Chapter 8
Case Study on Dealing with Missing Values

Background A case study is presented on prognostic modelling in patients with moderate and severe traumatic brain injury (TBI). Individual patient data from several studies were available to quantify predictor effects and to develop and validate prognostic models. Missing values were a key issue, since few studies recorded all predictors of interest. The use of single and multiple imputation methods is illustrated with a detailed description of the analyses in R software.

8.1 Introduction

8.1.1 Aim

Randomized controlled trials (RCTs) in TBI are complex due to the heterogeneity of the population. None of the multicentre RCTs conducted in this field over the past decades have convincingly shown benefit of new therapies in the overall population.[273,310] The overall aim of the study was to optimize the methodology of randomized clinical trials in the field of TBI, such that chances of demonstrating benefit with an effective new therapy or therapeutic agent would be maximized. This NIH sponsored project was labelled IMPACT: International Mission on Prognosis and Analysis of Clinical Trials in TBI.[271] Individual patient data from recent trials and observational studies were available.

Prognosis was central to the aims of the project. For example, prognostic models can be used for the efficient selection of patients (excluding those with an extreme prognosis, either very poor or very good) and for covariate adjustment of the treatment effect (with several advantages as described in Chap. 2).[189] In TBI, outcome is commonly assessed with the Glasgow outcome scale (GOS), which is an ordinal scale (Table 8.1).[218] The scale ranges from dead, through vegetative state, severe disability to moderate disability, and good recovery. In conventional analyses, the GOS is often dichotomized as mortality vs. survival (category 1 vs. 2–5), or as

E.W. Steyerberg, *Clinical Prediction Models*,
DOI:10.1007/978-0-387-77244-8_8, © Springer Science+Business Media, LLC 2009

Table 8.1 Definition of the Glasgow outcome scale[218,491]

Category	Label	Definition
1	Dead	Mortality from any cause
2	Vegetative	Unable to interact with environment; unresponsive
3	Severe disability	Conscious but dependent
4	Moderate disability	Independent, but disabled
5	Good recovery	Return to normal occupational and social activities; may have minor residual deficits

unfavourable vs. favourable (category 1, 2, 3 vs. category 4, 5), although it is preferable to exploit the ordinal nature of this scale. One approach is the "sliding dichotomy" analysis, in which the split for dichotomization of the GOS is differentiated according to the baseline prognosis established prior to randomization.[304] Another approach is to use a proportional odds model for the GOS as an ordered outcome (see Chap. 4).

We aimed to predict the dichotomized 6-month GOS. Missing data were a key problem in the prognostic analysis.[283] We focus on approaches for dealing with missing data.

8.1.2 Patient Selection

Our focus was on patients with severe TBI (Glasgow coma score, GCS 3–8), but cohorts that included patients with moderate TBI (GCS 9–12) were also considered. The GCS is a measure for the level of consciousness. An individual patient data meta-analysis of 11 studies was performed, including 8 randomized controlled trials (RCTs), and 3 relatively unselected prospective surveys, with the potential for analysing data on 9,205 patients. Complete outcome data were available for 8,719 of the 9,205 patients (95%). We further excluded children, leaving 8,530 patients for analysis. The studies are arbitrarily designated as A to K in Table 8.2. The meta-analysis was a continuation of analyses of two related RCTs (Tirilazad, Table 8.2: study ID A and B).[203]

8.1.3 Selection of Potential Predictors

Extensive univariate analyses were performed within the IMPACT study of potential predictors. In combination with a review of the literature we identified predictors for further multivariable analyses.[305] These predictors included demographic characteristics (age).[306] injury details (cause of injury),[62] secondary insults (hypoxia and hypotension),[284] clinical measures of injury severity (Glasgow coma scale and pupillary reactivity),[276] characteristics of the admission CT scan,[272] and laboratory values.[446] For prognostic modelling, a core set of three strong predictors emerges from the literature since the 1970s, consisting of age, motor score, and pupillary

Table 8.2 Availability of predictor values by study (A–K), as included in the IMPACT study (n=8,530)[271]

Study	A	B	C	D	E	F	G	H	I	J	K	Total
N	1,118	1,041	409	919	1,510	350	812	604	126	822	819	8,530
Core predictors												
Age (%)	100	100	100	100	100	100	100	100	100	100	100	100
Motor score (%)	100	100	100	100	100	100	100	100	100	100	100	100
Pupils (%)	93	95	97	0	98	100	92	100	0	96	99	85
Secondary insults												
Hypoxia (%)	88	89	100	93	0	0	98	100	67	99	0	64
Hypotension (%)	97	97	0	93	0	98	99	100	83	99	100	75
CT												
CT class (%)	99	99	100	99	0	0	0	0	100	98	99	61
tSAH (%)	97	95	99	99	100	73	0	87	100	95	100	87
EDH (%)	98	99	0	99	100	100	95	0	100	100	100	87
Cisterns (%)	89	87	99	99	0	0	0	86	100	0	0	45
Shift (%)	89	88	99	99	100	0	0	89	100	0	100	73
Laboratory values												
Glucose (%)	96	99	0	95	96	85	0	0	98	0	0	57
Sodium (%)	98	96	0	96	96	95	0	64	98	0	0	62
Hb (%)	99	98	0	90	30	97	0	0	93	0	0	45
Platelets (%)	0	0	0	90	29	0	0	40	93	0	0	19
Prothrombin time (%)	0	0	0	0	29	0	0	48	91	0	0	10

reactivity. We subsequently expanded this core model to a 7-predictor model by including secondary insults and CT characteristics (CT classification, traumatic subarachnoid haemorrhage).[203] Further modelling studies were performed with inclusion of more predictors, but are omitted here.

*8.1.4 Coding and Time Dependency of Predictors

An important issue was the definition of predictors across the 11 studies. Definitions varied between data sets. The data extraction was guided by a data dictionary and original study documentation, which standardized the format of variables entered into the pooled data set. A consistent set of categories for coding was sought for each variable by collapsing more extensive codings into a simpler format. For example, the presence of hypoxia on admission was collapsed into a binary coding present/absent, although some data sets contained a more detailed coding as No/Suspect/Definite. "Cause of injury" raised this same issue but in a more complex form, since many and different categories were considered per study.[276]

A further issue was related to the time of measurement of a predictor. We aimed to consider predictors that would be available when patients were to be enrolled in an RCT, in line with the overall aim of the project. An interesting example is the

motor score, which is the prognostically most important element of the GCS. Four time points for assessment were defined: pre-hospital, first hospital (in case of secondary referral), admission, and post stabilization. Most data sets had data for at least two of these time points. For prognostic analysis we aimed to select the latest reliable assessment on admission to correspond with a baseline assessment prior to randomization, i.e. the post-stabilization score. If this was missing we used the next reliable value going back in time (admission, first in-hospital, pre-hospital). However, sometimes the motor score is not clinically obtainable because of early sedation or paralysis, required for artificial ventilation. The motor score was then coded as a separate category ("9," untestable), rather than considered as a missing value. This approach made the motor score available for all patients.

It can be debated whether a more formal analysis should have been used for defining the baseline motor score; e.g. a multiple imputation procedure might have considered all four time points of the motor score, providing a formally imputed post-stabilization motor score. MI might also have provided estimates for the untestable patients ("category 9"). However, the necessity for sedation and paralysis is related to the severity of injuries. In this specific case, missingness in the sense of "untestable" may possibly be of prognostic relevance, and imputation of a virtual motor score for "untestable" patients was hence not considered appropriate.

8.2 Missing Values in the IMPACT Study

Missing values were present in the outcome and in predictors. We discuss dealing with both below.

8.2.1 Missing Values in Outcome

Data on 6-month outcome were available for 10 of the 11 studies. For one however, only the 3-month GOS was measured (study E). Since the GOS is assumed to be relatively stable between 3 and 6 months, we imputed missing 6-month GOS with the 3-month GOS. This approach is consistent with the way in which missing outcome had been imputed in a small number of patients in the individual studies (Last Value Carried Forward approach). We chose not to further attempt imputation of the 6-month GOS in the 5% of patients in whom outcome remained missing, as not to compromise the interpretation of our outcome measure.

A more formal MI procedure could have been followed, incorporating the GOS patterns over time as available in some of the studies (e.g. 1, 3, 6, 12 months), and correlations with predictors.

8.2.2 Quantification of Missingness of Predictors

Table 8.2 summarizes the availability of predictors within the 11 studies of the IMPACT database. The main reason for missingness was absence of a predictor within a given dataset. If the dataset included a predictor, availability was generally high. Data for age and motor score (including the untestable category) were complete, but some studies had no data for pupils (studies D and I, Table 8.2). If pupils were recorded, data were complete in >90% in most studies. Secondary insults (hypoxia and hypotension) had not been recorded in some studies, but if recorded, data were quite complete.

CT scans are usually performed within hours after admission, after stabilization of the patient. CT scans provide important diagnostic information, and are often classified according to the Marshall classification.[280] This classification was available in 7 of the 11 studies, for 61% of the 8,530 patients. Other important CT characteristics, such as traumatic subarachnoid haemorhage (tSAH) and the presence of an epidural haemaotoma (EDH) were available in slightly higher numbers of patients. The presence of EDH is illustrated in Fig. 8.1.

Laboratory values were available for only few studies (Table 8.2). Glucose, pH, sodium, and Hb levels were available for around 50% of the patients, but platelets and prothrombin time (which are related to blood clotting), were available for less

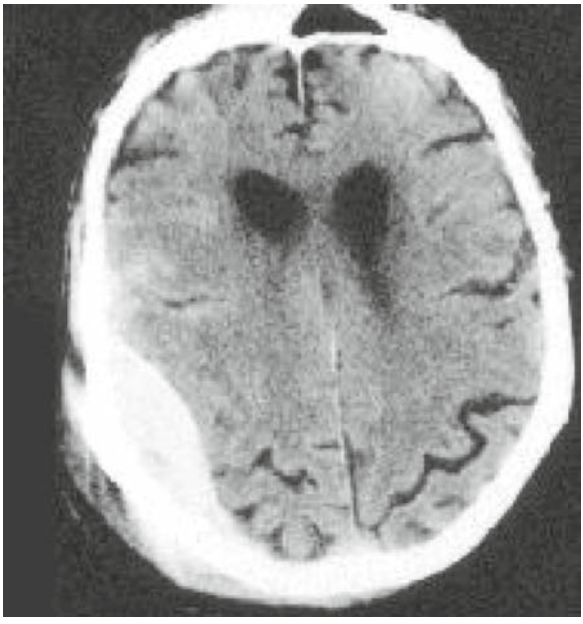

Fig. 8.1 Example of an epidural haematoma (EDH). An EDH is located directly under the skull and mainly causes brain damage due to compression. Consequently, prognosis is more favourable if it can be evacuated rapidly. A developing EDH is one of the greatest emergencies in neurosurgery

than 20% (Table 8.2). The latter percentages were so low that we did not consider these predictors for a prediction model; admittedly this judgment is arbitrary. A series of models was developed, with different selections of studies, based on availability of predictors per study.

8.2.3 Patterns of Missingness

We further examined patterns of missingness, following the steps discussed in Chap. 7.

a. How many missings occur for each potential predictor?

We used the `naclus` and `naplot` function to visualize missing value patterns. As was also noted in Table 8.2, missing values were most frequent for laboratory parameters and some CT characteristics (Fig. 8.2, left panel). Many patients had multiple missing values, e.g. 3,170 patients had 7 missing values, and 4 patients even had 12 missing values among the 15 predictors considered (Fig. 8.2, right panel).

b. Missing value mechanisms

For analysis of the mechanism of missingness we examined combinations of missing predictors, associations between predictors and missingness, and associations between outcome and missingness. As proposed by Harrell, we used the `naclus` function to visualize missing value patterns (Fig. 8.3).[174] We note that platelets and prothrombine time are often jointly missing, as also noted in Table 8.2. Characteristics of CT scans, such as shift and cisterns are often missing in combination, while also laboratory values are missing in such patients (hb, glucose, sodium, platelet, ptt).

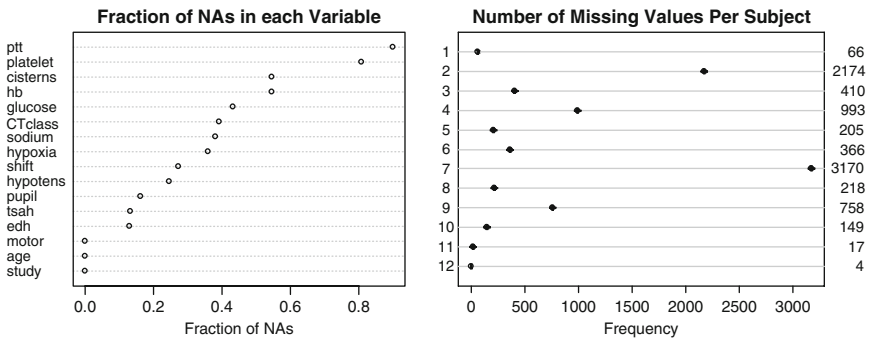

Fig. 8.2 Fraction of missing values per potential predictor (*left panel*), and number of missing values per subject (*right panel*)

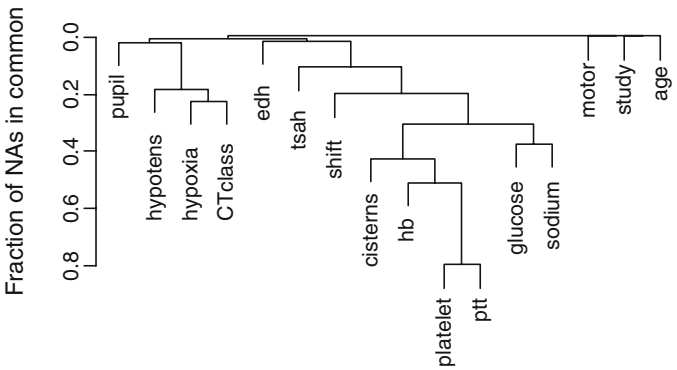

Fig. 8.3 Combinations of missing values in predictors ("NAs"), based on a hierarchical cluster analysis of missingness combinations

c. Associations between predictors and missingness

Table 8.2 demonstrates that missingness of most predictors strongly depends on study. We explored in detail whether there were other determinants of missingness for pupils, hypoxia, hypotension, CT class, tSAH, or EDH but no clear patterns were found (Fig. 8.4). Hence, no MAR on x patterns were evident.

d. Associations between outcome and missingness

Fig. 8.4 further demonstrates no clear associations between missingness and an unfavourable 6-month GOS outcome. To explore the relation between missingness and outcome in more detail, logistic regression models were constructed, but again no clear patterns were noted. Hence, there were no indications of an MAR on y mechanism.

e. Plausible mechanisms for missingness

The most plausible mechanism for missingness was that a predictor was simply not recorded for some studies. Within studies, a mechanism close to MCAR had occurred. We conclude that missingness was essentially MCAR, conditional on the study. Hence, we would like to stratify on study when making imputations. This is however logically impossible in situations that predictor values are 100% missing in a study, as study specific estimates cannot be derived.

We hence imputed values conditional on values of the other predictors, but not conditional on study. On the other hand, we excluded some studies from analyses if we judged that too many predictors were 100% missing in a study.

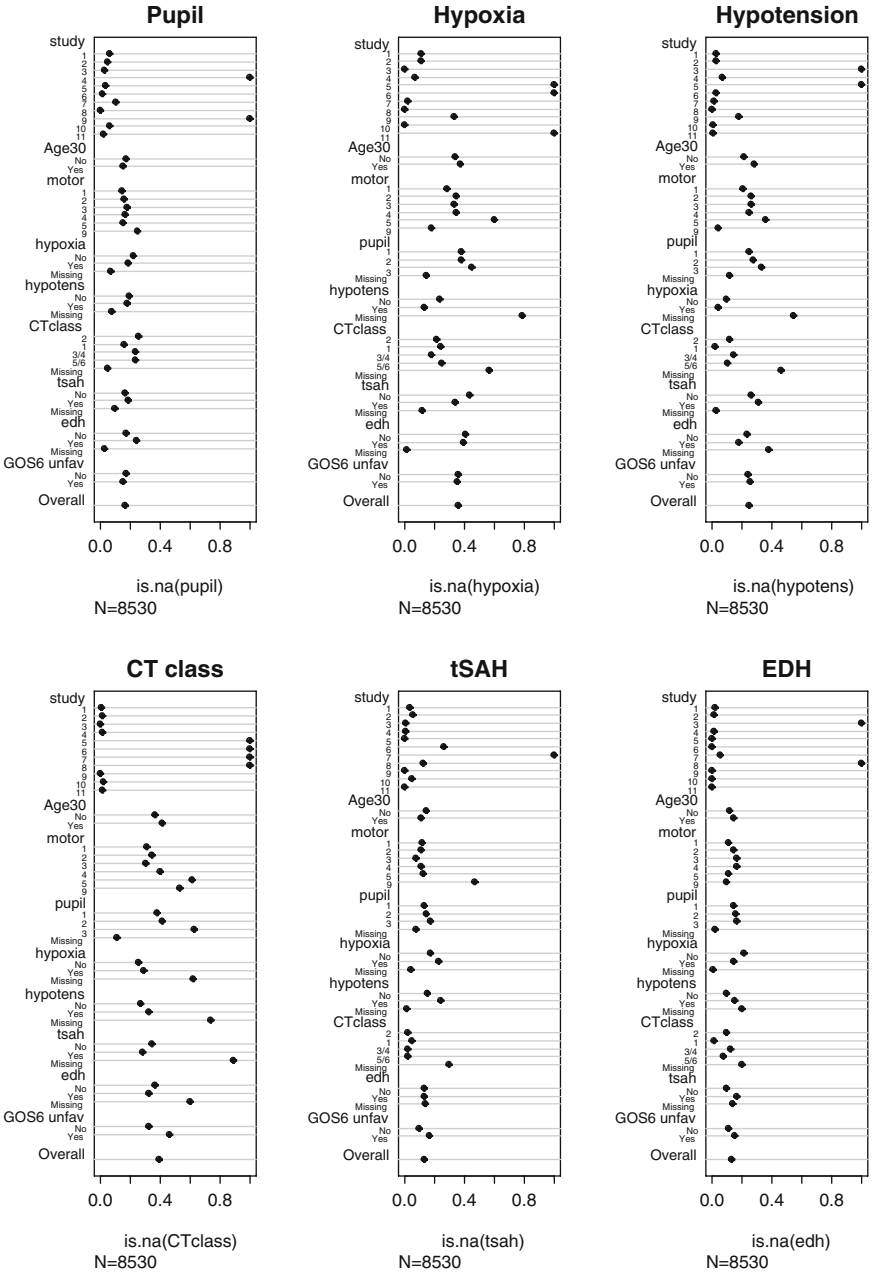

Fig. 8.4 Missingness in relation to study (numbered 1–11), other predictors (age to EDH), and outcome (GOS6). Study was the main determinant of missingness. Only weak associations were observed with other predictors, and no relationship with the 6-month outcome (GOS6)

8.3 Imputation of Missing Predictor Values

8.3.1 Correlations Between Predictors

In Chap. 7, we noted that multiple imputation became more relevant when predictors were correlated. Table 8.3 shows that the correlations between variables were generally modest, implying that both single and multiple imputation procedures may be considered. Some more substantial correlations (r>0.4) were noted among CT scan characteristics and between some laboratory values. The associations between cisterns/shift and the CT classification are to be expected, as these characteristics are used in the definition of the CT classification. Hb and platelets are correlated, as both will decrease following blood loss.

8.3.2 Imputation Model

An imputation model was considered that included all relevant potential predictors and the outcome (6-month GOS, in five categories). No auxiliary variables were used. The imputation model was fitted using the mice library and aregImpute from the Hmisc library in R. We show the commands below for illustration, with more details on the web site.

```
# mice imputation model for pmat as predictor matrix, with default
  settings
gm <- mice (TBIallR2, m = 10,
   imputationMethod  =c("polyreg",  "polyreg",  "pmm",  "polyreg",
   "polyreg",  "polyreg",  "logreg",  "logreg",  "logreg",  "pmm",
   "logreg",  "logreg",  "logreg",  "logreg",  "pmm",  "pmm",  "pmm",
   "pmm"), predictorMatrix = pmat, seed=1)
# aregImpute for data set TBIallR2, with default settings
g <- aregImpute (formula = ~d.gos+as.factor(trial) + age +
     as.factor(motorr) + as.factor(pupil) + as.factor(CTclass) +
     tsah + cisterns + shift + size + sdh + edh +
     hypoxia + hypotens + d.sysbpt + hbt + glucoset + sodiumt,
     n.impute = 10, data=TBIallR2)
```

Here, `d.gos` is the derived 6-month GOS; `trial` is the study; `age` is age in years; motorr is the Motor score; `pupil` is pupillary reactivity; `CTclass` is CT classification; `tsah` is presence of tSAH; `cisterns` is presence of compressed cisterns on CT; `shift` is shift ≥5 mm on CT; `size` is shift in mm; `sdh` and `edh` refer to subdural and epidural haematomas; `hypoxia` and `hypotens` refer to secondary insults; `d.sysbpt` is derived systolic blood pressure; `hbt` is truncated Hb; `glucoset` is truncated glucose; `sodiumt` is truncated sodium.

Table 8.3 Rank correlations between predictors, with correlations > 0.4 in bold

	Study	Age	Motor	Pupil	Hypoxia	Hypotens	CTclass	TSAH	EDH	Cisterns	Shift	Gluco	Sodiu	Hb	Platelet	Ptt
Study[a]	1	0.17	0.28	0.22	0.19	0.16	0.08	0.22	0.12	0.34	0.26	0.24	0.08	0.31	**0.52**	0.29
Age	0.17	1	−0.01	0.03	0.00	0.05	0.20	0.14	0.01	0.03	0.14	0.06	−0.07	−0.05	−0.17	0.03
Motor	0.28	−0.01	1	0.37	0.15	0.14	0.11	0.05	−0.02	0.21	0.11	0.11	−0.03	−0.06	−0.01	0.21
Pupil	0.22	0.03	0.37	1	0.14	0.17	0.22	0.11	−0.02	0.23	0.16	0.18	−0.06	−0.05	−0.03	0.11
Hypoxia	0.19	0.00	0.15	0.14	1	0.29	0.02	0.01	−0.05	0.02	−0.02	0.12	−0.01	−0.03	0.02	0.13
Hypotens	0.16	0.05	0.14	0.17	0.29	1	−0.02	0.05	−0.06	0.02	−0.03	0.15	0.02	−0.23	−0.14	0.31
Ctclass	0.08	0.20	0.11	0.22	0.02	−0.02	1	0.14	0.31	**0.44**	**0.48**	0.15	−0.06	−0.04	−0.01	−0.05
tSah	0.22	0.14	0.05	0.11	0.01	0.05	0.14	1	−0.04	0.13	0.07	0.10	−0.02	0.01	−0.04	0.12
EDH	0.12	0.01	−0.02	−0.02	−0.05	−0.06	0.31	−0.04	1	0.06	0.15	0.00	0.01	−0.05	−0.05	0.03
Cisterns	0.34	0.03	0.21	0.23	0.02	0.02	**0.44**	0.13	0.06	1	**0.51**	0.09	−0.03	−0.08	0.07	0.09
Shift	0.26	0.14	0.11	0.16	−0.02	−0.03	**0.48**	0.07	0.15	**0.51**	1	0.03	−0.01	−0.11	−0.10	0.08
Glucose	0.24	0.06	0.11	0.18	0.12	0.15	0.15	0.10	0.00	0.09	0.03	1	−0.13	−0.04	0.21	0.11
Sodium	0.08	−0.07	−0.03	−0.06	−0.01	0.02	−0.06	−0.02	0.01	−0.03	−0.01	−0.13	1	0.04	−0.08	0.06
Hb	0.31	−0.05	−0.06	−0.05	−0.03	−0.23	−0.04	0.01	−0.05	−0.08	−0.11	−0.04	0.04	1	**0.46**	−0.21
Platelet	**0.52**	−0.17	−0.01	−0.03	0.02	−0.14	−0.01	−0.04	−0.05	0.07	−0.10	0.21	−0.08	**0.46**	1	−0.34
Ptt	0.29	0.03	0.21	0.11	0.13	0.31	−0.05	0.12	0.03	0.09	0.08	0.11	0.06	−0.21	−0.34	1

[a]Based on generalized Spearman rank correlation as calculated with the `spearman2` function; other correlations based in standard Spearman rank correlation. All correlations were calculated with pairwise available patients

The `gm` and `g` objects each consist of ten imputed data sets of the IMPACT database. In total 18 variables were considered in the imputation model. Data were complete for the outcome (`d.gos`), `trial, age,` and motor score. With `aregImpute,` R^2 values are given to indicate how well each variable can be predicted from the other variables. R^2 values were very high for shift coded as a binary variable and `size` of `shift` in millimetres, which are by definition strongly correlated (shift defined as size ≥5 mm). Similarly, details of the imputations by `mice` can be inspected.

8.3.3 Distributions of Imputed Values

The distributions of imputed values in object *g* were checked for the plausibility of imputations (e.g. within a plausible range, no strange peaks, Fig. 8.5). The frequencies of categorical variables are shown as dot charts. For example, the first graph shows the imputations over ten sets for "pupil" (values 1, 2, 3), and the second for CTclass (values 1–6). For predictors that are treated as linear variables, the cumulative distribution is shown. For example, the third graph shows that imputed tSAH values were 0 in 60%, and 1 in 40%. Although size was considered as a linear variable, this does not imply that normality was assumed for the distribution; many values for "Imputed size" were zero. The before last graph shows imputations for glucose, which are truncated at 2 and 20, as in the original predictor definition.

8.4 Estimating Adjusted Effects

After imputation, we estimated the adjusted effects of each predictor of interest in turn, using imputed versions of other predictors. These other predictors are hence considered as potential confounders. We present all results for aregImpute for adjusted analyses; results with mice are only presented for the multivariable models. As confounders we considered seven predictors that had also shown convincing effects in previous TBI studies. These include the three core predictors (age, motor score, pupils), two secondary insults (hypoxia, hypotension), and two CT characteristics (CT classification and tSAH). The outcome was GOS at 6 months, dichotomized as unfavourable vs. favourable in logistic regression models. For illustration, we show the adjusted logistic regression coefficients of each of these predictors in turn (Table 8.4). We estimate adjusted effects in the complete cases (CC), as well as in completed data sets with single (SI) or multiple imputation (MI). Odds ratios can be calculated as $e^{\text{coefficient}}$.

Numbers of patients differ dramatically between the univariate and CC analyses, since only 2,428 patients had complete values for all 7 predictors considered. Per predictor, values were complete for some (age, motor score). Values were

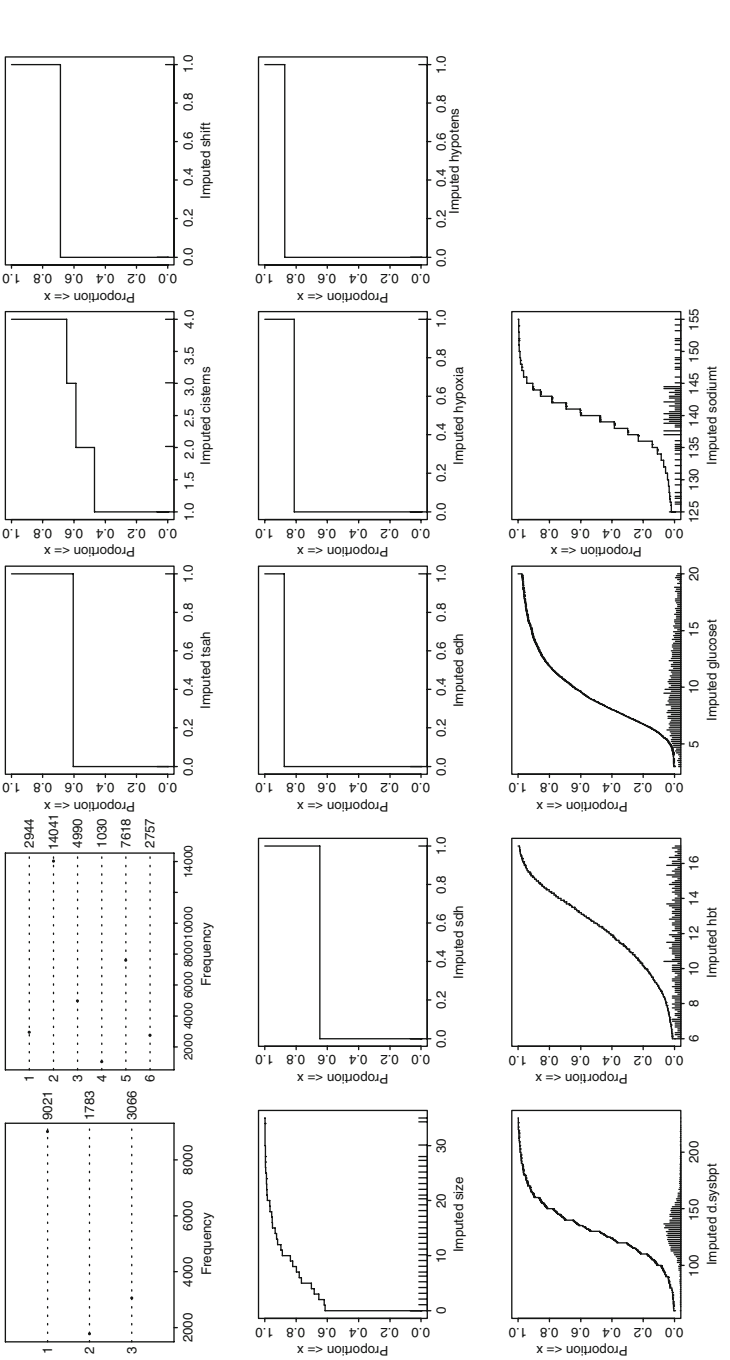

Fig. 8.5 Distribution of imputed values with aregImpute in the IMPACT study. The 14 imputations shown are for the predictors `pupil`, `CTclass`, `tasah`, `cisterns`, `shift`, `sdh`, `edh`, `hypoxia`, `hypolens`, `systbp`, `hb.glucose`, and `sodium`, respectively

Table 8.4 Logistic regression coefficients of predictors in univariate and adjusted analyses. Numbers are coefficients (SE)

	N	Univar	Adjusted		
	N	*N* = 5,192–8,530	CC, *n* = 2,428	SI, *n*=5,192–8,530	MI, *n* = 5,192–8,530
Age (per decade)	8,530	0.32 (0.015)	0.36 (0.033)	0.33 (0.018)	0.33 (0.018)
Motor score	8,530				
1 or 2		1.87 (0.065)	1.65 (0.160)	1.48 (0.074)	1.46 (0.075)
3		1.38 (0.077)	1.36 (0.157)	1.14 (0.086)	1.16 (0.087)
4		0.69 (0.065)	0.71 (0.128)	0.57 (0.071)	0.57 (0.072)
5 or 6		Zero (ref)	Zero (ref)	Zero (ref)	Zero (ref)
9		0.91 (0.112)	1.06 (0.259)	0.82 (0.127)	0.82 (0.128)
Pupillary reactivity	7,143				
Both pupil reactive		Zero (ref)	Zero (ref)	Zero (ref)	Zero (ref)
One non-reactive		0.97 (0.076)	0.51 (0.149)	0.56 (0.085)	0.57 (0.086)
Both non-reactive		1.77 (0.067)	0.94 (0.144)	1.18 (0.076)	1.18 (0.077)
Hypoxia	5,473	0.80 (0.072)	0.49 (0.125)	0.38 (0.085)	0.40 (0.087)
Hypotension	6,440	0.99 (0.070)	0.68 (0.133)	0.68 (0.084)	0.66 (0.085)
CT class	5,192				
1 or 2		Zero (ref)	Zero (ref)	Zero (ref)	Zero (ref)
3 or 4		1.08 (0.079)	0.77 (0.134)	0.78 (0.089)	0.77 (0.090)
5 or 6		0.96 (0.066)	0.67 (0.115)	0.55 (0.075)	0.54 (0.076)
Traumatic SAH	7,393	0.99 (0.050)	0.84 (0.101)	0.74 (0.057)	0.73 (0.058)

most incomplete for CT class (n = 5,192). The coefficients of most of the predictors were largest in univariate analyses, and smaller in adjusted analyses. This reflects the positive correlations between predictors (see Table 8.3). The estimates in adjusted analyses were largely similar for SI or MI, but were sometimes quite different from the CC analyses, e.g. smaller for motor score. The SEs in the CC analyses are higher than in the imputed analyses, reflecting smaller numbers. The MI analyses showed larger SEs than SI analyses, but differences were minor (3rd decimal).

Technical details of the model fitting are further discussed with detailed code for R programs. We first describe the modelling for complete predictors (age, motor), followed by the approach for predictors with missing values, such as pupils.

*8.4.1 Adjusted Analysis for Complete Predictors: Age and Motor Score

Age and motor score were completely available (n=8,530). Univariate effects can easily be estimated with logistic models:

```
lrm(d.unfav~as.factor(trial)+age, data=TBIall)
lrm(d.unfav~as.factor(trial)+as.factor(motorr),
data=TBIall)
```

The estimated regression coefficients are shown in Table 8.4.

Here, `d.unfav` refers to unfavourable GOS at 6 months, trial is the study indicator, such that analyses are stratified by study.

A CC model with adjustment for confounders included only 2,428 patients, due to exclusion of patients with any missing value for the other predictors (pupil, hypoxia, hypotens, CTclass, tSAH). Only patients from studies A, B, and J are included:

CC model:

```
lrm(formula = d.unfav ~ as.factor(trial) + age + as.factor(motorr) + as.factor(pupil)
+ hypoxia + hypotens + CTclass34 + CTclass56 + tsah, data = TBIall)

Frequencies of Missing Values Due to Each Variable
d.unfav trial   age  motorr  pupil hypoxia hypotens   CTclass34  CTclass56   tsah
0           0     0       0   1387    3057     2090        3338       3338   1137

Obs Max Deriv Model  L.R.   d.f.     P      C     Dxy   Gamma Tau-a     R2    Brier
2428         6e-010   840     14    0  0.823   0.645   0.646 0.315   0.393     0168

                         Coef          S.E.     Wald Z         P
      Intercept        -3.59911     0.186332    -19.32    0.0000
        trial=A        -0.14818     0.121132     -1.22    0.2212
        trial=J         0.07172     0.137206      0.52    0.6012
            age         0.03571     0.003348     10.67    0.0000
     motorr=1/2         1.64538     0.159743     10.30    0.0000
       motorr=3         1.35782     0.156498      8.68    0.0000
       motorr=4         0.71459     0.128421      5.56    0.0000
       motorr=9         1.06208     0.258924      4.10    0.0000
        pupil=2         0.51432     0.148866      3.45    0.0006
        pupil=3         0.94368     0.143710      6.57    0.0000
        hypoxia         0.49115     0.124781      3.94    0.0001
       hypotens         0.67864     0.133171      5.10    0.0000
      CTclass34         0.76777     0.134252      5.72    0.0000
      CTclass56         0.67493     0.114807      5.88    0.0000
           tsah         0.84091     0.101395      8.29    0.0000
```

In this specific case with complete data on age and motor score, fitting age and motor score with imputed data (SI or MI) is identical to fitting a model in the fully imputed data set (n=8,530). For SI, we create imputed data from the first MI data set in the `g` object, for example:

```
TBIall$pupil.i <- TBIall$pupil
TBIall$pupil.i[is.na(TBIall$pupil)] <- g$imputed$pupil[,1]
```

This is done for all predictors with missing values, with the extension "`.i`" added to indicate that we consider imputed data for a predictor.

SI model:

```
lrm (formula = d.unfav ~ as.factor(trial) + age + as.factor(motorr) + as.factor(pupil.i)
+ hypoxia.i + hypotens.i + CTclass34.i + CTclass56.i + tsah.i, data = TBIall)
```

```
Obs   Max Deriv Model  L.R.  d.f.  P     C    Dxy  Gamma  Tau-a   R2  Brier
8530  2e-009          2678    22  0 0.805 0.609  0.61 0.304  0.36  0.18

                    Coef       S.E.   Wald Z      P
     Intercept  -3.193738   0.111591  -28.62 0.0000
                …
           age   0.032630   0.001774   18.39 0.0000
    motorr=1/2   1.475716   0.074192   19.89 0.0000
      motorr=3   1.169648   0.085977   13.60 0.0000
      motorr=4   0.574532   0.071067    8.08 0.0000
      motorr=9   0.820593   0.126781    6.47 0.0000
     pupil.i=2   0.588143   0.076883    7.65 0.0000
     pupil.i=3   1.103948   0.068252   16.17 0.0000
     hypoxia.i   0.264818   0.068488    3.87 0.0001
    hypotens.i   0.670742   0.073482    9.13 0.0000
   CTclass34.i   0.570787   0.069579    8.20 0.0000
   CTclass56.i   0.491745   0.059293    8.29 0.0000
        tsah.i   0.723821   0.053876   13.43 0.0000
```

The MI model for age and motor score is fitted using the `fit.mult.impute` function, which automatically combines results over imputed data sets.

MI model:

```
fit.mult.impute(d.unfav ~ as.factor(trial) + age + as.factor(motorr) + as.factor(pupil)
   +hypoxia + hypotens + as.factor(CTclass == 3 | CTclass == 4) + as.factor (CTclass
   ==5 | CTclass == 6) + tsah, lrm, xtrans = g, data = TBIall)

Variance Inflation Facto rs Due to Imputation:

 Intercept trial=B trial=C trial=D  trial=E  trial=F  trial=G trial=H trial=I
    1.07     1.01    1.01    1.06     1.03     1.04     1.07    1.02    1.05
 trial=J trial=K  age  motorr=1/2 motorr=3 motorr=4 motorr=9 pupil=2 pupil=3 hypoxia
   1.01    1.05  1.02    1.03      1.02     1.03     1.02    1.53    1.56    1.76
 hypotens  CTclass=3/4  CTclass=5/6  tsah=TRUE
    1.15        1.59         1.23         1.1
  Obs   Max Deriv  Model  L.R.  d.f.  P     C     Dxy   Gamma  Tau-a   R2   Brier[1]
  8530    2e-009    2688          22   0   0.805  0.61  0.611  0.305 0.361  0.179

                 Coef      S.E.   Wald Z        P
  Intercept  -3.22374  0.116132  -27.76   0.0000
             …
         age   0.03321  0.001799   18.46   0.0000
  motorr=1/2   1.46459  0.075307   19.45   0.0000
    motorr=3   1.15620  0.086893   13.31   0.0000
    motorr=4   0.57085  0.072146    7.91   0.0000
    motorr=9   0.82014  0.128497    6.38   0.0000
     pupil=2   0.57979  0.095312    6.08   0.0000
     pupil=3   1.15770  0.085706   13.51   0.0000
     hypoxia   0.35117  0.090620    3.88   0.0001
    hypotens   0.63300  0.079237    7.99   0.0000
 CTclass=3/4   0.55875  0.088365    6.32   0.0000
 CTclass=5/6   0.47454  0.066028    7.19   0.0000
        tsah   0.73805  0.056594   13.04   0.0000
```

[1]These statistics are from the last fit with imputed data, in this case the tenth imputed data set

In conclusion, single and multiple imputation yielded very comparable results in this example: model statistics were similar (LR statistic, *c* statistic, R^2 estimate), as well as regression coefficients and standard errors.

*8.4.2 Adjusted Analysis for Incomplete Predictors: Pupils

Pupillary reactivity was recorded for 7,143 patients. This selection of patients was used in univariate and adjusted analyses.

Univariate analysis:

```
lrm(d.unfav ~ as.factor(trial) + as.factor(pupil), data = TBIall)
Frequencies of Missing Values Due to Each Variable
d.unfav trial pupil
0       0     1387
Obs Max Deriv Model L.R. d.f. P C     Dxy   Gamma Tau-a R2    Brier
7143 le-008         1097 10   0 0.708 0.417 0.441 0.208 0.19 0.213

               Coef     S.E.  Wald Z        P
 Intercept  -0.73594  0.06787  -10.84   0.0000
          …
   pupil=2   0.96801  0.07590   12.75   0.0000
   pupil=3   1.77194  0.06670   26.56   0.0000
```

Adjusted analysis following single imputation:

```
lrm(d. unfav ~ as.factor(trial) + age + as.factor(motorr) + as.factor(pupil) +
   hypoxia.i + hypotens.i + CTclass34.i + CTclass56.i + tsah.i, data = TBIall)
Obs Max Deriv Model L.R. d.f. P C     Dxy   Gamma Tau-a R2    Brier
7143 le-008         2403 20   0 0.814 0.628 0.629 0.314 0.381 0.175

               Coef      S.E.   Wald Z      P
Intercept   -3.25518  0.120631   -26.98 0.0000
…
pupil=2      0.55594  0.085164     6.53 0.0000
pupil=3      1.17557  0.076062    15.46 0.0000
…
```

For adjusted analyses, we can also use multiple imputations, e.g. from aregImpute. We first rename the predictor of interest (e.g. "`.o`" for "original") such that this predictor is not imputed:

```
TBIall$pupil.o <- TBIall$pupil
fit.mult.impute(d.unfav ~ as.factor(trial) + age + as.factor(motorr) + as.factor(pupil.o) +
hypoxia + hypotens + as.factor(CTclass==3|CTclass==4) + as.factor(CTclass==5|Ctclass
==6) + tsah, lrm, xtrans = g2, data = TBIall)
```

This original version of the pupil variable remains missing in 1,387 patients:

```
Frequencies of Missing Values Due to Each Variable
d.unfav   trial    age   motorr  pupil.o  hypoxia hypotens CTclass    tsah
   0          0      0        0     1387        0        0       0       0
Obs Max Deriv Model L.R. d.f. P C     Dxy   Gamma Tau-a R2    Brier
7143 9e-009         2386 20   0 0.813 0.626 0.627 0.313 0.379 0.176

               Coef      S.E.  Wald Z       P
 Intercept  -3.28470   0.12572  -26.13  0.0000
          …
 pupil.o=2   0.56648   0.08624    6.57  0.0000
 pupil.o=3   1.17570   0.07670   15.33  0.0000
          …
```

Again, the results obtained with single or multiple imputation procedures were very similar. Analyses for the other predictors with missing values were performed in a similar way. A series of papers presents further results for the other predictors with missing values.[62,272,276,284,306,446]

8.5 Multivariable Analyses

After studying adjusted effects per predictor, we are further interested in the multivariable effects of all predictors combined. We start with a core model, consisting of three predictors age, motor score, and pupils. All studies could reasonably be considered for this model, since they had age and motor score completely available (n=8,530). A CC analysis included 7,143 patients, because of 1,387 missing values for pupils. These 1,387 values led to exclusion of 1,387/8,530=16% of the patients, while they represented 5.4% of the required values for the three predictors.

Next, we considered a more extended model, including the seven predictors that were also used as confounders before: three core predictors plus secondary insults plus CT characteristics. It was not considered reasonable to include study #E in this analysis, since secondary insults and CT classification was not recorded in the database for this study. We hence considered 10 studies, with a total of 7,020 patients. These were included in SI and MI procedures. A CC analysis was possible with only 2,428 patients, representing a loss of 4,592 patients (65%), while only 13% of the required values were missing (6,426 of 7×7,020=49,140).

The multivariable coefficients are shown in Table 8.5, together with rounded prognostic scores. Scores were based on multiplying coefficients by 10, and rounding to whole numbers ("`round(10*fit$coef)`"). We note that the SI and MI coefficients and prognostic scores were largely similar. Scores never differed by more than 2 points. The CC analysis gave quite different estimates compared to SI or MI, demonstrating the substantial limitations of CC analyses. Prognostic scores with MI were lower for motor scores, larger for pupils. lower for hypoxia, and similar for CT characteristics.

8.6 Concluding Remarks

This case study illustrates how we may deal with missing values in assessing predictor effects (univariate and adjusted effects), and in multivariable modelling to derive prediction models. The difference in numbers of patients was dramatic between complete case and single or multiple imputed data. Since a reasonable imputation model could be constructed, we should have more confidence in the results after imputation (either SI or MI) than the CC results. The presented R code is available at the book's web site, and may be useful in implementing MI in other case studies.

Table 8.5 Multivariable regression coefficients and rounded prognostic scores for a 7-predictor model in the IMPACT study.

	CC, *n*=2,428		SI, *n*=7020				MI, *n*=7020			
			`aregImpute`		`mice`		`aregImpute`		`mice`	
	Coef	Score	Coef	Score	Coef	Score	Coef	Score	Coef	Score
Age (coef per decade score per 3 year)	0.36	1	0.32	1	0.33	1	0.31	1	0.34	1
Motor score										
1 or 2	1.65	17	1.44	14	1.54	14	1.42	14	1.57	16
3	1.36	14	1.12	11	1.13	11	1.11	11	1.18	12
4	0.71	7	0.54	5	0.54	5	0.53	5	0.57	6
5 or 6	Zero (ref)		Zero (ref)		Zero (ref)		Zero (ref)		Zero (ref)	
9	1.06	11	0.82	8	1.04	10	0.79	8	0.83	8
Pupillary reactivity										
Both pupil reactive	Zero (ref)		Zero (ref)		Zero (ref)		Zero (ref)		Zero (ref)	
One non-reactive	0.51	5	0.62	6	0.48	5	0.60	6	0.48	5
Both non-reactive	0.94	9	1.14	11	1.09	11	1.22	12	1.01	10
Hypoxia	0.77	8	0.30	3	0.37	4	0.38	4	0.33	3
Hypotension	0.67	7	0.68	7	0.61	6	0.64	6	0.58	6
CT Class										
1 or 2	Zero (ref)		Zero (ref)		Zero (ref)		Zero (ref)		Zero (ref)	
3 or 4	0.84	8	0.62	6	0.62	6	0.63	6	0.59	6
5 or 6	0.49	5	0.51	5	0.45	5	0.49	5	0.43	4
Traumatic SAH	0.68	7	0.72	7	0.64	6	0.74	7	0.62	6

Questions

8.1 Missingness mechanisms
We state that most predictors were missing complete at random (MCAR), conditional on study (Sect. 8.2.3 e).
(a) Does Table 8.2 support an MCAR mechanism?
(b) What do we learn from Fig. 8.4 with respect to MAR on x, or MAR on y mechanisms?
(c) Can we exclude a MNAR mechanism from the presented tables and figures?
(d) The imputation models did not include "study" as a variable. Why was this desirable, but not possible?

8.2 Imputation results (Sect. 8.4.1)
(a) For the MI model, the `aregImpute` imputation procedure lists "`Variance Inflation Factors Due to Imputation`." What do these factors refer to? When are they larger than 1? Which predictor has the largest VIF?
(b) Compare the predictor effects of age between the CC, SI, and MI models. When is the standard error estimated as the smallest?

8.3 Numbers in adjusted vs.multivariable analyses (Sect. 8.4.2 and 8.5)
The adjusted analysis for the predictor pupillary reactivity ("`pupil`") was performed with 7,143 patients (Sect. 8.4.2), while the multivariable analysis included 7,020 patients (Table 8.5).
(a) How did this difference arise?
(b) Do you agree with this approach? Or explain alternatives.