Combining Multiple Imputation and Inverse Probability Weighting

Shaun Seaman[1], Ian White[1], Andrew Copas[2,3]

[1] MRC Biostatistics Unit, Cambridge, UK

[2] MRC Clinical Trials Unit, London, UK

[3] Dept. Primary Care & Population Sciences, UCL, London, UK

Overview

- Inverse probability weighting (IPW) and multiple imputation (MI)
- Why combine IPW and MI?
- Are Rubin's Rules valid for IPW/MI?
- Application to 1958 British Birth Cohort
- Conclusion

IPW and MI

- Let $\boldsymbol{D}$ be variables in analysis model and $\boldsymbol{\theta}$ be parameters
- Three methods when missing values in $\boldsymbol{D}$:

  Complete Cases: Just analyse individuals with complete data on $\boldsymbol{D}$

  IPW: Like Complete Cases but weight by inverse of $P(\text{Comp Case})$

  MI: Stochastically fill in missing values using observed data

  Create multiple complete datasets

  Apply complete-data estimator to each

  Combine estimates (Rubin's Rules)

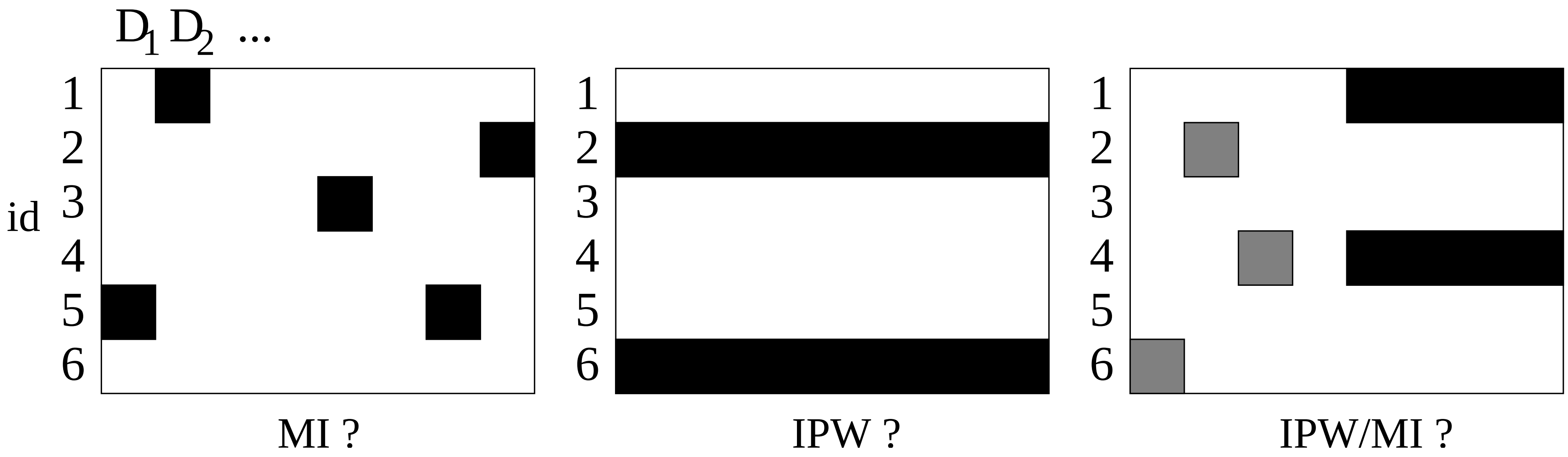
  (There are also doubly robust methods)

- IPW needs correct missingness model

  MI needs correct imputation model

- IPW also used for surveys with unequal sampling fractions

Why combine IPW and MI?



IPW/MI also for surveys with unequal sampling fractions and missing data

IPW/MI has been used (e.g. Priebe et al., 2004; Stansfeld et al., 2008)

Are Rubin's Rules valid for IPW/MI?

Rubin's Rules for MI:

How to combine parameter estimates and variance estimates from multiple imputed datasets to get overall estimate $\hat{\boldsymbol{\theta}}$ and estimate, $\hat{\boldsymbol{V}}$, of its variance

When complete-data estimator is mle, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{V}}$ are asymptotically unbiased (as sample size $\rightarrow \infty$)

Nielsen (2003) gives example of asymptotically biased $\hat{\boldsymbol{V}}$ when not mle

In IPW/MI, complete-data estimator is IPW estimator

Kim et al. (2006) showed Rubin's Rules valid for IPW/MI when $\boldsymbol{\theta}$ is a mean

We proved valid when $\boldsymbol{\theta}$ are parameters of a linear regression and missing outcome is imputed

Our simulation studies suggest also valid when

- linear regression with imputed covariate
- logistic regression with imputed outcome or covariate

Application to 1958 British Birth Cohort

- 17638 individuals born in UK during one week in 1958
  16334 still alive at age 45
  9014 (55%) participated in biomedical survey
- Thomas et al. (2007) regressed blood glucose (high/normal) at age 45 on BMI and waist size at age 45 and variables measured at birth
  7518 had observed glucose, BMI and waist size
  Thomas et al. imputed missing birth variables using ICE
- But are 7518 representative of 16334?
  We used IPW/MI
- Need model for P(glucose, BMI & waist size complete)
  Used predictors measured at birth, age 7 and 11 (e.g. birth weight, class)
  Disadvantaged less likely to be complete for glucose, BMI & waist size
  Weights: mean$= 2.5$; 95th centile$= 5.2$; max$= 23.1$

Predictor of	Thomas		IPW/MI		All MI	
High Glucose	log OR	SE	log OR	SE	log OR	SE
short gestation	0.45	0.22	0.46	0.23	0.44	0.20
pre-eclampsia	**0.47**	0.26	**0.55**	0.27	**0.47**	0.25
smoke in pregncy	0.02	0.14	0.04	0.14	0.04	0.14
mother overwt	**0.28**	0.15	**0.35**	0.15	**0.18**	0.12
manual at birth	0.37	0.17	0.44	0.18	0.39	0.17
low birth weight	−0.30	0.09	−0.30	0.09	−0.32	0.09
BMI age 45	0.04	0.02	0.02	0.02	0.03	0.02
waist (cm) age 45	0.07	0.01	0.07	0.01	0.07	0.01

Stronger relation between glucose and pre-eclampsia/mother overwt in disadvantaged than in advantaged

When use All MI, relation similar in individuals with imputed glucose

Ordinary IPW uses only 5673 (75% of 7518) with complete birth data

Conclusion

- IPW/MI more efficient than IPW
- MI more efficient still
- But some researchers not confident about imputing large blocks of missing data and may prefer IPW/MI
- Rubin's Rules valid for IPW/MI
- IPW can fail too when lots of missing data, but arguably missingness model easier to assess than imputation model