

PHUSE EU Connect 2020

INCOMPLETE DATA IN LONGITUDINAL STUDIES: CONTINUOUS AND DICHOTOMOUS OUTCOME

ANA GRECU
PAREXEL, ROMANIA

Outline

- I. Introduction
- II. Multiple imputation for continuous outcome
 - 1. Analysis model
 - 2. Imputation model
 - 3. Sensitivity analysis
- III. Multiple imputation for dichotomized outcome
 - 1. Multiple imputation and standard GEE
 - 2. Weighted GEE
- IV. Concluding comments

Introduction

Incomplete data analysis

- I. Valid methods for primary analysis:
 - 1. Direct likelihood
 - 2. Multiple imputation
 - 3. Weighted GEE (for dichotomized outcome)

- II. Non-valid methods for primary analysis:
 - 1. CC (Complete cases)
 - 2. LOCF (Last observation carried forward)
 - 3. BOCF (Baseline observation carried forward)

Introduction

Incomplete data taxonomy

I. Missing data mechanism

1. MCAR: probability of being observed is independent of the data
2. MAR: probability of being observed is independent of unmeasured data given the observed information
3. MNAR: probability of being observed depends on unmeasured data

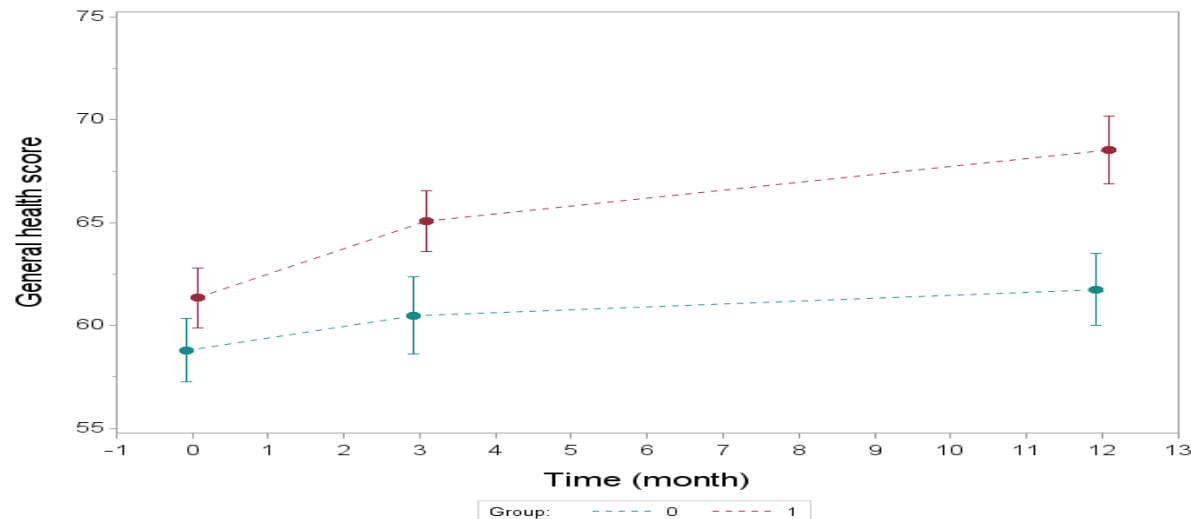
II. Missing data pattern

1. Monotone (dropout)
2. Non-monotone (intermittent)

Introduction

Case study: hypothetical clinical trial

- 410 subjects: control (200 patients) and active treatment (210 patients)
- Response: evolution in general health (0 to 100 score)
- Baseline patient characteristics: age, gender
- 2 post-baseline outcome measurements



MI for continuous outcome

Missing data pattern

Available data	Control group		Treatment group	
	N	%	N	%
1 st , 2 nd and 3 rd visits (completers)	134	67	156	74.3
Only 1 st visit (monotone)	40	20	33	15.7
1 st and 2 nd visits (monotone)	18	9	15	7.1
1 st and 3 rd visits (non-monotone)	8	4	6	2.9

- Monotone missing pattern
- Non-monotone missing pattern
- Overall missingness ~ 30% (number of patients with at least one missing value)

MI for continuous outcome

Analysis model

$$Y_{ij} = \beta_0 + \beta_1 Age_i + \beta_2 Gen_i + \beta_3 G_i * t_{ij} + b_{1i} + b_{2i}t_{ij} + \varepsilon_{ij}$$

Y_{ij} - General health score measured at occasion j for patient i

G_i - treatment group: 0 – control, 1 – active treatment

t_{ij} - measurement time for occasion j , ($j=0, 6, 12$)

$b_{1i}, b_{2i} \sim N(0, D)$ - random effects

$\varepsilon_{ij} \sim N(0, \Sigma_i); \Sigma_i = \sigma^2 I_{n_i}$

```
proc mixed data=mydata method=ml covtest ;
class id group gender timecont;
model genhealth = age gender group*time / s;
random intercept time/type=un subject=id g gcorr v vcorr;
repeated timecont/type=simple subject=id r rcorr ;
run;
```

MI for continuous outcome

Imputation model and number of imputations

- Decide on the method: FCS
- Predictors: all variables used in the analysis model + baseline measurement of the outcome
- Use auxiliary variables if available
- Use a sufficiently large number of imputations

```
ods graphics on;  
proc mi data=alldata seed=123456 simple out=step0 nimpute=50 round=0.1 minimum=0 maximum=100;  
class gender;  
by group ;  
fcs reg (gen2= age gender gen1);  
fcs reg(gen3= age gender gen2) plots=trace;  
var age gender gen1 gen2 gen3;  
run;
```

MI for continuous outcome

Imputation model and number of imputations

```
*** Analysis task;
```

```
proc mixed data=step02 method=ml;  
class id timeclass group gender;  
by _imputation_;  
model gen = age gender group*time / s covb;  
random intercept time/type=un subject=id g gcorr v vcorr;  
repeated timeclass/type=simple subject=id r rcorr;  
ods output solutionF=nlparms covb=gmcovb;  
run;
```

```
*** Inference task;
```

```
proc mianalyze parms=nlparms covb(effectvar=rowcol)=gmcovb wcov bcov tcov ;  
class group gender;  
modeleffects Intercept age gender group*time ;  
run;
```

MI for continuous outcome

Imputation model and number of imputations

Overview of the results

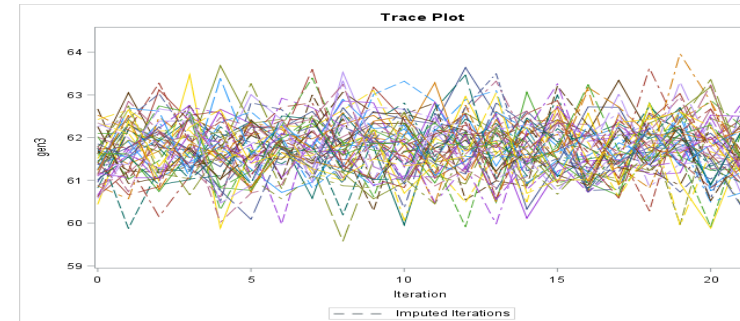
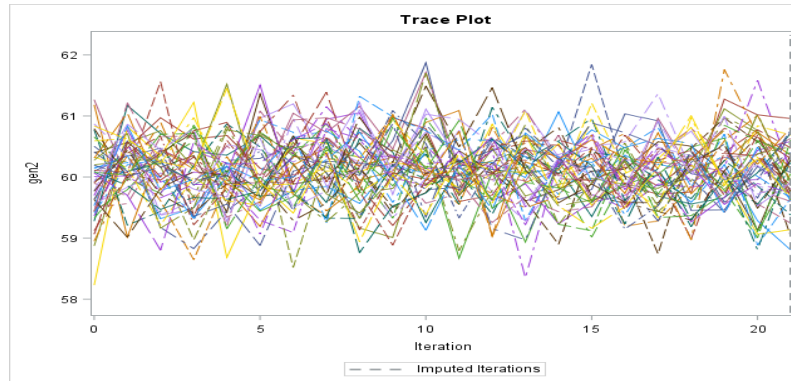
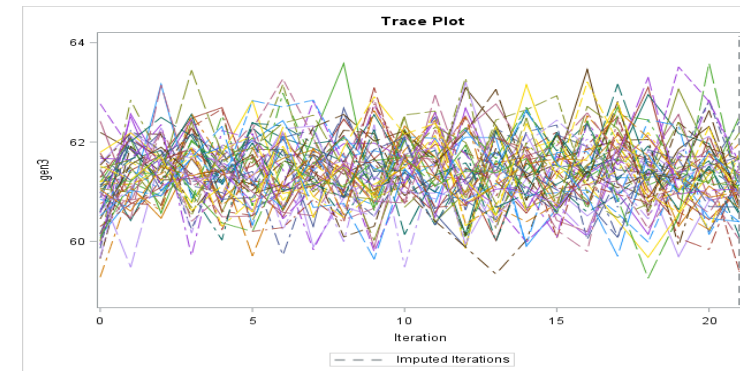
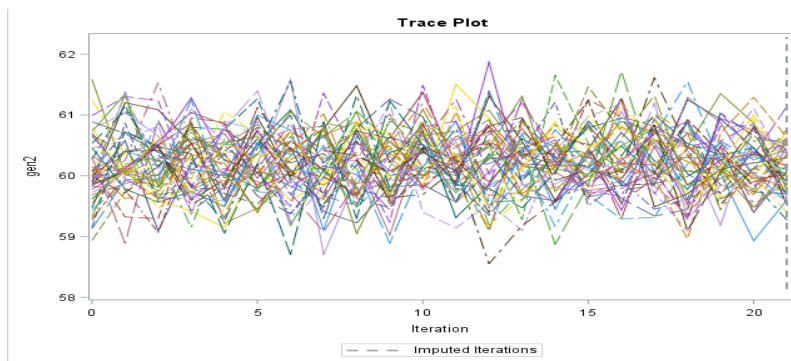
Effect	Parameter	Direct likelihood	MI: FCS m=30	MI: FCS m=50
Intercept	β_0	66.872 (4.524)	68.180 (4.559)	67.783 (4.553)
Age	β_1	-0.145 (0.096)	-0.178 (0.095)	-0.168 (0.095)
Gender (M)	β_2	-0.559 (0.864)	-0.388 (2.066)	-0.432 (2.064)
Time x group 0	β_3	-0.028 (0.106)	0.017 (0.114)	0.012 (0.108)
Time x group 1	β_3	0.227 (0.100)	0.216 (0.107)	0.215 (0.104)

➤ Relatively stable estimates with m=30/50

MI for continuous outcome

Imputation diagnostics

Random pattern in trace plots



MI for continuous outcome

Sensitivity analysis

Is MAR assumption plausible?



Sensitivity analysis

Deviate from MAR to check this assumption

MI for continuous outcome

Sensitivity analysis

➤ Control-based pattern imputation

```
proc mi data=alldata seed=123456 simple out=step01 nimpute=50 round=0.1;
class group gender;
var age gender gen1 gen2 gen3;
fcs reg;
mnar model (gen2/modelobs=(group='0'));
mnar model (gen3/modelobs=(group='0'));
run;
```

➤ Adjusted pattern imputation

```
proc sort data=alldata; by group; run;
proc mi data=help seed=123456 simple out=step01 nimpute=50 round=0.1;
class group gender;
var age gender gen1 gen2 gen3;
fcs reg;
mnar adjust (gen2/shift=2 adjustobs=(group='1'));
mnar adjust (gen3/shift=5 adjustobs=(group='1'));
by group;
run;
```

MI for continuous outcome

Sensitivity analysis

Overview of the results

Effect	Parameter	MAR m=50	MNAR Subgroup adjustment	MNAR Shift model
Intercept	β_0	67.783 (4.553)	67.882 (4.716)	68.681 (4.609)
Age	β_1	-0.168 (0.095)	-0.178 (0.098)	-0.189 (0.097)
Gender (M)	β_2	-0.432 (2.064)	0.194 (2.134)	0.073 (2.107)
Time x group 0	β_3	0.012 (0.108)	0.006 (0.106)	0.003 (0.105)
Time x group 1	β_3	0.215 (0.104)	0.209 (0.094)	0.363 (0.104)

Stable parameter estimates

MI for dichotomized outcome

A model for dichotomized outcome

$$\log\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \beta_0 + \beta_1 Age_i + \beta_2 Gender + \beta_3 t_{ij} + \beta_4 G_{i*} t_{ij}$$

Possible approaches

- Standard GEE after multiple imputation
- Weighted GEE → requires monotone missing pattern

MI for dichotomized outcome

MI and standard GEE

➤ Dichotomization of the outcome done after imputation

```
data step1;
set step1;
if genhealth=<50 then dgen=0;
if genhealth>50 then dgen=1;
run;

proc gee data=step2 descending;
class timeclass id ;
by _imputation_ ;
model dgen = age gender0 gender1 time timegroup / dist=binomial;
repeated subject= ID/ within=timeclass corr=un corrw modelse
covb;
ods output GEEEmpPEst=gmparms parminfo=gmpinfo GEERCov=gmcovb;
run;

proc mianalyze parms=gmparms covb=gmcovb parminfo=gmpinfo wcov
bcov tcov;
modeleffects Intercept age gender0 time timegroup;
run;
```

MI for dichotomized outcome

Weighted GEE

- Allows only monotone missing pattern (if possible use *MCMC impute=monotone* to impute non-monotone missingness)

```
proc gee data=step1 plots=histogram;  
class id group timeclass gender prevgenh;  
model dgen = age gender time timegroup / dist=binomial;  
repeated subject= ID/ within=timeclass corr=un corrw modelse;  
missmodel prevgenh group age gender time/type=obslevel; run;
```

- Dependence on covariates and previous outcome
(**prevgenh=lag(dgen)**)

MI for dichotomized outcome

Overview of the results

	GEE after MI		Weighted GEE	
Effect	Estimate (s.e.)	P-value	Estimate (s.e.)	P-value
Intercept	0.884 (0.415)	0.033	0.554 (0.428)	0.203
Age	-0.006 (0.008)	0.491	0.001 (0.009)	0.908
Gender (M)	-0.095 (0.179)	0.597	0.004 (0.103)	0.968
Time	0.002 (0.012)	0.862	-0.004 (0.013)	0.769
Time x Group 1	0.015 (0.016)	0.361	0.018 (0.017)	0.297

- Slightly different parameter estimates
- No important change in parameters significance

MI in longitudinal studies

Concluding comments

- Multiple imputation instead of single imputation
- Analysis model and imputation model does not have to be the same
- Use Rubin's rule to combine the imputed data
- Check the imputed variables and perform imputation diagnostics
- Assess plausibility of the MAR assumption

THANK YOU FOR YOUR ATTENTION!