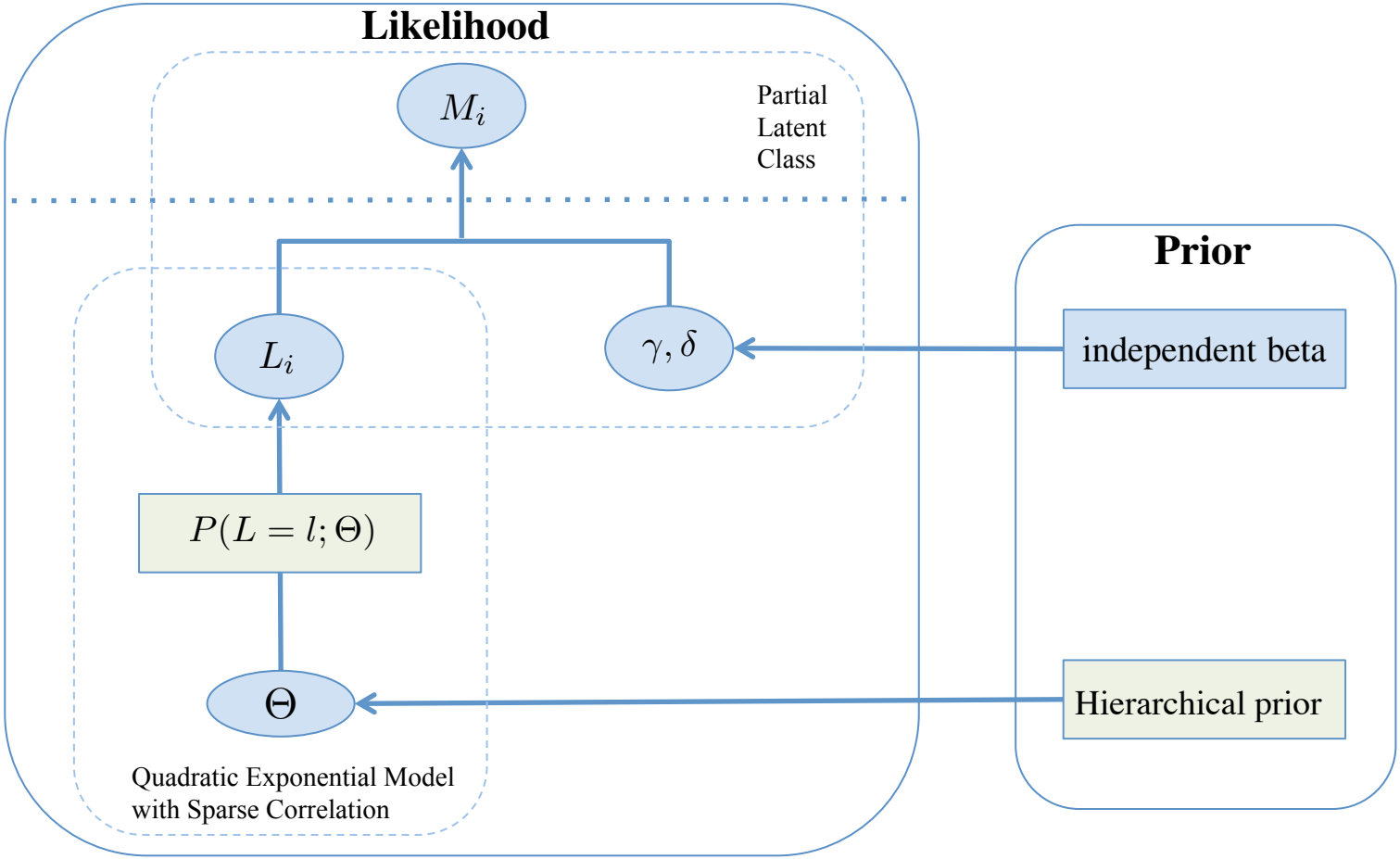
Model Hierarchy



Latent Variable

Regular Quadratic Exponential Model:

$$P(L = l; \Theta) = \exp\{\Theta_1^T l + \Theta_2^T u_2\}/A(\Theta)$$

$$\text{with } A(\Theta) = \sum_{l^* \in \{0,1\}^K} \exp\{\Theta^T l^*\}$$

$\theta_{ik}^{(1)} = X_i^T \beta_k$, where X_i is the vector of covariates.

Sparse Correlation:

$$\Theta_2 = \theta_2 \cdot (I_1, \ldots, I_{\binom{K}{2}})$$

Hierarchical Prior:

$$I_j \sim \text{Bernoulli}(p)$$

$$p \sim \text{Beta}(a, b)$$

Simulation Study

5 candidate etiological pathogens.
BS are available for each of them.
500 cases and 1000 controls in each data set.
200 independent data sets simulated.
Two binary covariates: age and severity.

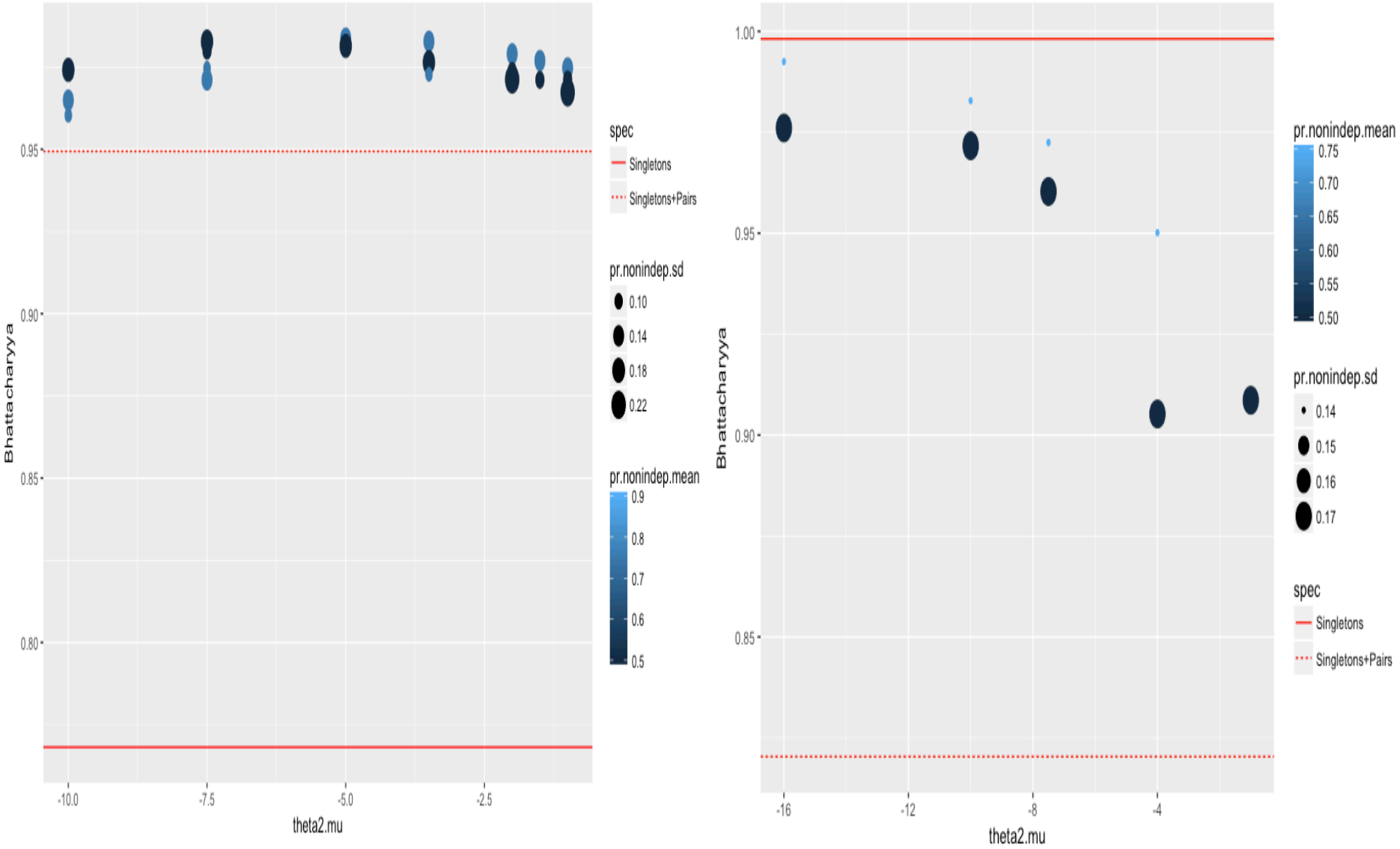
Simulation I-a:
Allow multiple pathogen infect lung at the same time. Pr(2 or more) ≈ 0.35
All pathogens have SS measurements.
Relatively low-quality data:
SS TPR ≈ 0.1 BS TPR ≈ 0.6 BS FPR ≈ 0.45

Simulation I-b:
Only pathogen D, E have SS measurements.

Simulation II:
Relatively high-quality data:
SS TPR ≈ 0.8 BS TPR ≈ 0.9 BS FPR ≈ 0.05

Simulation III:
Only single-pathogen infection is possible.

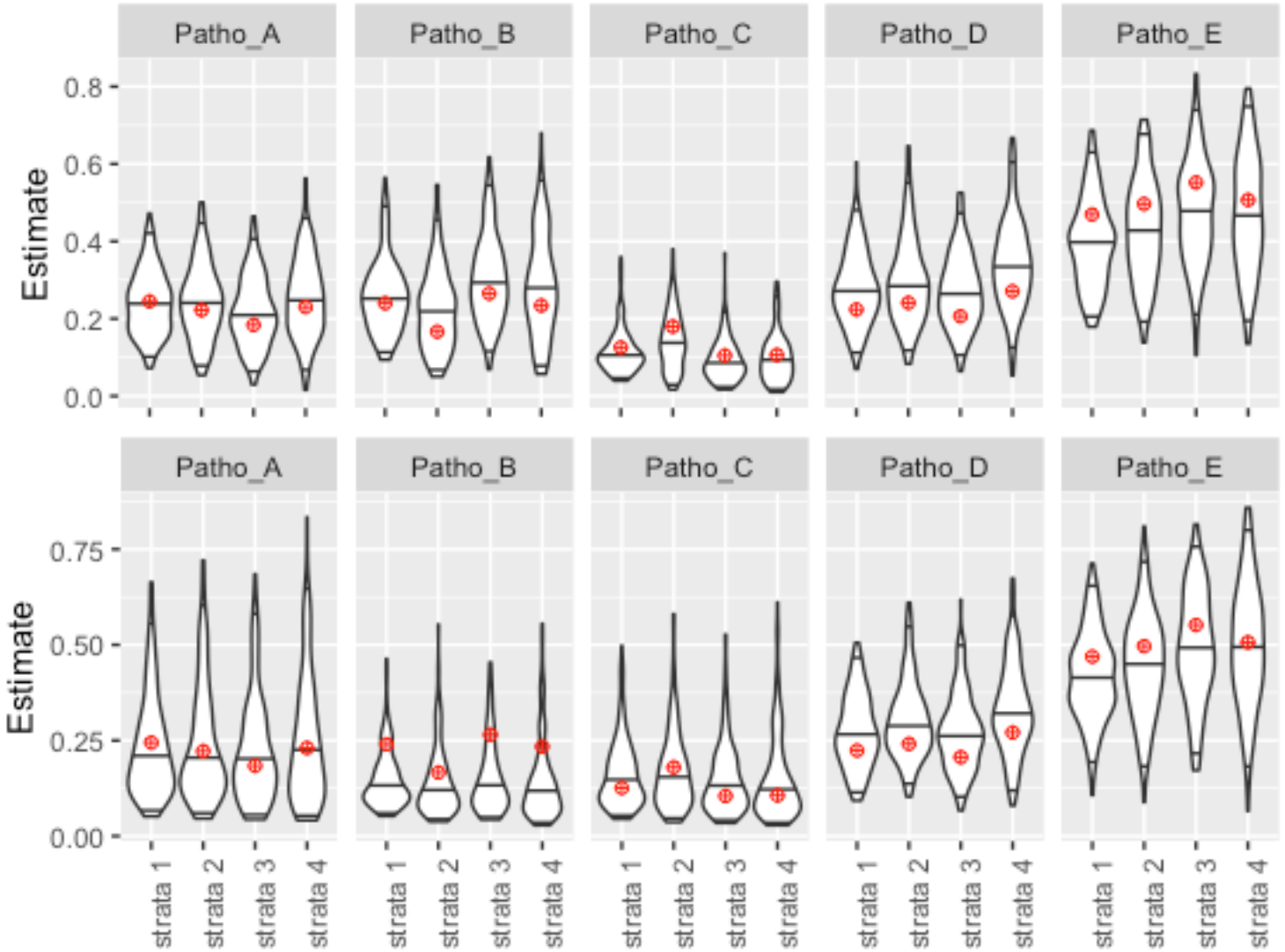
Simulation I and III vs. pLCM



Simulation I-a vs. I-b

SS available for all five pathogens:

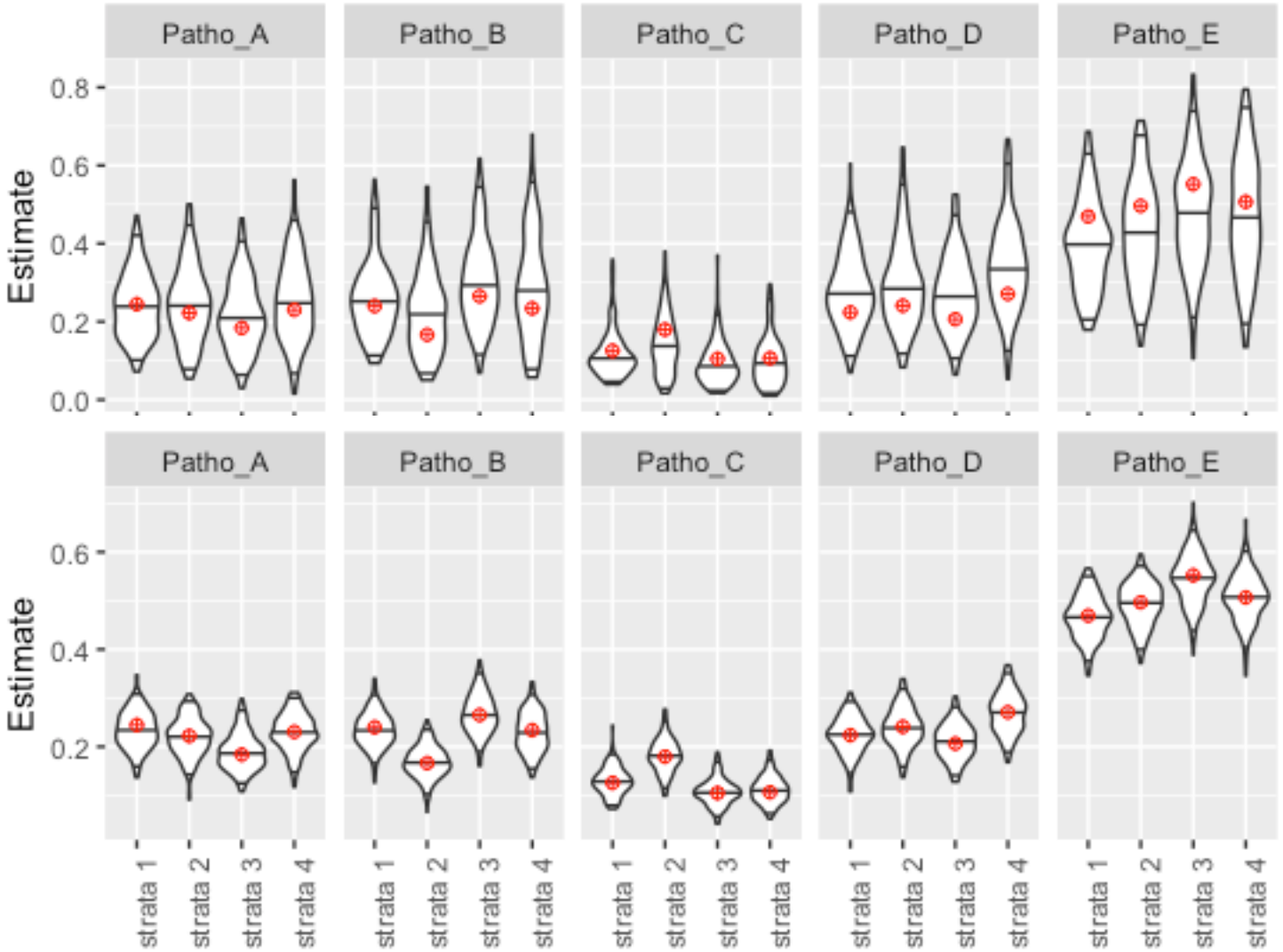
SS available for pathogen D and E only:



Simulation I-a vs. II

Low-quality data:

High-quality data:



Appendix

β_0	0.210	-0.282	-0.835	-0.205	1.068
β_{age}	-0.1	-0.5	0.5	0.2	0.1
β_{sev}	-0.3	0.2	-0.2	-0.1	0.3
$\beta_{a:s}$	0.4	0.3	-0.4	0.2	-0.2

Θ_2 = (-1.5, -1.5, -1.5, -1.5, 0.0, 0.0, -1.5, -1.5, -1.5, -1.5)

Mu	[1]	[2]	[3]	[4]	[5]
00	0.245	0.240	0.126	0.224	0.469
10	0.222	0.166	0.180	0.241	0.497
01	0.184	0.265	0.105	0.206	0.552
11	0.230	0.234	0.106	0.271	0.507

Pr(s)	[0]	[1]	[2]	[3]	[4]	[5]
00	0.090	0.550	0.329	0.031	0.000	0
10	0.085	0.555	0.328	0.031	0.000	0
01	0.081	0.558	0.329	0.032	0.000	0
11	0.076	0.536	0.351	0.036	0.001	0