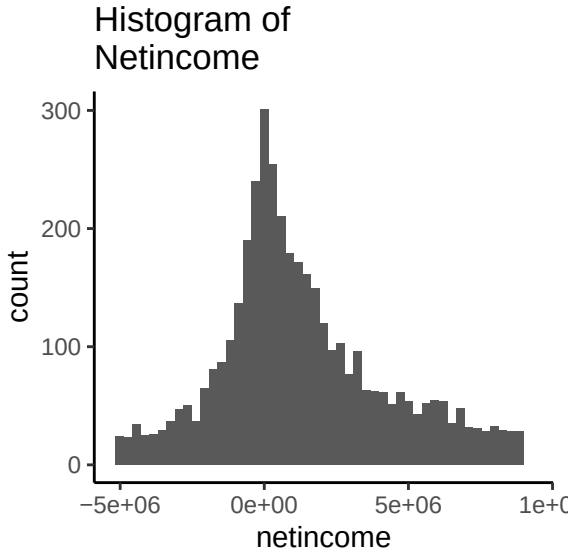
Lab 4

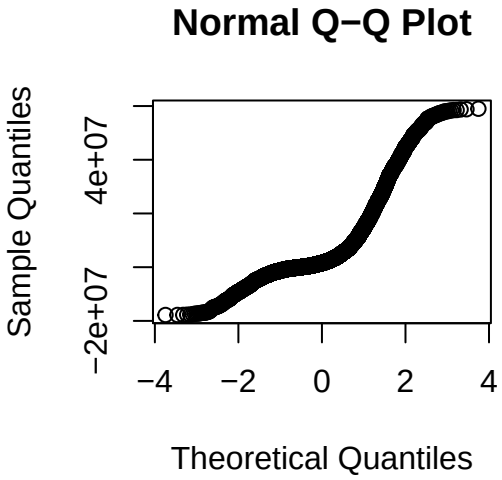
Aasha Reddy Shirley Mathur Bo Liu Marc Brooks

9/17/2021

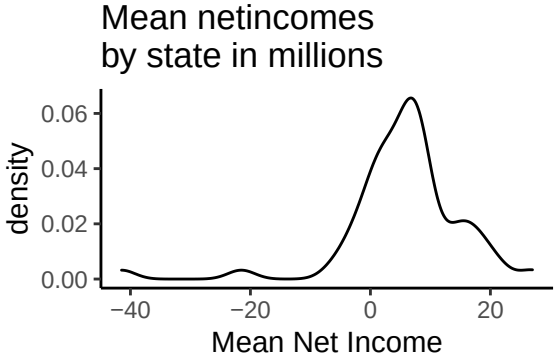
EDA

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).

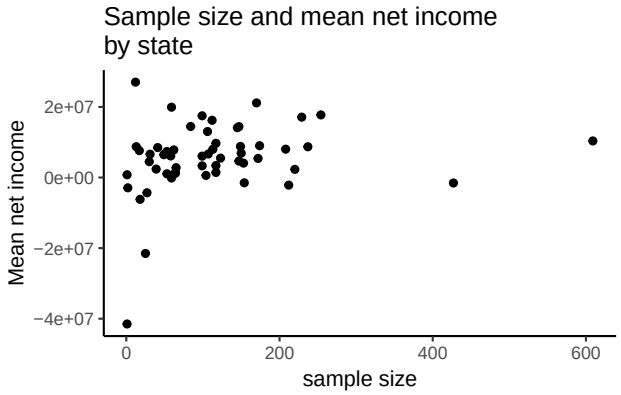


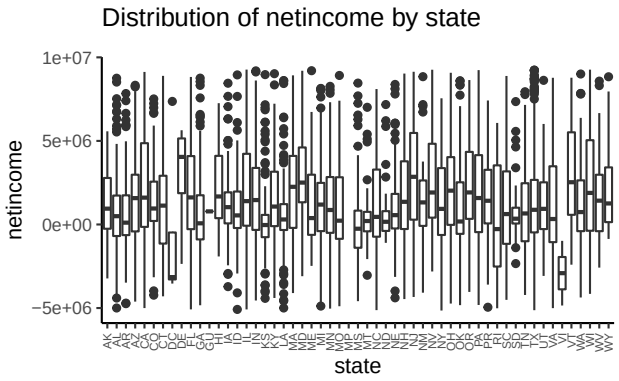


We can see below that the distribution of the mean net income by state is also relatively normal which is an assumption we care about for our model. This justifies a state specific intercept.

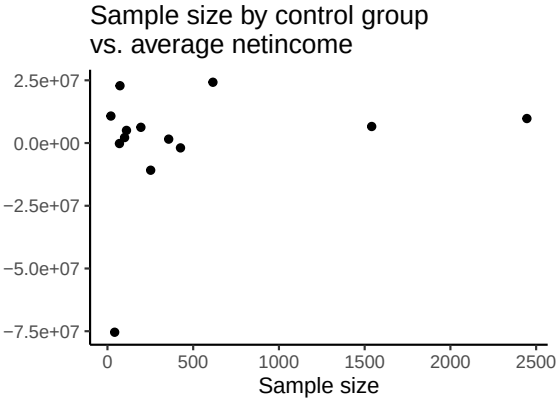


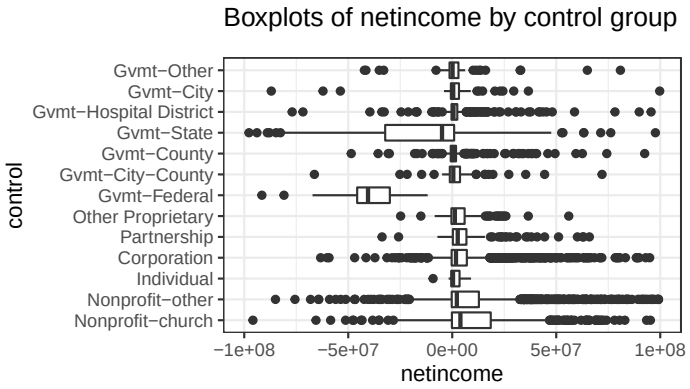
The boxplots below demonstrate that netincome varies by state, which again justifies a random intercept.





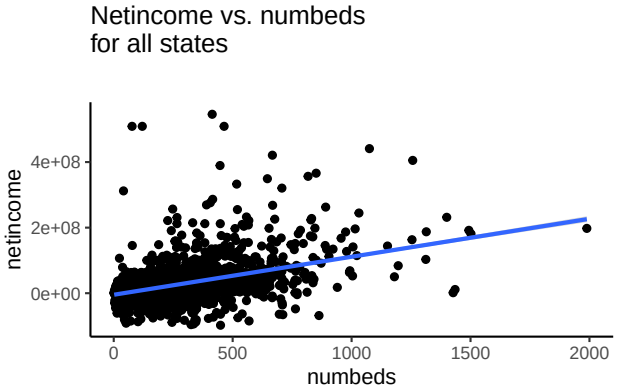
Although the scatter plot above shows that some states with smaller sample sizes have more extreme values, we decide not to disregard these potential outliers in our model.

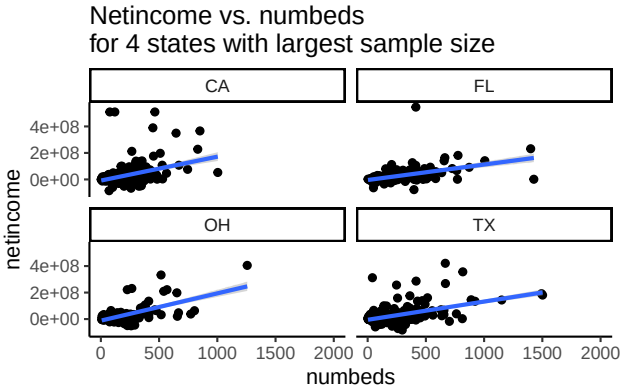




We also would like to examine the relationship between netincome and the control grouping variable. The plot above and on the left suggests that control groups have a sufficient sample size (the smallest being 42 observations). As we are using random effects, it also does not matter much how large our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

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

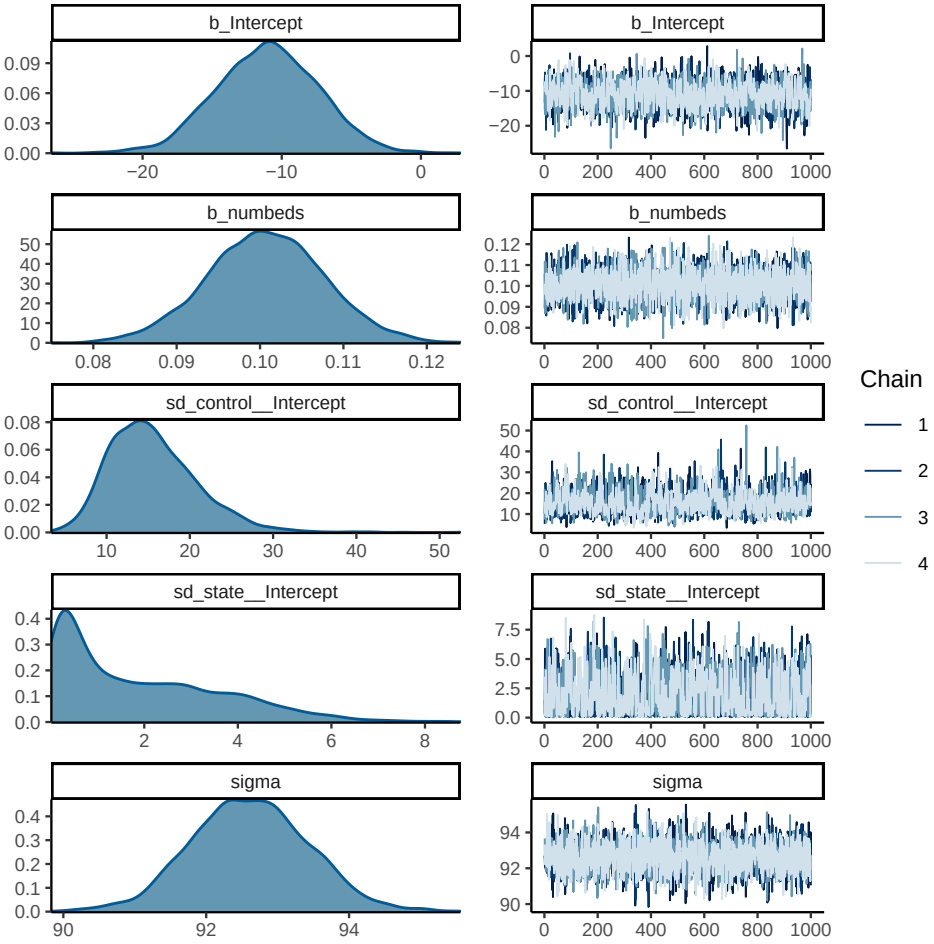
We set the below priors. We selected priors that looked appropriate for our model. For the tau's, we are setting very weak priors because we know very little about variance of the random effects. For the grand mean, we assume a normal prior of N(0, 5) since the means were relatively centered around 0.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

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

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

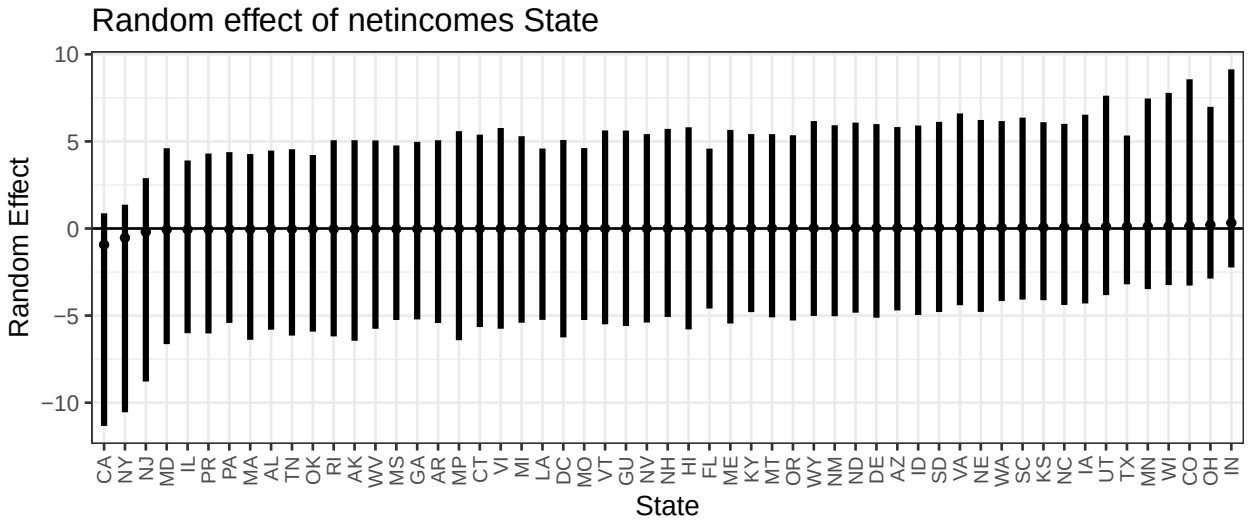
Evaluate diagnostics

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

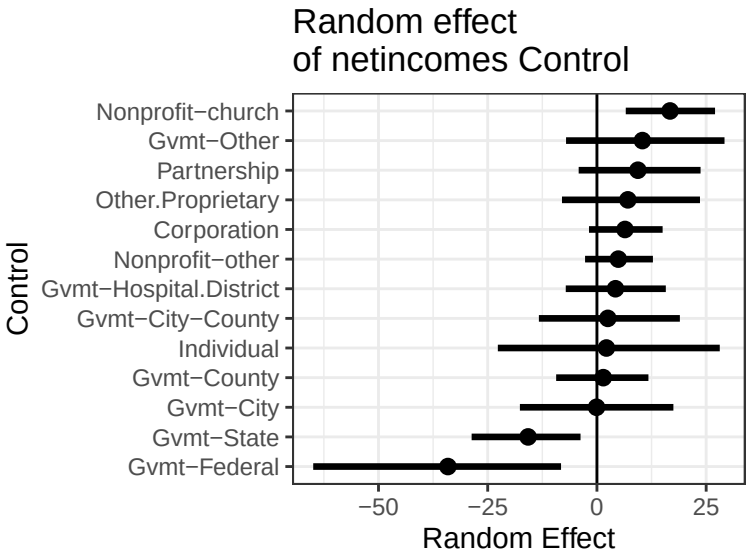


Results

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)	-11.024	-18.114	-4.028
numbeds	0.101	0.087	0.114
sigma	92.613	91.036	94.217
sd_control	15.596	7.298	27.367
sd_state	1.893	0.066	5.825

Table 2: Top (left) and bottom (right) 5 estimated group means of netincome by state

State	Group Mean	Lower	Upper	State	Group Mean	Lower	Upper
IN	-9.799	-17.984	0.303	CA	-13.133	-24.920	-4.925
OH	-10.227	-18.141	-1.148	NY	-12.630	-24.076	-4.610
CO	-10.390	-18.605	0.084	NJ	-11.892	-22.023	-3.979
WI	-10.430	-18.624	-0.632	MA	-11.377	-20.090	-2.890
UT	-10.457	-18.734	-0.948	IL	-11.355	-20.323	-3.425

Table 3: Estimated group means of netincome by control

control	Group Means	.lower	.upper
Nonprofit-church	5.774	-2.008	13.375
Gvmt-Other	-0.559	-17.643	17.941
Partnership	-1.698	-13.710	11.046
Other.Proprietary	-3.857	-18.151	11.577
Corporation	-4.547	-9.493	0.225
Nonprofit-other	-6.073	-10.484	-1.894
Gvmt-Hospital.District	-6.697	-16.549	2.784
Gvmt-City-County	-8.363	-23.949	6.643
Individual	-8.505	-34.159	16.801
Gvmt-County	-9.583	-18.083	-1.453
Gvmt-City	-10.922	-29.067	6.047
Gvmt-State	-26.894	-38.711	-15.614
Gvmt-Federal	-45.378	-76.305	-17.830