

Assessing mortality as an outcome in the MIMIC-III Database

BMIN503 – Final Project Presentation – December 5th, 2019

Presented by Silvia P. Canelón, PhD

Postdoctoral Researcher
Penn Dept. of Biostatistics, Epidemiology & Informatics

Background

- Secondary analysis of **Electronic Health Records** (EHR) can provide additional information about patients that are often under-represented in clinical studies, such as pregnant patients
- The U.S. has the worst rate of maternal deaths among developed nations, highlighting a critical area requiring additional research
- Pregnant patients could die or suffer other adverse outcomes during their care

Objectives

- Utilize a database containing medical records from intensive care units of a large medical center
 - Describe characteristics of the patient population
 - Study mortality as an outcome with a variety of patient care predictors
- Assess suitability of the MIMIC-III database to study **pregnancy-related outcomes** (i.e. cesarean section), particularly as relating to **mortality**

Data Source: MIMIC-III Critical Care Database

- Medical Information Mart for Intensive Care III
- Freely-available database with **de-identified** health-related data
- **Beth Israel Deaconess Medical Center** in Boston, Massachusetts
- **40,000+ patients** admitted to intensive care 2001-2012
- Includes demographics, vital sign measurements, laboratory test results, procedures, medications, caregiver notes, imaging reports, and mortality (both in and out of hospital) **across 26 tables**

Approach

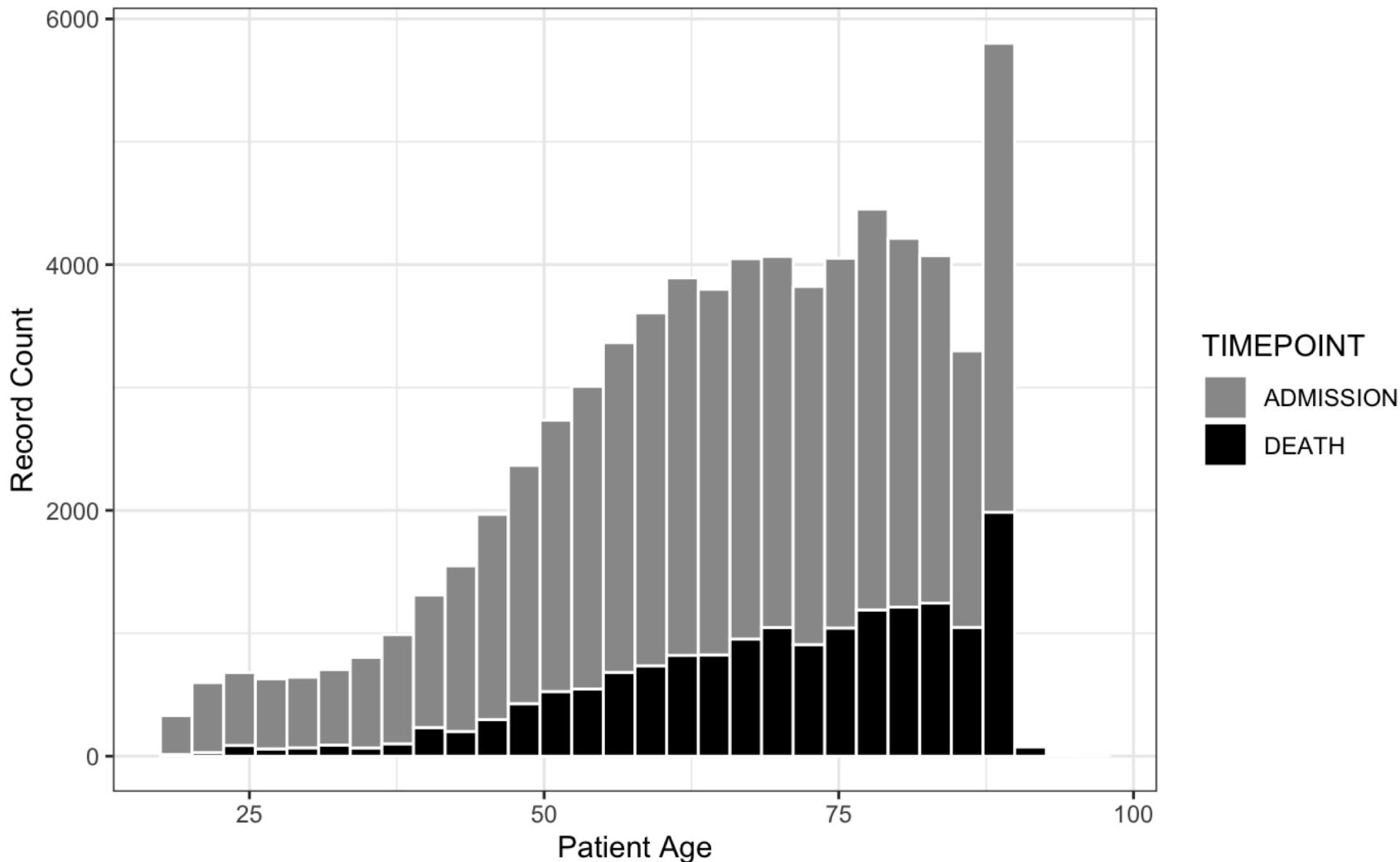
- 1.** Tidy the data!
 - 1.** Formatting/recoding variables
 - 2.** Joining data from multiple tables: admission records + patient information + diagnosis/procedure records + ICD code definitions
 - 3.** Identifying delivery and c-section records using ICD9 billing codes
- 2.** Describing the patient population
- 3.** Using logistic regression to study mortality as an outcome

Results: General record information

- Admission records: 58976
 - Adult records: 54362
- Patients: 46520
 - Adult patients: 38552
- Diagnosis records: 651047
- Procedure records: 240095

Results: Patient demographics – Age

- Average age at time of **admission**:
 64.02 ± 16.99 years
- Average age at time of **death**:
 70.13 ± 14.87 years



Results: Demographics – Sex/Gender, ethnicity, and insurance type

GENDER

GENDER	COUNT	PROP
F	23748	43.68
M	30614	56.32

ETHNICITY

ETHNICITY_BROAD	COUNT	PROP
AIAN	28	0.05
ASIAN	1263	2.32
BLACK	5173	9.52
CARIBBEAN ISLAND	9	0.02
HISPANIC	1838	3.38
MIDDLE EASTERN	41	0.08
MULTI RACE ETHNICITY	110	0.20
NHPI	14	0.03
OTHER	1198	2.20
PATIENT DECLINED TO ANSWER	467	0.86
PORTUGUESE	67	0.12
SOUTH AMERICAN	8	0.01
UNABLE TO OBTAIN	843	1.55
UNKNOWN/NOT SPECIFIED	4327	7.96
WHITE	38976	71.70

INSURANCE

INSURANCE	COUNT	PROP
GOV	1461	2.69
MEDICAID	4801	8.83
MEDICARE	30346	55.82
PRIVATE	17168	31.58
SELF PAY	586	1.08

Results: Outcomes – Delivery, c-section, mortality

DELIVERY

DELIVERY	COUNT	PROP
DELIVERY	53	0.1
NO DELIVERY	54309	99.9

C-SECTION

MORTALITY	COUNT	PROP
LIVED	47788	87.91
DIED	6574	12.09

MORTALITY

CSECTION	COUNT	PROP
CSECTION	41	0.08
NO CSECTION	54321	99.92

ADMISSION_TYPE

ADMISSION_TYPE	COUNT	PROP
ELECTIVE	8088	14.88
EMERGENCY	44853	82.51
URGENT	1421	2.61

Results: Outcomes – Delivery, c-section, mortality

DELIVERY

DELIVERY	COUNT	PROP
DELIVERY	53	0.1
NO DELIVERY	54309	99.9

C-SECTION

MORTALITY	COUNT	PROP
LIVED	47788	87.91
DIED	6574	12.09

MORTALITY

CSECTION	COUNT	PROP
CSECTION	41	0.08
NO CSECTION	54321	99.92

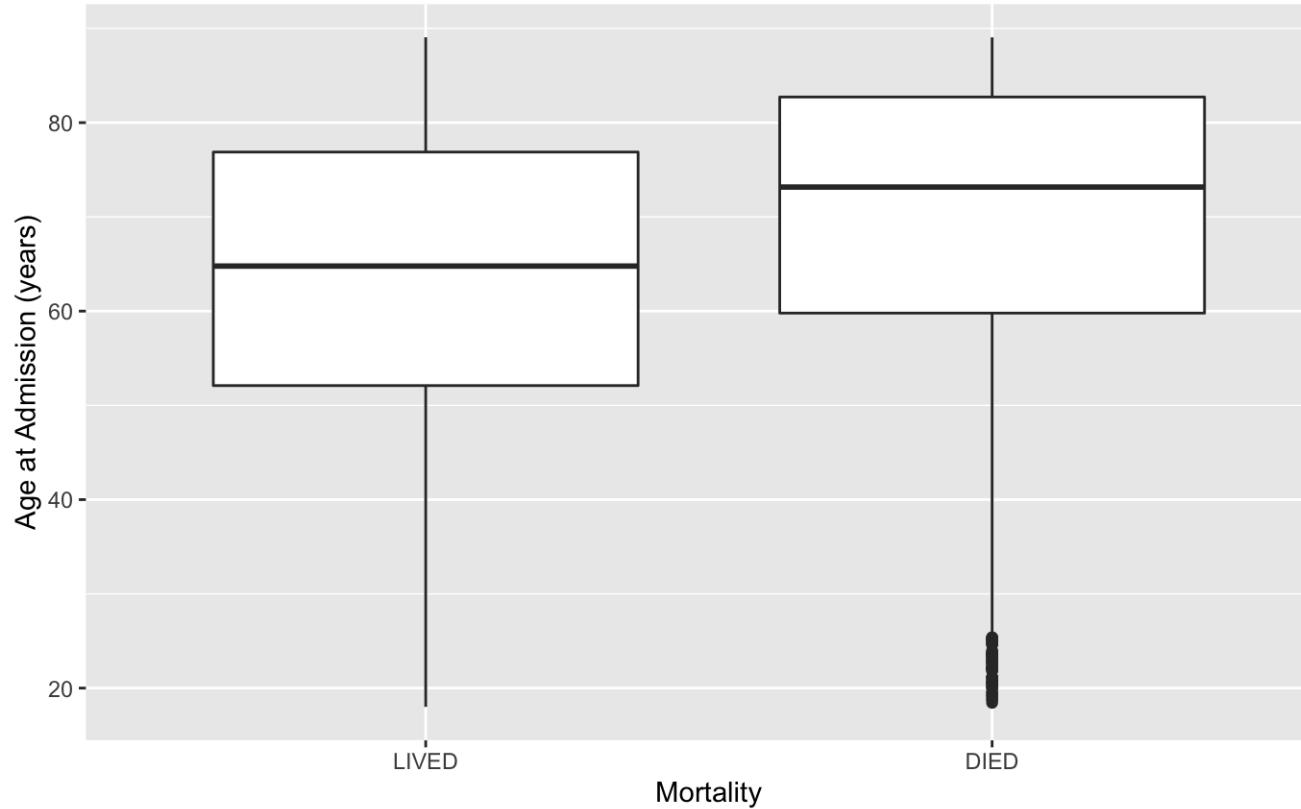
ADMISSION_TYPE

ADMISSION_TYPE	COUNT	PROP
ELECTIVE	8088	14.88
EMERGENCY	44853	82.51
URGENT	1421	2.61

AGE_ADM was positively associated with MORTALITY with p-value <2e-16

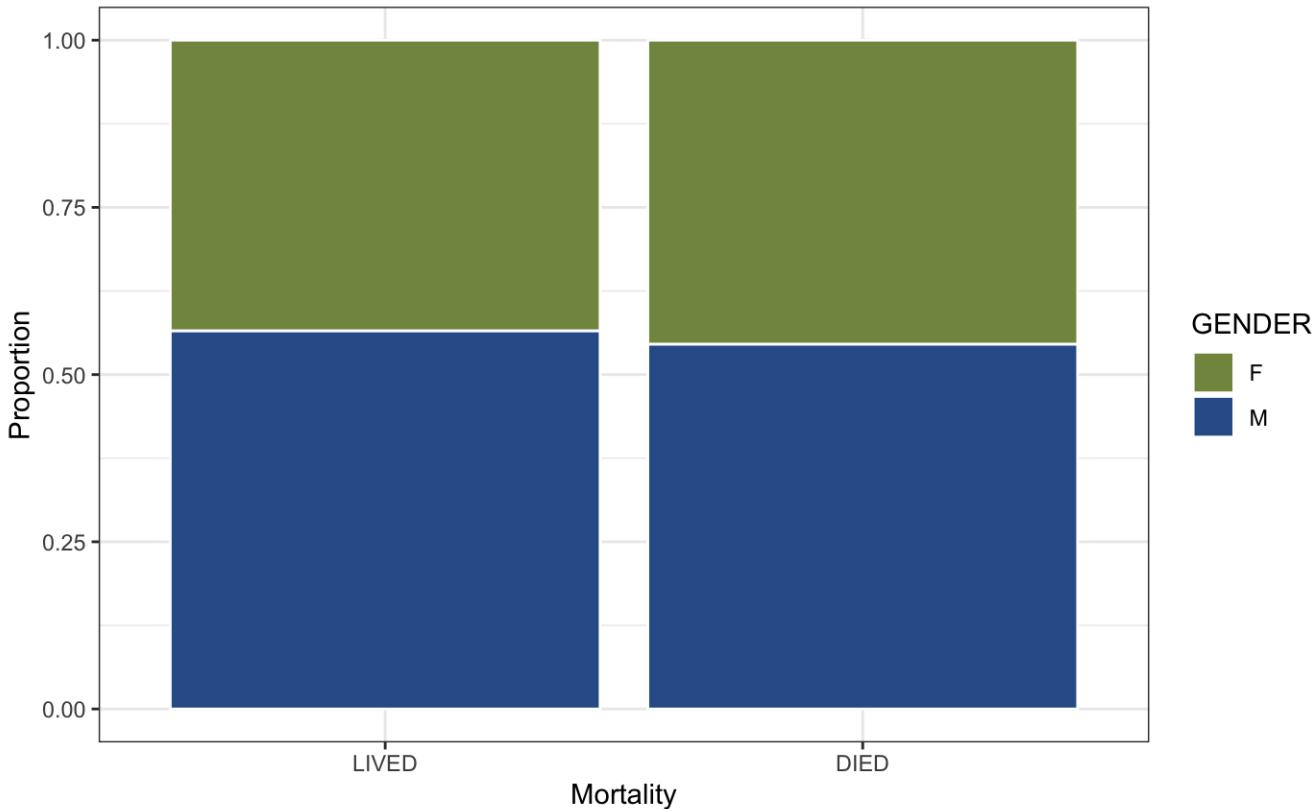
Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-3.7335953	0.0623189	-59.91	<2e-16 ***	
AGE_ADM	0.0262223	0.0008791	29.83	<2e-16 ***	



GENDER MALE was negatively associated with MORTALITY
with p-value 0.00226

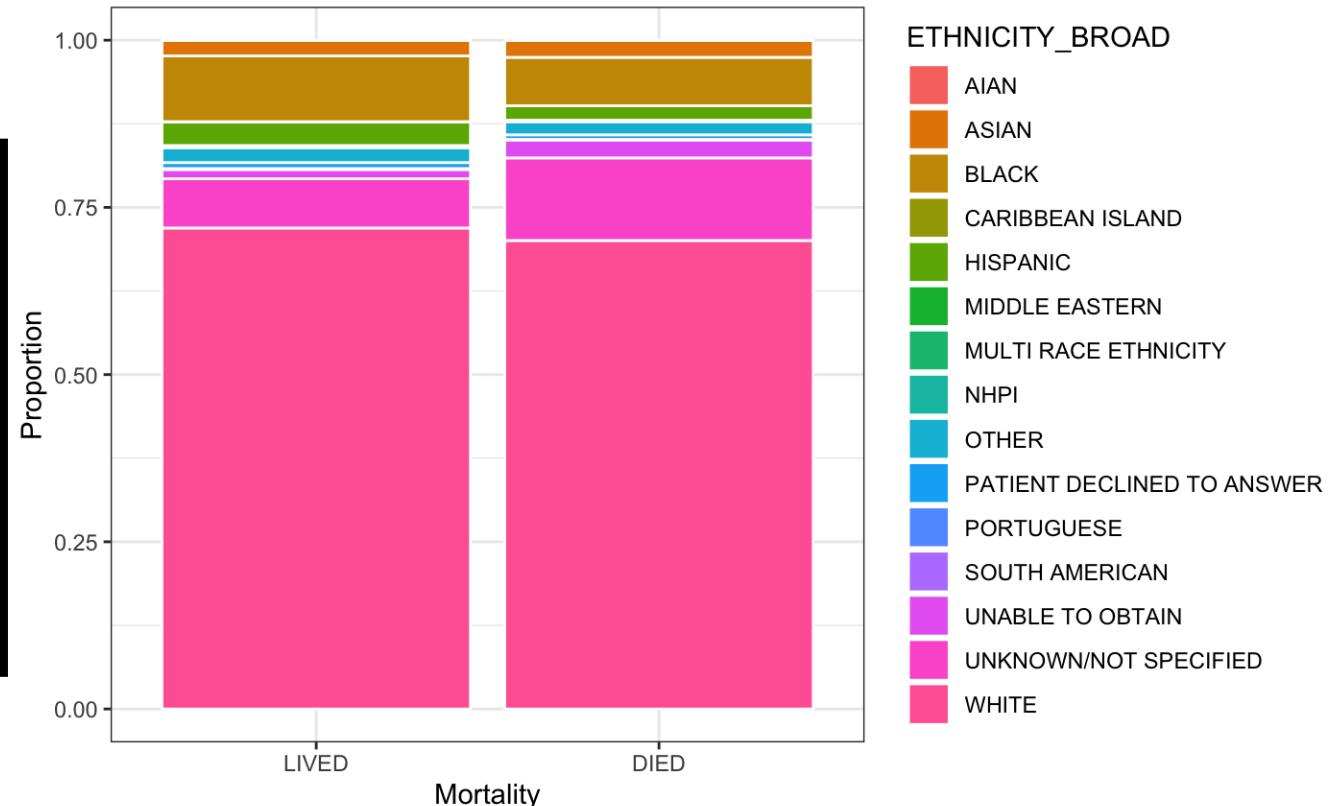
Coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.93881	0.01957	-99.075	< 2e-16	***
GENDERM	-0.08071	0.02643	-3.053	0.00226	**



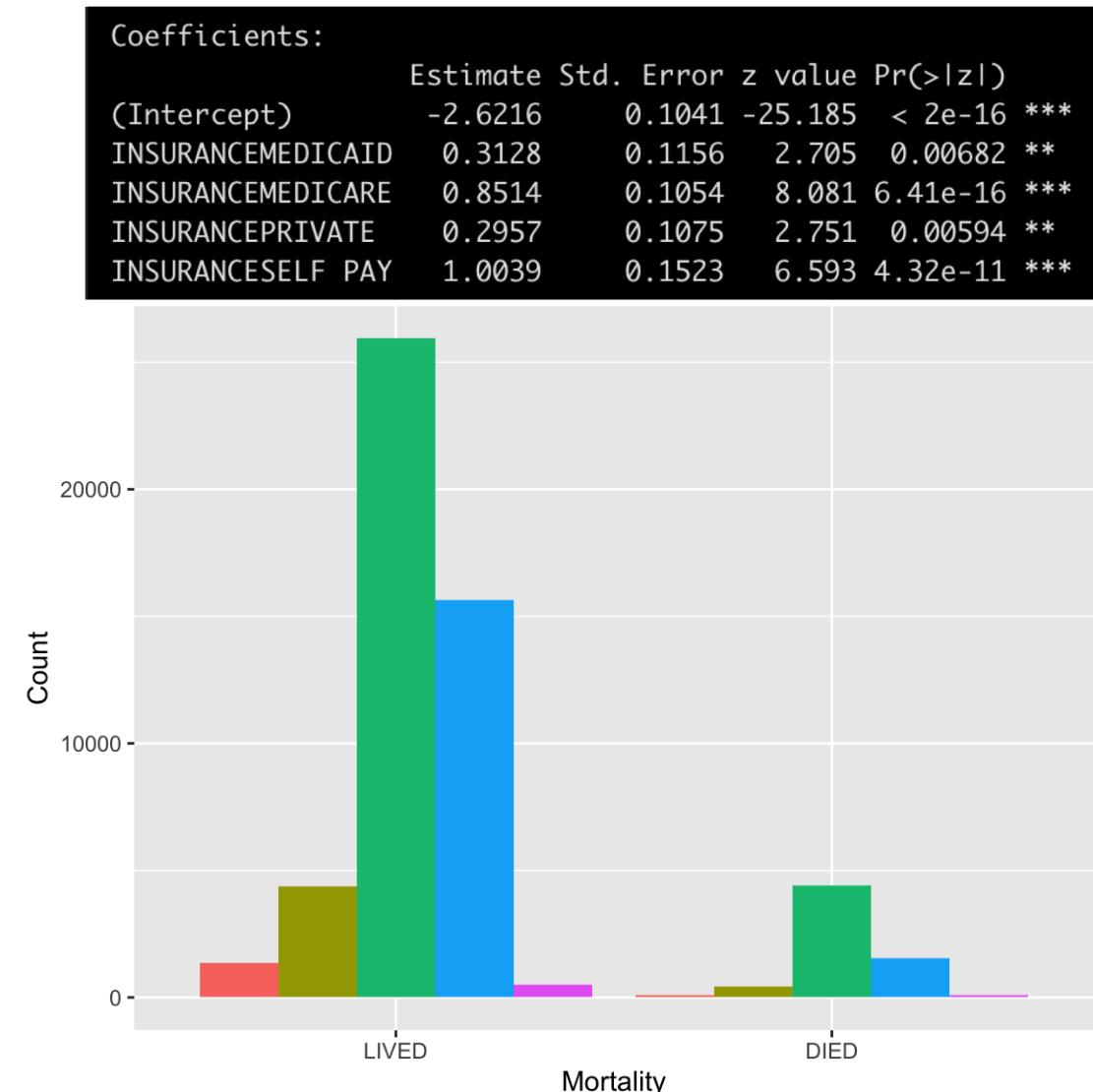
No significant association between any particular ETHNICITY and MORTALITY

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.120264	0.611010	-3.470	0.00052	***
ETHNICITY_BROADASIAN	0.224963	0.616688	0.365	0.71527	
ETHNICITY_BROADBLACK	-0.168998	0.612901	-0.276	0.78275	
ETHNICITY_BROADCARIBBEAN ISLAND	-10.445799	108.249623	-0.096	0.92313	
ETHNICITY_BROADHISPANIC	-0.359937	0.617224	-0.583	0.55979	
ETHNICITY_BROADMIDDLE EASTERN	-0.850151	0.948143	-0.897	0.36991	
ETHNICITY_BROADMULTI RACE ETHNICITY	-0.297632	0.703097	-0.423	0.67206	
ETHNICITY_BROADNHPI	1.203973	0.850490	1.416	0.15689	
ETHNICITY_BROADOTHER	0.005597	0.618078	0.009	0.99277	
ETHNICITY_BROADPATIENT DECLINED TO ANSWER	-0.118079	0.630813	-0.187	0.85151	
ETHNICITY_BROADPORTUGUESE	-0.198851	0.745918	-0.267	0.78979	
ETHNICITY_BROADSOUTH AMERICAN	-10.445799	114.815861	-0.091	0.92751	
ETHNICITY_BROADUNABLE TO OBTAIN	0.787957	0.616858	1.277	0.20147	
ETHNICITY_BROADUNKNOWN/NOT SPECIFIED	0.654969	0.612249	1.070	0.28472	
ETHNICITY_BROADWHITE	0.110193	0.611212	0.180	0.85693	



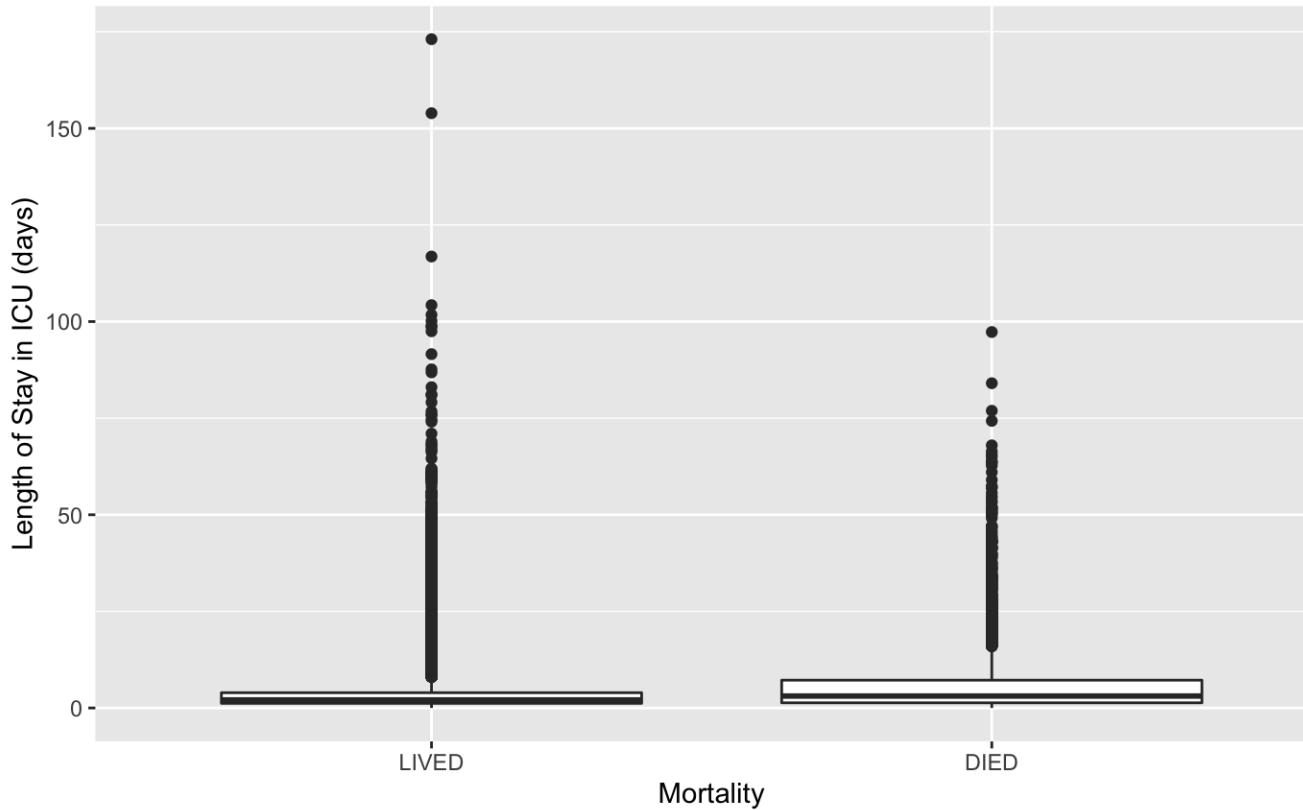
Significant association between all INSURANCE types and MORTALITY (i.e. having any insurance)



LOS_ICU was positively associated with MORTALITY with p-value <2e-16

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.15138	0.01594	-134.97	<2e-16 ***	
LOS_ICU	0.03879	0.00169	22.95	<2e-16 ***	



MORTALITY outcome with DELIVERY and CSECTION as predictors

- Non-significant positive association between

NO DELIVERY and MORTALITY

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-13.57	73.54	-0.184	0.854
DELIVERYNO DELIVERY	11.58	73.54	0.158	0.875

- Non-significant positive association between

NO CSECTION and MORTALITY

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-13.57	83.62	-0.162	0.871
CSECTIONNO CSECTION	11.58	83.62	0.139	0.890

Logistic regression including all categorical variables

- EMERGENCY and URGENT admissions were significantly associated with MORTALITY
- All INSURANCE types significantly associated with MORTALITY

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-15.43867	83.21577	-0.186	0.85282
ADMISSION_TYPEEMERGENCY	1.57660	0.06539	24.111	< 2e-16 ***
ADMISSION_TYPEURGENT	1.35167	0.10350	13.059	< 2e-16 ***
INSURANCEMEDICAID	0.31243	0.11639	2.684	0.00727 **
INSURANCEMEDICARE	0.83403	0.10651	7.831	4.86e-15 ***
INSURANCEPRIVATE	0.32421	0.10862	2.985	0.00284 **
INSURANCESELF PAY	0.84436	0.15376	5.491	3.99e-08 ***
GENDERM	-0.04235	0.02703	-1.567	0.11716
ETHNICITY_BROADASIAN	0.28434	0.62115	0.458	0.64712
ETHNICITY_BROADBLACK	-0.25079	0.61725	-0.406	0.68453
ETHNICITY_BROADCARIBBEAN ISLAND	-11.23008	172.85007	-0.065	0.94820
ETHNICITY_BROADHISPANIC	-0.31925	0.62160	-0.514	0.60754
ETHNICITY_BROADMIDDLE EASTERN	-0.68997	0.95312	-0.724	0.46912
ETHNICITY_BROADMULTI RACE ETHNICITY	-0.26377	0.70813	-0.372	0.70952
ETHNICITY_BROADNPI	1.23922	0.86300	1.436	0.15102
ETHNICITY_BROADOTHER	0.08627	0.62258	0.139	0.88980
ETHNICITY_BROADPATIENT DECLINED TO ANSWER	0.13867	0.63585	0.218	0.82737
ETHNICITY_BROADPORTUGUESE	-0.18083	0.75150	-0.241	0.80984
ETHNICITY_BROADSOUTH AMERICAN	-11.33813	187.59154	-0.060	0.95180
ETHNICITY_BROADUNABLE TO OBTAIN	0.75903	0.62143	1.221	0.22193
ETHNICITY_BROADUNKNOWN/NOT SPECIFIED	0.64843	0.61678	1.051	0.29311
ETHNICITY_BROADWHITE	0.09827	0.61568	0.160	0.87319
DELIVERYNO DELIVERY	11.21691	153.84693	0.073	0.94188
CSECTIONNO CSECTION	0.07140	174.90953	0.000	0.99967

Limitations

- Dates in the database have been shifted to protect patient confidentiality
- Date shifting removed:
 - Year - The year is randomly distributed between 2100 – 2200
 - Day of the month - The absolute day of the month is not preserved
 - Inter-patient information - Two patients in the ICU on 2150-01-01 were not in the ICU at the same time
- Not all `ICD9_CODES` in the database match to a ICD code
- Population is mostly older which means fewer pregnancies, so pregnancy-related outcomes are not easily studied

Acknowledgements

- **Mary Regina Boland**; Assistant Professor of Informatics;
Department of Biostatistics, Epidemiology, and Informatics
- **Ray Bai**; Postdoctoral Researcher; Department of Biostatistics,
Epidemiology, and Informatics
- **Lena Davidson**; Data Analyst; Department of Biostatistics,
Epidemiology, and Informatics