

WinBUGS code for imputation of missing data in Mendelian randomization studies

Stephen Burgess *

MRC Biostatistics Unit Technical report 2011/1

This is WinBUGS code for methods described in the manuscript “Missing data methods in Mendelian randomization studies with multiple instruments”, which has been recently accepted (conditional on technical review) by the American Journal of Epidemiology. The manuscript is currently available from the author.

Bayesian method incorporating correlation

```
model {
  alpha0 ~ dnorm(0, 0.000001) # priors for regression parameters
  beta ~ dnorm(0, 0.000001)
  beta0 ~ dnorm(0, 0.000001)
  xtau <- pow(xsd, -2) # priors for variance parameters
  xsd ~ dunif(0, 20)
  ytau <- pow(ysd, -2)
  ysd ~ dunif(0, 20)
  tauy <- ytau/(1-pow(rho,2)) # conditional precision given x[i]
  rho ~ dunif(-1, 1) # prior for correlation
  for(k in 1:K) { # index across IVs
    alpha[k] ~ dnorm(0, 0.000001) # prior for IV effects
  }
  for (i in 1:N) { # index across individuals
    xi[i] <- alpha0 + inprod(alpha[1:K], gene[i, 1:K])
    # phenotype regression in additive model across IVs
    x[i] ~ dnorm(xi[i], xtau) # normal model of phenotype
    muy[i] <- eta[i] + sqrt(xtau/ytau)*rho*(x[i]-xi[i])
    # conditional mean given x[i]
    y[i] ~ dnorm(muy[i], tauy) # normal model of outcome
    eta[i] <- beta0 + beta * xi[i] # unconditional mean of outcome
  } # beta is causal parameter of interest
}
```

*Correspondence to: stephen.burgess@mrc-bsu.cam.ac.uk

Multiple imputations method

```
model {
  alpha0 ~ dnorm(0, 0.000001)
  beta   ~ dnorm(0, 0.000001)
  beta0  ~ dnorm(0, 0.000001)
  xsig   ~ dunif(0, 20)
  xtau   <- pow(xsig, -2)
  ysig   ~ dunif(0, 20)
  ytau   <- pow(ysig, -2)
  tauy   <- ytau/(1-pow(rho,2))
  rho    ~ dunif(-1, 1)
  r      ~ dpick(1,10) # r indexes imputations
  for (j in 1:K) {
    alpha[k] ~ dnorm(0, 0.000001)
  }
  for (i in 1:N) {
    xi[i] <- alpha0 + alpha[1]*gene[i, 1, r] + alpha[2]*gene[i, 2, r]
              + alpha[3]*gene[i, 3, r] # phenotype regression uses current imputation
    x[i] ~ dnorm(xi[i], xtau)
    muy[i] <- eta[i] + sqrt(xtau/ytau)*rho*(x[i]-xi[i])
    y[i] ~ dnorm(muy[i], tauy)
    eta[i] <- beta0 + beta * xi[i]
  }
}
```

SNP imputation method

```
model {
  alpha0 ~ dnorm(0, 0.000001)
  beta   ~ dnorm(0, 0.000001)
  beta0  ~ dnorm(0, 0.000001)
  xsig   ~ dunif(0, 20)
  xtau   <- pow(xsig, -2)
  ysig   ~ dunif(0, 20)
  ytau   <- pow(ysig, -2)
  tauy   <- ytau/(1-pow(rho,2))
  rho    ~ dunif(-1, 1)
  for (k in 1:K) {
    alpha[k] ~ dnorm(0, 0.000001)
  }
  for (i in 1:N) {
    for (k in 1:K) {
      gene[i, k] ~ dcat(geneprobs[i, k, 1:3])
    } # geneprobs are posterior probabilities from genetic imputation
  }
}
```

```

xi[i] <- alpha0 + inprod(alpha[1:K], gene[i, 1:K])
x[i] ~ dnorm(xi[i], xtau)
muy[i] <- eta[i] + sqrt(xtau/ytau)*rho*(x[i]-xi[i])
y[i] ~ dnorm(muy[i], tauy)
eta[i] <- beta0 + beta * xi[i]
}
}

```

Multivariate latent variable model

```

model {
mu[1:K] ~ dnorm(mu0[1:K], Sigma0[1:K, 1:K])
Sigma[1:K, 1:K] ~ dwish(Sigma1[1:K, 1:K], K)
# priors for the haplotype distributions:
alpha0 ~ dnorm(0, 0.000001)
beta ~ dnorm(0, 0.000001)
beta0 ~ dnorm(0, 0.000001)
xsig ~ dunif(0, 20)
xtau <- pow(xsig, -2)
ysig ~ dunif(0, 20)
ytau <- pow(ysig, -2)
for (k in 1:K) {
alpha[k] ~ dnorm(0, 0.000001)
}
for (i in 1:N) {
psi1[i, 1:K] ~ dnorm(mu[1:K], Sigma[1:K, 1:K])
psi2[i, 1:K] ~ dnorm(mu[1:K], Sigma[1:K, 1:K])
# psi1 and psi2 are drawn from the same multivariate distribution
# and represent the two haplotypes
for (k in 1:K) {
gene[i, k] ~ dgene.aux(psi1[i, k], psi2[i,k])
} # gene values when known are entered as data, when unknown as NA
# missing data values are imputed from the multivariate haplotype model
xi[i] <- alpha0 + inprod(alpha[1:K], gene[i, 1:K])
x[i] ~ dnorm(xi[i], xtau)
muy[i] <- eta[i] + sqrt(xtau/ytau)*rho*(x[i]-xi[i])
y[i] ~ dnorm(muy[i], tauy)
eta[i] <- beta0 + beta * xi[i]
}
}

```