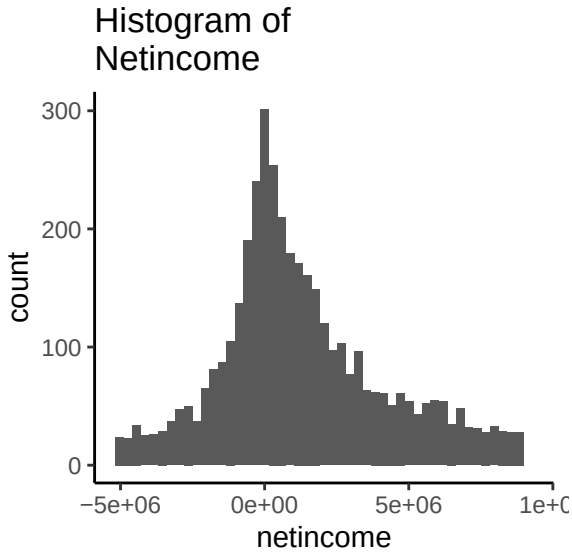
Lab 4

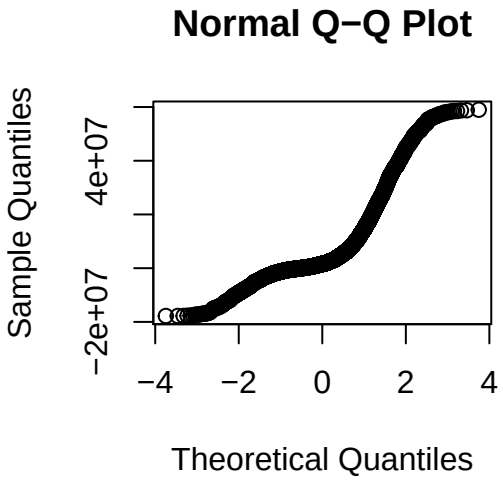
Aasha Reddy Shirley Mathur Bo Liu Marc Brooks

9/17/2021

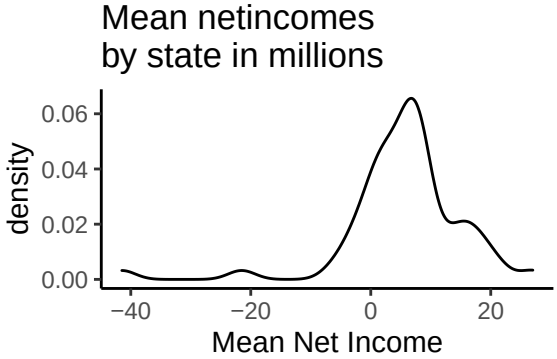
EDA

According to the histogram, the distribution of netincome is relatively normal (we restrict the x axis in the below graph). The distribution is somewhat right-skewed. We do not believe a log transformation of netincome is appropriate given this would eliminate all of the negative values. The QQ plot is roughly normal (in the below plot we trim off the top and bottom 5% of values).

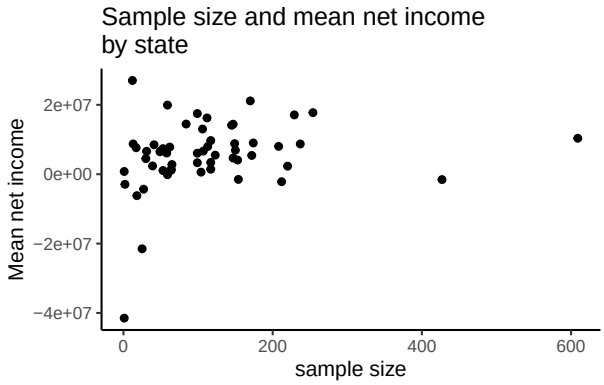


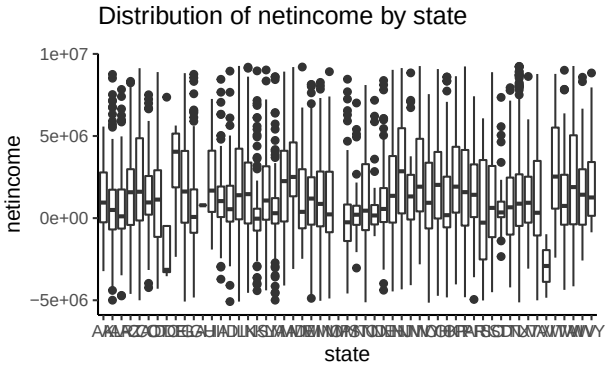


We can see below that the distribution of the mean net income by state is also relatively normal which is an assumption we care about for our model. This justifies a state specific intercept.

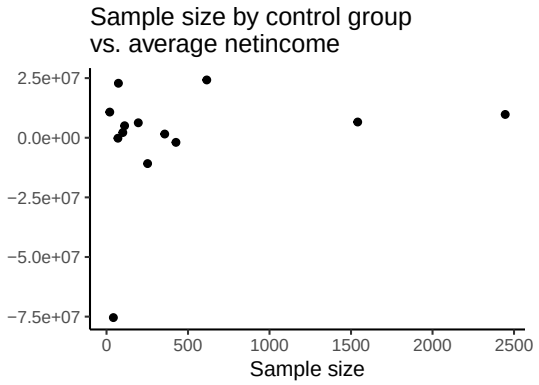


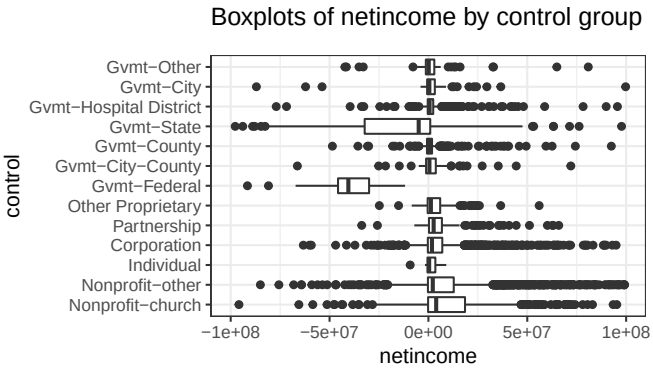
The boxplots below demonstrate that netincome varies by state, which again justifies a random intercept.





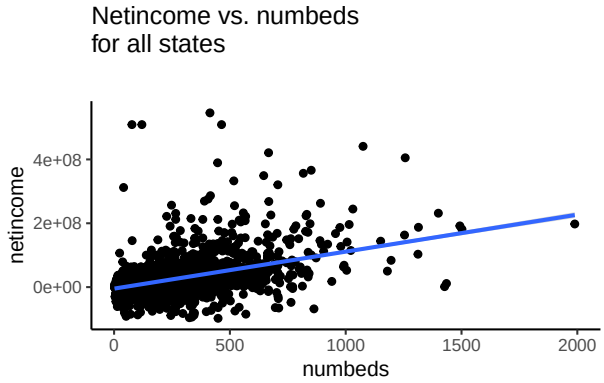
Although the scatter plot above shows that some states with smaller sample sizes have more extreme values, we decide not to disregard these potential outliers in our model.

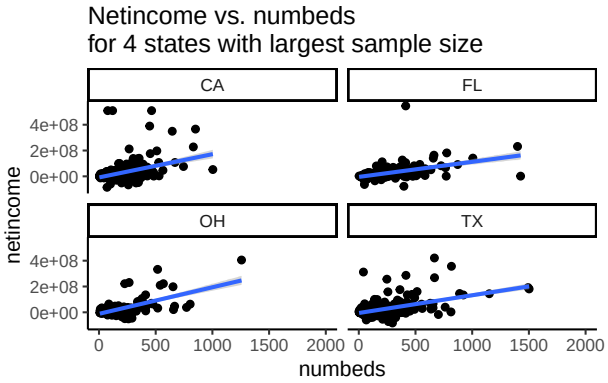




We also would like to examine the relationship between netincome and the control grouping variable. The plot above and on the left suggests that control groups have a sufficient sample size (the smallest being 42 observations). As we are using random effects, it also does not matter much how large our sample sizes are since we expect shrinkage towards the mean.

We can see based on the above boxplots of netincome by control group that most of the groups are centered around a similar mean. For this reason, we could try to include a random intercept.





Examining the relationship between netincome and number of beds, we see that the slope for each state is similar to the slope across all the states. This justifies a fixed effect for numbeds.

Model Specification

$$y_{ijk} = \mu + \alpha_j + \gamma_k + \beta * B_{ij} + \epsilon_{ijk}$$

where i indicates the hospital i, j indicates the state j and k is the type of control. B_{ij} is numbeds.

$$\alpha_j \sim \mathcal{N}(0, \tau_\alpha^2), \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2), \quad \gamma_k \sim \mathcal{N}(0, \tau_\gamma^2)$$

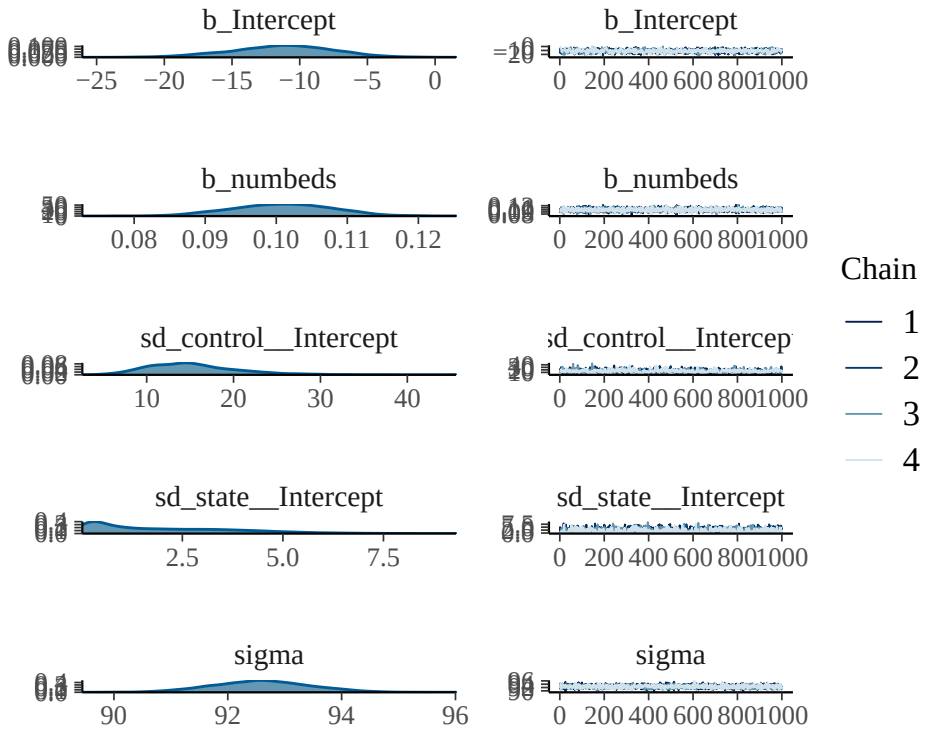
We set the below priors. We selected priors that looked appropriate for our model. For the tau's, we are setting very weak priors because we know very little about variance of the random effects. For the grand mean, we assume a normal prior of N(0, 5) since the means were relatively centered around 0. From our intuition, a variance of 5 makes sense for our prior on μ.

$$\mu \sim \mathcal{N}(0, 5), \quad \sigma^2 \sim \mathcal{IG}(0.5, 5), \quad \beta \sim \mathcal{N}(0, 5)$$
$$\tau_{alpha}^2 \sim \mathcal{IG}(0.1, 0.1), \quad \tau_{alpha}^2 \sim \mathcal{IG}(0.1, 0.1)$$

For our model, we scaled the netincome values by dividing them by 1,000,000.

Evaluate diagnostics

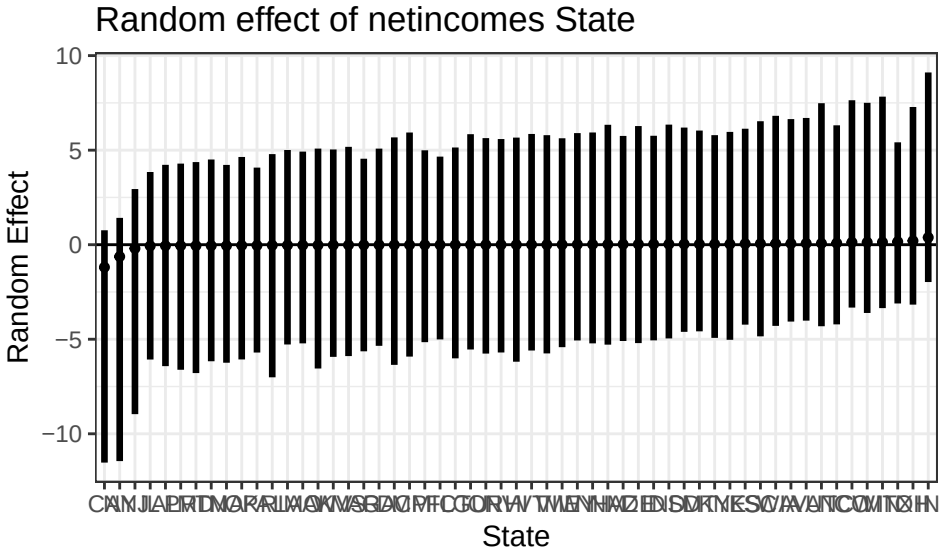
We can see from the below plots that the chains seem to converge.



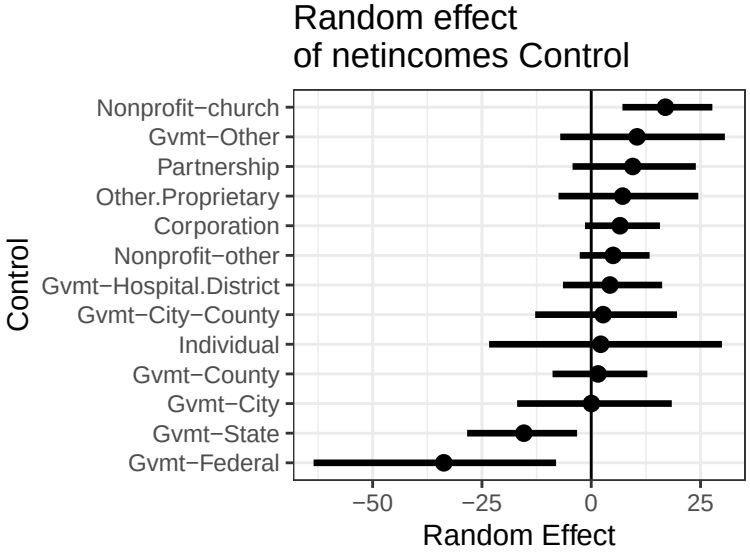
Results

For most of the states, the mean is very similar to each other. The random effects for state are relatively centered around 0. The intervals are also very similar across the states. No point estimates are outside of the other state intervals, so there is not enough evidence to say that the random effects are different between any

two hospitals (not really enough information to provide a ranking). We can provide a ranking based on the point estimate but there is a lot of uncertainty around this.



Government-federal and government-state owned hospitals seems to have on average a negative effect on netincome. Non-profit church owned hospitals seems to have on average a positive effect on netincome.



Below are our estimates of the parameters, including the 95% credible intervals. We can see that our estimate of μ, the grand mean is -11.321 million (credible interval of [-18.885, -4.311]). This suggests that the hospitals at baseline (no beds) are expected to on average lose money.

Our estimate of numbeds is 0.101 (credible interval of [0.087, 0.114]). We expect netincome to increase by 0.101 million on average for each additional bed.

This is evidence that health care costs too much money!

Table 1: Estimated posterior parameters

	Est	Lwr	Upr
Intercept (grand mean)	-11.321	-18.885	-4.311
numbeds	0.101	0.087	0.114
sigma	92.591	90.958	94.241
sd_control	15.285	7.074	27.497
sd_state	1.978	0.064	5.875

Table 2: Top 5 estimated group means of netincome by state

state	Group Means	.lower	.upper
CA	-13.540	-25.597	-5.314
NY	-12.898	-25.151	-5.030
NJ	-12.054	-22.507	-4.227
IL	-11.586	-20.703	-3.666
MD	-11.571	-21.202	-3.277

Table 3: Top 5 estimated group means of netincome by state

state	Group Means	.lower	.upper
IN	-10.054	-18.372	0.049
OH	-10.446	-18.817	-1.477
CO	-10.551	-19.031	-1.157
WI	-10.568	-19.329	-0.967
MN	-10.571	-18.978	-1.493

Table 4: Bottom 5 estimated group means of netincome by control

control	Group Means	.lower	.upper
Corporation	-4.504	-9.439	0.317
Gvmt-City	-10.856	-28.617	5.773
Gvmt-City-County	-8.504	-23.403	6.858
Gvmt-County	-9.669	-18.249	-1.282
Gvmt-Federal	-45.010	-74.975	-17.320
Gvmt-Hospital.District	-6.794	-16.226	2.687
Gvmt-Other	-0.684	-17.956	18.198
Gvmt-State	-26.740	-38.799	-15.098
Individual	-8.935	-35.535	18.776
Nonprofit-church	5.772	-1.867	13.354
Nonprofit-other	-6.173	-10.474	-1.641
Other.Proprietary	-4.002	-18.627	11.848
Partnership	-1.594	-14.504	10.878