SSOAR

Open Access Repository
www.ssoar.info

Multiple imputation in an international database of social science surveys

Longford, Nicholas T.

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:
Longford, N. T. (2000). Multiple imputation in an international database of social science surveys. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 46, 72-95. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-199340

Nutzungsbedingungen:
Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:
This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.
By using this particular document, you accept the above-stated conditions of use.

gesis Leibniz-Institut für Sozialwissenschaften

Mitglied der Leibniz-Gemeinschaft

Multiple imputation in an international database of social science surveys

by Nicholas T. Longford [1]

Zusammenfassung

In diesem Beitrag wird das Verfahren der multiplen Imputation anhand von Datensätzen aus dem International Social Survey Programme diskutiert. Da es in den meisten Variablen fehlende Werte gibt, die in vielen unterschiedlichen Kombinationen vorkommen, werden die Imputations in mehreren Schritten durchgeführt. Als erstes werden die Angaben bei den sozio-demografischen Merkmalen ersetzt, da es hier in der Regel nur relativ wenige fehlende Werte gibt. Bei Blöcken von Items werden nur deren Summenwerte geschätzt, was die Aufgabe der Imputation vereinfacht, ohne daß dabei auf wichtige Informationen verzichtet wird. Ein weiterer Vorteil dieser Vorgehensweise ist, daß die Anzahl der einzusetzenden Werte reduziert wird.

Abstract

This paper describes an implementation of the method of multiple imputation in the database of surveys in the International Social Science Programme. Since missing values occur for most variables, with a wide range of patterns, the imputations are carried out in stages, starting with background variables which, in general, have fewer missing values. For blocks of questionnaire items only their total scores are imputed, making the imputation task manageable without substantial loss of utility of the database, and reducing the size of the data files added to the database by the imputation procedure.

1 Introduction

Population surveys are an important source of information for social scientists. Such surveys are designed so as to be representative of the studied population and to yield infer-

1 Author's address: ***Nicholas T. Longford***, Department of Medical Statistics, De Montfort University, The Gateway, Leicester LE1 9BH, UK. Email: NTL@dmu.ac.uk. Research for this paper was conducted while the author was a visitor at the Zentralarchiv für Empirische Sozialforschung, Cologne University, Germany, supported by a grant from the Training and Mobility of Researchers Programme of the European Union.

ences with sufficient precision. These two general goals are often undermined by nonresponse, because it reduces the effective sample size. Also, the subsample of responders may be a poor representation of the studied population even when the original sample is a good one. As an example, suppose a simple random sample of subjects from a population has only 2% missing values on the variable of principal interest. A substantial bias in estimation of the mean can still result if the non-respondents tend to have the highest values. In more complex analyses, the rate on nonresponse is not a good indicator of the loss of information, because subjects' values exert uneven influences on the estimates. Nonresponse can be quite extensive, especially in surveys with many questionnaire items, some of which probe issues which the respondents may not be comfortable with or confident to discuss. Many analysts of such surveys rely on standard statistical software, such as SPSS and Minitab, or on specialist software packages, such as LISREL and EQS, which have limited provisions for incomplete data. So, most analyses are restricted to subjects with complete records (also known as *listwise deletion*), reducing the sample sizes substantially when a large number of variables is considered. In such a setting, a lot of information contained in the incomplete records is made no use of. By pairwise deletion, all recorded data are used, but inconsistencies, such as inadmissible matrices of crossproducts, can arise (***Little/Rubin*** 1987).

The method of multiple imputation (***Rubin*** 1987 and 1996, and ***Schafer*** 1996) was designed for large-scale databases in which missing values appear in a variety of patterns (configurations). The context considered is that of a database constructor and a number of secondary users. The constructor is privy to the details of the data collection procedures, including the causes of missing data, their coding, and the like. For the secondary users, the published materials accompanying the public-use version of the database are the principal source of information about the database. Of course, these materials may not contain all the details of how the survey was conducted, what kinds of contingencies were encountered and how they were dealt with.

Following the conduct of a typical survey, the constructor compiles the database from the collected data items, prepares a comprehensive documentation and background information, and provides these to (secondary) users. A typical user focuses on statistical analyses to address substantive research issues. For users not equipped to deal with the problem efficiently, missing data are a profound inconvenience. But even if they were equipped, a lot of effort in dealing with missing values would unnecessarily be duplicated. A more practical and economic solution is for the constructor to deal with the problem, in such a way that the secondary users would require no software tools other than those they would have employed had the data been complete. The users could then focus in the analyses on the substance of their research, without the distraction of the technical (statistical) issues associated with the missing data.

Multiple imputation has been designed for just such a setting. It has been successfully implemented in a number of government surveys, in the U.S.A. in particular, see ***Rubin/ Schenker*** (1991). The method has made much less impact in social science surveys; primary or secondary analysts of such surveys rarely address the issue of missing data in any integral manner. Multiple imputation, or any other approach to handling missing values, is difficult to implement because of varied patterns of missingness involving many background and outcome variables, most of them with multinomial distribution, such as in Likert scales.

The International Social Science Programme (ISSP) is a sequence of annual surveys in a number of countries. It began in 1985, and in 1999 it had more than thirty members (countries). In this paper, we describe an implementation of the multiple imputation in the 1995 surveys which were conducted in 23 countries. The surveys used similar sampling designs and translations of the questionnaire originally written in English. The theme of the questionnaire was 'National Identity'. ISSP is analysed by many secondary users, especially researchers in comparative (cross-national) social studies. They apply a range of statistical methods, using implementations in standard software (SPSS, Minitab, and the like), or in specialist packages, such as LISREL and EQS.

Longford (2000) applied multiple imputation, within the context of a sensitivity analysis, to a section of the survey containing six items on attitudes to immigration, each scored on a Likert scale 1–5. The imputations were based on a model that conditions on the observed responses in the section and aggregates similar score patterns to avoid having to deal with sparse tables, while compromising as little as possible the ideal of conditioning on all available information (***Rubin*** 1987).

Our goal is to design a multiple imputation procedure which could be implemented in the past as well as future surveys in ISSP with only minor adaptations and without having to deal with many special cases. For this purpose, a compromise has to be struck between the ideal of including all available variables in the model for missing data and the practicalities of handling large datasets of (mostly) categorical variables. In the models for missing data, we consider only a short-list of *prima facie* correlated variables.

The next section gives details of the ISSP database. Section 3 introduces the standard terminology for missing data. Section 4 describes the multiple imputation procedure. The concluding section discusses the constructor's and analyst's perspectives on multiple imputation.

2 The ISSP database

The data from the 1995 ISSP surveys is stored in a single file of 214 variables defined for the 28 456 subjects from 23 countries. Eighteen countries are from Europe, nine of them former communist countries (or their parts), four are developed countries in America (U.S.A. and Canada), Asia (Japan) and Oceania (New Zealand), and the remaining country is the Philippines. In three member countries, Australia, Northern Ireland, and Israel, the 1995 survey was not conducted. For the purposes of the surveys, West Germany and the former German Democratic Republic, as well as Northern Ireland and Great Britain, are treated as separate countries. The within-country sample sizes range from 612 (East Germany) and 994 (Ireland) to 1598 (Poland) and 2089 (the Netherlands).

Some of the variables in the database are derived from others. For instance, *Earnings* and *Family income* are recorded both in the currency of the country (rounded or grouped) and as an ordinal categorical variable. Several variables (*Region of the country* and *Party affiliation*), refer to a specific country, but the variables are defined for the entire sample, with values `Not applicable` for subjects from all the other countries. Thus, the effective number of variables is only about 80. In some cases, the prevalence of missing values is so high that imputation for them would not be very useful. Also, in some countries a few variables have not been recorded at all.

We distinguish between administrative, background, and response (opinion) variables. Administrative variables are the study number, respondent's identification, and indicator of the country; these contain no missing values. Background variables record the subject's sex, age, marital status, education, employment status, income, religious denomination, party affiliation, trade union membership, type of the community, household size, and region of the country. The response part of the questionnaire can be divided into eight sections of items inquiring about related matters; the items in most sections have a common lead-in passage. The responses within these sections tend to be correlated more highly than between the sections or with the background variables. Further, some items inquire about the languages spoken at home, ethnicity, and citizenship, which can be regarded as being on the borderline between background and response items.

Another set of questions inquires about language proficiency and the languages spoken by the subject at home. The items ask for the first, second, and third languages. The response options contain a list of languages, but the coding does not distinguish clearly among failures to respond and `Not applicable` (e. g., when only one language is spoken at home). For instance, in the survey for West Germany, there are only three and eight missing values for the ‘first language’ items, but the code for missing item and `Not applicable` dominates for the second-mention items (95% for the language spoken at home and 58.5%

for the language able to speak). Since the vast majority of first mentions are `German`, imputing `German` for the eleven missing values would seem appropriate.

3 Missing data

The list of values of a set of variables for a subject is called a *record*. In most datafiles a record is a row of the file, or its subset (a segment). The elements of the record, the individual values, are called *items*. Each item may be observed or missing (nonresponse). A record is said to be *complete* if each of its items is observed (no nonresponse), *incomplete* if at least one item is missing, and *empty* if all items are missing. Similarly, a variable is said to be observed completely, not completely, or not at all, if its values are observed, respectively, on every subject, some subjects, or no subjects in the sample.

For each item, we define the indicator of missingness. It is equal to 1 if the item is missing, and to 0 if it is recorded. The sequence of indicators corresponding to a record is called the *pattern of missingness*. For instance, the pattern of missingness for a subject on a set of four variables may be 1001 (first and fourth variables not recorded). In principle, any one of the $2^4 = 16$ patterns of missingness can occur. Similarly, we define the *score pattern* as the sequence of the values in a record. Clearly, it is meaningful to consider score patterns only for categorical variables, otherwise most score patterns are unique. The missingness pattern is a special case of a score pattern, with each variable being dichotomous. The number of possible score patterns can be quite large, especially when some of the variables have many categories or when the pattern is considered for many variables. In general, if H variables have $c_1, c_2, \ldots, c_H$ categories, there may be up to $C = c_1 \times c_2 \times \ldots \times c_H$ score patterns. However, the number of score patterns is often much smaller, not only because the sample size is smaller than C, or due to chance, but also because some combinations of categories are not feasible. For instance, for variables *Age* and *Education*, the score patterns corresponding to young (say, 16–20 years of age) and completed university education are not feasible.

The patterns of missingness are partially ordered. Pattern A is said to have more missingness than pattern B if any variable not observed in B is not observed in A either. Thus, pattern 1100 has more missingness than 0100, less missingness than 1101, but neither more nor less missingness than 0010, 1001, or 1011, even though it has, respectively, more, the same number, or fewer missing items.

Although we usually fail to collect all the items of data that the plan envisages, it is still instructive to consider this hypothetical dataset, called the *complete data*, because it usually has an easy-to-manage ‘rectangular’ structure of subjects and variables. It is practical to plan the analysis for the complete dataset; also, the sample size calculations are usually conducted for the complete dataset.

Figure 1: Analysis of complete records. An illustration.

Observed data

Id.	A	B	C	D	E
1	•	•	•	•	•
2	•	•	•	•	?
3	•	•	•	•	•
4	•	•	•	?	•
5	•	•	•	•	•
6	•	•	•	?	?
7	•	•	•	•	•
8	?	?	?	?	?

→

Analysed data

Id.	A	B	C	D	E
1	•	•	•	•	•
3	•	•	•	•	•
5	•	•	•	•	•
7	•	•	•	•	•

Notes: The record identification is given in the left-hand column (1–8). Available values are marked by • and missing values by ? .

The observed data are the complete data with the nonresponse superimposed on it. The procedure that we would have applied - had the data been complete - is called the *complete-data procedure*. Because of the missing values, its direct application to the observed data is usually not possible. A compromise frequently made is to apply the complete-data procedure on the subjects with complete records. This is referred to as the *analysis of complete records*, or listwise deletion. Although it is only as difficult to apply as the complete-data procedure would have been, it may not have some of the desirable properties of the complete-data analysis. First, transparently, the sample of responders (the subjects with complete records) is smaller than the original sample, so we can expect less precision in estimating any (population) quantity of interest. Second, much less transparently, the non-responders may not be as representative of the studied population as the entire sample is. The analysis of complete records is illustrated in Figure 1 on a dataset of subjects 1–8 and variables A–E. The four incomplete records are discarded and the analysis is based on only four records, even though only nine items out of 40 have missing values. Moreover, the (observed) values of A, B, and C may provide some information about the missing values for D and E.

All the available data are made use of in procedures based on *pairwise deletion*. A given total of crossproducts is calculated from subjects for whom both constituent variables are recorded. Such procedures are deficient as they fail to make use of the available information about the missing values. Also, anomalies such as non-positive definite matrices of crossproducts can arise (***Little/Rubin*** 1987).

Selection of the subjects into the sample is a random (sampling) process, which eliminates most of the members of the population from being included in the sample. Similarly, nonresponse can be regarded as another random process. The main qualitative difference between these two processes is that, at least in the ideal setting, we exercise probabilistic control over sampling (e. g., by applying simple random sampling without replacement). On the other hand, for a given set of data collection arrangements, nonresponse (missingness) is entirely outside our control. Furthermore, the data provide no information about the nonresponse process. So, assuming that the missingness process is simple random amounts to totally unwarranted optimism. At the same time, assuming that the nonresponders are vastly different from the responders, in an unknown way, is not very constructive.

For simplicity, we consider first the setting in which only one variable is subject to nonresponse. If the process of missingness were simple random, assigning each subject the same probability of nonresponse and one nonresponse not affecting the occurrence of another, the respondents would be a simple random sample from the studied population. Then the analysis of complete records would be as valid as if the procedure were applied to the dataset we intended to collect, free of any nonresponse. The only difference would be the reduced sample size.

Missing values are unlikely to arise according to a simple random process especially when we have no means of promoting it. A much less restrictive assumption is that the process of missingness is simple random within each group (stratum) defined by the values of completely observed variables. Such a process is called *stratified simple random*. With it, there are no systematic differences between responders and non-responders within either stratum. Obviously, the finer the stratification (the larger the number of strata), the less restrictive the assumption of stratified simple random process, and the more likely that it is appropriate.

In the literature on missing data (***Little/Rubin*** 1987 and ***Rubin*** 1987), the process with simple random sampling is referred to as *missing completely at random* (MCAR), the process with stratified simple random sampling as *missing at random* (MAR), and the complement of MAR as *missing not at random* (MNAR). MCAR is a special case of MAR, with the trivial stratification. MAR and MNAR processes should always be qualified by the stratification applied. The variables used in the stratification are called *conditioning* variables, and MAR is said to be conditioned on them. Although stratification is formally defined only for categorical variables, it can be extended to continuous variables by considering their coarsened (grouped) versions, or by a reference to a regression model. An important result about MAR is that if the complete-data procedure is valid (has little bias and the uncertainty about the estimates is also assessed without bias), then so is the corresponding analysis of complete records, so long as all the variables used in the procedure

have also been used for conditioning in MAR. But every analysis of data with missing values should carefully consider the possibility of MNAR, a ubiquitous threat to the validity of its conclusions. See ***Longford*** (2000) for an example.

The data at hand contain no information about the process of missingness, so MCAR can be assumed only when missingness is transparently unrelated to the outcome. MNAR may be acute in clinical trials when some subjects withdraw (drop out) because they anticipate poor effect of the assigned treatment.

3.1 Data imputation

An alternative to the analysis of complete records is to generate a replacement or *impute* a value for each missing item. The obvious appeal of this approach is that the rectangular structure of the complete data and the original sample size are restored, and the complete-data analysis can be applied without any alteration. One way of generating the imputed values is by assuming a MAR process. For simplicity of the description, suppose nonresponse occurs only for one variable, the outcome y, and the background variables $\boldsymbol{x}$ are observed completely. Then a model relating y to $\boldsymbol{x}$ is fitted to the complete records, and the replacements are defined as the values fitted to the nonresponders' values of $\boldsymbol{x}$. For instance, assuming that it is appropriate, an ordinary regression of the observed values of y on the corresponding values of the regressors $\boldsymbol{x}$ yields a vector of regression estimates $\boldsymbol{b}$; the fitted values $\boldsymbol{x}_m\boldsymbol{b}$ would be imputed for each missing value of y which corresponds to the regressors $\boldsymbol{x}_m$. In such an imputation, the size of the residual variation plays no role. Even if we choose the model correctly, and fit it efficiently, the replacements are ‘correct’ only on average; they do not display one important feature of the responders, namely *dispersion* around the model relating y to $\boldsymbol{x}$.

Whichever way we impute for the missing values, the resulting *completed dataset* will appear to contain more information than the (observed) incomplete dataset, and so the uncertainty indicated by the standard errors or confidence intervals will be understated.

Figure 2 illustrates imputation on the same fictitious dataset as used in Figure 1. A replacement is defined for each missing value, and after imputation the analysis is conducted on the completed dataset of eight subjects.

4 Multiple imputation

Rubin (1987 and 1996) designed the method of multiple imputation, originally for the setting of a single data constructor or archivist and a number of independently acting secondary (groups of) analysts. The analysts rely on standard statistical software or specialist packages that have no comprehensive provisions for handling missing values. The goals

Figure 2: Analysis with values imputed for missing items. An illustration.

	Completed dataset							**Replacements**				
Id.	A	B	C	D	E		Id.	A	B	C	D	E
1	•	•	•	•	•		1					
2	•	•	•	•	i1		2					i1
3	•	•	•	•	•		3					
4	•	•	•	i2	•	←	4				i2	
5	•	•	•	•	•		5					
6	•	•	•	i3	i4		6				i3	i4
7	•	•	•	•	•		7					
8	•	i5	i6	i7	i8		8		i5	i6	i7	i8

Notes: The record identification is given in the left-hand column (1–8). Available values are marked by • and the imputed values by i1, . . ., i8.

set out for multiple imputation are to enable the secondary analysts to deal with the problem of missing data, requiring

- no software, other than what would be used for complete-data analysis;
- no expertise in missing data issues in general or in relation to the studied database.

In multiple imputation, a set of replacements is generated for each missing value. The complete-data analysis applied to the dataset completed with the first set of replacements yields one set of results, the analysis of the dataset completed with the second replacements another set, and so on. Figure 3 gives an illustration for a dataset of seven records and four variables; only one variable has missing values (for three records). For each missing item, a set of five replacements, called *plausible values*, is generated. In other applications, the number of plausible values may be different.

The (five sets of) results are then averaged, with an appropriate inflation for the standard errors; see below. Generating the plausible values is usually a complex task comprising model formulation, model fitting and generating the plausible values from the model fit; see Section 4.1. In a typical setting, the constructors are best suited for this task, because they have easier access to the required expertise and information about the processes of missingness, and by taking care of missing data any duplication of effort among the secondary analysts is avoided. Also, by enabling analyses of higher quality, the constructor provides a more valuable product, without imposing greater demands on the users' technical expertise or software equipment.

Figure 3: Multiple imputation. An illustration.

Observed data						**Plausible (imputed) values**						
Id.	A	B	C	D		Id.	Var.	M1	M2	M3	M4	M5
1	•	•	•	•								
2	•	•	•	?	←	2	D	i11	i12	i13	i14	i15
3	•	•	•	•								
4	•	•	•	?	←	4	D	i21	i22	i23	i24	i25
5	•	•	•	•								
6	•	•	•	?	←	6	D	i31	i32	i33	i34	i35
7	•	•	•	•								

Datasets completed by imputations

	Completed dataset 1				Completed dataset 2				Completed dataset 3			
Id.	A	B	C	D	A	B	C	D	A	B	C	D
1	•	•	•	•	•	•	•	•	•	•	•	•
2	•	•	•	i11	•	•	•	i12	•	•	•	i13
3	•	•	•	•	•	•	•	•	•	•	•	•
4	•	•	•	i21	•	•	•	i22	•	•	•	i23
5	•	•	•	•	•	•	•	•	•	•	•	•
6	•	•	•	i31	•	•	•	i32	•	•	•	i33
7	•	•	•	•	•	•	•	•	•	•	•	•

	Completed dataset 4				Completed dataset 5			
1	•	•	•	•	•	•	•	•
2	•	•	•	i14	•	•	•	i15
3	•	•	•	•	•	•	•	•
4	•	•	•	i24	•	•	•	i25
5	•	•	•	•	•	•	•	•
6	•	•	•	i34	•	•	•	i35
7	•	•	•	•	•	•	•	•

A secondary analyst, instead of applying a complete-data procedure once, has to impute each set of plausible values (one set at a time), and apply the procedure to the dataset thus completed. So, the complete-data procedure has to be applied several times, requiring more computer time, but little additional time or effort of the analyst or programmer.

Let b_i , $i = 1, \ldots, K$, be the estimates from the K analyses ($K = 5$ in earlier discussion), and s_i be the corresponding (complete-data) standard errors. Then the estimator that applies to the incomplete data is

$$b = (b_1 + \ldots + b_K)/K \tag{1}$$

and the corresponding squared standard error is

$$s^2 = (s_1^{\ 2} + \ldots + s_K^{\ 2})/K + (1 + 1/K)\, B \tag{2}$$

where

$$B = \{(b_1 - b)^2 + \ldots + (b_K - b)^2\}/(K-1)$$

is the variance of the completed-data estimates over the sets of imputations. B is the between-imputation variance of the completed-data estimates; it can be interpreted as the contribution to the uncertainty due to the missing values. If an infinite number of sets of plausible values were imputed ($K \rightarrow +\infty$), the second term in (2) would be B or, more precisely, its expected value. Thus, B/K is an estimator of the information lost because only a finite number of sets of plausible values are used. Since B/K converges to zero very slowly as $K \rightarrow +\infty$, gains due to additional sets of plausible values diminish with K. In practice, $K = 5$ is often sufficient.

The constructor generates the plausible values from a model for missingness. A practical choice is a MAR model with as many conditioning variables as is feasible. If not all the (completely observed) background variables can be included in the model, variables that are correlated with the outcome variables should be preferred. Say, the model fit is a vector of regression parameter estimates, γ, with estimated sampling variance matrix Σ. In data on many subjects, normality of γ can usually be assumed. To generate a set of plausible values, a random draw is made from the (multivariate) normal distribution with mean γ and variance matrix Σ, yielding a plausible vector of regression parameters γ_1 . This vector defines a set of plausible fitted values, say, $\boldsymbol{x}\gamma_1$, for each non-respondent. The plausible values for the missing outcomes are generated so as to reflect the dispersion of the outcomes around the fit. Thus, if an ordinary regression is used to model MAR, the plausible values are $\boldsymbol{x}\gamma_1 + \varepsilon$, where ε are drawn from the normal distribution with zero mean, and variance $\sigma_1^{\ 2}$ itself drawn at random from the estimated sampling distribution of the residual variance σ^2. If the missing values are categorical, the model for them yields plausible probabilities, and a plausible value (category) is drawn using these probabilities. This process, starting with a draw of the plausible model parameters γ_2, γ_3, γ_4 and γ_5 is repeated to generate the other K–1 = 4 sets of plausible values.

The principal result underpinning the method of multiple imputation is that if the model for missing data is correctly specified and the complete-data analysis is efficient, then so is the average of the analyses of the data completed by the imputations (***Rubin*** 1987). Of course, we can never be certain that the model assumed is correct; it is extremely likely that it is not. However, the more general the model for missing values, the more plausible it is that the bias and loss of efficiency are modest. Trivial approaches to handling incomplete records, such as listwise and pairwise deletion and (single) mean imputation, can be interpreted as an imputation procedure with a very restrictive model, so they are inferior.

The method of multiple imputation has an obvious extension for multivariate outcomes. In many large-scale surveys, missing data occur in most variables, and so the method cannot be applied straightforwardly. Some ingenuity is called for, and the ideal of conditioning on all the available (background) information has to be compromised. For instance, the imputations can be organised in rounds, first imputing for variables with few missing values, and then proceeding to those with more missing values, conditioning on the variables for which plausible values have been generated in earlier rounds. See ***Longford*** et al. (2000) for an example.

4.1 The imputations in ISSP

Ideally, the imputations for a variable would be informed by the recorded values of all the other variables. This is not feasible to implement with ISSP because an unmanageably large number of patterns of missingness among the conditioning (informing) variables would have to be distinguished. Instead, we carry out the imputations in a sequence of rounds, starting with the background variables which tend to have fewer missing values.

Round 1

In the first round, we impute for *Sex* (in the database, variable `V200`, with values `male` and `female`), *Marital status* (`V202`, `married`, `widowed`, `divorced`, `separated`, and `never married`), and *Cohabitation with a steady partner* (`V203`, `yes`, `no`, and `not applicable`). For example, the survey for West Germany has sample size N = 1282 and 49 records (3.6%) are not complete for these three variables. Among the complete records there are 18 distinct score patterns and seven of the eight possible patterns of missingness. Most incomplete records (29) have a missing value for marital status only.

The patterns of missingness and score patterns for West Germany are summarized in Table 1. From the score patterns we see that category 0 of `V203` (question about cohabitation not applicable) occurs only with category 1 of `V202` (married). The score patterns for the incomplete records illustrate the complexity of the imputation task. For instance, the feasible

Table 1: The patterns of missingness and score patterns for the variables *Sex* (`V200`), *Marital status* (`V202`), and *Cohabitation with a steady partner* (`V203`) in the ISSP 1995 survey in West Germany.

A. Patterns of missingness

Pattern	*ooo*	*oo?*	*o?o*	*o??*	*?oo*	*?o?*	*???*
Count	1233	5	29	4	2	1	8

B. Score patterns (complete records)

Pattern	110	121	122	131	132	141	142	151	152
Count	461	4	6	8	19	2	5	47	111

Pattern	210	221	222	231	232	241	242	251	252
Count	362	7	64	12	13	6	4	32	70

C. Score patterns (incomplete records)

Pattern	12?	15?	1?1	1?2	1??	22?	2?1
Count	1	2	16	1	3	2	8

Pattern	2?2	2??	?10	??2	?52	???
Count	4	1	1	1	1	8

Notes: The coding of the variables: *Sex* (`V200`) 1 – male, 2 – female; *Marital status* (`V202`) 1 – married, 2 – widowed, 3 – divorced, 4 – separated, 5 – not married; *Cohabitation with a steady partner* (`V203`) 0 – not applicable (married, no partner), 1 – yes, 2 – no. The order of the digits in the score patterns and of the symbols *o* (observed), ? (missing), is (`V200`, `V202`, `V203`).

values for `V203` for the sole record with the pattern 12? are 1 and 2. With single imputation, we might choose 2 because among the complete records pattern 122 is more frequent than 121. As an alternative, we may leave it to the chance and draw at random from the binary distribution implied by the observed frequencies; $p = 0.4$ for pattern 121. This prob-

ability is only estimated, so we may reflect our uncertainty about the underlying value of p by drawing at random from the sampling distribution of p, which itself has to be estimated (approximately normal, with mean 0.4 and standard deviation 0.15). Further, the relatively large number of score patterns 1?1 should have a bearing on our consideration, because some of these records may have the complete-data pattern 121, but certainly not the other feasible pattern 122. These considerations are based on an implicit assumption of MAR. An example of an extreme MNAR is that all nonresponses to `V202` are from single subjects (code 5), where `Sex` is also missing, from men (code 2 on `V200`), and when `V203` is missing, from those living on their own (code 2).

In the multiple imputation for these three variables, we also assume that missing values arise at random (MAR). First, the multinomial probabilities of the score patterns in the complete data are estimated. For West Germany, this could be based on the 1233 complete records (96.2%).

In general, we should make use also of the incomplete non-empty records (for West Germany, 49 records, 3.8%). This is achieved by applying the EM algorithm, a general approach to estimation with incomplete data (***Dempster/Laird/Rubin*** 1977). The EM algorithm is an iterative procedure, with each iteration comprising two steps. In the application to estimating the probabilities of the patterns, the E-step estimates the complete version of each incomplete record, and the M-step evaluates the probabilities. Each incomplete record is associated with a set of feasible (complete-record) patterns. The result of an E-step are the estimated probabilities of (the expectations of belonging to) these feasible patterns for each incomplete record. In the M-step, these probabilities are treated as contributions to the counts for each category; from these (estimated) counts the estimated probabilities are calculated straightforwardly, by dividing by the sample size. Since there are few incomplete records, they contribute little information, and the EM algorithm converges very quickly; three iterations are sufficient.

The conditional distribution of the missing part of a record, given the observed part, is calculated from the unconditional distribution of the score patterns, by restricting them to the feasible patterns and normalizing their probabilities so that they add up to unity. For illustration, suppose there are only five score patterns, A–E, with probabilities $\boldsymbol{p} = (p_A, p_B, p_C, p_D, p_E)$

p_A	p_B	p_C	p_D	p_E	Total
0.12	0.22	0.15	0.45	0.06	1.00

and, for a given incomplete record, only patterns C and E are feasible. Then the conditional probabilities of these two patterns are $pr_C = p_C/(p_C + p_E) = 0.15/0.21 = 0.71$ and $pr_E = 0.29$.

Since the probabilities $\boldsymbol{p}$ are not known, a set of imputations is based on the *plausible* vector of probabilities $\boldsymbol{p}^*$, drawn from the estimated (approximate) sampling distribution of the estimator $\boldsymbol{p}$ of the probabilities. For each incomplete record, we then draw a completion from the conditional distribution of the missing part, given the observed part of the record. Normality of the estimator is justified when neither cell in the cross-classification of the three variables is small. With complete data, the variance matrix of $\boldsymbol{p}$ is estimated as $n^{-1}\{\mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^{\mathrm{T}}\}$, where n is the sample size. With incomplete data, the sampling variance cannot be evaluated analytically; it is bounded from below by what it would be had the data been complete, and from above by the estimated variance based on the complete records. In most instances, these two variances (variance matrices) differ only slightly. We use the estimated variance based on the complete records, preferring to err on the side of greater uncertainty about the missing values.

The imputation steps, drawing a plausible vector of probabilities $\boldsymbol{p}^*$ and a completion for each incomplete record, are replicated (repeated independently) $K = 5$ times. A practical way of organising the generation of one set of plausible values is by going through the missing patterns, and simulating a completion for each record with the given pattern.

Round 2

In the second round of imputation, missing values for *Education* are dealt with. *Education* is defined on an ordinal scale of seven points, from `None or still in education` to `Completed university education`. The model for imputed values conditions on *Sex* and *Marital status*, collapsed to a dichotomy, distinguishing only between those married and others. Thus, for the first set of plausible values, we complete the data for *Sex* and *Marital status* by imputing the first set of plausible values, and apply the procedure described for Round 1, generating one set of plausible values for *Education*. The other sets of plausible values for *Education* are generated by replication of this process; for set $i = 2, 3, 4, 5$:

1. the data for *Sex* and *Marital status* are completed by the *i*th set of plausible values (generated in Round 1);
2. the probabilities $\boldsymbol{p}$ of the score patterns for (*Sex*, *Marital status*, *Education*) are estimated;
3. a set of plausible probabilities $\boldsymbol{p}^*$ is drawn;
4. based on $\boldsymbol{p}^*$, a completion is generated for each incomplete record.

Further rounds. Background variables

Next, imputations are generated for the variables related to employment: *Current employment status* (10 categories), whether *Working for public or private sector* (four categories:

`Government employment`, `Public-owned firm`, `Private firm`, and `Self-employed`), whether *Self-employed* (four categories: `Not applicable`, `Self-employed`, `Working for someone else`, and `Combination`), and whether the subject *Supervises any employees* (three categories: `Not applicable`, `Yes`, and `No`). Although, the crosstabulation involves many cells, a large number of them are structurally empty, so their crosstabulation is, in effect, not a large table. The same conditioning (on *Sex* and dichotomized *Marital status*) is applied as for *Education*.

Although *Education* and *Current employment status* are associated, using one in the model for missing values of the other is problematic because a sparse table is obtained. For West Germany, one category of *Education* and two categories of *Current employment status* are not present in the data, and in the 48 cells of the crosstabulation, 19 cells have fewer than six records each, and only six cells have more than 50 records each. Further crossclassification (by *Sex* and *Marital status*) would make the table of counts even sparser. Yet, for West Germany, the number of missing values is very small, 12 for *Current employment status* and 18 for *Education*.

In the software developed in Splus (see Section 5.1), the conditioning variable is provided as an argument of a function, itself in the form of a function, so that its evaluation can make use of the imputed values for its missing items.

The same procedure is applied for further background variables: (personal) *Earnings* and *Family Income* (six and seven categories, respectively); *Subjective social class* (seven categories), *Trade union membership* (three categories) and *Party affiliation* (eight categories); *Religious denomination* (several categories, but only a few in every country), and *Attendance of services* (seven categories of frequency); and *Household size* (integers, 11 categories, with a ceiling for 11+). In the former three rounds of imputation, the following conditioning variables are used: *Education* (collapsed to two categories, whether completed secondary education or not) and whether *Supervises at work* (yes and others). For *Household size*, the conditioning is on *Cohabitation* (two categories), *Education* (whether completed secondary education), and *Attendance of religious services* (whether at least once a month). In each case, the completed version of the conditioning variables is used, completed by the corresponding set of plausible values generated in the earlier rounds.

Table 2: The numbers of missing values for the background variables. ISSP 1995, West Germany.

Round of imputation	Variable	Categories	Missing values	Conditioning variables
1	A. *Sex*	2	11	None
1	B. *Marital status*	5	41	None
1	C. *Cohabitation*	3	18	None
2	D. *Education*	7	18	A & B
3	E. *Employment status*	10	12	A & B
3	F. *Private/public*	5	38	A & B
3	G. *Self-employed?*	3	38	A & B
3	H. *Supervision*	3	16	A & B
4	I. *Earnings*	6	133	D & H
4	J. *Family income*	7	114	D & H
5	K. *Social class*	7	24	D & H
5	L. *Trade union*	2	39	D & H
5	M. *Party affiliation*	7	62	D & H
6	N. *Religion*	Many	18	D & H
6	O. *Attendance src's*	6	16	D & H
7	P. *Household size*	11	19	B, D, & O

Notes: Details of conditioning (aggregation of the categories) are given in the text. The numbers of categories are given as per definition, not specific to West Germany.

The sets of five imputations are stored in a single file, in which each record consists of the subject identification, the order number of the variable, and the set of five plausible values for the item. For West Germany, there are 618 missing values on the background variables considered, so this file contains a matrix of 618 rows and seven columns. Table 2 lists the numbers of missing values for each variable for West Germany.

Two important background variables not discussed in this procedure are *Age* and *Years of education*. They are effectively continuous, but for most purposes their categorical versions would suffice. For instance, age categories 16–24, 25–39, 40–64, and 65+ could be defined (and similarly for *Years of education*) and imputation of missing items for both variables conducted in a further round.

The Splus functions `MIca0` and `MIcat` implement the multiple imputation for a set of variables; `MIca0` is for no conditioning, and `MIcat` for conditioning on a single variable.

`MIcat` has the following arguments:

- the (incomplete) dataset (`dat`);
- the columns of `dat` for which plausible values are to be generated;
- the plausible values generated earlier (`MIm`);
- the number of imputations (`nmI`);
- the codes for missing values in `dat`;
- the function which evaluates the conditioning variable (`Condi`);

and several other arguments of technical nature (maximum number of iterations, convergence criterion, etc.). `MIca0` has the same arguments, except for `MIm` and `Condi`, which are not applicable.

An application of `MIca0` or `MIcat` yields the matrix of plausible values (`nmI`+2 columns), and the following information: number of iterations used, number of score categories, number of missingness patterns, number of complete records, total sample size, and the function used in conditioning. For instance, in the first round of imputation for West Germany, plausible values were generated for 70 missing items (a 70×7 matrix). Three iterations of the EM algorithm were required, there were 18 score patterns, six patterns of missingness (not counting the complete pattern), and 1233 complete records out of 1282.

In `MIcat`, a single conditioning variable is used. However, conditioning on several categorical variables is equivalent to conditioning on the crossclassification of the variables. For instance, if A and B are categorical variables with three and two categories, respectively, with respective values (1,2,3) and (0,1), then the categories of 2A+B correspond to the unique score patterns of (A,B); conditioning on A and B is equivalent to conditioning on the single variable 2A+B.

The imputed values from the different applications of `MIca0` and `MIcat` are stacked into a single matrix with seven columns. The plausible values in such a matrix can be explored by the function `Dimp`, which tabulates the values generated in each imputation. Thus, when five imputations are made, it produces five tables. For example, round 1 for West Germany yielded for the first set of plausible values the table

	Values						
Variable	0	1	2	3	4	5	Total
Sex	0	6	5	0	0	0	11
Marital status	0	10	3	8	2	18	41
Cohabitation	6	2	10	0	0	0	18

The output of `Dimp` also indicates that sets of five plausible values have been generated for each of 70 missing items, and gives the number of missing values for the three variables: 11 for *Sex*, 41 for *Marital status*, and 18 for *Cohabitation*.

The function `Imput` imputes into a dataset (`dat`) a given set of the plausible values (`imN`, an integer) from another dataset (`plv`). The result of applying `Imput` is a dataset of the same dimensions as `dat`, with the elements in the rows (records) and columns (variables) indicated in the first two columns of `plv` replaced in `dat` by the plausible values given in column 2+`imN`.

Imputation for 'National Identity' items

The response part of the questionnaire comprises blocks of items that have the same underlying theme. For instance, the five items of the first block (items `V44`–`V48`) share the lead-in passage

How close do you feel to your ... ?,

with the object of the question being the respondent's neighbourhood, town or city, county, country, and the continent. Each item is scored on a Likert scale (very close – 1, close – 2, neutral – 3, distant – 4, very distant – 5). Missing values are quite common in this and other blocks. For instance, in the survey for West Germany, only 586 subjects (45.7%) responded to all five items. Only a few of the $2^5 = 32$ patterns of missingness are absent, although most common are the patterns with only one item missing (509 records, 39.7%).

In contrast, the second block of items,

> *If you could improve your living and working conditions, would you be willing to move to another ... ?,*

(*neighbourhood, town, county, country, continent*), has many more complete records for West Germany, 1045, but a large proportion of the 237 incomplete records are empty. In fact, by far the most common patterns of missingness are ????? (59 cases) and *o*???? (37 cases).

Table 3 summarises the information about missing values in the blocks of questionnaire items, with information about missing values for West Germany. There are two different kinds of nonresponse: `Cannot choose` and similar (code 8), and `Refusal` (code 9). For simplicity, we do not distinguish between the two codes, and consider a single model for missing data. A more principled, but also more complex, approach would distinguish between the two codes.

In most analyses, a summary score of the responses of a subject to the items of such a block would be constructed. We therefore generate imputations only for these totals or, more generally, for linear combinations of the scores. The plausible values for these totals are more conveniently stored as records of eight sets of quintets of items each, one for each block. These records of plausible totals are defined not only for the (originally) incomplete records, but also for complete ones, as five copies of the same (observed) total. Thus, the plausible values of the within-block totals for West Germany are a 1282×40 matrix.

For the imputations, a MAR model is adopted, with conditioning on a selected set of background variables and on the total for the available items. Conditioning on the background variables means, in effect, that unrelated models are considered for each category defined by them. For the first two blocks of items, *Sex* and *Education* (whether completed secondary education) are conditioned on.

The imputations proceed in the order of increasing numbers of missing values, separately for each pattern of missingness. For records with one missing item, the total of the recorded scores is matched with the set of complete records that have the same (partial) total.

For records with two or more missing values, their partial totals on the available responses are matched with the records that have less missingness. For these ‘matching’ records, the total score is either available or plausible values have been generated earlier. Thus, the *i*th set of plausible totals for the ‘matched’ records is generated from the plausible distribution based on the *i*th set of plausible totals for the matching records. Finally, for the empty records, the matches are all the non-empty records, for each of which a set of plausible totals has been generated earlier.

The imputed total for an incomplete record is drawn from a plausible (conditional) distribution of the total scores of the complete records and records completed earlier. The estimators of the distributions of the various subtotals are correlated because they are based on overlapping subsamples and in some cases refer to the same components. It is impossible to keep track of these dependences and appropriately reflect them in the imputations. Instead, each plausible distribution is generated by an independent perturbation of the estimated vector of probabilities, erring on the side of overstating the uncertainty about the missing values.

Table 3: Blocks of items on National Identity in the 1995 ISSP and information on the patterns of missingness for West Germany.

Block of items	Number of items	Complete records (%)	Missing values (%)
1. How close do you feel . . .	5	586 (45.7)	1096 (17.1)
2. Move to improve conditions . . .	5	1045 (81.5)	696 (10.9)
3. Important to be . . .	7	1126 (87.8)	365 (4.1)
4. Nation – Country – Citizenship	6	945 (73.7)	607 (7.9)
5. Proud of . . .	11	769 (60.0)	1521 (10.8)
6. Traditions vs. Foreign countries	9	770 (60.1)	1030 (8.9)
7. Immigrants should (be) . . .	6	908 (70.8)	721 (9.4)
8. Ethnicity & citizenship	6	956 (74.6)	373* (9.7)
All items	55	186 (14.5)	6409* (9.6)

Notes: The blocks of items are: 1. – `V4`–`V8`, 2. – `V9`–`V14`, 3. – `V15`–`V21`, 4. – `V14` and `V22`–`V26`, 5. – `V27`–`V37`, 6. – `V38`–`V46`, 7. – `V47`–`V52`, and 8. – `V66`–`V71`. The bottom line (*All items*) gives the total for the numbers of items (55) and of missing values (6409).

*Only three items from the block 8 were administered in West Germany, so the total is for 52 items only.

In the imputations, we condition on *Sex* and on *Education*, collapsed to two categories, whether completed secondary education or not. This amounts to generating the imputations separately for each of the four categories defined by *Sex* and *Education*.

The same imputation process can be applied to the other blocks of items. It is practical to store the plausible totals in a single file in which the records contain the identification of the individuals and sets of *K* totals. For background variables, it is more practical to construct a file in which each record represents a missing item and the columns are subject identification, identification of the variable, and the set of five plausible values.

The procedure for generating plausible values has been implemented as function `MIsum` in Splus. Its arguments are: the dataset of incomplete records; the columns for which the totals are to be imputed; how the item scores are to be combined (e. g., as the total); the formula evaluating the (single) conditioning variable; the dataset of plausible values for the background variables; the number of imputations; the codes for the missing values; and some technical information. The function returns a matrix of plausible totals and the counts of records which have 0, 1, . . . missing values. Also, the number of records is indicated for which the total of the observed scores is not matched among the totals for the records with

less missingness within the same conditioning category. A record may be counted once for each imputation, certainly so when the background information for conditioning is complete. In case of a non-match a neighbouring total (±1) is considered as a substitute match. For West Germany, the largest number of non-matches encountered was 10 (out of 5×1282 = 6410 cases).

In block 8, three items were not administered in the survey for West Germany (coded as 0 for every subject), so imputation could be conducted in the same manner as for the background variables, using the function `MIcat`. For consistency, we generate the partial totals as for the other blocks.

5 Discussion

Implementing multiple imputation in a large-scale survey is a considerable burden, in its nature different from the other tasks of the constructor. Discussing its advantages should not therefore be constrained to the statistical perspective of efficient estimation with well-calibrated assessment of uncertainty. The first concern of the constructors' management is whether allocation of the additional funds can be justified and, perhaps, whether these costs could be recovered from the users. This concern leads to the problem of assigning a value to the precision in inferences (estimation) made with the public-release database. Most of this research is funded on the subsistence basis (by payment for the researcher's time, equipment, and expenses such as travel and administrative costs), and the researcher's output is usually not oriented for a specific application. In such a setting, the rather intangible outcomes of the research are very difficult to evaluate and the consequences of higher fees for the provision of the database to secondary analysts are difficult to anticipate.

Another concern of the constructor is that the multiple imputation is not guaranteed to be an improvement in all the analyses conducted by prospective secondary analysts, because the method is contingent on correct specification of the model for missingness. For instance, one might argue that the imputation should be postponed until a comprehensive and verified theory for the subject area is established. Imputation is not futile without such a theory. In any case, we cannot judge whether such a theory is on the horizon. More credible is the prospect of continual confirmation of unexplainable variation in the attitudes, perceptions, and opinions of human subjects in whatever population. This does not preclude the presence of some systematic, but far from dominant, differences among subpopulations and domains of issues, always affected by the context in which the survey takes place, as well as the details of how the interviewers are instructed, how interviews are conducted, and other details. If there were such a theory, handicapped by the reduced power due to inefficient use of incomplete records, its discovery would only be delayed. The additional costs and storage requirements for the plausible values should be set in the context of the costs of data collection, cleaning, coding, archiving, and documentation. The

research, design, and implementation of multiple imputation require a much more modest outlay. The storage requirement for multiple imputation for the 1995 ISSP survey for West Germany is two arrays (data files, or matrices) of dimensions 618×7 and 1282×41, amounting to about 57 000 items. In contrast, the original database contains 1282×214 =275 000 items, although its size could easily be reduced to about 110 000 without any loss of information. Thus, the additional storage requirement, although not trivial, is far from overwhelming. Note however, that our imputation scheme does not provide a set of plausible values for every missing item. The extent of missing values for the survey in West Germany is around the average for the 23 participating countries.

The following example illustrates the gains attained by using multiple imputation. A typical analysis for West Germany may be restricted to 700 complete cases; with multiple imputation, the effective sample size may rise to about 900. Thus, the standard errors of the estimated parameters would be reduced $\sqrt{(9/7)} = 1.13$ times. Had the data been complete for the variables used, the standard errors would be smaller $\sqrt{(1282/700)} = 1.35$ times, so multiple imputation is instrumental in recovering only a fraction of the missing information. Although the gain in precision is modest, it should be viewed in the context of a large number of conducted analyses. In particular, the constructor's challenge that implementing multiple imputation would be more attractive if several examples of analyses in which listwise deletion and multiple imputation yield different results is misplaced. The main benefit of multiple imputation is in small improvements, consistent over all the conducted analyses.

What if the multiple imputation is completely wrong? It is useful to distinguish two aspects of this question. First, that the model for missingness may be incorrectly specified, and second, that the procedure may be incorrectly implemented. With model specification, we do not rely on the actual values of the estimates obtained, but on the *estimated distribution* of the estimator. Thus, whatever model we select, it is more general than the MAR (MCAR) model implied by the casewise deletion. The implementation can be checked straightforwardly by inspecting the frequencies of the generated plausible values and their variation across the replications of the multiple imputation process.

Suggestions to experiment with deletions from a complete dataset to see how the results of the analyses are altered can be useful only if the process of multiple imputation is replicated, because multiple imputation is not meant to reproduce the complete-data results, but appropriately represent the additional uncertainty due to incomplete records. Neither is multiple imputation intended to reconstruct the missing values; it is intended to (near) optimally estimate population quantities.

It is difficult for multiple imputations to lead the analysis completely astray because the analysis still relies on the bulk of the data that are observed and left intact by multiple im-

putation. Some estimates may be further away from their underlying parameter values, but the *estimators* are very likely to have smaller mean squared errors, because, assuming valid (that is, efficient) complete-data analyses, the analyses conducted with multiple imputations use more information and are (almost) fully efficient under a wider range of models for the mechanism of missingness than the naive alternatives: casewise deletion, single imputation, or the use of all available items.

5.1 Software

The software implementing the described multiple imputation procedures was developed in Splus (***Becker/Chambers/Wilks*** 1988). Splus is a general environment for statistical computing and graphics. It encompasses an object-oriented programming language in which mathematical formulae are easy to transcribe and operations on data objects (matrices) require a very compact code. ***Venables/Ripley*** is an excellent text on Splus for a reader with little experience in statistical computing. The software could be developed in other statistical packages with flexible programming facilities (such as matlab or Gauss), but not in general purpose packages that provide only a limited set of modules or procedures (e. g., SPSS or SAS). The code, in the form of functions, can be obtained from the author (`ntl@dmu.ac.uk`) on request.

References

Becker, ***R. A***.; ***Chambers***, ***J. M***.; ***Wilks***, ***A. R***.; 1988: The New S Language. A Programming Environment for Data Analysis and Graphics. Wadsworth & Brooks/Cole, Pacific Grove, CA.

Dempster, ***A. P***.; ***Laird***, ***N. M***.; ***Rubin***, ***D. B***.; 1977: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Ser. B 39, 1–38.

Heitjan, ***D. F***.; ***Rubin***, ***D. B***.; 1991: Inference from coarse data via multiple imputation with application to age heaping. Journal of the American Statistical Association 85, 304–314.

Little, ***R. J. A***.; ***Rubin***, ***D. B***.; 1987: Statistical analysis of incomplete data. Wiley and Sons, New York.

Longford, ***N. T***.; 2000: Attitudes to immigration in an international social survey. Submitted.

Longford, ***N. T.; Ely***, ***M.; Hardy, R***.; Wadsworth, M. E. J.; 2000: Handling missing data in diaries of alcohol consumption. Journal of the Royal Statistical Society Ser. A 163, to appear.

Rosenbaum, ***P. R***.; 1995: Observational Studies. Springer-Verlag. New York.

Rubin, ***D. B***;. 1987: Multiple Imputation for Nonresponse in Surveys. Wiley and Sons, New York.

Rubin, ***D. B***.; 1996: Multiple imputation after 18+ years. Journal of the American Statistical Association 91, 473–489.

Rubin, ***D. B***.; Schenker, N.; 1991: Multiple imputation in health-care databases: An overview and some applications. *Statistics in Medicine* 10, 585–598.

Schafer, ***J. L***.; 1996: Analysis of Incomplete Multivariate Data. Chapman and Hall, New York.

Venables, ***W. N***.; ***Ripley***, ***B. D***.; 1999: Modern Applied Statistics with S-PLUS. Springer-Verlag, New York.