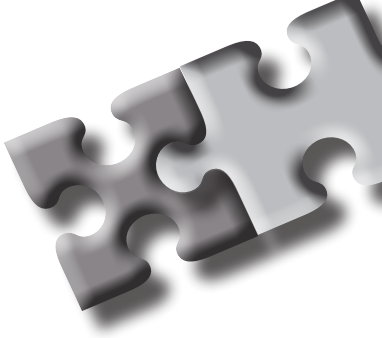

Chapter 2

Missing data: a systematic review of how they are reported and handled

Published: Eekhout, I., de Boer, M.R., Twisk, J.W.R., de Vet, H.C.W., & Heymans, M.W. (2012). Missing data: a systematic review of how they are reported and handled. Epidemiology, 23(5), 729-732.

Abstract

The objectives of this systematic review are to examine how researchers report missing data in questionnaires and to provide an overview of current methods for dealing with missing data. We included 262 studies published in 2010 in three leading epidemiological journals. Information was extracted on how missing data were reported, types of missing, and methods for dealing with missing data. 78% of studies lacked clear information about the measurement instruments. Missing data in multi-item instruments were not handled differently from other missing data. Complete-case analysis was most frequently reported (81% of the studies), and the selectivity of missing data was seldom examined. Although there are specific methods for handling missing data in item scores and in total scores of multi-item instruments, these are seldom applied. Researchers mainly use complete-case analysis for both types of missing data, which may seriously bias the study results.

Keywords: bias (epidemiology), data interpretation, statistical, epidemiological research design, missing data, questionnaires, regression analysis, research report

Introduction

Missing data are a problem in most epidemiological studies, especially with questionnaires containing multi-item instruments. Multi-item instruments measure one construct with multiple items (de Vet, Terwee, Mokkink, & Knol, 2011); for example, the CES-D uses 20 items to assess depressive symptoms (Radloff, 1977). On multi-item instruments, several or all item scores can be missing. Single-item instruments assess constructs by one question, for example pain by a visual analog scale. Missing cases, when eligible subjects do not fill out or return the questionnaire, can also occur (McKnight, McKnight, Sidani, & Figueredo, 2007).

Missing total scores on multi-item instruments are equivalent to missing scores on a single-item instrument. Commonly used methods to deal with such missingness are complete-case analysis, mean imputation, or single regression imputation. More advanced techniques that account for missing data uncertainty are multiple imputation or maximum likelihood estimation (Baraldi & Enders, 2010; DeSouza, Legedza, & Sankoh, 2009; Donders, van der Heijden, Stijnen, & Moons, 2006; Enders, 2010; Greenland & Finkle, 1995; McKnight et al., 2007). Specific methods have also been developed for missing-item scores in multi-item instruments, for example, person mean imputation, two-way imputation, response-function imputation, and multivariate normal imputation (Bernaards et al., 2003; Sijtsma & van der Ark, 2003; van Ginkel, Van der Ark, & Sijtsma, 2007).

Several reviews of missing data methods in medical and epidemiological studies have observed that complete-case analyses and single-imputation techniques are the most frequently used (Burton & Altman, 2004; Wood, White, & Thompson, 2004). These reviews have not distinguished between missing data in single-item and multi-item instruments and, for the latter, between missing several or all items. Previous reviews were published at least five years ago, and it might be expected that missing-data methods have improved. We review how recent epidemiological reports have handed missing data, and whether distinctions are made between types of missing data. We also provide an overview of current methods to handle various types of missing data.

Methods

The 2010 volumes of the American Journal of Epidemiology, Epidemiology, and International Journal of Epidemiology (Impact Factor > 5.0 (Reuters, 2009)) were searched by one researcher (IE) (846 articles). We selected articles in which studies used questionnaires to assess the predictors, covariates, or outcomes. 285 studies fulfilled the inclusion criteria (online Appendix, http://links.lww.com). In 4 studies the authors explicitly reported that no data were missing and 19 studies contained no

information on missing data, leaving 262 studies for analysis (Figure 2.1).

Information was extracted using an inventory list containing 28 items (Appendix 2.1). The list was based on the guidelines provided by Sterne et al. (2009) and the STROBE statement (Von Elm et al., 2008). The list assessed information on number and type of missing data, and the methods used to handle missing data. Items could be answered by "yes", "no", "unclear", "not applicable" or "no information."

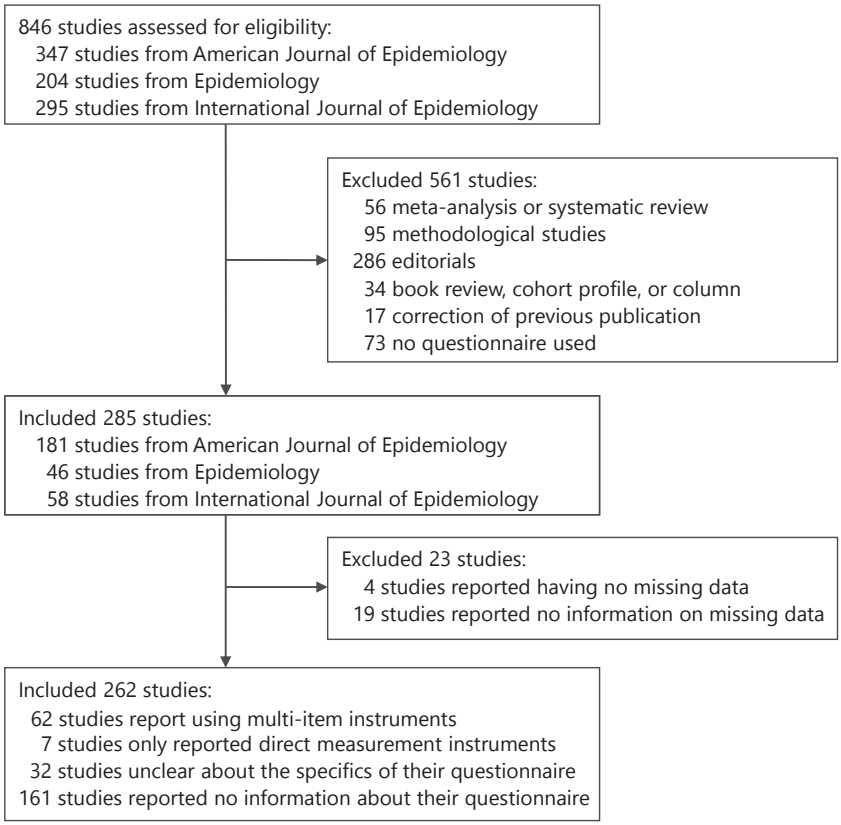

Figure 2.1. Study selection process of studies using questionnaires in all publications in 2010 in the American Journal of Epidemiology, Epidemiology, and the International Journal of Epidemiology

One reviewer (IE) evaluated all studies. 50 studies were randomly selected for independent assessment by another rater (MWH). Agreement was scored for the selection of studies and the items on the inventory list. Overall agreement was 75%, with discrepancies settled by consensus.

Results

In 262 studies having missing data, the type of missing data could not be clearly defined in 46% (Table 2.1). Missing data were most frequently reported for total scores (76%). Among the types of missing data, the average percent missing data was highest for missing cases, although with a wide range for all types. Results were similar for the subgroup of 62 studies that used a multi-item instrument, except that the percent that reported missing item scores was greater (19%) by definition; only multi-item instruments can have missing item scores.

Table 2.1.
Number and percentage of various types of missing data

Type of missing data	All studies (n=262)[a] %	Average percent missing (range)
Item scores	5	11(1 – 44)
Total scores	76	15 (<1 – 68)
Cases	32	26 (<1 – 82)
Unclear	46	13 (<1 – 62)

Note: [a] 105 studies reported multiple types of missing data.

22% of studies considered the possible selectivity of missing data (Table 2.2), presumably by examining differences in characteristics of responders and non-responders. Only 14% of studies that considered selection were specific about the assumed mechanism of missingness (e.g., missing at random).

Table 2.2.
Exploration of selectivity of missing data and reported mechanism in 262 studies that reported missing data

Methods Applied or Reported	% of Studies
Selectivity of missing data examined (n=262)	22
Method used to compare responders to non-responders (n=58)	
Differences in characteristics described in text	54
Descriptive statistics presented	22
Statistical test performed[a]	14
Unclear	10
Reported mechanism (n=58)	
MCAR	0
MAR	14
MNAR	0
Unclear	2
No Information[b]	85

Note: [a] T-test, Chi Square test or Logistic regression; [b] No information means that the assumed mechanism was not explicitly reported.

The methods used to handle missing data are presented in Table 2.3. Most studies (81%) performed a complete-case analysis. 14% used a single imputation technique (e.g., mean imputation, single regression imputation, last observation carried forward, etc.). Multiple imputation, full information maximum likelihood estimation and inverse probability weighting, which assume that data are missing at random, were reported in 8%, 2%, and 3% of the studies, respectively. Results were similar in studies with a multi-item instrument. Just 11% of all studies performed sensitivity analysis to investigate the influence of the handling of missing data on the study results.

Table 2.3.
Reported methods to handle missing data

	All studies (n=262)[a]	Multi-item instrument studies (n=62)[b]
	%	%
Complete-case analysis	81	79
Single imputation techniques	14	19
Including missing indicator as extra answer category	6	7
Inverse probability weighting	3	2
Multiple imputation	8	5
Full information maximum likelihood estimation	2	2
Unclear	13	13
No Information	2	3

Note: [a] 67 studies reported multiple methods to handle their missing data. [b] 16 studies reported multiple methods to handle their missing data.

Discussion

In a review of recent papers published in the three leading epidemiological journals, the routine approach to missing data was to include analysis of complete cases only. For many studies it was unclear whether missing data were from a single-item or multi-item instrument. In studies that use multi-item instruments, researchers generally did not pay attention to the different types of missing data and their corresponding handling methods. Methods designed to handle missing item scores in multi-item instruments, have the advantage that the total score of the construct can be estimated from other items within the same scale. When scores of single-item instruments or total scores are missing, information from other scales and variables is needed — which are usually less efficient. A review by van Ginkel, et al. (2010) found similar lack of distinction between methods to handle missing item scores in multi-item instruments and missing total scores. A broader appreciation of this difference might lead to the application of more valid missing data methods for

multi-item instruments.

More generally, complete-case analysis and single imputation techniques can bias study results, depending on the underlying mechanism of missingness (Huisman, 1999; Little & Rubin, 2002). Knowledge about the selectivity of missing data and the corresponding mechanism forms an important starting point to effectively handle missing data. A proper approach to missing data depends on whether the data are missing at random (MAR) or missing not at random (MNAR). Where data are missing at random (or can be assumed to be MAR), observed data can be used to estimate the missing values. When data are MNAR, such estimation is not possible. A third mechanism assumes that subjects with missing data are a random subset of the whole study sample and therefore even less prone to bias (MCAR) (Rubin, 1976). In the reviewed studies the average mean proportion of missing data was larger than 10%, which might lead to potential biased results. Even when the MCAR assumption holds, loss of power may cause unreliable estimates (Enders, 2010). Furthermore, for both MAR and MNAR data, complete-case analysis and single imputation methods can result in incorrect parameter estimates (Donders et al., 2006; Enders, 2010; Enders & Bandalos, 2001; Graham, 2009; Huisman, 2000; Roth, 1994). Recommended methods such as multiple imputation, full information maximum likelihood estimation, and inverse probability weighting were used in only 13% of studies. These techniques have been shown to work well when MCAR and MAR assumptions hold (Baraldi & Enders, 2010; van Buuren, 2010). In general, studies should report on how many cases their inferences are based. Where a substantial proportion of data are missing, (selective) missings can lead to biased results.

We used publications of three leading epidemiological journals to represent current epidemiological research. We expect that the practices described in these papers are at least as good as actual practice for the field as a whole. The majority of the studies were rated by only one rater, which is presumably less valid than if several raters had sought consensus. Also, our results are based only on the reported information in the studies. This might underestimate the actual extent of missing data. The lack of clarity in many studies over whether a multi-item instrument had been used makes it impossible to assess whether optimal methods for dealing with missing data in multi-item questions were applied.

The reporting of missing data in epidemiological studies is highly variable and mostly poor. Most epidemiologists do not distinguish between missing item scores and missing total scores in multi-item instruments, either in reporting their missing data or in the application of missing data methods. Many researchers may not be aware of the impact of the different types of missing data (i.e., item or total scores) on their study results.

References

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. Journal of School Psycholology, 48(1), 5-37.

Bernaards, C. A., Farmer, M. M., Qi, K., Dulai, G. S., Ganz, P. A., & Kahn, K. L. (2003). Comparison of Two Multiple Imputation Procedures in a Cancer Screening Survey. Journal of Data Science, 1(3), 19.

Burton, A., & Altman, D. G. (2004). Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines. British Journal of Cancer, 91(1), 4-8.

van Buuren, S. (2010). Item Imputation Without Specifying Scale Structure. Methodology, 6(1), 31-36.

DeSouza, C. M., Legedza, A. T., & Sankoh, A. J. (2009). An overview of practical approaches for handling missing data in clinical trials. Journal of Biopharmaceutical Statistics, 19(6), 1055-1073.

Donders, A. R., van der Heijden, G. J., Stijnen, T., & Moons, K. G. (2006). Review: a gentle introduction to imputation of missing values. Journal of Clinical Epidemiology, 59(10), 1087-1091.

Enders, C. K. (2010). Applied Missing Data Analysis. New York, NY: The Guilford Press.

Enders, C. K., & Bandalos, D. L. (2001). The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. Structural Equation Modeling: A Multidisciplinary Journal, 8(3), 430-457.

van Ginkel, J. R., Sijtsma, K., van der Ark, L. A., & Vermunt, J. K. (2010). Incidence of missing item scores in personality measurement, and simple item-score imputation. Methodology, 6(1), 17-30.

van Ginkel, J. R., Van der Ark, L. A., & Sijtsma, K. (2007). Multiple Imputation of Item Scores in Test and Questionnaire Data, and Influence on Psychometric Results. Multivariate Behavioral Research, 42(2), 387-414.

Graham, J. W. (2009). Missing data analysis: making it work in the real world. Annual Review of Psychology, 60, 549-576.

Greenland, S., & Finkle, W. D. (1995). A critical look at methods for handling missing covariates in epidemiological regression analyses. American Journal of Epidemiology, 142(12), 1255-1264.

Huisman, M. (1999). Item Nonresponse: Occurence, Causes, and Imputation of Missing Answers to Test Items. Leiden: DSWO Press.

Huisman, M. (2000). Imputation of Missing Item Responses: Some Simple Techniques. Quality & Quantity, 34(4), 331-351.

Little, R. J., & Rubin, D. B. (2002). Statistical Analysis with Missing Data (Second Edition ed.). Hoboken, NJ: John Wiley & Sons.

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). Missing Data A Gentle Introduction. New York: The Guilford Press.

Radloff, L. S. (1977). The CES-D Scale. Applied Psychological Measurement, 1(3), 385-401.

Reuters, T. (2009). Journal Citation Reports (Science Edition ed.). New York: ISI Web of Knowledge.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. Personnel Psychology, 47(3), 537-560.

Rubin, D. B. (1976). Inference and missing data. Biometrika, 63(3), 581-592.

Sijtsma, K., & van der Ark, L. A. (2003). Investigation and treatment of missing scores in test and questionnaire data. Multivariate Behavioral Research, 38(4), 505-528.

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., . . . Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ, 338, b2393.

de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). Measurement in Medicine. Cambridge: Cambridge University Press.

Von Elm, E., Altman, D. G., Egger, M., Pocock, S., Gotzsche, P. C., Vandenbroucke, J. P., & Initiative, S. (2008). The strengthening the reporting of observational studies in epidemiology (STROBE) statement. Journal of Clinical Epidemiology, 61, 344-349.

Wood, A. M., White, I. R., & Thompson, S. G. (2004). Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Clinical Trials, 1(4), 368-376.

Appendix 2.1|Inventory list

First Author:.....
Pages: .....

Study design
1. Type of design:
☐ Treatment studies
☐ RCT
☐ Non-randomized trial (quasi-experiment)
☐ Observational studies
☐ Cohort study
☐ Prospective cohort
☐ Retrospective cohort
☐ Case-control study
☐ Cross-sectional study

2. Researched population
☐ Patients
☐ Healthy individuals

3. If patients: type of Patients included:.....

4. Number of participants:.....

5. Principal analysis used in the study:
☐ Linear/logistic/Cox/Poisson regression
☐ Mixed models
☐ Generalized Estimating Equations (GEE)
☐ T-tests
☐ ANOVA / MANOVA / Repeated measures ANOVA
☐ Item response theory (IRT)
☐ Chi-square test
☐ Nonparametric test
☐ Other:.....

6a. A Questionnaire was used for the assessment of:
☐ Covariates/Predictor
☐ Outcome
☐ Both
☐ None

6b. Did the questionnaire consist of different items resulting in a total score (or total scores per dimension/subscale)?
☐ Yes
☐ No
☐ Unclear
☐ No information
☐ Not applicable

6c. Is more information available on the questionnaire (e.g., response/answer categories, score calculation, etc.)?
☐ Yes
☐ No

Missing data information:
7a. Is the percentage/number of missing data described?
☐ Yes
☐ No
☐ Unclear

7b. Is the location of missing data described?
☐ Yes
☐ No
☐ Unclear
☐ Not applicable

7c. What is the location of missing data presented?
☐ Missing in variables
☐ Missing total scores (also attrition)
☐ Missing items
☐ Missing cases (unit nonresponse: did not show up/return questionnaire)
☐ Planned missingness (e.g.. missing by design)
☐ Other:.....
☐ Unclear
☐ Not applicable

7d. What type of missing data is reported?
☐ Nonresponse (item/variable/unit)
☐ Dropout
☐ Attrition
☐ Lost to follow up
☐ Intermittent
☐ Other:.....
☐ Unclear
☐ Not applicable
☐ No Information

7e. What is the percentage of missings in the data reported?
Total score:.....
Item: .....
Cases:.....

8. Is the fraction of missing information presented?
☐ Yes
☐ No
☐ Unclear

9. Are the potential reasons for missing data discussed (e.g.. exhaustion, deceased, lack of motivation, lost in mail, etc.)?
☐ Yes
☐ No
☐ Unclear

10a. Was the missing data mechanism evaluated?
☐ Yes
☐ No
☐ Unclear

10b. Which method was used to test the mechanism?
☐ Differences in characteristics between missing and non-missing group described
☐ Analysis between cases with complete and missing data:
☐ Descriptive statistics (e.g., comparing means / percentages)
☐ Chi-square tests
☐ T-tests
☐ Univariate t-Test comparisons
☐ Little's MCAR test
☐ Logistic regression analysis with missing data as outcome
☐ Other:.....
☐ Unclear
☐ Not applicable

10c. What category of missing data mechanism is reported?
☐ MCAR
☐ MAR
☐ MNAR
☐ Other:.....
☐ Unclear
☐ No information
☐ Not applicable

Methods used to handle the missing data:
11. Handling method
☐ Missing total score/unit score methods
☐ Complete-case analysis
☐ Pairwise deletion
☐ Mean substitution/arithmetic mean imputation/unconditional mean imputation/median imputation
☐ Single regression imputation (e.g.. Stochastic)
☐ Hot-deck imputation – matching nonrespondents to resembling respondent
☐ Last value carried forward
☐ Multiple imputation
☐ Full Information Maximum Likelihood Estimation
☐ Missing item score methods
☐ Unconditional random imputation
☐ Item mean substitution/person mean substitution
☐ Corrected item mean substitution
☐ Two-way imputation
☐ Response-function imputation
☐ Multivariate normal imputation
☐ Fully conditional specification
☐ Similar Response pattern imputation
☐ Item correlation substitution
☐ Multiple response-function imputation
☐ Including a missing category
☐ Unclear
☐ No information

☐ Other.....
☐ Not Applicable

If multiple imputation is used (only 12a, 12b and 12c):
12a. Are the number of variables used in the imputation model clearly described?
☐ No
☐ No but normality discussed
☐ Yes
☐ Not applicable

12b. Is the number of multiple imputations presented?
☐ Yes
☐ No
☐ Unclear
☐ Not applicable

12c. Was the imputation process evaluated (i.e., convergence studied, imputed values compared with observed values, etc)?
☐ Yes
☐ No
☐ Unclear
☐ Not applicable

13. Is it described how non-normal / categorical variables were dealt with in the missing data method?
☐ Yes
☐ No
☐ Unclear
☐ Not applicable

14a.Was a sensitivity analysis performed to investigate the influence of how the missing data were handled on the study results?
☐ Yes
☐ No
☐ Unclear
☐ Not applicable

14b. What kind of sensitivity analysis was performed?
☐ Complete case analysis versus imputation method
☐ Different imputation techniques were compared
☐ Other:.....
☐ Not applicable

14c. Are the results of the sensitivity analysis clearly described?
☐ Yes
☐ No
☐ Unclear
☐ Not applicable

15. What Software package was used for the missing data method?
☐ SPSS
☐ AMOS – Structural equation modeling tool in SPSS
☐ EQS – Structural equation modeling software
☐ HLM – Hierarchical data modeling
☐ LISREL – Structural equation modeling software

- ☐ Mplus – Statistical modeling program
- ☐ SAS
- ☐ SOLAS for missing data analysis
- ☐ Stata
- ☐ EMCOV
- ☐ S-Plus
- ☐ R
- ☐ S-Plus en R packages
- ☐ Amelia - Amelia II: A Program for Missing Data
- ☐ NORM - analysis of multivariate normal datasets with missing values
- ☐ CAT - Analysis of categorical-variable datasets with missing values
- ☐ MICE - Multivariate imputation by chained equations
- ☐ MI - Missing Data Imputation and Model Checking
- ☐ MIX - Multiple imputation for Mixed Categorical and Continuous Data
- ☐ missMDA - Handling missing values with/in multivariate data analysis (principal component methods)
- ☐ mitools - Tools for multiple imputation of missing data
- ☐ mlmmm - ML estimation under multivariate linear mixed models with missing values
- ☐ mvnmle - ML estimation for multivariate normal data with missing values
- ☐ PAN - Multiple imputation for multivariate panel or clustered data
- ☐ MIXED
- ☐ No information
- ☐ Unclear
- ☐ Not applicable

16. What software was used for the primary/general analyses?

SPSS Version .....
SAS Version .....
Stata Version .....
Statistica Version .....
R Version.....
Mplus Version.....
S-plus Version.....
EQS Version .....
LISREL Version .....
SUDAAN Version .....
Other:.....

- ☐ No Information