Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

William C. Dube   Date

Quantifying excess patient risk for CLABSIs attributed to multiple central line use and consequential flaws in current performance measurement in the CMS Hospital Inpatient Quality Reporting Program

By

William C. Dube

Master of Public Health

Epidemiology

_____

Scott K. Fridkin, MD

Committee Chair

Quantifying excess patient risk for CLABSIs attributed to multiple central line use and consequential flaws in current performance measurement in the CMS Hospital Inpatient Quality Reporting Program

By

William C. Dube

B.A.

University of Colorado

2017

Thesis Committee Chair: Scott K. Fridkin, MD

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology

2019

Abstract

Quantifying excess patient risk for CLABSIs attributed to multiple central line use and consequential flaws in current performance measurement in the CMS Hospital Inpatient Quality Reporting Program

By William C. Dube

Although relatively rare, central-line associated blood-stream infections (CLABSI) are an important healthcare associated infection (HAI). As part of the goal to eliminate HAI, they are included in NHSN reporting, and are a performance measure used to determine hospital re-imbursement from the Center for Medicare and Medicaid Services (CMS). Current NHSN methodology for counting central-line days counts a patient with one line the same as a patient with two or more. This fails to adjust for the additional risk of developing a CLABSI from multiple lines.

We conducted a large, multi-facility retrospective cohort study, including over 52,000 patient admissions and over 61,000 central venous catheters (CVC) from January 2012 to December 2017. We analyzed these data using a propensity score matched logistic regression to compare CLABSI risk in patients with multiple concurrent lines and those with only one line, adjusting for other risk factors.

We found that having any concurrent CVC periods in an admission was associated with a 65% increase in the risk of developing a CLABSI (when using NHSN central-line days). If central-line days are counted as we recommend, the risk of concurrence becomes insignificant, because it is accounted for by the central-line day metric. In these data, the risk for each additional NHSN CVC day was 6%. In contrast, the per additional day risk with our recommended method was 3%.

These findings suggest that in order to make CLABSI risk adjustment fairer for facilities who are likely to have patients with more than one CVC at a time, NHSN central-line reporting methodology needs to be changed. 1 patient with *n* lines should contribute *n* central-line days.

Quantifying excess patient risk for CLABSIs attributed to multiple central line use and consequential flaws in current performance measurement in the CMS Hospital Inpatient Quality Reporting Program

By

William C. Dube

B.A.

University of Colorado

2017

Thesis Committee Chair: Scott K. Fridkin, M.D.

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2019

Acknowledgements

      I would first like to thank my advisor, Scott Fridkin for his feedback and patience in working on this project. I thank Chad Robichaux for extracting the data from the Emory Healthcare system. I would also like to thank my peers for their comradery, support, and assistance though out this process.

Table of Contents

Introduction

Central-line associated blood stream infections are blood stream infections that are attributed to the presence of a central venous catheter (CVC). Although relatively rare compared to other HAIs, central line-associated blood stream infections (CLABSI) are an important healthcare-associated infection (HAI)(1). These infections are important because of their clinical significance (up to 1 in 4 of these patients die from the infection), and their economic impacts as well. The Centers for Disease Control and Prevention (CDC) have established goals to try to eliminate CLABSI, and provide guidelines to government, hospitals, and clinicians on how they can help to reduce CLABSI incidence(2,3).

As part of working towards elimination of CLABSI, the Centers for Medicare & Medicaid Services (CMS) has implemented financial incentives to keep CLABSI rates down and reduce SIR. They do this by including NHSN CLABSI SIR in their healthcare-acquired condition score (HAC). Hospitals who are above the 75th percentile (in the worst quartile) in HAC receive a 1% reduction in their reimbursement from CMS (4).

Given the financial stakes, equitability in performance measures of HAIs is important to make sure that facilities are being evaluated fairly. In order to be effective, performance metrics should be both reliable and objective(5). They should provide the correct answer regardless of the context, accounting for differences in the degree of patient's risk for infection unrelated to actual performance of infection prevention between facilities and their patients. Previous discussions of improving the performance measure of CLABSI have often focused on concerns related to standardization of case ascertainment (5). However, we contend that the current methodology for quantifying the denominator as specified by NHSN protocols for performing

surveillance of CLABSI is flawed. The denominator reflects the sum of person-time at risk for CLABSI and are reflected as the sum of central line days (CLD) that occurred in the hospital location under surveillance for CLABSI.

Currently, NHSN defines CLD as the sum of patients on each day with any central-line present (as observed at the same time period each day). This means one patient with n CVCs on one day counts as one catheter day(6). However, it makes intuitive sense that a patient with more than one CVC at the same time (i.e., concurrent CVC use) would be at greater risk of CLABSI, as there are more breaches of biological barriers to organisms entering the circulatory system. If in fact the attributable risk for CLABSI among patients with concurrent CVC use is high, it would follow that hospital locations that care for patients requiring concurrent CVC use have higher risk patients compared to a second location with a similar NHSN designation but less concurrent CVC use – differences not accounted for with current NHSN risk adjustment methodology. Current risk adjustment accounts only for the category of patient location and number of CLD as defined above. Considering some acute care hospitals may care for patients which require concurrent CVC more often than their peer facilities, the risk adjustment methodology could be improved to allow more fair comparisons when calculating a performance metric. To that end, we argue that if patient's risk for CLABSI is measurably higher with concurrent CVC use, than one patient with n catheters for one day should count for n catheter days(6).

In order to evaluate the excess risk of a patient with more than one CVC, we utilized a multi-year, multi-hospital dataset. We used an advanced statistical approach to allow us to compare the risk of having concurrent CVCs versus having only one CVC at a time. These

techniques combined with our large sample size allowed us to make these comparisons while also controlling for several other risk factors.

Methods

Study Population

We conducted a retrospective cohort study among patients from four hospitals in the Emory Healthcare system in Atlanta, GA, USA. CVC data from January 1, 2012 to December 31, 2017 were extracted from Emory Healthcare's (EHC) clinical data warehouse (CDW) with Emory IRB approval. CVC level data were extracted from the CDW including line type, line site, insertion and removal dates, and encounter number (allowing joining with the encounter level data). Line level data were limited to CVCs that were inserted for at least two days. Patient encounter information for all admissions with at least one CVC were also extracted. This included discharge codes (allowing calculation of Charlson Comorbidity Index [CCI], orders for total parenteral nutrition (TPN) and chemotherapy, age, sex, admission date, and discharge date. To simplify the analytic approach, and focus the research question on typical adult inpatients, encounters were eliminated if they were in patients under 18 years of age, had a LOS of more than 49 days, or had more than 5 CVCs. Surveillance data for CLABSI as had been reported to NHSN by EHC were linked to patient encounters. The dataset was comprised of 64,575 line observations in 52,383 distinct patient admissions. Further information about the data source and preparation are in the appendix.

CVC Characteristics and Presence of Multiple Catheters

First, we defined line types into 4 types: port, PICC, dialysis, and temporary. Indicator variables were created for each type, for each line. We also quantified the duration of each line

by subtracting the insertion date from the removal date. We retained the insertion and removal

dates for each line in order to evaluate concurrence of lines in an encounter.

Presence of multiple catheters was assessed in two ways. First, we evaluated whether

there were any periods in an encounter where at least two CVCs were present. This was used to

create a dichotomous variable: any concurrence versus no concurrence. Second, we created

two variables for CLD. One where we employed the NHSN guidelines (one CLD for each day any

CVCs were present), and the other where we counted each line for each day (e.g. if two lines

were in for three days, that would be six CLD). We calculated the proportion of the total CLD

that was accounted for by the NHSN CLD for each encounter (NHSN CLD/total CLD). Further

details on the line level cleaning, and characteristics of those lines are available in the appendix.

Propensity Score Matching

In order to select a group of patients without concurrent CVCs that was most

comparable to our patients with concurrent CVCs, we conducted a propensity score match. This

process was conducted using the "MatchIt" package in R (7–9). Briefly, we used logistic

regression to determine a patient's likelihood of having concurrent CVCs (using our

dichotomous variable for any concurrence) given a set of covariates. These covariates included

hemodialysis, chemotherapy, TPN, and CCI. Then, based on the likelihood of concurrence, we

selected two non-concurrent encounters for every encounter that had concurrent CVCs. This

allowed us to use the propensity score in the model in order to control for all four of the factors

(10,11). Further details on the propensity score matching process are available in the appendix.

Statistical Analyses

All analyses were conducted in R, with various packages (7,12,13). First, we completed a

descriptive analysis, where compared all of our covariates in two ways. First, comparing them

between patients with concurrent lines at any point, and those without concurrence. Second,

between patients who had a CLABSI during their admission and those who did not. We also

compared NHSN CLD, total CLD, and proportion of total CLD accounted for by NHSN in these

two stratification schemes. Statistical significance of the differences in strata were evaluated

using chi-square tests for the categorical variables, two sample t-test for the age variable, and

Mann-Whitney tests for the three CLD variables (since they were not normally distributed).

Next, we completed a logistic regression model with a binomial distribution using our

propensity score matched data(10). The outcome of interest was incidence of a CLABSI within

the patient encounter. The predictors included the propensity score calculated in the matching

process, any CVC concurrence within the encounter, age, and sex. We also built a second model,

which was the same we used the total central line days instead of NHSN.

Results

Full Dataset

Over the 4-year study period, 52,383 encounters included patients over 18 years old that utilized a CVC for at least 2 days and were included in the analytic dataset. Among these patients, encounters were linked to 647 occurrences of CLABSI. The majority (499, 77%) of these CLABSIs occurred in patients that were exposed only to a single CVC; however, CLABSI patients often had at least 2 concurrent lines (148, 23%) during their encounter. These CVCs had a median dwell times of 14 days (IQR: 8,21) in CLABSI patients, compared to 5 days (IQR: 3, 11) in patients who did not develop a CLABSI (Table 1). Almost 1 in 4 CLABSI patients had concurrent CVCs at some point in their hospital stay (table 1). 3% of the patients with concurrent CVCs at any point during their admission had a CLABSI vs. 1% percent of patients without any concurrence (Table 2). The concurrent and non-concurrent groups were very imbalanced with respect to a number of covariates, this was improved on in the propensity matched data.

Matched analysis

The propensity matched dataset included 17,724 patients. Of those, 279 (2%) had a CLABSI during their admission. 5,908 (33%) patients had concurrent lines at some point in their stay, 11,816 (67%) did not. Among the 5,908 patients with concurrent lines, 148 (3%) had a CLABSI, while 131 (1%) of the 11,816 patients without any concurrence had a CLABSI during their stay. (Table 3, 4)

Propensity score matching improved balance across all of the covariates included in the score, however, TPN and chemotherapy were still significantly different between the two groups, although differences were small (Table 3). The median number of NHSN central line

days was 12 (IQR: 6, 19) for the group with concurrence and 5 (IQR: 3, 10) for the non-concurrent group. The median number of total central line days was 21 (IQR: 12, 34) vs. 5 (IQR: 3, 10) for the concurrent group and non-concurrent group, respectively; the percentage of total central line days accounted for by NHSN methodology in the concurrent group was a median of 53.6% (IQR: 50.0, 65.5; Table 3). The distribution of this variable is shown graphically in figure 1.

The unadjusted risk ratio for developing a CLABSI among patients with concurrent CVC use was 2.26 compared to propensity-matched patients without concurrence (Table 5). The median NHSN central line days among patients who had a CLABSI was 15 (IQR: 8, 22) vs. 7 (IQR: 4, 13) (Table 4). Median total central line days was 22 (IQR: 12, 34) for CLABSI patients compared to 8 (IQR: 4, 17) for patients who did not have a CLABSI in their admission (Table 4). For percentage of total line days captured by NHSN, the median was 75.7 (IQR: 50.0, 100.0) for patients who had a CLABSI, and 100.0 (IQR: 66.7, 100.0) for those who did not (Table 4).

Model Results

Since CLABSI is a rare outcome, the odds ratios from our logistic regression model can be used to approximate risk ratios(14). Patients with any concurrent CVC use were at 65% higher risk of developing an CLABSI than comparable patients without concurrent CVC use, with an adjusted risk ratio for patients with concurrent CVC use of 1.65 (95% CI: 1.27, 2.13; Table 5). The adjusted risk ratio for NHSN central line days was 1.06 (95% CI: 1.04, 1.07; Table 5). Using total line days instead of NHSN, and the adjusted risk ratio for concurrence became non-significant at 1.18 (95% CI: 0.87, 1.60; Table 5). The adjusted risk ratio for total central line days was 1.03 (95% CI: 1.03, 1.04; Table 5).

Discussion

Here, we show that risk of CLABSI is increased for people who have concurrent lines, and therefore, current methodology does not adequately risk adjust for these patients. This difference in risk is show by the difference in risk per day between the NHSN method, and our proposed method. With our proposed method, the risk of each central line day is halved, and the risk of having multiple lines is insignificant (because we are already accounting for it with our line days).

Our data show that current NHSN methodology does not account for the inherent increased risk of a CLABSI for a patient having multiple central lines in use during their period of central line use. Our analysis suggests that risk to be at least 65% higher among patients having more than a single central-line in use compared to those without concurrent line use, independent of the duration of central-line days as defined by NHSN. Relying on solely NHNS defined central-line days when comparing CLABSI experience between similar types of locations as reported to NHSN, does not account for this substantial difference in patient's risk.

In our second model, where we adjusted for total central line days, rather than NHSN central line days, we saw our any concurrence variable drop out as a significant predictor. This is because the increased risk from the concurrent lines is accounted for by the different line day reporting. Additionally, we observed a reduction in the risk ratio for a line day from 1.06 for NHSN to 1.03 when counting all device days. Taken together, this indicates that the method we are suggesting for quantifying central line days was able to account for the additional CLABSI risk from having multiple CVC. In contrast, NHSN does not account for this risk, which results in the device day risk of NHSN central line days being much higher than the device day risk when counting all eligible lines.

Our estimates for the increase in risk due to CVC concurrence are lower than another paper which sought to examine the influence of concurrence on individual level CLABSI risk (15). That study was small in comparison to this one and may have suffered some bias in their odds-ratios because of that. Further, our value almost reaches the bottom of their fairly wide confidence interval. Regardless of the magnitude, both studies show the same direction of association between concurrent CVCs and CLABSI.

A previous study found that counting CLDs in the same way we did resulted in a 53.6% increase when compared to the NHSN methodology in the same data, resulting in a 36% reduction in CLABSI rates in their population(6). This is an important note, showing that these differences have a large influence on the facility level rates. Our analysis provides an individual level explanation for the facility level effects they observed, adding another piece of evidence to the case for reporting all central-lines. For limitations, it is important to note that our characterization of concurrence was if two lines were present with at least one day of overlap. This likely gives a good sense of truly overlapping lines. However, due to the nature of the line data collection, it is possible that lines were recorded as being in longer than they actually were. There are techniques for validating these insertion and removal dates which may provide more assurance about length of overlap, compared to our one-day minimum overlap we imposed.

In summary, our findings indicate that current NHSN reporting methodology is not appropriate for accurately characterizing the denominator in a facility's CLABSI rate. In order to improve the performance measure to be fairer and more equitable, all eligible central lines should contribute their own central line days. This is supported by the underestimation of central line days we showed in the concurrent group, as well as the risk from concurrent central lines shown by our model where we used NHSN central line days.

References

1. Magill SS, O'Leary E, Janelle SJ, et al. Changes in Prevalence of Health Care–Associated Infections in U.S. Hospitals. *N. Engl. J. Med.* 2018;379.
2. CDC Vital Signs. 2011 1-4 p.
3. Cardo D, Dennehy PH, Halverson P, et al. Moving toward elimination of healthcare-associated infections : A call to action. *Am. J. Infect. Control* [electronic article]. 2010;38. (http://dx.doi.org/10.1016/j.ajic.2010.09.001)
4. Hospital-Acquired Condition Reduction Program Fiscal Year 2019 Fact Sheet. 2019 1-5 p.
5. Fridkin SK, Olmsted RN. Meaningful measure of performance : A foundation built on valid, reproducible findings from surveillance of health care-associated infections. *Am. J. Infect. Control* [electronic article]. 2011;39. (http://dx.doi.org/10.1016/j.ajic.2011.01.002)
6. Aslakson RA, Romig M, Jr SMG, et al. Effect of Accounting for Multiple Concurrent Catheters on Central Line – Associated Effect of Accounting for Multiple Concurrent Catheters on Central Line – Associated Bloodstream Infection Rates : 2011;
7. Team RC. R: A language and environment for statistical computing. 2018;(https://www.r-project.org/)
8. Ho DE, Imai K, King G, et al. MatchIt : Nonparametric Preprocessing for Parametric Causal Inference. 2011;42(8).
9. Hansen BB, Klopfer SO. Optimal full matching and related designs via network flows. *J. Comput. Graph. Stat.* 2006;15.
10. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav. Res.* 2011;46.
11. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*. 1983;70.
12. Grolemund G, Wickham H. Dates and Times Made Easy with lubridate. *J. Stat. Softw.* [electronic article]. 2011;40. (http://www.jstatsoft.org/v40/i03/)
13. Wickham H. tidyverse: Easily Install and Load the "Tidyverse". R package version 1.2.1. 2017;(https://cran.r-project.org/package=tidyverse)
14. Viera AJ. Odds Ratios and Risk Ratios : What's the Difference and Why Does It Matter ? *South. Med. J.* 2008;730–734.
15. Concannon C, van Wijngaarden E, Stevens V, et al. The Effect of Multiple Concurrent Central Venous Catheters on Central Line–Associated Bloodstream Infections. *Infect. Control Hosp. Epidemiol.* [electronic article]. 2014;35:1140–1146. (https://www.cambridge.org/core/product/identifier/S0899823X00192852/type/journal_article)

Tables & Figures

Table 1. Characteristics of Patients with CLABSI and Control Patients (Full dataset)

	CLABSI			
	Yes (n=647)	No (n=51,736)	P-value	Total (n=52,383)
Any time with overlapping lines				
Concurrent	148 (23 %)	5,760 (11 %)	**<0.0001**	5,908 (11 %)
Non-Concurrent	499 (77 %)	45,976 (89 %)		46,475 (89 %)
Sex			**0.002**	
Female	289 (45 %)	26,332 (51 %)		26,621 (51 %)
Male	358 (55 %)	25,404 (49 %)		25,762 (49 %)
Hemodialysis			**0.037**	
Yes	128 (20 %)	8,601 (17 %)		8,729 (17 %)
No	519 (80 %)	43,135 (83 %)		43,654 (83 %)
PICC			**0.024**	
Yes	213 (33 %)	19,310 (37 %)		19,523 (37 %)
No	434 (67 %)	32,426 (63 %)		32,860 (63 %)
Port			**0.880**	
Yes	161 (25 %)	12,701 (25 %)		12,862 (25 %)
No	486 (75 %)	39,035 (75 %)		39,521 (75 %)

Other			<0.0001	
Yes	309 (48 %)	18,077 (35 %)		18,386 (35 %)
No	338 (52 %)	33,659 (65 %)		33,997 (65 %)
TPN			<0.0001	
Yes	150 (23 %)	5,532 (11 %)		5,682 (11 %)
No	497 (77 %)	46,204 (89 %)		46,701 (89 %)
Chemotherapy			0.167	
Yes	29 (4 %)	1,764 (3 %)		1,793 (3 %)
No	618 (96 %)	49,972 (97 %)		50,590 (97 %)
Length of Stay (days)			<0.0001	
Median (Q1, Q3)	23.0 (15.0, 30.0)	9.00 (5.00, 16.0)		9.00 (5.00, 16.0)
Age (years)			0.539	
Median (Q1, Q3)	59.0 (45.0, 68.0)	59 .0 (45.0, 69.0)		59.0 (45.0, 69.0)
CCI			<0.0001	
Median (Q1, Q3)	5.00 (2.00, 8.00)	4.00 (2.00, 7.00)		4.00 (2.00, 7.00)
Missing	0 (0%)	2 (0.0%)		2 (0.0%)
NHSN Central Line Days			<0.0001	
Median (Q1, Q3)	14.0 (8.00, 21.0)	5.00 (3.00, 10.0)		5.00 (3.00, 11.0)

All Central Line Days		<0.0001	
Median (Q1, Q3)	16.0 (9.00, 26.0)	6.00 (3.00, 11.0)	6.00 (3.00, 12.0)
% Central Line Days Accounted for by NHSN		<0.0001	
Median (Q1, Q3)	100 (100, 100)	100 (100, 100)	100 (100, 100)

Table 2. Characteristics of Patients who had a multiple line at any time during their stay compared to those who never had multiple lines (Full dataset)

	Concurrent (n=5,908)	Non-Concurrent (n=46,475)	P-value	Total (n=52,383)
Central Line-Associated Blood-stream Infection			<0.0001	
Yes	148 (3 %)	499 (1 %)		647 (1 %)
No	5,760 (97 %)	45,976 (99 %)		51,736 (99 %)
Sex			<0.0001	
Female	2,780 (47 %)	23,841 (51 %)		26,621 (51 %)
Male	3,128 (53 %)	22,634 (49 %)		25,762 (49 %)
TPN			<0.0001	
Yes	985 (17 %)	4,697 (10 %)		5,682 (11 %)
No	4,923 (83 %)	41,778 (90 %)		46,701 (89 %)
Chemotherapy			<0.0001	
Yes	106 (2 %)	1,687 (4 %)		1,793 (3 %)
No	5,802 (98 %)	44,788 (96 %)		50,590 (97 %)
Hemodialysis			<0.0001	
Yes	2,843 (48 %)	5,886 (13 %)		8,729 (17 %)
No	3,065 (52 %)	40,589 (87 %)		43,654 (83 %)

PICC			**<0.0001**	
Yes	2,700 (46 %)	16,823 (36 %)		19,523 (37 %)
No	3,208 (54 %)	29,652 (64 %)		32,860 (63 %)
Port			**0.302**	
Yes	1,418 (24 %)	11,444 (25 %)		12,862 (25 %)
No	4,490 (76 %)	35,031 (75 %)		39,521 (75 %)
Other			**<0.0001**	
Yes	4,208 (71 %)	14,178 (31 %)		18,386 (35 %)
No	1,700 (29 %)	32,297 (69 %)		33,997 (65 %)
Length of Stay (days)			**<0.0001**	
Median (Q1, Q3)	17.0 (10.0, 25.0)	9.00 (5.00, 15.0)		9.00 (5.00, 16.0)
Age (years)			**0.177**	
Median (Q1, Q3)	58.0 (46.0, 68.0)	59.0 (45.0, 69.0)		59.0 (45.0, 69.0)
CCI			**<0.0001**	
Median (Q1, Q3)	5.00 (3.00, 8.00)	4.00 (2.00, 7.00)		4.00 (2.00, 7.00)
Missing	0 (0%)	2 (0%)		2 (0%)
NHSN Central Line Days			**<0.0001**	
Median (Q1, Q3)	12.0 (6.00, 19.0)	5.00 (3.00, 9.00)		5.00 (3.00, 11.0)

All Central Line Days			<0.0001	
Median (Q1, Q3)	21.0 (12.0, 34.0)	5.00 (3.00, 9.00)		6.00 (3.00, 12.0)
% Central Line Days Accounted for by NHSN			<0.0001	
Median (Q1, Q3)	53.6 (50.0, 65.5)	100 (100, 100)		100 (100, 100)

Table 3. Comparison of patients with concurrent lines vs. patients without concurrent lines (2:1 propensity score balanced data)

	Concurrent (n=5,908)	Non-Concurrent (n=11,816)	P-value	Total (n=17,724)
Central Line-Associated Blood-stream Infection			<0.0001	
Yes	148 (3 %)	131 (1 %)		279 (2 %)
No	5,760 (97 %)	11,685 (99 %)		17,445 (98 %)
Sex			<0.0001	
Female	2,780 (47 %)	5,984 (51 %)		8,764 (49 %)
Male	3,128 (53 %)	5,832 (49 %)		8,960 (51 %)
Hemodialysis			0.987	
Yes	2,843 (48 %)	5,683 (48 %)		8,526 (48 %)
No	3,065 (52 %)	6,133 (52 %)		9,198 (52 %)
Port			<0.0001	
Yes	1,418 (24 %)	1,541 (13 %)		2,959 (17 %)
No	4,490 (76 %)	10,275 (87 %)		14,765 (83 %)
Other			<0.0001	
Yes	4,208 (71 %)	2,294 (19 %)		6,502 (37 %)
No	1,700 (29 %)	9,522 (81 %)		11,222 (63 %)

PICC			<0.0001	
Yes	2,700 (46 %)	2,753 (23 %)		5,453 (31 %)
No	3,208 (54 %)	9,063 (77 %)		12,271 (69 %)
TPN			<0.0001	
Yes	985 (17 %)	1,259 (11 %)		2,244 (13 %)
No	4,923 (83 %)	10,557 (89 %)		15,480 (87 %)
Chemotherapy			0.199	
Yes	106 (2 %)	180 (2 %)		286 (2 %)
No	5,802 (98 %)	11,636 (98 %)		17,438 (98 %)
Length of Stay (days)			<0.0001	
Median (Q1, Q3)	17.0 (10.0, 25.0)	9.00 (5.00, 15.0)		11.0 (6.00, 19.0)
Age (years)			0.003	
Median (Q1, Q3)	58.0 (46.0, 68.0)	59.0 (46.0, 69.0)		59.0 (46.0, 68.0)
CCI			<0.0001	
Median (Q1, Q3)	5.00 (3.00, 8.00)	5.00 (3.00, 7.00)		5.00 (3.00, 7.00)
NHSN Central Line Days			<0.0001	
Median (Q1, Q3)	12.0 (6.00, 19.0)	5.00 (3.00, 10.0)		7.00 (4.00, 13.0)
All Central Line Days			<0.0001	
Median (Q1, Q3)	21.0 (12.0, 34.0)	5.00 (3.00, 10.0)		8.00 (4.00, 18.0)

**% Central Line Days
Accounted for by NHSN**   **<0.0001**

Median (Q1, Q3)   53.6 (50.0, 65.5)   100 (100, 100)   100 (65.5, 100)

Table 4. Characteristics of Patients who had a multiple line at any time during their stay compared to those who never had multiple lines (2:1 propensity score balanced data)

	Yes (n=279)	No (n=17,445)	P- Value	Total (n=17,724)
Any time with overlapping lines			<0.0001	
Concurrent	148 (53 %)	5,760 (33 %)		5,908 (33 %)
Non-Concurrent	131 (47 %)	11,685 (67 %)		11,816 (67 %)
Sex			0.097	
Female	123 (44 %)	8,641 (50 %)		8,764 (49 %)
Male	156 (56 %)	8,804 (50 %)		8,960 (51 %)
Hemodialysis			0.473	
Concurrent	128 (46 %)	8,398 (48 %)		8,526 (48 %)
Non-Concurrent	151 (54 %)	9,047 (52 %)		9,198 (52 %)
Port			<0.0001	
Yes	72 (26 %)	2,887 (17 %)		2,959 (17 %)
No	207 (74 %)	14,558 (83 %)		14,765 (83 %)
Other			<0.0001	
Yes	147 (53 %)	6,355 (36 %)		6,502 (37 %)
No	132 (47 %)	11,090 (64 %)		11,222 (63 %)
PICC			0.086	
Yes	82 (29 %)	5,371 (31 %)		5,453 (31 %)
No	197 (71 %)	12,074 (69 %)		12,271 (69 %)

	Yes (n=279)	No (n=17,445)	P- Value	Total (n=17,724)
TPN			**<0.0001**	
Yes	65 (23 %)	2,179 (12 %)		2,244 (13 %)
No	214 (77 %)	15,266 (88 %)		15,480 (87 %)
Chemotherapy			**0.129**	
Yes	6 (2 %)	280 (2 %)		286 (2 %)
No	273 (98 %)	17,165 (98 %)		17,438 (98 %)
Length of Stay (days)			**<0.0001**	
Median (Q1, Q3)	24.0 (18.0, 32.0)	11.0 (6.00, 18.0)		11.0 (6.00, 19.0)
Age (years)			**0.779**	
Median (Q1, Q3)	59.0 (46.0, 68.0)	59.0 (46.0, 68.0)		59.0 (46.0, 68.0)
CCI			**0.117**	
Median (Q1, Q3)	5.00 (3.00, 8.00)	5.00 (3.00, 7.00)		5.00 (3.00, 7.00)
NHSN Central Line Days			**<0.0001**	
Median (Q1, Q3)	15.0 (9.00, 23.0)	6.00 (3.00, 13.0)		7.00 (4.00, 13.0)
All Central Line Days			**<0.0001**	
Median (Q1, Q3)	23.0 (12.0, 34.0)	8.00 (4.00, 17.0)		8.00 (4.00, 18.0)

	Yes (n=279)	No (n=17,445)	P- Value	Total (n=17,724)
% Central Line Days Accounted for by NHSN			**<0.0001**	
Median (Q1, Q3)	81.3 (50.0, 100)	100 (66.7, 100)		100 (65.5, 100)

Table 5. Crude and adjusted risk ratios for matched dataset

	Risk Ratio (95% CI)		
Variable	**Crude**	**Adjusted with NHSN**	**Adjusted with Total**
Any concurrence	2.26 (1.79, 2.85)	1.65 (1.27, 2.13)	1.18 (0.87, 1.60)
NHSN central line days	-	1.06 (1.04, 1.07)	-
Total central line days	-	-	1.03 (1.03, 1.04)
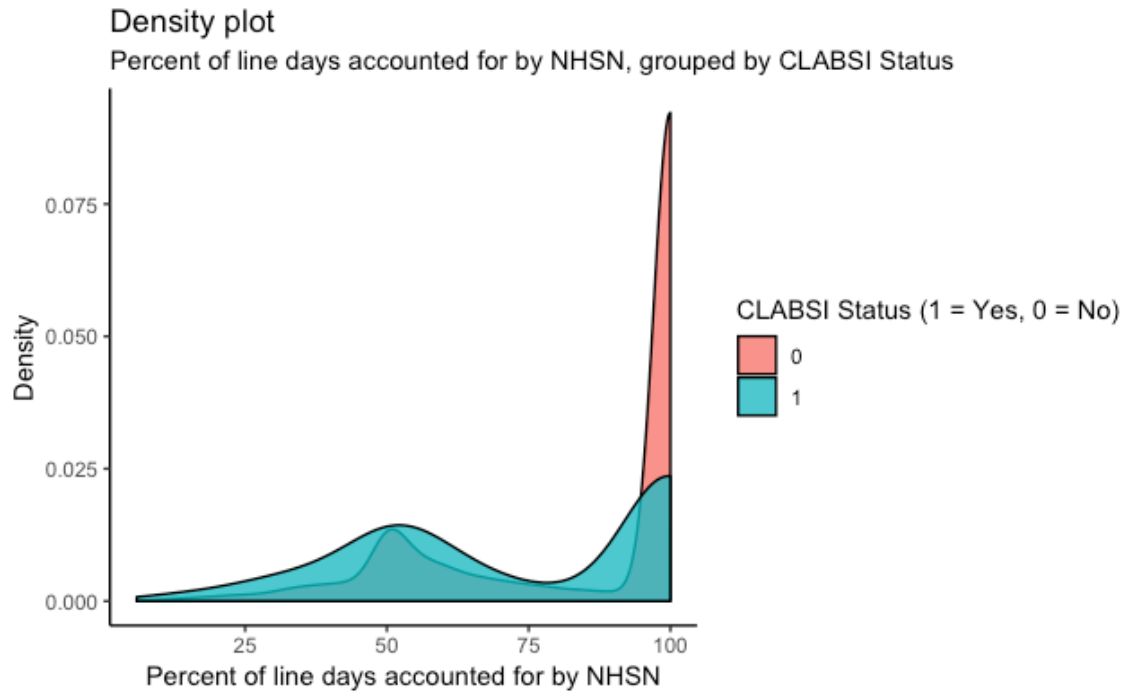
Figure 1. Percentage of total central line days accounted for by NHSN, stratified by patient disease status. For both CLABSI and non-CLABSI patients, much of the total device time is accounted for by NHSN. However, there is more spread in the CLABSI patients, with a larger proportion of these patients having less central line time accounted for by NHSN methodology.

Effect of Multiple Lines on CLABSI Risk

Data Cleaning and Preperation

Will Dube

3/13/2019

Data Recevied

We received data from Emory Healthcare's Clinical Data Warehouse and Theradoc for Emory University Hospital, Emory University Hospital Midtown, Emory John's Creek Hospital, and Emory St. Joseph's Hospital. Data included all central lines and central line associated blood stream infections (CLABSI) from January 2012 to December 2017 (IRB #). All patient names were removed, and medical record numbers and encounter numbers were modified in a standardized way in order to identify all data prior to analysis.

Description of Each Data Set

CLABSI CLABSI date, admission date, deidentified patient encounter number (a unique identifier for that particular admission), and hospital where the CLABSI occurred.

Line data Modified encounter number, line start date, line stop date, line ID, hospital, line type, line site, and number of lumens.

The modified encounter number was the same across all data sets, which allowed merging of the data sets for analysis.

Demos Modified patient encounter number, modified medical record number, birth date, age, sex, admission date, and discharge date.

CCI CCI score (both total and broken up by each component) as well as modified encounter number.

Chemo Modified encounter number for each patient who had an order for chemotherapy during that encounter.

TPN Modified encounter number for each patient who had an order for total parenteral nutrition during that encounter.

Preparing line data to be condensed to the encounter level (i.e. long to wide)

The line level dataset starts with **118,080** observations.

Dwell time and length of stay

First, I created variables for dwell time (duration = stop date - start date) and length of stay (LOS = discharge date - admission date). LOS will be modified later to be truncated at time of CLABSI for patients with CLABSIs.

Indicator variable creation

I have changed the "Dialysis/pheresis, non-tunneled" and "Dialysis/pheresis, tunneled" to "dialysis" using excel. Now I will create an indicator variable, "hd" where dialysis = 1, else = 0. Also creating indicators for presence of line types or sites of interest (port, picc, dialysis, more than 1 lumen, non-traditional sites: abdominal, trans-lumbar, and trans-abdominal.

Then, I made a variable non.trad, which is a sum of all the non-traditional line types across an encounter. This will allow us to know if any non-traditional lines were present in a patient (after we've removed those lines from our data set).

Removing non-traditional and not recorded line sites

We have decided to remove non-traditional line sites (abdominal, translumbar, transhepatic) as well as thoughs without a site recorded. **This eliminates 1,311 line obesrvations (n = 118,080 –> 116,769)**.

```
setwd("~/Desktop/CLABSI Project MPH/data cleaning")

cvcdata <- read.csv("deidentified_line_data.csv")

demos <- read.csv('demo_deid.csv')

library(tidyverse)

## ── Attaching packages ──────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.0.0     ✔ purrr   0.2.5

## ✔ tibble  1.4.2     ✔ dplyr   0.7.6

## ✔ tidyr   0.8.1     ✔ stringr 1.3.1

## ✔ readr   1.1.1     ✔ forcats 0.3.0

## ── Conflicts ─────────────────────────────────────────────── tidyverse_conflicts() ──

## ✖ dplyr::filter() masks stats::filter()
```

```
## ✖ dplyr::lag()    masks stats::lag()
```

```r
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```r
library(gmodels)
```

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
## The following object is masked from 'package:base':
##
##     date
```

```r
library(ggplot2)
# creating duration variable
cvcdata$start <- ymd(cvcdata$start)
cvcdata$stop <- ymd(cvcdata$stop)
cvcdata$duration <- (cvcdata$stop - cvcdata$start)

#changing dates in demos to date format, calculating los
demos$admit <- as.Date(demos$admit)
demos$dc <- as.Date(demos$dc)
demos$los <- (demos$dc - demos$admit)
```

```r
# Creating indicator variables

library(stringr)

cvcdata <- cvcdata %>%

  mutate(hd = ifelse(str_detect(line.type,

                      "dialysis"),1, 0)) %>%

  mutate(picc = ifelse(str_detect(line.type, 'PICC'),

                      1, 0)) %>%

  mutate(port = ifelse(str_detect(line.type, 'Port'),

                      1, 0)) %>%

  mutate(lumens = as.character(lumens)) %>%

  mutate(gt1lumen = ifelse(str_detect(lumens,'no'), 0,
1)) %>%

  mutate(abdomen = ifelse(str_detect(line.site,

                      "Abdomen"), 1, 0)) %>%

  mutate(lumbar = ifelse(str_detect(line.site,

                      "lumbar"), 1, 0)) %>%

  mutate(hepatic = ifelse(str_detect(line.site,

                      "hepatic"), 1, 0)) %>%

  mutate(not.rec = ifelse(str_detect(line.site,

                      "Recorded"), 1, 0)) %>%

  group_by(m.encounter) %>%

  mutate(not.rec.tot = sum(not.rec)) %>%

  mutate(non.trad = sum(abdomen, lumbar, hepatic))


#removing non-trad and not-recorded

cvcdata1 <- cvcdata %>%
```

```
filter(line.site != "Abdomen R") %>%

filter(line.site != "Abdomen L") %>%

filter(line.site != "Transhepatic") %>%

filter(line.site != "Translumbar") %>%

filter(line.site != "Not Recorded")
```

Overlap in same site

Next, I would like to take overlapping dates from the same encounter and the same site, and make them one line episode (keeping the earliest start, and the latest stop). I am also summing their values for the indicator variables to make sure we retain that information.

(n = 116,769 –> n = 103,983)

```
cvcdata1.1 <- cvcdata1 %>%

    arrange(m.encounter, line.site, start) %>%

    group_by(m.encounter, line.site) %>%

    mutate(indx = c(0, cumsum(as.numeric(lead(start)) >

                    cummax(as.numeric(stop)))[-n()]))%>%

    group_by(m.encounter,line.site, indx) %>%

    mutate(start = min(start),

    stop = max(stop), dial = sum(hd),

    port = sum(port), picc = sum(picc)) %>%

    distinct(m.encounter, line.site, indx, .keep_all = T)
%>%

    mutate(prev.site = indx)
```

Eliminating short duartions/patients under 18

Next, I will fix the zero duation line episodes. I will group them by line site and encounter number. Then make the start date the earliest date, and stop date the latest date. Then, I will retain only one of these observations per encounter and line site. I will also remove those that are still <2. Additionally, I will eliminate entries for patients who are under 18 years old.

```
cvcdata.zero <- cvcdata1.1 %>%

    filter(duration == '0') %>%

    group_by(m.encounter, line.site) %>%

    mutate(start = min(start),

    stop = max(stop)) %>%

    distinct(m.encounter, line.site, .keep_all = T) %>%

    mutate(duration = (stop-start)) %>%

    filter(duration >= 2)
```

```
    #Now, I will remove the observations with durations that
are less than two from the #original dataset. Then, I will
add the observations for which the duration has changed
#back in. Also adding demos and removing people under 18.
```

```
cvcdata1.2 <- cvcdata1.1 %>%

    filter(duration >= 2) %>%

    bind_rows(cvcdata.zero) %>%

    left_join(demos, by = 'm.encounter') %>%

    filter(age >17)
```

After removing durations < 2 and age < 18, there are now 68,596 line observations.

Removing duplicates

Now, I will remove duplicates based on encounter, line site, and start date (keeping the one with longest duration). There should be relatively few of these, since we have already done some adjusting of same site with overlap.

```
cvcdata1.3 <- cvcdata1.2 %>%

    arrange(m.encounter, desc(duration)) %>%

    distinct(m.encounter, line.site, start, .keep_all = T)
```

After removing duplicates based on encounter, line site, and start date, there are 68,589 line observations.

Adding variable for sequence and removing Wesley Woods patients.

After removing Wesley Woods Patients, there are **68,249** unique line episodes.

```r
cvcdata1.4 <- cvcdata1.3 %>%

  group_by(m.encounter) %>%

  arrange(start) %>%

  mutate(seq = row_number()) %>%

  ungroup %>%

  #select(-X.x, -X.y) %>%

  filter(hospital != "Wesley Woods Hospital") %>%

  filter(line.type != "Transvenous pacemaker") %>%

  mutate(m.line.type = if_else(str_detect(line.type,
c('Port')), 0, 9),

         m.line.site = if_else(str_detect(line.site,
c('Arm')), 0, 9)) %>%

  mutate(m.line.site = if_else(str_detect(line.site,
c('IJ', 'SC', 'Chest')),

                               1, m.line.site)) %>%

  mutate(m.line.site = if_else(str_detect(line.site,
c('SC')),

                               1, m.line.site)) %>%

  mutate(m.line.site = if_else(str_detect(line.site,
c('Chest')),

                               1, m.line.site)) %>%

  mutate(m.line.site = if_else(str_detect(line.site,
'Fem'),

                               2, m.line.site)) %>%
```

```
    mutate(m.line.type = if_else(str_detect(line.type,
c('PICC')),

                            1, m.line.type)) %>%

    mutate(m.line.type = if_else(str_detect(line.type,
c('MLC')), 2, m.line.type)) %>%

    mutate(m.line.type = if_else(str_detect(line.type,
c('PA')), 2, m.line.type)) %>%

    mutate(m.line.type = if_else(str_detect(line.type,
c('Transvenous')),

                                2, m.line.type)) %>%

    mutate(m.line.type = if_else(str_detect(line.type,
c('CVC')), 2, m.line.type)) %>%

    mutate(m.line.type = if_else(str_detect(line.type,
c('Introducer')), 2, m.line.type)) %>%

    mutate(m.line.type = if_else(str_detect(line.type,
c('dialysis')), 3, m.line.type)) %>%

    mutate(m.line.site = if_else(m.line.site == 9, 1,
m.line.site))

  ## Warning in stri_detect_regex(string, pattern,
opts_regex = opts(pattern)):

  ## longer object length is not a multiple of shorter
object length

  CrossTable(cvcdata1.4$line.site, cvcdata1.4$m.line.site)

  ##

  ##

  ##    Cell Contents

  ## |-------------------------|

  ## |                       N |

  ## | Chi-square contribution |

  ## |             N / Row Total |
```

```
## |                 N / Col Total |

## |                 N / Table Total |

## |-------------------------|

##

##

## Total Observations in Table:  68249

##

##

##                          | cvcdata1.4$m.line.site

## cvcdata1.4$line.site |         0 |         1 |
2 | Row Total |

## --------------------|-----------|-----------|--------
---|-----------|

##          Chest L |         0 |      3329 |
0 |      3329 |

##                          |   830.433 |   462.136 |
200.182 |           |

##                          |     0.000 |     1.000 |
0.000 |     0.049 |

##                          |     0.000 |     0.071 |
0.000 |           |

##                          |     0.000 |     0.049 |
0.000 |           |

## --------------------|-----------|-----------|--------
---|-----------|

##          Chest R |         0 |      9178 |
0 |      9178 |

##                          |  2289.491 |  1274.102 |
551.898 |           |
```

```
   ##                              |     0.000 |     1.000 |
0.000 |     0.134 |

   ##                              |     0.000 |     0.195 |
0.000 |           |

   ##                              |     0.000 |     0.134 |
0.000 |           |

   ## --------------------|-----------|-----------|--------
---|-----------|

   ##               L Arm |     5106 |        0 |
0 |       5106 |

   ##                     | 11530.400 | 3525.249 |
307.038 |           |

   ##                     |     1.000 |     0.000 |
0.000 |     0.075 |

   ##                     |     0.300 |     0.000 |
0.000 |           |

   ##                     |     0.075 |     0.000 |
0.000 |           |

   ## --------------------|-----------|-----------|--------
---|-----------|

   ##               L Fem |        0 |        0 |
1387 |       1387 |

   ##                     |   345.993 |   957.603 |
20375.038 |           |

   ##                     |     0.000 |     0.000 |
1.000 |     0.020 |

   ##                     |     0.000 |     0.000 |
0.338 |           |

   ##                     |     0.000 |     0.000 |
0.020 |           |

   ## --------------------|-----------|-----------|--------
---|-----------|
```

```
##                          LIJ |          0 |       5399 |
0 |       5399 |

##                              | 1346.803 |   749.497 |
324.657 |              |

##                              |    0.000 |      1.000 |
0.000 |      0.079 |

##                              |    0.000 |      0.115 |
0.000 |              |

##                              |    0.000 |      0.079 |
0.000 |              |

## --------------------|-----------|-----------|--------
---|-----------|

##                          LSC |          0 |       3249 |
0 |       3249 |

##                              |  810.477 |   451.031 |
195.371 |              |

##                              |    0.000 |      1.000 |
0.000 |      0.048 |

##                              |    0.000 |      0.069 |
0.000 |              |

##                              |    0.000 |      0.048 |
0.000 |              |

## --------------------|-----------|-----------|--------
---|-----------|

##                        R Arm |      11919 |          0 |
0 |      11919 |

##                              | 26915.558 |  8229.033 |
716.722 |              |

##                              |    1.000 |      0.000 |
0.000 |      0.175 |

##                              |    0.700 |      0.000 |
0.000 |              |
```

```
##                                |      0.175 |      0.000 |
0.000 |            |

##  --------------------|-----------|-----------|--------
---|-----------|

##                      R Fem |          0 |          0 |
2717 |       2717 |

##                                |    677.767 |   1875.852 |
39912.746 |            |

##                                |      0.000 |      0.000 |
1.000 |      0.040 |

##                                |      0.000 |      0.000 |
0.662 |            |

##                                |      0.000 |      0.000 |
0.040 |            |

##  --------------------|-----------|-----------|--------
---|-----------|

##                       RIJ |          0 |      19918 |
0 |      19918 |

##                                |   4968.629 |   2765.044 |
1197.724 |            |

##                                |      0.000 |      1.000 |
0.000 |      0.292 |

##                                |      0.000 |      0.423 |
0.000 |            |

##                                |      0.000 |      0.292 |
0.000 |            |

##  --------------------|-----------|-----------|--------
---|-----------|

##                       RSC |          0 |       6047 |
0 |       6047 |

##                                |   1508.450 |    839.453 |
363.623 |            |
```

```
   ##                                   |       0.000 |      1.000 |
0.000 |      0.089 |

   ##                                   |       0.000 |      0.128 |
0.000 |           |

   ##                                   |       0.000 |      0.089 |
0.000 |           |

   ## --------------------|-----------|-----------|--------
---|-----------|

   ##            Column Total |      17025 |      47120 |
4104 |      68249 |

   ##                                   |       0.249 |      0.690 |
0.060 |           |

   ## --------------------|-----------|-----------|--------
---|-----------|

   ##

   ##

CrossTable(cvcdata1.4$line.type, cvcdata1.4$m.line.type)

   ##

   ##

   ##    Cell Contents

   ## |-------------------------|

   ## |                       N |

   ## | Chi-square contribution |

   ## |           N / Row Total |

   ## |           N / Col Total |

   ## |         N / Table Total |

   ## |-------------------------|

   ##
```

```
   ##

   ## Total Observations in Table:  68249

   ##

   ##

   ##                                | cvcdata1.4$m.line.type

   ## cvcdata1.4$line.type |          0 |          1 |
2 |          3 | Row Total |

   ## ---------------------|-----------|-----------|--------
---|-----------|-----------|

   ##              dialysis |          0 |          0 |
0 |     10989 |     10989 |

   ##                                |  2125.215 |  3520.879 |
3573.530 | 48040.376 |           |

   ##                                |     0.000 |     0.000 |
0.000 |     1.000 |     0.161 |

   ##                                |     0.000 |     0.000 |
0.000 |     1.000 |           |

   ##                                |     0.000 |     0.000 |
0.000 |     0.161 |           |

   ## ---------------------|-----------|-----------|--------
---|-----------|-----------|

   ##           Double Port |      2247 |          0 |
0 |          0 |      2247 |

   ##                                |  7559.279 |   719.939 |
730.705 |   361.797 |           |

   ##                                |     1.000 |     0.000 |
0.000 |     0.000 |     0.033 |

   ##                                |     0.170 |     0.000 |
0.000 |     0.000 |           |

   ##                                |     0.033 |     0.000 |
0.000 |     0.000 |           |
```

```
## ---------------------|-----------|-----------|--------
---|-----------|-----------|

##        Hypothermia CVC |         0 |         0 |
73 |         0 |        73 |

##                       |    14.118 |    23.389 |
102.222 |    11.754 |           |

##                       |     0.000 |     0.000 |
1.000 |     0.000 |     0.001 |

##                       |     0.000 |     0.000 |
0.003 |     0.000 |           |

##                       |     0.000 |     0.000 |
0.001 |     0.000 |           |

## ---------------------|-----------|-----------|--------
---|-----------|-----------|

##             Introducer |         0 |         0 |
1807 |         0 |      1807 |

##                       |   349.464 |   578.963 |
2530.346 |   290.951 |           |

##                       |     0.000 |     0.000 |
1.000 |     0.000 |     0.026 |

##                       |     0.000 |     0.000 |
0.081 |     0.000 |           |

##                       |     0.000 |     0.000 |
0.026 |     0.000 |           |

## ---------------------|-----------|-----------|--------
---|-----------|-----------|

##                   MLC |         0 |         0 |
12470 |         0 |     12470 |

##                       |  2411.633 |  3995.392 |
17461.766 |  2007.836 |           |

##                       |     0.000 |     0.000 |
1.000 |     0.000 |     0.183 |
```

```
##                              |      0.000 |      0.000 |
0.562 |      0.000 |              |

##                              |      0.000 |      0.000 |
0.183 |      0.000 |              |

## --------------------|-----------|-----------|--------
---|-----------|-----------|

##            PA catheter |          0 |          0 |
3701 |          0 |       3701 |

##                              |    715.754 |   1185.802 |
5182.518 |    595.910 |              |

##                              |      0.000 |      0.000 |
1.000 |      0.000 |      0.054 |

##                              |      0.000 |      0.000 |
0.167 |      0.000 |              |

##                              |      0.000 |      0.000 |
0.054 |      0.000 |              |

## --------------------|-----------|-----------|--------
---|-----------|-----------|

##                    PICC |          0 |      17534 |
0 |          0 |      17534 |

##                              |   3390.984 |  25275.191 |
5701.909 |   2823.208 |              |

##                              |      0.000 |      1.000 |
0.000 |      0.000 |      0.257 |

##                              |      0.000 |      0.802 |
0.000 |      0.000 |              |

##                              |      0.000 |      0.257 |
0.000 |      0.000 |              |

## --------------------|-----------|-----------|--------
---|-----------|-----------|

##        PICC, tunneled |          0 |       4333 |
0 |          0 |       4333 |
```

```
##                                | 837.980 |  6246.002 |
1409.055 |   697.671 |           |

##                                |   0.000 |     1.000 |
0.000 |     0.000 |     0.063 |

##                                |   0.000 |     0.198 |
0.000 |     0.000 |           |

##                                |   0.000 |     0.063 |
0.000 |     0.000 |           |

## --------------------|-----------|-----------|--------
---|-----------|-----------|

##          Single Port |     10952 |         0 |
0 |         0 |     10952 |

##                                | 36844.338 |  3509.024 |
3561.498 |  1763.418 |           |

##                                |   1.000 |     0.000 |
0.000 |     0.000 |     0.160 |

##                                |   0.830 |     0.000 |
0.000 |     0.000 |           |

##                                |   0.160 |     0.000 |
0.000 |     0.000 |           |

## --------------------|-----------|-----------|--------
---|-----------|-----------|

##          Tunneled CVC |         0 |         0 |
4143 |         0 |      4143 |

##                                | 801.235 |  1327.418 |
5801.451 |   667.078 |           |

##                                |   0.000 |     0.000 |
1.000 |     0.000 |     0.061 |

##                                |   0.000 |     0.000 |
0.187 |     0.000 |           |

##                                |   0.000 |     0.000 |
0.061 |     0.000 |           |
```

```
   ## --------------------|-----------|-----------|--------
---|-----------|-----------|

   ##          Column Total |     13199 |     21867 |
22194 |     10989 |     68249 |

   ##                       |     0.193 |     0.320 |
0.325 |     0.161 |           |

   ## --------------------|-----------|-----------|--------
---|-----------|-----------|

   ##

   ##
```
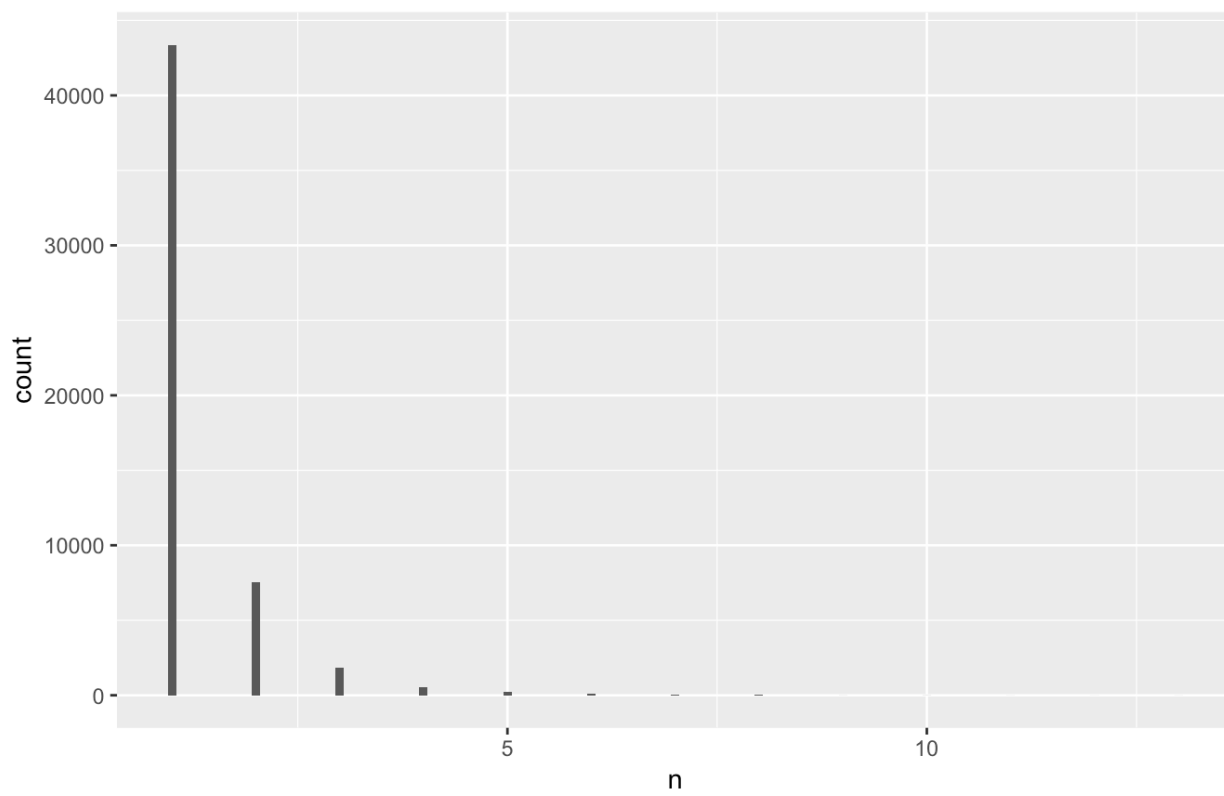
Rexamining data to determine cutoffs for LOS and number of lines per encounter

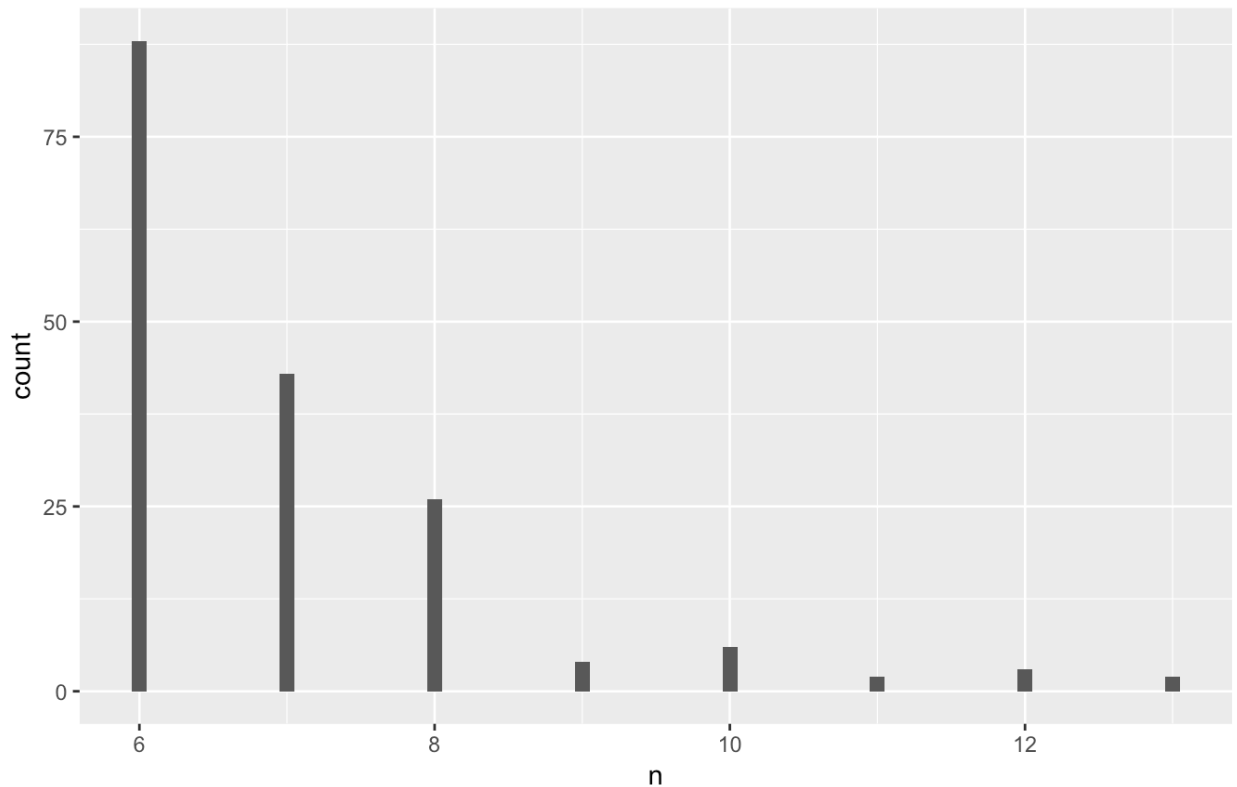Number of Lines per Encounter

Lines per encounter

Number of unique line episodes per encounter

Lines per encounter n>5
Number of unique line episodes per encounter

Given that very few encounters have more than 5 lines, theses are likely not relevant for these analyses. As such, I have dicided to restrict to 5 or fewer

By hospital statistics at line level

Histograms for LOS and quantile information

```
## List of 1
##  $ legend.position: num [1:2] 0.1 0.4
##  - attr(*, "class")= chr [1:2] "theme" "gg"
##  - attr(*, "complete")= logi FALSE
##  - attr(*, "validate")= logi TRUE
## Don't know how to automatically pick scale for object
of type difftime. Defaulting to continuous.
## Warning: Removed 10 rows containing non-finite values
(stat_bin).
```
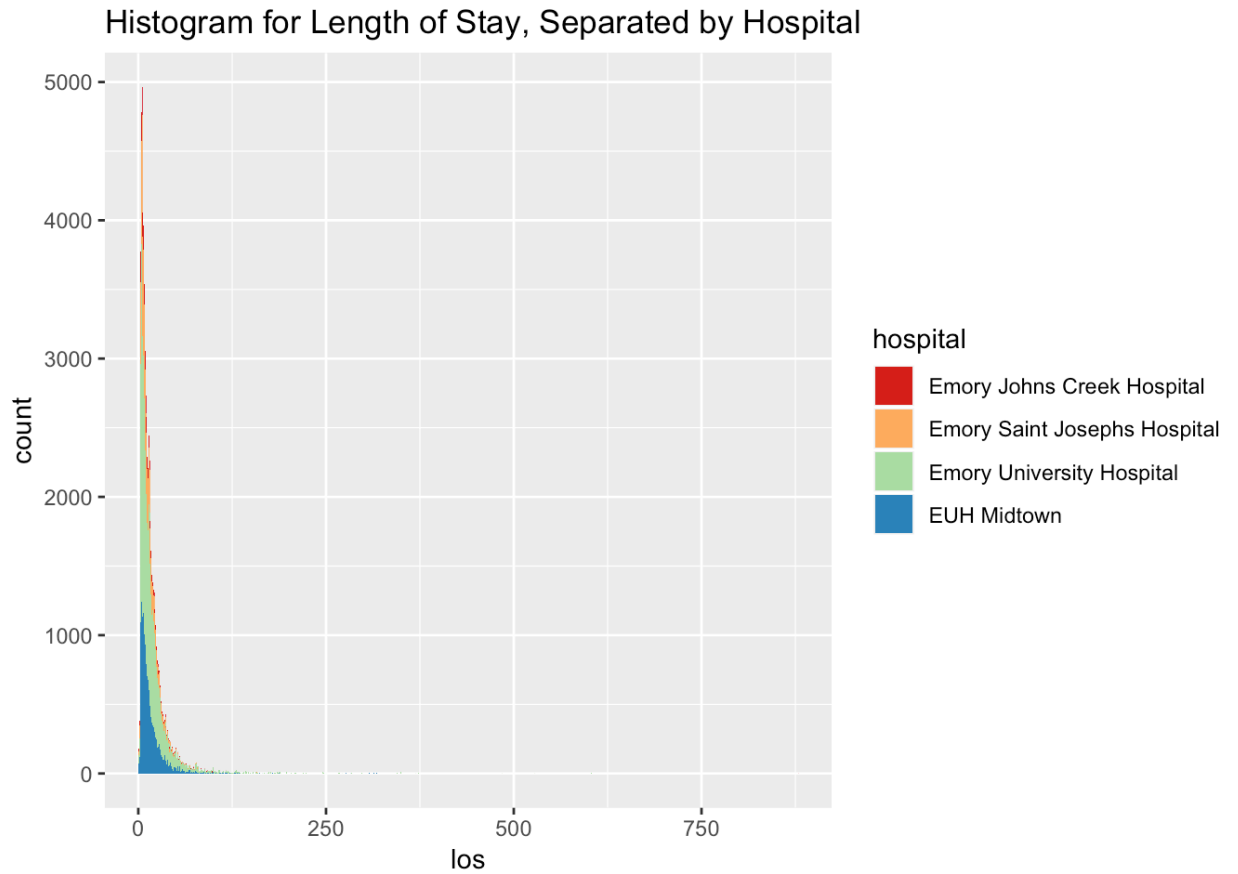
Histogram for Length of Stay, Separated by Hospital


```
## List of 1

##  $ legend.position: num [1:2] 0.1 0.4

##  - attr(*, "class")= chr [1:2] "theme" "gg"

##  - attr(*, "complete")= logi FALSE

##  - attr(*, "validate")= logi TRUE

## Don't know how to automatically pick scale for object
of type difftime. Defaulting to continuous.
```

Histogram for Length of Stay, Separated by Hospital
Where LOS is less than 150 days
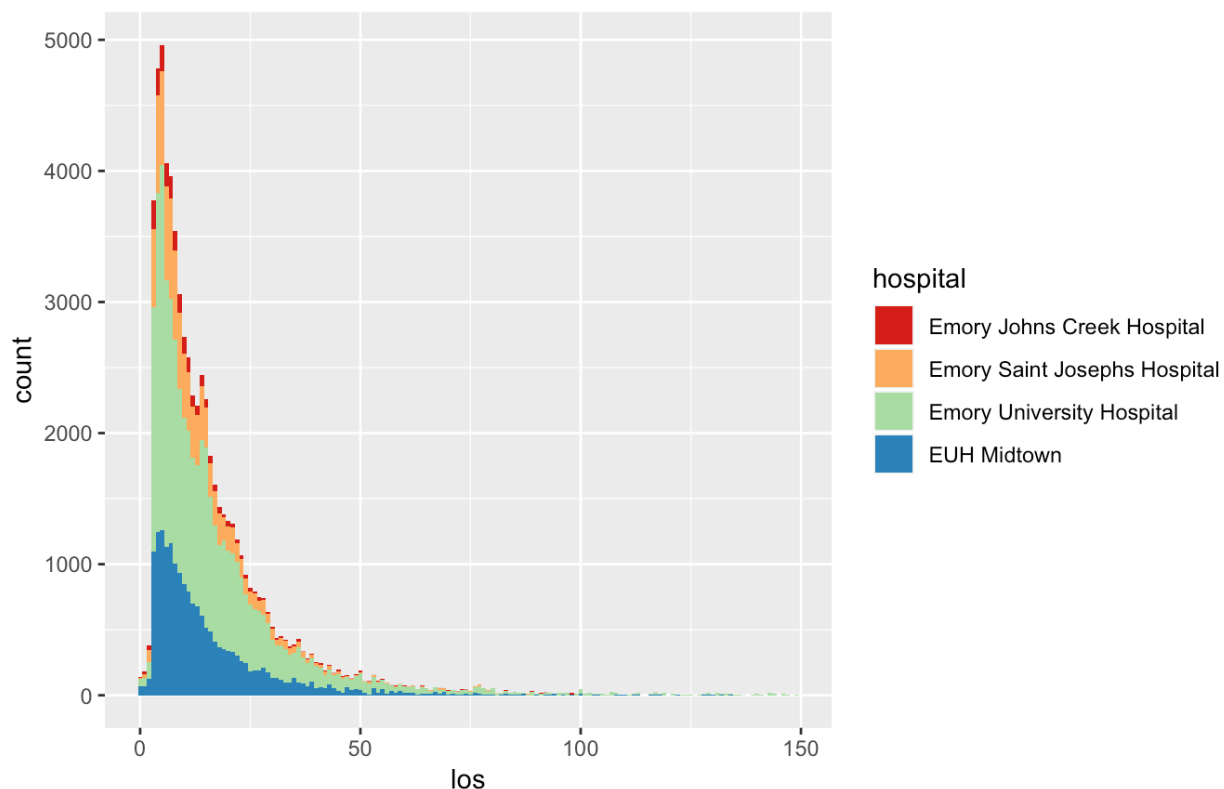

```
## List of 1

##  $ legend.position: num [1:2] 0.1 0.4

##  - attr(*, "class")= chr [1:2] "theme" "gg"

##  - attr(*, "complete")= logi FALSE

##  - attr(*, "validate")= logi TRUE

## Don't know how to automatically pick scale for object
of type difftime. Defaulting to continuous.
```

Since the 95th percentile for LOS is 49 days, we decided to elminate everything above 49 days. This reduces the data to 64,867 line observations

Converting data from long to wide (going up to encounter level)

Bringing in clabsi data and clabsi date which has been modified in excel so that it matches the removal date of the line with the closest date.

Now, I will eliminate line observations which started after the CLABSI date. **This reduces the data to 64,575 line observations.**

```r
library(lubridate)

line_level_clabsi <- line_level %>%

  left_join(m.clabsi.date, by = "m.encounter") %>%

  filter(m.clabsi.date != is.na(m.clabsi.date))


line_level1 <- line_level_clabsi%>%

  filter(start < m.clabsi.date)


line_level2 <- line_level_clabsi %>%

  anti_join(line_level1, by = "line.id")


line_level3 <- line_level %>%

  anti_join(line_level2, by = "line.id") %>%

   select(-c(indx, insertion.info, line.id, abdomen,

          lumbar, hepatic, not.rec, non.trad)) %>%

  mutate(hd = as.integer(hd), duration =
as.integer(duration),

        picc = as.integer(picc), port =
as.integer(port),

        gt1lumen = as.integer(gt1lumen), not.rec.tot =
as.integer(not.rec.tot),

        prev.site = as.integer(prev.site), birth =
as.Date(birth),

        admit = as.Date(admit, origin = "1970-01-01"),

        dc = as.Date(dc, origin = "1970-01-01"), los =
as.integer(los),
```

```
        m.line.type = as.integer(m.line.type),
m.line.site = as.integer(m.line.site)) %>%

    select(-c(X.x, X.y))
```

Now, the data are ready to be switched to the encounter level (long to wide), this takes our **64,575** line observations and organizes them together on **53,275** encounter level rows. In these data there are **647 CLABSI patients.**

```
    # now going from long to wide

    encounter_level <-

        dcast(melt(line_level3, id.vars=c("m.encounter",
"seq"), factorsAsStrings = FALSE),

            m.encounter~variable+seq)

    ## Warning: attributes are not identical across measure
variables; they will

    ## be dropped

    ## Warning: cannot avoid coercion of factors when measure
attributes not

    ## identical

    # now I can join the two

    encounter_level <- m.clabsi.date %>%

        left_join(encounter_level, by = 'm.encounter') %>%

        select(-c(hospital_2, hospital_3, hospital_4,
hospital_5,

                birth_2, birth_3, birth_4, birth_5,

                los_2, los_3, los_4, los_5,

                age_2, age_3, age_4, age_5,

                admit_2, admit_3, admit_4, admit_5,

                dc_2, dc_3, dc_4, dc_5,

                sex_2, sex_3, sex_4, sex_5,
```

```
            dial_1, dial_2, dial_3, dial_4, dial_5,

            not.rec.tot_2, not.rec.tot_3, not.rec.tot_4,
not.rec.tot_5)) %>%

            mutate(m.clabsi.date = as.Date(m.clabsi.date))
```

```
  summary(encounter_level$m.clabsi.date)

  ##          Min.      1st Qu.        Median         Mean
3rd Qu.

  ## "2012-11-26" "2014-07-22" "2015-12-02" "2015-10-07"
"2016-12-09"

  ##          Max.         NA's

  ## "2017-12-31"      "52515"
```

```
#glimpse(encounter_level)
```

CLABSI line durations

Now, I am adjusting line durations to be truncated at date of CLABSI

```
clabsi_enc <- encounter_level %>%

   filter(m.clabsi.date != is.na(m.clabsi.date)) %>%

   mutate(m.clabsi.date =
as.Date(as.integer(m.clabsi.date), origin = '1970-1-1'),

          start_1 = as.Date(as.integer(start_1), origin =
'1970-1-1'),

          start_2 = as.Date(as.integer(start_2), origin =
'1970-1-1'),

          start_3 = as.Date(as.integer(start_3), origin =
'1970-1-1'),

          start_4 = as.Date(as.integer(start_4), origin =
'1970-1-1'),

          start_5 = as.Date(as.integer(start_5), origin =
'1970-1-1'),
```

```r
          stop_1 = as.Date(as.integer(stop_1), origin =
'1970-1-1'),

          stop_2 = as.Date(as.integer(stop_2), origin =
'1970-1-1'),

          stop_3 = as.Date(as.integer(stop_3), origin =
'1970-1-1'),

          stop_4 = as.Date(as.integer(stop_4), origin =
'1970-1-1'),

          stop_5 = as.Date(as.integer(stop_5), origin =
'1970-1-1'),


          duration_1 = as.integer(duration_1),

          duration_2 = as.integer(duration_2),

          duration_3 = as.integer(duration_3),

          duration_4 = as.integer(duration_4),

          duration_5 = as.integer(duration_5))


  #### recalculating durations for clabsi patients ####

  clabsi_enc <- clabsi_enc %>%

    mutate(duration_1 = (m.clabsi.date-start_1),

          duration_2 = (m.clabsi.date-start_2),

          duration_3 = (m.clabsi.date-start_3),

          duration_4 = (m.clabsi.date-start_4),

          duration_5 = (m.clabsi.date-start_5))

  #glimpse(clabsi_enc)


  #### eliminating negative line durations
```

```r
clabsi_enc <- clabsi_enc %>%

  mutate(duration_1 = ifelse(duration_1 > 2, duration_1,
NA),

          duration_2 = ifelse(duration_2 > 2, duration_2,
NA),

          duration_3 = ifelse(duration_3 > 2, duration_3,
NA),

          duration_4 = ifelse(duration_4 > 2, duration_4,
NA),

          duration_5 = ifelse(duration_5 > 2, duration_5,
NA))


  #### loading in and changing date class for non-clabsi
people

  non.clabsi <- encounter_level %>%

    anti_join(clabsi_enc, by = "m.encounter") %>%

    mutate(start_1 = as.Date(as.integer(start_1), origin =
'1970-1-1'),

          start_2 = as.Date(as.integer(start_2), origin =
'1970-1-1'),

          start_3 = as.Date(as.integer(start_3), origin =
'1970-1-1'),

          start_4 = as.Date(as.integer(start_4), origin =
'1970-1-1'),

          start_5 = as.Date(as.integer(start_5), origin =
'1970-1-1'),

          stop_1 = as.Date(as.integer(stop_1), origin =
'1970-1-1'),

          stop_2 = as.Date(as.integer(stop_2), origin =
'1970-1-1'),

          stop_3 = as.Date(as.integer(stop_3), origin =
'1970-1-1'),
```

```
        stop_4 = as.Date(as.integer(stop_4), origin =
'1970-1-1'),

        stop_5 = as.Date(as.integer(stop_5), origin =
'1970-1-1'),

        duration_1 = as.integer(duration_1),

        duration_2 = as.integer(duration_2),

        duration_3 = as.integer(duration_3),

        duration_4 = as.integer(duration_4),

        duration_5 = as.integer(duration_5))


enc <- non.clabsi %>%

  bind_rows(clabsi_enc)


#glimpse(enc)
```

How many lines are in each of the type and sites, for lines 1, 2, 3, 4, and 5

For m.line.site:

0 = Arm

1 = Internal jugular, subclavian, or chest

2 = Femoral

For m.line.type:

0 = Port

1 = PICC

2 = MLC, PA, Tunneled CVC, and Introducer

3 = Dialysis

```
library(gmodels)

CrossTable(enc$m.line.type_1, enc$m.line.site_1)
```

```
## 

## 

##      Cell Contents

## |-------------------------|

## |                       N |

## | Chi-square contribution |

## |             N / Row Total |

## |             N / Col Total |

## |           N / Table Total |

## |-------------------------|

## 

## 

## Total Observations in Table:  52383

## 

## 

##                          | enc$m.line.site_1

## enc$m.line.type_1 |          0 |          1 |          2 | Row Total |

## ------------------|-----------|-----------|-----------|-----------|

##                 0 |       648 |     11707 |         0 |     12355 |

##                   |  1950.180 |  1077.134 |   589.883 |           |

##                   |     0.052 |     0.948 |     0.000 |     0.236 |

##                   |     0.049 |     0.319 |     0.000 |           |
```

```
##                            |      0.012 |      0.223 |      0.000
|             |

##  -----------------|----------|----------|-----------
|----------|

##                     1 |      12462 |       4216 |        120
|      16798 |

##                       | 16019.948 |  4844.630 |    579.967
|             |

##                       |      0.742 |      0.251 |      0.007
|      0.321 |

##                       |      0.945 |      0.115 |      0.048
|             |

##                       |      0.238 |      0.080 |      0.002
|             |

##  -----------------|----------|----------|-----------
|----------|

##                     2 |         77 |      15101 |       1078
|      16256 |

##                       |  3941.007 |  1211.890 |    117.406
|             |

##                       |      0.005 |      0.929 |      0.066
|      0.310 |

##                       |      0.006 |      0.412 |      0.431
|             |

##                       |      0.001 |      0.288 |      0.021
|             |

##  -----------------|----------|----------|-----------
|----------|

##                     3 |          4 |       5667 |       1303
|       6974 |

##                       |  1748.190 |   125.236 |   2825.953
|             |
```

```
## |         0.001 |     0.813 |     0.187
|    0.133 |

## |         0.000 |     0.154 |     0.521
|          |

## |         0.000 |     0.108 |     0.025
|          |

## ------------------|-----------|-----------|-----------
|-----------|

##    Column Total |    13191 |    36691 |    2501
|    52383 |

## |         0.252 |     0.700 |     0.048
|          |

## ------------------|-----------|-----------|-----------
|-----------|

##

##

CrossTable(enc$m.line.type_2, enc$m.line.site_2)

##

##

##   Cell Contents

## |-------------------------|

## |                      N |

## | Chi-square contribution |

## |             N / Row Total |

## |             N / Col Total |

## |           N / Table Total |

## |-------------------------|

##
```

```
##

## Total Observations in Table:  9211

##

##

##                                | enc$m.line.site_2

## enc$m.line.type_2 |          0 |          1 |          2 | Row Total |

## ------------------|-----------|-----------|-----------|-----------|

##                0 |         30 |        616 |          0 |       646 |

##                  |    106.887 |     90.869 |     64.172 |           |

##                  |      0.046 |      0.954 |      0.000 |     0.070 |

##                  |      0.013 |      0.103 |      0.000 |           |

##                  |      0.003 |      0.067 |      0.000 |           |

## ------------------|-----------|-----------|-----------|-----------|

##                1 |       2257 |        716 |         25 |      2998 |

##                  |   3039.328 |    782.268 |    249.913 |           |

##                  |      0.753 |      0.239 |      0.008 |     0.325 |

##                  |      0.981 |      0.119 |      0.027 |           |

##                  |      0.245 |      0.078 |      0.003 |           |
```

```
## -----------------|----------|----------|----------|----------|

##                2 |       12 |     2940 |      335 |     3287 |

##                  |  796.944 |  299.324 |    0.220 |          |

##                  |    0.004 |    0.894 |    0.102 |    0.357 |

##                  |    0.005 |    0.490 |    0.366 |          |

##                  |    0.001 |    0.319 |    0.036 |          |

## -----------------|----------|----------|----------|----------|

##                3 |        1 |     1724 |      555 |     2280 |

##                  |  567.321 |   38.747 |  476.483 |          |

##                  |    0.000 |    0.756 |    0.243 |    0.248 |

##                  |    0.000 |    0.288 |    0.607 |          |

##                  |    0.000 |    0.187 |    0.060 |          |

## -----------------|----------|----------|----------|----------|

##     Column Total |     2300 |     5996 |      915 |     9211 |

##                  |    0.250 |    0.651 |    0.099 |          |

## -----------------|----------|----------|----------|----------|

##
```

```
##

CrossTable(enc$m.line.type_3, enc$m.line.site_3)

##

##

##    Cell Contents

## |-------------------------|

## |                       N |

## | Chi-square contribution |

## |             N / Row Total |

## |             N / Col Total |

## |           N / Table Total |

## |-------------------------|

##

##

## Total Observations in Table:  2037

##

##

##                       | enc$m.line.site_3

## enc$m.line.type_3 |          0 |          1 |          2 | Row Total |

## -----------------|-----------|-----------|-----------|-----------|

##                 0 |        9 |       66 |        0 |       75 |

##                   |    3.799 |    5.541 |    8.505 |          |

##                   |    0.120 |    0.880 |    0.000 |    0.037 |
```

```
##                        |     0.019 |     0.049 |     0.000 |           |

##                        |     0.004 |     0.032 |     0.000 |           |

## -----------------|----------|----------|-----------|-----------|

##                1 |       449 |       133 |         5 |       587 |

##                  |   746.423 |   166.718 |    56.943 |           |

##                  |     0.765 |     0.227 |     0.009 |     0.288 |

##                  |     0.970 |     0.099 |     0.022 |           |

##                  |     0.220 |     0.065 |     0.002 |           |

## -----------------|----------|----------|-----------|-----------|

##                2 |         5 |       665 |        84 |       754 |

##                  |   161.526 |    56.698 |     0.026 |           |

##                  |     0.007 |     0.882 |     0.111 |     0.370 |

##                  |     0.011 |     0.495 |     0.364 |           |

##                  |     0.002 |     0.326 |     0.041 |           |

## -----------------|----------|----------|-----------|-----------|

##                3 |         0 |       479 |       142 |       621 |
```

```
##                             |    141.150 |     11.822 |     72.751
|            |

##                             |      0.000 |      0.771 |      0.229
|      0.305 |

##                             |      0.000 |      0.357 |      0.615
|            |

##                             |      0.000 |      0.235 |      0.070
|            |

## ------------------|-----------|-----------|-----------
|-----------|

##       Column Total |        463 |       1343 |        231
|       2037 |

##                             |      0.227 |      0.659 |      0.113
|            |

## ------------------|-----------|-----------|-----------
|-----------|

##

##

CrossTable(enc$m.line.type_4, enc$m.line.site_4)

##

##

##     Cell Contents

## |-------------------------|

## |                       N |

## | Chi-square contribution |

## |           N / Row Total |

## |           N / Col Total |

## |         N / Table Total |

## |-------------------------|
```

```
##

##

## Total Observations in Table:  491

##

##

##                            | enc$m.line.site_4

## enc$m.line.type_4 |          0 |          1 |          2 | Row Total |

## ------------------|-----------|-----------|-----------|-----------|

##                 0 |          0 |         13 |          0 |        13 |

##                   |      2.145 |      1.609 |      1.695 |           |

##                   |      0.000 |      1.000 |      0.000 |      0.026 |

##                   |      0.000 |      0.038 |      0.000 |           |

##                   |      0.000 |      0.026 |      0.000 |           |

## ------------------|-----------|-----------|-----------|-----------|

##                 1 |         81 |         32 |          1 |        114 |

##                   |    205.675 |     29.081 |     12.927 |           |

##                   |      0.711 |      0.281 |      0.009 |      0.232 |

##                   |      1.000 |      0.092 |      0.016 |           |
```

```
##                            |     0.165 |     0.065 |     0.002 |           |

## -----------------|-----------|-----------|-----------|-----------|

##                     2 |         0 |       144 |        15 |       159 |

##                            |    26.230 |     9.114 |     1.581 |           |

##                            |     0.000 |     0.906 |     0.094 |     0.324 |

##                            |     0.000 |     0.416 |     0.234 |           |

##                            |     0.000 |     0.293 |     0.031 |           |

## -----------------|-----------|-----------|-----------|-----------|

##                     3 |         0 |       157 |        48 |       205 |

##                            |    33.819 |     1.088 |    16.945 |           |

##                            |     0.000 |     0.766 |     0.234 |     0.418 |

##                            |     0.000 |     0.454 |     0.750 |           |

##                            |     0.000 |     0.320 |     0.098 |           |

## -----------------|-----------|-----------|-----------|-----------|

##      Column Total |        81 |       346 |        64 |       491 |

##                            |     0.165 |     0.705 |     0.130 |           |
```

```
## ------------------|----------|----------|----------
|----------|

##

##

CrossTable(enc$m.line.type_5, enc$m.line.site_5)

##

##

##     Cell Contents

## |-------------------------|

## |                       N |

## | Chi-square contribution |

## |             N / Row Total |

## |             N / Col Total |

## |           N / Table Total |

## |-------------------------|

##

##

## Total Observations in Table:  110

##

##

##                    | enc$m.line.site_5

## enc$m.line.type_5 |          0 |          1 |          2
| Row Total |

## ------------------|----------|----------|----------
|----------|

##               0 |          1 |          1 |          0
|          2 |
```

```
##                              |       2.183 |       0.193 |       0.200 
|           |

##                              |       0.500 |       0.500 |       0.000 
|     0.018 |

##                              |       0.071 |       0.012 |       0.000 
|           |

##                              |       0.009 |       0.009 |       0.000 
|           |

## -----------------|----------|----------|-----------|----------|

##                           1 |         12 |         13 |          0 
|        25 |

##                              |      24.439 |       2.066 |       2.500 
|           |

##                              |       0.480 |       0.520 |       0.000 
|     0.227 |

##                              |       0.857 |       0.153 |       0.000 
|           |

##                              |       0.109 |       0.118 |       0.000 
|           |

## -----------------|----------|----------|-----------|----------|

##                           2 |          0 |         33 |          7 
|        40 |

##                              |       5.091 |       0.141 |       2.250 
|           |

##                              |       0.000 |       0.825 |       0.175 
|     0.364 |

##                              |       0.000 |       0.388 |       0.636 
|           |

##                              |       0.000 |       0.300 |       0.064 
|           |
```

```
   ## ------------------|----------|----------|----------
|----------|

   ##                 3 |        1 |       38 |        4
|        43 |

   ##                   |    3.655 |    0.686 |    0.021
|           |

   ##                   |    0.023 |    0.884 |    0.093
|    0.391 |

   ##                   |    0.071 |    0.447 |    0.364
|           |

   ##                   |    0.009 |    0.345 |    0.036
|           |

   ## ------------------|----------|----------|----------
|----------|

   ##     Column Total |       14 |       85 |       11
|       110 |

   ##                   |    0.127 |    0.773 |    0.100
|           |

   ## ------------------|----------|----------|----------
|----------|

   ##

   ##
```

Lastly, I am going to make some changes to variable formats

```
enc <- enc %>%

  mutate(clabsiyn = if_else(as.integer(m.clabsi.date) >
0, 1, 0),

          hd_1 = as.integer(hd_1),

          hd_2 = as.integer(hd_2),

          hd_3 = as.integer(hd_3),

          hd_4 = as.integer(hd_4),
```

```r
        hd_5 = as.integer(hd_5),


        picc_1 = as.integer(picc_1),

        picc_2 = as.integer(picc_2),

        picc_3 = as.integer(picc_3),

        picc_4 = as.integer(picc_4),

        picc_5 = as.integer(picc_5),


        port_1 = as.integer(port_1),

        port_2 = as.integer(port_2),

        port_3 = as.integer(port_3),

        port_4 = as.integer(port_4),

        port_5 = as.integer(port_5),


        gt1lumen_1 = as.integer(gt1lumen_1),

        gt1lumen_2 = as.integer(gt1lumen_2),

        gt1lumen_3 = as.integer(gt1lumen_3),

        gt1lumen_4 = as.integer(gt1lumen_4),

        gt1lumen_5 = as.integer(gt1lumen_5),


        not.rec.tot_1 = as.integer(not.rec.tot_1),


        prev.site_1 = as.integer(prev.site_1),

        prev.site_2 = as.integer(prev.site_2),

        prev.site_3 = as.integer(prev.site_3),
```

```r
          prev.site_4 = as.integer(prev.site_4),

          prev.site_5 = as.integer(prev.site_5),


          birth_1 = as.Date(as.integer(birth_1), origin =
"1970-01-01"),

          admit_1 = as.Date(as.integer(admit_1), origin =
"1970-01-01"),

          dc_1 = as.Date(as.integer(dc_1), origin = "1970-
01-01"),

          age_1 = as.integer(age_1),

          los_1 = as.integer(los_1),


          m.line.type_1 = as.integer(m.line.type_1),

          m.line.type_2 = as.integer(m.line.type_2),

          m.line.type_3 = as.integer(m.line.type_3),

          m.line.type_4 = as.integer(m.line.type_4),

          m.line.type_5 = as.integer(m.line.type_5),


          m.line.site_1 = as.integer(m.line.site_1),

          m.line.site_2 = as.integer(m.line.site_2),

          m.line.site_3 = as.integer(m.line.site_3),

          m.line.site_4 = as.integer(m.line.site_4),

          m.line.site_5 = as.integer(m.line.site_5),


          duration_1 = (stop_1) - (start_1),

          duration_2 = (stop_2) - (start_2),

          duration_3 = (stop_3) - (start_3),
```

```
            duration_4 = (stop_4) - (start_4),

            duration_5 = (stop_5) - (start_5)) %>%

    mutate(hd.tot = ifelse(hd_1 | hd_2 | hd_3 | hd_4 | hd_5
== 1, 1, 0))
```

```
    enc$hd.tot[is.na(enc$hd.tot)] <- 0

    enc$clabsiyn[is.na(enc$clabsiyn)] <- 0

    #glimpse(enc)
```

Overlap variables

Now, I will create the overlap variables. First, I will create variables looking at each possible overlap combination (1 and 2, 1 and 3, ..., 4 and 5).

```
    #### show overlap between two lines

    enc_overlap <- enc %>%

      mutate(o12 = ifelse((stop_1) - (start_2) >= 1,

                          (stop_1) - (start_2), NA)) %>%

      mutate(o13 = ifelse((stop_1) - (start_3) >= 1,

                          (stop_1) - (start_3), NA)) %>%

      mutate(o14 = ifelse((stop_1) - (start_4) >= 1,

                          (stop_1) - (start_4), NA)) %>%

      mutate(o15 = ifelse((stop_1) - (start_5) >= 1,

                          (stop_1) - (start_5), NA)) %>%

      mutate(o23 = ifelse((stop_2) - (start_3) >= 1,

                          (stop_2) - (start_3), NA)) %>%

      mutate(o24 = ifelse((stop_2) - (start_4) >= 1,

                          (stop_2) - (start_4), NA)) %>%

      mutate(o25 = ifelse((stop_2) - (start_5) >= 1,
```

```
                              (stop_2) - (start_5), NA)) %>%

        mutate(o34 = ifelse((stop_3) - (start_4) >= 1,

                              (stop_3) - (start_4), NA)) %>%

        mutate(o35 = ifelse((stop_3) - (start_5) >= 1,

                              (stop_3) - (start_5), NA)) %>%

        mutate(o45 = ifelse((stop_4) - (start_5) >= 1,

                              (stop_4) - (start_5), NA))
```

Now I will create all of the variables I need to calculate NHSN central line days, this means that I will need to make it so that the overlapping periods only count as one line when I add all the lines up.

```
    # creating m.stop, trying to only count one line per day
for mult lines

    enc_overlap <- enc_overlap %>%

      mutate(o12.1 = ifelse(is.na(o12), 0, o12),

             o13.1 = ifelse(is.na(o13), 0, o13),

             o14.1 = ifelse(is.na(o14), 0, o14),

             o15.1 = ifelse(is.na(o15), 0, o15),

             o23.1 = ifelse(is.na(o23), 0, o23),

             o24.1 = ifelse(is.na(o24), 0, o24),

             o25.1 = ifelse(is.na(o25), 0, o25),

             o34.1 = ifelse(is.na(o34), 0, o34),

             o35.1 = ifelse(is.na(o35), 0, o35),

             o45.1 = ifelse(is.na(o45), 0, o45),


             m.stop_4 = as.Date(ifelse(o45.1 > 0, stop_5,
stop_4),

                                 origin = '1970-1-1'),
```

```r
                m.stop_3 = as.Date(ifelse(o34.1 > 0, m.stop_4,
stop_3),

                                    origin = '1970-1-1'),

              m.stop_3 = as.Date(ifelse(o35.1 > 0 & stop_5 >
m.stop_3, stop_5, m.stop_3),

                                    origin = '1970-1-1'),

              m.stop_2 = as.Date(ifelse(o23.1 > 0, m.stop_3,
stop_2), origin = '1970-1-1'),

                m.stop_2 = as.Date(ifelse(o24.1 > 0 & m.stop_4 >
m.stop_2, m.stop_4, m.stop_2),

                                    origin = '1970-1-1'),

              m.stop_2 = as.Date(ifelse(o25.1 > 0 & stop_5 >
m.stop_2, stop_5, m.stop_2),

                                    origin = '1970-1-1'),

              m.stop_1 = as.Date(ifelse(o12.1 > 0, m.stop_2,
stop_1), origin = '1970-1-1'),

                m.stop_1 = as.Date(ifelse(o13.1 > 0 & m.stop_3 >
m.stop_1, m.stop_3, m.stop_1),

                                    origin = '1970-1-1'),

              m.stop_1 = as.Date(ifelse(o14.1 > 0 & m.stop_4 >
m.stop_1, m.stop_4, m.stop_1),

                                    origin = '1970-1-1'),

              m.stop_1 = as.Date(ifelse(o15.1 > 0 & stop_5 >
m.stop_1, stop_5, m.stop_1),

                                    origin = '1970-1-1'))


   enc_overlap <- enc_overlap %>%

     mutate(m.duration_1 = as.numeric(m.stop_1 - start_1),

            m.duration_2 = ifelse(o12.1 > 0, 0, m.stop_2 -
start_2),
```

```
        m.duration_3 = ifelse(o13.1 | o23.1 > 0, 0,
m.stop_3 - start_3),

        m.duration_4 = ifelse(o14.1 | o24.1 | o34.1 > 0,
0, m.stop_4 - start_4),

        m.duration_5 = ifelse(o15.1 | o25.1 | o35.1 |
o45.1 > 0, 0, duration_5),


        # removing later line if overlap exists

        m.duration_2 = ifelse(o12.1 == 0, m.duration_2,
NA),

        m.duration_3 = ifelse(o13.1 + o23.1 == 0,
m.duration_3, NA),

        m.duration_4 = ifelse(o14.1 + o24.1 + o34.1 ==
0, m.duration_4, NA),


        # adding a yes no for any concurrency

        multyn = ifelse(o12 | o13 | o14 | o15 | o23 |
o24 | o25 | o34 | o35 | o45 > 2,

        1,0))


  enc_overlap$multyn[is.na(enc_overlap$multyn)] <- 0

  #glimpse(enc_overlap)
```

Now, I can make the variables for NHSN duration (only one line day no matter how many lines are in), and total line days (each line contributes a line day each day that it is in). Finally, I am removing erroneuous encounters where the durations = 0, which is a by product of the joining process. This leaves us with our final dataset of **52,383** patient encounters.

```
  #glimpse(enc_overlap)

  enc_overlap <- enc_overlap %>%

    group_by(m.encounter) %>%
```

```
    mutate(NHSN_CLD = sum(m.duration_1, m.duration_2,
m.duration_3,

                            m.duration_4, m.duration_5, na.rm
= TRUE),

          CLD_new = sum(duration_1, duration_2,
duration_3,

             duration_4, duration_5, na.rm = TRUE),

          CLD_diff = CLD_new - NHSN_CLD,

          pct_CLD_NHSN=
(NHSN_CLD/as.integer(CLD_new))*100)


  enc_overlap_final <- enc_overlap %>%

    filter(duration_1 != is.na(duration_1))


  summary(as.integer(enc_overlap_final$CLD_new))

  ##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.

  ##   2.000   3.000   6.000   9.329  12.000 171.000

  summary(as.integer(enc_overlap_final$NHSN_CLD))

  ##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.

  ##   2.000   3.000   5.000   8.064  11.000  89.000

  summary(enc_overlap_final$pct_CLD_NHSN)

  ##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.

  ##   5.882 100.000 100.000  95.030 100.000 100.000

  glimpse(enc_overlap_final)

  ## Observations: 52,383

  ## Variables: 133

  ## $ m.encounter      <dbl> 12103275, 15966363,
15967013, 15967049, 176...
```

```
   ## $ m.clabsi.date        <date> NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA...

   ## $ X                    <int> 1, 5, 6, 7, 9, 10, 13, 15,
16, 19, 20, 28, ...

   ## $ cci_mi_score         <int> 1, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ cci_chf_score        <int> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ cci_pvd_score        <int> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ cci_cvd_score        <int> 0, 0, 0, 1, 0, 0, 0, 0, 0,
1, 0, 0, 0, 0, 0...

   ## $ cci_dementia_score <int> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ cci_cpd_score        <int> 0, 0, 0, 0, 0, 0, 0, 0, 0,
1, 0, 1, 1, 1, 1...

   ## $ cci_rheum_score      <int> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ cci_pud_score        <int> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ cci_mliverdz_score <int> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ cci_dm_score         <int> 1, 0, 0, 0, 1, 1, 1, 0, 0,
0, 1, 0, 0, 1, 0...

   ## $ cci_dmcomp_score   <int> 0, 0, 0, 0, 0, 2, 2, 0, 0,
0, 0, 0, 0, 2, 0...

   ## $ cci_hemipleg_score <int> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 2, 0, 0, 0, 0...

   ## $ cci_renal_score      <int> 2, 0, 0, 2, 2, 2, 2, 0, 0,
0, 0, 0, 0, 2, 2...

   ## $ cci_malig_score      <int> 0, 2, 2, 2, 0, 0, 0, 2, 2,
2, 0, 0, 0, 0, 0...
```

```
   ## $ cci_sliverdz_score <int> 0, 0, 0, 0, 0, 3, 0, 3, 0,
0, 0, 3, 0, 0, 0...

   ## $ cci_mettumor_score <int> 0, 6, 6, 6, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ cci_aids_score    <int> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ cci               <int> 4, 8, 8, 11, 3, 8, 5, 5, 2,
4, 3, 4, 1, 6, ...

   ## $ chemo             <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ tpn               <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ m.mrn             <int> 1130553, 1110050, 1110050,
1110050, 967217,...

   ## $ hospital_1        <chr> "EUH Midtown", "EUH
Midtown", "EUH Midtown"...

   ## $ line.type_1       <chr> "MLC", "Single Port",
"Single Port", "Singl...

   ## $ line.type_2       <chr> NA, NA, NA, NA, NA,
"dialysis", NA, NA, NA,...

   ## $ line.type_3       <chr> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ line.type_4       <chr> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ line.type_5       <chr> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ line.site_1       <chr> "RIJ", "Chest R", "Chest
R", "Chest R", "R ...

   ## $ line.site_2       <chr> NA, NA, NA, NA, NA, "R
Fem", NA, NA, NA, NA...

   ## $ line.site_3       <chr> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...
```

```
## $ line.site_4        <chr> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ line.site_5        <chr> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ lumens_1           <chr> "2L", "no", "no", "no",
"no", "2L", "2L", "...

## $ lumens_2           <chr> NA, NA, NA, NA, NA, "2L",
NA, NA, NA, NA, N...

## $ lumens_3           <chr> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ lumens_4           <chr> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ lumens_5           <chr> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ start_1            <date> 2013-10-08, 2016-12-27,
2017-01-12, 2017-0...

## $ start_2            <date> NA, NA, NA, NA, NA, 2016-
05-16, NA, NA, NA...

## $ start_3            <date> NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA...

## $ start_4            <date> NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA...

## $ start_5            <date> NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA...

## $ stop_1             <date> 2013-10-14, 2016-12-30,
2017-01-17, 2017-0...

## $ stop_2             <date> NA, NA, NA, NA, NA, 2016-
05-23, NA, NA, NA...

## $ stop_3             <date> NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA...

## $ stop_4             <date> NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA...
```

```
## $ stop_5              <date> NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA...

## $ duration_1          <time> 6 days, 3 days, 5 days, 21
days, 7 days, 8...

## $ duration_2          <time> NA days, NA days, NA days,
NA days, NA day...

## $ duration_3          <time> NA days, NA days, NA days,
NA days, NA day...

## $ duration_4          <time> NA days, NA days, NA days,
NA days, NA day...

## $ duration_5          <time> NA days, NA days, NA days,
NA days, NA day...

## $ hd_1                <int> 0, 0, 0, 0, 0, 1, 1, 0, 0,
0, 0, 0, 0, 0, 0...

## $ hd_2                <int> NA, NA, NA, NA, NA, 1, NA,
NA, NA, NA, NA, ...

## $ hd_3                <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ hd_4                <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ hd_5                <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ picc_1              <int> 0, 0, 0, 0, 1, 0, 0, 0, 0,
0, 1, 1, 0, 1, 1...

## $ picc_2              <int> NA, NA, NA, NA, NA, 0, NA,
NA, NA, NA, NA, ...

## $ picc_3              <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ picc_4              <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ picc_5              <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...
```

```
   ## $ port_1              <int> 0, 1, 1, 1, 0, 0, 0, 1, 1,
1, 0, 0, 1, 0, 0...

   ## $ port_2              <int> NA, NA, NA, NA, NA, 0, NA,
NA, NA, NA, NA, ...

   ## $ port_3              <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ port_4              <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ port_5              <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ gt1lumen_1          <int> 1, 0, 0, 0, 0, 1, 1, 0, 0,
0, 1, 0, 0, 1, 1...

   ## $ gt1lumen_2          <int> NA, NA, NA, NA, NA, 1, NA,
NA, NA, NA, NA, ...

   ## $ gt1lumen_3          <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ gt1lumen_4          <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ gt1lumen_5          <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ not.rec.tot_1       <int> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 1, 0, 0...

   ## $ prev.site_1         <int> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ prev.site_2         <int> NA, NA, NA, NA, NA, 0, NA,
NA, NA, NA, NA, ...

   ## $ prev.site_3         <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ prev.site_4         <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ prev.site_5         <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...
```

```
## $ sex_1            <chr> "Female", "Female",
"Female", "Female", "Fe...

## $ birth_1          <date> 1944-05-04, 1947-02-25,
1947-02-25, 1947-0...

## $ age_1            <int> 69, 69, 69, 70, 64, 63, 63,
71, 71, 76, 70,...

## $ admit_1          <date> 2013-10-02, 2016-12-27,
2017-01-12, 2017-0...

## $ dc_1             <date> 2013-10-15, 2016-12-31,
2017-01-18, 2017-0...

## $ los_1            <int> 13, 4, 6, 22, 15, 9, 7, 18,
9, 4, 26, 6, 14...

## $ m.line.type_1    <int> 2, 0, 0, 0, 1, 3, 3, 0, 0,
0, 1, 1, 0, 1, 1...

## $ m.line.type_2    <int> NA, NA, NA, NA, NA, 3, NA,
NA, NA, NA, NA, ...

## $ m.line.type_3    <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ m.line.type_4    <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ m.line.type_5    <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ m.line.site_1    <int> 1, 1, 1, 1, 0, 1, 1, 1, 1,
1, 0, 0, 1, 1, 0...

## $ m.line.site_2    <int> NA, NA, NA, NA, NA, 2, NA,
NA, NA, NA, NA, ...

## $ m.line.site_3    <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ m.line.site_4    <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ m.line.site_5    <int> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...
```

```
## $ clabsiyn        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

## $ hd.tot          <dbl> 0, 0, 0, 0, 0, 1, 1, 0, 0,
0, 0, 0, 0, 0, 0...

## $ o12             <dbl> NA, NA, NA, NA, NA, 7, NA,
NA, NA, NA, NA, ...

## $ o13             <dbl> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ o14             <dbl> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ o15             <dbl> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ o23             <dbl> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ o24             <dbl> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ o25             <dbl> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ o34             <dbl> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ o35             <dbl> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ o45             <dbl> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

## $ o12.1           <dbl> 0, 0, 0, 0, 0, 7, 0, 0, 0,
0, 0, 0, 0, 0, 0...

## $ o13.1           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

## $ o14.1           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

## $ o15.1           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...
```

```
   ## $ o23.1              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ o24.1              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ o25.1              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ o34.1              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ o35.1              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ o45.1              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...

   ## $ m.stop_4           <date> NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA...

   ## $ m.stop_3           <date> NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA...

   ## $ m.stop_2           <date> NA, NA, NA, NA, NA, 2016-
05-23, NA, NA, NA...

   ## $ m.stop_1           <date> 2013-10-14, 2016-12-30,
2017-01-17, 2017-0...

   ## $ m.duration_1       <dbl> 6, 3, 5, 21, 7, 8, 5, 16,
9, 3, 25, 5, 12, ...

   ## $ m.duration_2       <dbl> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ m.duration_3       <dbl> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ m.duration_4       <dbl> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ m.duration_5       <dbl> NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA,...

   ## $ multyn             <dbl> 0, 0, 0, 0, 0, 1, 0, 0, 0,
0, 0, 0, 0, 0, 0...
```

```
   ## $ NHSN_CLD              <dbl> 6, 3, 5, 21, 7, 8, 5, 16,
9, 3, 25, 5, 12, ...

   ## $ CLD_new               <time> 6 days, 3 days, 5 days, 21
days, 7 days, 1...

   ## $ CLD_diff              <time> 0 days, 0 days, 0 days, 0
days, 0 days, 7 ...

   ## $ pct_CLD_NHSN          <dbl> 100.00000, 100.00000,
100.00000, 100.00000,...

   write.csv(enc_overlap_final,

            "~/Desktop/CLABSI Project MPH/analysis/Final
analytic set/final_pre_match_dataset.csv")
```

Propensity Score Balancing

Will Dube

3/21/2019

Propensity score matching

Motivation for matching

Now, I will present tables of the concurrent and non-concurrent groups to illustrate the imbalances of covariates between the two groups. Propensity score matching will allow us to select a group of patients with no concurrent lines who's covariates are more similar to those patients who have at least one concurrent line.

Propensity score match

Now, using logistic regression, we will evaluate each encounter's likelihood of having any concurrence according to severity of illness, hemodialysis, chemotherapy, and total parenteral nutrition. This will allow us to select a group of encounters with no concurrence that is more equally balanced to the concurrent group with respect to those covariates. Then we will just use the likelihood of being "exposed" to concurrent lines in our model for CLABSI. Since "MatchIt" requires that none of our variables have "NAs", I am removing any records that are missing data. This leaves us with **52,381** encounters (only eliminated two eligible encounters).

```
##

## Call:

## matchit(formula = multyn ~ hd.tot + tpn + chemo + cci,
data = match,

##      method = "nearest", ratio = 2)

##

## Summary of balance for all data:

##            Means Treated Means Control SD Control Std.
Mean Diff. eCDF Med

## distance          0.2088        0.1006      0.0863
0.7165    0.2761

## hd.tot            0.4812        0.1267      0.3326
0.7096    0.1773

## tpn               0.1667        0.1011      0.3014
0.1761    0.0328

## chemo             0.0179        0.0363      0.1870
-0.1383    0.0092

## cci               5.4958        4.5288      3.3461
0.2993    0.0125

##              eCDF Mean eCDF Max

## distance      0.2233    0.3651

## hd.tot        0.1773    0.3546
```

```
## tpn          0.0328   0.0657

## chemo        0.0092   0.0184

## cci          0.0420   0.1552

##

##

## Summary of balance for matched data:

##            Means Treated Means Control SD Control Std.
Mean Diff. eCDF Med

## distance        0.2088        0.1914      0.1249
0.1153   0.0001

## hd.tot          0.4812        0.4810      0.4997
0.0005   0.0001

## tpn             0.1667        0.1066      0.3086
0.1614   0.0301

## chemo           0.0179        0.0152      0.1225
0.0204   0.0014

## cci             5.4958        5.0661      3.0344
0.1330   0.0044

##            eCDF Mean eCDF Max

## distance    0.0220   0.0962

## hd.tot      0.0001   0.0003

## tpn         0.0301   0.0602

## chemo       0.0014   0.0027

## cci         0.0188   0.0693

##

## Percent Balance Improvement:

##          Std. Mean Diff. eCDF Med eCDF Mean eCDF Max

## distance        83.9010  99.9694   90.1321  73.6426
```
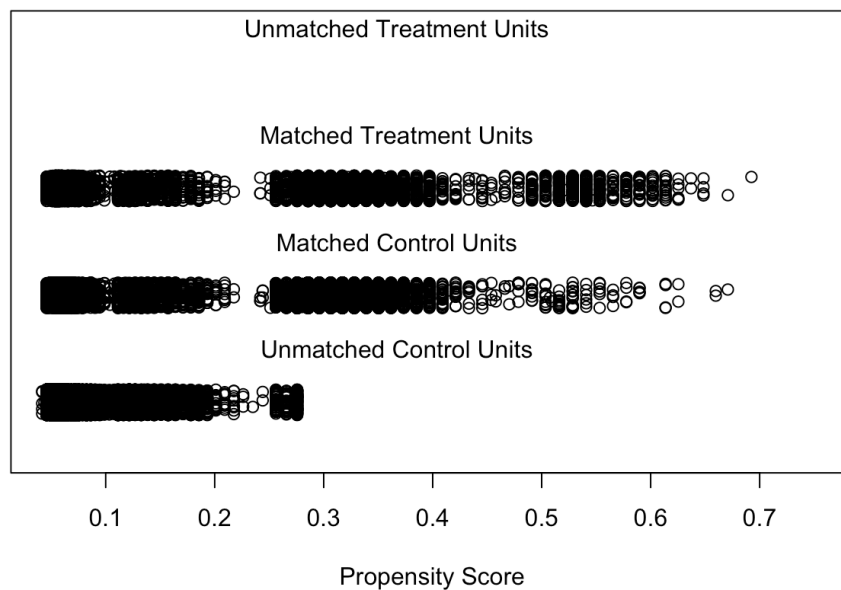
```
## hd.tot              99.9284   99.9284    99.9284   99.9284

## tpn                  8.3484    8.3484     8.3484    8.3484

## chemo               85.2486   85.2486    85.2486   85.2486

## cci                 55.5660   64.8174    55.3599   55.3402

##

## Sample sizes:

##             Control Treated

## All           46473    5908

## Matched       11816    5908

## Unmatched     34657       0

## Discarded         0       0
```
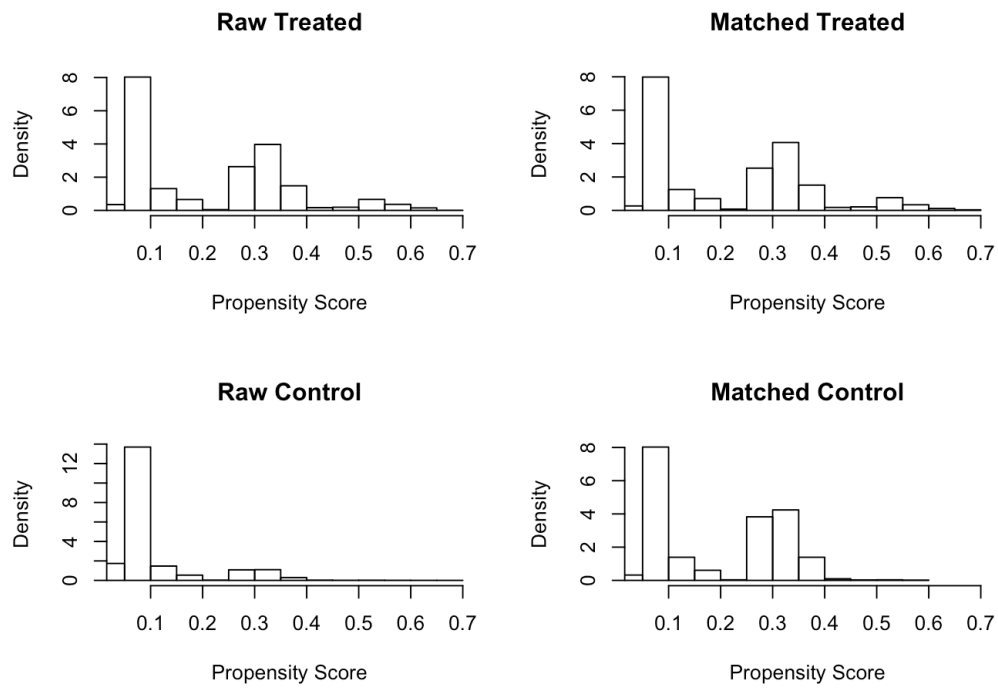
Distribution of Propensity Scores


```
## [1] "To identify the units, use first mouse button; to
stop, use second."

## integer(0)
```

Raw Treated

Matched Treated

Raw Control

Matched Control

Now I am exporting this matched data to a .csv for further analysis