Web-based Supplementary Materials for *Bayesian Variable Selection for Latent Class Models* by Joyee Ghosh, Amy H. Herring and Anna Maria Siega-Riz

Web Figure 1

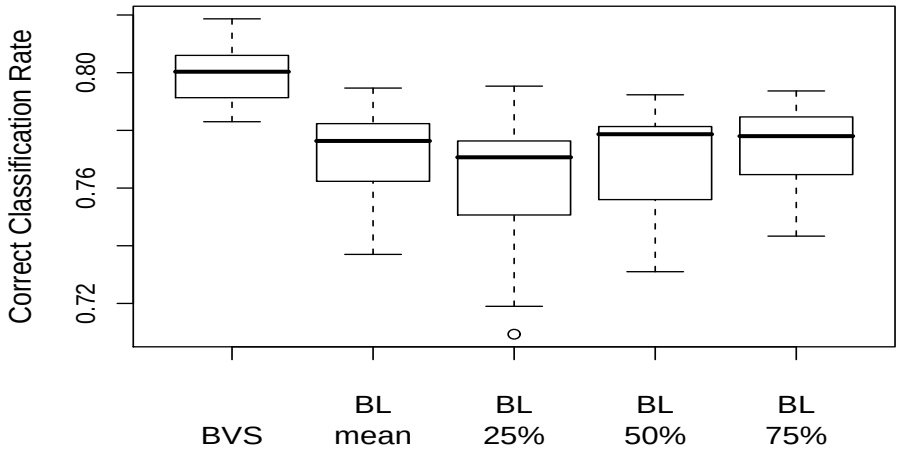

Figure 1: Boxplots showing proportion of correctly classified subjects using the latent class model for 25 (different) simulated datasets with i) Bayesian variable selection (BVS), ii-v) Bayesian Lasso (BL) for four choices of the penalty parameter: mean, 25th, 50th, and 75th percentiles of the prior distribution considered by Park and Casella (2008)

Web Figure 2

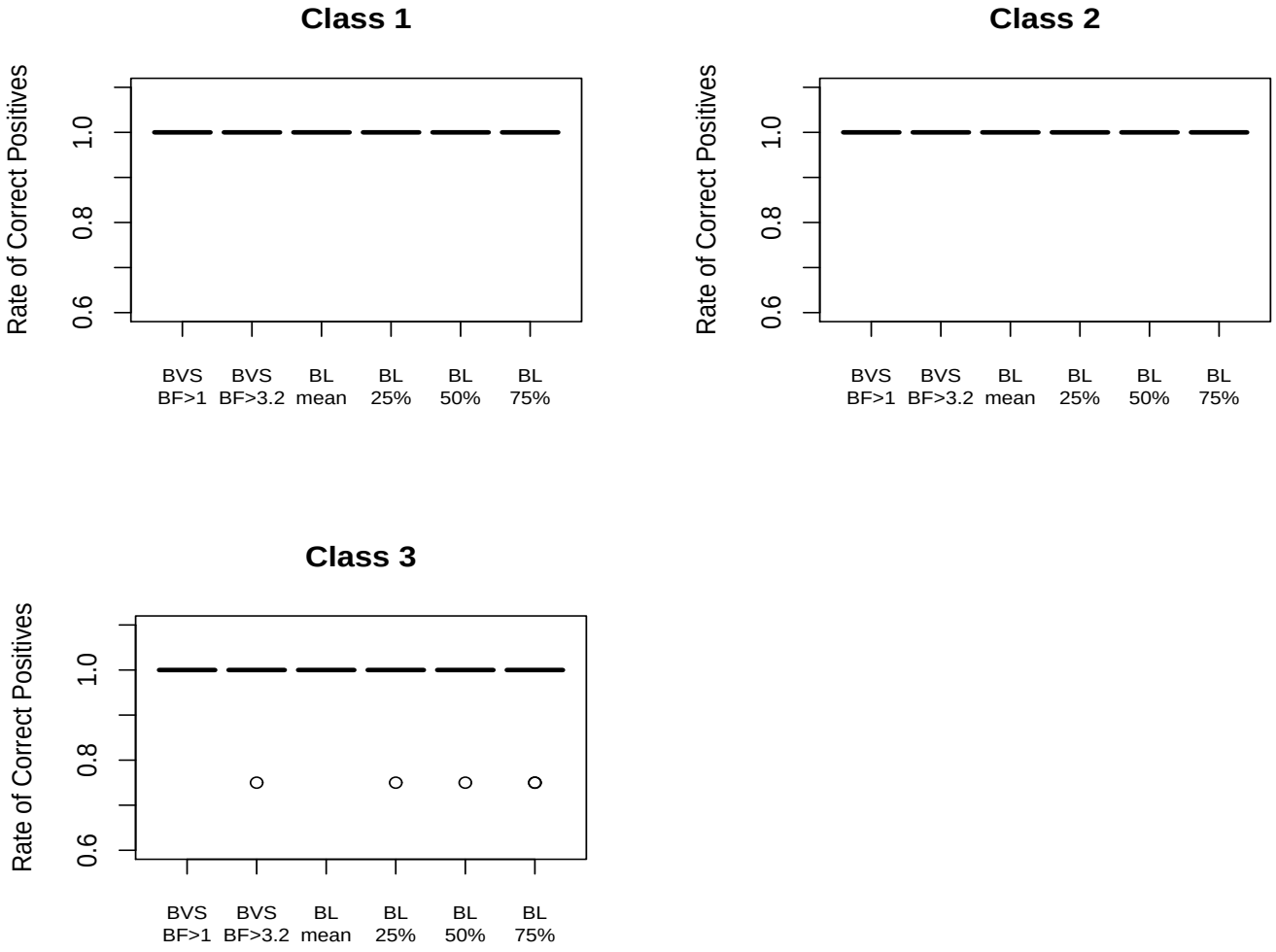

Figure 2: Rate of correct positives (proportion of true covariates flagged as important) using the latent class model for 25 (different) simulated datasets with i-ii) Bayesian variable selection (BVS) for Bayes factor thresholds at 1 and 3.2, iii-vi) Bayesian Lasso (BL) for four choices of the penalty parameter: mean, 25th, 50th, and 75th percentiles of the prior distribution considered by Park and Casella (2008)

Web Figure 3

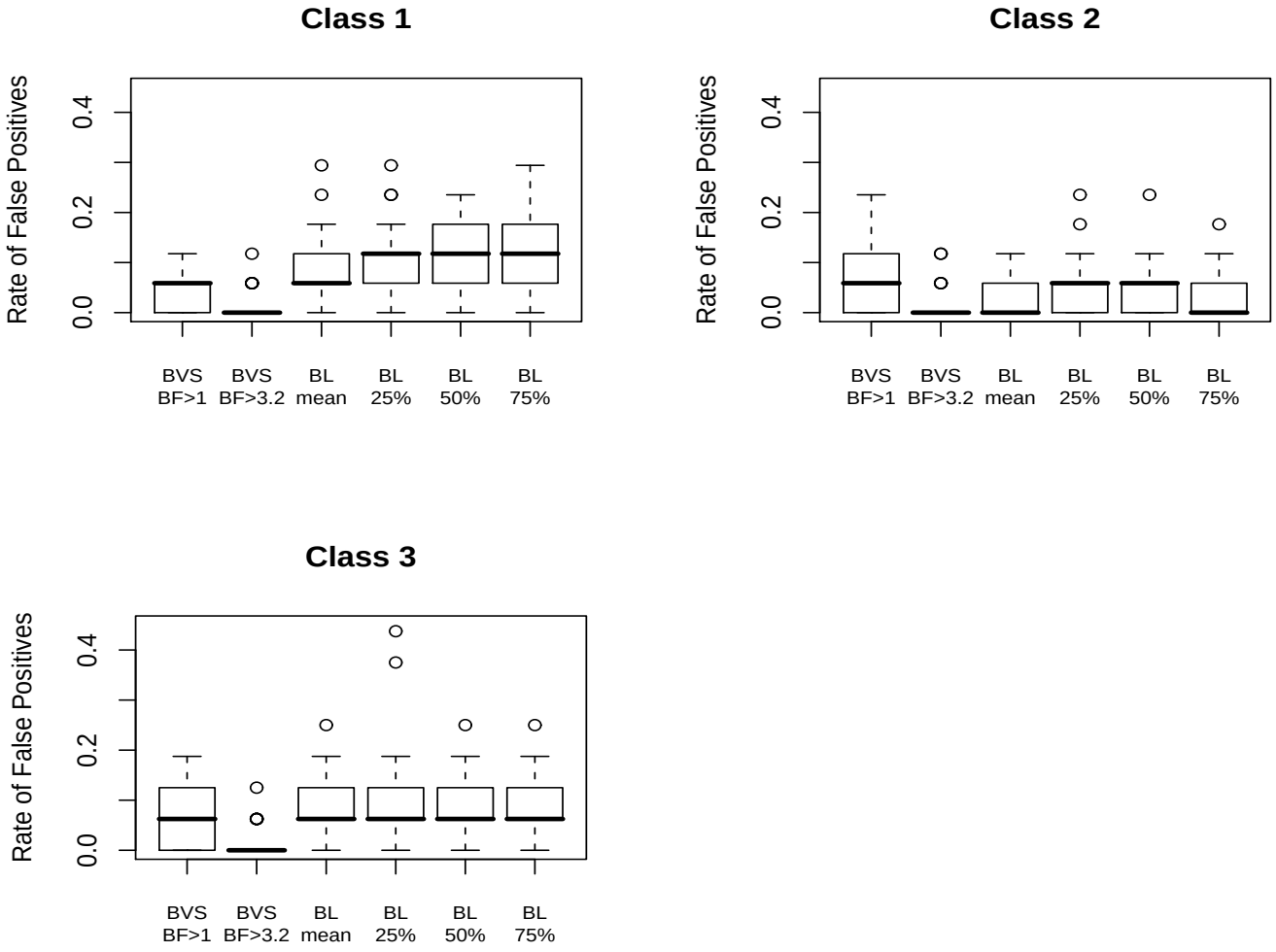

Figure 3: Rate of false positives (proportion of null covariates flagged as important) using the latent class model for 25 (different) simulated datasets with i-ii) Bayesian variable selection (BVS) for Bayes factor thresholds at 1 and 3.2, iii-vi) Bayesian Lasso (BL) for four choices of the penalty parameter: mean, 25th, 50th, and 75th percentiles of the prior distribution considered by Park and Casella (2008)

Web Table 1

Mean Squared Errors (MSE) in estimating $\boldsymbol{\beta}_k$

	$\boldsymbol{\beta}_1$			$\boldsymbol{\beta}_2$			$\boldsymbol{\beta}_3$		
	True	BVS	BL mean	True	BVS	BL mean	True	BVS	BL mean
1	0.80	0.10	0.24	0.30	0.12	0.31	0.30	0.18	1.58
2	1.00	0.03	0.25	0.00	0.01	0.16	1.00	0.04	0.36
3	2.00	0.08	0.68	0.00	0.02	0.09	-2.00	0.11	1.61
4	0.50	0.01	0.05	0.00	0.02	0.05	0.80	0.04	0.32
5	0.00	0.00	0.02	0.00	0.02	0.06	0.90	0.06	0.33
6	0.00	0.00	0.02	0.00	0.00	0.04	0.00	0.01	0.11
7	0.00	0.00	0.02	0.00	0.00	0.03	0.00	0.01	0.08
8	0.00	0.00	0.03	0.00	0.01	0.03	0.00	0.00	0.07
9	0.00	0.00	0.03	0.00	0.00	0.02	0.00	0.01	0.10
10	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.01	0.09
11	0.00	0.00	0.01	0.00	0.01	0.03	0.00	0.01	0.11
12	0.00	0.01	0.04	0.00	0.01	0.04	0.00	0.00	0.05
13	0.00	0.00	0.04	0.00	0.01	0.02	0.00	0.02	0.13
14	0.00	0.00	0.01	0.00	0.00	0.03	0.00	0.01	0.08
15	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.09
16	0.00	0.00	0.03	0.00	0.01	0.03	0.00	0.01	0.10
17	0.00	0.00	0.03	0.00	0.00	0.02	0.00	0.00	0.06
18	0.00	0.00	0.02	0.00	0.01	0.05	0.00	0.00	0.10
19	0.00	0.01	0.06	-1.00	0.04	0.13	0.00	0.00	0.09
20	0.00	0.01	0.07	1.70	0.05	0.33	0.00	0.01	0.19
21	0.00	0.00	0.10	-2.00	0.11	0.33	0.00	0.01	0.18

Table 1: True values of class-specific regression coefficients, $\boldsymbol{\beta}_k$, for the simulation study and the corresponding mean squared errors (MSE) calculated from 25 (different) simulated datasets, using Bayesian variable selection (BVS) and Bayesian Lasso with the penalty parameter set equal to the mean (BL mean) of the prior distribution considered by Park and Casella (2008)

Web Figure 4

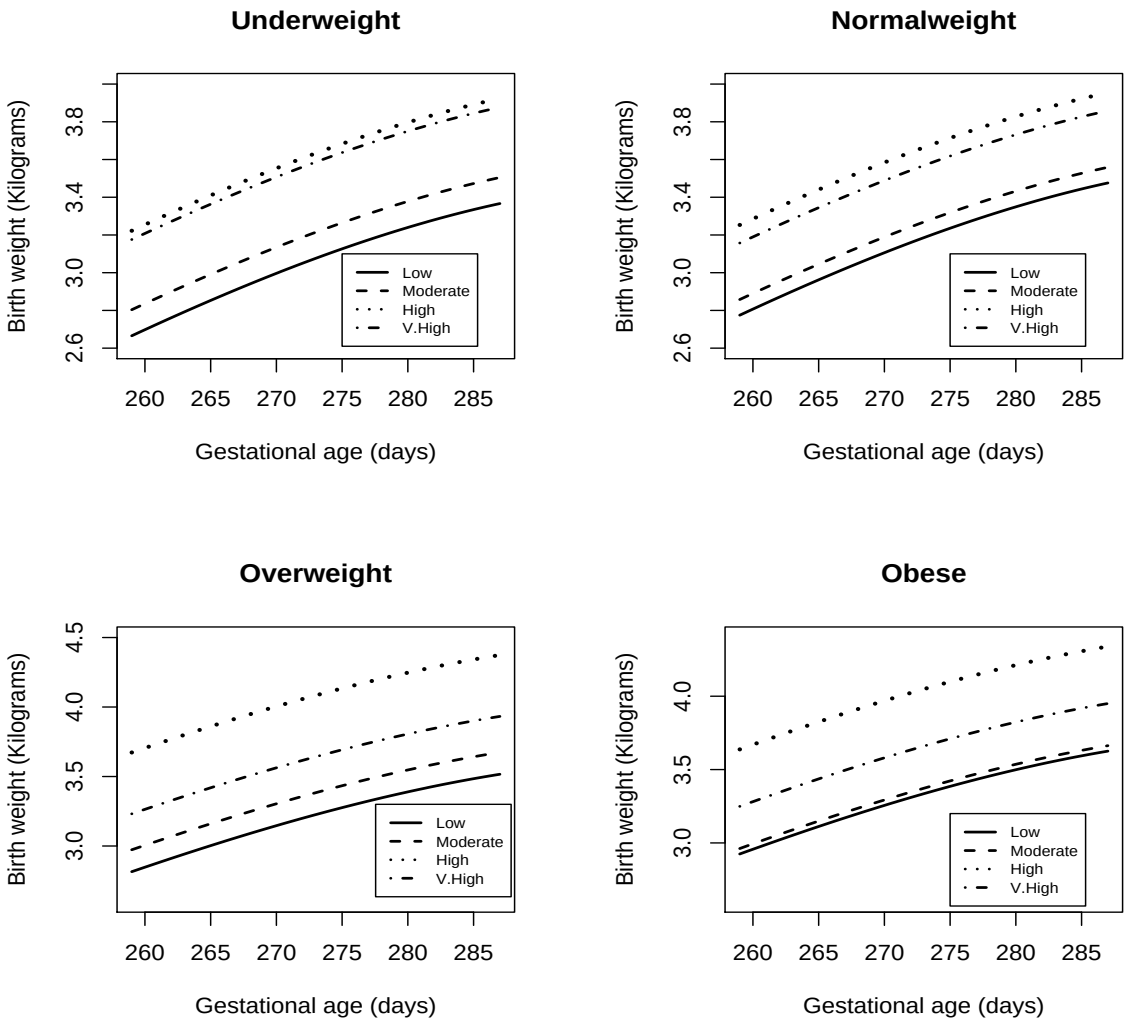

Figure 4: Expected birth weight as a function of gestational age for latent class model, grouped by BMI categories

Web Figure 5

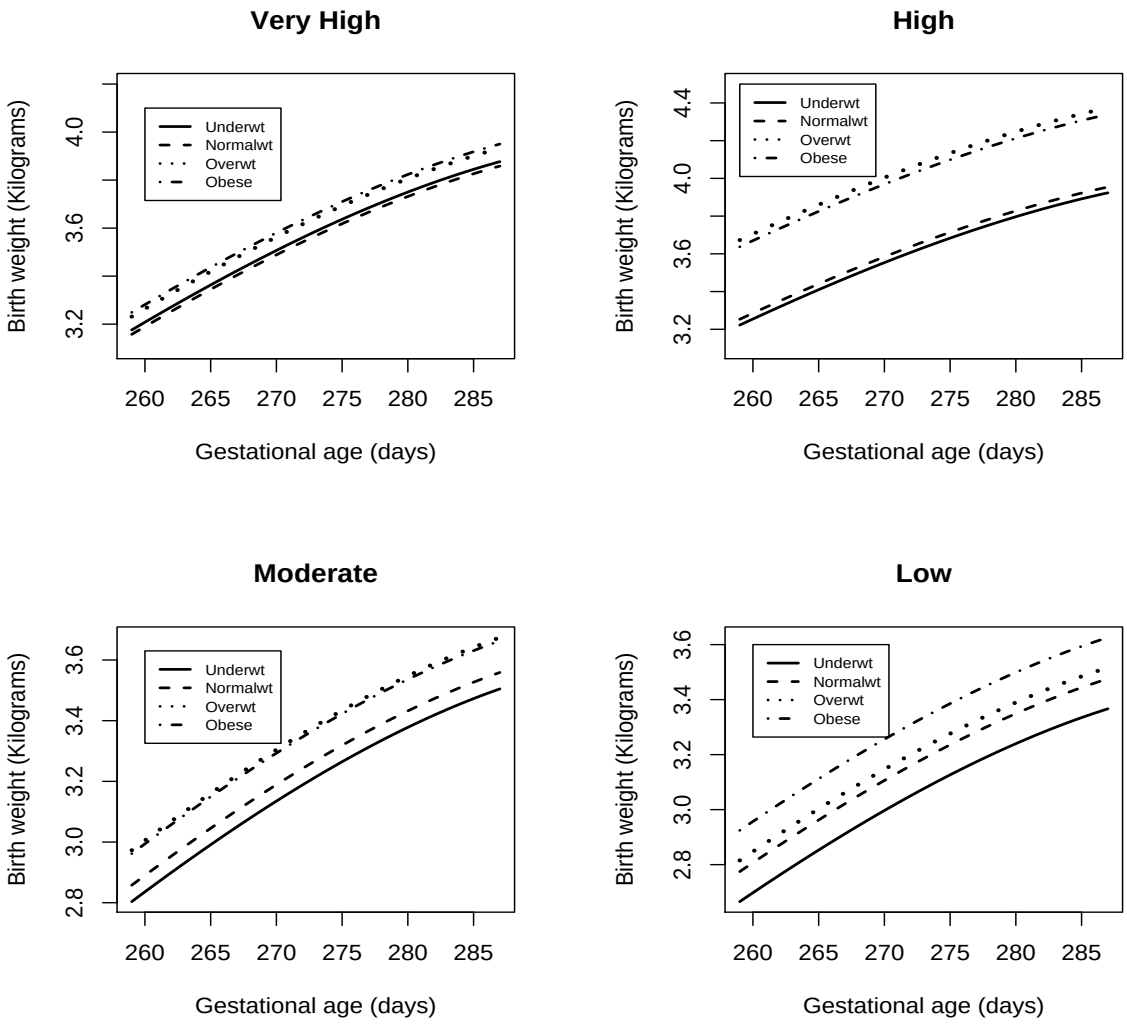

Figure 5: Expected birth weight as a function of gestational age for latent class model, grouped by latent classes

Web Figure 6

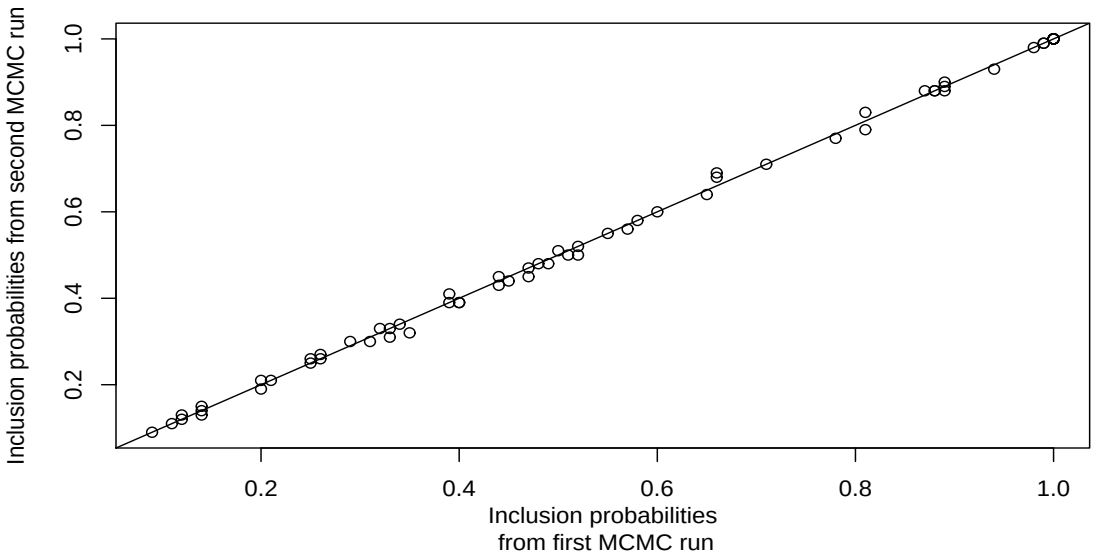

Figure 6: Convergence of marginal posterior inclusion probabilities of class-specific inclusion indicators for covariates, for the real data

Web Figure 7

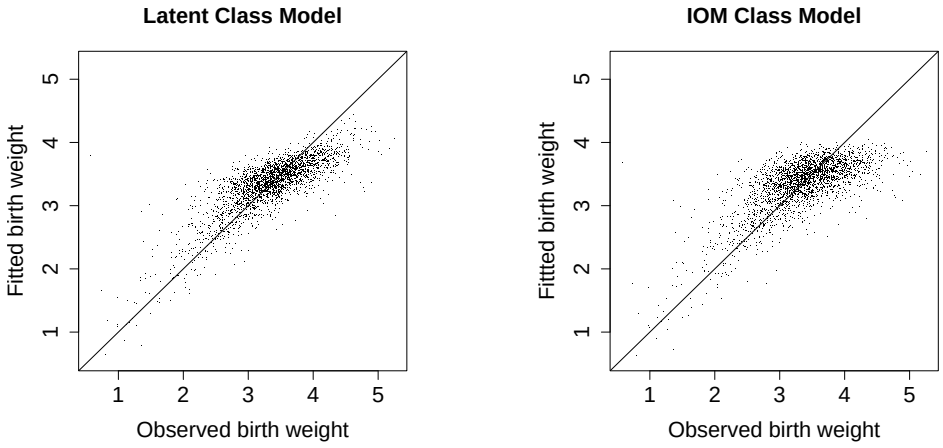

Figure 7: Observed and fitted values of birth weight (in Kilograms) using latent class model and Institute of Medicine (IOM) class model; the correlation between observed and fitted values for latent and IOM class models are 0.81 and 0.70 respectively; the square-root of mean squared error (RMSE) are 0.35 and 0.42 respectively

Web Figure 8

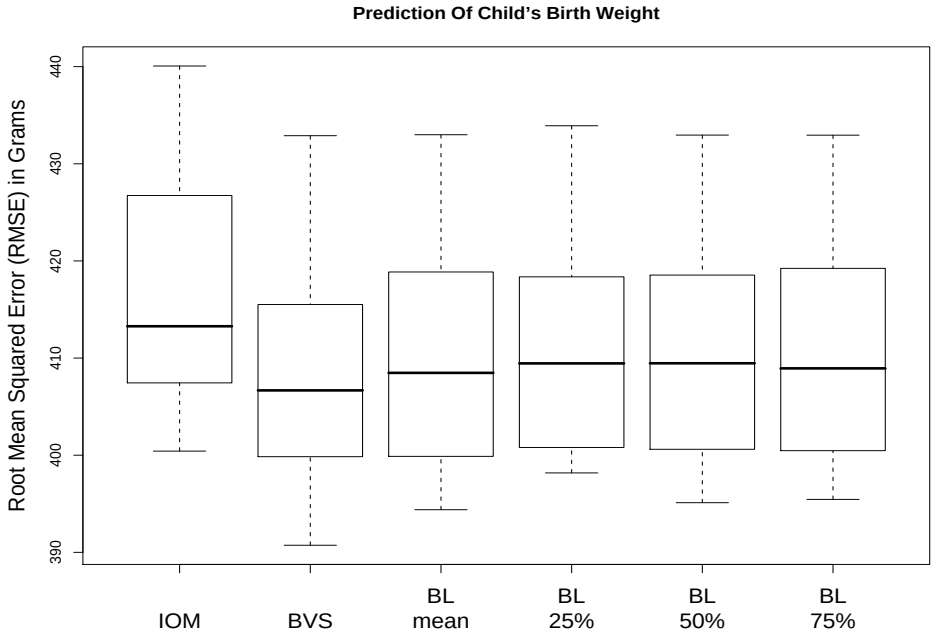

Figure 8: Boxplots of root mean squared error (RMSE) for 10 different test samples of size 266 women for i) Institute of Medicine class model (IOM), ii) latent class model with Bayesian variable selection (BVS), iii-vi) latent class model with Bayesian Lasso (BL) for four choices of the penalty parameter: mean, 25th, 50th, and 75th percentiles of the prior distribution considered by Park and Casella (2008)

Web Table 2

Table 2: Table showing posterior means and 95% credible intervals for birth weight regression coefficients with *inadequate* (iom.inad) and *excessive* (iom.ex) IOM class indicators as covariates

	$\hat{E}(\alpha_j\|Y)$	95% C.I.
Intercept	3.19	(3.13 , 3.25)
gestational age	0.38	(0.37 , 0.4)
(gestational age)^2	-0.06	(-0.07 , -0.04)
gestational age)^3	-0.04	(-0.05 , -0.02)
non-smoker	0.13	(0.09 , 0.17)
nulliparous	-0.14	(-0.17 , -0.11)
male	0.1	(0.08 , 0.14)
mom race black	-0.16	(-0.19 , -0.12)
underweight	-0.06	(-0.13 , 0.01)
overweight	0.06	(-0.08 , 0.19)
obese	0	(-0.14 , 0.14)
iom.inad	-0.04	(-0.11 , 0.03)
iom.ex	0.13	(0.08 , 0.18)
iom.inad*underweight	-0.02	(-0.15 , 0.1)
iom.ex*underweight	0.02	(-0.08 , 0.12)
iom.inad*overweight	-0.1	(-0.31 , 0.11)
iom.ex*overweight	0	(-0.15 , 0.15)
iom.inad*obese	0.05	(-0.11 , 0.2)
iom.ex*obese	0.12	(-0.02 , 0.27)