3. Method

The six major research questions described in Chapter 1 are addressed using secondary data. The option of using secondary data is inextricably linked with the financial advantages it offers. However, disadvantages of data gathered by other research teams such as operationalisation of concepts or the specific sample composition must also be taken into account. Therefore, the decision to make use of secondary data for the present study was the result of careful consideration of the drawbacks and the advantages offered by the available data. In the first and second section of this chapter the longitudinal dataset and the procedure of data gathering are described. Because only married couples are eligible for our research purpose, the third section discusses the selection of this sub-sample. Since non-response is a pervasive problem in longitudinal datasets, the fourth section evaluates the external and internal validation of the sample. However, besides non-response of respondents, item non-response might also pose difficulties. The fourth section therefore additionally addresses the procedure that was used to save cases with missing values.

3.1. DUTCH PANEL DATA

The present study makes use of cross-sectional and longitudinal data from the research project "Child-rearing and family in the Netherlands". This national survey was undertaken to examine Dutch familial and marital life (Gerris et al, 1992; 1993). The focus is upon subjective characteristics of primary relationships as actualized in parent-child and husband-wife interactions. Properties of the broader social environment are measured both in more objective as well as subjective terms. The central aim is to relate characteristics of the environment to the internal functioning of the family.

To select the families, a multi-stage sampling method was followed. In a first stage, a sample was taken of all Dutch municipalities distinguished by regional zone and degree of urbanization. Four regional zones were distinguished: (1) the North, comprising the provinces of Groningen, Friesland and Drenthe; (2) the East comprising the provinces of Overijssel, Flevo-

land and Gelderland; (3) the South comprising the provinces of Zeeland, Noord-Brabant and Limburg and (4) the West, comprising the provinces of Utrecht, Zuid-Holland and Noord-Holland.

To measure the degree of urbanization, the classification of the Centraal Bureau van de Statistiek was employed (Gerris et al, 1992). Category "A" refers to rural municipalities with 20-50% of the economically active population employed in the agrarian sector. Category "B" represents the urbanized rural municipalities with a maximum of 20.000 inhabitants, in which more than half of its active population works in the manufacturing industry. Also included in this category are the resident municipalities of commuters where 30% of the economically active population is commuter and where 60% of the resident male population (of at least 14 years old) has been born elsewhere. Categories "C1" to "C4" include country towns, small or medium sized, with a population of 20.000 up to 100.000 inhabitants in the urban center, containing urban institutions and services. The final category "C5" comprises large towns with over 100.000 inhabitants in the urban center. For a more extended description of this classification see Gerris et al. (1992).

The number of inhabitants in the specific combination of regional zone and degree of urbanization was divided by the sample fraction, indicating the number of families to select. The sample fraction equals the population divided by one thousand; for the Netherlands this is 14.804. For practical reasons it was decided to approach a minimum of 10 families in each selected city.

Families were selected on the basis of their target child. Therefore, in a second stage, a random sample of children was taken from the municipal registers in the selected municipalities. The target children were selected in such a way that in each city as many boys as girls and as many children aged 9 to 12 as aged 13 to 16 were selected. The parents were automatically selected via the target child. In most cases, the parents are the biological father and mother of the target child, but also stepparents and one-parent families were recruited. Twice as many families than ultimately intended were chosen. If the address of the child and the parents was not the same, for example because the child lived in an institution, the family was replaced by the next one on the list. The same principle was used when a family refused to participate. In total, 1829 families were selected for an interview.

Initially the interview time was calculated to be one and a half hour but finally turned out to be two hours and fifteen minutes. That might be the reason why more families than expected refused to participate in the study or failed to finish the interview. This increased the cost of the field-work, and therefore it was decided to eliminate part of the interview and decrease the number of families to be included. In the end, 788 families participated in the project, which is a response rate of 43 % (1.829/788). This ratio is comparable to the response rate in other panel studies such as the Panel Study of Belgian Households (PSBH) but indicates the difficulty of attracting families to participate in a longitudinal research project.

Families who refused to participate were asked for their reasons. The reasons are presented in Table 3.1. As can be seen, the majority of the non-participants reported not to be interested or having no time.

Table 3.1
Reasons Why Families Refused to Participate In the Research Project in 1990 and 1995

Reasons to refuse	Number of families in 1990	Number of families in 1995
Family was not interested or no time	761	72
Family was approached three times without success	79	10
Address is unknown or not correct	43	6
Respondent was died, ill or on holiday	32	7
The child was too young or too old for being eligible	22	0
Linguistic problem	76	1
Invasion of privacy	0	7
Couple was divorced or problems in the family	0	8
Prior research was too long ago		
Otherwise	0	9
	28	23
Total	1.041	143

Source: Gerris et al., 1992; 1998

However, despite the reduction in the number of families, the sample was representative regarding regional zone and degree of urbanization. In the upper line of Table 3.2, the number (left column) and percentages (right column) of families that had to participate to guarantee a representative sample are presented. The bottom line shows the number and percentages of families that actually participated.

Table 3.2
Number and Percentage of Families that Should Have Participated
and that Actually Participated in the Research Project

	North		East		South		West		Total	
	N	%	N	%	N	%	N	%	N	%
A	25	3.2	21	2.7	17	2.1	27	3.4	90	11.4
	33	4.1	26	3.3	11	1.4	17	2.2	87	11.0
B	24	3.0	65	8.2	91	11.5	119	15.1	298	37.8
	15	1.9	64	8.1	101	12.8	132	16.8	312	39.6
C1-C4	27	3.4	44	5.6	55	7.0	86	10.9	212	26.9
	33	4.2	29	3.7	49	6.2	96	12.2	207	26.3
C5	9	1.2	30	3.8	32	4.0	118	15.0	188	23.9
	11	1.4	36	4.6	30	3.8	105	13.3	182	23.1
Total	84	10.7	160	20.3	194	24.6	350	44.4	788	100
	92	11.6	155	19.7	191	24.2	350	44.5	788	100

Source: Gerris et al., 1992

Of the 656 families who agreed in 1990 to participate in the second wave, 627 could be traced five years later. The other families had moved or were untraceable. 484 of the contacted families agreed to participate again (1356 of the 1881 contacted family members). This adds up to a response rate of 77% at the household level (Gerris et al., 1998). The reasons for refusal are

presented in Table 3.1. As in the first wave, the most frequently reported argument is "no time" or "not interested".

This sample still proved to be representative for regional zone but not for degree of urbanization. It appeared that primarily participants from the bigger cities refused to participate for the second time in the research project.

3.2. PROCEDURE

The interviewers responsible for the collection of the data were selected on the basis of their experience with administering surveys. Instruction meetings were organized to extensively inform the interviewers about the project in general and specifically about the interview and the questionnaires. The fieldwork started in August 1990 and lasted until January 1991. The selected families received a letter in which the purpose of the study was explained (Gerris et al., 1992). Within three days after receiving this letter, the interviewer contacted the families and made an appointment with both parents and the target child.

Various kinds of data were gathered during the interview. Demographic information on the parents as well as on the family as a whole was asked to both parents together by means of structured interviews. Background information such as spousal age, age of the target child, marital duration, educational level, employment status and whether or not one's parents had been divorced were included in the demographic questionnaire.

Other variables regarding attitudes, family characteristics and personal properties were measured with questionnaires consisting of 7-point Likert scale items. Most of the scales range from "not at all applicable" to "completely applicable".

Both husband and wife had to fill in separately a questionnaire about parenting including among others, parental satisfaction, parental stress and feelings of restriction imposed by the parental role. They also filled in a questionnaire about characteristics of the marital relationship. Items measuring marital satisfaction, negative communication as well as open communication were included.

Socio-cultural orientations on diverse aspects such as gender roles, autonomy and work orientations, were measured by means of 5-point

Likert scales ranging from 1 = "very important" to 5 = "unimportant" or 1 = "totally agree" to 5 = "totally disagree". The questionnaires on cultural orientation were left with the family. The interviewer explained how to fill out these questionnaires and requested that the family members completed them separately. After sending these questionnaires back frank, the target child received a reward of 25 Dutch Guilders.

In the second wave of the research project, some additional measures were included. Of interest for the purpose of this study is the Personal Attributes Questionnaire measuring femininity and masculinity as well as a questionnaire on task allocation in the household.

Each of the aforementioned instruments and measures will be described in depth in the subsequent studies addressed in the following chapters.

3.3. SELECTION AND CHARACTERISTICS OF THE MARRIED SUB-SAMPLE

In order to examine married couples, a sub-sample from the Dutch dataset had to be selected. Because the original dataset also includes cohabitating and one-parent households, we removed these observations from the sample. The resulting sample of married couples, however, still included first and higher order marriages. Although their relational satisfaction may not be that different, previous studies have found that the marital processes and characteristics in first and higher order marriages may differ (Booth & Edwards, 1992). Therefore, only first marriages were selected. Because no question was directed to the issue of first and higher order marriages, we reconstructed the first married couples by using indicators about the number of times the child experienced divorce or death of (one of) the parents.

In the next step, ethnic mixed marriages were removed from the sample.[1] Hondius (2003) has shown that couples consisting of partners with a different ethnic background develop specific strategies to cope with their interpersonal differences and with the reactions from their environ-

---

[1] Because 96% of the fathers and 95% of the mothers reported to be born in the Netherlands, this selection criterion did not result in a strong reduction of the sample size.

ment (Hondius, 2003). Moreover, Dutch statistics demonstrate that marriages between autochthonous and allochthonous partners have the highest divorce risks (Janssen, 2003). In comparison to non-mixed marriages, couples in mixed marriages are more confronted with specific problems relating to differences in preferences and behavior and may therefore confound the results with respect to the factors that differentiate between more and less satisfied spouses.

In sum, first marriages consisting of husbands and wives with a Dutch nationality were eligible for the study. The selection procedure resulted in a research group of 646 couples in 1990 and 386 couples in 1995.

In 1990, the average age in our sample is 42.48 (SD = 4.88) years for husbands and 40.06 (SD = 4.17) for wives. About 90% of the husbands and wives included in our study were born between 1940 and 1955, roughly representing the baby boom generation. At the onset of the study, couples were married on average for 17 years (SD = 3.37). The range in marital duration varies from 1 year to 31 years, with about 90% married between 11 and 22 years. About 85% of the respondents report that they did not experience a parental divorce during their childhood. Couples have on average 2.48 children (SD = 1.04). In one out of two families the target child included in the study is the oldest child whereas in another 37% of the families the child is second in rank. The mean age of the target child is 12.8 years (SD = 2.20). As can be seen in Table 3.3, about a quarter of the male population enjoyed higher vocational or university education, while for women this figure was approximately one out of eight. In comparison to men, women are situated more often in the lower and middle-low educational levels. This gender pattern recurs when considering occupational status. About 16% of women are employed in unskilled jobs whereas this only holds for 4% of men. The reverse is true when considering higher occupations. It appears that 9% of the employed women occupy a higher profession; for men this is 18%. In general, the labor market participation of Dutch fathers and mothers show large discrepancies. 95% of the husbands are employed whereas less than half of the wives have a job. The discrepancy also becomes clear when taking into account the number of hours officially worked. On average, husbands work 38.4 (SD = 3.43) hours while wives work 19.5 (SD = 10.2). Hence, the couples in our study

primarily represent single earner- and main-earner households. Their average monthly family income is between 1250 and 1625 euro.

Table 3.3
Distribution of Men and Women According to Educational and Occupational Level (1990) in the Research Project

	Women	Men
Educational level		
Lower	13.2	6.0
Middle-low	51.6	46.3
Middle-high	24.2	24.8
Higher	12.0	22.9
Occupational status		
Unskilled job	15.7	3.8
Skilled job	10.1	17.8
Lower employee	50.8	24.4
Small tradesman	5.6	11.6
Moderate employee	8.6	24.1
Higher professions	9.3	18.3

3.4. VALIDATION OF THE SAMPLE AND MISSING CASES

In survey research in general and in panel research in particular, non-response is an important obstacle. The difficulty lies in the fact that this non-response might not be ad random but characterized by selectivity, and hence threatening the validity of a study (Billiet, 1993; Scott, 1995, Wothke, 2000). Selectivity signifies that respondents with particular characteristics are more likely to refuse their participation than those characterized by another profile. Studies show that individuals who refuse to participate have a typical profile: male, low education and unmarried (Taris, 2000).

A sample characterized by strong bias is often more homogeneous with respect to the variables of interest, because the latter might be related

with the non-response. Therefore, selective non-response can strongly distort the research results, and thus strongly reduce possibility of generalization to the entire population. Hence, one should always presuppose that the non-response is systematically associated with the research subject of interest (Billiet, 1993).

Attention to the non-response pattern is particularly relevant when using a longitudinal design. The advantages of longitudinal research and panel research in specific are offset by the risk of loosing panel members throughout the study (Dennis & Li, 2003). The possibility that lost panel members differ from all-waves-participants with respect to their attitudes and behavior towards the issue under study, cast doubt on the validity of the data (Dennis & Li, 2003). Therefore, this section aims to evaluate the non-response of our sample. In the first paragraph, characteristics of the participants of the first measurement wave are compared with characteristics of the Dutch population, i.e. the external validation of our sample. The second paragraph addresses the internal validation of our sample by evaluating the characteristics of the panel members who only participated at the first measurement wave with those of the participants who remained in the study.

3.4.1. External Validation

As mentioned before, 43% of the selected households participated at the first measurement wave, indicating a high non-response by international standards. Latten and De Graaf (1997), however, allege that the public willingness to participate in survey studies is lower in the Netherlands than in most other countries. Therefore, a response of 43% does not diverge greatly from typical rates in other Dutch studies conducted by the Netherlands Central Bureau of Statistics. For example in the Fertility and Family Survey (FFS) conducted in 1993, the response rate at the household level was 48,5%. A closer examination of the non-response in the FFS revealed that men participated less than women, younger women refused more often than older women and older men refused more often than younger men.

Because of the high refusal rate in the Netherlands, it is even more important to evaluate in depth the bias introduced by non-response in the sample used in this study. Unfortunately, the reasons for refusal notwith-

standing, no registration was made of the demographic information of the refusers. Therefore an indirect method is required to gain insight in the specificity of the sample. To evaluate the non-response, characteristics of the respondents are compared with the distribution of these characteristics in the total population.

As already indicated, the representativeness of the sample in 1990 is ensured with respect to the degree of urbanization and regional zone. In 1995 the sample was no longer significant with respect to the degree of urbanization. Among other things, this may relate to the higher chance of divorce in bigger cities (Weeda, 1981).

Besides regional zone and urbanization, we are interested in background characteristics such as educational level and employment characteristics. This comparison, however, poses serious difficulties as the sample of married couples was obtained by means of a sample drawn on the basis of children's age. Hence, information is required of a very delineated subgroup of the population, i.e. first married couples with at least one child aged between 9 and 16 years old. These data are not available and therefore, as a second best solution, we use less specific data to situate our respondents according to the following characteristics: (1) age at first marriage, (2) employment status and (3) educational level of the spouses.

On average, the couples in our sample got married in 1973. In the Netherlands the mean age at first marriage in the period 1970-1975 was about 25 years for men and 23 years for women (see Chapter 1). In our sample, mean ages at marriage are 25.11 for men and 22.68 for women, which is fairly comparable with the mean age in the population.

With respect to spousal employment status, it appears that 47% of the mothers and 95% of the fathers in our sample have a paid job. To evaluate these proportions, the figures reported by Niphuis-Nell & de Beer (1997) about the labor market participation of fathers and mothers may be interesting.[2] Considering the gross labor market participation[3] of women with children in 1990, the figures show that 49% of mothers with a youngest child of 6 years or older were employed. With respect to fathers' employ-

---

[2] The figures presented by Niphuis-Nell & De Beer are computed on the basis of data of the Central Bureau of Statistics.
[3] The gross labor market participation is defined as working for at least one hour or seeking employment for at least one hour

ment, Niphuis-Nell & de Beer (1997) only discuss figures for 1995. About 90% of fathers with a youngest child between 6 and 17 years old participate in the labor market. Comparing these proportions with the figures in our sample, we can assume that the labor market participation of the couples included in our study approximates the employment situation in the population.

Of the employed spouses in our study, 5% of the husbands work less than 35 hours per week whereas this is true for 88% of the wives. Considering the figures presented by Niphuis-Nell & de Beer (1997) regarding couples with children between 6 and 17 years old, it appears that 6% of the fathers are employed for less than 35 hours whereas for mothers this is about 91%. Therefore, we can reasonably assume that the distribution of part-time versus full-time employment in our sample of employed parents reflects that of the population.

To evaluate the educational level of the respondents with that of the population, we divided the different educational levels into three categories representing lower, secondary, and higher education. The distribution is presented in Table 3.4. Since no specific cohort figures of married people with children are available, we compared our results with the distribution of educational levels within the active population aged 45-64 in 1996 as well as in 2002 (CBS, 2003). The population aged 45-64 in 1996 was born between 1932 and 1951; in 2002 they refer to the group born between 1938 and 1957. Hence, this population approximately covers the cohorts included in our study. However, in contrast to the group selected in our study, the general population also includes unmarried people or married individuals without children as well as people who do not have the Dutch nationality. Therefore, the comparison is only a rough indication and must be interpreted cautiously. Table 3.4 shows that women who are lower or higher educated appear to be underrepresented in our study in favor of women who received secondary education. In contrast, lower educated men are somewhat over represented at the expense of higher educated men. Husbands who received secondary education, however, appear to be a good reflection of the proportion in the population.

Table 3.4
Educational Level of the Active Population Aged 45-64 Years in the
Netherlands, According to Sex (%)

	Sample		Active population in 1996		Active population in 2002	
	Women	Men	Women	Men	Women	Men
Lower education	42.92	38.63	56.8	36.9	49.4	34.0
Secondary education	46.07	38.47	28.7	38.9	32.1	37.9
Higher education	12.0	22.9	14.4	23.9	18.4	27.9

Source: CBS, 2003

3.4.2. Internal Validation

To assess the selectivity of the panel attrition in our study, we compare the structural, cultural and marital characteristics of the panel members who remained in the sample during the period under study – this is between 1990 and 1995 – with those who only participated at the first measurement wave. Three separate logistic regressions are conducted each testing if particular individual properties relevant to the phenomenon under study, could predict a couple's participation only at the first measurement wave. The first model assesses the degree to which *one-time* participation depends on socio-demographic characteristics of the spouses. The second model examines whether both wave participants hold different cultural orienta-tions compared to respondents who only participated once and whether they have different network support and parental experiences. Finally, the third model tests whether differences exist with respect to spouses' marital characteristics in terms of marital satisfaction and communication.

Regarding the socio-demographic characteristics, the first logistic regression includes the following variables of both husbands and wives: age, educational level, occupational status, net family income, income satis-faction, year of marriage, parental divorce and whether one works or not.

The $\chi^2(14) = 21.4724$ with $p = .0901$ is not significant at a .05 level. However, three variables appeared to be significant: husbands ($p = .003$) and wives' income dissatisfaction ($p = .02$), and the educational level of the wife ($p = .07$). The results demonstrate that the respondents participating twice consist of husbands who are more satisfied with the family income, of wives who are less satisfied with this income and of higher educated wives.

With regard to cultural orientations, we included spousal attitudes towards (a) parenting as stressful, (b) parenting as personal restricting, (c) parenting as a rewarding task, (d) togetherness and (e) freedom in the intimate relationship, (f) sex roles, and (g) traditional family values. Moreover, the degree to which husbands and wives perceive to have access to network support is included in the model as well. This model is significant with $\chi^2(16) = 26.9396$ with $p = .0422$. Closer examination shows that husbands' attitudes towards sex roles ($p = .01$) and traditional family values ($p = .01$) differ significantly between the couples who remained in the study and those who dropped out. The latter endorse less traditional family values but are more traditionally orientated towards sex roles.

Regarding marital characteristics, it is of particular importance to assess whether the couples that dropped out of the study were significantly less satisfied with their marriage or display different communication behavior than those who remained in the study. The factors that tend to lower marital satisfaction may also tend to drive persons out of the married population. Therefore, it might be expected that husbands and wives who separated between Time 1 and Time 2 partially account for the dropout of the panel. To examine this assumption, we tested a model including both spouses' marital satisfaction as well as their perceptions of marital communication. This model, however, was not significant with $\chi^2(6) = 3.3468$ and $p = .7642$, indicating that the panel dropout is not due to the satisfaction or communication profile of the respondents.

Note that we established the nature of the panel attrition by testing three separate multivariate models. However, to fully capture the impact of our panel attrition on the various research problems and models dealt with in this study, we will reconduct the above analysis for each particular research problem. Specifically, this means that for each research problem addressed in the following chapters, a logistic regression, including all the

variables of interest to that particular part of the study, will be conducted and evaluated.

3.4.3. Item Non-Response

Apart from the refusals of respondents to participate in a survey, observations can also be lost because of item non-response. The latter is a pervasive problem in empirical research because analyses rest upon a much smaller sample size than the initial number of observations selected. Therefore the question arises whether this limited number of observations is still representative. In this paragraph, we elaborate on the procedure that was followed to save cases with missing values that would be lost in common listwise deletion approaches as used in regression analysis.

Missingness can be either related or unrelated to the dependent or independent variables in the analysis. Data are considered to be *missing completely at random* (MCAR) if the missing value on a variable is unrelated to the dependent variable or to any other variable in the study (Allison, 2001; Jamshidian, 2004). A weaker variant of this assumption is *missing at random* (MAR), referring to the idea that a missing value on a variable depends on the values of variables without missings (observed characteristics). When the missingness on a variable is additionally related to the missing values on other variables, this is referred to as *missing not at random* (MNAR).

In statistical packages such as SAS, the default strategy to deal with item non-response is listwise deletion, excluding the observations with one or more missings on the variables of interest (Jamshidian, 2004). In spite of the attractive simplicity of this strategy, it is not obvious that respondents who have missing values on a particular variable are indifferent compared to those with observed variables. Listwise deletion may result in biased estimates under the conditions of MAR or MNAR (Jamshidian, 2004). Instead of ignoring observations with item non-response, one can also recycle the observation by imputing the missing value. The simplest technique is replacing the missing value by the mean of cases that have observed data on the variable of interest. However, this method may lead to biased estimates of variances (Allison, 2001; Wothke, 2000). The most advanced techniques of imputation are (1) the method of full information maximum Likelihood (FIML) using all the information of the observed

data to estimate missing values, and (2) multiple imputation methods, replacing missing values on a given variable with a set of simulated values, generating different complete datasets. Each of these datasets is analyzed and the results are considered together to draw inferences (Allison, 2001; Jamshidian, 2004). It is clear that this strategy is time consuming and cumbersome and therefore not employed in this study.

To save cases with missing values, a single imputation method was used for variables with maximum 10 % missings. The imputation was conducted as follows. For a given variable that requires imputation of missing values (from here on referred to as the imputation variable), it is assumed that whether or not an observation has a missing value depends on some observed characteristics of the observation (MAR). Therefore, the first step determines predictor variables for the missingness on the imputation variable. These predictor variables are denoted matching variables. The general idea is to impute missing values on the imputation variable with values from other observations that are as similar as possible to the imputed observation with respect to their scores on the matching variables. More specifically, imputation is performed as follows.[4]

Suppose there are p matching variables, denoted $X_1, \dots , X_p$ and that the imputation variable is denoted Y.

1. Standardize the matching variables. The standardized matching variables are denoted $Z_1, \dots , Z_p$.
2. For every observation i with a non missing value on the imputation variable, compute:

$$SS_i = \sum_{j=1}^{p} (Z_{j,i} - Z_{j,m})^2$$

    and m denotes the observation that requires imputation on the imputation variable
3. Detect the n observations for which SS_i is minimal.
4. If $n=1$ then $Y_m=Y_i$, i.e. the observation m which requires imputation on the imputation variable receives the value of the single observation with minimal SS_i.
5. If $n>1$, then a variance ratio is calculated as follows:

---

[4] I would like to thank my former colleague Dr. Jerry Welkenhuysen-Gybels for developing and elaborating the imputation procedure.

$$V = \frac{S_{Y,n}^2}{S_{Y,total}^2}$$

where $S_{Y,n}^2$ refers to the variance of the imputation variable Y for the n observations with minimal SS_i and $S_{Y,total}^2$ denotes the variance of the imputation variable for the entire sample.

If $V < 1$ then one of the n observations with minimal SS_i is randomly selected to serve as the donor for imputing the Y value of the observation m. If $V > 1$, no imputation is performed. The latter condition is used to ensure that the variability of the Y variable among the n possible donor observations is not too high as compared to the variability of the entire sample.

Note that no imputation occurs if (1) observation m also has a missing value for at least one of the matching variables; (2) there are no donor observations without missing values on the matching variables and (3) the variance ratio V is larger than 1. The imputation was only conducted for the variables indicating individuals' score on a particular scale and not for the separate items of each scale

In comparison to imputation based on observed covariates of the imputation variable - instead of the covariates of the missingness on the imputation variable - the advantage of this imputation strategy is that the associations between the observed variables in the data set are not further strengthened by the imputation procedure.[5]

---

[5] The main conclusions drawn from the regression analyses before and after imputing the missing variables were comparable, indicating that no strong bias was caused by the imputation method.