

HANDLING MISSING DATA IN CLINICAL TRIALS: TECHNIQUES AND METHODS



Pennidhi Karlakunta & Naveen Kommuru

BARDS

OBJECTIVES

- Effects of missing data
- Pattern of missing data
- Missing data mechanisms
- Prevention of missing data
- Methods to handle missing data
- Multiple imputation: Case study

EFFECTS OF MISSING DATA

- Power
- Variability
- Bias

PATTERN OF MISSING DATA

- Monotonic
- Arbitrary
- Matrix sampling

MISSING DATA MECHANISMS

- Missing completely at random (MCAR)
- Missing at random (MCR)
- Missing not at random (MNAR)

PREVENTION OF MISSING DATA

- Trial outcomes
- Minimizing dropouts
- Data collection for dropouts
- Actions for Investigators and site personnel
- Targets for acceptable rates of missing data

METHODS TO HANDLE MISSING DATA

- Complete case analysis
- Single imputation
- Inverse probability weighing
- Likelihood based analysis
- Event time analysis
- Non-responder imputation
- Multiple imputation

MULTIPLE IMPUTATION

Multiple imputation (MI) is a method for obtaining estimates and correct inferences for statistics ranging from simple descriptive statistics to the parameters of complex multivariate models.

Three steps:

- Independently impute missing values in the original dataset M times.
- Analysis of M imputed datasets.
- Combining the output results from above steps, to generate final estimates, standard errors and confidence intervals.

WHY MULTIPLE IMPUTATION

- Model-based
- Multivariate
- Multiple independent repetitions
- Robust
- Very usable

HOW MANY MULTIPLE IMPUTATION REPETITIONS ARE NEEDED?

Rate of missing	Repetitions to achieve minimum 95% of statistical efficiency
< 20%	M=5 to M=10
30% to 50%	M=20 to M=30

METHODS FOR MULTIPLE IMPUTATION

Missing Data Pattern	Variable Type	Method	PROC MI Statement in SAS
Monotone	Continuous	Linear regression predictive mean matching propensity score	MONOTONE REG MONOTONE REGPMM MONOTONE PROPENSITY
	Binary	Logistic regression	MONOTONE LOGISTIC
	Nominal	Discriminant function	MONOTONE DISCRIM
Arbitrary	Continuous	With continuous covariates: MCMC monotone method MCMC full-data imputation	MCMC IMPUTE=MONOTONE MCMC IMPUTE=FULL
	Continuous	With mixed covariates: FCS regression FCS predictive mean matching	FCS REG FCS REGPMM
	Binary	FCS logistic regression	FCS LOGISTIC
	Nominal	FCS discriminant function	FCS DISCRIM

CASE STUDY

Report with complete case analysis

Analysis of Change from Baseline in EORTC QLQ-C30 Global Health Status/QoL at Week 24
(FAS Population)

Treatment	Baseline		Week 24		Change from Baseline at Week 24	
	N	Mean (SD)	N	Mean (SD)	N	LS Mean (95% CI) [†]
Treatment A	202	66.87 (18.982)	125	74.53 (16.943)	222	4.65 (-1.55, 7.75)
Treatment B	69	69.57 (19.742)	46	70.83 (18.151)	74	0.07 (-4.73, 4.88)
Pairwise Comparison					Difference in LS Means (95% CI)	p-Value
Treatment A vs. Treatment B					4.57 (-0.79, 9.94)	0.0943
[†] Based on cLDA model with the PRO scores as the response variable, and treatment by study visit interaction, stratification factors (prior auto-SCT (yes, no) and disease status after frontline therapy (primary refractory, relapsed less than 12 months, relapsed 12 months or more)) as covariates. For baseline and Week 24, N is the number of subjects in each treatment group with non-missing assessments at the specific time point; for change from baseline, N is the number of subjects in the analysis population in each treatment group. Database Cutoff Date: 16JAN2020						

Source: [adam-adsl; adpro]

CASE STUDY (Cont..)

SAMPLE ADPRO ANALYSIS DATASET

	USUBJID	TRT01P	STRATA1N	PARAM	PARAMCD	BASE	AVAL	CHG	CHGCAT1	AVISIT
1	XXXX-XXX_XX00A	Treatment A	6	Global health status/QoL	QL2	.	83.333	.		Week 6
2	XXXX-XXX_XX00A	Treatment A	6	Global health status/QoL	QL2	.	83.333	.		Week 12
3	XXXX-XXX_XX00A	Treatment A	6	Global health status/QoL	QL2	.	83.333	.		Week 18
4	XXXX-XXX_XX00B	Treatment A	4	Global health status/QoL	QL2	91.667	91.667	.		Baseline
5	XXXX-XXX_XX00B	Treatment A	4	Global health status/QoL	QL2	91.667	91.667	0	Stable	Week 18
6	XXXX-XXX_XX00B	Treatment A	4	Global health status/QoL	QL2	91.667	83.333	-8.334	Stable	Week 24
7	XXXX-XXX_XX00C	Treatment A	6	Global health status/QoL	QL2	50	50.000	.		Baseline
8	XXXX-XXX_XX00C	Treatment A	6	Global health status/QoL	QL2	50	66.667	16.667	Improved	Week 6
9	XXXX-XXX_XX00C	Treatment A	6	Global health status/QoL	QL2	50	66.667	16.667	Improved	Week 12
10	XXXX-XXX_XX00C	Treatment A	6	Global health status/QoL	QL2	50	66.667	16.667	Improved	Week 18

CASE STUDY (Cont..)

RESTRUCTURING ANALYSIS DATASET

```
proc transpose data=adpro out=adpro_t prefix=y;  
  by usubjid trt01pn trt01p STRATA1 paramcd param;  
  id avisitn;  
  var aval;  
run;
```

USUBJID	TRT01P	PARAMCD	PARAM	Y2	Y3	Y4	Y1	Y5
XXXX-XXX_XX00A	Treatment A	QL2	Global health status/QoL	83.333	83.333	83.333	.	.
XXXX-XXX_XX00B	Treatment A	QL2	Global health status/QoL	.	.	91.667	91.667	83.333
XXXX-XXX_XX00C	Treatment A	QL2	Global health status/QoL	66.667	66.667	66.667	50.000	.

CASE STUDY (Cont..)

CHECKING THE MISSING DATA PATTERN

```
proc mi data=adplda_t nimpute=0 simple;
  class paramcd trt01pn;
  fcs;
  var y1-y5 paramcd trt01pn;
run;
```

Missing Data Patterns														
Group	Y1	Y2	Y3	Y4	Y5	PARAMCD	TRT01PN	Freq	Percent	Group Means				
										Y1	Y2	Y3	Y4	Y5
1	X	X	X	X	X	X	X	131	44.26	67.875328	73.345977	74.109397	73.345962	73.218756
2	X	X	X	X	.	X	X	42	14.19	67.857190	65.674667	66.865071	64.880952	.
3	X	X	X	.	X	X	X	15	5.07	68.333267	72.222133	71.666667	.	71.666533
4	X	X	X	.	.	X	X	41	13.85	64.024415	67.479707	60.365878	.	.
5	X	X	.	X	X	X	X	4	1.35	81.249750	60.416750	.	72.916750	72.916750
6	X	X	.	X	.	X	X	4	1.35	70.833500	72.916750	.	77.083250	.
7	X	X	.	.	.	X	X	18	6.08	67.592500	60.648111	.	.	.

CASE STUDY (Cont..)

MODEL BASED MULTIPLE IMPUTATION (Step 1)

```
proc mi data=adpro_t out=adpro_mi seed=3475 nimpute=50
minmaxiter=1000 minimum=. 0 0 0 0 0 maximum=. 100 100 100 100 100;
    by paramcd trt01pn;
    class STRATA1;
    FCS REG (y1-y5);
    var STRATA1 y1-y5;
run;
```

IMPUTATION	USUBJID	TRT01P	PARAMCD	PARAM	Y2	Y3	Y4	Y1	Y5
1	XXXX-XXX_XX00A	Treatment A	QL2	Global health status/QoL	83.333	83.333	83.333	83.884	92.218
1	XXXX-XXX_XX00B	Treatment A	QL2	Global health status/QoL	89.132	76.850	91.667	91.667	83.333
1	XXXX-XXX_XX00C	Treatment A	QL2	Global health status/QoL	66.667	66.667	66.667	50.000	71.330
2	XXXX-XXX_XX00A	Treatment A	QL2	Global health status/QoL	83.333	83.333	83.333	81.288	93.142
2	XXXX-XXX_XX00B	Treatment A	QL2	Global health status/QoL	73.708	91.064	91.667	91.667	83.333
2	XXXX-XXX_XX00C	Treatment A	QL2	Global health status/QoL	66.667	66.667	66.667	50.000	50.737
3	XXXX-XXX_XX00A	Treatment A	QL2	Global health status/QoL	83.333	83.333	83.333	87.961	59.878
3	XXXX-XXX_XX00B	Treatment A	QL2	Global health status/QoL	96.062	91.519	91.667	91.667	83.333
3	XXXX-XXX_XX00C	Treatment A	QL2	Global health status/QoL	66.667	66.667	66.667	50.000	72.608

CASE STUDY (Cont..)

BACK TO ADAM BDS STRUCTURE

```
proc transpose data=adpro_mi out=adpromi;
  by _imputation_ trt01pn trt01p usubjid STRATA1 parcat2 paramcd
  paramn param;
run;
```

IMPUTATION	USUBJID	TRT01P	STRATA1N	PARAM	PARAMCD	BASE	AVAL	CHG	CHGCAT1	AVISIT
1	XXXX-XXX_XX00A	Treatment A	6	Global health status/QoL	QL2	83.54324	83.333	-0.2102396	Stable	Week 6
1	XXXX-XXX_XX00A	Treatment A	6	Global health status/QoL	QL2	83.54324	83.333	-0.2102396	Stable	Week 12
1	XXXX-XXX_XX00A	Treatment A	6	Global health status/QoL	QL2	83.54324	83.333	-0.2102396	Stable	Week 18
1	XXXX-XXX_XX00A	Treatment A	6	Global health status/QoL	QL2	83.54324	83.54324	0	Stable	Baseline
1	XXXX-XXX_XX00A	Treatment A	6	Global health status/QoL	QL2	83.54324	45.70502	-37.83822	Deteriorated	Week 24
1	XXXX-XXX_XX00C	Treatment A	6	Global health status/QoL	QL2	50	66.667	16.667	Improved	Week 6
1	XXXX-XXX_XX00C	Treatment A	6	Global health status/QoL	QL2	50	66.667	16.667	Improved	Week 12
1	XXXX-XXX_XX00C	Treatment A	6	Global health status/QoL	QL2	50	66.667	16.667	Improved	Week 18
1	XXXX-XXX_XX00C	Treatment A	6	Global health status/QoL	QL2	50	50	0	Stable	Baseline
1	XXXX-XXX_XX00C	Treatment A	6	Global health status/QoL	QL2	50	47.18979	-2.8102134	Stable	Week 24

CASE STUDY (Cont..)

ANALYSIS WITH MULTIPLE IMPUTED DATASET (Step 2)

```
proc means data= adpromi MEAN STD median min max q1;  
  by _imputation_ trt01pn trt01p;  
  var aval base;  
run;
```

```
proc mixed data= adpromi;  
  by _imputation_;  
  class avisitn usubjid STRATA1N;  
  model aval=avisitn STRATA1N v2t1 v3t1 v4t1 v5t1 / ddfm=kr;  
  repeated avisitn / subject=usubjid type=un R;  
  estimate "P1: Week 24; TRT: 1" avisitn -1 0 0 0 1 v5t1 1 /  
    divisor=1 cl alpha=0.05;  
  estimate "P1: Week 24; TRT: 2" avisitn -1 0 0 0 1/  
    divisor=1 cl alpha=0.05;  
  estimate "P2: Week 24; TRT: 1 - 2" v5t1 1 / cl alpha=0.05;  
run;
```

CASE STUDY (Cont..)

Results from PROC MEANS

	_TRTNAM	BIGN	_IMPUTATION_	NB	MEANB	STDB	NV	MEANV	STDV	LABEL	ESTIMATE	STDERR
1	Treatment A	222	1	222	67.639703048	18.543853223	222	72.521502239	17.472059936	P1: Week 24; TRT: 1	4.6801	1.3535
2	Treatment A	222	2	222	67.466466093	18.862392895	222	70.628639731	17.97807418	P1: Week 24; TRT: 1	2.9400	1.2920
3	Treatment A	222	3	222	66.415047728	18.999528841	222	70.444187458	18.601435802	P1: Week 24; TRT: 1	3.6335	1.3533
4	Treatment A	222	4	222	66.838788003	18.933755662	222	70.803907265	19.200848462	P1: Week 24; TRT: 1	3.6908	1.3811
5	Treatment A	222	5	222	66.524958789	19.025194811	222	71.336788971	17.523955587	P1: Week 24; TRT: 1	4.3633	1.3449
6	Treatment A	222	6	222	67.013206753	18.774523193	222	70.46192564	17.616884323	P1: Week 24; TRT: 1	3.2280	1.2996

Results from PROC MIXED

	IMPUTATION	LABEL	ESTIMATE	STDERR	DF	TVALUE	PROBT	ALPHA	LOWER	UPPER	_BY_TRT	_1STTRT	_2NDTRT
1	1	P2: Week 24; TRT: 1 - 2	8.0335	2.3208	289	3.46	0.0006	0.05	3.4658	12.6013	TRT1_2	1	2
2	2	P2: Week 24; TRT: 1 - 2	4.4622	2.2132	286	2.02	0.0447	0.05	0.1060	8.8184	TRT1_2	1	2
3	3	P2: Week 24; TRT: 1 - 2	5.6605	2.3819	292	2.38	0.0181	0.05	0.9726	10.3484	TRT1_2	1	2
4	4	P2: Week 24; TRT: 1 - 2	4.1921	2.4069	288	1.74	0.0826	0.05	-0.5453	8.9295	TRT1_2	1	2
5	5	P2: Week 24; TRT: 1 - 2	5.0368	2.2596	290	2.23	0.0266	0.05	0.5895	9.4842	TRT1_2	1	2
6	6	P2: Week 24; TRT: 1 - 2	1.6045	2.1621	286	0.74	0.4586	0.05	-2.6511	5.8601	TRT1_2	1	2

CASE STUDY (Cont..)

POOLING THE RESULTS (Step 3)

```
proc mianalyze data=est_sum;
  by _trtnam;
  modeleffects estimate meanb meanv;
  stderr stderr stdb stdv;
  ods output ParameterEstimates=est0sum;
run;
```

	_TRTNAM	ESTIMATE	MEANB	MEANV	STDB	STDV	LOWER	UPPER	BIGN	NB	NV
1	Treatment A	4.146304	66.873930	71.375857	18.816351	17.704110	0.81327	7.4793	222	222	222
2	Treatment B	-0.339998	69.082885	67.676018	19.748969	19.359143	-4.90083	4.2208	74	74	74

```
proc mianalyze data=est_comp;
  modeleffects estimate;
  stderr stderr;
  ods output ParameterEstimates=est0comp;
run;
```

	NIMPUTE	ESTIMATE	STDERR	LOWER	UPPER	DF	TVALUE	PROBT	ALPHA	LABEL	_BY_TRT
1	50	4.486302	2.661546	-0.74058	9.713181	611.5	1.69	0.0924	0.05	P2: Week 24; TRT: 1-2	TRT1_2

CASE STUDY (Cont..)

Report with multiple imputation

Analysis of Change from Baseline in EORTC QLQ-C30 Global Health Status/QoL with Multiple Imputation Based on MAR at Week 24 (FAS Population)

Treatment	Baseline		Week 24		Change from Baseline at Week 24	
	N	Mean (SD)	N	Mean (SD)	N	LS Mean (95% CI) [†]
Treatment A	222	66.87 (18.816)	222	71.38 (17.704)	222	4.15 (0.81, 7.48)
Treatment B	74	69.08 (19.749)	74	67.68 (19.359)	74	-0.34 (-4.90, 4.22)
Pairwise Comparison					Difference in LS Means (95% CI)	p-Value
Treatment A vs. Treatment B					4.49 (-0.74, 9.71)	0.0924
[†] Based on cLDA model with the PRO scores as the response variable, and treatment by study visit interaction, stratification factors (prior auto-SCT (yes, no) and disease status after frontline therapy (primary refractory, relapsed less than 12 months, relapsed 12 months or more)) as covariates. For baseline and Week 24, N is the number of subjects in each treatment group with non-missing assessments at the specific time point; for change from baseline, N is the number of subjects in the analysis population in each treatment group. Database Cutoff Date: 16JAN2020						

Source: [adam-adsl; adplda]

CASE STUDY (Cont..)

Comparison

Analysis of Change from Baseline in EORTC QLQ-C30 Global Health Status/QoL at Week 24
(FAS Population)

Treatment	Baseline		Week 24		Change from Baseline at Week 24	
	N	Mean (SD)	N	Mean (SD)	N	LS Mean (95% CI) [†]
Treatment A	202	66.87 (18.982)	125	74.53 (16.943)	222	4.65 (-1.55, 7.75)
Treatment B	69	69.57 (19.742)	46	70.83 (18.151)	74	0.07 (-4.73, 4.88)
Pairwise Comparison					Difference in LS Means (95% CI)	p-Value
Treatment A vs. Treatment B					4.57 (-0.79, 9.94)	0.0943

Analysis of Change from Baseline in EORTC QLQ-C30 Global Health Status/QoL with Multiple Imputation Based on MAR at Week 24
(FAS Population)

Treatment	Baseline		Week 24		Change from Baseline at Week 24	
	N	Mean (SD)	N	Mean (SD)	N	LS Mean (95% CI) [†]
Treatment A	222	66.87 (18.816)	222	71.38 (17.704)	222	4.15 (0.81, 7.48)
Treatment B	74	69.08 (19.749)	74	67.68 (19.359)	74	-0.34 (-4.90, 4.22)
Pairwise Comparison					Difference in LS Means (95% CI)	p-Value
Treatment A vs. Treatment B					4.49 (-0.74, 9.71)	0.0924

CONCLUSION

- As explained in our case study, the pattern of the missing data is identified as 'Arbitrary' using PROC MI and we decided to use the FCS REG imputation method since variable type is 'continuous' and has mixed covariates.
- Report created with multiple imputation could be used as a supporting report as part of the sensitivity analysis, which can be requested by the submission agencies or the internal committees.
- As per our case study results, we observe that the statistical inferences in report with MI are close to the statistical inferences from the main analysis Report (complete case analysis).

THANK YOU



MERCK



Q & A

