

Aurélien Vesin
Elie Azoulay
Stéphane Ruckly
Lucile Vignoud
Kateřina Rusinová
Dominique Benoit
Marcio Soares
Paulo Azevedo-Maia
Fekri Abroug
Judith Benbenishty
Jean Francois Timsit

Reporting and handling missing values in clinical studies in intensive care units

Received: 5 November 2012
Accepted: 25 April 2013
Published online: 18 May 2013
© Springer-Verlag Berlin Heidelberg and
ESICM 2013

Electronic supplementary material

The online version of this article
(doi:[10.1007/s00134-013-2949-1](https://doi.org/10.1007/s00134-013-2949-1)) contains
supplementary material, which is available
to authorized users.

A. Vesin · S. Ruckly · L. Vignoud ·
J. F. Timsit

Albert Bonniot Institute, U823, University
Grenoble 1, Grenoble, France

A. Vesin · S. Ruckly · J. F. Timsit
Biostatistical Department, Outcomerea
Organization, Paris, France

E. Azoulay
Medical ICU, St Louis University Hospital,
Paris, France

K. Rusinová
Medical ICU, Prague University Hospital,
Prague, Czech Republic

D. Benoit
Department of Intensive Care Medicine,
Ghent University Hospital, Ghent, Belgium

M. Soares
Instituto Nacional de Câncer,
Rio De Janeiro, Brazil

P. Azevedo-Maia
Department of Anesthesia and Intensive
Care, Hospital de S.João, Porto, Portugal

F. Abroug
Intensive Care Unit, CHU Fatouma
Bourguiba, Monastir, Tunisia

J. Benbenishty
Hadassah Medical Centre, Jerusalem, Israel

J. F. Timsit (✉)
Medical ICU, University Grenoble 1,
University Hospital Albert Michallon,
La Tronche, 38000 Grenoble, France
e-mail: jftimsit@chu-grenoble.fr
Tel.: +33-4-76768779
Fax: +33-4-76765569

Abstract *Introduction:* Missing values occur in nearly all clinical studies, despite the best efforts of the investigators, and cause frequently unrecognised biases. Our aims were (1) to assess the reporting and handling of missing values in the critical care literature; (2) to describe the impact of various techniques for handling missing values on the study results; (3) to provide guidance on the management of clinical study analysis in case of missing data. *Methods:* We reviewed 44 published manuscripts in three critical care research journals. We used the Conflicus study database to illustrate how to handle missing values. *Results:* Among 44 published manuscripts, 16 (36.4 %) provided no information on whether missing data occurred, 6 (13.6 %) declared having no missing data, 20 (45.5 %) reported

that missing values occurred but did not handle them and only 2 (4.5 %) used sophisticated missing data handling methods. In our example using the Conflicus study database, we evaluated correlations linking job strain intensity to the type and proportion of missing values. Overall, 8 % of data were missing; however, using only complete cases would have resulted in discarding 24 % of the questionnaires. A greater number and a higher percentage of missing values for a particular variable were significantly associated with a lower job strain score (indicating greater stress). Among respondents who fully completed the job strain questionnaire, the comparison of those whose questionnaires did and did not have missing values showed significant differences in terms of age, number of children and country of birth. We provided an algorithm to manage clinical studies analysis in case of missing data. *Conclusion:* Missing data are common and generate interpretation biases. They should be reported routinely and taken into account when modelling data from clinical studies.

Keywords Missing data · ICU · Multiple imputation · Job strain · Bias

Introduction

In clinical studies, missing data are common despite the best efforts of the investigators [1, 2]. Although missing data are virtually unavoidable, their existence is not always reported. In many studies, the existence of missing data is simply ignored. Another common strategy, which is the default setting of statistical software, consists in deleting observations that have missing values.

A missing value occurs when data for a variable or question are not collected, not available or not appropriate, leading to an empty cell in the data set. Sometimes, missing information may constitute a meaningful value (as when the response is “I don’t know”) that deserves to be analysed. Ignoring or inadequately handling missing values can lead to biases and to loss of statistical power [3–5]. In randomised controlled trials (RCTs), patients with adverse events often discontinue their treatment and therefore fail to undergo an evaluation of the primary end-point. Consequently, the population of patients with data on the primary end-point is not representative of the initial population. In extreme cases, the bias and loss of power can cancel out or even reverse the treatment effect. Also, sample size calculations are often performed without accounting for incomplete observations, and the sample included in the statistical analysis is therefore smaller than expected. Missing data affect not only RCTs, but also all observational studies.

The missing value issue was first investigated in depth in the 1970s, most notably by Rubin, who developed multiple imputation in 1987 [6]. As the need for practical solutions to handle missing values became recognised, research in the field expanded and the number of theoretical and applied studies increased at a fast pace. Today, strategies for handling missing data in clinical research continue to generate vigorous controversy.

The patterns of missing values in a data set can be classified into three main categories [4] depending on the relationship between the process causing the absence of data and the measured or unmeasured variables. The effect of ignoring the missing data on the reliability of the statistical analysis varies across the three categories.

In the missing completely at random (MCAR) pattern, the probability of the value being missing is independent from all patient characteristics. When representing the data set as a large matrix, the missing values are scattered randomly throughout the matrix. Consequently, the subsample of observations without missing data is representative of the original sample. This pattern is unlikely to be observed. In the missing at random (MAR) pattern, the probability of data being missing is dependent on other observed covariates. In a survey, for instance, young adults are more likely not to declare income; therefore, income is MAR on age. In the third pattern, not missing at random (NMAR), it is assumed that the

probability of data being missing cannot be explained solely by other observed covariates but depends instead on the incomplete information itself. For example, it may be assumed that persons with high incomes are less likely to communicate their incomes; therefore, the non-response probability depended on the nature of the missing value.

Analyses such as missing indicator modelling can be performed to distinguish between the two random mechanisms (MCAR and MAR) but do not allow testing of NMAR assumptions.

The objectives of this study were to assess the frequency of missing data reporting in the published critical care literature and to illustrate, on the basis of a large data set from the Conflicus study [7], the consequences of handling versus ignoring missing data and the impact of various techniques for handling missing data. Recommendations on how to handle missing data are provided.

Methods

Review of the ICU literature

To evaluate the reporting and handling of missing data in clinical studies, we reviewed the articles published in the October 2010 issues of three major critical care journals, *Intensive Care Medicine* (ICM), *American Journal of Respiratory and Critical Care Medicine* (AJRCCM) and *Critical Care Medicine* (CCM). All articles providing a statistical analysis on a patient population were reviewed, regardless of study design or size (observational study or clinical trial). Each article was carefully reviewed by two readers (AV and JFT) working independently of each other. The evaluation was standardised by having the readers answer three questions: (a) Were there any missing values? (b) Was information on missing values provided in the methods or results section? (c) Was a specific strategy used to handle the missing data and, if yes, which one? The readers then met to discuss their reviews and to reach a consensus.

The Conflicus study

The Conflicus study is a 1-day cross-sectional study designed in 2006 by the European Society of Intensive Care Medicine (ESICM) from 306 ICUs in 26 countries. The ESICM ethics section prepared a questionnaire to collect data on ICU conflicts, to be completed by all staff members working in each participating ICU on 7 December 2006. The primary objective of the study was to examine the frequency, characteristics and risk factors of conflicts in ICU [7]. As job strain was known to be strongly associated with conflicts in the ICU, an ancillary

objective was to identify risk factors for job strain in ICU workers.

Job strain was measured using a 12-item scale derived from the Job Content Questionnaire [8]. This scale explores three domains (job demand, job control and social support). The score is the difference between well-being (represented by job control score plus social support score) and the job stress (job demand score). A high score indicates a low level of stress. In addition to respondent characteristics, each ICU was asked to provide detailed information on the ICU and hospital. Country characteristics were retrieved from the World Health Organisation website (<http://www.who.int/research/en/>). For the sake of simplicity, we decided to focus only on respondents' characteristics.

Among the 7,771 respondents initially included, 7,209 (93 %) completed the job strain questionnaire. For the

remaining questionnaires the level of missingness of other variables was higher than 94 %, making imputation of job strain hazardous.

Because the questionnaires were self-completed by the ICU workers, missing data were expected. As reported in Table 1, the frequency of missing data for each of the 17 items ranged from 0 to 8 %. The first step in dealing with missing data is to identify the nature of the missing information.

Statistical analysis of Conflicus study data

Respondent characteristics were described using *n* (%) for qualitative variables, and median (interquartile range, IQR) or mean \pm SD for quantitative variables. The frequency and percentage of missing values for each variable were collected.

Table 1 Characteristics of respondents when information was available and frequencies of missing values per variable

Variable	Missing, <i>n</i> (%)	Modalities	Data ^b
Age	197 (2.7)		34 [28; 42]
Sex	30 (0.4)	Female	5,079 (70.7)
Single	259 (3.6)		2,807 (40.4)
Number of children	443 (6.1)		1 [0; 2]
Born in the country of current employment	62 (0.9)		6,293 (88.1)
Years of experience in the ICU	222 (0.3)		6 [2; 12]
Degree of religiosity	290 (4.0)	0: not religious	2,074 (30)
		1	833 (12)
		2	966 (14)
		3	1,677 (24.2)
		4	663 (9.6)
		5: very religious	706 (10.2)
Graduated in the country of current employment	255 (3.5)		6,387 (91.8)
Job title	160 (2.2)	Nurse	3,235 (45.9)
		Nurse assistant	1,130 (16)
		Head nurse	267 (3.8)
		Physiotherapist	352 (5)
		Junior physician	504 (7.1)
		Senior physician	582 (8.3)
		Consultant	318 (4.5)
		Director	143 (2)
		Other	518 (7.3)
Average hours of work per week	129 (1.8)		40 [36; 50]
Number of weeks since last vacation	416 (5.8)		12 [5; 17]
Average hours of work per day	151 (2.1)		8 [8; 12]
Formal training in ethics or communication	0 (0)		3,123 (43.3)
Number of end-of-life patients cared for within the last week	364 (5.0)		1 [0; 2]
Number of treatment-limitation decisions shared within the last week	607 (8.4)		0 [0; 1]
Number of deaths within the last week	387 (5.4)		0 [0; 2]
Receiving antidepressant therapy	61 (0.8)		575 (8)
Job strain score ^a	0 (0)		5 [3; 6]
Number of conflicts	0 (0)	0	2,301 (31.9)
		1	4,906 (68.1)
		≥ 2	2 (0)

^a A low score reflects a high level of stress

^b The data are *n* (%) for qualitative variables and median [25 % quartile; 75 % quartile] for quantitative variables

A missingness indicator was built for each item by assigning a value of 1 in the event of missing data for a respondent and of 0 otherwise. We then used the Wilcoxon non-parametric test to assess the relationship between the job strain score and the missingness indicator for each item.

Respondent-related variables associated with the job strain score were identified using a linear mixed model including random effects for country and ICU and discarding missing values. We then built a multivariate model including all respondent-related variables yielding p values smaller than 0.10 by univariate analysis. Observations with at least one missing value for one included variable were withdrawn from model fitting (this corresponded to the default process when using SAS software, i.e. “complete-case analysis” or “listwise deletion”). Then, the multivariate model was rebuilt twice, by replacing missing values with the median value for quantitative variables or the most frequent modality for qualitative variables and by using multiple imputations with IVEware. Multiple imputations create several copies of the original data set, in which missing values are imputed by values that differ slightly across the copies. This approach reflects the uncertainty regarding the imputed value. IVEware uses sequential regression multiple imputation (SRMI). Briefly, it specifies an imputation model for each incomplete variable, using other variables as predictors. The imputed variables are used in subsequent imputation models, and the process is repeated until convergence occurs (see electronic supplementary material 1 and 2 for details).

We did not check for interaction among risk factors in the multivariate model and did not address the issue of multiple testing. Post-imputation diagnostics were used for multiple imputation process assessment [9]. We assumed that our data were MAR. It was an intermediate compromise between plausibility (multiple imputation methods behave well with MAR which is more realistic than the MCAR assumption) and complexity (considering NMAR would need more complex consideration going beyond the scope of this article). For both imputation techniques, the percentage of variation of estimates [and standard errors (SEs)] from the complete case analysis were computed.

All statistical analyses were performed using the MIXED and MIANALYZE procedures in SAS software v9.3 (Cary, NC, USA) and IVEware v0.1 (Survey Research Centre Institute for Social Research, University of Michigan, MI, USA).

Results

Literature review

The detailed characteristics of the articles are reported in the electronic supplementary material 6. Of the 44

articles, 4 (9 %) were clinical trials and the remaining 40 (91 %) were observational studies. The median (IQR) number of analysed subjects was 749 (138–3,213). From the 44 articles, 16 (36.4 %) provided no information on whether missing data occurred, 6 (13.6 %) declared having no missing data, 20 (45.5 %) reported that missing values occurred but did not handle them (complete case analyses) and only 2 (4.5 %) used sophisticated statistical methods to manage missing data (multiple imputation and missing variable indicator, respectively).

Conflicus data analysis

The characteristics of respondents when information was available and frequencies of missing values per variable are described in Table 1. The proportion of missing data for independent variables ranged between 0 and 8.4 % (median = 2.9 %). Taking the missing-data pattern into consideration, multivariate analysis using the software default complete-case approach and including all 18 variables would have excluded 24 % of the data set observations (data not shown).

When studying the relationship between the missing-data pattern and job strain, we found that numerous questionnaire items with missing values were associated with the job strain score. Significantly higher job strain scores (i.e. lower stress) were noted in respondents with missing values for the following items: age ($p < 0.0001$), number of children ($p = 0.003$), degree of religiosity ($p = 0.007$), average number of work hours per week ($p = 0.001$), number of weeks since last vacation ($p = 0.02$), number of patients cared for during life-sustaining treatment withdrawal within the last week ($p = 0.038$), number of patients who died within the last week ($p = 0.019$) and use of antidepressant therapy by the respondent ($p = 0.043$) (Table 2). Also, the number of missing items per respondent was significantly associated with the respondent's job strain score (Fig. 1). The final multivariate model obtained for the complete-case analysis and the same models built using the two different imputation methods are shown in Table 3. Diagnostics analyses on the values imputed using multiple imputation were performed but did not reveal critical behaviour (electronic supplementary material 3 and 4).

Although the estimates obtained using the three methods seemed similar, estimates varied from 0.2 to 60 % between complete case analysis and imputed-data analyses. Similarly, SEs varied from 7.1 to 18.2 %. Variables showing the highest estimates variation were age (up to 60 % in multiple imputation and up to 39 % for median/most frequent modality imputation method), and the number of children (respectively 45 and 52 %). All estimates and SEs variation rates are shown in electronic supplementary material 5. The SEs were lowest using the

Table 2 Differences in respondent job strain scores based on the missing values

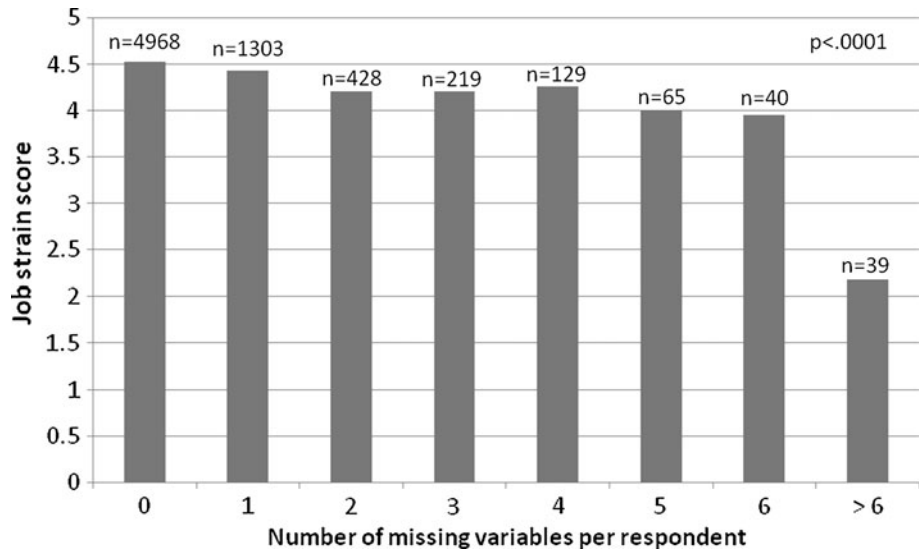
Questionnaire item	Job strain score ^a in respondents who answered the relevant item (n) mean ± SD	Job strain score ^a in respondents who failed to answer the relevant item (n) mean ± SD	p value ^b
Age	(7,012) 4.48 ± 2.5	(197) 3.25 ± 2.8	<0.0001
Sex	(7,179) 4.45 ± 2.5	(30) 3.87 ± 3.4	0.4
Single	(6,950) 4.45 ± 2.5	(259) 4.34 ± 2.8	0.5
Number of children	(6,766) 4.48 ± 2.5	(443) 4.06 ± 2.6	0.003
Born in the country of current employment	(7,147) 4.45 ± 2.5	(62) 4.00 ± 3.0	0.3
Years of experience in the ICU	(6,987) 4.45 ± 2.5	(222) 4.40 ± 2.7	0.9
Degree of religiosity	(6,919) 4.47 ± 2.5	(290) 4.02 ± 2.6	0.007
Graduated in the country of current employment	(6,954) 4.45 ± 2.5	(255) 4.37 ± 2.7	0.8
Job title	(7,049) 4.46 ± 2.5	(160) 4.05 ± 2.8	0.11
Average hours of work per week	(7,080) 4.46 ± 2.5	(129) 3.71 ± 2.6	0.001
Number of weeks since last vacation	(6,793) 4.47 ± 2.5	(416) 4.15 ± 2.6	0.02
Average hours of work per day	(7,058) 4.46 ± 2.5	(151) 4.18 ± 2.8	0.3
Number of end-of-life patients cared for within the last week	(6,845) 4.46 ± 2.5	(364) 4.28 ± 2.7	0.3
Number of treatment-limitation decisions shared within the last week	(6,602) 4.47 ± 2.5	(607) 4.22 ± 2.6	0.038
Number of deaths within the last week	(6,822) 4.47 ± 2.5	(387) 4.10 ± 2.7	0.019
Receiving antidepressant therapy	(7,148) 4.46 ± 2.5	(61) 3.69 ± 2.8	0.043

There were no missing values for the items “formal training in ethics or communication” and “number of conflicts”

^a A low score reflects a high level of stress

^b Wilcoxon non-parametric test

Fig. 1 Histogram of the number of missing variables per respondent according to the job strain score



median/most frequent modality imputation method, in keeping with the fact that this method artificially decreases the variability within a variable by using the same value to replace all missing data. SEs were largest with the complete-case model because of the smaller

respondent sample (24 % fewer respondents). Most of the multiple imputation SEs were intermediate between those of the other two methods, as the number of observations used was 100 % and the uncertainty of imputations was taken into account.

Table 3 Risk factors for job strain in the Conflicus study ICU workers

Variable	Complete-case analysis		Median/most frequent modality imputation		Multiple imputation	
	<i>n</i> = 5,504 (76.3 %)		<i>n</i> = 7,209 (100 %)		<i>n</i> = 7,209 (100 %)	
	Beta (SE)	<i>p</i> value	Beta (SE)	<i>p</i> value	Beta (SE)	<i>p</i> value
Age in years						
<28	0 (-)		0 (-)		0 (-)	
28–34	-0.349 (0.095)	0.0003	-0.245 (0.083)	0.0033	-0.216 (0.085)	0.0107
34–42	-0.195 (0.113)	0.0849	-0.180 (0.097)	0.0632	-0.086 (0.101)	0.3936
≥42	-0.200 (0.132)	0.1295	-0.123 (0.112)	0.2712	-0.080 (0.117)	0.4924
Female	-0.130 (0.072)	0.0709	-0.104 (0.064)	0.1031	-0.100 (0.065)	0.1204
Number of children						
0	0 (-)		0 (-)		0 (-)	
1	-0.050 (0.088)	0.5708	-0.024 (0.072)	0.7394	-0.065 (0.078)	0.4085
≥2	0.212 (0.084)	0.0120	0.161 (0.075)	0.0327	0.117 (0.076)	0.1248
Born in the country of current employment	-0.495 (0.103)	0.0000	-0.352 (0.089)	0.0000	-0.373 (0.089)	0.0000
Years of experience in the ICU						
<2	0 (-)		0 (-)		0 (-)	
2–5	-0.531 (0.090)	0.0000	-0.570 (0.080)	0.0000	-0.561 (0.080)	0.0000
6–11	-0.471 (0.106)	0.0000	-0.530 (0.090)	0.0000	-0.550 (0.093)	0.0000
≥12	-0.464 (0.118)	0.0000	-0.462 (0.102)	0.0000	-0.480 (0.104)	0.0000
Job title						
Physician	0 (-)		0 (-)		0 (-)	
Nurse	-1.041 (0.088)	0.0000	-1.039 (0.079)	0.0000	-1.046 (0.079)	0.0000
Physiotherapist	-0.099 (0.151)	0.5097	-0.106 (0.137)	0.4393	-0.122 (0.137)	0.3717
Other	-0.541 (0.144)	0.0002	-0.683 (0.125)	0.0000	-0.701 (0.126)	0.0000
>4 months since last vacations	0.247 (0.071)	0.0005	0.222 (0.064)	0.0005	0.222 (0.064)	0.0005
Average hours of work >8/day	0.285 (0.071)	0.0000	0.272 (0.063)	0.0000	0.288 (0.064)	0.0000
Formal training in ethics or communication	-0.359 (0.062)	0.0000	-0.350 (0.055)	0.0000	-0.355 (0.055)	0.0000
End-of-life patients cared for within the last week						
0	0 (-)		0 (-)		0 (-)	
1	-0.270 (0.085)	0.0015	-0.228 (0.072)	0.0015	-0.261 (0.079)	0.0009
>1	-0.574 (0.081)	0.0000	-0.521 (0.073)	0.0000	-0.551 (0.074)	0.0000
At least one treatment-limitation decision shared within the last week	-0.357 (0.074)	0.0000	-0.353 (0.067)	0.0000	-0.380 (0.067)	0.0000
Receiving antidepressant therapy	0.760 (0.109)	0.0000	0.818 (0.098)	0.0000	0.815 (0.098)	0.0000
At least one conflict in the ICU	-0.504 (0.072)	0.0000	-0.541 (0.063)	0.0000	-0.537 (0.062)	0.0000

High job strain scores means a lower stress; negative estimates suggest greater stress. For multiple imputations, sequential regression multiple imputation (SRMI) was used; the imputation models varied with the type of variable to be imputed; 200 imputation data sets were created and variable-specific bounds were defined for

imputation values (see electronic supplementary material 2 for details on parameterisation and SAS code sample). The variation of both imputation methods estimates and standard errors are shown in electronic supplementary material 5

Discussion

A cross-sectional analysis of a representative subset of the ICU literature at one specific time point showed that more than one-third of the articles did not mention data completeness and that missing data were noticed in almost half of these studies. Also, missing data were handled very rarely. In our illustrative case, discarding incomplete observation would have led to the exclusion of 24 % of the responders, who were also the ones observing stronger job strain. Then, we did observe changes in parameter estimates and SEs (thus in results interpretation) after using missing data imputation. Finally, we proposed an algorithm to help the clinician through the statistical analysis in the event of missing data.

Our analysis had some limitations. First, the review of critical care literature was based on three major critical care journals but consequently was not exhaustive of the overall critical care literature. Considering the statistical analysis, the assumption of MAR missingness may be unrealistic in our data. Also, we discarded 7 % of persons who completed neither job strain items nor other questions, and who were probably very different from the 7,209 respondents of our analysis sample. Another limitation was that the multiple testing issue was not addressed in our article. Although this can (as missing data) lead to erroneous conclusions, this matter was beyond of the scope of our objectives. We proposed the use of multiple imputation using SRMI because it was adapted to the survey design of our data. However, this method may not

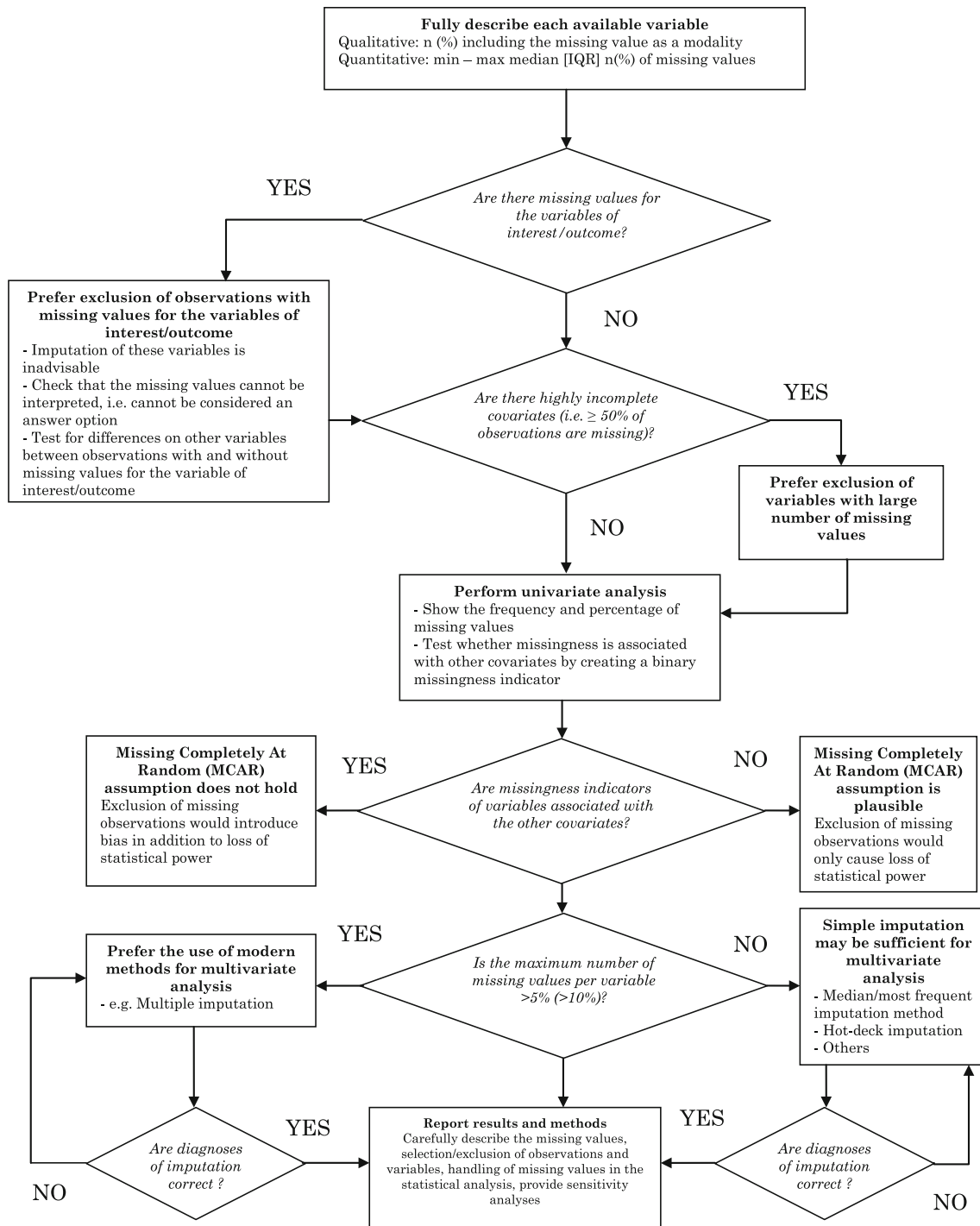


Fig. 2 Suggested algorithm for analysing data sets with missing values

be adapted for other specific study designs in clinical research, for which other suitable methods must be preferred. Finally, despite all efforts that can be provided to ensure appropriate missing data completion, their accuracy will never reach that of the true data, highlighting the

fact that one should prioritize efforts on data collection rather than on post data completion.

We could not find other estimations of the frequency of missing data in clinical studies in the literature. Our partial literature review was sufficient to unmask the

existence of this issue in many cases. Although awareness among clinicians of the impact of missing data has improved in recent years, little practical information is available about how to report and handle missing data or about the impact of inadequate reporting and handling on the validity of the data. In recent years, guidelines were developed to help improve the reporting of missing data in health research (see Equator Network website for reporting guidelines), and to address the missing value issue in RCTs [10, 11]. There is no unanimous agreement on missing data handling strategy. Thus, we proposed an algorithm based on our experience for analysing data sets with missing values (Fig. 2).

From our analysis we saw that discarding observations with missing values raises two major problems: the induction of bias (leading to erroneous results) and loss of statistical power (leading to loss of precision). As shown by our analysis of the Conflicus study, job strain was greater in respondents who supplied incomplete data and increased with the number of unanswered items. Thus, confining the analysis to complete cases would not only diminish statistical power by eliminating 24 % of the population, but also bias the results by producing spuriously low job strain levels. More specifically, on the basis of a strict interpretation of the 5 % significance threshold, ICU workers having at least two children would have significantly increased their job strain score of 0.21 points, whereas multiple imputation analysis revealed a non-significant increase of 0.11 points (50 % less). Another important issue when considering biases introduced by missing data in a clinical study such as Conflicus is the possible existence of specific interactions between questionnaires with missing data (i.e. respondents giving no replies for a given variable) and the outcome variable (i.e. the job strain level). Several levels of interaction may occur. First, it is likely that the characteristics of respondents providing no information on their personal life (having children, being happy or living alone) may differ between respondents with high versus low job strain, leading to bias for the identification of job strain risk factors based on NMAR data. Second, reasons for not replying to a question may be directly related to the outcome variable. Job strain may prevent clinicians

from building a relationship with a partner and from having children; in this case, not having children is a consequence and not a cause of high job strain.

Many methods exist to cope with missing data; however, the simplest may result in unrealistic interpretations and the most sophisticated may appear daunting to use.

Complete-case analysis (or listwise deletion) is the most widely used method and consists in deleting observations that have missing values. Complete-case analysis is the default setting of standard statistical software. It produces unbiased estimates only with MCAR data, as the analysed subpopulation remains representative in this situation. Nevertheless, even with MCAR data, the loss of power induced by the deletion of observations may be problematic if the original data set is not large enough. The use of simple imputation methods (median or most frequent value, hot deck imputation) artificially reduces data variability and thus artificially increases confidence in the results obtained. These methods should be avoided.

We encourage clinicians to use multiple imputation using SRMI (or chained equation) for several reasons : first, the overall principles of multiple imputation and the SRMI are understandable; second, this approach uses well-known models (linear, Poisson, logistic regressions models); third, the fitting of the imputation models is based on common guidelines of regression modelling, as is the post-diagnostics of the fitted model; fourth, this approach allows one to specify bounds or to perform conditional imputation (number of cigarettes smoked imputed only for 'ever smokers'). Additional information and an introduction to different alternative methods are available in electronic supplementary material 1.

In conclusion, the existence of missing data is a common problem in medical studies. Ignoring or not handling this problem is the most often encountered behaviour in the clinical literature but amounts to concealing part of the results. This mainly biases study results and reduces statistical power that might result in erroneous study conclusions. We strongly encourage clinicians to report missing data and to use appropriate statistical techniques for handling missing data. To this purpose we provided a decision algorithm to handle missing data in a clinical study analysis.

References

1. Ibrahim JG, Chu H, Chen MH (2012) Missing data in clinical studies: issues and methods. *J Clin Oncol* 30:3297–3303
2. King G, Honaker J, Joseph A, Scheve K (2001) Analyzing incomplete political science data: an alternative algorithm for multiple imputation. *Am Political Sci Rev* 95:49–69
3. Molnar FJ, Man-Son-Hing M, Hutton B, Fergusson DA (2009) Have last-observation-carried-forward analyses caused us to favour more toxic dementia therapies over less toxic alternatives? A systematic review. *Open Med* 3:e31–e50
4. Rubin DB (1976) Inference and missing data. *Biometrika* 63:581–592
5. Touloumi G, Babiker AG, Pocock SJ, Darbyshire JH (2001) Impact of missing data due to drop-outs on estimators for rates of change in longitudinal studies: a simulation study. *Stat Med* 20:3715–3728

-
6. Rubin DB (1987) Multiple imputation for nonresponse in surveys. Wiley classics library, vol 81. Wiley, New York
 7. Azoulay E, Timsit J-F, Sprung CL, Soares M, Rusinová K, Lafabrie A, Abizanda R, Svantesson M, Rubulotta F, Ricou B, Benoit D, Heyland D, Joynt G, Français A, Azevedo-Maia P, Owczuk R, Benbenishty J, de Vita M, Valentin A, Ksomos A, Cohen S, Kompan L, Ho K, Abroug F, Kaarlola A, Gerlach H, Kyprianou T, Michalsen A, Chevret S, Schlemmer B (2009) Prevalence and factors of intensive care unit conflicts: the Conflicus study. *Am J Respir Crit Care Med* 180:853–860
 8. Karasek R, Baker D, Marxer F, Ahlbom A, Theorell T (1981) Job decision latitude, job demands, and cardiovascular disease: a prospective study of Swedish men. *Am J Public Health* 71:694
 9. Abayomi K, Gelman A, Levy M (2008) Diagnostics for multivariate imputations. *J R Stat Soc Ser C* 57:273–291
 10. Altman DG (2009) Missing outcomes in randomized trials: addressing the dilemma. *Open Med* 3:e51
 11. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Frangakis C, Hogan JW, Molenberghs G, Murphy SA, Neaton JD, Rotnitzky A, Scharfstein D, Shih WJ, Siegel JP, Stern H (2012) The prevention and treatment of missing data in clinical trials. *N Engl J Med* 367:1355–1360