NTNU
Norwegian University of Science and Technology

Multippel imputering

Bergen, 4 oktober 2018

Stian Lydersen
Professor i medisinsk statistikk,
Regionalt kunnskapssenter for barn og unge - psykisk helse og barnevern (RKBU Midt-Norge), NTNU

Versjon 5 oktober 2018

1

Innhold:

1. Introduksjon: Hva er manglende data?

Mekanismer for manglende data: MCAR, MAR, MNAR
Multippel imputering og alternative metoder: Complete case, enkel imputering, multippel imputering (MI), mixed model, full maximum likelihood (FIML).

2. MI imputeringsmodell:

Valg av variable i imputeringsmodellen
Interaksjoner og ikke lineære effekter
Hvor mange imputerte datasett?

3. MI analysemodell:

Rubins regler
Hvilke analyser og parametre kan håndteres?

Eksempler vil bli vist i SPSS og Stata

2

1. Introduksjon

- Hva er manglende data?
- Mekanismer for manglende data: MCAR, MAR, MNAR
- Multippel imputering og alternative metoder: Complete case, enkel imputering, multippel imputering (MI), mixed model, full maximum likelihood (FIML).

3

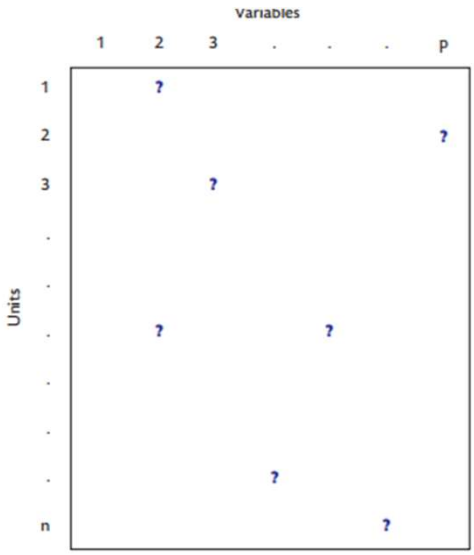
3

Missing data:

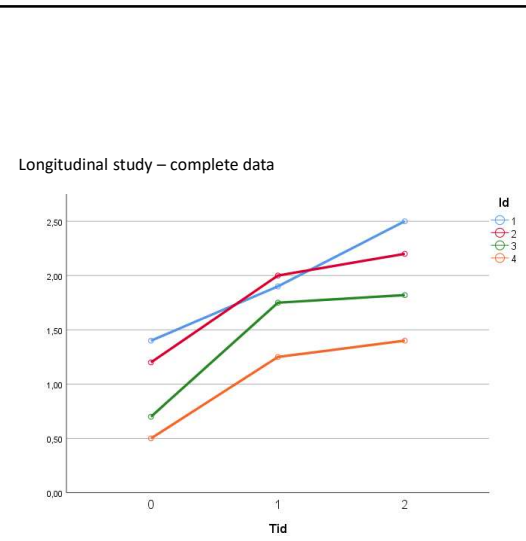
- ”Holes” in the data matrix which ideally should be complete
- Usually, these are data we intended to collect, but for some reason did not.
- There exists a meaningful value which was not recorded.

4

4

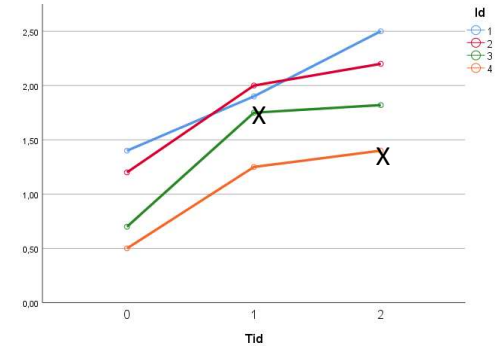

Variables
1 2 3 . . . p
Units
1
2
3
n

5


Longitudinal study – complete data
Id
1
2
3
4
2,50
2,00
1,50
1,00
0,50
0,00
0
1
2
Tid

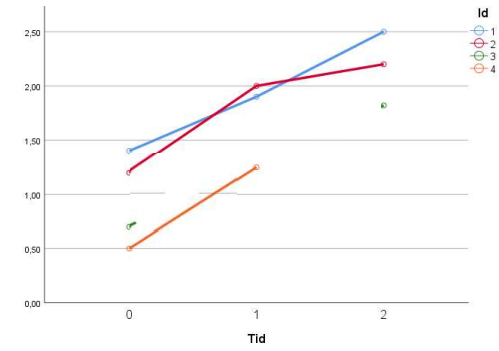
6

Longitudinal study – missing data



7

Longitudinal study – missing data



8

Missing data mechanism

Let R denote what is missing, for example 0 (1) if the corresponding value is observed (missing).

The probability distribution of R has been called

- Missing data mechanism
- Probability of nonresponse
- response mechanism
- missingness mechanism
- probability of missingness
- distribution of missingness

9

Types of missing data (Missing data mechanism)	The probability that a data value is missing (unobserved) can depend on
MCAR Missing Completely at Random	Neither observed or unobserved values
MAR Missing at Random (Ignorable nonresponse)	Only observed values
MNAR Missing Not at Random (Nonignorable nonresponse)	Unobserved values (and observed values)

10

Types of missing data (Sterne et al. 2009)

- Missing completely at random—There are no systematic differences between the missing values and the observed values. For example, blood pressure measurements may be missing because of breakdown of an automatic sphygmomanometer
- Missing at random—Any systematic difference between the missing values and the observed values can be explained by differences in observed data. For example, missing blood pressure measurements may be lower than measured blood pressures but only because younger people may be more likely to have missing blood pressure measurements
- Missing not at random—Even after the observed data are taken into account, systematic differences remain between the missing values and the observed values. For example, people with high blood pressure may be more likely to miss clinic appointments because they have headaches

11

11

Missing data may be viewed as two stage sampling:

- First, there is a (potentially) complete data set.
- Second, some of the values are missing due to «sampling» by the missing data mechanism.

12

12

Y_{Com}

$$\begin{bmatrix} Y_{11} & Y_{12} & Y_{13} \\ Y_{21} & Y_{22} & Y_{23} \\ Y_{31} & Y_{32} & Y_{33} \\ Y_{41} & Y_{42} & Y_{43} \end{bmatrix}$$

R

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Y_{Obs}

$$\begin{bmatrix} Y_{11} & Y_{12} & Y_{13} \\ Y_{21} & Y_{22} & Y_{23} \\ Y_{31} & & Y_{33} \\ Y_{41} & Y_{24} & \end{bmatrix}$$

Y_{mis}

$$\begin{bmatrix} & & \\ & & \\ & Y_{32} & \\ & & Y_{43} \end{bmatrix}$$

13

Types of missing data

$Y_{com} = (Y_{obs}, Y_{mis})$

MAR: The distribution of missingness does not depend on Y_{mis}:

$P(R|Y_{com}) = P(R|Y_{obs})$

MCAR: It does not depend on Y_{obs} either:

$P(R|Y_{com}) = P(R)$

14

Plausibility and implications of MAR

- Planned missingness is usually MCAR or MAR
  - Certain sequential designs
  - Multiple questionnaire forms
- MAR may (in principle) be tested by obtaining follow-up data from non-respondents
- Else: No way to test if MAR holds:
  MAR is an unverifyable assumption
- In some situations, erroneous assuming MAR has small impact on results. Generally, assuming MAR introduces less bias than assuming MCAR. (refs in Schafer & Graham 2002)

15

15

Impute:
To fill in data values (usually missing data) with values that are thought to be sensible.

Day, S: 2007: Dictionary for clinical trials, 2nd ed, Wiley

16

Some traditional methods and some recommended methods. (Unbiased when)

- Complete case analysis, available case analysis (MCAR)
- Single imputation
  - Mean substitution (never)
  - Averaging available items on a scale (?)
  - LOCF (Last Observation Carried Forward) (never)
  - Defining «missing» as a data value (never)
  - Using logical structure in a questionnaire (can be OK)
  - Proper single imputation such as the EM (Expectation-Maximation algortithm) (MAR but underestimates uncertainty)
- Multiple Imputation (MI) (MAR)

continues on next slide …

17

17

Some traditional methods and some recommended methods (continued). (Unbiased when)

- Full model based analysis (full information maximum likelihood)
  - Mixed model (MAR)
  - Generalized Estimating Equations (GEE) (MCAR)
  - Structural equation modelling (SEM) (MAR)
- Weighting procedures (mainly in surveys) (MAR)
- Models for MNAR (MNAR if the unverifiable assumptions are correct)
  - Selection models
  - Pattern mixture models

18

18

CLINICAL EPIDEMIOLOGY www.jasn.org

Combining GFR and Albuminuria to Classify CKD Improves Prediction of ESRD

Stein I. Hallan,*† Eberhard Ritz,‡ Stian Lydersen,* Solfrid Romundstad,*† Kurt Kvenild,§ and Stephan R. Orth‖

Departments of *Cancer Research and Molecular Medicine and §Community Medicine and General Practice, Faculty of Medicine, Norwegian University of Science and Technology, and †Department of Medicine, Division of Nephrology, St. Olav University Hospital, Trondheim, Norway; and ‡Department of Medicine, Division of Nephrology, Ruperto Carola University Heidelberg, Heidelberg, and ‖Dialysis Center Bad Aibling, Bad Aibling, and Department of Internal Medicine II, University of Regensburg, Regensburg, Germany

ABSTRACT

Despite the high prevalence of chronic kidney disease (CKD), relatively few individuals with CKD progress to ESRD. A better understanding of the risk factors for progression could improve the classification system of CKD and strategies for screening. We analyzed data from 65,589 adults who participated in the Nord-Trøndelag Health (HUNT 2) Study (1995 to 1997) and found 124 patients who progressed to ESRD after 10.3 yr of follow-up. In multivariable survival analysis, estimated GFR (eGFR) and albuminuria were independently and strongly associated with progression to ESRD: Hazard ratios for eGFR 45 to 59, 30 to 44, and 15 to 29 ml/min per 1.73 m^2 were 6.7, 18.8, and 65.7, respectively ($P < 0.001$ for all), and for micro- and macroalbuminuria were 13.0 and 47.2 ($P < 0.001$ for both). Hypertension, diabetes, male gender, smoking, depression, obesity, cardiovascular disease, dyslipidemia, physical activity and education did not add predictive information. Time-dependent receiver operating characteristic analyses showed that considering both the urinary albumin/creatinine ratio and eGFR substantially improved diagnostic accuracy. Referral based on current stages 3 to 4 CKD (eGFR 15 to 59 ml/min per 1.73 m^2) would include 4.7% of the general population and identify 69.4% of all individuals progressing to ESRD. Referral based on our classification system would include 1.4% of the general population without losing predictive power (*i.e.*, it would detect 65.6% of all individuals progressing to ESRD). In conclusion, all levels of reduced eGFR should be complemented by quantification of urinary albumin to predict optimally progression to ESRD.

J Am Soc Nephrol 20: 1069–1077, 2009. doi: 10.1681/ASN.2008070730

Times Cited: 275
(from Web of Science Core Collection)

Highly Cited Paper

19

Example from:

Hallan, S. I., Ritz, E., Lydersen, S., Romundstad, S., Kvenild, K., & Orth, S. R. Combination of estimated glomerular filtration rate and albuminuria provides best prediction of kidney failure: Results of the HUNT II study, Norway. Journal of the American Society of Nephrology, 2009.

Cox proportional hazards regression with time to kidney failure (CKD stage 5) as dependent variable.

HUNT II (Helseundersøkelsen i Nord-Trøndelag), 1995-1997. Follow-up until 2007.

20

92939 persons, 20 years and older, were invited. 65589 (70.6%) responded.
124 kidney failures.
8360 were hypertensive or had diabetes mellitus. These were asked to deliver urine samples, and 88.6% did so. In addition, a random 5% sample of non-diabetic non-hypertensive subjects (n=2,861) was also asked to deliver urine samples; 75.6% did so.

Hence: For 95% of the non-diabetic non-hypertensive subjects, urine samples were
Missing at random (MAR) by design.

21

Variable	n	% missing
Follow-up time	65589	0,0
Age	65589	0,0
Male sex	65589	0,0
Low education	61369	6,4
Depression	58423	10,9
Smoking	64395	1,8
Low physical activity	57881	11,8
Diabetes mellitus	64693	1,4
CVD	64624	1,5
BMI	64306	2,0
Waist circumference	64022	2,4

22

Variable	n	% missing
Systolic BP	64708	1,3
Diastolic BP	64708	1,3
Cholesterol	65158	0,7
HDL-Cholesterol	65155	0,7
GLUCOSE	65158	0,7
Triglycerides	65158	0,7
Creatinine	65158	0,7
eGFR [1)]	65158	0,7
ACR [2)]	9703	85,2

[1)] estimated glomerular filtration rate

[2)] Albumin creatinin ratio (from urine sample)
Not requested (Missing by design): 82,8 %
Requested, but not deliverd: 2,5%

23

Comprehensive geriatric care for patients with hip fractures: a prospective, randomised, controlled trial

Anders Prestmo, Gunhild Hagen*, Olav Sletvold, Jorunn L Helbostad, Pernille Thingstad, Kristin Taraldsen, Stian Lydersen, Vidar Halsteinli, Turi Saltnes, Sarah E Lamb, Lars G Johnsen, Ingvild Saltvedt*

Summary

Background Most patients with hip fractures are characterised by older age (>70 years), frailty, and functional deterioration, and their long-term outcomes are poor with increased costs. We compared the effectiveness and cost-effectiveness of giving these patients comprehensive geriatric care in a dedicated geriatric ward versus the usual orthopaedic care.

Methods We did a prospective, single-centre, randomised, parallel-group, controlled trial. Between April 18, 2008, and Dec 30, 2010, we randomly assigned home-dwelling patients with hip-fractures aged 70 years or older who were able to walk 10 m before their fracture, to either comprehensive geriatric care or orthopaedic care in the emergency department, to achieve the required sample of 400 patients. Randomisation was achieved via a web-based, computer-generated, block method with unknown block sizes. The primary outcome, analysed by intention to treat, was mobility measured with the Short Physical Performance Battery (SPPB) 4 months after surgery for the fracture. The type of treatment was not concealed from the patients or staff delivering the care, and assessors were only partly masked to the treatment during follow-up. This trial is registered with ClinicalTrials.gov, number NCT00667914.

Findings We assessed 1077 patients for eligibility, and excluded 680, mainly for not meeting the inclusion criteria such as living in a nursing home or being aged less than 70 years. Of the remaining patients, we randomly assigned 198 to comprehensive geriatric care and 199 to orthopaedic care. At 4 months, 174 patients remained in the comprehensive geriatric care group and 170 in the orthopaedic care group; the main reason for dropout was death. Mean SPPB scores at 4 months were 5·12 (SE 0·20) for comprehensive geriatric care and 4·38 (SE 0·20) for orthopaedic care (between-group difference 0·74, 95% CI 0·18–1·30, p=0·010).

Interpretation Immediate admission of patients aged 70 years or more with a hip fracture to comprehensive geriatric care in a dedicated ward improved mobility at 4 months, compared with the usual orthopaedic care. The results suggest that the treatment of older patients with hip fractures should be organised as orthogeriatric care.

Funding Norwegian Research Council, Central Norway Regional Health Authority, St Olav Hospital Trust and Fund for Research and Innovation, Liaison Committee between Central Norway Regional Health Authority and the Norwegian University of Science and Technology, the Department of Neuroscience at the Norwegian University of Science and Technology, Foundation for Scientific and Industrial Research at the Norwegian Institute of Technology (SINTEF), and the Municipality of Trondheim.

Lancet 2015; 385: 1623–33
Published Online February 5, 2015
http://dx.doi.org/10.1016/S0140-6736(14)62409-0
See Comment page 1594

*Joint first authors
Department of Neuroscience (A Prestmo MD, O Sletvold PhD, Prof J L Helbostad PhD, P Thingstad MSc, K Taraldsen PhD, L G Johnsen PhD, I Saltvedt PhD), **Department of Public Health and General Practice** (G Hagen MPhil, V Halsteinli PhD), **and Regional Centre for Child and Youth Mental Health and Child Welfare** (Prof S Lydersen PhD), **Norwegian University of Science and Technology, Trondheim, Norway; Department of Geriatrics** (A Prestmo, O Sletvold, I Saltvedt), **Department of Clinical Services** (Prof J L Helbostad), **and Department of Orthopaedics**

Times Cited: 126
(from Web of Science Core Collection)
Highly Cited Paper

24

Example from:

Prestmo, A., Hagen, G., Sletvold, O., Helbostad, J.L., Thingstad, P., Taraldsen, K., Lydersen, S., Halsteinli, V., Saltnes, T., Lamb, S.E., Johnsen, L.G., & Saltvedt, I. "A randomised trial of comprehensive geriatric care in hip-fracture patients." *The Lancet*, In press, 2014.

Hip fracture patients > 70 years. RCT of Comprehensive Geriatric care (CGC) versus usual ortopaedic care (OC)

397 patients assessed at baseline, 1 month, 4 months and 1 year.

25

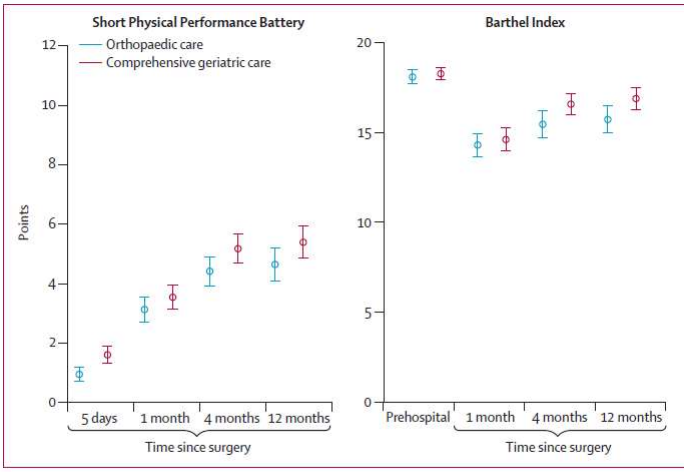


Figure 2: **Mobility, activities of daily living, and instrumental activities of daily living**
Data are mean, 95% CI. ADL=activities of daily living.

26

Missing data:

- Partially missing data at a time point:
  - Typically <1% missing.
  - Single imputation using the EM algorithm.
- No data at a time point:
  - About 15% to 30% missing.
  - Mixed model analysis.

"We used single imputation with the Expectation Maximation (EM) algorithm for imputation of single missing items on questionnaires and performance tests, using scores from the same time-point as predictors. … Linear mixed models for repeated measurements were performed with SPPB, BI, CDR, NEAS, EQ-5D-3L and MMSE as dependent variables, controlling for age, sex and femoral neck fractures."

27

27

Missing data on scales:

Barthel index:
An ordinal scale with 10 items used to measure performance in activities of daily living.

Missing data:

Time point	complete	10 missing	1 missing	2 missing	sum	proportion missing except cases with 10 missing	Complete or max 2 missing
1	365	10	19	3	397	0,00646	387
2	326	49	21	1	397	0,006609	348
3	318	64	15	0	397	0,004505	333
4	288	97	10	2	397	0,004667	300

Among cases with complete, 1 missing or 2 missing, the proportion missing is only 0.5% to 0.7%. Hence, I use single imputation with the EM algorithm on these, using the other Barthel scores from the same time point as predictors.
Some of the imputed values are slightly out of range. These are set to the range (0-1, 0-2, 0-3, respectively).

28

	Comprehensive geriatric care		Orthopaedic care		Difference	
	N	Mean (SE)	N	Mean (SE)	Estimate (95% CI)	p value
Hospital	198		199			
Mobility						
Short Performance Physical Battery	183	1·61 (0·19)	161	1·04 (0·20)	0·56 (0·20 to 1·10)	0·042
1 month	187		183			
Mobility						
Short Performance Physical Battery	173	3·59 (0·19)	160	3·09 (0·20)	0·50 (−0·05 to 1·05)	0·08
Timed Up and Go	140	31·32 (1·53)	120	32·80 (1·66)	−1·48 (−5·92 to 2·95)	0·51
Cognition						
4 months	174		170			
Mobility						
Short Physical Performance Battery	165	5·12 (0·20)	160	4·38 (0·20)	0·74 (0·18 to 1·30)	0·010
Timed Up and Go	153	24·05 (1·47)	136	25·94 (1·56)	−1·90 (−6·09 to 2·31)	0·38
Cognition						
Clinical Dementia Rating scale	159	3·59 (0·35)	145	4·38 (0·36)	−0·79 (−1·70 to 0·20)	0·12
Mini Mental Status Examination	165	23·92 (0·44)	156	22·83 (0·46)	1·10 (−0·15 to 2·34)	0·08
Activities of daily living						
Barthel index	168	16·31 (0·29)	165	15·30 (0·29)	1·01 (0·21 to 1·81)	0·013

Table 3: Clinical assessments

29

Prestmo et al (2014), Table 3.

Primary endpoint: Short Performance Physical Battery (SPPB) at 4 months.

Note that the extent of missing data is made clear by reporting *n* for each outcome at each time point.

The mixed model analysis utilized all data in the estimation, for example also for patients without SPPB data at 4 months.

30

Efficacy and Safety of Individualized Coaching After Stroke: the LAST Study (Life After Stroke)

A Pragmatic Randomized Controlled Trial

Torunn Askim, PhD; Birgitta Langhammer, PhD; Hege Ihle-Hansen, MD, PhD; Mari Gunnes, MSc; Stian Lydersen, PhD; Bent Indredavik, MD, PhD; on behalf of the LAST Collaboration Group*

Background and Purpose—The evidence for interventions to prevent functional decline in the long term after stroke is lacking. The aim of this trial was to evaluate the efficacy and safety of an 18-month follow-up program of individualized regular coaching on physical activity and exercise.

Methods—This was a multicentre, pragmatic, single-blinded, randomized controlled trial. Adults (age ≥18 years) with first-ever or recurrent stroke, community dwelling, with modified Rankin Scale <5, and no serious comorbidities were included 10 to 16 weeks poststroke. The intervention group received individualized regular coaching on physical activity and exercise every month for 18 consecutive months. The control group received standard care. Primary outcome was the Motor Assessment Scale at end of intervention (18-month follow-up). Secondary measures were Barthel index, modified Rankin Scale, item 14 from Berg Balance Scale, Timed Up and Go test, gait speed, 6-minute walk test, and Stroke Impact Scale. Other outcomes were adverse events and compliance to the intervention assessed by training diaries and the International Physical Activity Questionnaire.

Results—Three hundred and eighty consenting participants were randomly assigned to individualized coaching (n=186) or standard care (n=194). The mean estimated difference on Motor Assessment Scale in favor of control group was −0.70 points (95% confidence interval, −2.80, 1.39), P=0.512. There were no differences between the groups on Barthel index, modified Rankin Scale, or Berg Balance Scale. The frequency of adverse events was low in both groups. Results from International Physical Activity Questionnaire and training diaries showed increased activity levels but low intensity of the exercise in the intervention group.

Conclusions—The regular individualized coaching did not improve maintenance of motor function or the secondary outcomes compared with standard care. The intervention should be regarded as safe. Despite the neutral results, the health costs related to the intervention should be investigated.

Clinical Trial Registration—URL: https://www.clinicaltrials.gov. Unique identifier: NCT01467206.

(***Stroke***. **2018;49:426-432. DOI: 10.1161/STROKEAHA.117.018827.**)

Key Words: cerebrovascular disorders ■ compliance ■ exercise ■ life style ■ rehabilitation ■ secondary prevention

31

The primary end point was motor function measured by MAS at 18-month follow-up. We used ANCOVA for primary and secondary end points, with measurement at 18 months as dependent variable, and treatment group, sex, hospital site, stroke severity, age, and measurement at baseline as covariates. The Mann–Whitney U test was used for data that were not normally distributed.

We were aiming for an intention to treat analysis approach. For instrument scales with no more than half of the items missing, the missing values were singly imputed using the expectation–maximization algorithm on these. In the primary analysis, participants who had died before follow-up were imputed as zero on all scales except mRS, Timed Up and Go test and Stroke Impact Scale. We used multiple imputation to impute all other missing values, with m=100 imputations as recommended by van Buuren.[30] A sensitivity analysis was done to determine whether participants who were dead at 18 months affected the outcome.

Prespecified subgroup analyses were performed according to the stratification variables (stroke severity [mRS 0–2 versus 3–4], age <80 years, and recruitment site) in addition to sex and cognitive status (Mini-Mental State Examination <25), with a separate ANCOVA for each subgroup.

32

Complete case analysis and available case analysis

- Complete case analysis (also called case deletion or listwise deletion)
  - Only use cases with complete data on all the variables to be used.
- Available case analysis (alo called pairwise deletion or pairwise inclusion)
  - In each analysis, use as many cases as possible (with complete data for the analysis at hand)
- Default in many computer programs.
- Introduces bias unless data are MCAR.

33

33

Altman & Bland (BMJ, 2007):
" … complete case analysis: … When only a very few observations are missing little harm will be done"

Schafer J. L. 1997, "Analysis of incomplete multivariate data" Chapman & Hall, London, page 1:
"When incomplete cases comprise only a small fraction of all cases (say, five percent or less) then case deletion may be a perfectly reasonable solution to the missing-data problem."

Bjørnstad & Lydersen (2012): "However, it is problematic to set up a general rule as to what is a small fraction in this context. That depends on how much the missing data mechanism departs from MCAR."

34

Mean substitution:

- For subject missing data on a variable, fill in the mean for the subjects with data on the variable.
- NEVER OK to do this
- Note that this means averaging across subjects. Averaging *within* subjects (items on a scale) can be OK

35

35

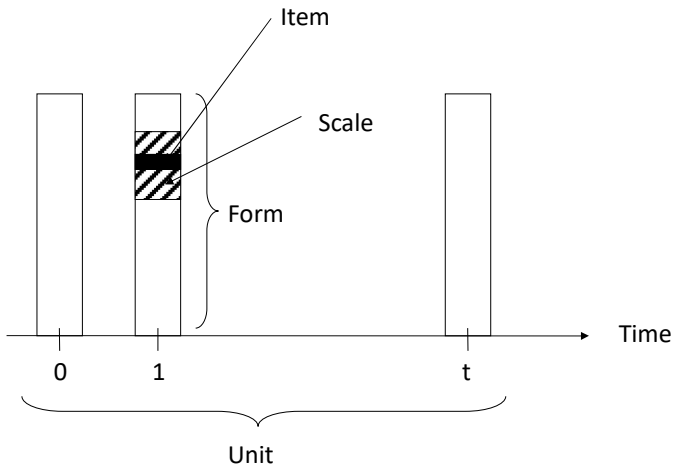
Averaging available items on a scale
Example:

- 36-Item Short Form Survey (SF-36) is a generic quality of life instrument.
- Consists of eight scales with 2 to 10 items each:
  - physical functioning
  - role limitations due to physical problems
  - bodily pain
  - general health perceptions
  - Vitality
  - Social functioning
  - role limitations due to emotional problems
  - mental health
- Recommended in the manual: On each scale, compute the average score if at least 50% of the items are available

36

36

Example: Quality of Life questionnaires



37

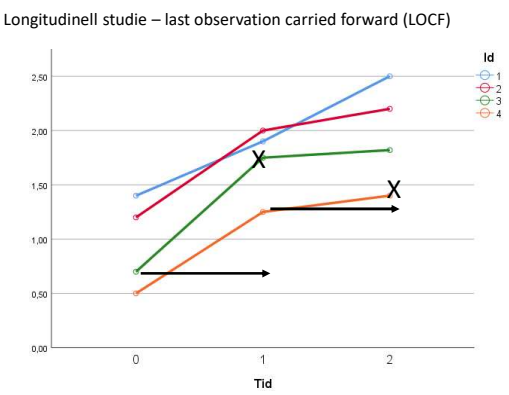
Beregning av gjennomsnitt i SPSS

Mean(q21, q22, q23, q24).
beregner hvis minst en av variablene er gitt

(q21+ q22 + q23 + q24)/4
beregner bare hvis alle variablene er gitt

Mean.2(q21, q22, q23, q24).
beregner hvis minst 2 av verdiene er gitt

38



39

Last observation carried forward (LOCF, LVCF)

As LOCF is neither valid under general assumptions nor based on statistical principles, it is not a sensible method, and should not be used. (Carpenter and Kenward 2015)

"LOCF" is an assumption that is rarely clinically plausible." (O'Kelly and Ratitch 2014)

"This method is attractive because it is simple, but it has little else to recommend it." (Vickers and Altman 2013)

40

Last observation carried forward (LOCF, LVCF)

" … LOCF is dubious. The method has long been used in clinical trials. The U.S. Food and Drug Administration (FDA) has traditionally viewed LOCF as the preferred method of analysis, considering it conservative and less prone to selection than listwise deletion. However, ((Molenberghs and Kenward 2007) pp 47 – 50) show that the bias can operate in both directions, and that LOCF can yield biased estimates even under MCAR." (van Buuren 2018)

41

Defining «missing» as a data value

For example, if smoking has the categories 0 (no) and 1 (yes), one could add an additional category 2 (missing), and regard this as three nominal categories with no missing answers. Such approaches have the potential to introduce bias and are not recommended, see Horton and Kleinman (2007) and Carpenter and Kenward (2015, page 33) and references therein.

42

42

Using logical structures in the questionnaire

Example: The HUNT 2 questionnaire includes several questions about smoking habits

"Do you smoke daily at present?"

"If you smoked earlier, how long ago did you quit smoking?"

If the first question is unanswered, and the second question is answered, one can deduce that the person does not smoke daily at present. Originally, 15% of the subjects did not answer the question about daily smoking. Assuming that the answers were internally consistent, it was possible to fill in most of the missing values, resulting in only 2% missing in daily smoking (Hallan et al 2009).

43

43

Single imputation: The EM (Expectation – Maximation) Algorithm for ML estimation

- Assume a multivariate distribution (usually normal)
- Fill in missing data with a best guess
- Estimate the parameters for the complete data set
- Re-guess missing data with the estimated parameters
- Repeat until convergence

- May need many iterations

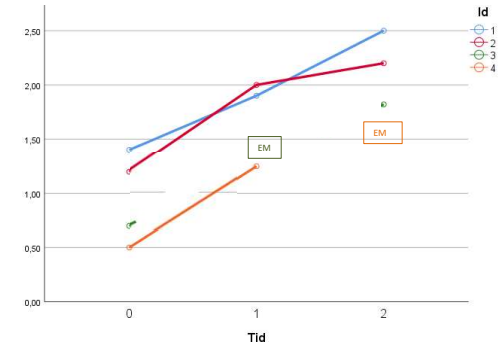
- Available in many statistical software packages

- Unbiased if MAR but underestimates uncertainty

44

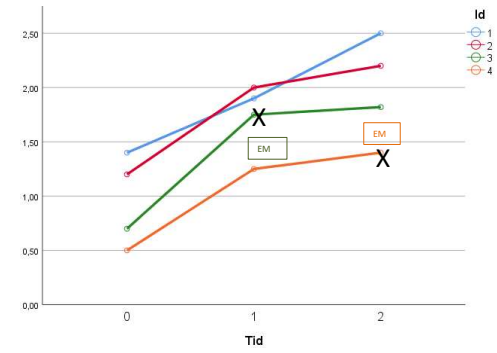
44

Single EM imputation



45

Single EM imputation



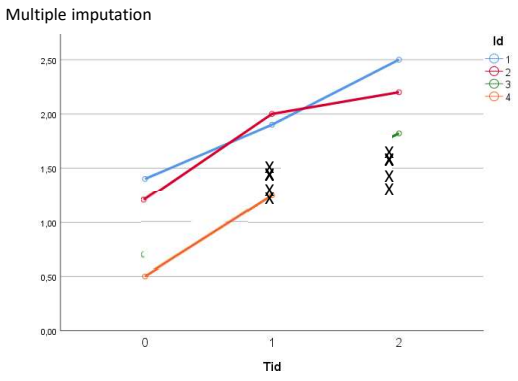
46

MI (Multiple Imputation)

- Create m > 1 (for example m=20) data sets by single imputation from the conditional distribution (Imputation model)
- Analyse each data set by a complete data method (Analysis model)
- Combine the results using simple artihmetric to obtain overall estimates reflecting missing data uncertainty and finite-sample variations.

47

47



48

MI - advantages

- Retains the attractive of single imputation from conditional distribution
- A single imputed set may be randomly atypical
- Does not underestimate uncertainty
- Unlike other Monte Carlo methods, few repetitions are needed.

49

49

Rubin's (1987) rules for combining estimates and variances

Q = the population quantity of interest, $U = Var(\hat{Q})$

m estimates $\hat{Q}^{(j)}$, $U^{(j)}$, for j = 1, …, m

Estimate for Q:

$$\bar{Q} = \frac{1}{m}\sum_{j=1}^{m}\hat{Q}^{(j)}$$

50

Average within-imputation variance

$$\bar{U} = \frac{1}{m}\sum_{j=1}^{m} U^{(j)}$$

Between-imputation variance

$$B = \frac{1}{m-1}\sum_{j=1}^{m}\left[\hat{Q}^{(j)} - \bar{Q}\right]^2$$

Total variance:

$$T = \bar{U} + \left(1 + \frac{1}{m}\right)B$$

51

Student's t approximation for confidence intervals and tests for Q

$$\frac{\bar{Q} - Q}{\sqrt{T}} \sim t_{\upsilon}$$

where

$$\upsilon = (m-1)\left[1 + \frac{\bar{U}}{(1+m^{-1})B}\right]^2$$

52

2. MI imputeringsmodell (mm)

- Valg av variable i imputeringsmodellen
- Interaksjoner og ikkelineære effekter
- Hvor mange imputerte datasett?

53

53

Study the amount of and patterns of missing data

- Calculate amount of missing data for the relevant variables. STROBE, Descriptive data, 14b: «Indicate the number of participants with missing data for each variable of interest”
- Study missing data patterns

STROBE Statement—checklist of items that should be included in reports of observational studies.
https://www.strobe-statement.org/index.php?id=available-checklists

54

54

Multiple imputation

Analysis model:

$$y \mid x_1\, x_2\, x_3 \cdots x_p$$

Imputation model:
A set of regression equations
(usually linear, binary logistic regression, nominal logistic regression,
or ordinal logistic regression)

$$x_1 \mid y \quad x_2\, x_3 \cdots x_p$$
$$x_2 \mid y\, x_1 \quad x_3 \cdots x_p$$
$$\vdots$$
$$x_p \mid y\, x_1\, x_2 \cdots x_{p-1}$$

55

Straightforward if missingness is monotone (x_{ij} missing $\Rightarrow$ x_{ik} is missing for $k > j$)

$$x_1 \mid y$$
$$x_2 \mid y\, x_1$$
$$x_3 \mid y\, x_1\, x_2$$
$$\vdots$$
$$x_p \mid y\, x_1\, x_2 \cdots x_{p-1}$$

Else: Use chained equations, solved iteratively using Markov Chain Monte Carlo (MCMC)

$$x_1 \mid y \quad x_2\, x_3 \cdots x_p$$
$$x_2 \mid y\, x_1 \quad x_3 \cdots x_p$$
$$\vdots$$
$$x_p \mid y\, x_1\, x_2 \cdots x_{p-1}$$

56

MI using chained equations

- Idea: Mimic the conditional distribution of the missing values given the joint distribution
- Automatic procedure:
  - Uses Markov Chain Monte Carlo (MCMC) technique
  - Initial guesses for the missing values are inserted.
  - Use the equations to improve the predicted missing values
  - Repeat until convergence (not always achieved)
- Even when convergence is achieved, it can happen that it converges at some other distribution
- Simulation studies confirm that the procedure works (surprisingly) well

57

57

Predictors in the imputation model

- Include all variables to be used in the main analysis model(s). Failure to do so may bias the analysis.
- The outcome variable in the analysis model must be included as a predictor in the imputation model
- Possibly include (auxilliary) variables not in the analysis model:
  - Include predictors of missingness.
  - Include variables associated with the variable to be imputed
- Rather use a rich model than a model with few covariates.
- But: Limit the number of predictors, to avoid mathematical instability. The default maximum in SPSS is 100. Usually wise to stay well below 100.

58

58

Predictors in the imputation model

- Possible interactions and nonlinear effects must be handled appropriately.
- Some algorithms require categorical variables to be coded 0, 1, …, k
- When the main analysis is lifetime analysis such as Cox regression, include both the time t and censoring indicator as predictors. Some authors recommend using ln(t). But the literature on methods for censored data is rapidly evolving (van Buuren 2018, page 93 to 96 and page 270)
- Multilevel analysis models: The multilevel structure must be included in the imputation model.

59

59

Interactions and nonlinear effects in the analysis model:

Interaction:
Includes the term x_1x_2 in addition to the main effect x_1 and x_2.

Nonlinear effect:
For example, x_3 and x_3^2.

60

Traditional advice ("passive imputation"):
Compute the terms x_1x_2 and x_3^2 after x_1, x_2, and x_3 have been imputed.

But this may induce bias:
Although y is a linear function of x_1x_2 , and of x_3 and x_3^2 in the main analysis model: Still, x_1, x_2, and x_3 are NOT linear functions of y in the imputation model.

Possible remedies:
- JAV (Just another variable): Treat x_1x_2 and x_3^2 as if they are separate variables, not functions of x_1, x_2, and x_3 .
- Dichotomous variable in the interaction (f.ex. sex): Split file in two and impute separately, then combine the imputed files.
- See also (van Buuren 2012) and (Carpenter & Kenward 2013).

For RCT, Sullivan et al (2018) generally recommend imputing separately in the treatment arms.

61

In MI algorithms, continuous variables are generally assumed normally distributed. What if the variable to be imputed has a skewed distribution or limited range:

Examples:
- Concentration of a substance in a liquid
- Likert scale, for example from 0 (or 1) to k

Possible solutions:
a) Non-rounded regression (including out of range values)
b) Impute on transformed variable (fex log(x) or log(x+c) or sqrt(x))
c) Post-imputation rounding
d) Truncated regression
e) Predictive mean matching
f) Combining b) with c), d) or e)

Note that the range restrictions in the MI menu in SPSS use during-imputation rejection of out of range values. This may be similar to d), but I expect it to introduce bias. I do not recommend it.

62

Skewed or limited range variables:

Varying advice exists in the literature.

(Rodwell et al. 2014): "… the best method to impute limited-range variables is to impute on the raw scale with no restrictions to the range, and with no post-imputation rounding. … Although this imputation method results in some implausible values, it appears to be the most consistent method with low bias and reliable coverage … "

The purpose of MI is not to create sensible data sets, but sensible estimates.

63

Impute the outcome variable?

Under MAR, there are generally no benefits to impute the outcome, and for a low number of imputations the results may even be somewhat more variable because of simulation error. There is an important exception to this. If we have access to an auxiliary complete variable that is not part of the model and that is highly correlated with the outcome, imputation can be considerably more efficient than complete case analysis, resulting in more precise estimates and shorter confidence intervals. A common scenario where this occurs is if we have a cheap outcome measure for everyone, and an expensive measure for a subset.

In many data sets, missing data also occur in the independent variables. In these cases, we need to impute the outcome variable since its imputed version is needed to impute the independent variables.

https://stats.stackexchange.com/questions/46226/multiple-imputation-for-outcome-variables (Stef van Buuren, 13 Jan 2013)

64

Impute the outcome variable?

«Suppose that the complete-data model is a regression with outcome Y and predictors X. If the missing data occur in Y only, complete-case analysis and multiple imputation are equivalent, so then complete-case analysis is preferred since it is easier, more efficient and more robust (Von Hippel, 2007). … Multiple imputation gains advantage over complete-case analysis if additional predictors for Y are available that are not part of X. The efficiency of complete-case analysis declines if X contains missing values, which may result in inflated type II error rates.» (Van Buuren 2018, page 57)

65

Proper MI reflecting the uncertainty in the model parameters

A single imputation is drawn from $P(Y_{mis} \mid Y_{obs}; \hat{\theta})$

MI:

simulate m plausible values $\theta^{(1)},...,\theta^{(m)}$
draw $Y_{mis}^{(t)}$ from $P[Y_{mis} \mid Y_{obs}; \theta^{(t)}]$ for t=1,…,m

Bayesian approach with a prior distribution for θ
is natural but not essential

66

Imputation of (residual) value given predictive value. Alternatives:

- Draw from the estimated regression model
- Predictive mean matching (PMM): Draw and outcome randomly from the *k* cases with close predictive value. (Default is *k*=1 in Stata)
  - Does not return values outside the observed range
  - Robust against misspecification of the imputation model
  - Works best with large data sets
  - May perform poosly in small data sets, especially with a small number of discrete values

67

67

How many imputations *m*?

- The classic advice was *m* = 3 to 5.
- Bjørnstad & Lydersen (2012) generally recommend *m* = 20. But a higher number may be required to report p-values with, say, 2 digits accuracy.
- Van Buuren (2018) reviews relevant work. «It could be beneficial to set *m* higher, in the range 20 to 100.»
- If you use *m*=100, you are on the safe side.

68

68

Set random seed

- Set random seed (to for example 12345) to make it possible to exactly reproduce results :
  - SPSS: set seed = 12345.
  - Stata: Option rseed(12345)

69

69

3. MI analysemodell

- Rubins regler
- Hvilke analyser og parametre kan håndteres?

70

70

Rubin's (1987) rules for combining estimates and variances

Q = the population quantity of interest, $U = Var(\hat{Q})$

m estimates $\hat{Q}^{(j)}$, $U^{(j)}$, for j = 1, …, m

Estimate for Q:

$$\bar{Q} = \frac{1}{m}\sum_{j=1}^{m}\hat{Q}^{(j)}$$

71

Average within-imputation variance

$$\bar{U} = \frac{1}{m}\sum_{j=1}^{m}U^{(j)}$$

Between-imputation variance

$$B = \frac{1}{m-1}\sum_{j=1}^{m}\left[\hat{Q}^{(j)} - \bar{Q}\right]^2$$

Total variance:

$$T = \bar{U} + \left(1 + \frac{1}{m}\right)B$$

72

Student's t approximation for confidence intervals and tests for Q

$$\frac{\bar{Q}-Q}{\sqrt{T}} \sim t_{\upsilon}$$

where

$$\upsilon = (m-1)\left[1+\frac{\bar{U}}{(1+m^{-1})B}\right]^2$$

73

Relative increase in variance due to nonresponse:

$$r = \frac{\left(1+\frac{1}{m}\right)B}{\bar{U}}$$

Fraction missing information (FMI): The part of the total variance of the estimate for Q which is attributed to missing values

$$\gamma = \frac{r+2/(\nu+3)}{1+r}$$

74

Which statistics (quantities) can be combined using Rubin's rules?

- Straighforward:
  - Mean, standard deviation, proportion, regression coefficient, area under the ROC curve
- Log transformed (that is the regression coefficient):
  - Odds ratio, hazard ratio
- Correlation: Use the Fisher z transformation
- Not applicable: Statistics whose values change systematically with sample size, such as:
  - P-values, likelihood ratio statistic

75

75

Check plausibility of results …

- Outliers among the m estimates?
- Check fraction missing information

76

76

Eventuelle ekstra slides om longitudinelle data:

77

Alternative methods for logitudinal data:

- Repeated measures ANOVA
- Mixed models
- Generalized estimating equations (GEE)

Some structural equation models (SEM), including growth models, are generalizations of mixed models, and relevant with latent variables beyond random intercept and random slope models.

78

78

Repeated measures ANOVA

- Only complete cases are included in the analysis
- Unbiased only if data are missing completely at random (MCAR)
- The underlying mathematical model is not transparent
- Was an attractive method before computers became powerful (Personal communication with Garrett M. Fitzmaurice)
- Ought to be in the museum.

79

79

Chance, vil 18, no 3, 2005

Should we quit using repeated measures analysis of variance?

Repeated Measures ANOVA, R.I.P.?

Charles E. McCulloch

It is a difficult experiment to run and to analyze: What are the effects of alcohol on sleepiness and does a hormone, pregnenolone, which has been shown to enhance memory in rat experiments, help alleviate the sleepiness? Each person is tested under each of four conditions on four different visits in random order: a placebo for the drug and for the hormone, alcohol alone, hormone alone, and the combination. Each subject is also queried multiple times within a visit in the minutes after alcohol (or placebo) ingestion. Some subjects drop out of the protocol without completing all the conditions and some of the sleepiness scores are not recorded within a visit because of difficulties executing the protocol. How should the data be analyzed?

This is an opportunity for the professional statistician to pull out any of a number of impressive and more recent tools of the trade: generalized estimating equations, mixed model analyses, imputation, and inverse probability weighting. Gone

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \alpha\beta_{jk} + p_i + \varepsilon_{ijk}, \quad (1)$$

where μ is the overall mean, α_j represents the alcohol effect, β_k the pregnenolone effect, $\alpha\beta_{jk}$ the interaction effect, p_i is the person effect, and ε_{ijk} is an error term. This model hypothesizes simple person effects that raise or lower (if the effect is negative) the average sleepiness in all four conditions. Interest focuses on the interaction, because the scientific question is whether pregnenolone helps to reduce the sleep-inducing effect of alcohol.

With a mean, error term and four explanatory factors in the model, the analysis of variance would partition the variability in Y_{ijk} into four sources: person, alcohol, pregnenolone, and the interaction. An ANOVA would generate a table as outlined in Table 1. A typical assumption is that the ε_{ijk} are all independent and follow a $N(0, \sigma^2)$ distribution and that the rest of the terms in (1) are fixed. If these assumptions are correct, the

80

Mixed models

- Includes all subjects, also those with missing data at some time point(s)
- Unbiased under the less restrictive missing at random (MAR) assumption
- Transparent mathematical model

81

81

GEE: Generalized estimating equations

- A useful alternative to Mixed models, especially for categorical outcome such as binary data (logistic regression) or count data (Poisson regression).
- Unbiased only if data are MCAR

82

82

References

Altman, D.G. & Bland, J.M. 2007. Statistical notes: Missing data. *BMJ*, 334, 424

Askim, T., Langhammer, B., Ihle-Hansen, H., Gunnes, M., Lydersen, S., & Indredavik, B. 2018. Efficacy and Safety of Individualized Coaching After Stroke: the LAST Study (Life After Stroke): A Pragmatic Randomized Controlled Trial. *Stroke*, 49, 426-432

Bjørnstad, J. & Lydersen, S. 2012, "Missing data," *In Medical statistics in clinical and epidemiological research*, M. Veierød, S. Lydersen, & P. Lakke, eds., Oslo: Gyldendal akademisk, pp. 429-461.

Carpenter, J.R. & Kenward, M.G. 2013. *Multiple imputation and its application* Chichester, West Sussex, John Wiley & Sons.

Carpenter, J. R. & Kenward, M. G. 2015, "Development of Methods for the Analysis of Partially Observed data and Critique of ad hoc Methods," *In Handbook of Missing Data Methodology*, G. Molenberghs et al., eds., CRC / Champan Hall, pp. 23-46.

83

References

Day, S. 2007. *Dictionary for clinical trials*, 2nd ed ed. Chichester, England, John Wiley & Sons.

Hallan, S.I., Ritz, E., Lydersen, S., Romundstad, S., Kvenild, K., & Orth, S.R. 2009. Combination of estimated glomerular filtration rate and albuminuria provides best prediction of kidney failure: Results of the HUNT II study, Norway. *Journal of the American Society of Nephrology*, 20, 1069-1077

Horton, N.J. & Kleinman, K.P. 2007. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *American Statistician*, 61, (1) 79-90

Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M.G., Mallinckrodt, C., & Carroll, R.J. 2004. Analyzing incomplete longitudinal clinical trial data. *Biostatistics.*, 5, (3) 445-464

Molenberghs, G., Fitzmaurice, G., Kleinman, K.P., Tsiatis, A., & Verbeke, G. 2015. *Handbook of Missing Data Methodology* Chapman & Hall/CRC.

Molenberghs, G. & Kenward, M.G. 2007. *Missing data in Clinical Studies* Chichester, Wiley.

84

References

O'Kelly, M. & Ratitch, B. 2014. *Clinical trials with missing data: A guide for practitioners* Chichester, Wiley.

Prestmo, A., Hagen, G., Sletvold, O., Helbostad, J.L., Thingstad, P., Taraldsen, K., Lydersen, S., Halsteinli, V., Saltnes, T., Lamb, S.E., Johnsen, L.G., & Saltvedt, I. 2015. Comprehensive geriatric care for patients with hip fractures: a prospective, randomised, controlled trial. *Lancet*, 385, (9978) 1623-1633

Rodwell, L., Lee, K.J., Romaniuk, H., & Carlin, J.B. 2014. Comparison of methods for imputing limited-range variables: a simulation study. *BMC.Med.Res.Methodol.*, 14, 57

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

Schafer, J.L. 1997. *Analysis of incomplete multivariate data* London, Chapman & Hall.

Schafer, J.L. & Graham, J.W. 2002. Missing data: Our view of the state of the art. *Psychological Methods*, 7, (2) 147-177

85

References

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

Sullivan, T.R., White, I.R., Salter, A.B., Ryan, P., & Lee, K.J. 2018. Should multiple imputation be the method of choice for handling missing data in randomized trials? *Stat.Methods Med.Res.*, 27, (9) 2610-2626

van Buuren, S., Boshuizen, H.C., & Knook, D.L. 1999. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat.Med.*, 18, (6) 681-694

van Buuren, S. 2018. *Flexible imputation of missing data* 2nd ed. Boca Raton, FL, CRC Press.

Vickers, A.J. & Altman, D.G. 2013. Statistics notes: missing outcomes in randomised trials. *BMJ*, 346, f3438

86