_0$: overall percentage missing data; Min: minimum percentage missing data on an item; Max: maximum percentage missing data on an item; d$Q_0$: number of items for which the proportion missing differs significantly from $Q_0$; *nMD*: number of items having no missing values.
[a] The subsample of 435 non-respondents.

### 3.2. OCCURRENCE OF MISSING DATA

Table III shows the results of the investigation of the occurrence of item non-response (INR) in the second part of the questionnaire containing the scales. For example, there are 1275 respondents ($N$) who answered some or all of the items ($k$) of the MPI-DLV. Of these respondents, 26.1% had at least one item missing (unit INR). The mean number of missing items was 0.86 (mean INR) and the overall percentage missing in the scale is 6.6% ($Q_0$). The observed percentages missing per item have a minimum of 2.6% (min) and a maximum of 14.5% (max). There are seven items for which the proportion missing differs significantly from $Q_0$ (d$Q_0 = 7$), and there are no completely observed items ($nMD = 0$).

Table III illustrates that items of the MPI-DLV are the most frequently skipped. On the other hand, the items of the GARS, show the largest differences between proportions missing for every item. This heterogeneity of the proportions missing for every item can be indicative for the non-randomness of the missing data mechanism (Huisman, 1998).

The second part of Table III shows the results for the subsample of 435 non-respondents. Apart from the larger non-response rates, the same picture emerges as in the first part of the table.

When handling item non-response, the nature of the missing data mechanism is very important (Little & Rubin, 1987; Graham & Donaldson, 1993). Covariate information can be used to make assumptions about the missing data mechanism, and, therefore can be used in the analyses. Analysis of variance is used to determine

*Table IV.* Correlates of item non-response – significant main effects only. The effect of every covariate is corrected for the effect of the other covariates.

| | Sex | Age | Education | Condition | Fill in |
|---|---|---|---|---|---|
| Part I | | + | − | +/− | − |
| GARS | + | + | | | |
| MPI-DLV | | + | − | | − |
| NHP | | + | − | +/− | − |
| Change | | + | | +/− | |

Sex: 1 female, 2 male; Age: 1 <25, 2 25–40, 3 40–65, 4 65–80, 5 >80; Education: 1 primary school, 2 LBO/MAVO, 3 MBO, 4 HAVO/VWO, 5 HBO/University; Condition: 1 knee, 2 hip, 3 shoulder/elbow, 4 ankle/foot/wrist/hand; Fill in: 1 someone else, 2 self.

whether the number of missing values per scale varies by sex, age, educational level, or orthopaedic condition, and which person filled in the questionnaire. In Table IV the significant main effects and their direction can be found, here + indicates that the number of missing items increases with higher levels of the covariate, and − indicates a decreasing number of missing items. For the covariate 'condition' different directions were found in different parts of the questionnaire.

The covariate sex does not have a significant effect on the missing data, except for the GARS. However, contrary to earlier findings (Huisman, 1998), men tend to have more missing values than women. In the entire questionnaire age is also an important factor for missingness and older respondents tend to have more missing values. In the different parts of the questionnaire the other three covariates have significant effects also. Additionally, some significant second-order interactions were found, especially the interactions between sex and 'fill in', and between age and education.

### 3.3. CAUSES OF ITEM NON-RESPONSE

Item non-response can have several different reasons and for the treatment of missing data it is important to know the cause. Although some respondents state the reason for not answering an item, the causes of missing data are usually not known at the time the questionnaires are returned. The telephone follow-up was used to detect causes for missing data. When the non-respondent was unwilling to give answers to skipped questions, the reason for missingness was probed.

The causes of item non-response can be divided into six groups which are not always mutually exclusive.

*Missing by design.* Although branching prevents a respondent from having to answer questions which do not apply to him/her, there still are some questions inapplicable for respondents. Some reasons given (literally) by the non-respondents are:

- did not have to wait/accident/acute admission/waiting list short;
- no physician/already treated by surgeon;
- not (yet) operated on/still recovering/hospitalized while waiting;
- scale-items inapplicable: no work/no partner/no children/no pain.

*Inapplicable item.* The non-respondent wrongfully thinks that an item does not apply to him/herself:

- question irrelevant for waiting list problem/I was already unable to do that before waiting;
- scale-items inapplicable: I never do that/my wife does that/my mother does that/I need help doing that/I use appliances.

*Cognitive task too difficult.* The non-respondent has problems remembering situations which occurred earlier or understanding the question(s):

- have had several operations for same or different complaint;
- have several complaints/complaints other than orthopaedic ones;
- question too difficult/do not understand question/strange question.

*Refuse to respond.* The non-respondent refused to answer a question without giving a (clear) reason for it.

*Don't know*. Non-respondent gave non-substantive answer: don't know.

*Inadequate score.* The response did not fit in the given response categories:

- respondent gave two answers to one question;
- gave answers like: sometimes/a few/not always/depends on situation.

Table V shows the percentage non-respondents with missing data from one or more categories. A difference is made between: (1) what is known about the cause of missingness based on the comments of respondents and answers given on earlier questions when the questionnaire is returned, and (2) what is known after the follow-up (either by telephone or mail). The percentages do not sum to 100% because a non-respondent can state different reasons for different skipped questions.

The categories *inapplicable item*, *cognitive task too difficult*, and *refuse to respond* are taken together in the remainder of this section. This new category is

*Table V.* Number of and percentage non-respondents who mentioned one or more causes of item non-response falling in one of the six categories. First the information after the returning of the questionnaires is given, second the extra information obtained by re-approaching.

| | | Questionnaire | | Re-approaching | | Both | |
|---|---|---|---|---|---|---|---|
| | *N* | 263 | (%) | 172 | (%) | 435 | (%) |
| **Design**: | did not have to wait | 17 | 7 | 10 | 6 | 27 | 6 |
| | no physician | 11 | 4 | 12 | 7 | 23 | 5 |
| | not operated | 16 | 6 | 11 | 6 | 27 | 6 |
| | inapplicable | 37 | 14 | 43 | 25 | 80 | 18 |
| **Inapl**: | irrelevant for waiting | 8 | 3 | 7 | 4 | 15 | 3 |
| | scale-item inapplicable | 40 | 15 | 57 | 33 | 97 | 22 |
| **Cogn**: | several operations | 15 | 6 | 13 | 8 | 28 | 6 |
| | several complaints | 12 | 5 | 7 | 4 | 19 | 4 |
| | question too difficult | 15 | 6 | 22 | 13 | 37 | 9 |
| **Ref**: | refused to answer | 13 | 5 | 10 | 6 | 23 | 5 |
| **Don't**: | don't know | 45 | 17 | 37 | 22 | 82 | 19 |
| **Inad**: | two answers given | 30 | 11 | 19 | 11 | 49 | 11 |
| | sometimes/a few etc. | 48 | 18 | 46 | 27 | 94 | 22 |
| | unknown | 182 | 69 | 80 | 47 | 262 | 60 |

called *item non-response.* Hence, four main categories of missing data remain, namely, *missing by design, item non-response, don't know,* and *inadequate score.*

Of the 340 re-approached non-respondents, 172 gave one or more reasons for not answering the question(s). From the rest of the sample of 451 non-respondents information about the cause of missingness is obtained only by comments written on the questionnaire or follows from answers to earlier questions.

From Table V it follows that re-approaching provides extra information on all four main categories of missing data. The percentages of non-respondents with missing data belonging to these categories are 29, 43, 19, and 31%, respectively. From 82% of the non-respondents at least one reason for missingness was obtained. Most missing data belong to the category *item non-response* (43%), of which the subcategory *inapplicable* is the largest, 25% of the non-respondents state this reason. A substantial part of the causes refer to *missing by design* (29%). This is not due to respondents, but to incorrect branching of the questionnaire. Except for the category *don't know,* all categories of causes occur most frequently in the scales.

*Table VI.* Results of the re-approaching of non-respondents – new answers to previously skipped questions.

| | | | |
|---|---|---|---|
| Selected non-respondents | 435 | of which 95 not followed-up | |
| Followed-up | 340 | | |
| By *telephone* | 279: | 219 | of which responded (answers and/or reasons for missingness). |
| | | 211 | of which gave answers to previously skipped questions, of which 67% answered all skipped questions. |
| By *mail* | 61: | 37 | of which responded (answers and/or reasons for missingness). |
| | | 35 | of which gave answers to previously skipped questions, of which 69% answered all skipped questions. |
| Total | | 246 | respondent gave new answers, |
| of which | | 231 | answered (some or all) skipped scale items and |
| | | 150 | answered (some or all) skipped non-scale items. |

### 3.4. NEW ANSWERS

Table VI shows the results of the re-approaching of non-respondents with respect to obtaining answers to previously skipped questions.

From Table VI it follows that 211 (96%) of the 219 non-respondents followed-up by telephone gave answers to some or all previously skipped questions. In the follow-up by mail 95% of the non-respondents gave new answers. In all, the follow-up resulted in answers on skipped questions from 246 non-respondents of whom 231 answered skipped scale items. Of those 231 non-respondents, most had missing data for the MPI-DLV (61%) and least for the questions concerning change (22%). Note that in contrast to Table III, non-respondents who skipped every item of one scale are taken into account. The percentages respondents with missing data in the four categories of missing data are of the same order as those in Table V except for *don't know*. This category occurs most frequently in Part I of the questionnaire.

Logistic regression was used to predict whether, or not, non-respondents, with missing data in one of the categories mentioned earlier, will give an answer to an item previously skipped. In all, three different models were tested: (I) a model with all categories of missing data, (II) a model with all categories of missing data and covariates, and (III) a model with only significant main effects from model (II) with a stepwise procedure. The results are presented in Table VII.

*Table VII.* Results of the logistic regression – significant main effects only.

| | Missing data categories | | | | | | Covariates | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | design | inapl | cogn | ref | don't | inad | sex | age | educ | cond | fill in |
| (I) | | + | – | | | | | | | | |
| (II) | – | + | | | | + | | | | [a] | |
| (III) | – | + | – | | | + | | | | [a] | |

[a] Three dummy variables for condition, **significant**.

Whether or not a non-respondent will give an answer when followed-up is correctly predicted in 61, 65 and 63% of the cases, respectively. The categories of missing data are most important for this prediction. In Table VII only the direction of the significant main effects is presented. From the table it follows that when respondents have missing data in the categories *missing by design* and *cognitive task too difficult*, the probability of additional responses decreases. The same holds for *don't know*, but the effect was not significant. With missing data in the categories *inapplicable, inadequate score*, and *refuse*, the probability of additional responses increases, although the last one not statistically significant. The direction of the change in probability of response for the covariates is the same as in Table IV, however, only the covariate condition has a significant effect.

The difference between the latent outcome variables for respondents and the followed-up non-respondents was tested. The mean score on the GARS proved to be significantly different for both groups, especially the score on the ADL subscale. The follow-ups tend to have higher scores, indicating more restrictions with activities of daily living. The scores on the subscales of the NHP also showed significant differences between respondents and follow-ups, except for the subscale 'amount of energy'. A higher score for the follow-ups means that for the NHP they have more problems with the trait measured by the scale. Finally, there were no significantly different mean scores for the two subscales of the MPI-DLV. All tests were corrected for sex and age.

## 4. Discussion

The collection of additional data from non-respondents is the best way one can obtain information about the missing data mechanism. From a statistical point of view, the best method to collect new data is re-approaching a random sample of non-respondents. Obtaining a random sample of non-respondents, however, is difficult to accomplish. In this paper we ignored the fact that the sample may not be random and that the additional data were not collected in a manner identical to the original approach of the respondents.

The re-approaching was useful in gaining better insight into the problem of missing data in two ways. First, probing for reasons for skipping questions re-

sulted in a more complete understanding of the nature of the missing data patterns. Although the grouping of the causes in Section 3.3 is rather item specific, because the reasons for missingness are dependent upon the question being asked, one can detect from these categories whether the respondent or the questionnaire caused the missingness, and therefore can be more certain about the mechanism. In this study we found that a substantial part of the non-response was due to the questionnaire; respondents could not answer certain questions because these did not apply to them, or their response did not match one of the response options offered. But also the respondents themselves caused a substantial part of the missing data; they could not remember what had happened, or thought (wrongly) the question did not apply to them. It is, however, hard to isolate one cause. There are many different causes for many different items, even for items belonging to the same scale.

Second, the additional data from the follow-ups not only resulted in a larger sample, but can also be used to investigate the (systematic) differences between the respondents and the non-respondents. The covariates with different scores for respondents and non-respondents (for instance, age and education) can be used to analyze the data under the assumption of 'missing at random' (Little & Rubin, 1987), which allows missingness to be dependent on completely observed covariates. The observed difference in latent traits for some scales (GARS, NHP) can be used to make assumptions about the missing data mechanism (e.g., respondents with more emotional problems have more missing values for this subscale of the NHP), which should be included in the analysis.

We found that the assumption of randomly missing data does not hold in the waiting list study. The missingness is not only dependent on certain covariates, but in some cases also on the latent outcome variable. Therefore the follow-up is useful for properly handling of the missing data (see also Rubin, 1987, Sections 6.6 and 6.7, and Graham & Donaldson, 1993). Graham & Donaldson (1993) even suggest that researchers should build an additional data collection into their projects. The extra effort and costs will result in better estimations of scores and the sample of non-respondents does not need to be large for finding solutions for the problem of missing data.

## Acknowledgement

The authors are grateful to Ruud Teunter for his assistance in the telephone interviews, and to Ivo Molenaar for his comments on earlier versions of this paper. Mark Huisman's research is supported by the Netherlands Research Council (NWO), Grant 575-67-048.

## References

Graham, J. W. & Donaldson, S. I. (1993). 'Evaluating interventions with differential attrition: The importance of non-response mechanisms and use of follow-up data', *Journal of Applied Psychology* 78: 119–128.

Huisman, M. (1998). 'Missing data in behavioral sciences research: Investigation of a collection of data sets', *Kwantitatieve Methoden* 57: 69–93.

Hunt, S. M., McKenna, S. P. & McEwen, J. (1993). 'Nottingham health profile (NHP)', in C. Koenig-Zahn, J. W. Furer & B. Tax (eds), *Het meten van de gezondheidstoestand; 1-Algemene gezondheid.* Assen: Van Gorcum, pp. 100–114.

Kempen, G. I. J. M., Doeglas, D. M. & Suurmeijer, Th. P. B. M. (1993). *Het meten van problemen met zelfredzaamheid op verzorgend en huishoudelijk gebied met de Groningen Activity Restriction Scale* (GARS): *een handleiding* [Measuring problems with independent functioning in daily life with the Groningen Activity Restriction Scale]. Groningen: Northern Center for Healthcare Research (NCG).

Krol, B. (1996). *Beleefd wachten. Een onderzoek naar de wachtduur bij orthopedische patiënten* [Waiting politely. Investigation of the duration of waiting lists for orthopaedic patients]. Groningen: Groeneland verzekeringen, Northern Center for Healthcare Research, University of Groningen.

Little, R. J. A. & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Little, R. J. A. & Schenker, N. (1995). 'Missing data', in G. Arminger, C. C. Clogg & M. E. Sobel (eds), *Handbook of Statistical Modeling for the Social and Behavioral Sciences.* New York: Plenum Press, pp. 39–75.

Lousberg, H. B. (1994). 'Chronic pain: Multiaxial diagnostics and behavioral mechanisms', Ph.D. dissertation, University of Limburg.

Rao, P. S. R. S. (1983). 'Callbacks, follow-ups, and repeated telephone calls', in W. G. Madow, I. Olkin & D. B. Rubin (eds), *Incomplete Data in Sample Surveys, Vol. II: Theory and Bibliographies*. New York: Academic Press, pp. 33–44.

Rubin, D. B. (1987). *Multiple Imputation for Non-response in Surveys.* New York: Wiley.