Simulated Risk Model

James T. Bang

September 5, 2022

Data Generation and Estimation Equations

Data Generation

The data for observed patient outcomes of readmission, home health care (HHC) assignment, patient demographics, prior hospitalization, diagnosis, and payer type come from a single hospital in the READI study (Weiss, et al., 2019).

Predictive models of risk attempt to assign patients into risk categories based on data known prior to discharge, and hospitals increasingly use such risk models to assign patients to post-discharge interventions like HHC. The challenge is simulating the patient's latent "risk" for the adverse outcome of readmission. We propose that this is not the same as the *ex post **likelihood*** of readmission. We also propose that patients with higher risk of adverse outcomes - including readmission - will be more likely to receive HHC assignments, and that the effectiveness of the clinician's decision to refer the patient to interventions like HHC at discharge masks this *ex ante **risk***. Hence, there will be a bias in the using the observed likelihood to measure risk, and therefore a bias in the observed effectiveness of the intervention of assigning HHC.

To simulate this bias, suppose that "treated" HHC patients exhibit an *ex post* readmission *prevalence*, π, but that the treatment carries a proportional *effectiveness* equal to η. Then, the overall *risk* of readmission would be:

$$\rho = \frac{\pi}{1-\eta}.$$

We can express the *bias* of any *ex ante* prediction based on pre-discharge observables as the true risk (ρ) as a proportion of the fraction of HHC patients who did *not* eventually readmit themselves. This bias, B equals:

$$B = \frac{\rho}{1-\pi}.$$

Substituting (1) into (2) and simplifying, we get:

$$B = \frac{\pi}{1-\pi}\frac{1}{1-\eta}$$

Estimation Equations

The equations to estimate the *observed* rate of readmission and home health assignment, respectively, are:

$$HHC = \beta_{10} + \beta_{11}Age + \beta_{12}Prior + \sum_{r=1}^{R}\rho_{1r}Race_r + \sum_{h=1}^{H}\theta_{1h}Ethnicity + \sum_{p=1}^{P}\pi_{1p}Paytype_p$$
$$+ \sum_{m=1}^{M}\delta_{1m}Diagnosis + \alpha_{1j} + \epsilon_{1i},$$

$$Risk = \beta_{20} + \beta_{21}Age + \beta_{22}Prior + \sum_{r=1}^{R}\rho_{2r}Race_r + \sum_{h=1}^{H}\theta_{2h}Ethnicity + \sum_{p=1}^{P}\pi_{2p}Paytype_p$$
$$+ \sum_{m=1}^{M}\delta_{2m}Diagnosis + \alpha_{2j} + \epsilon_{2i},$$

and

$$Readmission = \beta_{30} + \beta_{31}Age + \beta_{32}Prior + \beta_{33}HHC + \sum_{r=1}^{R}\rho_{3r}Race_r + \sum_{h=1}^{H}\theta_{3h}Ethnicity$$
$$+ \sum_{p=1}^{P}\pi_{3p}Paytype_p + \sum_{m=1}^{M}\delta_{3m}Diagnosis + \alpha_{3j} + \epsilon_{3i},$$

where α_j represents the fixed effect for hospital j.

Simulation Results.

We simulated the model 1000 times using 1000 observations randomly selected from the dataset described above in each iteration. We present the replicated results for the impact of length of stay on *likelihood* of readmission and the *likelihood* of readmission in Table 1.

We present the replicated results for the impact of length of stay on *risk* of readmission of readmission in Table 2.

Figure 1 presents a kernel density plot of the distributions of the two outcomes.

Figure 1: Distribution of Length of Stay Coefficient

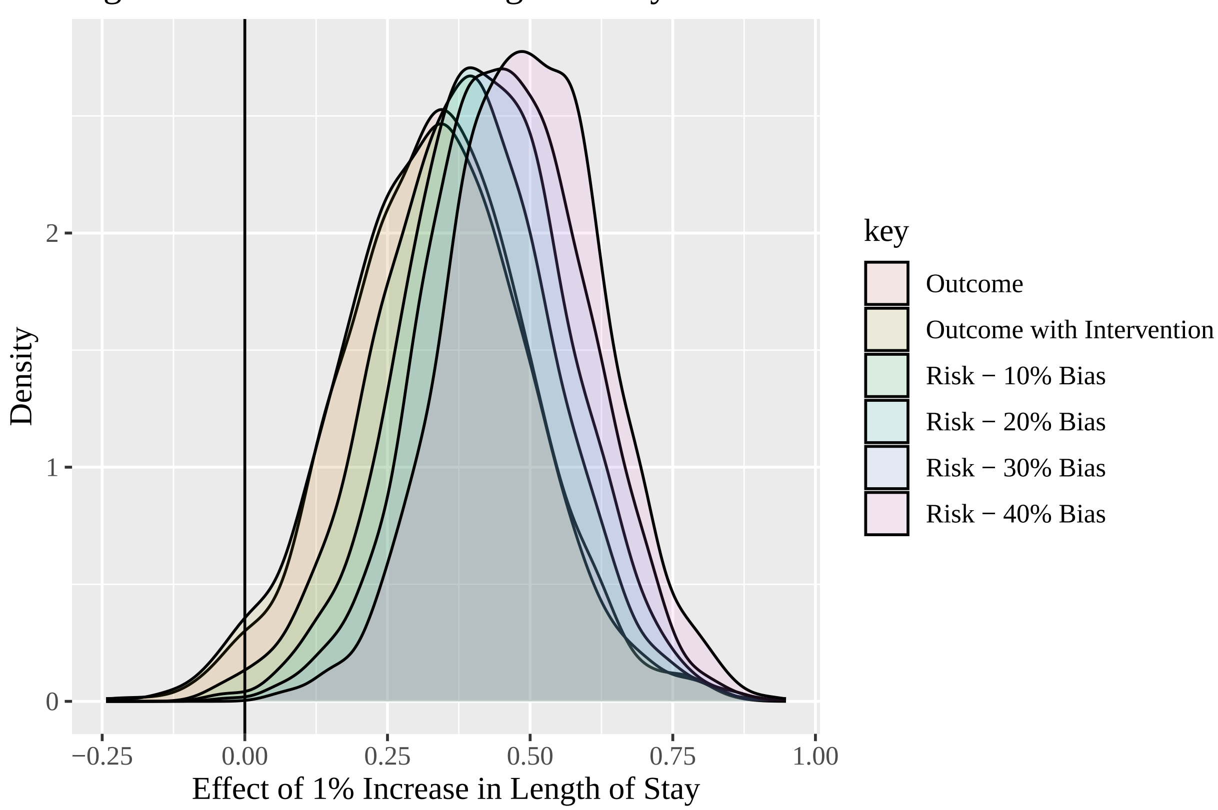

Table 1:

	Dependent variable:		
	hh	read30_01	
	(1)	(2)	(3)
medsugptSurgical	0.163	−0.530*	−0.534*
	(0.250)	(0.318)	(0.319)
icustayICU Stay	0.343	−0.243	−0.247
	(0.245)	(0.305)	(0.305)
genderFemale	0.227	−0.073	−0.076
	(0.224)	(0.241)	(0.242)
raceAsian	0.110	0.176	0.176
	(0.286)	(0.312)	(0.313)
raceBlack	−0.297	0.013	0.015
	(0.312)	(0.323)	(0.323)
ethnicityHispanic	−0.353	0.401	0.403
	(0.738)	(0.412)	(0.414)
maritalsMarried	−0.174	−0.094	−0.094
	(0.217)	(0.243)	(0.243)
mortriskModerate	0.352	0.292	0.291
	(0.317)	(0.342)	(0.343)
mortriskMajor	0.387	0.433	0.429
	(0.396)	(0.434)	(0.435)
severityModerate	0.372	0.240	0.238
	(0.360)	(0.366)	(0.366)
severityMajor	0.624	0.448	0.443
	(0.434)	(0.444)	(0.446)
age_cats45-54	0.486	0.119	0.116
	(0.456)	(0.394)	(0.395)
age_cats55-64	0.848**	0.020	0.015
	(0.417)	(0.369)	(0.371)
age_cats65-74	1.240***	0.026	0.017
	(0.408)	(0.388)	(0.390)
age_cats75+	1.767***	0.049	0.034
	(0.423)	(0.392)	(0.398)
lnlos	0.861***	0.337**	0.331**
	(0.151)	(0.159)	(0.163)
hhHHC			0.061
			(0.302)
Constant	−4.973***	−3.385***	−3.376***
	(0.872)	(0.684)	(0.689)
Observations	1,000	1,000	1,000
Log Likelihood	−335.269	−315.279	−315.020
Akaike Inf. Crit.	738.538	698.559	700.040
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 2:

	Dependent variable:			
	risk			
	(1)	(2)	(3)	(4)
medsugptSurgical	−0.443	−0.372	−0.316	−0.270
	(0.297)	(0.279)	(0.264)	(0.254)
icustayICU Stay	−0.184	−0.129	−0.086	−0.050
	(0.286)	(0.267)	(0.253)	(0.242)
genderFemale	−0.043	−0.017	0.002	0.020
	(0.230)	(0.217)	(0.208)	(0.201)
raceAsian	0.166	0.163	0.157	0.150
	(0.295)	(0.285)	(0.274)	(0.265)
raceBlack	−0.012	−0.036	−0.054	−0.074
	(0.304)	(0.294)	(0.287)	(0.281)
ethnicityHispanic	0.343	0.303	0.267	0.236
	(0.400)	(0.389)	(0.384)	(0.383)
maritalsMarried	−0.099	−0.108	−0.116	−0.121
	(0.230)	(0.217)	(0.212)	(0.204)
mortriskModerate	0.293	0.299	0.303	0.309
	(0.326)	(0.314)	(0.304)	(0.290)
mortriskMajor	0.412	0.407	0.404	0.400
	(0.417)	(0.400)	(0.385)	(0.370)
severityModerate	0.246	0.258	0.266	0.273
	(0.349)	(0.336)	(0.323)	(0.313)
severityMajor	0.456	0.466	0.477	0.487
	(0.422)	(0.411)	(0.399)	(0.383)
age_cats45-54	0.132	0.152	0.167	0.184
	(0.381)	(0.372)	(0.364)	(0.352)
age_cats55-64	0.075	0.128	0.169	0.215
	(0.351)	(0.341)	(0.331)	(0.321)
age_cats65-74	0.115	0.192	0.268	0.338
	(0.371)	(0.364)	(0.352)	(0.339)
age_cats75+	0.201	0.335	0.447	0.543
	(0.376)	(0.361)	(0.350)	(0.340)
lnlos	0.386**	0.423***	0.456***	0.494***
	(0.150)	(0.143)	(0.138)	(0.134)
Constant	−3.373***	−3.367***	−3.356***	−3.362***
	(0.650)	(0.624)	(0.593)	(0.573)
Observations	1,000	1,000	1,000	1,000
Log Likelihood	−329.351	−351.133	−381.002	−398.738
Akaike Inf. Crit.	726.703	770.267	830.005	865.476
Note:				*p<0.1; **p<0.05; ***p<0.01