The racial disparity conundrum of HIV risk among men who have sex with men: Bayesian approaches for correcting misclassification and residual confounding

Goldstein ND, Burstyn I, Welles SL

eAppendix

Additional Methods

Specification of Bayesian Correction for Misclassification
Purported misclassification of MSM behavior was corrected via Bayesian adjustment previously implemented by a three-equation approach, based upon previous work (Goldstein, 2015):

(Eq. 1) *Outcome model*
$$logit\{\Pr(Y=1|X_1, \ldots X_5)\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$
(Eq. 2) *Exposure model*
$$logit\{r=\Pr(X_1=1|X_2, \ldots X_5)\} = \alpha_0 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5$$
(Eq. 3) *Misclassification model*
$$\Pr(X^*_1=1|Y,X_1) = \begin{cases} \text{if } Y = 0: r\text{Sn}_0 + (1 - r)(1 - \text{Sp}_0) \\ \text{if } Y = 1: r\text{Sn}_1 + (1 - r)(1 - \text{Sp}_1) \end{cases}$$

Eq. 1 represents the adjusted log odds of self-reported HIV infection (Y) given the true MSM behavior (X_1), adjusted for potential confounding (X_2-X_5) by "being in a recent relationship", "having a history of sexually transmitted diseases", "having been sexually abused", and "recent drug use"; eq. 2 represents the log odds of true MSM behavior, given the aforementioned covariates; and eq. 3 relates the observed apparent MSM behavior (X^*_1) to the conditional probability of true MSM behavior (r) by the Sn and Sp of the exposure predictor (allowing for differential misclassification), derived from BAAMHS.

Prior distributions were required for all α's, β's, Sn and Sp. The priors for Sn, Sp, β_1 are described separately below. The priors for all other parameters are described in the sections specific to latent confounding and effect modification.

Prior Distributions for Sn, Sp, β_1
Sn, Sp: Sensitivity (Sn) and specificity (Sp) of male sexual partner gender as an indicator of MSM behavior (anal intercourse with another man) was derived from BAAMHS (Goldstein, 2015). For the priors on Sn and Sp, we used Beta distributions and computed the parameters (a and b) of the Beta distributions from the Sn/Sp 2x2 tables as follows. For sensitivities, a=the number of true positives and b=the number of false negatives; and for specificities, a=the number of true negatives and b=the number of false positives. We allowed for differential misclassification and thus had seperate Sn/Sp estimates for HIV positive and negative individuals.

β_1: β_1 represents the log odds of MSM behavior (anal intercourse with another man) associated with self-reported HIV positivity, derived from BAAMHS. The β_1 distribution was estimated via multivariable logistic regression with the mean equivalent to the log odds and the corresponding variance. The adjusted OR in BAAMHS of HIV infection associated with anal intercourse was 6.1 (95% CI: 2.5, 14.8). As we used hierarchical priors for modeling the residual confounding, the variance for this distribution was sampled from ξ, forming the basis of the prior $\beta_1 \sim \text{Normal}(\log(6.1), \xi)$.

Specification for Bayesian Correction for Latent Confounding: Stratified Models
An unmeasured confounder, U, was added to the exposure and outcome models, yielding:
(Eq. 4) *Outcome model w/ unmeasured covariate*
$$logit\{\Pr(Y=1|X_1, \ldots X_5, U_i)\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 U_i$$
(Eq. 5) *Exposure model w/ unmeasured covariate*
$$logit\{r=\Pr(X_1=1|X_2, \ldots X_5, U_i)\} = \alpha_0 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5 + \alpha_6 U_i$$

Priors for β_0, β_2, … β_6 and α_0, α_2, … α_6 used hierarchical non-informative Gaussian priors with zero mean and variance (ξ) sampled from the Gamma distribution: $\xi \sim \text{Gamma}(10,(\log(6)/2)^2)$. Using hierarchical priors allows modeling of latent confounding based upon known confounding from the included covariates; that is, if there is strong confounding from what is known, there is possible strong confounding from what is not known, and vice-versa for weak effects.

Specification for Bayesian Correction for Latent Confounding: All Men Models

Eqs. 4 & 5 were expanded to include a racial group term as follows:

(Eq. 6) *Outcome model w/ unmeasured covariate and interaction term*

$$logit\{\Pr(Y=1|X_1, \ldots X_6, U_i)\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 U_i + \beta_7 X_6$$

(Eq. 7) *Exposure model w/ unmeasured covariate and interaction term*

$$logit\{r=\Pr(X_1=1|X_2, \ldots X_6, U_i)\} = \alpha_0 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5 + \alpha_6 U_i + \alpha_7 X_6$$

Where X_6 represents confounding by racial group. Priors for β_7 and α_7 used hierarchical non-informative Gaussian priors with zero mean and ξ variance.

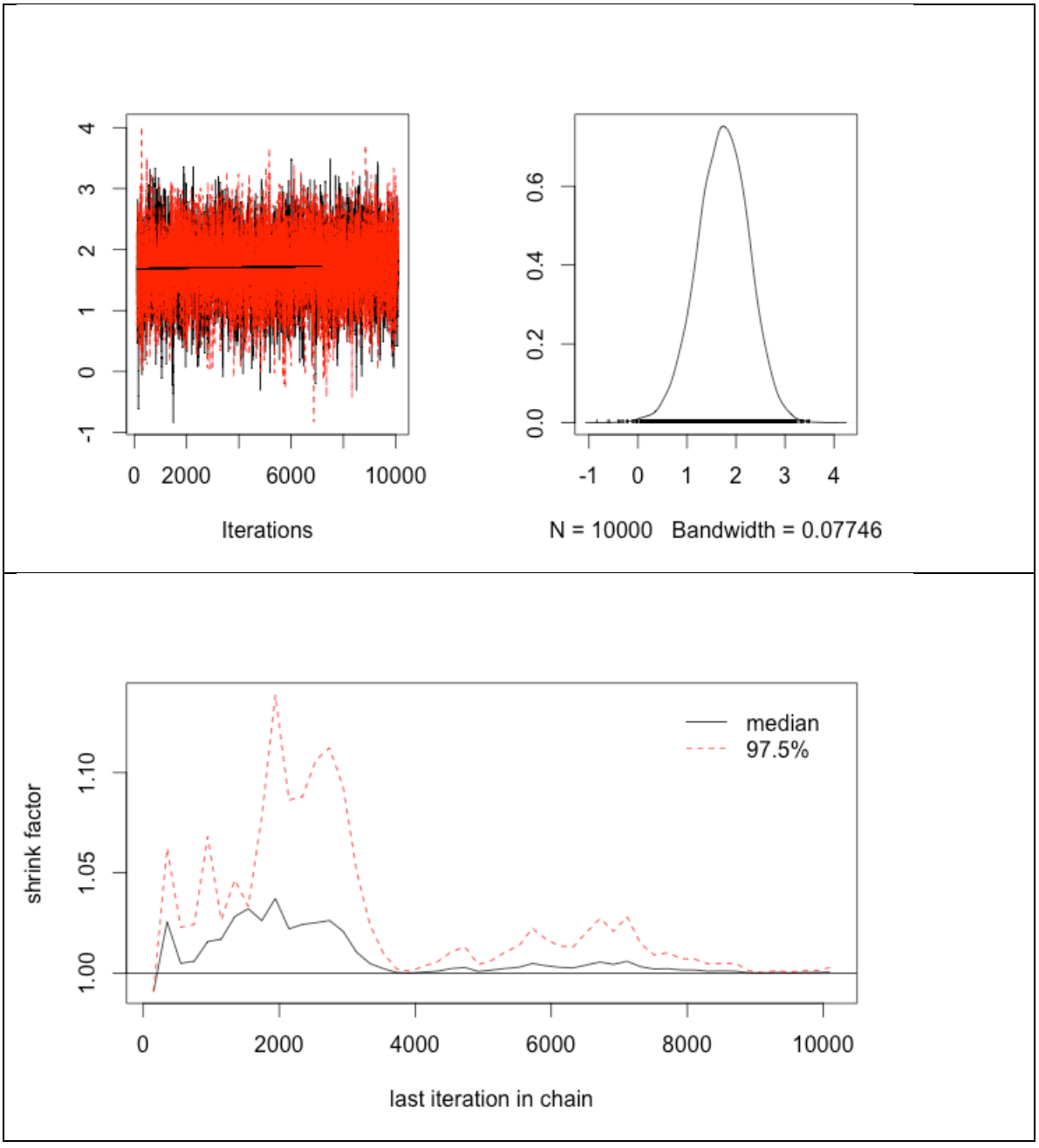
We modeled the exposure/outcome relationship specified in eq. 4 stratified by racial group and the all men model model specified in eq. 6 (minus the effects of U_i *in both models*) using logistic regression and presented the results as the naïve analysis for comparison with the Bayesian analysis.

Specification of Sensitivity Analysis

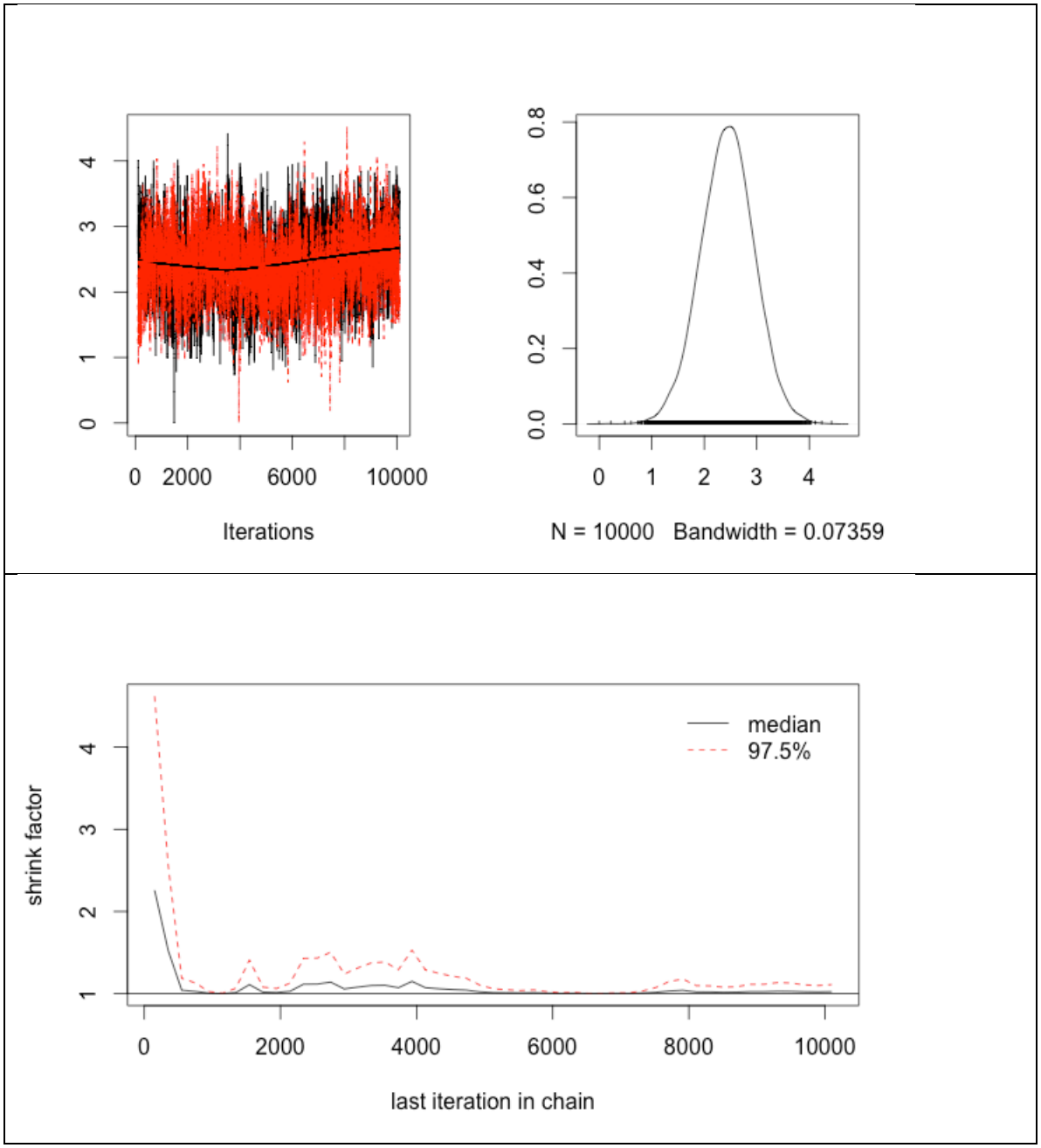
We fixed the OR to $\exp(\beta_1)$ from the all men model (eq. 6) and found compatible sets of prevalence estimates of U_i among MSM and non-MSM necessary to converge the point estimates of β_1 from the stratified analyses (eq. 4).

Bayesian MCMC Convergence Diagnostics

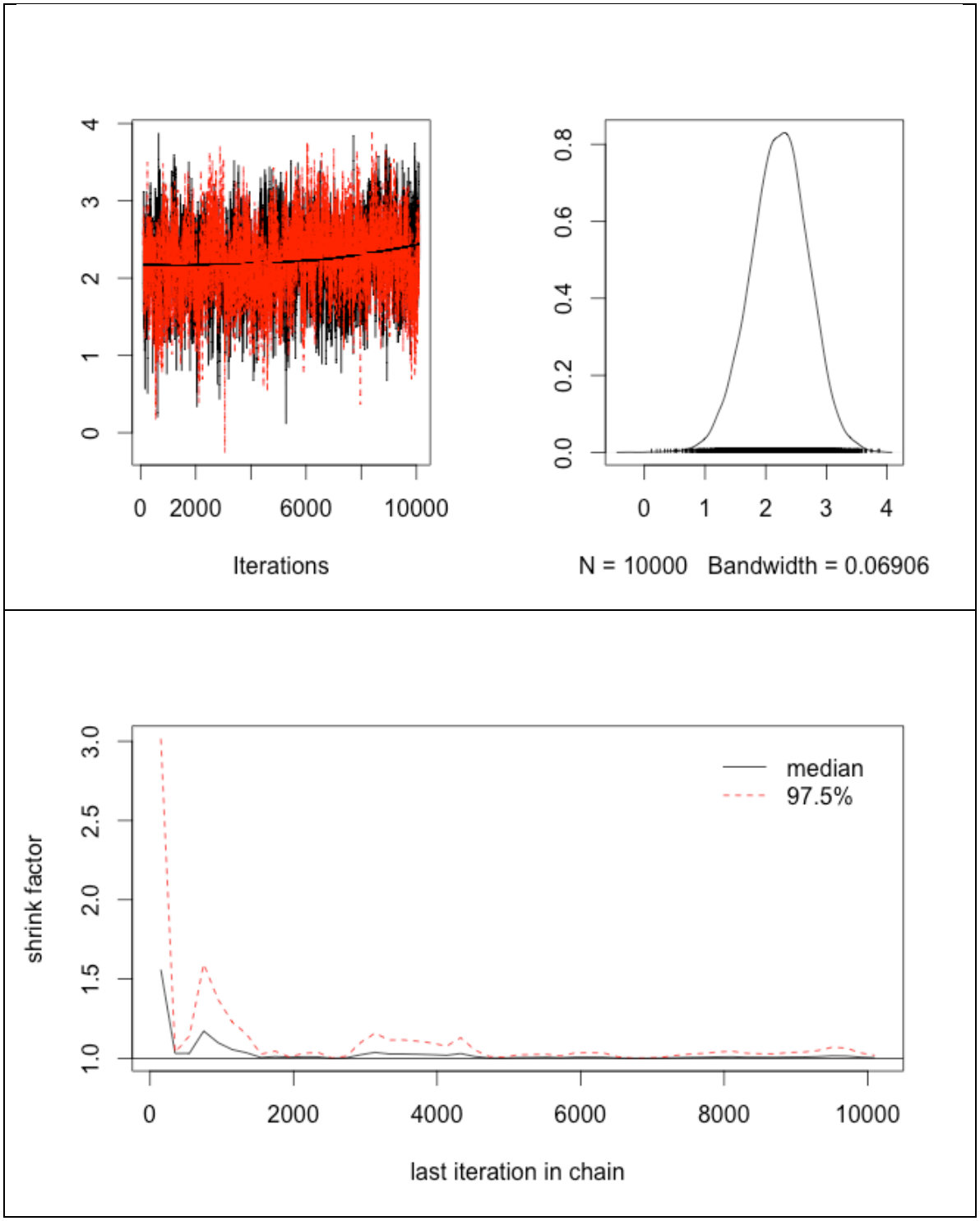
Posterior distribution and model diagnostics of β_{MSM}, black men (stratified model).



Posterior distribution and model diagnostics of β_{MSM}, white men (stratified model).



Posterior distribution and model diagnostics of β_{MSM}, black and white men (all men model).



Annotated R Code

Stratified model

```
model {
  for (i in 1:n) {

    #outcome model, log odds of hiv_aids given these predictors
    hiv_aids[i] ~ dbern(p_hiv_aids[i])
    logit(p_hiv_aids[i]) <-
b0+b1*msm[i]+b2*relationship[i]+b3*std[i]+b4*abuse_sexual[i]+b5*recent_drug[i]+b6*U[i]

    #exposure models, log odds of true msm status given these predictors
    msm[i] ~ dbern(p_msm[i])
    logit(p_msm[i]) <- a0+a1*relationship[i]+a2*std[i]+a3*abuse_sexual[i]+a4*recent_drug[i]+a5*U[i]

    #measurement model, imputing the true msm status given the measurement error
    msm.star[i] ~ dbern(p_msm.star[i])
    p_msm.star[i] <- sn.msm.hivneg*msm[i]*(1-hiv_aids[i])+(1-msm[i])*(1-sp.msm.hivneg)*(1-hiv_aids[i])
+ sn.msm.hivpos*msm[i]*(hiv_aids[i])+(1-msm[i])*(1-sp.msm.hivpos)*(hiv_aids[i])

    #prevalence models of potential confounders
    relationship[i] ~ dbern(p_relationship[i])
    logit(p_relationship[i]) <- prev.relationship

    std[i] ~ dbern(p_std[i])
    logit(p_std[i]) <- prev.std

    abuse_sexual[i] ~ dbern(p_abuse_sexual[i])
    logit(p_abuse_sexual[i]) <- prev.abuse_sexual

    recent_drug[i] ~ dbern(p_recent_drug[i])
    logit(p_recent_drug[i]) <- prev.recent_drug

    #prevalence model of unknown counfounder
    U[i] ~ dbern(p_U[i])
    logit(p_U[i]) <- prev.U
  }

  #priors
  #for normal distribution, provide (mean, precision=(1/variance))
  #for beta distribution, provide (alpha, beta)

  #for independent priors use dnorm(0,1/(fixed variance, e.g. 10)
  #for hierarchical priors use dnorm(0,1/(random variance sample from inverse chi sq distribution); see:
http://www.ncbi.nlm.nih.gov/pubmed/18226747
  #for prevalence priors, use independent prior dnorm(0,1/10)

  #instead of inverse chi2 (not in JAGS) use gamma (do not need to take recip), see:
http://www.cs.berkeley.edu/~jordan/courses/260-spring10/lectures/lecture5.pdf, and
http://www.stat.ubc.ca/~gavin/WinBUGSdocs/WinBUGS%20lectures%20.pdf
  xi ~ dgamma(10,((log(6)/2)^2)) # hierarchical prior variance for outcome vars
  rho ~ dgamma(10,((log(6)/2)^2)) # hierarchical prior variance for exposure vars

  b0 ~ dnorm(0,xi)
  b1 ~ dnorm(1.81,xi) #informed prior from BAAHMS
```

```
  b2 ~ dnorm(0,xi)
  b3 ~ dnorm(0,xi)
  b4 ~ dnorm(0,xi)
  b5 ~ dnorm(0,xi)
  b6 ~ dnorm(0,xi) # log odds for relationship: U --> Y
  a0 ~ dnorm(0,rho)
  a1 ~ dnorm(0,rho)
  a2 ~ dnorm(0,rho)
  a3 ~ dnorm(0,rho)
  a4 ~ dnorm(0,rho)
  a5 ~ dnorm(0,rho) # log odds for relationship: U --> X
  prev.relationship ~ dnorm(0,1/10)
  prev.std ~ dnorm(0,1/10)
  prev.abuse_sexual ~ dnorm(0,1/10)
  prev.recent_drug ~ dnorm(0,1/10)
  prev.U ~ dnorm(0,1/10)
  sn.msm.hivneg ~ dbeta(45,6) #partner, add beta(1,1)
  sp.msm.hivneg ~ dbeta(518,20) #partner, add beta(1,1)
  sn.msm.hivpos ~ dbeta(15,2) #partner, add beta(1,1)
  sp.msm.hivpos ~ dbeta(17,1) #partner, add beta(1,1)
}
```

All men model

```
model {
  for (i in 1:n) {

    #outcome model, log odds of hiv_aids given these predictors
    hiv_aids[i] ~ dbern(p_hiv_aids[i])
    logit(p_hiv_aids[i]) <-
b0+b1*msm[i]+b2*black[i]+b4*relationship[i]+b5*std[i]+b6*abuse_sexual[i]+b7*recent_drug[i]+b8*U[i]

    #exposure models, log odds of true msm status given these predictors
    msm[i] ~ dbern(p_msm[i])
    logit(p_msm[i]) <-
a0+a1*black[i]*a2*relationship[i]+a3*std[i]+a4*abuse_sexual[i]+a5*recent_drug[i]+a6*U[i]

    #measurement model, imputing the true msm status given the measurement error
    #allows for differential misclassification by HIV status as well as different estimates by race
    msm.star[i] ~ dbern(p_msm.star[i])
    p_msm.star[i] <- black[i]*(sn.black.msm.hivneg*msm[i]*(1-hiv_aids[i])+(1-msm[i])*(1-
sp.black.msm.hivneg)*(1-hiv_aids[i]) + sn.black.msm.hivpos*msm[i]*(hiv_aids[i])+(1-msm[i])*(1-
sp.black.msm.hivpos)*(hiv_aids[i])) + (1-black[i])*(sn.nonblack.msm.hivneg*msm[i]*(1-hiv_aids[i])+(1-
msm[i])*(1-sp.nonblack.msm.hivneg)*(1-hiv_aids[i]) + sn.nonblack.msm.hivpos*msm[i]*(hiv_aids[i])+(1-
msm[i])*(1-sp.nonblack.msm.hivpos)*(hiv_aids[i]))

    #prevalence models of potential confounders
    black[i] ~ dbern(p_black[i])
    logit(p_black[i]) <- prev.black

    relationship[i] ~ dbern(p_relationship[i])
    logit(p_relationship[i]) <- prev.relationship

    std[i] ~ dbern(p_std[i])
    logit(p_std[i]) <- prev.std

    abuse_sexual[i] ~ dbern(p_abuse_sexual[i])
    logit(p_abuse_sexual[i]) <- prev.abuse_sexual

    recent_drug[i] ~ dbern(p_recent_drug[i])
    logit(p_recent_drug[i]) <- prev.recent_drug

    #prevalence model of unknown counfounder
    U[i] ~ dbern(p_U[i])
    logit(p_U[i]) <- prev.U
  }

  #priors
  #for normal distribution, provide (mean, precision=(1/variance))
  #for beta distribution, provide (alpha, beta)

  #for independent priors use dnorm(0,1/(fixed variance, e.g. 10)
  #for hierarchical priors use dnorm(0,1/(random variance sample from inverse chi sq distribution); see:
http://www.ncbi.nlm.nih.gov/pubmed/18226747
  #for prevalence priors, use independent prior dnorm(0,1/10)
```

```
  #instead of inverse chi2 (not in JAGS) use gamma (do not need to take recip), see:
http://www.cs.berkeley.edu/~jordan/courses/260-spring10/lectures/lecture5.pdf, and
http://www.stat.ubc.ca/~gavin/WinBUGSdocs/WinBUGS%20lectures%20.pdf
  xi ~ dgamma(10,((log(6)/2)^2)) # hierarchical prior variance for outcome vars
  rho ~ dgamma(10,((log(6)/2)^2)) # hierarchical prior variance for exposure vars

  b0 ~ dnorm(0,xi)
  b1 ~ dnorm(1.81,xi) #informed prior from BAAHMS
  b2 ~ dnorm(0,xi)
  b4 ~ dnorm(0,xi)
  b5 ~ dnorm(0,xi)
  b6 ~ dnorm(0,xi)
  b7 ~ dnorm(0,xi)
  b8 ~ dnorm(0,xi) # log odds for relationship: U --> Y
  a0 ~ dnorm(0,rho)
  a1 ~ dnorm(0,rho)
  a2 ~ dnorm(0,rho)
  a3 ~ dnorm(0,rho)
  a4 ~ dnorm(0,rho)
  a5 ~ dnorm(0,rho)
  a6 ~ dnorm(0, rho) # log odds for relationship: U --> X
  prev.black ~ dnorm(0,1/10)
  prev.relationship ~ dnorm(0,1/10)
  prev.std ~ dnorm(0,1/10)
  prev.abuse_sexual ~ dnorm(0,1/10)
  prev.recent_drug ~ dnorm(0,1/10)
  prev.U ~ dnorm(0,1/10)

  #can have different SN/SP for black,nonblack men although keep it same right now
  sn.black.msm.hivneg ~ dbeta(45,6) #partner, add beta(1,1)
  sp.black.msm.hivneg ~ dbeta(518,20) #partner, add beta(1,1)
  sn.black.msm.hivpos ~ dbeta(15,2) #partner, add beta(1,1)
  sp.black.msm.hivpos ~ dbeta(17,1) #partner, add beta(1,1)
  sn.nonblack.msm.hivneg ~ dbeta(45,6) #partner, add beta(1,1)
  sp.nonblack.msm.hivneg ~ dbeta(518,20) #partner, add beta(1,1)
  sn.nonblack.msm.hivpos ~ dbeta(15,2) #partner, add beta(1,1)
  sp.nonblack.msm.hivpos ~ dbeta(17,1) #partner, add beta(1,1)
}
```