Bayesian measurement error models for air pollution and health studies

Erica Ponzi, Paolo Vineis, Marta Blangiardo

Epidemiology, Biostatistics and Prevention Institute, University of Zürich
Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom

Bayesian Biostatistics, 22nd June 2018

Environmental epidemiology

- Effect of environmental exposures on health conditions
- Used in the definition of health care policies and interventions
- Exposures: air and water pollutants, individual and ambient concentrations
- Risk assessment: health conditions and development of diseases
- High throughput molecular data: omics data (metabolome, genome, transcriptome,...)

Environmental epidemiology and the exposome

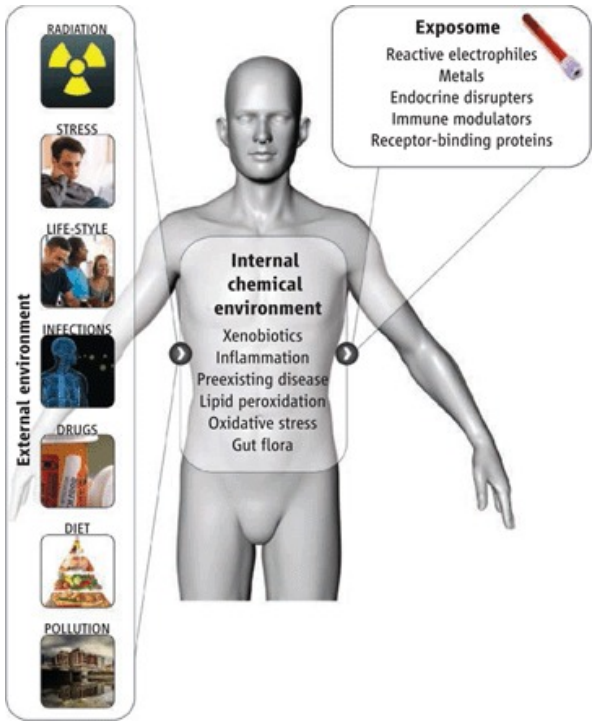


Figure: Overview of the exposome[1]

[1] **Rappaport and Smith, 2010.**

Measurement error in environmental exposures

- Particulary challenging on Traffic Related Air Pollution (TRAP)
- Different sources of ME:
  - Instrument imprecision
  - Difference between ambient concentration and individual exposure
  - Difference between average personal exposure (at home/work address,...) and actual individual exposure
- Mostly classical error

$$\mathbf{W} = \mathbf{X} + \mathbf{U}$$

$$\mathbf{U} \sim \mathrm{N}(0, \sigma^2_{em}\mathbf{I})$$

Bayesian measurement error models

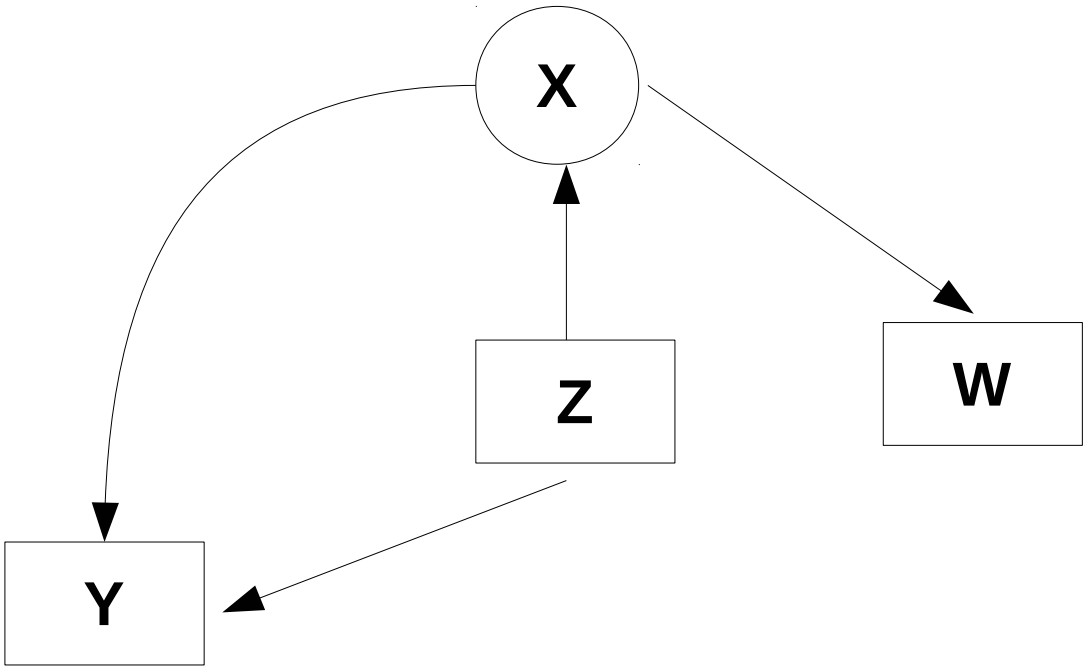


Figure: Graphical structure of the error models: $\boldsymbol{X}$ is the true covariate, $\boldsymbol{W}$ is the error-prone proxy, $\boldsymbol{Z}$ is an exact observable covariate and $\boldsymbol{Y}$ is the response[2].

[2] **Muff et al, 2015**.

Bayesian measurement error models

- Level 1: regression model $\boldsymbol{Y}|\boldsymbol{v}, \boldsymbol{\theta_1}$ of the response variable $\boldsymbol{Y}$ as if $\boldsymbol{X}$ were observed, depending on unknown parameters $\boldsymbol{v}$ and hyperparameters $\boldsymbol{\theta_1}$
- Level 2: latent error model $\boldsymbol{X}|\boldsymbol{W}, \boldsymbol{\theta_2}$ depending on hyperparameters $\boldsymbol{\theta_2}$
- Level 3: priors defined for the hyperparameters $(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$

Oxford Street II Study

- Randomized cross-over trial
- Part of Exposomics consortium
- 60 volunteers walking in Hyde Park (traffic free) and Oxford Street (highly polluted)
- 3 measurements per walking session: 6 measurements per individual
- Volunteers divided in three groups: healthy (20), COPD (20), IHD (20)
- Information on age, sex, BMI, caffeine intake, ...
- TRAP assessment: PM_{10}, $PM_{2.5}$, $CBLK$, UFP and NO_2
- Omics measurements: metabolomics, transcriptomics and adductomics

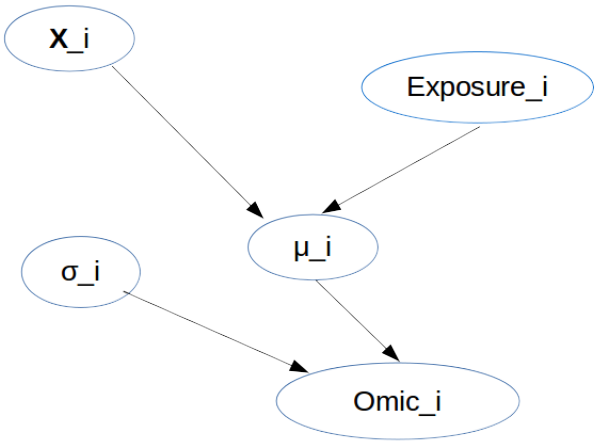
Model on TRAP and metabolomic signals

- On each pollutant
- 5749 metabolic features: selected features based on previous studies
- Fixed effects: sex, age, BMI, disease, caffeine intake

$$m_{ij} \sim \mathrm{N}(\mu_{ij}, \sigma_e^2)$$

$$\mu_{ij} = \alpha + \boldsymbol{\beta}^\top \boldsymbol{X}_{ij} + \beta_{expo} E_{ij} + ID_i + T * L_{ij}$$

- $\boldsymbol{E}$ = exposure to pollutant

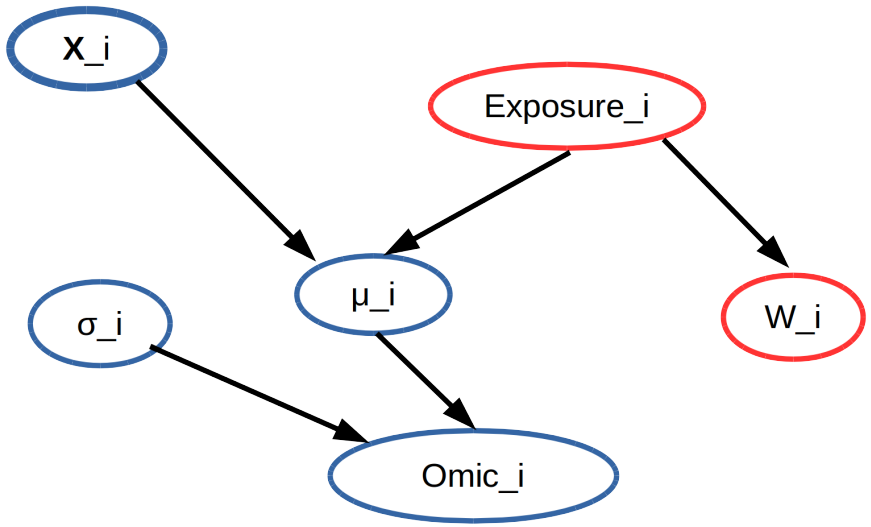


Measurement error on TRAP model

- ME on exposures

$$
\begin{aligned}
m_{ij} &\sim \mathrm{N}(\mu_{ij}, \sigma_e^2) \\
\mu_{ij} &= \alpha + \boldsymbol{\beta}^\top \boldsymbol{X}_{ij} + \beta_{expo} E_{ij} + ID_i + T * L_{ij} \\
W_{ij} &= E_{ij} + U_{ij} \\
U_{ij} &\sim \mathrm{N}(0, \sigma_U^2)
\end{aligned}
$$

- Normal priors to regression coefficients
- Inverse gamma priors to precisions of random effects



Dependency across omic signals

- Metabolic features are not independent in the same individual
- Multivariate response in the Bayesian model

$$\boldsymbol{m} \sim \boldsymbol{MVN}(\mathbf{0}, \Sigma)$$

$$\begin{bmatrix} \boldsymbol{m}(1) \\ \boldsymbol{m}(2) \\ \dots \\ \boldsymbol{m}(n) \end{bmatrix} = \boldsymbol{\mu} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta_{expo}}\boldsymbol{E} + \boldsymbol{Z}\boldsymbol{r}$$

$$\boldsymbol{W} = \boldsymbol{E} + \boldsymbol{U}$$

- Wishart prior for Σ

Dependency across different exposures

- TRAP exposures are not independent on the same individual and location
- Correlated exposures in the Bayesian model

$$\boldsymbol{m} \sim \boldsymbol{MVN}(\mathbf{0}, \Sigma)$$

$$\begin{bmatrix} \boldsymbol{m}(1) \\ \boldsymbol{m}(2) \\ \dots \\ \boldsymbol{m}(n) \end{bmatrix} = \mu + \boldsymbol{X}\beta + \beta_{\boldsymbol{expo_1}}\boldsymbol{E_1} + \beta_{\boldsymbol{expo_2}}\boldsymbol{E_2} + \dots + \beta_{\boldsymbol{expo_n}}\boldsymbol{E_n} + \boldsymbol{Z}\boldsymbol{r}$$

$$\boldsymbol{W} = \boldsymbol{E} + \boldsymbol{U}$$

Naive vs error-corrected BHM

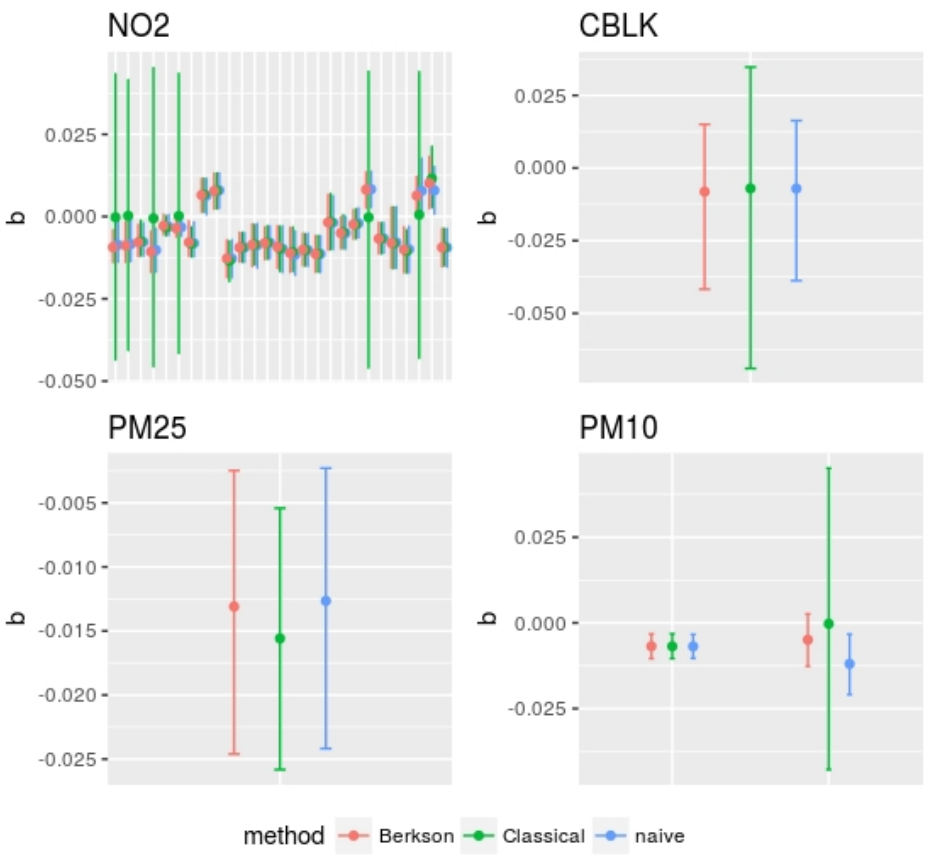


Figure: Regression coefficients with and without error modeling in JAGS. A classical and a Berkson error are assumed.

Multi-omics BHM

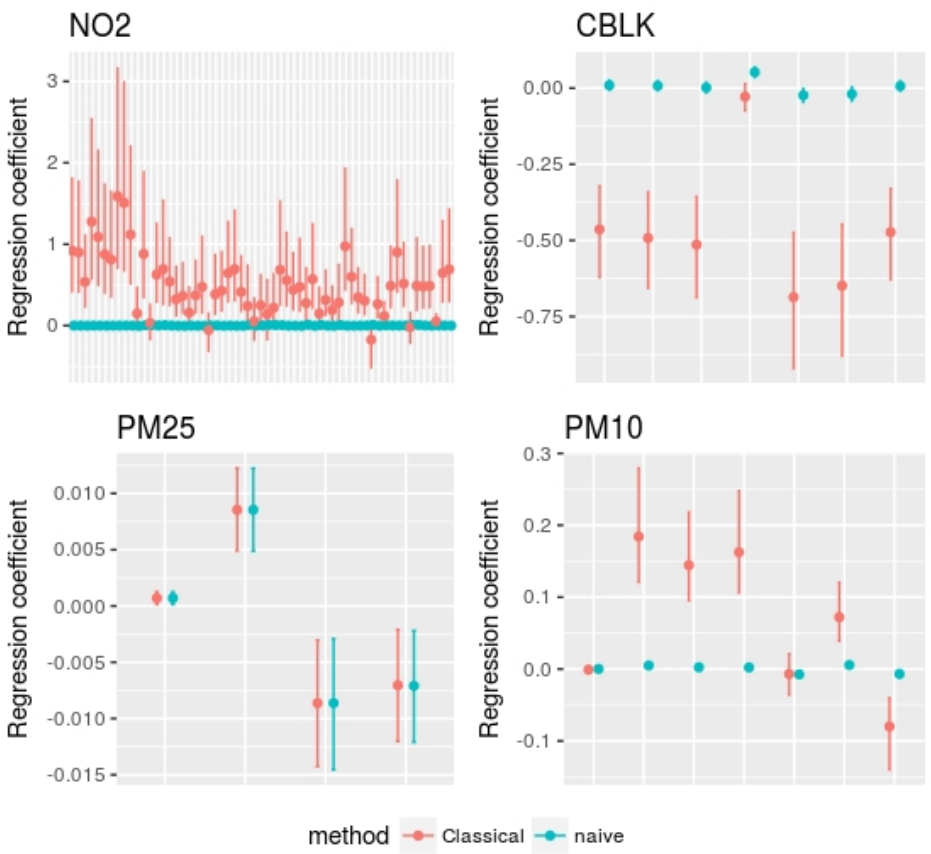


Figure: Regression coefficients with and without classical error modeling in `JAGS`. A correlated structure among omic signals is assumed.

Correlated exposure BHM

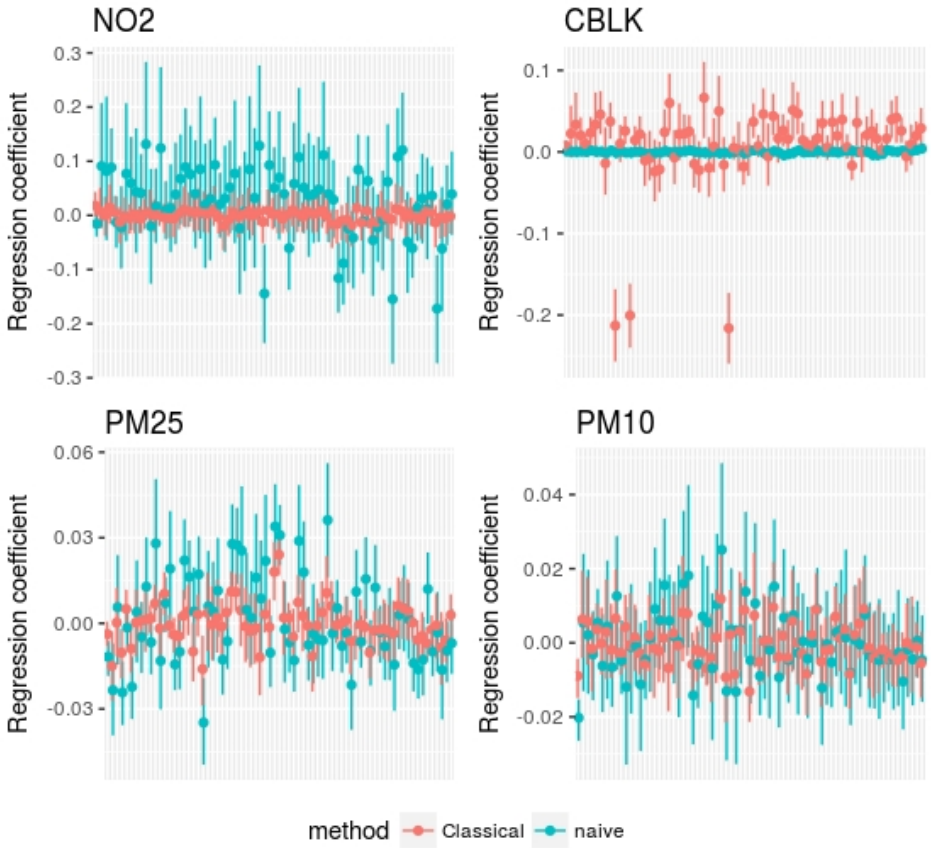


Figure: Regression coefficients with and without classical error modeling in `JAGS`. Dependencies among omic signals and among different TRAP exposures are assumed.

Conclusions and future work

- Effect of classical error
- No effect of Berkson error

Conclusions and future work

- Effect of classical error
- No effect of Berkson error
- BHM flexibility: possible to account for dependencies
- Among omics: allows to identify more signals, no multiple testing problems
- Among pollutants
- Classical ME included in both models

Conclusions and future work

- Effect of classical error
- No effect of Berkson error
- BHM flexibility: possible to account for dependencies
- Among omics: allows to identify more signals, no multiple testing problems
- Among pollutants
- Classical ME included in both models
- Implemented in `JAGS`
- Working on `INLA` implementation: some difficulties with multiple omics, faster for single omic models

Thank you!