**Text S1**

**The EM-IWLS Algorithm for Fitting Hierarchical Generalized Linear Models**

We describe our EM-IWLS algorithm for fitting the hierarchical generalized linear models with the linear predictor,

, (S1)

the link function *h*, and the data distribution , where *β*0 is the intercept, *xij*and *zij*represent observed values of ungrouped and grouped variables, respectively, *Xi*contains all variables, and *β* is a vector of all the coefficients and the intercept. For simplicity, we denote  and, where is the total number of variables. We consider the following prior distributions:

 with  set to a large value, ,

, ,  (S2)

For grouped variables, 

For ungrouped variables, *bk*[*j*] is a constant

where ν and *a* are constants, the subscript *k*[*j*] indexes the group *k* that the *j*-th predictor belongs to.

We fit the above hierarchical GLMs by incorporating a flexible EM algorithm into the iteratively weighted least squares (IWLS) for fitting classical GLMs. Our EM algorithm treats the unknown variances and the hyperparameters  and  as missing data. At each step of the iteration, we replace the terms involving the missing values by their conditional expectations, and then update the parameters () by maximizing the expected value of the joint log-posterior density

 (S3)

For the E-step of the algorithm, we take the expectation of the above joint log-posterior density with respect to the conditional posterior distributions,,  and . With the proposed priors, these conditional posterior distributions can be derived as

 (S4)

 (S5)

 (S6)

Therefore, we have the conditional expectations

 (S7)

 (S8)

 (S9)

Since only the terms  () include both the parameters and the missing values, only the conditional expectations  directly affect the M-step.

 In the M-step, we update () by maximizing , where, and  for . This is equivalent to solving the generalized linear model with the normal priors [1-3]. Following the usual IWLS algorithm for fitting GLMs (as implemented in the glm function in R), we approximate the generalized linear model likelihood  by the weighted normal likelihood

 (S10)

 where the ‘normal response’  and ‘weight’ *wi* are called the pseudo-response and pseudo-weight, respectively. The pseudo-response and pseudo-weight are calculated by

 (S11)

where,,, , and  is the current estimate of .

We then incorporate the prior into the weighted normal likelihood as an ‘additional data point’  (the prior mean) with corresponding ‘explanatory variables’ equal to 0 expect *xj*which equals 1 and a ‘residual variance’ [1,4]. Therefore, we can update by running the augmented weighted regression

  (S12)

where  is the vector of all  and all (*J* + 1) prior means 0,  is constructed by the design matrix *X* of the regression  and the identity matrix *I(J+1)*,and  is the diagonal matrix of all pseudo-weights and prior variances. With the augmented, this regression is identified and thus the resulting estimate is well defined and has finite variance, even if the original data are high-dimensional and have collinearity or separation that would result in nonidentifiability of the classical maximum likelihood estimate.

Therefore, we obtain the estimate of , , and its variance

. If a dispersion parameter,, is present, we can update  at each step of the iteration by

 (S13)

In summary, our EM-IWLS algorithm proceeds as follows:

1. Initialize () with some plausible values. For example, we can set = 0, ,  = 1, = 0.5, and  =  = 0.5.
2. For *t* = 1, 2, ··· :

E-step: Calculate the conditional expectations , , and .

M-step:

1. Based on the current value of, calculate the pseudo-data  and the pseudo-weights  using (S11).
2. Update by running the augmented weighted regression (S12).
3. If  is present, update  using (S13).

We apply the criterion in the glmfunction to assess convergence, i.e., , where  =  is the estimate of deviance at the *t*th iteration), and  is a small value (say 10-5). At convergence of the algorithm, we obtain the latest estimates (,) and the covariance matrix . As in the classical framework, the *p*-values for testing the hypotheses H0:  can be calculated using the statistics , which approximately follows a standard normal distribution or a Student-*t* distribution with *n* degrees of freedom, if the dispersion  is not or in the model, respectively.

**References**

1. Gelman A, Jakulin A, Pittau MG, Su YS (2008) A weakly informative default prior distribution for logistic and other regression models. Annals of Applied Statistics 2: 1360-1383.

2. Yi N, Banerjee S (2009) Hierarchical generalized linear models for multiple quantitative trait locus mapping. Genetics 181: 1101-1113.

3. Yi N, Kaklamani VG, Pasche B (2011) Bayesian analysis of genetic interactions in case-control studies, with application to adiponectin genes and colorectal cancer risk. Ann Hum Genet 75: 90-104.

4. Gelman A, Carlin J, Stern H, Rubin D (2003) Bayesian data analysis. London: Chapman and Hall.