5 APPENDIX

5.1 Priors for the binary EXNEX model

5.1.1 Weakly-informative prior distributions

The following prior specifications are meant to be weakly-informative. If solid prior information is available for any of the model parameters, informative priors should be envisaged instead. For the nonexchangeable component (12), each stratum-specific prior is normal,

$$\theta_j \sim N(m_w, v_w) \text{ with probability } p_{j3}, \tag{19}$$

with a mean m_w equal to the log-odds of a plausible guess (p_w) for the response probability, and a variance v_w that corresponds to approximately one observation:

$$\theta_j \sim N(\log(p_w/(1-p_w)), 1/p_w + 1/(1-p_w)) \tag{20}$$

The priors for the EX parameters are chosen as follows:

- τ_1 and τ_2: half-normal (HN) priors with scale parameter 1,

  $$\tau_1, \tau_2 \sim \text{HN(scale=1)} \tag{21}$$

  which imply prior 95%-intervals (0.031,2.24). On the log-odds scale, this range of τ values allows for small to very large between-strata heterogeneity (see [27] and Table 2 of section 2.1). For a small number of strata, this prior will lead to rather conservative borrowing. Depending on the context, more optimistic priors could be considered: for example, a HN(scale=0.5) prior, for which prior information is still fairly weak (95%-interval: 0.016-1.12).

- The means μ_1 and μ_2 of the exchangeability distributions follow normal priors $N(m_1, v_1)$ and $N(m_2, v_2)$, centered at plausible values m_1 and m_2, e.g., an uninteresting ("null") and interesting ("alternative") value.

- The variances of v_1 and v_2 are derived from two inputs: the priors for τ (21) and the assumed marginal variances for θ_j. For example, v_1 is obtained as follows. First the marginal variance of θ_j is specified. In the weakly-informative setting, we assume it to be approximately equivalent to one observation:

$$V_1(\theta_j) = 1/p_1 + 1/(1-p_1) \tag{22}$$

where $p_1 = \text{expit}(m_1)$. From the law of total variance, it then follows that

$$\begin{aligned} V_1(\theta_j) &= E_{\mu_1,\tau_1} V(\theta_j|\mu_1,\tau_1) + V_{\mu_1,\tau_1} E(\theta_j|\mu_1,\tau_1) \\ &= E_{\tau_1}(\tau_1^2) + V_{\mu_1}(\mu_1) \\ &= s^2 + v_1, \end{aligned}$$

since $E_{\tau_1}(\tau_1^2)$ is equal to s^2 times the expectation of a $\chi^2(1)$ distribution. Thus, the prior variance of μ_1 follows as

$$v_1 = V_1(\theta_j) - s^2 \tag{23}$$

5.1.2 Priors for application 1

For application 1 of section 3.1, the following weakly-informative prior distributions are obtained:

$$\begin{aligned} \tau_1 &\sim \text{HN(scale=1)} \\ \mu_1 &\sim N(\text{logit}(0.1), 3.18^2) \\ \tau_2 &\sim \text{HN(scale=1)} \\ \mu_2 &\sim N(\text{logit}(0.3), 1.94^2) \end{aligned}$$

The weakly-informative prior distributions under nonexchangeability are

$$\theta_j \sim N(\text{logit}(0.2), 2.5^2)$$

With these specifications, the prior analysis shows wide prior 95%-intervals $(0.01, 0.98)$ for $\pi_j = 1/(1 + \exp(-\theta_j))$.

5.2 Prior specifications for application 2

The priors for the parameters under nonexchangeability were chosen as follows:

- Prior for $\log(\alpha)$: $m_{w1} = \text{logit}(0.13)$, $s_{w_1} = 2$. For the DLT rate at $d^\star = 24$, this corresponds to a unit-information prior on the log-odds scale. On the DLT rate scale, the corresponding prior median is 0.13 and the 95%-interval (0.003,0.87).
- Prior for $\log(\beta)$: $m_{w2} = 0, s_{w2} = 1$. For the slope parameter β, the prior has the following interpretation: when doubling the dose, the odds of a DLT is multiplied by $2^{\exp(0)} = 2$ (median), and the 95%-interval for this multiplier is (1.1,137), which is weakly-informative since it allows for flat to steep curves.

The prior for the exchangeability parameters were chosen as follows

- Priors for τ_1, τ_2: $\tau_1 \sim \text{log-normal}(-1.386, 0.354^2)$, with 95%-interval (0.125,0.5), which allows for small to substantial between-strata heterogeneity for α; $\tau_2 \sim \text{log-normal}(-2.079, 0.354^2)$, with 95%-interval (0.063,0.25), which allows for small to substantial between-strata heterogeneity for β.
- Prior for ρ: uniform(-1,1)
- Priors for μ_1, μ_2: $\mu_1 \sim \text{N}(\text{logit}(0.13), 1.98^2)$, $\mu_2 \sim \text{N}(0, 0.99^2)$. The priors are similar to the ones under non-exchangeabilty, and thus weakly-informative.

5.3 WinBUGS code for the binary EXNEX model

```
model{
# prior distributions for EX-parameters
for (jj in 1:Nexch) {
  mu[jj] ~dnorm(mu.mean[jj],mu.prec[jj])
  prior.tau.prec[jj] <- pow(tau.HN.scale[jj],-2)
  tau[jj] ~ dnorm(0,prior.tau.prec[jj])I(0.001,)
  prec.tau[jj] <- pow(tau[jj],-2)
}
# log-odds parameters under EX
for (jj in 1:Nexch) {
  for (j in 1:Nstrata) {
        re[jj,j] ~ dnorm(0,prec.tau[jj])
        LogOdds[jj,j] <- mu[jj]+re[jj,j]
    }
}
# log-odds parameters under NEX
for (j in 1:Nstrata)  {
  LogOdds[Nmix,j] ~ dnorm(nex.mean,nex.prec)
}
# latent mixture indicators:
# exch.index: categorial 1,...,Nmix=Nexch+1
# exch: Nstrata x Nmix matrix of 0/1 elements
for (j in 1:Nstrata) {
  exch.index[j] ~ dcat(pMix[1:Nmix])
    for (jj in 1:Nmix)  {
      exch[j,jj] <- equals(exch.index[j],jj)
    }
}
# pick theta
for (j in 1:Nstrata)  {
    theta[j] <- LogOdds[exch.index[j],j]
}
# likelihood part
for (i in 1:Nstrata)  {
  logit( p[i] ) <- theta[i]
  p.success[i] <- step(p[i]-p.cut)
  r[i] ~ dbin(p[i],n[i])
}
}

# data list for example of section 2
list(
Nexch=2, Nmix=3,
 pMix = c(0.5,0,0.5),
# alternative weights for EX and EXNEX-2
# pMix = c(1,0,0),
# pMix = c(0.25,0.25,0.5),
Nstrata=10,
```

```
n = c(15,13,12,28,29,29,26, 5, 2, 20),
# original data
r = c( 2, 0, 1, 6, 7, 3, 5, 1, 0, 3),
# scenario 1 and 2 data
#r = c( 7, 0, 1, 6, 7, 3, 5, 1, 0, 3),
#r = c( 7, 0, 1, 3, 3, 3, 8, 2, 1, 6),
# prior means and precisions for EX parameter mu
mu.mean=c(-1.735,0.847), mu.prec=c(0.146,0.266),
# scale parameter of Half-Normal prior for tau
tau.HN.scale=c(1,1),
# NEX priors; make them strata-specific if needed
nex.mean=-1.734, nex.prec=0.128,
p.cut = 0.3
)

# initial values, 2 chains
list(mu = c(-0.5, 0.5))
list(mu = c(0, 0))

# results: 10000 burn-in followed by 40000 updates
node  mean    sd      MC error  2.5%      median   97.5%   start  sample
p[1]  0.1508  0.06311 2.896E-4  0.03888   0.1478   0.2908  10001  80000
p[2]  0.05313 0.05934 2.879E-4  3.056E-4  0.02692  0.1964  10001  80000
p[3]  0.1275  0.06748 3.513E-4  0.01312   0.1274   0.2681  10001  80000
p[4]  0.1881  0.05922 2.64E-4   0.09395   0.18     0.3286  10001  80000
p[5]  0.2052  0.06516 3.719E-4  0.1058    0.1946   0.3626  10001  80000
p[6]  0.1314  0.05111 3.147E-4  0.03814   0.1308   0.2351  10001  80000
p[7]  0.1766  0.05694 2.16E-4   0.08298   0.1701   0.31    10001  80000
p[8]  0.1781  0.1021  3.638E-4  0.02812   0.1617   0.4579  10001  80000
p[9]  0.1309  0.1125  4.567E-4  0.001021  0.1227   0.4232  10001  80000
p[10] 0.1573  0.05789 2.557E-4  0.05504   0.1533   0.2873  10001  80000
```