Characterizing the patients, hospitals, and data quality of the eICU Collaborative Research Database - Supplementary digital content

Supplemental Table 1: Tables from the eICU-CRD used in the analysis.

eICU-CRD Table	Features investigated
admissiondrug	Medications used prior to admission
admissiondx	Admitting diagnosis per APACHE scoring
allergy	Allergies documented during admission
apacheApsVar	Variables for the Acute Physiology Score component of APACHE
apachePatientResult	APACHE scores and predictions, and actual outcomes
carePlanGeneral	Documentation related to care planning updated over encounter
diagnosis	Diagnoses documented by clinical staff, along with ICD codes
hospital	Data regarding characteristics of participating hospitals
InfusionDrug	Details of drug infusions from nursing flowsheet
nurseCharting	Semi-structured data entered by bedside nurse
patient	Patient demographics, admission, and discharge details
physicalExam	Physical exam data (mostly structured) entered into HER
respiratoryCare	Data related to airway and ventilation
treatment	Structured data regarding treatments with time and date
vitalAperiodic	Vital signs data recorded at irregular intervals
vitalPeriodic	Vital signs data recorded at regular intervals

Supplemental Table 2: Ranges for plausible vital signs readings. These were determined based on logic and clinical reasoning.

Vital Sign	Realistic Range
Heart Rate	0-300 bpm inclusive
Respiratory Rate	0-100 inclusive
Oxygen Saturation	50-100% inclusive
Temperature	15-45 C (59-113 F) inclusive
Central Venous Pressure	0-50 cm H_2O inclusive
Blood Pressure	Systolic 0-300 mmHg; Diastolic 0-200 mmHg; Mean 0-190; Systolic > mean and diastolic; Diastolic < mean; Systolic > 4mmHg greater than diastolic

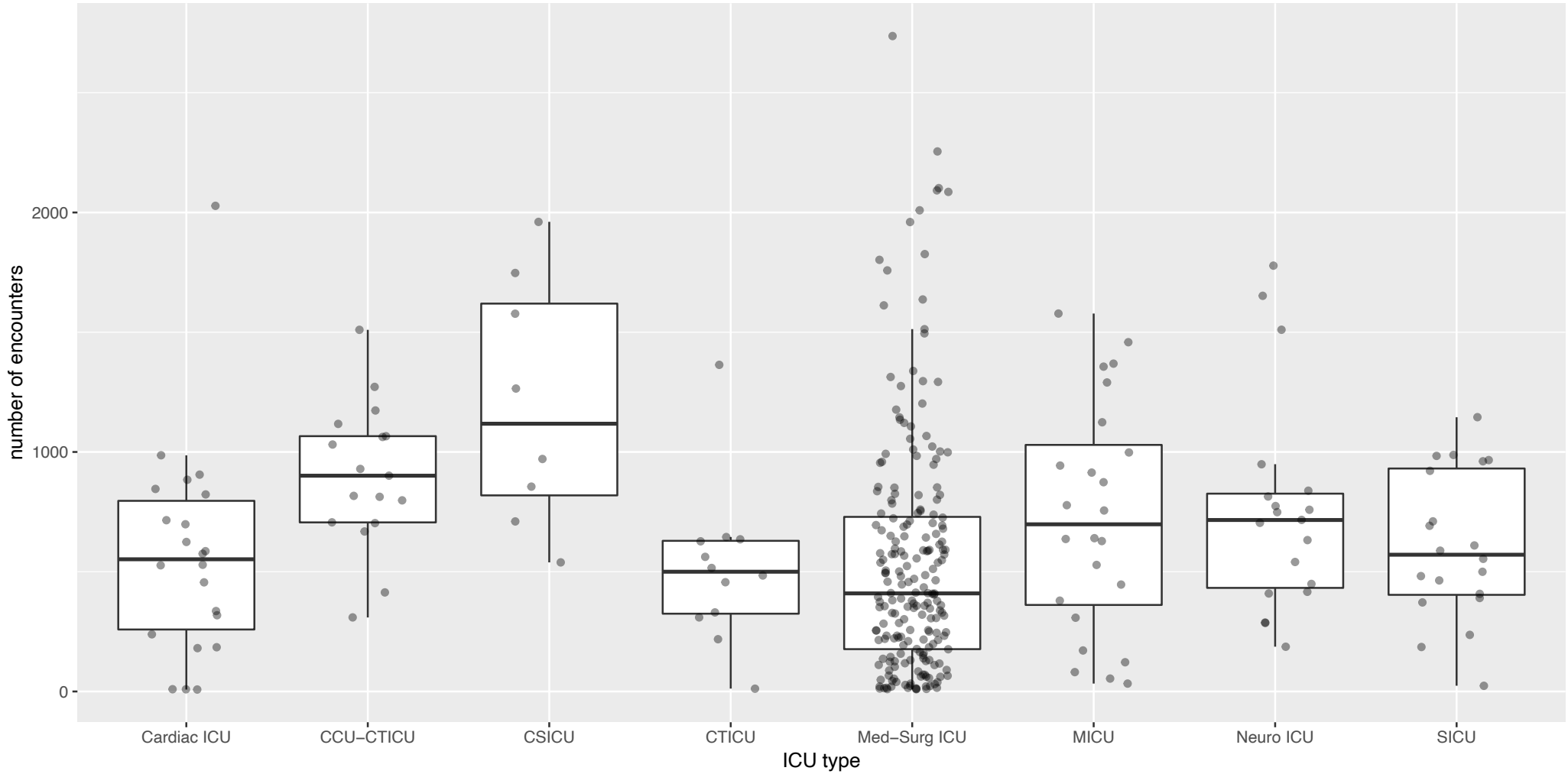
Supplemental Table 3: Hospitals represented in the eICU-CRD

	# Hospitals	Teaching	Non-teaching
Region			
Midwest	70	8	62
South	56	7	49
West	43	1	42
Northeast	13	2	11
Unknown	26	1	25
Size			
< 100 beds	46	0	46
100-249 beds	62	2	60
250-499 beds	35	3	32
≥ 500 beds	23	12	11
Unknown	42	2	40

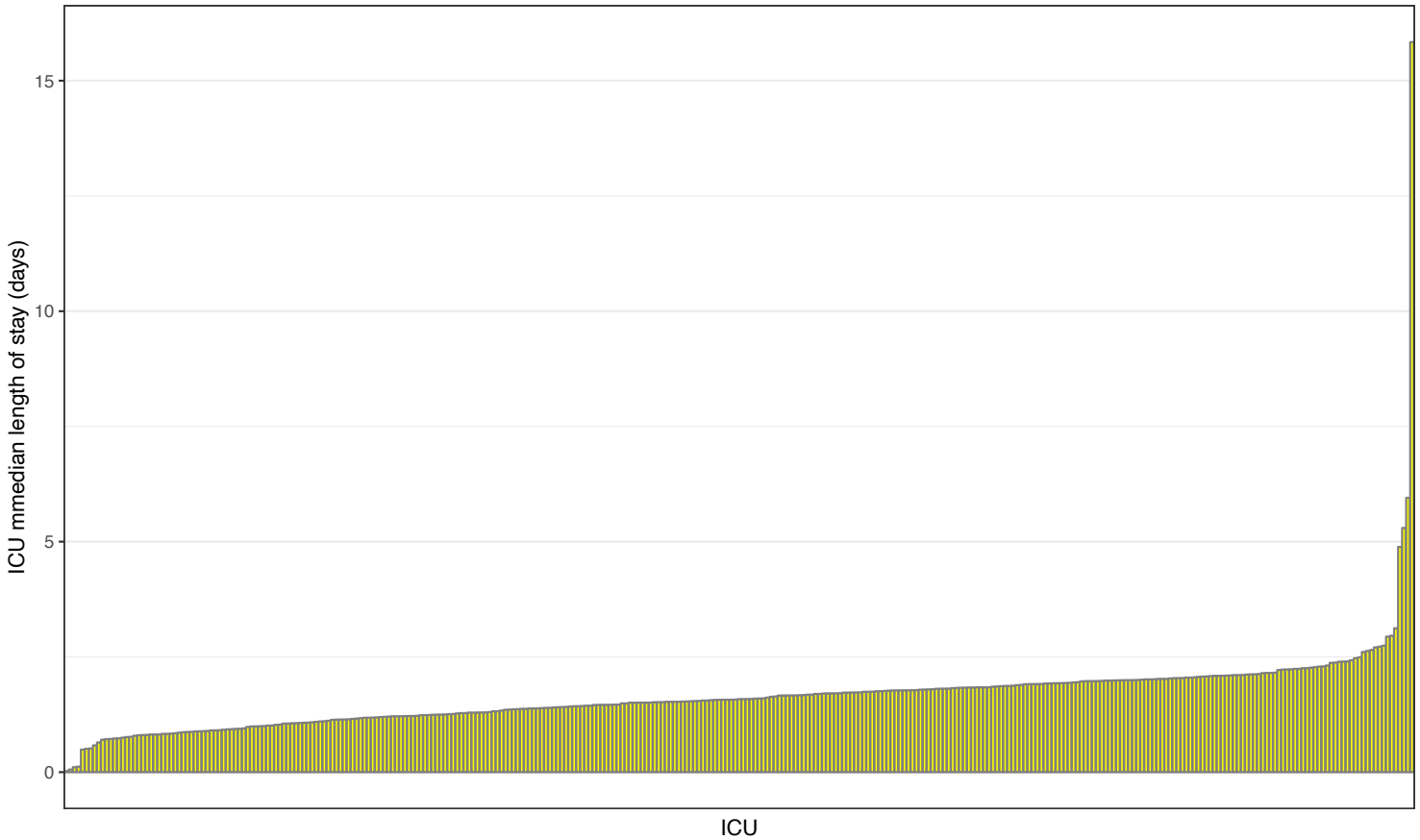
Supplemental Table 4: Providers in the eICU-CRD

Physician specialty	Practitioners
internal medicine	21460
hospitalist	20513
Specialty Not Specified	17742
critical care medicine (CCM)	17491
pulmonary/CCM	12790
cardiology	10049
unknown	8625
surgery-general	6707
surgery-cardiac	6050
pulmonary	5578
family practice	4920
surgery-neuro	3371
surgery-trauma	2444
neurology	2066
other	1778
surgery-vascular	1402
nephrology	730
surgery-critical care	500
emergency medicine	499
obstetrics/gynecology	441
anesthesiology/CCM	418
orthopedics	388
otolaryngology	343
urology	303
oncology	247
gastroenterology	240
surgery-orthopedic	209
hematology/oncology	180
surgery-transplant	168
anesthesiology	165
infectious disease	156
surgery-plastic	151
radiology	142
endocrinology	123
hematology	41
psychiatry	27

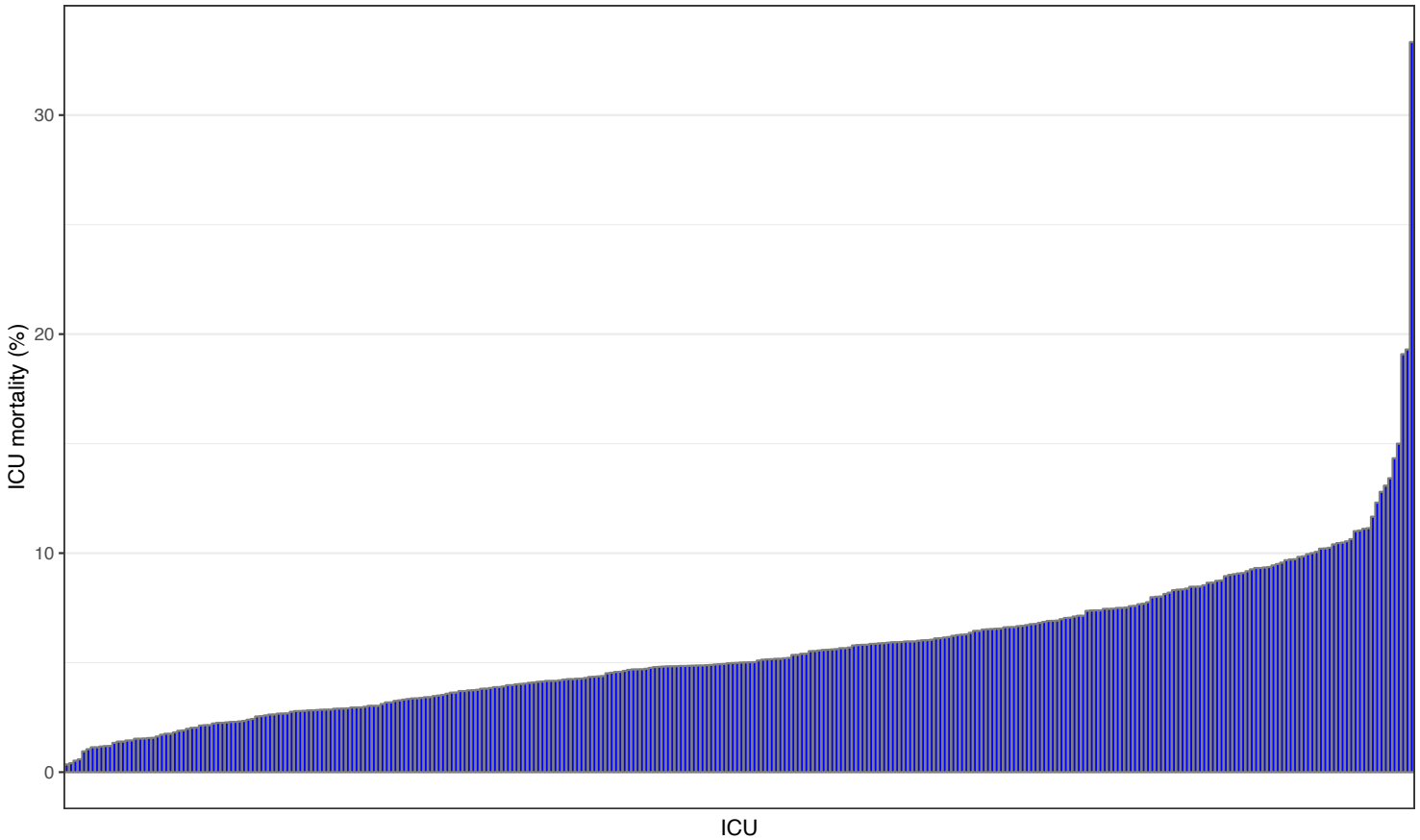
rheumatology	16
surgery-otolaryngology head & neck	16
nurse practitioner	10
ophthalmology	9
surgery-oral	8
physical medicine/rehab	5
ethics	3
nurse	3
allergy/immunology	2
dermatology	1
respiratory therapist	1
surgery-pediatric	1



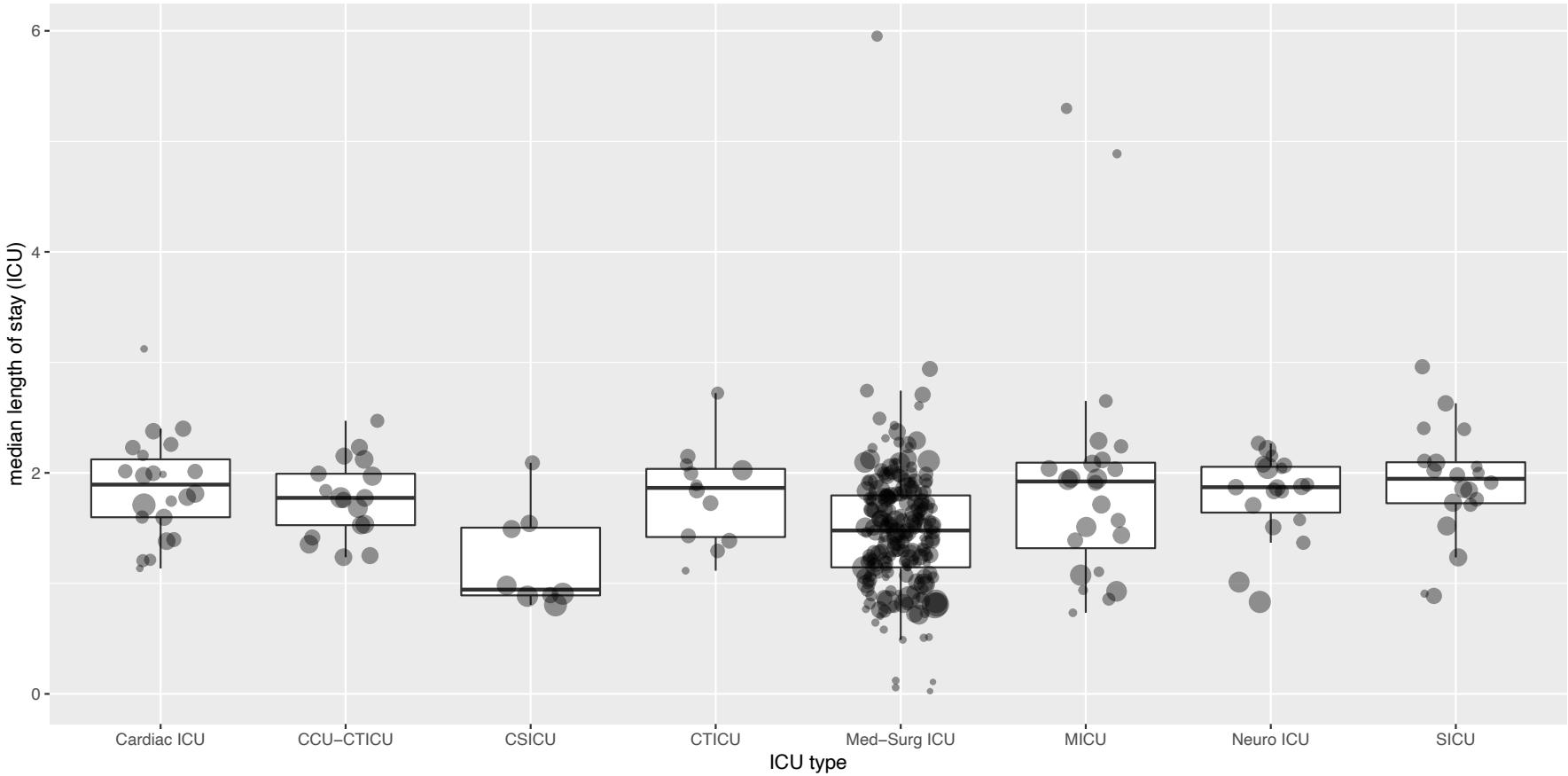
Supplementary Figure 1. Number of encounters per ICU, according to ICU type. Each dot represents an individual ICU. Boxplots show the median and interquartile range (IQR) , with whiskers extending to the most extreme value within 1.5 x IQR of the hinge.



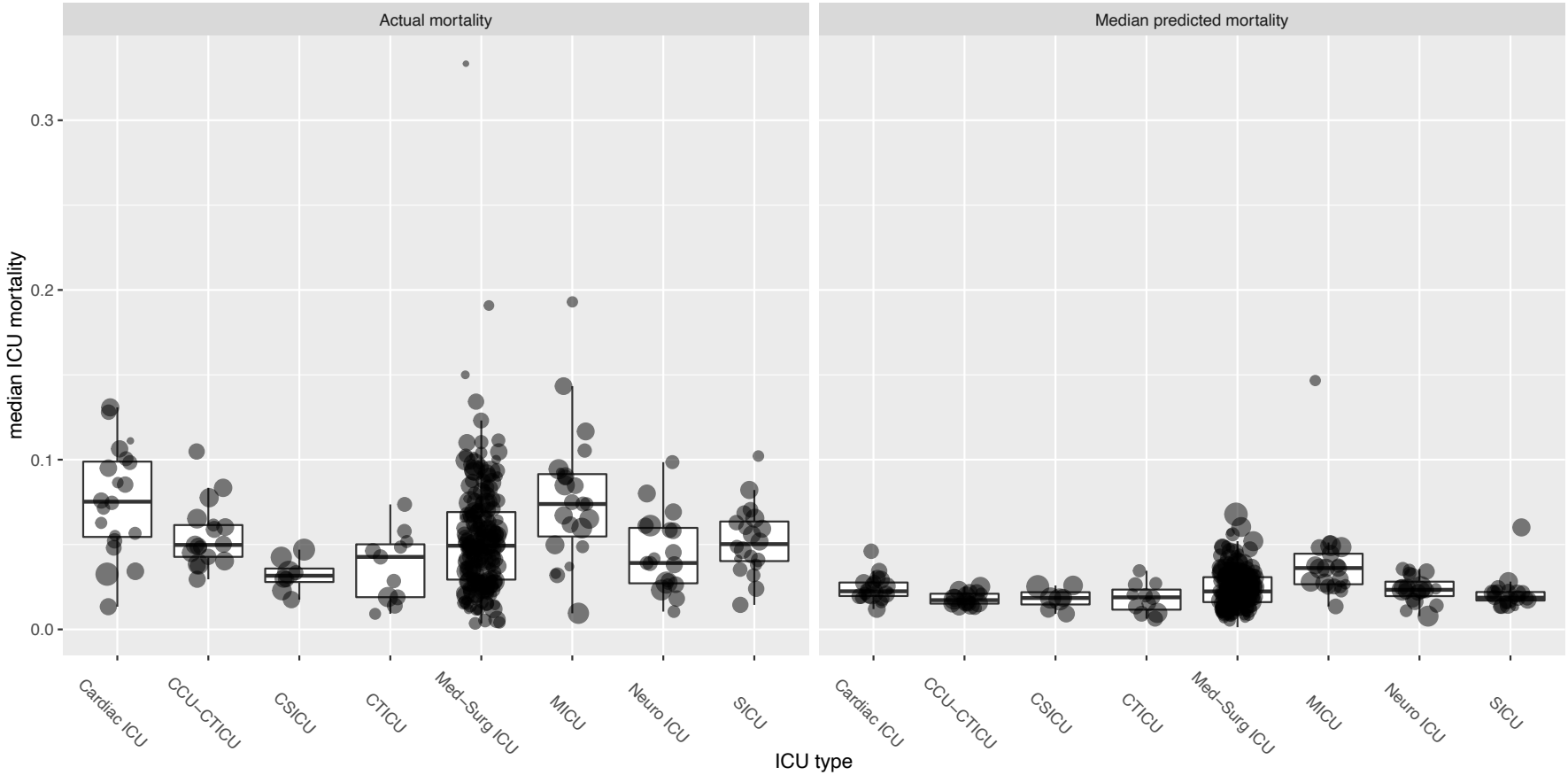
Supplementary Figure 2: Median length of stay for each fo the 335 ICUs in the eICU-CRD database



Supplementary Figure 3: Median mortality for each fo the 335 ICUs in the eICU-CRD database



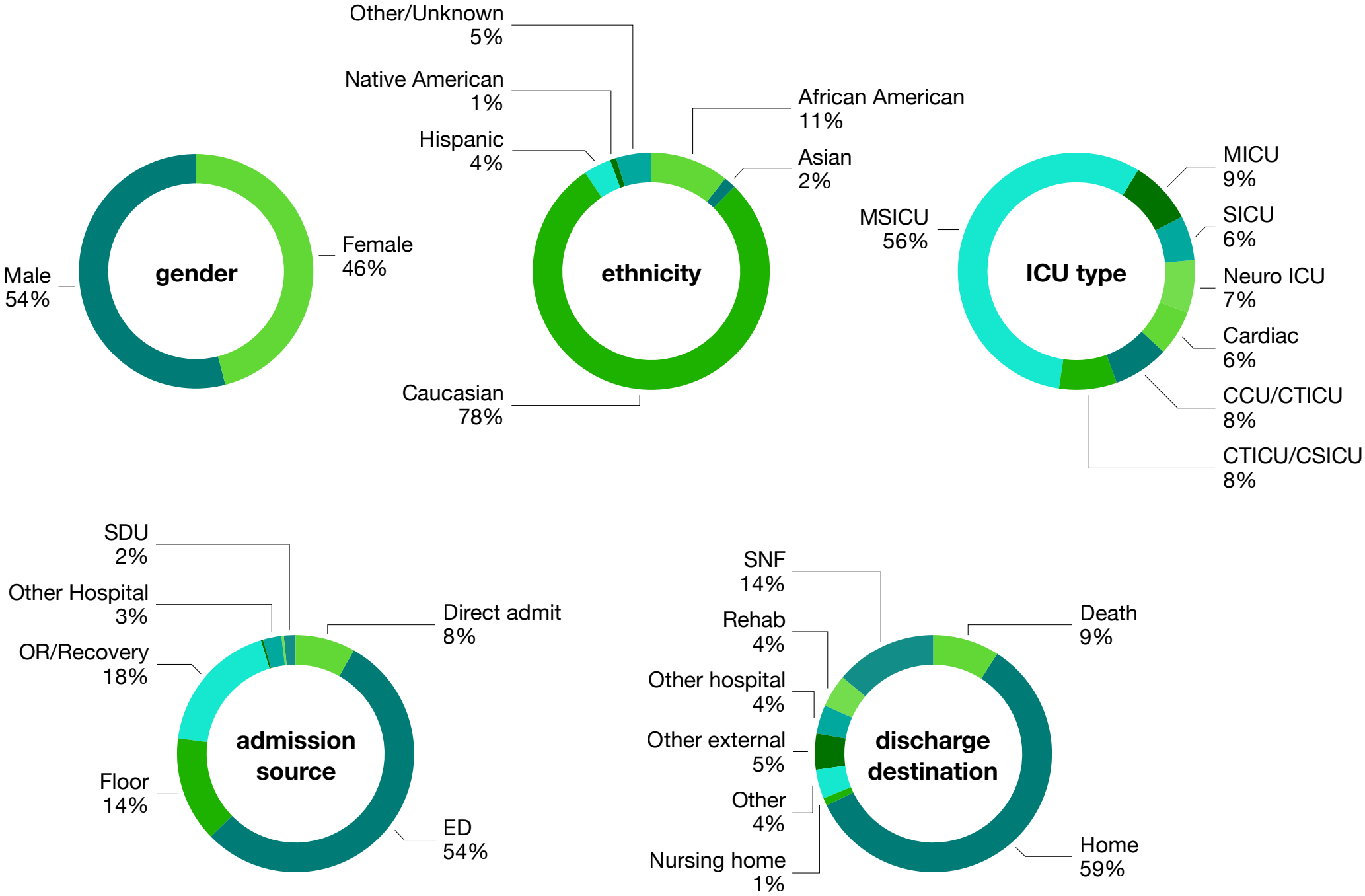
Supplemental Figure 4: Median LOS in each ICU. Each dot represents one ICU. The size of the dot reflects the number of encounters for that unit. Plot excludes one small ICU that had a median LOS 15.8 days (131 encounters)



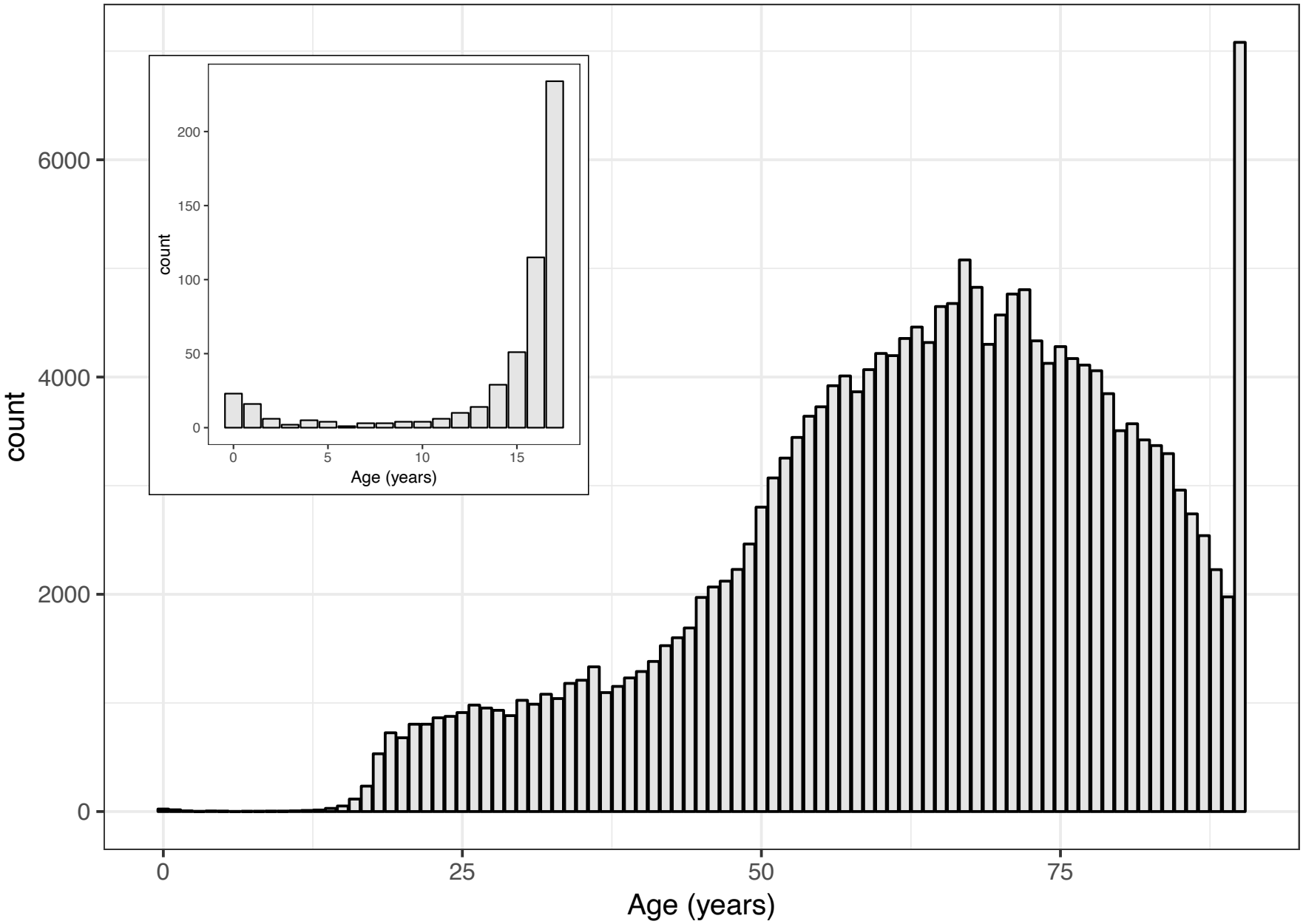
Supplementary Figure 5. Median mortality in each ICU. Each dot represents one ICU. The size of the dot reflects the number of encounters for that unit. The predicted mortality (right) is based on median APACHE IVa score.

Supplemental Table 5: Patient characteristics in the eICU-CRD

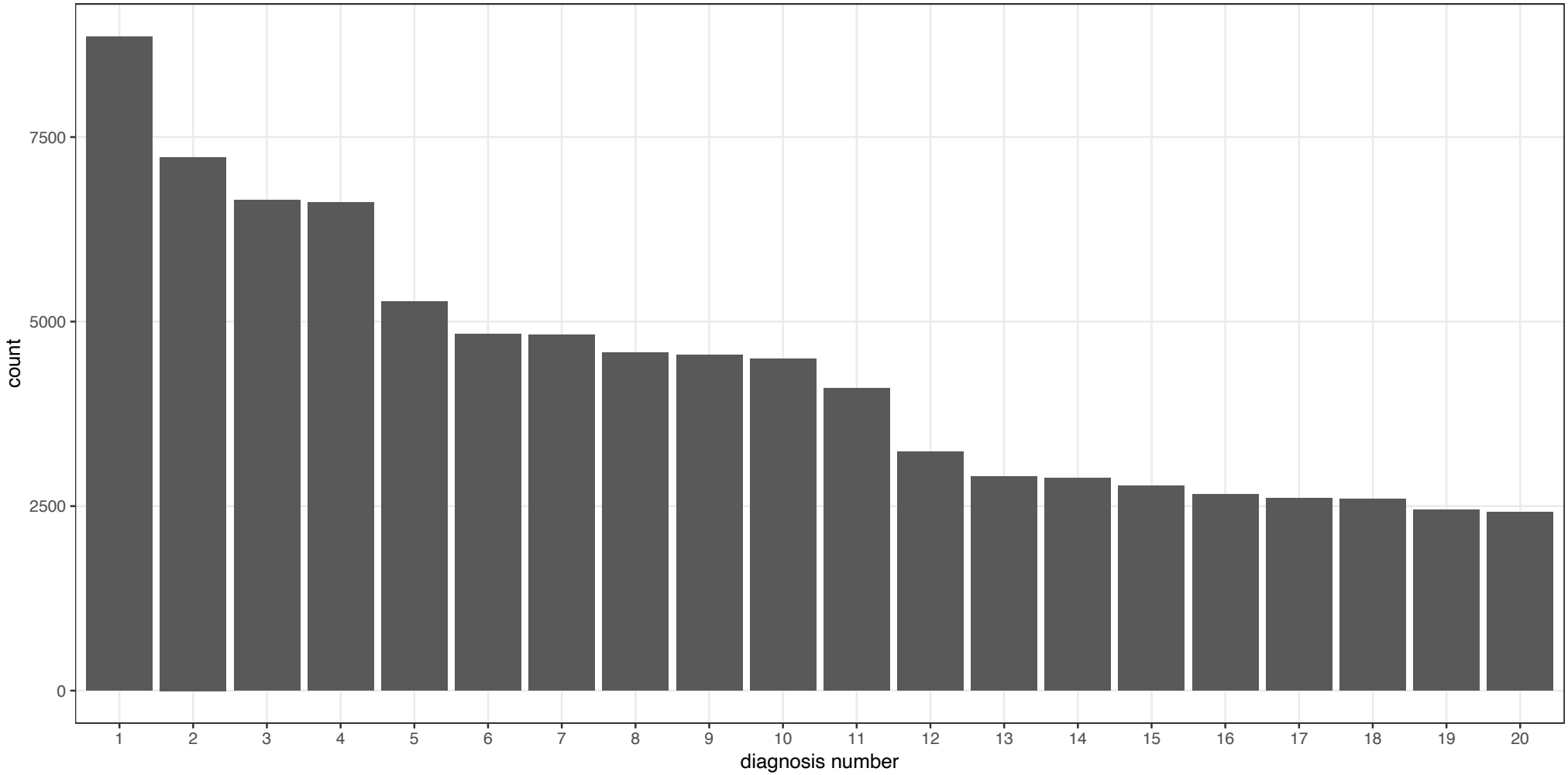
Variable	Count (% of category)
Demographics	
Age (years)	65.0 (IQR 53 to 76)
Age missing/unknown	95 (0.047%)
Gender	
Male	10,8379 (54.0%)
Female	92,303 (46.0%)
Other	8 (0.00398%)
Gender missing/unknown	169 (0.0841%)
Ethnicity	
African American	21308 (10.6)
Asian	3270 (1.63)
Caucasian	155285 (77.3)
Hispanic	7464 (3.72)
Native American	1700 (0.846)
Other	9542 (4.75)
Ethnicity missing/unknown	2290 (1.14%)
Admission source	
Direct admit	12369 (8.17%)
ED	82504 (54.5%)
Floor	21796 (14.4%)
OR/Recovery	928 (0.613%)
Other hospital	3825 (2.53%)
Step-down unit (SDU)	2413 (1.59%)
Admission source missing/unknown	27560 (18.2%)
Discharge destination	
Death	18004 (9.29%)
Home	116816 (60.3%)
Skilled nursing facility (SNF)	27582 (14.2%)
Rehab	8905 (4.59%)
Other hospital	7807 (4.03%)
Other external	9757 (5.03%)
Other	7888 (1.49%)
Nursing home	2067 (1.07%)
Discharge destination missing/unknown	2033 (1.01%)
ICU type	
Cardiac ICU	12467 (6.21%)
CCU/CTICU	15290 (7.61%)
CTIUC/CSICU	9625 (4.79%)
MSICU	6158 (3.07%)
SICU	113222 (56.4%)
MICU	12181 (6.06%)
Neuro ICU	17465 (8.70%)
ICU type missing/unknown	14451 (7.19%)



Supplementary Figure 6. Patient gender, ethnicity, ICU type, admission source, and hospital discharge destination for each ICU encounter



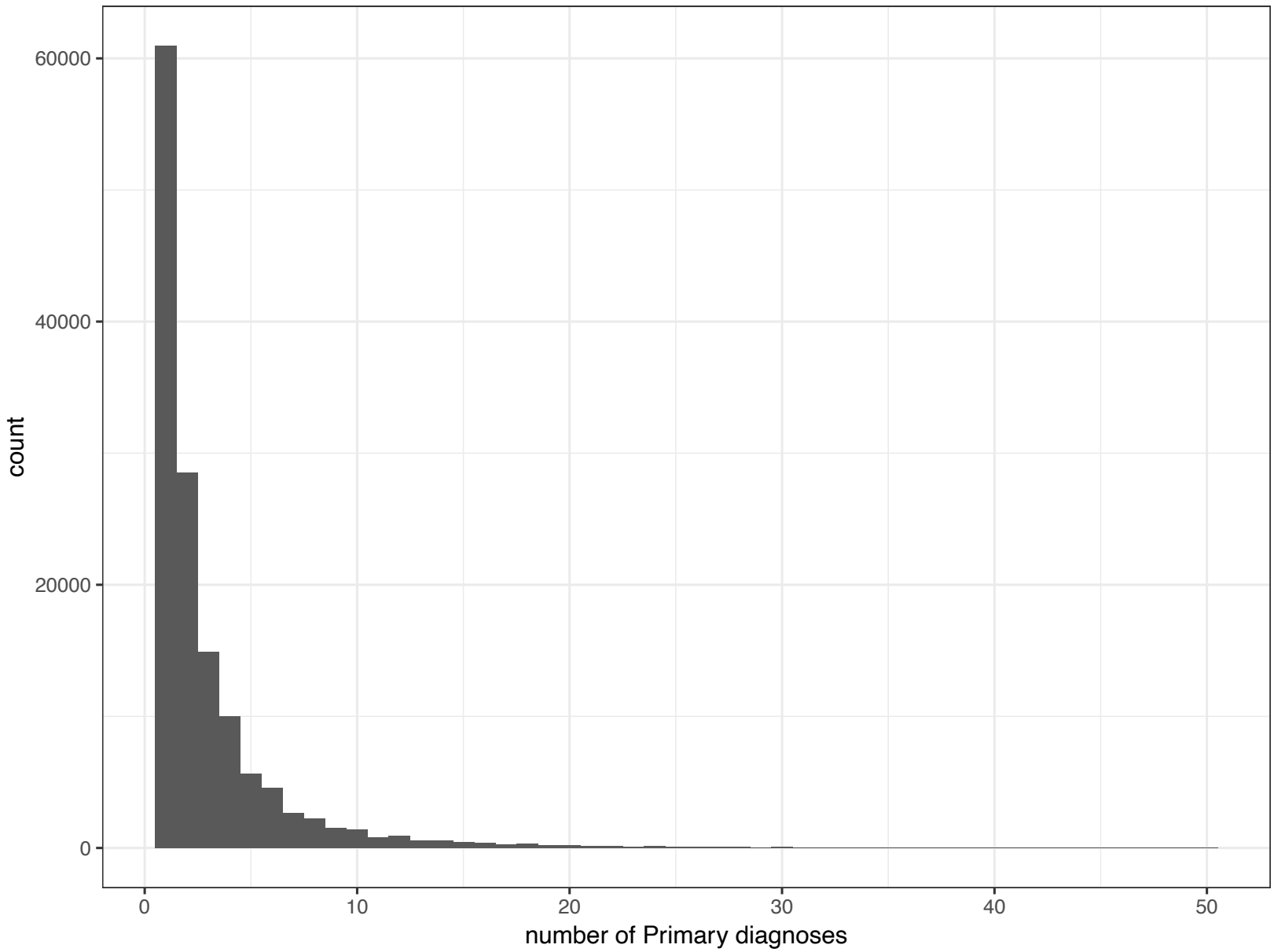
Supplementary Figure 7. Histogram showing ages of patients in the eICU-CRD. Inset shows enhanced histogram for patients aged <18 yrs. The large bar at the end represents "age >89 years", which is a category used by eICU-CRD for very elderly patients in order to minimize the risk of identifying patients based on advanced age.



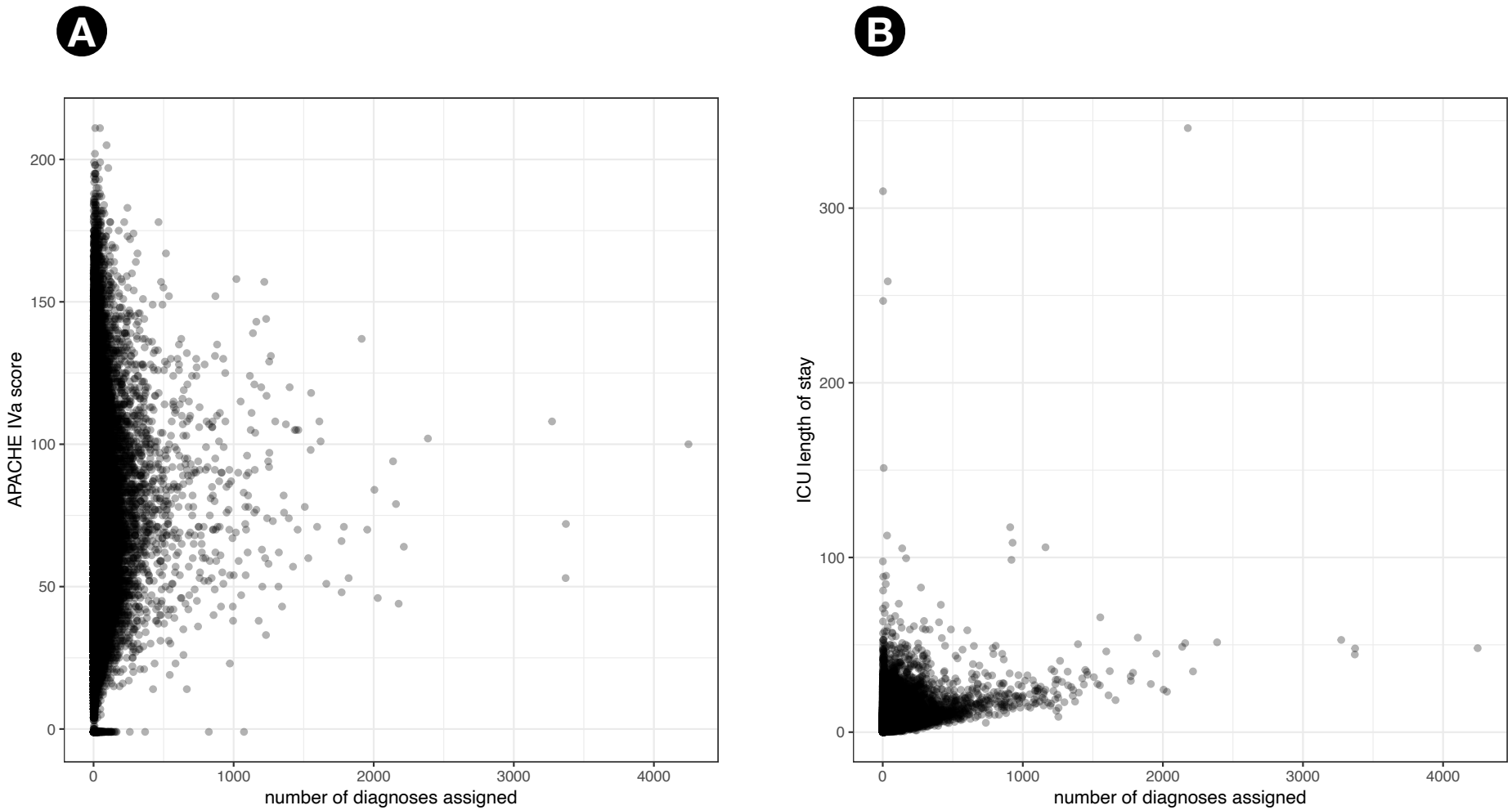
Supplementary Figure 8. Top APACHE diagnoses in the eICU-CRD encounters. 1 - sepsis, pulmonary; 2- Infarction, acute myocardial (MI); 3 - CVA, cerebrovascular accident; 4 - CHF, congestive heart failure; 5 - Sepsis, renal/UTI (including bladder); 6 - Rhythm disturbance (atrial, supraventricular); 7 - Diabetic ketoacidosis; 8 - Cardiac arrest (with or without respiratory arrest); 9 - CABG alone, coronary artery bypass grafting; 10 - Emphysema/bronchitis; 11 - Pneumonia, bacterial; 12 - Bleeding, upper GI; 13 - Coma/change in level of consciousness; 14 - Sepsis, GI; 15 - Seizures (primary-nostructural brain disease); 16 - Angina, unstable; 17 - Sepsis, unknown; 18 - Arrest, respiratory (without cardiac arrest); 19 - Hemorrhage/hematoma, intracranial; Bleeding, lower GI.

Supplementary Table 6: Top APACHE diagnoses

Admission diagnosis	Count
Sepsis, pulmonary	8862
Infarction, acute myocardial (MI)	7228
CVA, cerebrovascular accident/stroke	6647
CHF, congestive heart failure	6617
Sepsis, renal/UTI (including bladder)	5273
Rhythm disturbance (atrial, supraventricular)	4827
Diabetic ketoacidosis	4825
Cardiac arrest (with or without respiratory arrest; for respiratory arrest see Respiratory System)	4580
CABG alone, coronary artery bypass grafting	4543
Emphysema/bronchitis	4494
Pneumonia, bacterial	4101
Bleeding, upper GI	3238
Coma/change in level of consciousness (for hepatic see GI, for diabetic see Endocrine, if related to cardiac arrest, see CV)	2902
Sepsis, GI	2881
Seizures (primary-no structural brain disease)	2776
Angina, unstable (angina interferes w/quality of life or meds are tolerated poorly)	2658
Sepsis, unknown	2602
Arrest, respiratory (without cardiac arrest)	2599
Hemorrhage/hematoma, intracranial	2450
Bleeding, lower GI	2423



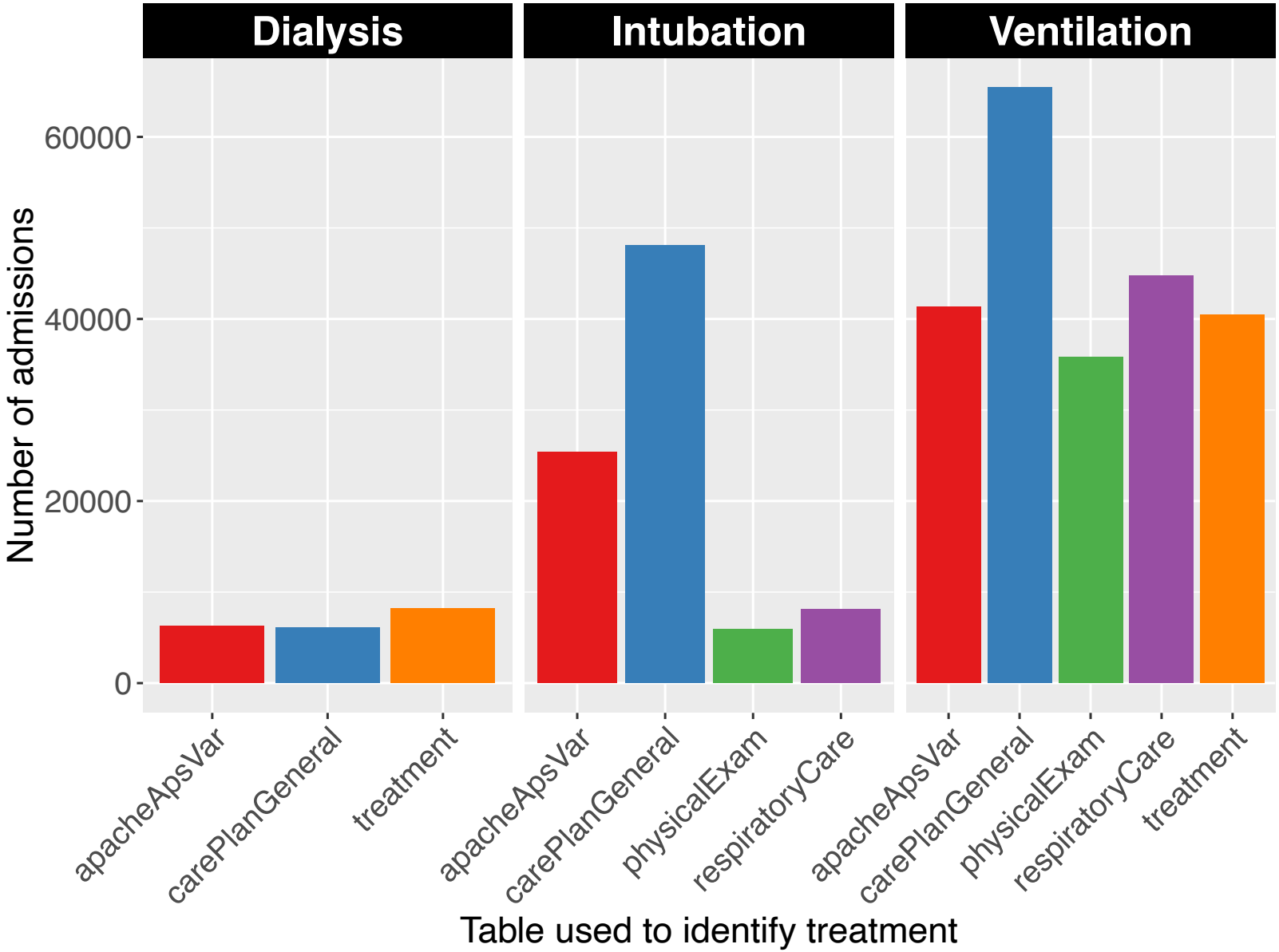
Supplementary Figure 9: Histogram showing the number of encounters with a given number of primary diagnoses assigned. Most encounters had just one primary diagnosis assigned, but some had as many as 499. This plot excludes 236 encounters that had > 50 primary diagnoses assigned.



Supplementary Figure 10: Scatter plots showing the relationship between number of diagnoses assigned and (A) APACHE Iva score, as well as (B) ICU length of stay.

Supplementary Table 7: Top EHR diagnoses

Top EMR Diagnoses	Count
pulmonary|respiratory failure|acute respiratory failure	23497
cardiovascular|shock / hypotension|sepsis	17387
cardiovascular|shock / hypotension|sepsis|severe	12890
cardiovascular|shock / hypotension|septic shock	12583
cardiovascular|ventricular disorders|congestive heart failure	7442
cardiovascular|cardiac arrest|cardiac arrest	7035
cardiovascular|cardiac surgery|s/p CABG < 7 days	6718
pulmonary|pulmonary infections|pneumonia	6451
endocrine|glucose metabolism|DKA	6118
cardiovascular|chest pain / ASHD|acute coronary syndrome|acute myocardial infarction (no ST elevation)	5409
pulmonary|respiratory failure|acute respiratory distress	5397
neurologic|disorders of vasculature|stroke|ischemic stroke	4799
neurologic|altered mental status / pain|change in mental status	4665
cardiovascular|arrhythmias|atrial fibrillation|with rapid ventricular response	4630
gastrointestinal|GI bleeding / PUD|upper GI bleeding	4611
infectious diseases|systemic/other infections|sepsis	4401
pulmonary|disorders of the airways|acute COPD exacerbation	4159
gastrointestinal|GI bleeding / PUD|GI bleeding	4102
renal|disorder of kidney|acute renal failure	4073
cardiovascular|chest pain / ASHD|acute coronary syndrome|acute myocardial infarction (with ST elevation)	3925



Supplementary Figure 11: Rates of dialysis, intubation, and ventilation vary by database table within the eICU-CRD. Using APACHE data from the first 24 hours (red), rates are lower than if the *carePlanGeneral* (blue) table is used, which documents the entire ICU stay. Rates of interventions are unambiguous in *apacheApsVar* (red), where they are designated by a "1" or a "0". In *carePlanGeneral* (blue), *treatment* (orange), *physicalExam* (green), and *respiratoryCare* (purple) interventions are identified by different indicators, leading to variations in the estimates.

Supplemental Table 8: Medications included in the list of vasoactive agents

Epinephrine	Norepinephrine
Phenylephrine	Dopamine
Dobutamine	Vasopressin
Milrinone	Nitroprusside
Nitroglycerine	Diltiazem
Esmolol	Isoproterenol
Labetalol	Nicardipine
Epoprostanol	Amiodarone

Supplemental Table 9: Results for vital signs recordings in *vitalPeriodic* datatable (automatically entered five minute medians)

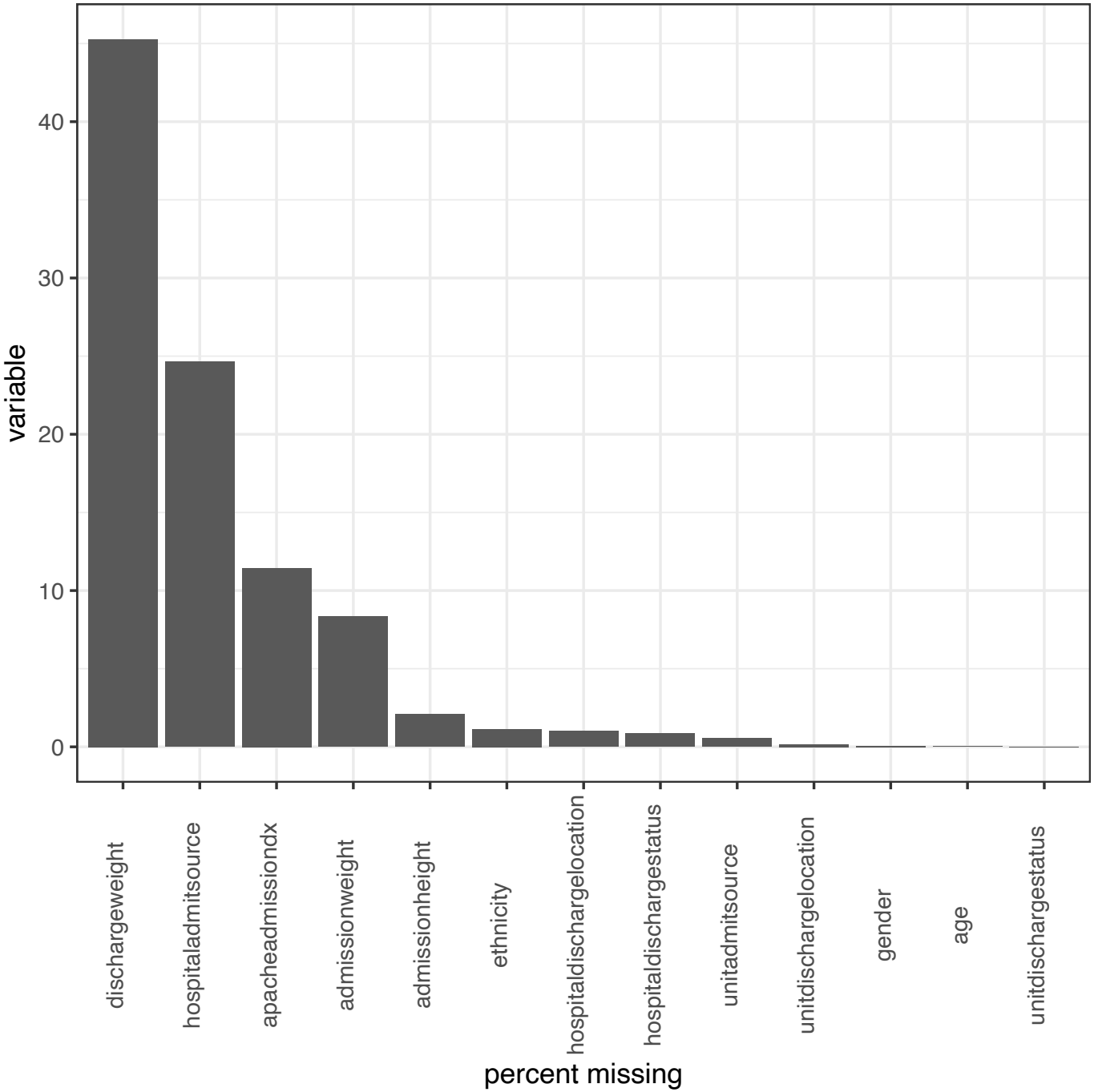
Vital Sign	Number of entries	Number implausible	Percent within realistic range	Mean excluding outliers	Mean including outliers
Heart Rate	145,979,794	0	100%	85.8	85.8
Respiratory Rate	128,501,032	22,659	99.98%	20.09	20.11
Oxygen Saturation	132,908,266	65,448	99.95%	96.7	96.7
Temperature	13,203,289	403,930	96.9%	37.1ºC	39.0ºC
Central Venous Pressure	19,157,758	2,895,336	84.9%	12.0	37.9
Blood Pressure	27,832,823	236,099	99.2%	123/59.3	123/59.6

Supplemental Table 10: Results for vital signs recordings in *vitalAperiodic* datatable

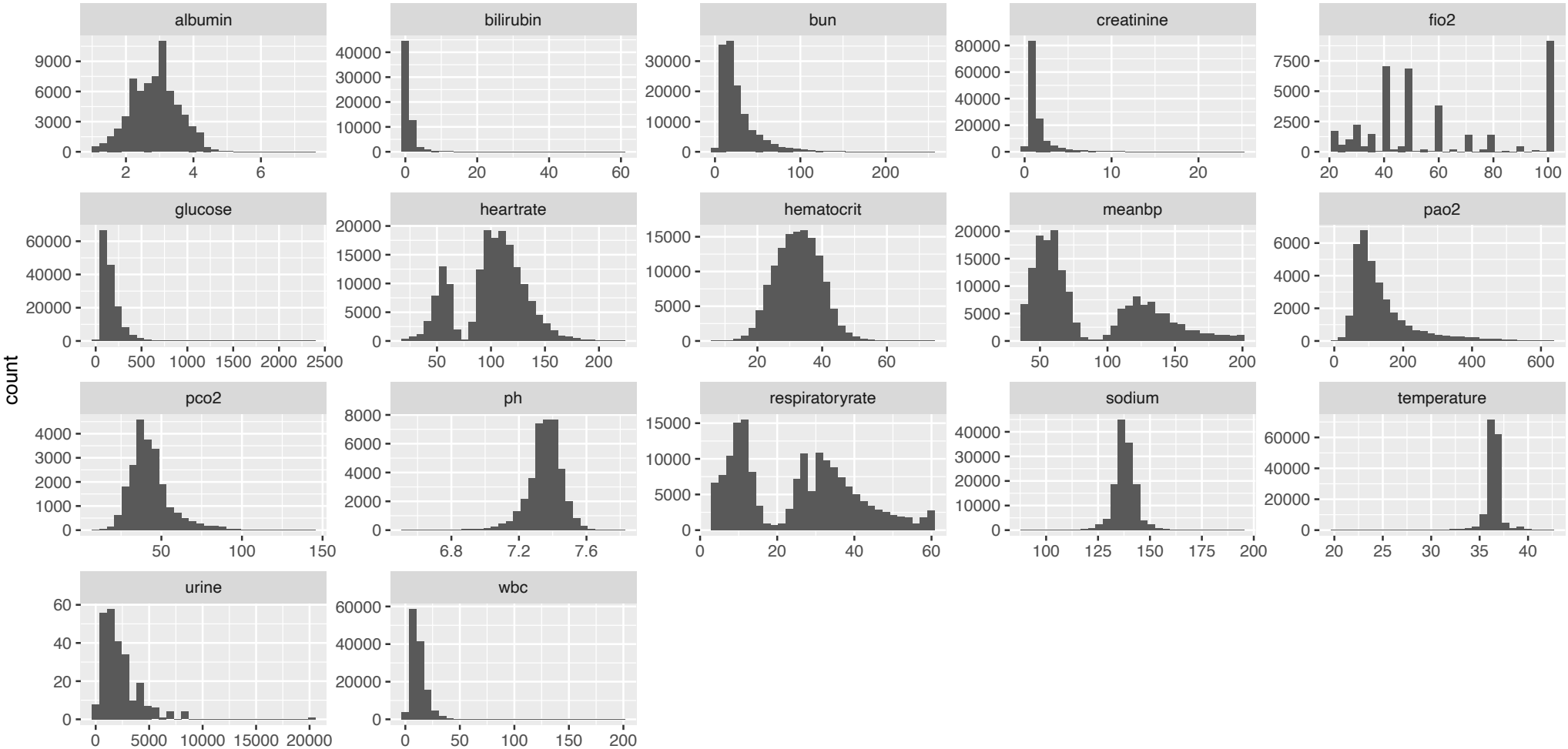
Vital Sign	Number of entries	Number implausible	Percent within realistic range	Mean excluding outliers	Mean including outliers
Blood Pressure	22,034,012	4,747	99.98%	121/65	121/65

Supplemental Table 11: Results for vital signs recordings in *nurseCharting* datatable (manually entered)

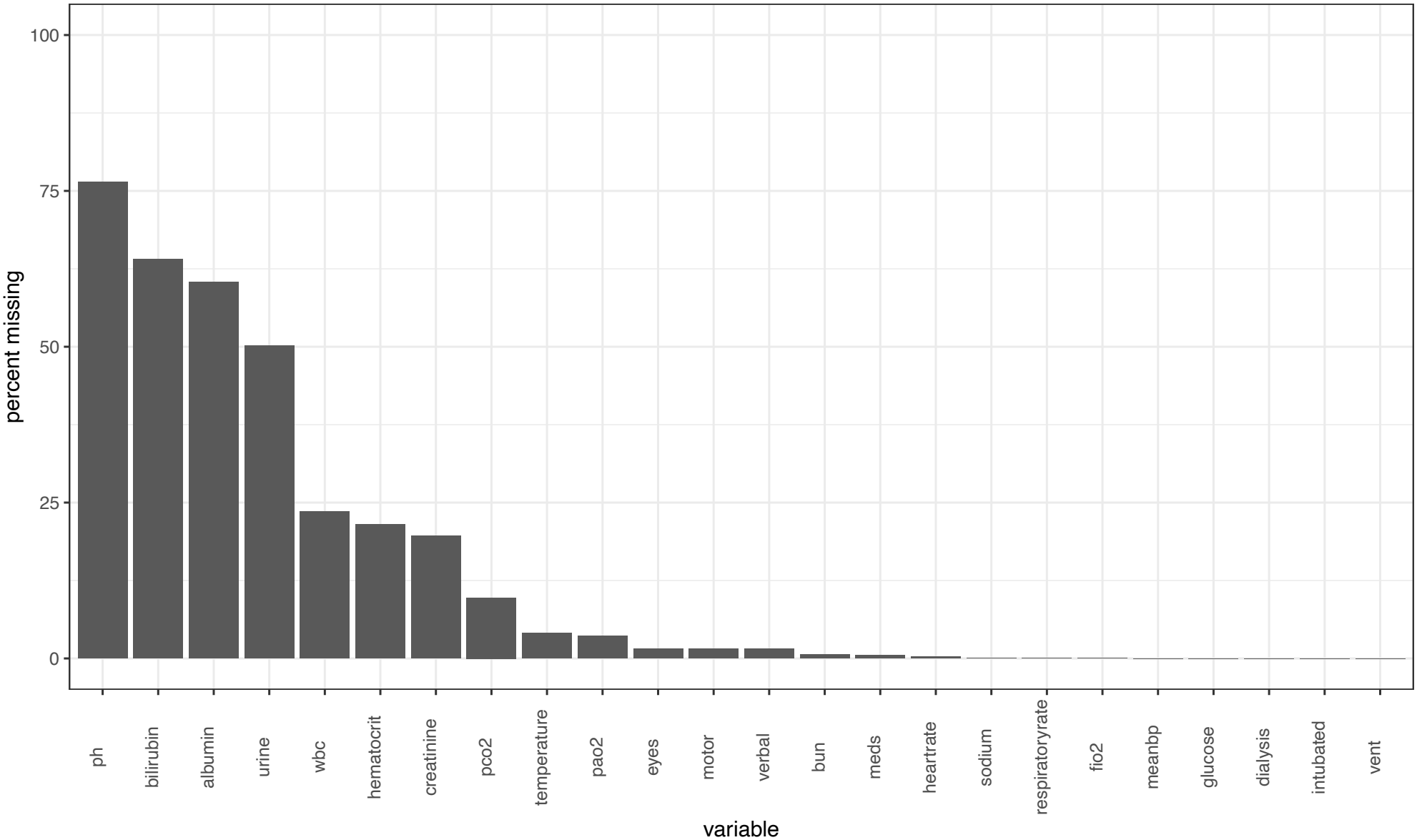
Vital Sign	Number of entries	Number implausible	Percent within realistic range	Mean excluding outliers	Original mean (includes outliers)
Heart Rate	17,565,063	4,909	99.97%	86.6	86.6
Oxygen Saturation	14,661,045	8,037	99.9% (99.95 if 4 digits)	96.6	96.6
Temperature in Celsius	6,267,541	24,885	99.6%	36.9	37.3
Temperature in Fahrenheit	6,267,330	24,824	99.6%	98.4F = 36.9C	98.9F=37.2C
Central Venous Pressure	1,093,874	47,099	95.7%	12.0	18.6
Blood Pressure	11,297,430	14,517	99.9%	122/65.5	122/65.5



Supplemental Figure 12: Rates of missingness from the “patients” table.



Supplemental Figure 13: Histograms of the physiological values used to calculate APACHE IV scores. All values are the best/worst from the first 24 hours of the stay. Missing values not represented.



Supplemental Figure 14: Rates of missingness from among the APACHE variables