

Identifying and Interpreting Subgroups in Health Care Utilization Data with Count Mixture Regression Models - Online Supplement

Christoph F. Kurz¹ and Laura A. Hatfield²

¹Institute of Health Economics and Health Care Management, Helmholtz Zentrum München

²Department of Health Care Policy, Harvard Medical School

Appendix

Component	N	high overlap				medium overlap				low overlap			
		β_0	β_1	β_2	ψ	β_0	β_1	β_2	ψ	β_0	β_1	β_2	ψ
1	3000	0.0	0.5	-0.5	1.00	0.0	0.5	-0.5	1.00	0.0	0.5	-0.5	1.00
2	5000	1.0	0.0	0.0	13.25	1.8	0.0	0.0	25.75	3.2	0.0	0.0	50.75
3	3000	1.7	-0.5	0.5	25.50	2.4	-0.5	0.5	50.50	3.9	-0.5	0.5	100.50
4	4000	2.0	0.0	0.0	37.75	2.7	0.0	0.0	75.25	4.3	0.0	0.0	150.25
5	2000	2.3	0.5	-0.5	50.00	3.0	0.5	-0.5	100.00	4.6	0.5	-0.5	200.00

Table 1: Data-generating parameters for the simulation study. N indicates the number of simulated observations from each component.

	Comp. 1					Comp. 2				
	β_0	β_1	β_2	ψ	c_1	β_0	β_1	β_2	ψ	c_2
est.	0.0	0.4	-0.5	1.2	0.59	3.0	0.5	-0.5	83.7	0.41
truth	0.0	0.5	-0.5	1.0	0.60	3.0	0.5	-0.5	100	0.40

Table 2: Comparison of true and estimated parameters for the case with 2 components and medium overlap, where the number of mixture components was accurately estimated by MD-NB. β 's indicate intercept and regression coefficients and c 's are the mixture weights. All estimates are based on posterior medians. Because the estimated mixture components are not always in the same order, estimated components were hand-matched to their corresponding best-fitting component.

	Comp. 1					Comp. 2					Comp. 3					Comp. 4				
	β_0	β_1	β_2	ψ	c_1	β_0	β_1	β_2	ψ	c_2	β_0	β_1	β_2	ψ	c_3	β_0	β_1	β_2	ψ	c_4
est.	-0.1	0.6	-1.0	3.2	0.18	1.0	0.0	0.0	8.9	0.45	2.0	0.0	0.0	11.1	0.21	2.3	0.5	-0.5	61.9	0.16
truth	0.0	0.5	-0.5	1.0	0.21	1.0	0.0	0.0	13.25	0.36	2.0	0.0	0.0	37.75	0.29	2.3	0.5	-0.5	50.0	0.14

Table 3: Comparison of true and estimated parameters for the case with 4 components and high overlap, where the number of mixture components was accurately estimated by MD-NB. β 's indicate intercept and regression coefficients and c 's are the mixture weights. All estimates are based on posterior medians. Because the estimated mixture components are not always in the same order, estimated components were hand-matched to their corresponding best-fitting component.

	Comp. 1					Comp. 2					Comp. 3					Comp. 4					Comp. 5				
	β_0	β_1	β_2	ψ	c_1	β_0	β_1	β_2	ψ	c_2	β_0	β_1	β_2	ψ	c_3	β_0	β_1	β_2	ψ	c_4	β_0	β_1	β_2	ψ	c_5
est.	0.0	0.6	-1.0	1.5	0.12	1.3	0.0	0.1	6.8	0.34	1.6	-0.6	0.5	18.1	0.26	2.1	-0.2	0.1	39.3	0.17	2.3	0.5	-0.5	72.5	0.12
truth	0.0	0.5	-0.5	1.0	0.18	1.0	0.0	0.0	13.25	0.29	1.7	-0.5	0.5	25.5	0.18	2.0	0.0	0.0	37.75	0.24	2.3	0.5	-0.5	50.0	0.12

Table 4: Comparison of true and estimated parameters for the case with 5 components and high overlap, where the number of mixture components was accurately estimated by MD-NB. β 's indicate intercept and regression coefficients and c 's are the mixture weights. All estimates are based on posterior medians. Because the estimated mixture components are not always in the same order, estimated components were hand-matched to their corresponding best-fitting component.

	Comp. 1					Comp. 2					Comp. 3					Comp. 4					Comp. 5				
	β_0	β_1	β_2	ψ	c_1	β_0	β_1	β_2	ψ	c_2	β_0	β_1	β_2	ψ	c_3	β_0	β_1	β_2	ψ	c_4	β_0	β_1	β_2	ψ	c_5
est.	-0.1	0.4	-0.7	1.5	0.10	1.8	0.0	0.0	6.9	0.34	2.4	-0.5	0.5	22.5	0.27	2.8	0.0	0.0	42.3	0.17	3.0	0.5	-0.5	66.2	0.13
truth	0.0	0.5	-0.5	1.0	0.18	1.0	0.0	0.0	13.25	0.29	1.7	-0.5	0.5	25.5	0.18	2.0	0.0	0.0	37.75	0.24	2.3	0.5	-0.5	50.0	0.12

Table 5: Comparison of true and estimated parameters for the case with 5 components and medium overlap, where the number of mixture components was accurately estimated by MD-NB. β 's indicate intercept and regression coefficients and c 's are the mixture weights. All estimates are based on posterior medians. Because the estimated mixture components are not always in the same order, estimated components were hand-matched to their corresponding best-fitting component.

	Comp. 1					Comp. 2					Comp. 3					Comp. 4					Comp. 5				
	β_0	β_1	β_2	ψ	c_1	β_0	β_1	β_2	ψ	c_2	β_0	β_1	β_2	ψ	c_3	β_0	β_1	β_2	ψ	c_4	β_0	β_1	β_2	ψ	c_5
est.	0.1	0.5	-0.6	1.5	0.21	3.2	0.0	0.0	3.6	0.20	3.9	-0.5	0.5	43.3	0.23	4.3	-0.1	0.0	81.0	0.22	4.3	0.1	-0.3	208.3	0.14
truth	0.0	0.5	-0.5	1.0	0.18	3.2	0.0	0.0	50.75	0.29	3.9	-0.5	0.5	100.5	0.18	4.3	0.0	0.0	150.25	0.24	4.6	0.5	-0.5	200.0	0.12

Table 6: Comparison of true and estimated parameters for the case with 5 components and low overlap, where the number of mixture components was accurately estimated by MD-NB. β 's indicate intercept and regression coefficients and c 's are the mixture weights. All estimates are based on posterior medians. Because the estimated mixture components are not always in the same order, estimated components were hand-matched to their corresponding best-fitting component.

	Comp. 1					Comp. 2				
	β_0	β_1	β_2	ψ	c_1	β_0	β_1	β_2	ψ	c_2
est.	0.0	0.4	0.5	1.0	0.60	2.3	-0.7	-0.5	60.8	0.40
truth	0.0	0.5	-0.5	1.0	0.60	3.0	0.5	-0.5	100	0.40

Table 7: Comparison of true and estimated parameters for the case with 2 components and high overlap, where the number of mixture components was accurately estimated by LOG-NB. β 's indicate intercept and regression coefficients and c 's are the mixture weights. All estimates are based on posterior medians. Because the estimated mixture components are not always in the same order, estimated components were hand-matched to their corresponding best-fitting component.

	Comp. 1					Comp. 2				
	β_0	β_1	β_2	ψ	c_1	β_0	β_1	β_2	ψ	c_2
est.	0.0	0.5	0.5	1.0	0.60	4.6	-0.6	-0.5	221	0.40
truth	0.0	0.5	-0.5	1.0	0.60	4.6	0.5	-0.5	100	0.40

Table 8: Comparison of true and estimated parameters for the case with 2 components and low overlap, where the number of mixture components was accurately estimated by LOG-NB. β 's indicate intercept and regression coefficients and c 's are the mixture weights. All estimates are based on posterior medians. Because the estimated mixture components are not always in the same order, estimated components were hand-matched to their corresponding best-fitting component.

	Comp. 1					Comp. 2					Comp. 3				
	β_0	β_1	β_2	ψ	c_1	β_0	β_1	β_2	ψ	c_2	β_0	β_1	β_2	ψ	c_3
est.	0.3	0.2	-0.4	22.4	0.37	0.9	0.3	-0.3	5.7	0.46	2.3	0.5	-0.5	159	0.17
truth	0.0	0.5	-0.5	1.0	0.3	1.0	0.0	0.0	13.25	0.5	2.3	0.5	-0.5	50	0.2

Table 9: Comparison of true and estimated parameters for the case with 3 components and high overlap, where the number of mixture components was accurately estimated by LOG-NB. β 's indicate intercept and regression coefficients and c 's are the mixture weights. All estimates are based on posterior medians. Because the estimated mixture components are not always in the same order, estimated components were hand-matched to their corresponding best-fitting component.

3 Comp.	Comp. 1					Comp. 2					Comp. 3				
	β_0	β_1	β_2	ψ	c_1	β_0	β_1	β_2	ψ	c_2	β_0	β_1	β_2	ψ	c_3
est.	0.7	0.2	-0.3	20.1	0.35	1.8	0.1	-0.2	35	0.47	3.0	0.5	-0.5	203	0.2
truth	0.0	0.5	-0.5	1.0	0.3	1.8	0.0	0.0	25.75	0.5	3.0	0.5	-0.5	100	0.2

Table 10: Comparison of true and estimated parameters for the case with 3 components and medium overlap, where the number of mixture components was accurately estimated by LOG-NB. β 's indicate intercept and regression coefficients and c 's are the mixture weights. All estimates are based on posterior medians. Because the estimated mixture components are not always in the same order, estimated components were hand-matched to their corresponding best-fitting component.

	Comp. 1					Comp. 2					Comp. 3					Comp. 4					Comp. 5				
	β_0	β_1	β_2	ψ	c_1	β_0	β_1	β_2	ψ	c_2	β_0	β_1	β_2	ψ	c_3	β_0	β_1	β_2	ψ	c_4	β_0	β_1	β_2	ψ	c_5
est.	0.1	0.5	-0.6	3.4	0.21	3.2	0.0	0.0	28.9	0.23	3.7	-0.5	0.6	61.3	0.22	4.3	0.0	0.0	102.0	0.22	4.7	0.5	-0.5	232.4	0.12
truth	0.0	0.5	-0.5	1.0	0.18	3.2	0.0	0.0	50.75	0.29	3.9	-0.5	0.5	100.5	0.18	4.3	0.0	0.0	150.25	0.24	4.6	0.5	-0.5	200.0	0.12

Table 11: Comparison of true and estimated parameters for the case with 5 components and low overlap, where the number of mixture components was accurately estimated by LOG-NB. β 's indicate intercept and regression coefficients and c 's are the mixture weights. All estimates are based on posterior medians. Because the estimated mixture components are not always in the same order, estimated components were hand-matched to their corresponding best-fitting component.

	Comp. 1				Comp. 2				Comp. 3				Comp. 4			
	β_0	β_1	β_2	c_1	β_0	β_1	β_2	c_2	β_0	β_1	β_2	c_3	β_0	β_1	β_2	c_4
est.	2.3	0.2	-0.2	0.26	2.3	0.2	-0.2	0.38	2.3	0.1	-0.2	0.28	2.3	0.2	-0.2	0.08
truth	0.0	0.5	-0.5	0.21	1.0	0.0	0.0	0.36	2.0	0.0	0.0	0.29	2.3	0.5	-0.5	0.14

Table 12: Comparison of true and estimated parameters for the case with 4 components and medium overlap, where the number of mixture components was accurately estimated by MLE-NB. β 's indicate intercept and regression coefficients and c 's are the mixture weights. Because the estimated mixture components are not always in the same order, estimated components were hand-matched to their corresponding best-fitting component.

	Comp. 1				Comp. 2				Comp. 3				Comp. 4			
	β_0	β_1	β_2	c_1	β_0	β_1	β_2	c_2	β_0	β_1	β_2	c_3	β_0	β_1	β_2	c_4
est.	-0.3	0.5	-0.6	0.20	3.9	0.2	-0.2	0.77	4.4	0.2	-0.2	0.02	4.5	0.5	-0.5	0.01
truth	0.0	0.5	-0.5	0.21	1.0	0.0	0.0	0.36	2.0	0.0	0.0	0.29	2.3	0.5	-0.5	0.14

Table 13: Comparison of true and estimated parameters for the case with 4 components and low overlap, where the number of mixture components was accurately estimated by MLE-NB. β 's indicate intercept and regression coefficients and c 's are the mixture weights. Because the estimated mixture components are not always in the same order, estimated components were hand-matched to their corresponding best-fitting component.

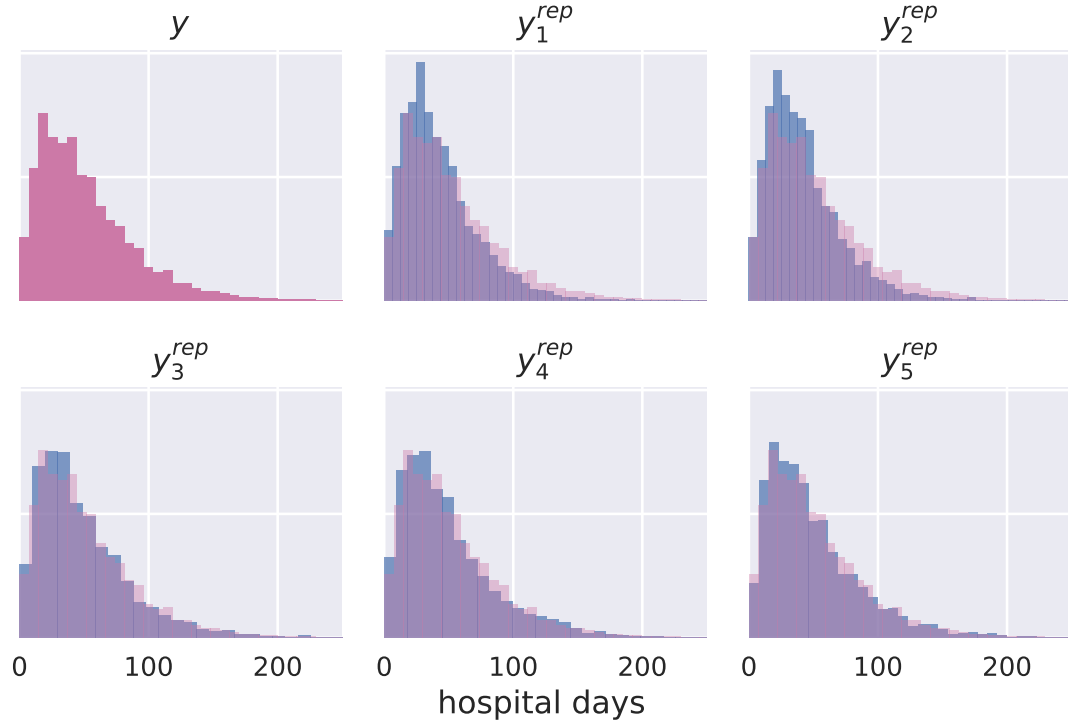


Figure 1: Density histograms for replicated outcome y^{rep} from simulating data from the posterior predictive distribution using the observed predictors. True outcome y from the AOK data set for comparison.

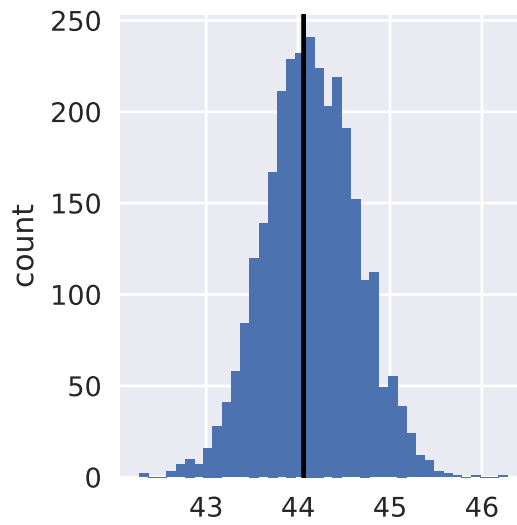


Figure 2: Histogram of means based on 1000 replicated data sets from the posterior predictive distribution. The black line marks the observed mean of hospital days from the AOK data set.

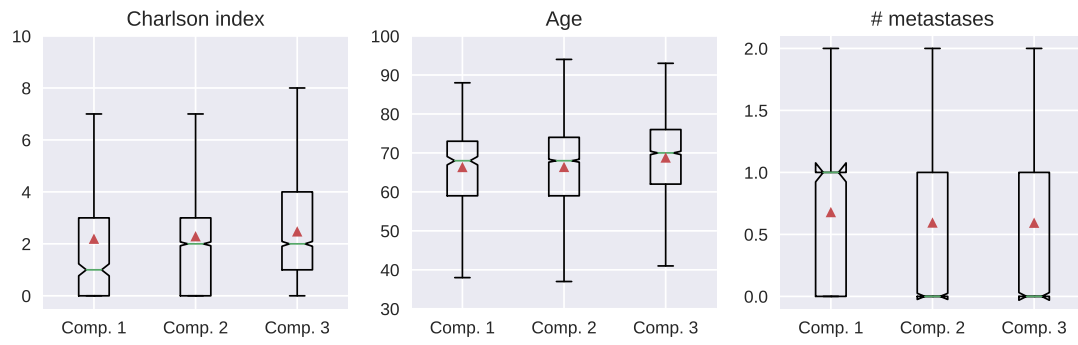


Figure 3: Boxplots for Charlson comorbidity index, age, and the number of metastases after using hard component assignments based on the MD-ZINB for the AOK data set. The red triangle marks the mean.