Web-based Supplementary Materials for "Power and Commensurate Priors for Synthesizing Aggregate and Individual Patient-Level Data in Network Meta-Analysis"

Hwanhee Hong, Haoda Fu, and Bradley P. Carlin

1 AD and IPD integrative NMA modeling using the contrast-based parameteterization

We assume the same likelihoods as those in the main manuscript for aggregated-level data (AD) and individual patient-level data (IPD). The likelihood functions are written as

$$y_{ik}^{AD} \sim \mathrm{N}(\mu_{ik}^{AD}, se_{ik}^2), i = 1, \ldots, I^{AD} \tag{A.1}$$

$$\text{and } y_{ijk}^{IPD} \sim \mathrm{N}(\mu_{ijk}^{IPD}, \sigma_{ik}^2), i = I^{AD} + 1, \ldots I, \tag{A.2}$$

where μ_{ik}^{AD} and μ_{ijk}^{IPD} are the underlying true means from AD and IPD, respectively, se_{ik} is the known (but often estimated) standard error in AD, and σ_{ik} is the unknown within-study standard deviation for IPD. Here, $i = 1, \ldots, I$, $k = 1, \ldots, K$, and $j = 1, \ldots n_i$ index study, treatment, and patient, respectively.

For network meta-analysis (NMA) models in contrast-based parameterization, we adopt the missing data framework proposed by Hong et al. (2016). In this parameterization, it is tricky to handle uncommon baseline treatments across studies. However, under the missing data framework, we can assume that every study contains a reference treatment (e.g., placebo or standard of care), usually coded as $k = 1$. When we have only AD, we model the unknown parameter μ_{ik}^{AD} in (A.1) as

$$\begin{aligned} \mu_{ik}^{AD} &= \phi_i^{AD} + \gamma_i^{AD} \bar{x}_{i1} && \text{if } k = 1 \\ \mu_{ik}^{AD} &= \phi_i^{AD} + \delta_{ik}^{AD} + \gamma_i^{AD} \bar{x}_{i1} + \beta_{ik}^{AD} \bar{x}_{ik} && \text{if } k \neq 1, \end{aligned} \tag{A.3}$$

where ϕ_i^{AD} is the reference treatment effect in study i, γ_i is the covariate effect for the reference treatment, δ_{ik}^{AD} and β_{ik}^{AD} are the random relative treatment effects (i.e., mean differences) and the

relative treatment-by-covariate interactions between treatments k and 1 in study i, respectively, with $\delta_{i1}^{AD} \equiv 0$ and $\beta_{i1}^{AD} \equiv 0$. Under a consistency assumption, the random treatment effects and treatment-by-covariate interaction effects are often assumed to follow normal distributions. That is, $\boldsymbol{\delta}_i^{AD} = (\delta_{i2}^{AD}, \ldots, \delta_{iK}^{AD})^T \overset{ind}{\sim} N(\mathbf{d}^{AD}, \boldsymbol{\Sigma})$ and $\boldsymbol{\beta}_i^{AD} = (\beta_{i2}^{AD}, \ldots, \beta_{iK}^{AD}) \overset{ind}{\sim} N(\mathbf{b}^{AD}, \boldsymbol{\Sigma}_x)$, where $\mathbf{d}^{AD} = (d_2^{AD}, \ldots, d_K^{AD})$ and $\mathbf{b}^{AD} = (b_2^{AD}, \ldots, b_K^{AD})$. The parameter of interest is d_k^{AD}, the mean relative treatment effect between treatments k and 1, and we interpret b_k^{AD} as the change in the difference between treatments k and 1 per unit increase in the covariate $\bar{x}_{ik}$. For hyperprior specifications, we assign vague normal priors to γ_i^{AD} and d_k^{AD}; we assume the b_k^{AD} to be exchangeable, $b_k^{AD} \overset{ind}{\sim} N(b^{AD}, \tau_b^2)$, where b^{AD} and τ_b follow vague normal and uniform prior distributions, respectively. Finally, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_x$ are $(K-1) \times (K-1)$ matrices where $\boldsymbol{\Sigma} = diag(\tau^2) + \rho\tau^2 I(i \neq j)$ and $\boldsymbol{\Sigma}_x = diag(\tau_x^2) + \rho_x \tau_x^2 I(i \neq j)$, with hyperprior distributions $\rho, \rho_x \sim$Uniform$(0,1)$ and $\tau, \tau_x \sim$Uniform$(0,10)$. Alternatively, $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Sigma}_x^{-1}$ can follow vague Wishart distributions.

When we have only IPD, we assume (A.2) to be the likelihood, and the unknown parameters μ_{ijk}^{IPD} can be modeled as

$$\begin{aligned} \mu_{ijk}^{IPD} &= \phi_i^{IPD} + \gamma_i^{IPD} x_{ij1} && \text{if } k = 1 \\ \mu_{ijk}^{IPD} &= \phi_i^{IPD} + \delta_{ik}^{IPD} + \gamma_i^{IPD} x_{ij1} + \beta_{ik}^{IPD} x_{ijk} && \text{if } k \neq 1, \end{aligned} \tag{A.4}$$

where all parameters have similar interpretations as the corresponding parameters in (A.3). We assume that $\boldsymbol{\delta}_i^{IPD} = (\delta_{i2}^{IPD}, \ldots, \delta_{iK}^{IPD})^T \overset{ind}{\sim} N(\mathbf{d}^{IPD}, \boldsymbol{\Psi})$ and $\boldsymbol{\beta}_i^{IPD} = (\beta_{i2}^{IPD}, \ldots, \beta_{iK}^{IPD}) \overset{ind}{\sim} N(\mathbf{b}^{IPD}, \boldsymbol{\Psi}_x)$, where $\mathbf{d}^{IPD} = (d_2^{IPD}, \ldots, d_K^{IPD})$ and $\mathbf{b}^{IPD} = (b_2^{IPD}, \ldots, b_K^{IPD})$. We define $\boldsymbol{\Psi} = diag(\psi^2) + \psi^2 I(i \neq j)$ and $\boldsymbol{\Psi}_x = diag(\psi_x^2) + \varphi_x \psi_x^2 I(i \neq j)$. We assign vague hyperprior distributions for all parameters, similar to those used above.

We can now adaptively integrate AD and IPD under the contrast-based parameterization. For the basic integrative model (the counterpart of Section 3.2.1 in the main manuscript) that assumes exchangeability of treatment effects in AD and IPD, we can replace δ_{ik}^{AD} and δ_{ik}^{IPD} with δ_{ik}, and β_{ik}^{AD} and β_{ik}^{IPD} with β_{ik} in (A.3) and (A.4). Then we fit the two models simultaneously by

assigning priors to δ_{ik} and β_{ik} similarly as above.

For integrative models using power priors, we use the same parameter specification as used in the basic integrative model. Suppose $\mathbf{\Theta}$ is a vector having all model parameters including d_k and β_{ik}. The power prior for $\mathbf{\Theta}$ given the likelihood for $\{y_{ik}^{AD}\}$ can be written as

$$\pi(\mathbf{\Theta} \mid \{y_{ik}^{AD}\}, \alpha_{ik}) \propto \left[\prod_{i=1}^{I^{AD}} L(\mathbf{\Theta} \mid y_{ik}^{AD})^{\alpha_{ik}}\right] \pi_0(\mathbf{\Theta}), \tag{A.5}$$

where $\pi_0(\mathbf{\Theta})$ is often assumed to be uniform and the $\alpha_{ik} \in (0, 1)$ control the level of borrowing from AD on the analysis. Then the posterior distribution of $\mathbf{\Theta}$ can be written as

$$q(\mathbf{\Theta} \mid \{y_{ijk}^{IPD}\}, \{y_{ik}^{AD}\}, \alpha_{ik}) \propto \left[\prod_{i=1}^{I^{AD}} L(\mathbf{\Theta} \mid y_{ik}^{AD})^{\alpha_{ik}}\right] \left[\prod_{i=I^{AD}+1}^{I} \prod_{j=1}^{n_i} L(\mathbf{\Theta} \mid y_{ijk}^{IPD})\right] \pi_0(\mathbf{\Theta}). \tag{A.6}$$

To adopt commensurate priors, we use the model specified in (A.3) and (A.4) with the IPD and AD superscripts. We assign commensurate priors for d_k^{IPD} and b_k^{IPD} to borrow information from the AD as

$$d_k^{IPD} \overset{ind}{\sim} \mathrm{N}(d_k^{AD}, \eta_k^2) \text{ and } b_k^{IPD} \overset{ind}{\sim} \mathrm{N}(b_k^{AD}, \eta_{bk}^2), \tag{A.7}$$

where the commensurability parameters η_k and η_{bk} follow spike-and-slab hyperpriors as described in the main manuscript.

2 Diabetes data analysis results in arm-based covariate-unadjusted models

Web Figure A1 displays treatment effect estimates and associated 95% credible intervals obtained from 10 different unadjusted models. We again use colors to distinguish groups of models: black is for AD models, orange is for IPD models, blue is for the basic integrative model, green is for integrative models with power priors, and red is for integrative models with commensurate priors. Compared to Figure 4 in the main manuscript, the overall trend of unadjusted treatment

effect estimates is similar to that of the corresponding adjusted models. However, the adjusted models tend to provide wider credible intervals, specifically for treatments having evidence from a single AD study. This is expected because those treatments do not have enough information to estimate the interactions and this adds uncertainty into estimates.

Web Figure A2 displays rankings of the 28 treatments across the 10 unadjusted models using SUCRA. As in the main manuscript, the ranks in the IPD model and the other models are not comparable because there are only 5 treatments to be compared in the IPD model. Across all models except the IPD model, GLY+TRO is the best treatment and MET+GLY is the runner-up. MET is the best treatment in the IPD model, but it is one of least-favored treatments in the AD1 and AD2 models, and emerges as 13^{th} or 7^{th} best treatment under the I-pp1 to I-pp4 and I-cp2 models, respectively. Overall, the rankings when adapting power priors are different from when adapting commensurate priors, except for a few treatments such as GLY+TRO, MET+SIT, and MET+GLY.

3 Additional data analysis results in arm-based covariate-adjusted models

Web Figure A3 displays the distributions of the two baseline covariates and their correlation with the outcome. Study-level baseline covariate information available from AD studies is limited because the distributions of means of baseline covariates across studies do not vary much compared to the distributions of patient baseline covariates available from IPD studies. In addition, Panel (b) shows a positive association between baseline age and outcome from AD studies (black solid line), while the IPD studies (orange solid lines) show little to no association. This shows that using baseline covariate information from AD studies only might result in an ecological fallacy and biased inference.

Web Table A1 shows estimates of covariate-by-treatment interactions in all covariate-adjusted models, $\widehat{b}$, within- and across-trial interactions, $\widehat{b}^w$ and $\widehat{b}^a$, and their difference measuring a degree of ecological fallacy. Age-by-treatment interaction is close to zero in all models except the AD1

model. Baseline HbA1c-by-treatment interaction is statistically significant except under the AD1 and AD2 models. The estimated measures of ecological fallacy of baseline HbA1c under models assuming ecological fallacy are significant, though treatment effect estimates do not change much compared to those obtained under models assuming no ecological fallacy.

4 JAGS code for integrative models

4.1 The covariate-adjusted I-vague model

```
## yIPD: outcome from IPD
## yAD: outcome from AD
## x1.IPD: age (centered at 57) from IPD
## x1.AD: mean age (centered at 57) from AD
## x2.IPD: baseline A1c (centered at 9)
## x2.AD: mean baseline A1c (centered at 9)

model {

  ## Fit IPD

  for (i in 1:NIPD) {
    yIPD[i] ~ dnorm(meanIPD[study[i],person[i]], precIPD[study[i],drug[i]])
    meanIPD[study[i],person[i]] <- mu.IPD[drug[i]] + v.IPD[study[i],drug[i]] +
                                   beta.x1.IPD[study[i],drug[i]]*x1.IPD[i] +
                                   beta.x2.IPD[study[i],drug[i]]*x2.IPD[i]
  }

  for (i in 1:NSIPD) {
    for(k in 1:5) {
      precIPD[i,k] <- 1/pow(sdIPD[i,k],2)
      sdIPD[i,k] ~ dunif(0.01, 10)
    }
  }

  for (j in 1:NSIPD) {
    v.IPD[j, 1:5] ~ dmnorm(zero.AB.IPD[1:5], invR.IPD[1:5, 1:5])
    beta.x1.IPD[j, 1:5] ~ dmnorm(b.x1.IPD[1:5], invR.x1.IPD[1:5, 1:5])
    beta.x2.IPD[j, 1:5] ~ dmnorm(b.x2.IPD[1:5], invR.x2.IPD[1:5, 1:5])
  }

  invR.IPD[1:5, 1:5] <- inverse(R.IPD[ , ])
  invR.x1.IPD[1:5, 1:5] <- inverse(R.x1.IPD[ , ])
  invR.x2.IPD[1:5, 1:5] <- inverse(R.x2.IPD[ , ])
```

```
for(j in 1:5){
  for(k in 1:5){
    R.IPD[j, k]<-tau.IPD[1]*tau.IPD[1]*pow(rho.IPD[1],step(abs(j-k)-0.5))
    R.x1.IPD[j, k]<-tau.IPD[2]*tau.IPD[2]*pow(rho.IPD[2],step(abs(j-k)-0.5))
    R.x2.IPD[j, k]<-tau.IPD[3]*tau.IPD[3]*pow(rho.IPD[3],step(abs(j-k)-0.5))
  }
}

for (j in 1:3) {
  tau.IPD[j] ~ dunif(0.01,10)
  rho.IPD[j] ~ dunif(0,1)
}

## Fit AD

for (i in 1:NAD) {
  yAD[i] ~ dnorm(meanAD[s[i],t[i]], precAD[i])
  precAD[i] <- n[i]/pow(sdAD[i],2)
  meanAD[s[i],t[i]] <- mu.AD[t[i]] + v[s[i],t[i]] +
                       beta.x1.AD[s[i],t[i]]*x1.AD[i] +
                       beta.x2.AD[s[i],t[i]]*x2.AD[i]
}

for (j in 1:NSAD) {
  v[j, 1:27] ~ dmnorm(zero.AB[1:27], invR.AD[1:27, 1:27])
  beta.x1.AD[j, 1:27] ~ dmnorm(b.x1.AD[1:27], invR.x1.AD[1:27, 1:27])
  beta.x2.AD[j, 1:27] ~ dmnorm(b.x2.AD[1:27], invR.x2.AD[1:27, 1:27])
}

invR.AD[1:27, 1:27] <- inverse(R.AD[ , ])
invR.x1.AD[1:27, 1:27] <- inverse(R.x1.AD[ , ])
invR.x2.AD[1:27, 1:27] <- inverse(R.x2.AD[ , ])

for(j in 1:27){
  for(k in 1:27){
    R.AD[j, k]<-tau.AD[1]*tau.AD[1]*pow(rho.AD[1],step(abs(j-k)-0.5))
    R.x1.AD[j, k]<-tau.AD[2]*tau.AD[2]*pow(rho.AD[2],step(abs(j-k)-0.5))
    R.x2.AD[j, k]<-tau.AD[3]*tau.AD[3]*pow(rho.AD[3],step(abs(j-k)-0.5))
  }
}

for (j in 1:3) {
  tau.AD[j] ~ dunif(0.01,10)
  rho.AD[j] ~ dunif(0,1)
}

## Redefine mean parameters including interactions
```

```
for(i in 1:27) {
  mu.AD[i] <- mu[i]
  b.x1.AD[i] <- b.x1[i]
  b.x2.AD[i] <- b.x2[i]
}
for(i in 1:4) {
  mu.IPD[i] <- mu[i]
  b.x1.IPD[i] <- b.x1[i]
  b.x2.IPD[i] <- b.x2[i]
}
mu.IPD[5] <- mu[28]
b.x1.IPD[5] <- b.x1[28]
b.x2.IPD[5] <- b.x2[28]

## Prior to b.x1[] b.x2[]
for (k in 1:NT) {
  b.x1[k] ~ dnorm(bb.x[1],precx[1])
  b.x2[k] ~ dnorm(bb.x[2],precx[2])
}

for (i in 1:2) {
  bb.x[i] ~ dnorm(0, 0.001)
  precx[i] <- 1/pow(taux[i],2)
  taux[i] ~ dunif(0.01,10)
}

for (k in 1:NT) {
  mu[k] ~ dnorm(0, 0.001)
}

## rank
for (k in 1:NT) { T[k] <- mu[k] }
T.rank <- rank(T)
for (k in 1:NT) {
  for(j in 1:NT) {
    best[k,j] <- equals(T.rank[k],j)
  }
}
for (k in 1:NT) { best1[k] <- equals(T.rank[k],1) }

# Prediction
# at age 57

for (k in 1:NT) {
  for (j in 1:8) {
    pred[k,j] <- mu[k] + b.x1[k]*(57-57) + b.x2[k]*(temp.x2[j]-9)
  }
}
```

```
}
```

4.2 The covariate-adjusted I-pp3 model

```
## a0[i]: the power parameter $alpha_{ik}$

model {

  ## Fit IPD

  for (i in 1:NIPD) {
    yIPD[i] ~ dnorm(meanIPD[study[i],person[i]], precIPD[study[i],drug[i]])
    meanIPD[study[i],person[i]] <- mu.IPD[drug[i]] + v.IPD[study[i],drug[i]] +
                                   beta.x[1]*x1.IPD[i] + beta.x[2]*x2.IPD[i]
  }

  for (i in 1:NSIPD) {
    for(k in 1:5) {
      precIPD[i,k] <- 1/pow(sdIPD[i,k],2)
      sdIPD[i,k] ~ dunif(0.01, 10)
    }
  }

  for (j in 1:NSIPD) {
    v.IPD[j, 1:5] ~ dmnorm(zero.AB.IPD[1:5], invR.IPD[1:5, 1:5])
  }

  invR.IPD[1:5, 1:5] <- inverse(R.IPD[ , ])

  for(j in 1:5){
    for(k in 1:5){
      R.IPD[j, k]<-tau.IPD*tau.IPD*pow(rho.IPD,step(abs(j-k)-0.5))
    }
  }

  tau.IPD ~ dunif(0.01,10)
  rho.IPD ~ dunif(0,1)

  ## Fit AD
  const <- 10000
  for (i in 1:NAD) {
    meanAD[s[i],t[i]] <- mu.AD[t[i]] + v[s[i],t[i]] +
                         beta.x[1]*x1.AD[i] + beta.x[2]*x2.AD[i]

    # log likelihood and use zeros trick
    sdsd[i] <- sdAD[i]/sqrt(n[i])
    phi[i] <- a0[i]*( log(sdsd[i]) +
```

```
                    0.5*pow((yAD[i]- meanAD[s[i],t[i]])/sdsd[i] , 2) ) + const
    zeros[i] ~ dpois(phi[i])
  }

  for (j in 1:NSAD) {
    v[j, 1:27] ~ dmnorm(zero.AB[1:27], invR.AD[1:27, 1:27])
  }

  invR.AD[1:27, 1:27] <- inverse(R.AD[ , ])

  for(j in 1:27){
    for(k in 1:27){
      R.AD[j, k]<-tau.AD[1]*tau.AD*pow(rho.AD,step(abs(j-k)-0.5))
    }
  }

  tau.AD ~ dunif(0.01,10)
  rho.AD ~ dunif(0,1)

  ## Redefine mean parameters including interactions
  for(i in 1:27) {
    mu.AD[i] <- mu[i]
  }
  for(i in 1:4) {
    mu.IPD[i] <- mu[i]
  }
  mu.IPD[5] <- mu[28]

  ## Prior to beta.x
  for (i in 1:2) { beta.x[i] ~ dnorm(0,0.01) }


  for (k in 1:NT) {
    mu[k] ~ dnorm(0, 0.001)
  }
}
```

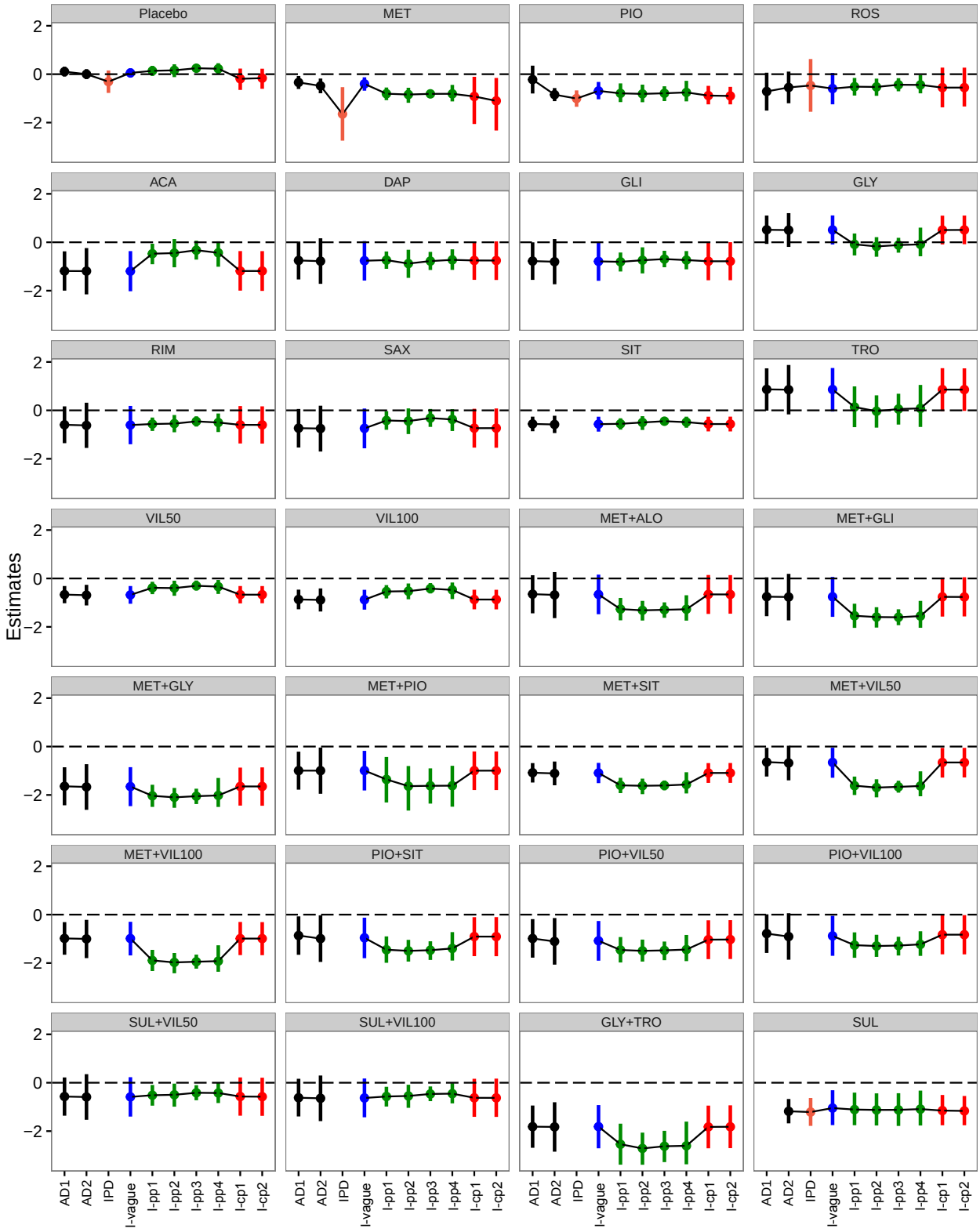
4.3 The covariate-adjusted I-cp2 model

```
model {

  ## Fit IPD

  for (i in 1:NIPD) {
    yIPD[i] ~ dnorm(meanIPD[study[i],person[i]], precIPD[study[i],drug[i]])
    meanIPD[study[i],person[i]] <- mu.IPD[drug[i]] + v.IPD[study[i],drug[i]] +
                                   beta.x1.IPD[study[i],drug[i]]*x1.IPD[i] +
                                   beta.x2.IPD[study[i],drug[i]]*x2.IPD[i]
```

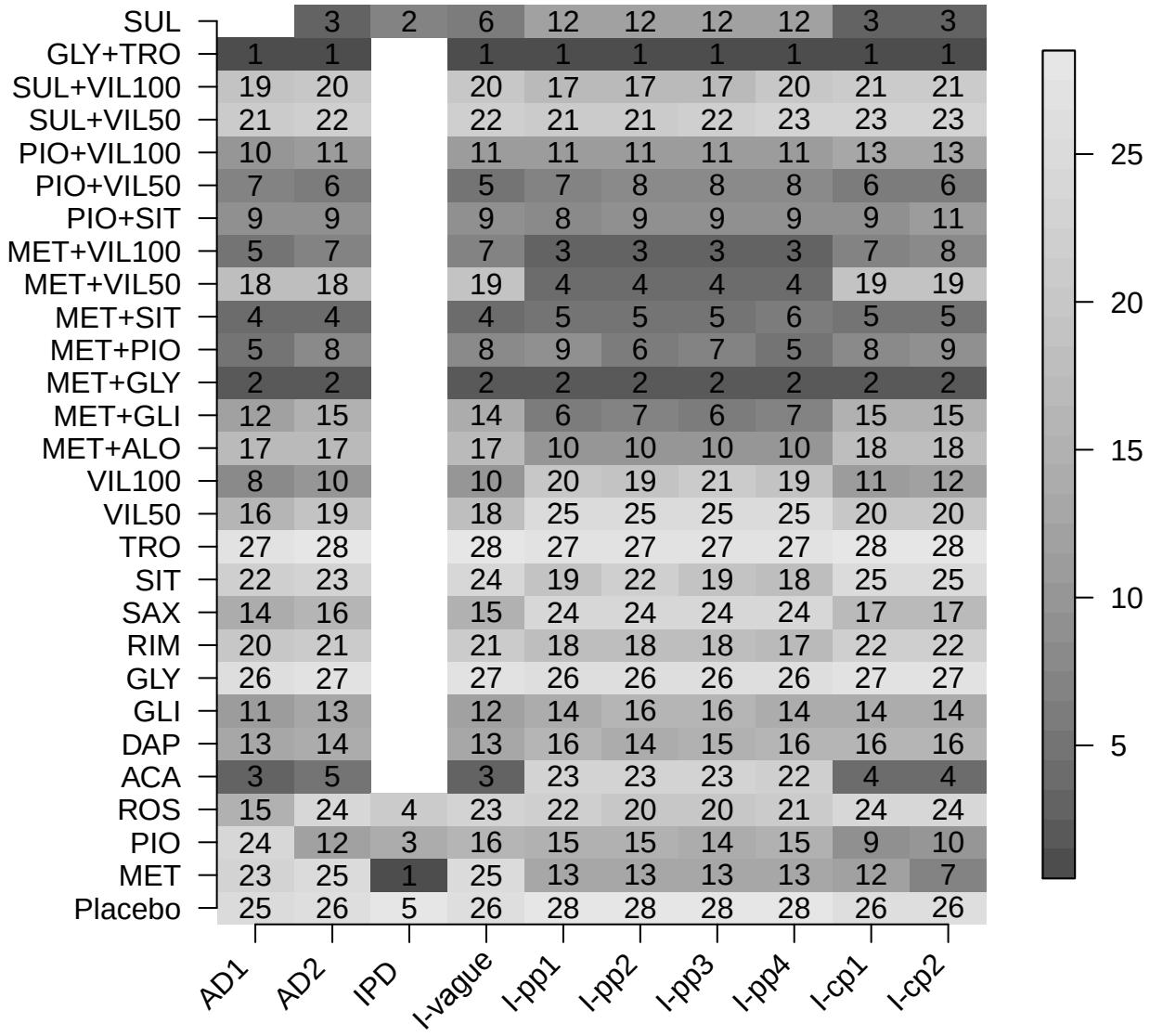
```
}

for (i in 1:NSIPD) {
  for(k in 1:5) {
    precIPD[i,k] <- 1/pow(sdIPD[i,k],2)
    sdIPD[i,k] ~ dunif(0.01, 10)
  }
}

for (j in 1:NSIPD) {
  v.IPD[j, 1:5] ~ dmnorm(zero.AB.IPD[1:5], invR.IPD[1:5, 1:5])
  beta.x1.IPD[j, 1:5] ~ dmnorm(b.x1.IPD[1:5], invR.x1.IPD[1:5, 1:5])
  beta.x2.IPD[j, 1:5] ~ dmnorm(b.x2.IPD[1:5], invR.x2.IPD[1:5, 1:5])
}

invR.IPD[1:5, 1:5] <- inverse(R.IPD[ , ])
invR.x1.IPD[1:5, 1:5] <- inverse(R.x1.IPD[ , ])
invR.x2.IPD[1:5, 1:5] <- inverse(R.x2.IPD[ , ])

for(j in 1:5){
  for(k in 1:5){
    R.IPD[j, k]<-tau.IPD[1]*tau.IPD[1]*pow(rho.IPD[1],step(abs(j-k)-0.5))
    R.x1.IPD[j, k]<-tau.IPD[2]*tau.IPD[2]*pow(rho.IPD[2],step(abs(j-k)-0.5))
    R.x2.IPD[j, k]<-tau.IPD[3]*tau.IPD[3]*pow(rho.IPD[3],step(abs(j-k)-0.5))
  }
}

for (j in 1:3) {
  tau.IPD[j] ~ dunif(0.01,10)
  rho.IPD[j] ~ dunif(0,1)
}

## Fit AD

for (i in 1:NAD) {
  yAD[i] ~ dnorm(meanAD[s[i],t[i]], precAD[i])
  precAD[i] <- n[i]/pow(sdAD[i],2)
  meanAD[s[i],t[i]] <- mu.AD[t[i]] + v[s[i],t[i]] +
                       beta.x1.AD[s[i],t[i]]*x1.AD[i] +
                       beta.x2.AD[s[i],t[i]]*x2.AD[i]
}

for (j in 1:NSAD) {
  v[j, 1:27] ~ dmnorm(zero.AB[1:27], invR.AD[1:27, 1:27])
  beta.x1.AD[j, 1:27] ~ dmnorm(b.x1.AD[1:27], invR.x1.AD[1:27, 1:27])
  beta.x2.AD[j, 1:27] ~ dmnorm(b.x2.AD[1:27], invR.x2.AD[1:27, 1:27])
}
```

```
invR.AD[1:27, 1:27] <- inverse(R.AD[ , ])
invR.x1.AD[1:27, 1:27] <- inverse(R.x1.AD[ , ])
invR.x2.AD[1:27, 1:27] <- inverse(R.x2.AD[ , ])

for(j in 1:27){
  for(k in 1:27){
    R.AD[j, k]<-tau.AD[1]*tau.AD[1]*pow(rho.AD[1],step(abs(j-k)-0.5))
    R.x1.AD[j, k]<-tau.AD[2]*tau.AD[2]*pow(rho.AD[2],step(abs(j-k)-0.5))
    R.x2.AD[j, k]<-tau.AD[3]*tau.AD[3]*pow(rho.AD[3],step(abs(j-k)-0.5))
  }
}

for (j in 1:3) {
  tau.AD[j] ~ dunif(0.01,10)
  rho.AD[j] ~ dunif(0,1)
}

# spike and slab now
for(i in 1:4) {
  mu.IPD[i] ~ dnorm(mu.AD[i], sstau[i])
  b.x1.IPD[i] ~ dnorm(b.x1.AD[i], sstau.x1[i])
  b.x2.IPD[i] ~ dnorm(b.x2.AD[i], sstau.x2[i])
}

# sstau <- 1000    # *forces* commensurability of AD and IPD
# sstau <- 0.001   # essentially disconnects AD and IPD
# sstau ~ dgamma(.01, .01)  # Standard WinBUGS vague hyperprior

for (k in 1:4) {
  tee[k,1] ~ dnorm(20,1)     # R_tau is (essentially) 20
  tee[k,2] ~ dgamma(0.1, 0.1)I(0.1, 5)  # replacement for the true slab
  flip[k] ~ dbern(0.5)     # p_tau is 0.5
  pick[k] <- flip[k] + 1
  sstau[k] <- tee[k,pick[k]]
  sstau_rev[k] <- sqrt(1/sstau[k])

  tee.x1[k,1] ~ dnorm(20,1)     # R_tau is (essentially) 20
  tee.x1[k,2] ~ dgamma(0.1, 0.1)I(0.1, 5)  # replacement for the true slab
  flip.x1[k] ~ dbern(0.5)     # p_tau is 0.5
  pick.x1[k] <- flip.x1[k] + 1
  sstau.x1[k] <- tee[k,pick.x1[k]]
  sstau_rev.x1[k] <- sqrt(1/sstau.x1[k])

  tee.x2[k,1] ~ dnorm(20,1)     # R_tau is (essentially) 20
  tee.x2[k,2] ~ dgamma(0.1, 0.1)I(0.1, 5)  # replacement for the true slab
  flip.x2[k] ~ dbern(0.5)     # p_tau is 0.5
  pick.x2[k] <- flip.x2[k] + 1
  sstau.x2[k] <- tee[k,pick.x2[k]]
```

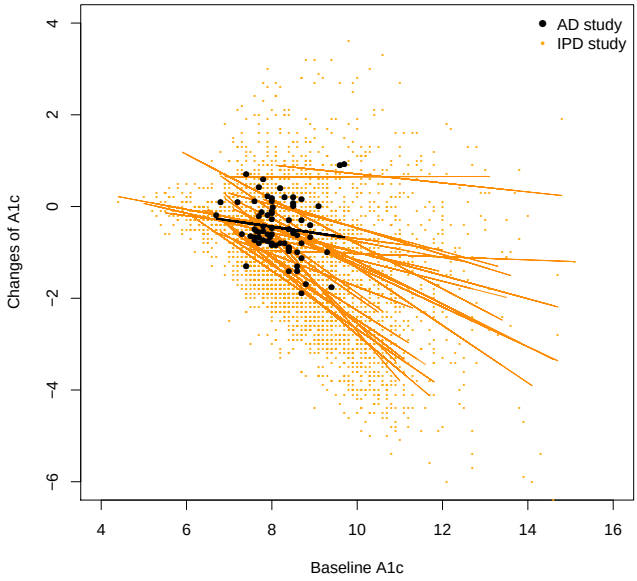
```
    sstau_rev.x2[k] <- sqrt(1/sstau.x2[k])
  }

  ## Redefine mean parameters including interactions
  for(i in 1:27) {
    mu.AD[i] <- mu.temp[i]
    b.x1.AD[i] <- b.x1.temp[i]
    b.x2.AD[i] <- b.x2.temp[i]
  }
  mu.IPD[5] <- mu.temp[28]     # just for easy coding
  b.x1.IPD[5] <- b.x1.temp[28]
  b.x2.IPD[5] <- b.x2.temp[28]

  for (k in 1:NT) { mu.temp[k] ~ dnorm(0, 0.001) }

  ## Prior to b.x1.AD[] b.x2.AD[]
  for (k in 1:NT) {
    b.x1.temp[k] ~ dnorm(bb.x[1],precx[1])
    b.x2.temp[k] ~ dnorm(bb.x[2],precx[2])
  }

  for (i in 1:2) {
    bb.x[i] ~ dnorm(0, 0.001)
    precx[i] <- 1/pow(taux[i],2)
    taux[i] ~ dunif(0.01,10)
  }

  for(k in 1:4) {
    mu[k] <- mu.IPD[k]
    b.x1[k] <- b.x1.IPD[k]
    b.x2[k] <- b.x2.IPD[k]
  }
  for(k in 5:28) {
    mu[k] <- mu.temp[k]
    b.x1[k] <- b.x1.temp[k]
    b.x2[k] <- b.x2.temp[k]
  }
}
```



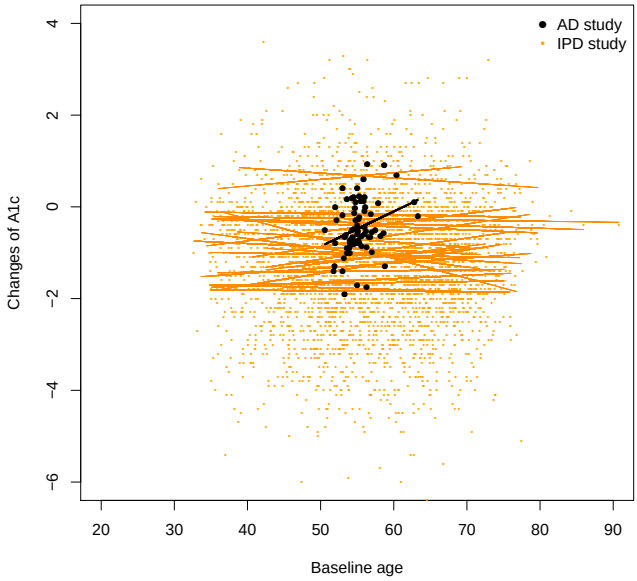
Web Figure A1: Treatment effect estimates obtained from unadjusted models.



Web Figure A2: Rankings of 28 treatments based on the SUCRA under 10 different covariate-unadjusted models.



(a) Baseline A1c by changes of A1c



(b) Baseline age by changes of A1c

Web Figure A3: Scatter plots and linear regression lines of the A1c change outcome and baseline covariates by studies and treatments: (a) baseline A1c by changes of A1c; (b) baseline age and changes of A1c. The black solid thick lines are obtained from the regression of the study-level outcome on the study-level baseline covariates across all AD studies. The orange lines are obtained from the regression of the patient-level outcome on the patient-level baseline covariates for each treatment in IPD studies.

Adjusted	Baseline age			Baseline HbA1c		
model	$\widehat{b}_{age}$ or $\widehat{b}^{w}_{age}$	$\widehat{b}^{a}_{age}$	$\widehat{b}^{w}_{age}$ - $\widehat{b}^{a}_{age}$	$\widehat{b}_{A1c}$ or $\widehat{b}^{w}_{A1c}$	$\widehat{b}^{a}_{A1c}$	$\widehat{b}^{w}_{A1c}$ - $\widehat{b}^{a}_{A1c}$
AD1	0.120 (0.07)	-	-	0.027 (0.19)	-	
AD2	0.017 (0.05)	-	-	-0.218 (0.13)	-	
IPD	-0.004 (0.01)	-	-	**-0.448 (0.11)**	-	
I-vague	-0.002 (0.01)	-	-	**-0.356 (0.10)**	-	
I-pp1	-0.001 (0.00)	-	-	**-0.422 (0.01)**	-	
I-pp2	-0.001 (0.00)	-	-	**-0.421 (0.01)**	-	
I-pp3	-0.001 (0.00)	-	-	**-0.421 (0.02)**	-	
I-pp4	-0.001 (0.00)	-	-	**-0.421 (0.01)**	-	
I-cp1	0.034 (0.05)	-	-	**-0.344 (0.11)**	-	
I-cp2	0.041 (0.05)	-	-	**-0.296 (0.12)**	-	
IPDef	-0.001 (0.00)	-0.076 (0.10)	0.075 (0.10)	**-0.426 (0.02)**	-0.068 (0.17)	**-0.358 (0.17)**
Ief-vague	-0.001 (0.00)	0.017 (0.03)	-0.017 (0.03)	**-0.426 (0.02)**	-0.166 (0.10)	**-0.260 (0.10)**
Ief-pp1	-0.001 (0.00)	0.007 (0.03)	-0.008 (0.03)	**-0.426 (0.02)**	-0.172 (0.11)	**-0.254 (0.11)**
Ief-pp3	-0.001 (0.00)	-0.010 (0.03)	0.009 (0.03)	**-0.426 (0.02)**	-0.168 (0.12)	**-0.258 (0.12)**
Ief-cp2	-0.001 (0.00)	-0.059 (0.09)	0.058 (0.09)	**-0.426 (0.02)**	-0.087 (0.15)	**-0.339 (0.15)**

Web Table A1: Estimates of interactions ($\widehat{b}$) and differences between within- and across-trial interactions ($\widehat{b}^{w} - \widehat{b}^{a}$) for baseline age and HbA1c under adjusted models with associated standard error in parentheses. Significant estimates are in bold.