

# Missingness and Bayesian modeling

BMTRY 763

# Missingness

- Often we find missing data in health studies
- Bayesian modeling has natural ways to deal with missingness
- Missing completely at Random (MCAR):
  - Mechanism completely unrelated to the data (given covariates) observed or otherwise
- Missing at random (MAR)
  - Mechanism unrelated to unobserved predictors but related to observed data
- Non-random Missingness (NRM)
  - Some mechanism has induced the missingness and it must be modelled

# Outcome missingness

- For MCAR or MAR, we can use imputation to estimate the missing data
- It is natural to use **Data Augmentation** within MCMC for this
- This uses the predictive distribution of the data

1) sample  $\theta$  from  $Post(\theta | x, y^{obs}, y^{pred}) \propto [y^{obs}, y^{pred} | \theta, x]. Pr(\theta)$

2) generate  $[y^{pred} | y^{obs}, x, \theta]$  from the likelihood

3) repeat 1 and 2

# Predictor missingness

- Predictor missingness is different
- Predictors are usually assumed to be fixed (given) and so there is no distribution associated with them.
- Essentially we have to assume a prior distribution for the predictor
- For example, a PM<sub>2.5</sub> measure (μg/l) is missing and we assume  $PM_{2.5} \sim N(\mu, \tau)$
- This will treat any missing data in PM<sub>2.5</sub> as parameters to be estimated
- This can be done within MCMC

# OpenBUGS and missingness

- Missing data is denoted as NA
- Outcome missingness is *\*automatically\** imputed within MCMC
- Predictor missingness requires a prior distribution
  
- Missing data must be initialised. This can be done using 'load inits' or 'Gen inits'