



INTEGO
GROUP

Multiple imputation as a valid way of dealing with missing data

Vadym Kalinichenko, Intego Group, Kharkiv, Ukraine



Amsterdam 2019





Introduction

Missing data and Multiple imputation



Monotone missing pattern

Y1	Y2	Y3	Y4
X	X	X	O
X	X	O	O
X	O	O	O

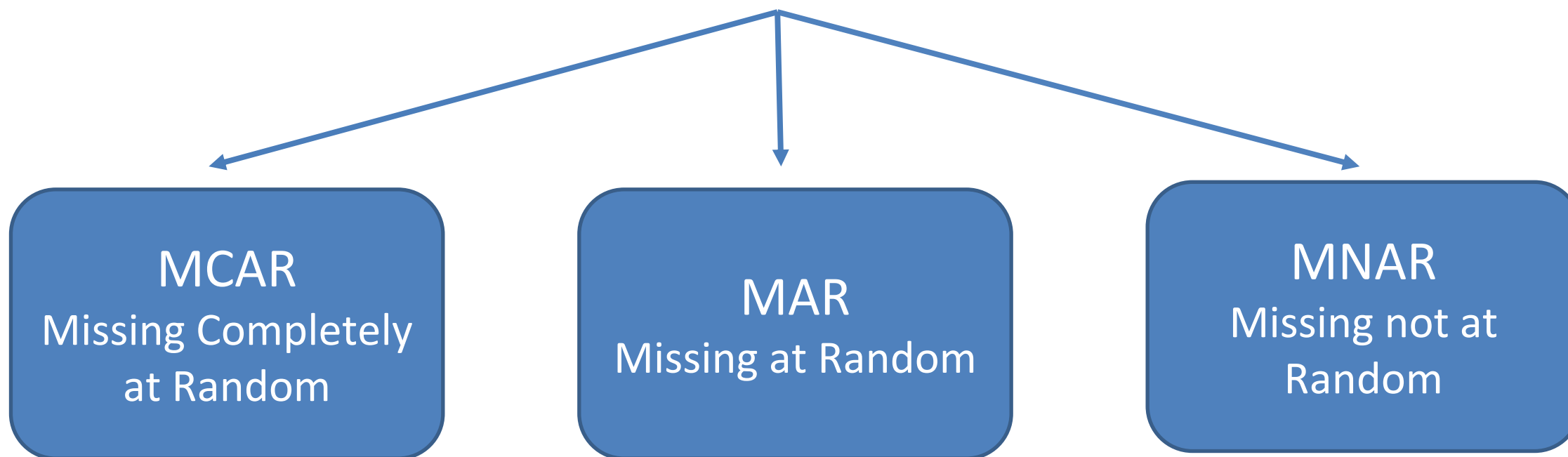
Non-monotone missing pattern

Y1	Y2	Y3	Y4
X	X	X	X
X	X	O	O
X	O	X	O

Crucial difference in MI for non-monotone vs. monotone data:

- **non-monotone:**
 - All the data imputed together (multivariate response)
 - Same distributional assumptions for categorical and continuous data
- **monotone:**
 - Using one-variable-at-a-time imputation
 - Different distributional assumptions for continuous and categorical data (as it would normally be done)

Missing data mechanisms



Missing Completely at Random, MCAR

- No relationship between the missingness of the data and any values, observed or missing
- Missing data points are a random subset of the data
- There is nothing systematic going on that makes some data more likely to be missing than others

Missing at Random, MAR

- There is a systematic relationship between the propensity of missing values and the observed data, but not the missing data
- The missingness mechanism does not depend on the unobserved data

Missing Not at Random, MNAR

- There is a relationship between the propensity of a value to be missing and its values



Multiple Imputation Implementation



Multiple Imputations (MI)

Introduced by Rubin (1976, 1987) for missing data in surveys

Main features of standard MI:

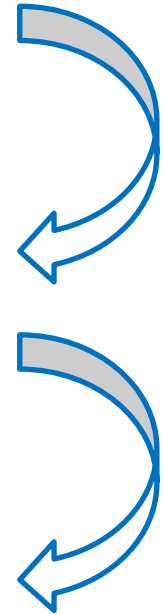
- Imputed values are randomly drawn from a distribution estimated based on non-missing data (imputation model)
- Missing values are replaced with multiple “plausible” values (multiple imputed datasets are produced)
- Each imputed dataset is analyzed separately, and then the results are combined taking into account uncertainty (variability) of different imputations

Multiple imputation steps

The missing data are filled in **m** times to generate **m** complete data sets

The **m** complete data sets are analyzed by using standard procedures

The results from the **m** complete data sets are combined for the inference



SUBJID	TRTN	TRT	Baseline	Week 1	Week 2	Week 4	Week 6	Week 8
1001	1	Treatment A	xx	xx	xx	xx	o	o
1002	2	Treatment B	xx	xx	xx	xx	xx	xx
1003	1	Treatment A	xx	xx	xx	o	xx	xx
1004	2	Treatment B	xx	xx	xx	xx	xx	o
1005	1	Treatment A	xx	xx	xx	xx	o	xx
1006	2	Treatment B	xx	xx	xx	o	o	o

Primary endpoint: change from baseline in sleep quality (daily values on a scale from 1 to 10, averaged) at Week 8

Primary analysis: ANCOVA with the baseline as a covariate (the most commonly used covariate)

```
/*examine missing patterns*/
proc mi data = inds_tr nimpute = 0;
    var vis1 - vis6;
run;
```

Missing Data Patterns														
Group	vis1	vis2	vis3	vis4	vis5	vis6	Freq	Percent	Group Means					
									vis1	vis2	vis3	vis4	vis5	vis6
1	X	X	X	X	X	X	348	69.60	5.4339	5.4683	5.5260	5.4550	5.6712	5.6233
2	X	X	X	X	X	.	19	3.80	5.1526	6.614	5.1794	4.6033	5.8673	.
3	X	X	X	X	.	X	21	4.20	5.8809	4.7800	4.7276	5.6033	.	6.1085
4	X	X	X	X	.	.	3	0.60	6.4000	5.9700	6.7100	5.1100	.	.



Pattern of Missingness	Type of Imputed Variable	Recommended Methods
Monotone	Continuous	● Parametric: - Regression method (MONOTONE REG) - Predictive mean matching (MONOTONE REGPMM) ● Non-parametric: - Propensity score (MONOTONE PROPENSITY)
Monotone	Categorical (Binary/Ordinal)	Logistic regression (MONOTONE LOGISTIC)
Monotone	Categorical (Binary/Ordinal)	Discriminant function method (MONOTONE DISCRIM)
Non-monotone	Continuous	Markov Chain Monte Carlo (MCMC) - Full imputation - Partial - to obtain monotone pattern
Non-monotone	Categorical	No recommended method available

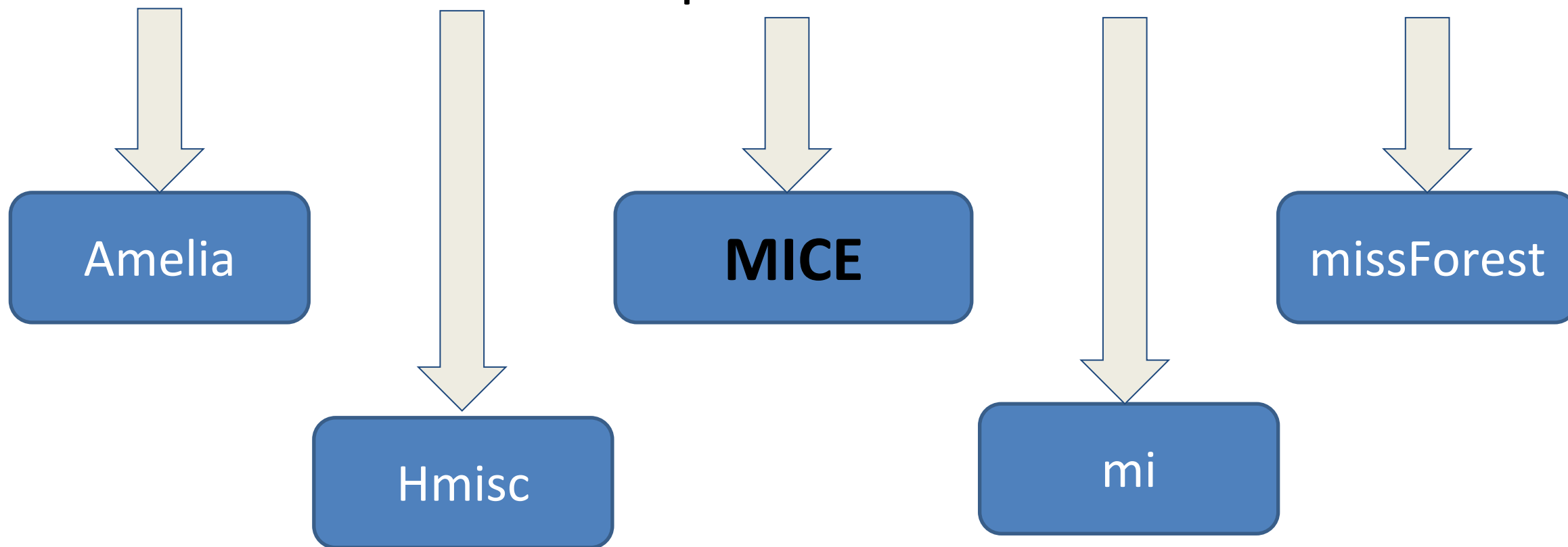
```
/*partial imputation to get monotone missing pattern*/  
proc mi data = inds_tr out = data_mono nimpute = 5;  
    var vis1 - vis6;  
    mcmc impute = monotone chain = multiple;  
    by trtn;  
run;  
  
/*perform multiple imputation of the remaining data*/  
proc mi data = data_mono out = mono_imp_reg nimpute = 1;  
    by _Imputation_;  
    var trtn vis1 - vis6;  
    class trtn;  
    monotone regression;  
run;
```



```
/*ANCOVA for each imputed dataset*/  
proc mixed data = mono_imp_reg;  
  class trtn;  
  model chg_6 = trtn vis1 / solution;  
  lsmeans trtn / diff = control('1') cl;  
  ods output diffs = lsdiffs lsmeans = lsm solutionf = parms;  
  by _Imputation_; /*perform analysis in each imputed dataset*/  
run;  
  
/*combining estimates with PROC MIANALYZE*/  
proc mianalyze parms(classvar = full) = parms;  
  class trtn;  
  modeleffects Intercept trtn vis1;  
  ods output ParameterEstimates = combined_parms;  
run;
```



5 most commonly used R packages for missing value imputation

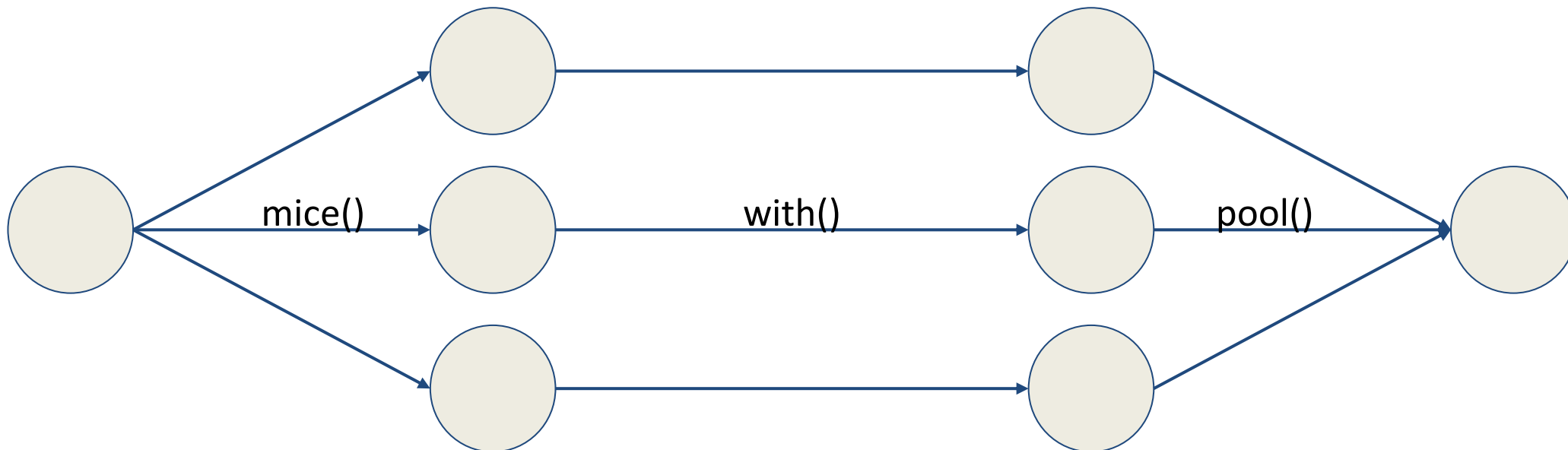


incomplete data

imputed data

analysis results

pooled results



data frame

mids

mira

mipo

```
# using mice for looking at missing data pattern
> md.pattern(input)

# imputation method for change from baseline
> meth["chg6"] <- "~ I(vis6 - vis1)"

# mice call
> imputed_Data <- mice(inds_tr3, m=5, maxit = 50, meth = meth)
```

subjid	trtn	vis1	vis2	vis3	vis4	vis5	vis6	chg6
""	""	"pmm"	"pmm"	"pmm"	"pmm"	"pmm"	"pmm"	"~ I(vis6 - vis1)"

```
# create An Analysis Of Variance Model  
> fit <- with(data = imputed_Data, exp = lm(chg6 ~ vis1 + trtn))
```

TRTN	Treatment group	Categorical Variable
VIS1	Baseline	Continuous Predictor Variable
CHG6	Change from baseline	Response Variable

```
#combine results of all 5 models  
> summary(pool(fit), conf.int = TRUE)
```

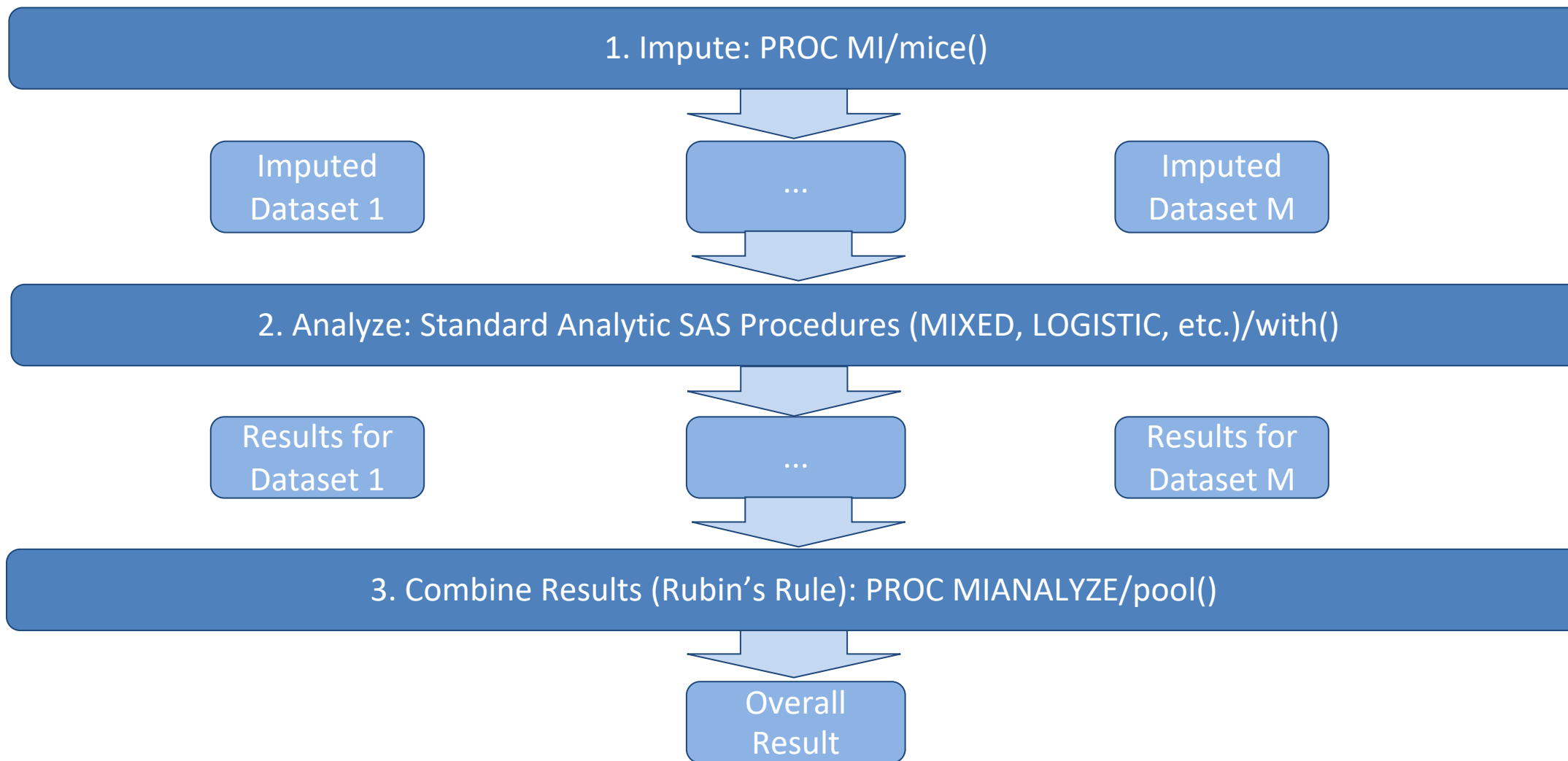




Summary



Multiple Imputation





Conclusion



Advantages of Multiple Imputation:

- Results in unbiased estimates, providing more validity than other approaches to missing data
- Uses all available data, preserving sample size and statistical power
- Allows to use standard complete-data methods
- Increases efficiency of estimation
- Provides valid inferences (for variance estimators) under an assumed model for nonresponse
- May be used with standard statistical software



THANK YOU

Vadym Kalinichenko
vadim.kalinichenko@intego-group.com,
Kharkiv, Ukraine
www.intego-group.com

