2018

CONSULTATION EXPERIENCE IN PUBLIC HEALTH

Hannah Bellamy
University of Kentucky, hclifton77@uky.edu

STUDENT AGREEMENT:

I represent that my capstone and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's capstone including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Hannah Bellamy, Student

Dr. Heather Bush, Committee Chair

Dr. Corrine Williams, Director of Graduate Studies

CONSULTATION EXPERIENCE IN PUBLIC HEALTH

Hannah Bellamy

Spring 2018

Table of Contents

Summary

The following portfolio is an overview of my consulting experience in public health. My experiences include making frequency tables, producing summary statistics, conducting Chi-Square and T-Test, and reproducing a demographics table. Working with investigators, I have conducted a power analysis for a pilot study, assisted with a study design consultation, used the Kappa coefficient to test for inter-rater reliability, presented survey results with answer distributions, and produced summary statistics for an investigator's research poster.

Working in the Applied Statistics Laboratory and with investigators has taught me the process of consulting. It begins by meeting with an investigator and determining their research needs. This can be many things, including study design, data management, and data analysis. Some investigators come with data and a study design already in place, while others will come to the Applied Statistics Laboratory for each step of the project, from the beginning to end. After determining an investigator's needs and at what point they are in their research, members of the Applied Statistics Laboratory are assigned tasks to assist investigators with their research. In addition to data management and study design, data analysis is often conducted so investigators can receive a report to use for their research results.

Being a part of the Applied Statistics Laboratory and assisting with consultations gave me the opportunity to improve my programming skills. As seen in this portfolio, I wrote numerous SAS programs and utilized various procedures and statements, including higher-level programming with SAS Macros. I also had the opportunity to obtain real-world practice conducting data management and data analysis. An important part of working with investigators was learning how to interpret statistical output, which is found in the "Critical Thinking" portions of each project in this portfolio. In the lab I also had the opportunity to follow up with investigators and produce reports which displays data in a form that is useful to the investigator, such as tables and graphs. The final page of my portfolio includes a poster made by an investigator I worked with. She utilized tables and graphs which I produced in her results section.

I enjoyed my consulting experience in public health and plan to use many of the skills I learned in consulting in the future. I expect to use the numerous SAS skills I learned at my future job, along with what I have learned about study design, data management, and data analysis.

Project 1: Frequency Tables

Description

The purpose of this project was to use proc freq in SAS to create frequency tables for race, marital status, education level, income level, self-report smoking status, smoking from data of pregnant women.

Report: Frequency Tables of Pregnant Women Data

race	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	220	36.48	220	36.48
1	383	63.52	603	100.00
Frequency Missing = 6				

marital	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	209	52.91	209	52.91
1	186	47.09	395	100.00
Frequency Missing = 214				

school	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	142	36.04	142	36.04
1	252	63.96	394	100.00
Frequency Missing = 215				

income	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	157	40.78	157	40.78
1	82	21.30	239	62.08
2	146	37.92	385	100.00
Frequency Missing = 224				

selfsmoke	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	466	76.90	466	76.90
1	140	23.10	606	100.00
Frequency Missing = 3				

smoke1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	279	73.04	279	73.04
1	103	26.96	382	100.00
Frequency Missing = 227				

smoke2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	197	72.96	197	72.96
1	73	27.04	270	100.00
Frequency Missing = 339				

smoke3	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	191	73.75	191	73.75
1	68	26.25	259	100.00
Frequency Missing = 350				

smoke4	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	298	74.87	298	74.87
1	100	25.13	398	100.00
Frequency Missing = 211				

Critical Thinking

The output appears to reflect the data well. In the frequency tables, the number of smokers and nonsmokers in each trimester is clearly defined, with all missing data excluded from the frequency. The correct frequency tables show, for every trimester and post-partum time, about 73% of subjects were non-smokers and 27% were smokers. This was pretty consistent throughout each trimester and appeared to vary by about 1%. It appears that as women progress from first-trimester through post-partum, the percentage of non-smokers declines, with the largest decline being from the first to second trimester. The frequency tables all present demographics the investigator is interested in, and could be incorporated in table 1 of a publication.

To double-check my data, I used proc print to display all the imported data. I compared the imported data to the excel sheet to ensure SAS interpreted it correctly, especially watching how missing data was interpreted. I also used proc contents with the function varnum to see which variables were numeric or character in SAS. During the data step, I also watched how the missing data was handled by checking the table viewer in SAS, to make sure SAS was not converting missing data to zero, which would influence the results of the frequency tables.

Other alternatives to importing the data include using the SAS import wizard, or converting the excel file first to a .csv file or another format. The frequency tables could be made using labels on the variables, so that they are defined more clearly in SAS. Categorical variables, such as race, could be converted from numeric to character values for a different display of the data, depending on how the investigator wishes to continue with the analysis.

Summary

This project taught me how to use many procedures in SAS including proc freq, the ods output function, proc import, proc contents, and practice with if/then statements. These procedures and statements are a basis for SAS programming. This project also taught me the importance of data checks. Using proc print and proc contents are two ways data can be displayed to spot check for errors. Producing and interpreting the frequency tables from this data taught me one way to present demographics data to an investigator, which could be incorporated as part of table 1 in a publication.

SAS Code

```
/*import the data*/
proc import out=pregwomen1
      datafile="C:\Users\HClif\Desktop\Fall 2017\BST 693 Consulting\Lecture 
1\Copy of PregnantWomen.xlsx"
      dbms=xlsx replace;
      getnames=yes ;
      sheet="pregnant";
run;

/*check data*/
proc print data=pregwomen1;
run;
proc contents data=pregwomen1 varnum;
run;

/*fix data- watch missing values*/
data pregwomen2;
set pregwomen1;
/*race*/
if q38=0|q38=2|q38=3|q38=4 then race=0 /*nonwhite*/;
else if q38=1 then race=1 /*white*/;
/*marital status*/
if q39=0|q39=1|q39=2|q39=3|q39=5|q39=6|q39=7 then marital=0 /*not married*/;
else if q39=4 then marital=1 /*married*/;
/*education level*/
if q41=1|q41=2|q41=3|q41=4 then school=0 /*high school or less*/;
else if q41=5|q41=6|q41=7 then school=1 /*above high school*/;
/*income level*/
if q42=1|q42=2|q42=3|q42=4 then income=0 /*0*/;
if q42=5|q42=6|q42=7 then income=1 /*1*/;
else if q42=8|q42=9 then income=2 /*2*/;
/*self-report smoking status*/
if q20=1|q20=3 then selfsmoke=0 /*nonsmokers*/;
else if q20=2 then selfsmoke=1 /*smoker*/;
```

```
/*Smoking 1st trimester*/
if (nic>=0 and nic<3) then smoke1=0 /*nonsmoker*/;
else if (nic>=3 and nic<=6) then smoke1=1 /*smoker*/;
/*Smoking 2nd trimester*/
if (w2nic>=0 and w2nic<3) then smoke2=0 /*nonsmoker*/;
else if (w2nic>=3 and w2nic<=6) then smoke2=1 /*smoker*/;
/*Smoking 3rd trimester*/
if (w3nic>=0 and w3nic<3) then smoke3=0 /*nonsmoker*/;
else if (w3nic>=3 and w3nic<=6) then smoke3=1 /*smoker*/;
/*Smoking post pregnancy*/
if (w4nic>=0 and w4nic<3) then smoke4=0 /*nonsmoker*/;
else if (w4nic>=3 and w4nic<=6) then smoke4=1 /*smoker*/;
run;

/*Frequency Tables*/
ods rtf file="C:\Users\HClif\Desktop\Fall 2017\BST 693 Consulting\Homework
1\Assignment1_HannahBellamy.rtf ";
title 'Frequency Tables of Pregnant Women Data';
proc freq data=pregwomen2;
tables race /*race*/ marital /*marital status*/ school /*education levels*/
income /*income level*/
         selfsmoke /*self-report smoking status*/  smoke1 /*smoking status
1st trimest*/ smoke2 /*smoking status 2st trimest*/
         smoke3 /*smoking status 3st trimest*/ smoke4 /*smoking status post
pregnancy*/;
run;
ods rtf close;
```

Project 2: Means and Summary Statistics

Description

The purpose of this project was to use proc means to create summary statistics including n, mean, standard deviation, and nmiss for the variable IL_10 in the first and second trimesters, using data collected from pregnant women.

Report: Summary Statistics for Variable IL_10

Variable	N	Mean	Std Error	N Miss
IL_10_time1_n	255	27.3521643	4.6007090	354
IL_10_time2_n	223	28.0083229	4.9730791	386

variable	mean	se	num	nm
IL 10 First Trimester	27.3522	4.60071	255	354
IL 10 Second Trimester	28.0083	4.97308	223	386

Critical Thinking

In the summary statistics Interleukin 10, a continuous variable, was measured during the first and second trimesters. The mean, standard error, sample size, and number missing was produced for the first and second trimesters. The table shows the mean for the first trimester was slightly lower than the mean of the second trimester, though the standard errors were somewhat large at 4.6 in the first trimester and approximately 5 in the second semester. Both the trimester had similar sample sizes, at 255 in the first and 223 in the second.

To double-check my analysis I usually look at the data view table in SAS. I find this is convenient as long as I remember to close it. When creating new variables, such as dropping the “<” from IL_10 variables and converting them from character to numeric, I usually spot check a few sections of the view table to make sure it converted correctly. In this case using proc contents with varnum would also be helpful, because it would show if the variable is character or numeric in SAS. In the summary statistics, producing a box and whisker plot for each trimester might be a good way to visually display the data and compare the means and spread of the data.

Summary

This project taught me how to create a table of summary statistics from data. I learned how to use the proc means procedure in SAS in order to create and stack tables of summary statistics for output. Producing and interpreting the frequency tables from this data taught me one way to present summary statistics to an investigator, which could be incorporated in the results section of a publication.

SAS Code

```
/*Import data*/
proc import out=pregwomen1
      datafile="C:\Users\HClif\Desktop\Fall 2017\BST 693 Consulting\Homework 
2\Copy of PregnantWomen.xlsx"
      dbms=xlsx replace;
      getnames=yes ;
      sheet="pregnant";
run;

/*Check data*/
proc print data=pregwomen1;
run;
proc contents data=pregwomen1 varnum;
run;


ods rtf file="C:\Users\HClif\Desktop\Fall 2017\BST 693 Consulting\Homework 
2\Assignment2_HannahBellamy.rtf ";


/*Summary Statistics*/
/*Fix IL_10 variables- character to numeric*/
data pregwomen4;
set pregwomen2;
IL1C= compress(IL_10_time1,'<','');
IL2C= compress(IL_10_time2,'<','');
IL_10_time1_n= input(IL1C,8.);
IL_10_time2_n= input(IL2C,8.);
run;
/*Check data*/
/*proc print data=pregwomen4;
run;
proc contents data=pregwomen4 varnum;
```

```
run;*/
/*Run Summary Stats*/
title 'IL 10 Summary Statistics';
proc means data=pregwomen4 n mean stderr nmiss;
      var IL_10_time1_n IL_10_time2_n;
      output out=IL_stats
      mean=m_IL1 m_IL2
      stderr=se_IL1 se_IL2
      n=n_IL1 n_IL2
      nmiss=nm_IL1 nm_IL2 ;
run;
/*First Trimester Table*/
data IL_stats_2;
      length variable $50;
      set IL_stats;
      keep m_IL1 se_IL1 n_IL1 nm_IL1 variable;
      rename m_IL1=mean se_IL1=se n_IL1=num nm_IL1=nm;
      variable='IL 10 First Trimester';
run;
/*Second Trimester Table*/
data IL_stats_3;
      length variable $50;
      set IL_stats;
      keep m_IL2 se_IL2 n_IL2 nm_IL2 variable;
      rename m_IL2=mean se_IL2=se n_IL2=num nm_IL2=nm;
      variable='IL 10 Second Trimester';
run;
/*Table- Summary Stats*/
title2 'Table of Summary Statistics';
data IL_report;
      set IL_stats_2 IL_stats_3;
run;
proc print data=IL_report noobs;
run;

ods rtf close;
```

Project 3: Chi-Square and T-Test

Description

The purpose of this project was to create a frequency table for ACL status by smoking status and use it to perform a chi-square test, create a summary statistics table for BMI and use it to conduct a t-test, and combine this information into one demographics table to present.

Report: ACL Status Demographics Table

	ACL Injury n=318	No ACL Injury n=282	Test Statistic (Degree of Freedom), P-Value
Non-Smoking Frequency (Column Percent)	204 (54.55%)	170 (45.45%)	$X^2(1)$ =0.25, P-Value=0.62
Smoking Frequency (Column Percent)	101 (52.33%)	92 (47.67%)	
Mean BMI (Standard Error)	28.46 (0.39)	29.05 (0.39)	t(593)=1.06, p-value=0.2882
Median BMI (Minimum, Maximum)	28.41 (17.00, 39.84)	29.00 (17.28, 39.98)	

Missing Values: 33 missing Smoking Status values, 5 missing BMI values

Critical Thinking

Chi-Square test tells us if there is an association between two categorical variables, in this case ACL and smoking status. The categories in both of these variables are binary- ACL status may be injury or no injury, and smoking status may be smoking or non-smoking. The Chi-Square test determines if ACL and smoking status are associated by measuring the difference between the expected counts and the observed counts calculated from the data. In this test, the null hypothesis is that ACL and smoking status are not related, while the alternative is that there is an association between the two. In the table above, you can see the two are likely not related because the column percentages in the ACL group and No ACL group are close. For example, when looking at Smoking Frequency, the column percentages are 52% for ACL and 47% for No ACL. Since they are very close it is unlikely the difference should not be statistically significant in the Chi-Square test. After running this data in SAS, we can conclude that there is not an association between ACL and smoking status at the .05 level. Because the p-value for this test was 0.62, we fail to reject the null hypothesis, that the two are not related

The t-test tells us if there is an association between two variables which are continuous and categorical, in this case ACL status and BMI. The t-test determines if there is enough evidence to show that the mean BMI of each ACL category, ACL injury and no ACL injury, is significantly different. The mean BMI of the ACL injury group was 28.46, while the mean BMI of the no ACL injury group was 29.05. The null hypothesis is that the means are the same, while the alternative hypothesis is that the means differ significantly. The p-value of this t-test was 0.28, therefore we fail to reject the null hypothesis, that the means do not differ, at the .05 level. In other words, according to this test BMI level is not associated with ACL status.

Summary

This project taught me how to conduct both chi-square and t-test in SAS, and how to combine multiple tables into one for clear presentation. I learned how to use proc freq to conduct a chi square test, use proc ttest, and use proc report. Producing and interpreting these two tests taught me one way to present a demographics table.

SAS Code

```
/*Import data. n=600*/
proc import out=acl datafile='C:\Users\HClif\Desktop\Fall 2017\BST 693 Consulting\Lecture 6\Copy of acl.xlsx' dbms=xlsx replace;
	sheet='data';
	getnames=yes;
run;
/*Check data*/
proc contents data=acl varnum;
run;
/*format data*/
proc format;
	value aclf 0="No ACL Injury"
	           1="ACL Injury";
	value smokef 0="Non-Smoker"
	           1="Smoker";
run;
/*Smoking status Variables*/
data acl2;
set acl;
```

```
if (nic>=0 and nic<3) then smoke=0 /*nonsmoker*/;
else if (nic>=3 and nic<=6) then smoke=1 /*smoker*/;
run;
/*Check BMI data*/
proc sort data=acl2 out=BMIcheck;
by bmi;
run;
/*note: BMI this dataset ranges >17 to <40. Typos are outside these ranges.*/
/*Set BMI typos to missing*/
data acl3;
set acl2;
if bmi<17 then bmi=".";
if bmi>40 then bmi=".";
run;
/*Re-Check BMI data*/
proc sort data=acl3 out=BMIcheck2;
by bmi;
run;
/*BMI values missing=5*/
/*Check Smoking data*/
proc sort data=acl3 out=SMOKEcheck;
by smoke;
run;
/*Smoke values missing=33*/
data acl4;
set acl3;
run;
/*Edited data: n=600, mising: smoke=33 bmi=5,dataset acl4*/

/*Smoking & ACL*/
/*frequency table- conti_abys; chisquare chisq_abys*/
proc freq data=acl4;
      tables acl*smoke / nopercent norow chisq out=conti_abys outpct;
      ods output ChiSq=chisq_abys;
      format smoke smokef. acl aclf.;
run;
/*dataset- conti2_abys with variable: count (columns percent %)*/
data conti2_abys;
      length nper $100;
      set conti_abys;
      nper=compress(put(COUNT,8.0))||" ("||compress(put(PCT_COL,8.2))||"%)";
      keep nper acl smoke;
      if cmiss (of _all_) then delete;
run;
/*transponse frequency data from long to wide by smoke*/
proc sort data=conti2_abys;
      by smoke;
run;
proc transpose data=conti2_abys out=t_abys;
      by smoke; /*Variable in rows*/
      id acl; /*variable in columns*/
      var nper; /*variable in cells*/
run;
/*Create p-value column- pull from chisq_abys*/
/*create new data set that merges frequency data together*/
data pval_abys;
      set chisq_abys;
```

```
      if Statistic="Chi-Square"; *just pull the chi sq p-value;
      length test_stat $100; *set length for test statistic and for stacking
p value column;
      test_stat="X^{super 2}("||compress(put(DF,8.0))||")
="||compress(put(Value,8.2))||", P-Value="||compress(put(Prob,PVALUE6.2));
      keep test_stat ;
run;
data tfix_abys; /*fix transposed data so it can easily stack*/
      set t_abys;
      length varlabel $100;
      if Smoke=0 then varlabel="Non-Smoking Frequency (Column Percent)";
      else if Smoke=1 then varlabel="Smoking Frequency (Column Percent)";
      drop _name_ smoke;
run;
/*merge the tables*/
data freq_abys;
      merge tfix_abys pval_abys ;
run;

/*BMI & ACL*/
/*mean, std err, n, median, min, max by acl*/
proc sort data=acl4 out=acl5;
by acl;
run;
proc means data=acl5 mean stderr n median min max;
by acl;
var bmi;
output out=ave_bbya mean=m_bmi stderr=se_bmi n=num_bmi median=med_bmi
min=mini_bmi max=maxi_bmi;
format acl aclf.;
run;
/*creating variable mst for mean standard error*/
data ave2_bbya;
      set ave_bbya;
      length mst $100; /*if you were going character to numeric, you would
use input BEST12*/
      meansterr= compress(put(m_bmi,8.2))||"
("||compress(put(se_bmi,8.2))||")";
      medminmax= compress(put(med_bmi,8.2))||"
("||compress(put(mini_bmi,8.2))||", "||compress(put(maxi_bmi,8.2))||")";
      keep acl meansterr medminmax;
run;
/*transpose- ACL columns*/
proc transpose data=ave2_bbya out=t_bbya;
      id acl ; /*columns variable*/
      var meansterr medminmax; /*cell variable*/
run;
data tfix_bbya;
      set t_bbya;
      length varlabel $100;
      if _NAME_="meansterr" then varlabel="Mean (Standard Error)";
      if _NAME_="medminmax" then varlabel="Median (Minimum, Maximum)";
      drop _NAME_;
run;
/*ANOVA*/
proc ttest data=work.acl5;
class acl;
```

```
var bmi;
ods output Equality=eqvar_bbya TTests=ttest_bbya Statistics=stats_bbya;
run;
/*create a table merging by variable*/
data ttestfix_bbya;
merge ttest_bbya eqvar_bbya;
by variable;
/*keep correct test*/
    if probf < .05 and method="Pooled" then delete;
    if probf >= .05 and method="Satterthwaite" then delete;
    length test_stat $100 ; *keep all lengths the same;
test_stat="t("||compress(put(DF,8.0))||")="||compress(put(tValue,8.2))||", p-
value="||compress(put(probt,pvalue6.4));
keep test_stat;
run;
/*merge tranposed and p-value data*/
data mean_bbya;
      merge tfix_bbya ttestfix_bbya;
run;

/*Combine final datasets*/
data report ;
      set freq_abys mean_bbya; *final data sets for frequencies and means;
run;
/*Column totals*/
proc freq data=acl5;
      tables acl;
run;

options nodate;
ods rtf file="C:\Users\HClif\Desktop\Fall 2017\BST 693 Consulting\Homework
3\Assignment4_HannahBellamy_Edited.rtf";
title "ACL Status Demographics Table";
ods escapechar="^";*define what character you put in for superscript;
proc report data= report split='|'; *nowindows headline headskip;
      columns varlabel ACL_Injury No_ACL_Injury test_stat;
      define varlabel / display left " "; *appear how listed in the data set=
display variable, appear how they are (other: order, group, etc.);
      *empty quotes, no column title;
      define ACL_Injury / display center "ACL Injury | n=318";
      define No_ACL_Injury / display center "No ACL Injury | n=282";
      define test_stat / display center "Test Statistic (Degree of Freedom),
P-Value";
run;
title; *turns off title in use;
ods rtf close;
```

Project 4: Demographics Table

Description

The purpose of this project was to create a demographics table with Neonatal Abstinence Syndrome and No Neonatal Abstinence Syndrome columns. The demographics to be included were race, treatment status, gestational age (GA) and birth weight. Using macros for frequency and summary statistics was recommended for this project. Chi square and t-tests were used to determine if there were any statistically significant differences between infants with Neonatal Abstinence Syndrome and infants without.

Report: Neonatal Abstinence Syndrome Demographics Table

	Neonatal Abstinence Syndrome n=170	No Neonatal Abstinence Syndrome n=430	Test Statistic (Degree of Freedom), P-Value
Non-White Frequency (Column Percent)	28 (16.47%)	24 (5.58%)	$X^2(1)$ =18.25, P-Value=<.01
White Frequency (Column Percent)	142 (83.53%)	406 (94.42%)	
Control Frequency (Column Percent)	72 (42.35%)	245 (56.98%)	$X^2(1)$ =10.46, P-Value=<.01
Treatment Frequency (Column Percent)	98 (57.65%)	185 (43.02%)	
Mean GA (Standard Error)	34.20 (0.18)	39.47 (0.07)	t(219)=27.10, p-value=<.0001
Median GA (Minimum, Maximum)	34.18 (30.09, 37.97)	39.46 (37.02, 42.00)	
Mean BirthWeight (Standard Error)	1753.39 (34.06)	3812.13 (36.34)	t(501)=41.34, p-value=<.0001
Median BirthWeight (Minimum, Maximum)	1783.39 (1029.96, 2492.66)	3855.41 (2511.85, 4997.78)	

Critical Thinking

The chi-square test indicates if there is an association between two categorical variables. Looking at race variables, it appears there is a statistically significant association between race and neonatal abstinence syndrome (NAS) status. The p-value of this test is <.01, indicating the association is statistically significant at the .05 level. Looking at the treatment variable, there also appears to be a statistically significant association between treatment status and NAS. The p-value of this test is <.01, indicating it is statistically significant at the .05 level.

The t-test determines if there is an association between a continuous variable and a categorical variable. In this case the variables are gestational age (GA) and neonatal abstinence syndrome (NAS) status. There appears to be a statistically significant association between mean gestational age and NAS status at the .05 level. Babies with NAS have a lower mean gestational age at 34.2 weeks, while babies without NAS had a much higher mean gestational age at 39.47 weeks. Results were similar for mean birth weight. Birth weight and NAS status were statistically significant at the .05 level, indicating there is an association between the two. Mean birth weight for NAS babies was 1753g while babies without NAS averaged much higher at 3812g.

Summary

This project taught me how to produce a larger demographics table in SAS and look for statistically significant associations. In addition, I learned how to use SAS macros and create more complicated, reproducible code. Learning SAS macros allows for much more efficient programming.

SAS Code

```
/*import data*/
proc import out=nas datafile='C:\Users\HClif\Desktop\Fall 2017\BST 693
Consulting\Homework 5\NAS.xlsx' dbms=xlsx replace;
      sheet='data';
      getnames=yes;
run;
/*Check data*/
proc contents data=nas varnum;
run;
/*format data*/
proc format;
      value nasf 0="No Neonatal Abstinence Syndrome"
                     1="Neonatal Abstinence Syndrome";
run;
/*fix grp variable*/
data nas;
set nas;
if grp="treatment" then grp="Treatment";
run;
data nas;
set nas;
if grp="Control" then grp1=0;
```

```
if grp="Treatment" then grp1=1;
run;
/*fix birth weight variable*/
data nas;
set nas;
BirthWeight=input(birth_weight,8.);
run;

/*DEMOGRAPHICS TABLE*/
/*categorical macro*/
%macro freqtab(data,rowvar,colvar,newdata,format1,row1,row2);
proc freq data=&data;
      tables &rowvar*&colvar / nopercent norow chisq out=conti_&newdata
outpct;
      ods output ChiSq=chisq_&newdata;
      format &colvar &format1;
run;
data conti2_&newdata;
      length nper $100; *nper doesnt matter because it will disappear;
      set conti_&newdata;
      nper=compress(put(COUNT,8.0))||" ("||compress(put(PCT_COL,8.2))||"%)";
      keep nper &rowvar &colvar;
run;
proc sort data=conti2_&newdata;
      by &rowvar;
run;
proc transpose data=conti2_&newdata out=t_&newdata;
      by &rowvar;
      id &colvar;
      var nper;
run;
data pval_&newdata;
      set chisq_&newdata;
      if Statistic="Chi-Square";
      length test_stat $100;
      test_stat="X^{super 2}("||compress(put(DF,8.0))||")
="||compress(put(Value,8.2))||", P-Value="||compress(put(Prob,PVALUE6.2));
      keep test_stat ;
run;
data tfix_&newdata;
      set t_&newdata;
      length varlabel $100;
      if &rowvar=0 then varlabel="&row1 Frequency (Column Percent)";
      else varlabel="&row2 Frequency (Column Percent)";
      drop _name_ &rowvar;
run;
data freq_&newdata;
      merge tfix_&newdata pval_&newdata ;
      *format ;
run;
%mend;
/*Frequency NASxRace, Chisq*/
%freqtab(nas,race, nas,rxn,nasf.,Non-White,White);
/*Frequency NASxTreatment Group, Chisq*/
%freqtab(nas,grp1,nas,grpxn,nasf.,Control,Treatment);
/*Stack frequency tables*/
data freq_report;
```

```
set freq_rxn freq_grpxn;
run;

/*continuous macro*/
%macro ttest(data, rowvar, colvar, newdata, format);
proc sort data=&data;
      by &colvar;
run;
proc means data=&data mean stderr n median min max;
      by &colvar;
      var &rowvar;
      output out=ave_&newdata mean=m_&rowvar stderr=se_&rowvar n=num_&rowvar
median=med_&rowvar min=mini_&rowvar max=maxi_&rowvar;

      format &colvar &format;
run;
data ave2_&newdata;
      set ave_&newdata;
      length mst $100;
      mst= compress(put(m_&rowvar,8.2))||"
("||compress(put(se_&rowvar,8.2))||")";
      medminmax= compress(put(med_&rowvar,8.2))||"
("||compress(put(mini_&rowvar,8.2))||",
"||compress(put(maxi_&rowvar,8.2))||")";
      keep &colvar mst medminmax;
run;
proc transpose data=ave2_&newdata out=t_&newdata;
      id &colvar ;
      var mst medminmax;
run;
data tfix_&newdata;
      set t_&newdata;
      length varlabel $100;
      if _NAME_="mst" then varlabel="Mean &rowvar (Standard Error)";
      if _NAME_="medminmax" then varlabel="Median &rowvar (Minimum,
Maximum)";
      drop _NAME_;
run;
proc ttest data=&data ;
      class &colvar ;
      var &rowvar ;
      ods output Equality=eqvar_&newdata TTests=ttest_&newdata
Statistics=stats_&newdata;
run;
data ttestfix_&newdata;
    merge ttest_&newdata eqvar_&newdata (keep=variable ProbF); /*Tells SAS to
only keep the p-value within eqvar_&newdata*/
    by variable;
      /*keep correct test*/
    if probf < .05 and method="Pooled" then delete;
    if probf >= .05 and method="Satterthwaite" then delete;
    length test_stat $100 ; *keep all lengths the same;
      test_stat="t("||compress(put(DF,8.0))||")="||compress(put(tValue,8.2))|
|", p-value="||compress(put(probt,pvalue6.4));
      keep test_stat ;
run;
data mean_&newdata;
```

```
      merge tfix_&newdata ttestfix_&newdata;
run;
%mend;
/*Summary Statistics NASxGestational Age, ttest, p-value*/
%ttest(nas,GA,nas,gaxn,nasf.);
/*Summary Statistics NASxBirth Weight, ttest, p-value*/
%ttest(nas,BirthWeight,nas,bwxn,nasf.);

/*Stacking summary stats*/
data mean_report;
set mean_gaxn mean_bwxn;
run;

/*Stacking tables for report*/
data report;
set freq_report mean_report;
run;
/*Column totals*/
proc freq data=nas;
      tables nas;
run;

options nodate;
ods rtf file="C:\Users\HClif\Desktop\Fall 2017\BST 693 Consulting\Homework
5\Assignment5_HannahBellamy.rtf";
title "Neonatal Abstinence Syndrome Demographics Table";
ods escapechar="^";
proc report data=report split='|';
      columns varlabel Neonatal_Abstinence_Syndrome
No_Neonatal_Abstinence_Syndrome test_stat;
      define varlabel / display left " ";
      define Neonatal_Abstinence_Syndrome / display center "Neonatal
Abstinence Syndrome | n=170";
      define No_Neonatal_Abstinence_Syndrome / display center "No Neonatal
Abstinence Syndrome | n=430";
      define test_stat / display center "Test Statistic (Degree of Freedom),
P-Value";
run;
```

Project 5: Reproduce a Demographics Table

Description

The purpose of this project was to write my own SAS code to reproduce a Table 1. Table 1: Patient Demographics was given, along with the data file.

Table 1: Patient Demographics

	MFM (N=161)	OB (N=198)	FP/NP (N=43)	Test Statistic (Degrees of Freedom), p-value
Type1 (Column Percent)	33 (20.5%)	1 (0.5%)	0	$X^2(6)$=138.89, p-value=<.0001
Type2 (Column Percent)	42 (26.1%)	10 (5.1%)	0	
GestA1 (Column Percent)	22 (13.7%)	103 (52.0%)	34 (79.1%)	
GestA2 (Column Percent)	64 (39.8%)	84 (42.4%)	9 (20.9%)	
Mean Age (Standard Error)	30.13 (0.5)	31.44 (0.5)	31.01 (1.1)	F(2,399)=1.96, p-value=0.1418
Primiparous (Column Percent)	34 (21.1%)	71 (35.9%)	13 (30.2%)	$X^2(2)$=9.32, p-value=0.0095
Not Primiparous (Column Percent)	127 (78.9%)	127 (64.1%)	30 (69.8%)	
Mean BMI (Standard Error)	32.29 (0.6)	29.75 (0.5)	28.58 (0.7)	F(2,399)=7.07, p-value=0.0010
Tobacco Use (Column Percent)	22 (13.7%)	28 (14.1%)	4 (9.3%)	$X^2(2)$=0.72, p-value=0.6963
No Tobacco Use (Column Percent)	139 (86.3%)	170 (85.9%)	39 (90.7%)	
Alcohol (Column Percent)	5 (3.1%)	0	0	$X^2(2)$=7.58, p-value=0.0226
No Alcohol (Column Percent)	156 (96.9%)	198 (100.0%)	43 (100.0%)	
Family History of Diabetes (Column Percent)	99 (61.5%)	76 (38.4%)	18 (41.9%)	$X^2(2)$=19.72, p-value=<.0001
No Family History of Diabetes (Column Percent)	62 (38.5%)	122 (61.6%)	25 (58.1%)	
Mean OGTT Result (Standard Error)	175.34 (4.0)	167.83 (1.9)	183.21 (6.1)	F(2,399)=3.33, p-value=0.0368
Family History of Preterm Birth (Column Percent)	47 (29.2%)	28 (14.1%)	2 (4.7%)	$X^2(2)$=19.53, p-value=<.0001
No History of Preterm Birth (Column Percent)	114 (70.8%)	170 (85.9%)	41 (95.3%)	
Hypertensive Disorder (Column Percent)	76 (47.2%)	65 (32.8%)	11 (25.6%)	$X^2(2)$=10.87, p-value=0.0044
No Hypertensive Disorder (Column Percent)	85 (52.8%)	133 (67.2%)	32 (74.4%)	
Nephropathy (Column Percent)	13 (8.1%)	3 (1.5%)	0	$X^2(2)$=11.99, p-value=0.0025
No Nephropathy (Column Percent)	148 (91.9%)	195 (98.5%)	43 (100.0%)	
Neuropathy (Column Percent)	2 (1.2%)	0	0	$X^2(2)$=3.01, p-value=0.2222
No Neuropathy (Column Percent)	159 (98.8%)	198 (100.0%)	43 (100.0%)	
Retinopathy (Column Percent)	1 (0.6%)	0	0	$X^2(2)$=1.50, p-value=0.4722
No Retinopathy (Column Percent)	160 (99.4%)	198 (100.0%)	43 (100.0%)	

Data are represented as mean (SE) or n(%). *Indicates statistical significance at 95% confidence level.

Report: Reproduction of Table 1: Patient Demographics

	MFM (N=161)	OB (N=198)	FP/NP (N=43)	Test Statistic (Degree of Freedom), P-Value
GestA1 (Column Percent)	22 (13.66%)	103 (52.02%)	34 (79.07%)	$X^2(6)$ =138.89, P-Value=<.0001
GestA2 (Column Percent)	64 (39.75%)	84 (42.42%)	9 (20.93%)	
Type 1 (Column Percent)	33 (20.50%)	1 (0.51%)	0	
Type 2 (Column Percent)	42 (26.09%)	10 (5.05%)	0	
Mean Age (Standard Error)	30.13 (0.5)	31.44 (0.5)	31.01 (1.1)	F(2,399)=1.96, p-value=0.1418
No Primiparous (Column Percent)	127 (78.88%)	127 (64.14%)	30 (69.77%)	$X^2(2)$ =9.32, P-Value=0.0095
Primiparous (Column Percent)	34 (21.12%)	71 (35.86%)	13 (30.23%)	
Mean BMI (Standard Error)	32.29 (0.6)	29.75 (0.5)	28.58 (0.7)	F(2,399)=7.07, p-value=0.0010
No Tobacco Use (Column Percent)	139 (86.34%)	170 (85.86%)	39 (90.70%)	$X^2(2)$ =0.72, P-Value=0.6963
Tobacco Use (Column Percent)	22 (13.66%)	28 (14.14%)	4 (9.30%)	
No Alcohol (Column Percent)	156 (96.89%)	198 (100.00%)	43 (100.00%)	$X^2(2)$ =7.58, P-Value=0.0226
Alcohol (Column Percent)	5 (3.11%)	0	0	
No Family History of Diabetes (Column Percent)	62 (38.51%)	122 (61.62%)	25 (58.14%)	$X^2(2)$ =19.72, P-Value=0.0001
Family History of Diabetes (Column Percent)	99 (61.49%)	76 (38.38%)	18 (41.86%)	
Mean OGTT (Standard Error)	175.34 (4.0)	167.83 (1.9)	183.21 (6.1)	F(2,399)=3.33, p-value=0.0368
No History of Preterm Birth (Column Percent)	114 (70.81%)	170 (85.86%)	41 (95.35%)	$X^2(2)$ =19.53, P-Value=0.0001
History of Preterm Birth (Column Percent)	47 (29.19%)	28 (14.14%)	2 (4.65%)	
No Hypertensive Disorder (Column Percent)	85 (52.80%)	133 (67.17%)	32 (74.42%)	$X^2(2)$ =10.87, P-Value=0.0044
Hypertensive Disorder (Column Percent)	76 (47.20%)	65 (32.83%)	11 (25.58%)	
No Nephropathy (Column Percent)	148 (91.93%)	195 (98.48%)	43 (100.00%)	$X^2(2)$ =11.99, P-Value=0.0025
Nephropathy (Column Percent)	13 (8.07%)	3 (1.52%)	0	
No Neuropathy (Column Percent)	159 (98.76%)	198 (100.00%)	43 (100.00%)	$X^2(2)$ =3.01, P-Value=0.2222
Neuropathy (Column Percent)	2 (1.24%)	0	0	
No Retinopathy (Column Percent)	160 (99.38%)	198 (100.00%)	43 (100.00%)	$X^2(2)$ =1.50, P-Value=0.4722
Retinopathy (Column Percent)	1 (0.62%)	0	0	

Data are represented as mean (SE) or n(%). *Indicates statistical significance at 95% confidence level.

Critical Thinking

I was able to reproduce all of Table 1. Originally when I ran the data my numbers were slightly off. After doing data checks I noticed my sample size was slightly larger than that of Table 1, and there was missing data in some of the cells. In order to replicate Table 1 I deleted the rows of data with any missing values. In practice, this is probably not the best choice, and it would be better to continue with a few cells of missing data. However, doing this allowed me to replicate the table.

The things I did to replicate it included deleting all rows with missing data, changing the format of age, bmi and ogtt from character to numeric, using the format of 8.4 for the p-value, match the titles, and create macros so similar variables ran the same code. I tried to add plenty of comments to my code so that if someone needed to reproduce the table or a similar one again in the future they could see what I did and easily adjust my code. For structure I put data changes (such as the character to numeric change) directly after reading in the data, with data checks before and after. I think it is important to have the data checks at multiple points so someone else running the code does not forget to rerun them and can notice the changes made.

Summary

This project taught me how to take data and reproduce an example of a Table 1 or demographics table. It was somewhat challenging to write code which reproduces someone else's table, since doing one step differently will result in a different table. For example, in order to reproduce this table I had to delete all rows with missing data, which I did not think to do on the first try. This project also taught me the importance of commenting code well and organizing it clearly, so if someone in the future wanted to reproduce a table I made they could use my code and understand what I did to get identical results.

SAS Code

```
/*import data*/

proc import out=health datafile="C:\Users\HClif\Desktop\Fall 2017\BST 693 
Consulting\Reproducibality Project\data.csv" dbms=csv replace;
	getnames=yes;
run;
```

```
/*check data*/
proc contents data=health varnum;
run;
proc print data=health;
run;
proc means data=health;
run;
proc freq data=health nlevels;
run;
/*Data Fix*/
*character to numeric (var1?, age, bmi, ogtt);
data health;
set health;
agen=input(age,8.);
bmin=input(BMI,8.);
ogttn=input(OGTT,8.);
run;
*delete rows with missing data;
data health;
set health;
if cmiss (of _all_) then delete;
run;
/*Re-Check Data*/
proc contents data=health varnum;
run;
proc means data=health;
run;
proc freq data=health nlevels;
run;

/*Frequency Tables*/
*frequency table diabetes;
proc freq data=health;
tables diabetes*pvgrp / nopercent norow chisq out=conti_diabetes outpct;
ods output ChiSq=chisq_diabetes;
run;
*data with variable count (columns percent %);
data conti2_diabetes;
length nper $100;
set conti_diabetes;
nper=compress(put(COUNT,8.0))||" ("||compress(put(PCT_COL,8.2))||"%)";
keep nper diabetes pvgrp;
run;
*transpose long to wide;
proc sort data=conti2_diabetes;
by diabetes;
run;
proc transpose data=conti2_diabetes out=t_diabetes;
by diabetes; *rows;
id pvgrp; *columns;
var nper; *cells;
run;
*p-value column;
data pval_diabetes;
set chisq_diabetes;
if Statistic="Chi-Square";
length test_stat $100;
```

```
test_stat="X^{super 2}("||compress(put(DF,8.0))||")
="||compress(put(Value,8.2))||", P-Value="||compress(put(Prob,8.4));
keep test_stat;
run;
*fix transposed data for stacking;
data tfix_diabetes;
set t_diabetes;
length varlabel $100;
if diabetes="GDMA 1" then varlabel="GestA1 (Column Percent)";
if diabetes="GDMA 2" then varlabel="GestA2 (Column Percent)";
if diabetes="Type 1" then varlabel="Type 1 (Column Percent)";
if diabetes="Type 2" then varlabel="Type 2 (Column Percent)";
drop _name_ diabetes;
run;
*merge tables;
data freq_diabetes;
merge pval_diabetes tfix_diabetes ;
run;

/*Macro for Frequency Tables*/
%macro freq(var1,var2);
*frequency table primi;
proc freq data=health;
tables &var1*pvgrp / nopercent norow chisq out=conti_&var1 outpct;
ods output ChiSq=chisq_&var1;
run;
*data with variable count (columns percent %);
data conti2_&var1;
length nper $100;
set conti_&var1;
nper=compress(put(COUNT,8.0))||" ("||compress(put(PCT_COL,8.2))||"%)";
keep nper &var1 pvgrp;
run;
*transpose long to wide;
proc sort data=conti2_&var1;
by &var1;
run;
proc transpose data=conti2_&var1 out=t_&var1;
by &var1; *rows;
id pvgrp; *columns;
var nper; *cells;
run;
*p-value column;
data pval_&var1;
set chisq_&var1;
if Statistic="Chi-Square";
length test_stat $100;
test_stat="X^{super 2}("||compress(put(DF,8.0))||")
="||compress(put(Value,8.2))||", P-Value="||compress(put(Prob,8.4));
keep test_stat;
run;
*fix transposed data for stacking;
data tfix_&var1;
set t_&var1;
length varlabel $100;
if &var1="Yes" then varlabel="&var2 (Column Percent)";
if &var1="No" then varlabel="No &var2 (Column Percent)";
```

```
drop _name_ &var1;
run;
*merge tables;
data freq_&var1;
merge tfix_&var1 pval_&var1;
run;
%mend;
%freq(primi,Primiparous);
%freq(tobac,Tobacco Use);
%freq(alc,Alcohol);
%freq(hist_diab,Family History of Diabetes);
%freq(hist_pre,History of Preterm Birth);
%freq(hyper,Hypertensive Disorder);
%freq(neph,Nephropathy);
%freq(neuro,Neuropathy);
%freq(retin,Retinopathy);

/*Macro for Means Tables*/
%macro oneway(var1,var2);
*mean std err agen;
proc sort data=health;
by pvgrp;
run;
proc means data=health mean stderr n;
by pvgrp;
var  &var1;
output out=ave_&var1 mean=m_&var1 stderr=se_&var1;
run;
*create mean standard error variable mst;
data ave2_&var1;
set ave_&var1;
length mst $100;
mst= compress(put(m_&var1,8.2))||" ("||compress(put(se_&var1,8.1))||")";
keep pvgrp mst;
run;
*transpose long to wide;
proc transpose data=ave2_&var1 out=t_&var1;
id pvgrp; *columns;
var mst; *cells;
run;
*fix for stacking;
data tfix_&var1;
set t_&var1;
length varlabel $100;
varlabel="Mean &var2 (Standard Error)";
drop _NAME_;
run;
*F stat and p-value;
proc glimmix data=health ;
      class pvgrp ;
      model &var1 = pvgrp ;
      lsmeans pvgrp/pdiff; *default lsmeans and stderr;
      ods output diffs=diff_&var1 lsmeans=ls_&var1 tests3=t3_&var1 ;*tests3
for type 3 analysis;
run;
*F-value test stat column;
data fval_&var1;
```

```
set t3_&var1;
length test_stat $100;
test_stat="F("||compress(put(NumDF,8.0))||","||compress(put(DenDF,8.0))||")="
||compress(put(FValue,8.2))||", p-value="||compress(put(ProbF,8.4)) ;
keep test_stat;
run;
*fix transposed data for stacking;
data tfix_&var1;
set t_&var1;
length varlabel $100;
varlabel="Mean &var2 (Standard Error)";
drop _NAME_;
run;
*merge tables;
data final_&var1;
merge tfix_&var1 fval_&var1;
run;
%mend;
%oneway(agen, Age);
%oneway(bmin, BMI);
%oneway(ogttn, OGTT);

/*Create Report*/
*stack multiple datasets;
data report;
set freq_diabetes final_agen freq_primi final_bmin freq_tobac freq_alc
freq_hist_diab final_ogttn freq_hist_pre
freq_hyper freq_neph freq_neuro freq_retin;
run;
*******for column totals in titles;
proc freq data=health;
tables pvgrp;
run;
*run output and input column totals below N=x;
****************;
ods rtf file="C:\Users\HClif\Desktop\Fall 2017\BST 693
Consulting\Reproducibality Project\ReproducibilityReport_HannahBellamy.rtf";
title "Table 1: Patient Demographics";
ods escapechar="^"; *defines superscript;
proc report data=report split="|";
columns varlabel MFM OB FP_NP test_stat;
define varlabel / display left " "; *blank;
define MFM / display center "MFM | (N=161)";
define OB / display center "OB | (N=198)";
define FP_NP / display center "FP/NP | (N=43)";
define test_stat / display center "Test Statistic (Degree of Freedom), P-
Value";
run;
title;
ods text="Data are represented as mean (SE) or n(%). *Indicates statistical
significance at 95% confidence level.";
ods rtf close;
***************;
```

Project 6: Study Design Consultation

Description

An investigator requested a meeting for study design, data management, and data analysis with the Applied Statistics Laboratory. He described how he was interested in conducting a study using hospital records of people with hydrocephalus, a brain condition which requires a cerebrospinal fluid drain to be placed surgically placed. Afterwards, sometimes a shunt should be placed- patients are monitored during follow-up to determine if this is the best option. His research questions were:

1. Is a CT scan necessary, or is clinical exam sufficient to determine if temporary shunt placement is needed?
2. In patients who initially received no shunt, what factors led to them require delayed shunt placement?

The study design was observational and retrospective, since he was using previous hospital records. The analysis goals were to determine which factors lead to delayed shunt placement (such as measurements from CT scans and other clinical observations of the patient), and if a CT scan is clinically necessary. It was recommended he use RedCap for data storage after his study was approved by IRB. The investigator did not follow-up after this meeting.

Report

The investigator did not return with data, so no report or RedCap form was produced.

Critical Thinking

This investigator came with a significant amount of background information and needed help pairing down what his main research question was and the variables that would be relevant. I think investigators often come with large amounts of data, or feel they will need to collect information on a large number of variables to find something significant. It was informative to see how Statisticians help investigators determine clear research questions and define what investigators need to accomplish their study.

Summary

Although the investigator did not return with data, I still learned a lot from this consultation. Investigators may not want to change their previous method of data entry, such as this investigator who preferred excel sheets. Although efforts were made to explain how to use RedCap, investigators may not be interested in taking the time to learn a new program or method. Statisticians can explain that RedCap is beneficial because it is a very secure form of data entry. If investigators still decline to use preferred methods, they can be counseled on other methods of securely storing sensitive data, like hospital records. In addition, I learned are that investigators may be very busy and not follow-up for a long time. Although this investigator did not return while I was in the ASL it is possible they may return later. Because of this, it is important to keep through and clear notes about past meetings.

Project 7: Power Analysis Consultation

Description

An investigator requested a meeting for study design and power analysis for a pilot study with the Applied Statistics Laboratory. She described how she was funded to conduct a pilot study about factors which contribute to an asthma diagnosis. Her research questions were:

1. What is the sample size needed for this pilot study?
2. What environmental factors contribute to an asthma diagnosis in the Appalachian population?

The investigator planned to collect data, by asking patients about their medical history. The data would be collected one time. Primary variables to be collected were mechanical ventilation status, smoking status, pollution, and home location. Other demographic variables would also be collected. For this pilot study, she planned to use the medical history to determine rates of asthma among two groups- those patients with no lung issues and those with lung issues.

Report: Power Analysis

Alpha	Power	Mean Lung Issues	Mean No Lung Issues	Standard Deviation	N Total	N Per Group
0.05	0.8	0.6	0.3	0.4	58	29
0.05	0.8	0.6	0.3	0.5	90	45

Critical Thinking

Above are the results for two different power analyses, both at 80% power but with different standard deviations. Standard deviations 0.4 and 0.5 were used, as specified by the investigator. The output shows that with a standard deviation of 0.4, the recommended total sample size is 58 and the sample size for each of the two groups is 29. With a standard deviation of 0.5, the recommended total sample size is 90 and the sample size for each of the two groups is 45.

Summary

Power analysis was conducted and provided to the investigator. The investigator reported being funded for a study with a sample size of 125, so will be able to conduct this pilot study. She was excited to get the power analysis back and planned to use RedCap to collect her data.

SAS Code

```
/*Power analysis*/
proc power;
      onewayanova
    groupmeans = 0.6 | 0.3
    stddev = 0.4 0.5
    ntotal = .
    power = 0.8;
      ods output output=power;
run;
data power2;
set power;
npergrp=NTotal/2;
label npergrp="N Per Group" Mean1="Lung Issues" Mean2="No Lung Issues";
keep Alpha NominalPower Mean1 Mean2 StdDev NTotal npergrp;
run;

options nodate;
ods rtf file="S:\ASL Current\ASL
Requests_Current\jlstur0.1264\Reports_Output\Power Analysis and Output.rtf";
title "Power Analysis";
ods text="Below are the results for two different power analyses, both at 80%
power but with different standard deviations.
Standard deviations 0.4 and 0.5 were used.  The output shows with a standard
deviation of 0.4, the recommended total sample size is 58
and the sample size for each of the two groups is 29. With a standard
deviation of 0.5, the recommended total sample size is 90
and the sample size for each of the two groups is 45. ";
proc report data=power2 split="|";
columns Alpha NominalPower Mean1 Mean2 StdDev NTotal npergrp;
define NominalPower / display "Power" center;
define Mean1 / display "Mean | Lung Issues" center;
define Mean2 / display "Mean | No Lung Issues" center;
define StdDev / display "Standard Deviation" center;
run;
title;
ods rtf close;
```

Project 8: Inter-Rater Reliability Consultation

Description

An investigator requested a meeting for data management, and data analysis with the Applied Statistics Laboratory. She created a smartphone app for doctors to input medical student review into. She wanted to know if there was any improvement in quality of medical student review comparing before the app and after the introduction of the app. To determine this, she had four different raters rate the quality of review before and after. The raters scored each review in 7 distinct categories: QI, CCERR, Category 1 through 4. These were quality scores the investigator determined were applicable to medical student reviews. For this portion of her project, her research question was: Does the rating of review quality vary among raters?

To determine this, I conducted Cohen's Kapp to test inter-rater reliability. This test compares raters to each other and gives a kappa estimate statistic, which is an indicator of the degree of agreement between two reviewers.

Report: Cohen's Kappa for Testing Inter-rater Reliability

Reliability for CCERR1 by CCERR2

Frequencies CCERR1 Score Category	Frequencies CCERR2 Score Category 0 Low	Frequencies CCERR2 Score Category 1 Medium	Frequencies CCERR2 Score Category 2 High	Kappa Estimate (95% CI)
0	16	1	0	0.619 (0.406,0.832)
1	6	17	1	
2	0	1	1	

Reliability for CCERR1 by CCERR3

Frequencies CCERR1 Score Category	Frequencies CCERR3 Score Category 0 Low	Frequencies CCERR3 Score Category 1 Medium	Frequencies CCERR3 Score Category 2 High	Kappa Estimate (95% CI)
0	17	0	0	0.190 (0.025,0.355)
1	18	6	0	
2	0	2	0	

Reliability for CCERR1 by CCERR4

Frequencies CCERR1 Score Category	Frequencies CCERR4 Score Category 0 Low	Frequencies CCERR4 Score Category 1 Medium	Frequencies CCERR4 Score Category 2 High	Kappa Estimate (95% CI)
0	16	1	0	0.509 (0.283,0.734)
1	11	13	0	
2	0	0	2	

Reliability for CCERR2 by CCERR3

Frequencies CCERR2 Score Category	Frequencies CCERR3 Score Category 0 Low	Frequencies CCERR3 Score Category 1 Medium	Frequencies CCERR3 Score Category 2 High	Kappa Estimate (95% CI)
0	21	1	0	0.211 (0.001,0.422)
1	14	5	0	
2	0	2	0	

Reliability for CCERR2 by CCERR4

Frequencies CCERR2 Score Category	Frequencies CCERR4 Score Category 0 Low	Frequencies CCERR4 Score Category 1 Medium	Frequencies CCERR4 Score Category 2 High	Kappa Estimate (95% CI)
0	19	3	0	0.432 (0.186,0.679)
1	8	10	1	
2	0	1	1	

Reliability for CCERR3 by CCERR4

Frequencies CCERR3 Score Category	Frequencies CCERR4 Score Category 0 Low	Frequencies CCERR4 Score Category 1 Medium	Frequencies CCERR4 Score Category 2 High	Kappa Estimate (95% CI)
0	26	9	0	0.348 (0.100,0.597)
1	1	5	2	
2	0	0	0	

Reliability for QI1 by QI2

Frequencies QI1 Score Category	Frequencies QI2 Score Category 0 Low	Frequencies QI2 Score Category 1 Medium	Frequencies QI2 Score Category 2 High	Kappa Estimate (95% CI)
0	17	4	0	0.748 (0.562,0.934)
1	1	18	1	
2	0	0	2	

Reliability for QI1 by QI3

Frequencies QI1 Score Category	Frequencies QI3 Score Category 0 Low	Frequencies QI3 Score Category 1 Medium	Frequencies QI3 Score Category 2 High	Kappa Estimate (95% CI)
0	18	3	0	0.663 (0.451,0.875)
1	4	15	1	
2	0	0	2	

Reliability for QI1 by QI4

Frequencies QI1 Score Category	Frequencies QI4 Score Category 0 Low	Frequencies QI4 Score Category 1 Medium	Frequencies QI4 Score Category 2 High	Kappa Estimate (95% CI)
0	19	2	0	0.656 (0.438,0.874)
1	6	14	0	
2	0	0	2	

Reliability for QI2 by QI3

Frequencies QI2 Score Category	Frequencies QI3 Score Category 0 Low	Frequencies QI3 Score Category 1 Medium	Frequencies QI3 Score Category 2 High	Kappa Estimate (95% CI)
0	16	2	0	0.590 (0.370,0.810)
1	6	15	1	
2	0	1	2	

Reliability for QI2 by QI4

Frequencies QI2 Score Category	Frequencies QI4 Score Category 0 Low	Frequencies QI4 Score Category 1 Medium	Frequencies QI4 Score Category 2 High	Kappa Estimate (95% CI)
0	16	2	0	0.504 (0.269,0.739)
1	9	13	0	
2	0	1	2	

Reliability for QI3 by QI4

Frequencies QI3 Score Category	Frequencies QI4 Score Category 0 Low	Frequencies QI4 Score Category 1 Medium	Frequencies QI4 Score Category 2 High	Kappa Estimate (95% CI)
0	20	2	0	0.658 (0.444,0.871)
1	5	13	0	
2	0	1	2	

Reliability for cat_1_1 by cat_1_2

Frequencies cat_1_1 Score Category	Frequencies cat_1_2 Score Category 0 Low	Frequencies cat_1_2 Score Category 2 High	Kappa Estimate (95% CI)
0	33	0	0.782 (0.549,1.000)
2	3	7	

Reliability for cat_1_1 by cat_1_3

Frequencies cat_1_1 Score Category	Frequencies cat_1_3 Score Category 0 Low	Frequencies cat_1_3 Score Category 2 High	Kappa Estimate (95% CI)
0	33	0	0.697 (0.428,0.967)
2	4	6	

Reliability for cat_1_1 by cat_1_4

Frequencies cat_1_1 Score Category	Frequencies cat_1_4 Score Category 0 Low	Frequencies cat_1_4 Score Category 2 High	Kappa Estimate (95% CI)
0	33	0	0.782 (0.549,1.000)
2	3	7	

Reliability for cat_1_2 by cat_1_3

Frequencies cat_1_2 Score Category	Frequencies cat_1_3 Score Category 0 Low	Frequencies cat_1_3 Score Category 2 High	Kappa Estimate (95% CI)
0	36	0	0.909 (0.735,1.000)
2	1	6	

Reliability for cat_1_2 by cat_1_4

Frequencies cat_1_2 Score Category	Frequencies cat_1_4 Score Category 0 Low	Frequencies cat_1_4 Score Category 2 High	Kappa Estimate (95% CI)
0	35	1	0.829 (0.600,1.000)
2	1	6	

Reliability for cat_1_3 by cat_1_4

Frequencies cat_1_3 Score Category	Frequencies cat_1_4 Score Category 0 Low	Frequencies cat_1_4 Score Category 2 High	Kappa Estimate (95% CI)
0	36	1	0.909 (0.735,1.000)
2	0	6	

Reliability for cat_2_1 by cat_2_2

Frequencies cat_2_1 Score Category	Frequencies cat_2_2 Score Category 0 Low	Frequencies cat_2_2 Score Category 2 High	Kappa Estimate (95% CI)
0	2	2	0.205 (-0.135,0.546)
2	7	32	

Reliability for cat_2_1 by cat_2_3

Frequencies cat_2_1 Score Category	Frequencies cat_2_3 Score Category 0 Low	Frequencies cat_2_3 Score Category 2 High	Kappa Estimate (95% CI)
0	3	1	0.550 (0.162,0.937)
2	3	36	

Reliability for cat_2_1 by cat_2_4

Frequencies cat_2_1 Score Category	Frequencies cat_2_4 Score Category 0 Low	Frequencies cat_2_4 Score Category 2 High	Kappa Estimate (95% CI)
0	1	3	0.289 (-0.209,0.788)
2	1	38	

Reliability for cat_2_2 by cat_2_3

Frequencies cat_2_2 Score Category	Frequencies cat_2_3 Score Category 0 Low	Frequencies cat_2_3 Score Category 2 High	Kappa Estimate (95% CI)
0	2	7	0.119 (-0.210,0.448)
2	4	30	

Reliability for cat_2_2 by cat_2_4

Frequencies cat_2_2 Score Category	Frequencies cat_2_4 Score Category 0 Low	Frequencies cat_2_4 Score Category 2 High	Kappa Estimate (95% CI)
0	1	8	0.114 (-0.176,0.405)
2	1	33	

Reliability for cat_2_3 by cat_2_4

Frequencies cat_2_3 Score Category	Frequencies cat_2_4 Score Category 0 Low	Frequencies cat_2_4 Score Category 2 High	Kappa Estimate (95% CI)
0	1	5	0.194 (-0.202,0.589)
2	1	36	

Reliability for cat_3_1 by cat_3_2

Frequencies cat_3_1 Score Category	Frequencies cat_3_2 Score Category 0 Low	Frequencies cat_3_2 Score Category 2 High	Kappa Estimate (95% CI)
0	20	6	0.626 (0.397,0.856)
2	2	15	

Reliability for cat_3_1 by cat_3_3

Frequencies cat_3_1 Score Category	Frequencies cat_3_3 Score Category 0 Low	Frequencies cat_3_3 Score Category 2 High	Kappa Estimate (95% CI)
0	20	6	0.247 (-0.047,0.542)
2	9	8	

Reliability for cat_3_1 by cat_3_4

Frequencies cat_3_1 Score Category	Frequencies cat_3_4 Score Category 0 Low	Frequencies cat_3_4 Score Category 2 High	Kappa Estimate (95% CI)
0	21	5	0.459 (0.187,0.732)
2	6	11	

Reliability for cat_3_2 by cat_3_3

Frequencies cat_3_2 Score Category	Frequencies cat_3_3 Score Category 0 Low	Frequencies cat_3_3 Score Category 2 High	Kappa Estimate (95% CI)
0	21	1	0.578 (0.346,0.809)
2	8	13	

Reliability for cat_3_2 by cat_3_4

Frequencies cat_3_2 Score Category	Frequencies cat_3_4 Score Category 0 Low	Frequencies cat_3_4 Score Category 2 High	Kappa Estimate (95% CI)
0	20	2	0.579 (0.341,0.817)
2	7	14	

Reliability for cat_3_3 by cat_3_4

Frequencies cat_3_3 Score Category	Frequencies cat_3_4 Score Category 0 Low	Frequencies cat_3_4 Score Category 2 High	Kappa Estimate (95% CI)
0	24	5	0.591 (0.340,0.843)
2	3	11	

Reliability for cat_4_1 by cat_4_2

Frequencies cat_4_1 Score Category	Frequencies cat_4_2 Score Category 0 Low	Frequencies cat_4_2 Score Category 2 High	Kappa Estimate (95% CI)
0	27	1	0.120 (-0.105,0.344)
2	13	2	

Reliability for cat_4_1 by cat_4_3

Frequencies cat_4_1 Score Category	Frequencies cat_4_3 Score Category 0 Low	Frequencies cat_4_3 Score Category 2 High	Kappa Estimate (95% CI)
0	26	2	0.436 (0.155,0.716)
2	8	7	

Reliability for cat_4_1 by cat_4_4

Frequencies cat_4_1 Score Category	Frequencies cat_4_4 Score Category 0 Low	Frequencies cat_4_4 Score Category 2 High	Kappa Estimate (95% CI)
0	28	0	0.167 (-0.042,0.376)
2	13	2	

Reliability for cat_4_2 by cat_4_3

Frequencies cat_4_2 Score Category	Frequencies cat_4_3 Score Category 0 Low	Frequencies cat_4_3 Score Category 2 High	Kappa Estimate (95% CI)
0	34	6	0.442 (0.098,0.785)
2	0	3	

Reliability for cat_4_2 by cat_4_4

Frequencies cat_4_2 Score Category	Frequencies cat_4_4 Score Category 0 Low	Frequencies cat_4_4 Score Category 2 High	Kappa Estimate (95% CI)
0	39	1	0.365 (-0.196,0.925)
2	2	1	

Reliability for cat_4_3 by cat_4_4

Frequencies cat_4_3 Score Category	Frequencies cat_4_4 Score Category 0 Low	Frequencies cat_4_4 Score Category 2 High	Kappa Estimate (95% CI)
0	34	0	0.311 (-0.027,0.650)
2	7	2	

Critical Thinking

Cohen's Kappa was estimated to determine the degree of agreement among two reviewers. Typically, how Kappa is interpreted:

≤0 = Disagreement

0.01-0.20 = Slight Agreement

0.21-0.40 = Fair Agreement

0.41-0.60 = Moderate Agreement

0.61-0.80 = Substantial Agreement

0.81-1 = Almost perfect Agreement

Note that CCERR and QI were each grouped into the categories Low, Medium, and High because these ratings had greater than six response types. Category 1 through Category 4 variables were grouped into Low or High categories because they had six or less response options.

For example, in the table "Reliability for CCERR1 by CCERR2", 16 subjects received a score of "Low" from both reviewer 1 and reviewer 2. In the same table, 17 subjects received a score of "medium" from both reviewer 1 and reviewer 2, and one subject received a score of "high" from both reviewers. Here the kappa estimate is 0.619 which is substantial agreement, since reviewers generally gave subjects the same score.

As another example, in the table "Reliability for cat_1_1 by cat_1_2" the level of agreement is compared between raters 1 and 2 for the cat_1 variable. 33 subjects received a score of "low" from both rater 1 and rater 2, and 7 subjects scored a "high" from both raters 1 and 2. The kappa estimate, or level of agreement, is substantial at 0.782.

One of the tables for reliability between raters had almost perfect agreement, which would be a kappa estimate of .81 to 1.0. This was table Cat_1_2 by Cat_1_3 with kappa 0.909. Many had substantial agreement with Kappa .61 to .80 or moderate agreement with Kappa .41 to .60. They are highlighted above. Substantial reliability between raters was found most often in variables QI and Cat_1. All raters had at least moderate agreement with each other in the QI and Cat_1.

Summary

I found working with this data was somewhat of a challenge. Each variable had a different scale to rate the quality of reviews, for example the variable CCERR could be rated on a scale of 0 to 35. Before conducting the Kappa test I had to categorize the reviewer's ratings for each variable. I created a macro to do this. There were four raters, so I created a second macro to conduct the Kappa test to do a pairwise comparison, which compares each rater to the other, for each variable. I created a third macro to run proc report, this simplified the process since there were so many tables. This consultation work taught me that SAS macros can be a very valuable tool when producing many tables. Additionally, I learned about the Kappa coefficient and could report the inter-rater reliability results to this investigator for her research.

SAS Code

```
/*import data*/
proc import out=raters datafile='S:\ASL Current\ASL
Requests_Current\rmto225.1165\Data\Raw Data\evaluation.xlsx'
      dbms=xlsx replace;
      sheet='All Data';
    getnames=yes;
run;

/*Check Data*/
proc contents data=raters varnum;
run;
proc means data=raters min max mean maxdec=0;
var _numeric_;
run;

/*Categorize data
0 low, 1 medium, 2 high*/
data raters;
set raters;
*CCERR;
if (CCERR1>=7 and CCERR1<=15 ) then CCERR1=0;
if (CCERR1>=16 and CCERR1<=25 ) then CCERR1=1;
if (CCERR1>=26 and CCERR1<=35 ) then CCERR1=2;
if (CCERR2>=7 and CCERR2<=15 ) then CCERR2=0;
if (CCERR2>=16 and CCERR2<=25 ) then CCERR2=1;
if (CCERR2>=26 and CCERR2<=35 ) then CCERR2=2;
if (CCERR3>=7 and CCERR3<=15 ) then CCERR3=0;
if (CCERR3>=16 and CCERR3<=25 ) then CCERR3=1;
if (CCERR3>=26 and CCERR3<=35 ) then CCERR3=2;
if (CCERR4>=7 and CCERR4<=15 ) then CCERR4=0;
if (CCERR4>=16 and CCERR4<=25 ) then CCERR4=1;
if (CCERR4>=26 and CCERR4<=35 ) then CCERR4=2;
*QI;
```

```
if (QI1>=0 and QI1<=2 ) then QI1=0;
if (QI1>=3 and QI1<=5 ) then QI1=1;
if (QI1>=6 and QI1<=7 ) then QI1=2;
if (QI2>=0 and QI2<=2 ) then QI2=0;
if (QI2>=3 and QI2<=5 ) then QI2=1;
if (QI2>=6 and QI2<=7 ) then QI2=2;
if (QI3>=0 and QI3<=2 ) then QI3=0;
if (QI3>=3 and QI3<=5 ) then QI3=1;
if (QI3>=6 and QI3<=7 ) then QI3=2;
if (QI4>=0 and QI4<=2 ) then QI4=0;
if (QI4>=3 and QI4<=5 ) then QI4=1;
if (QI4>=6 and QI4<=7 ) then QI4=2;
*Cat1;
if Cat_1_1=0 then Cat_1_1=0;
if Cat_1_1>=1 then Cat_1_1=2;
if Cat_1_2=0 then Cat_1_2=0;
if Cat_1_2>=1 then Cat_1_2=2;
if Cat_1_3=0 then Cat_1_3=0;
if Cat_1_3>=1 then Cat_1_3=2;
if Cat_1_4=0 then Cat_1_4=0;
if Cat_1_4>=1 then Cat_1_4=2;
*Cat2;
if Cat_2_1=0 then Cat_2_1=0;
if Cat_2_1>=1 then Cat_2_1=2;
if Cat_2_2=0 then Cat_2_2=0;
if Cat_2_2>=1 then Cat_2_2=2;
if Cat_2_3=0 then Cat_2_3=0;
if Cat_2_3>=1 then Cat_2_3=2;
if Cat_2_4=0 then Cat_2_4=0;
if Cat_2_4>=1 then Cat_2_4=2;
*Cat3;
if Cat_3_1=0 then Cat_3_1=0;
if Cat_3_1>=1 then Cat_3_1=2;
if Cat_3_2=0 then Cat_3_2=0;
if Cat_3_2>=1 then Cat_3_2=2;
if Cat_3_3=0 then Cat_3_3=0;
if Cat_3_3>=1 then Cat_3_3=2;
if Cat_3_4=0 then Cat_3_4=0;
if Cat_3_4>=1 then Cat_3_4=2;
*Cat4;
if Cat_4_1=0 then Cat_4_1=0;
if Cat_4_1>=1 then Cat_4_1=2;
if Cat_4_2=0 then Cat_4_2=0;
if Cat_4_2>=1 then Cat_4_2=2;
if Cat_4_3=0 then Cat_4_3=0;
if Cat_4_3>=1 then Cat_4_3=2;
if Cat_4_4=0 then Cat_4_4=0;
if Cat_4_4>=1 then Cat_4_4=2;
run;

/*Adjust data to Square table for proc freq*/
%macro freqfix(score,rater1, rater2, rater3, rater4,var);
data &var;
&rater1=&score; &rater2=&score; &rater3=&score; &rater4=&score;
WT=0; output;
run;
```

```
%mend;
*CCERR fix;
%freqfix (0, CCERR1, CCERR2, CCERR3, CCERR4,  CCERR0);
%freqfix (1, CCERR1, CCERR2, CCERR3, CCERR4,  CCERR1);
%freqfix (2, CCERR1, CCERR2, CCERR3, CCERR4,  CCERR2);
*QI fix;
%freqfix (0, QI1, QI2, QI3, QI4,  QI0);
%freqfix (1, QI1, QI2, QI3, QI4,  QI1);
%freqfix (2, QI1, QI2, QI3, QI4,  QI2);
*Cat_1_ fix;
%freqfix (0, cat_1_1, cat_1_2, cat_1_3, cat_1_4, cat1_0);
%freqfix (2, cat_1_1, cat_1_2, cat_1_3, cat_1_4, cat1_2);
*Cat_2_ fix;
%freqfix (0, cat_2_1, cat_2_2, cat_2_3, cat_2_4, cat2_0);
%freqfix (2, cat_2_1, cat_2_2, cat_2_3, cat_2_4, cat2_2);
*Cat_3_ fix;
%freqfix (0, cat_3_1, cat_3_2, cat_3_3, cat_3_4, cat3_0);
%freqfix (2, cat_3_1, cat_3_2, cat_3_3, cat_3_4, cat3_2);
*Cat_4_ fix;
%freqfix (0, cat_4_1, cat_4_2, cat_4_3, cat_4_4, cat4_0);
%freqfix (2, cat_4_1, cat_4_2, cat_4_3, cat_4_4, cat4_2);
*merge tables;
data raters_freqfix;
set raters
ccerr0 ccerr1 ccerr2
QI0 QI1 QI2
cat1_0  cat1_2
cat2_0  cat2_2
cat3_0  cat3_2
cat4_0  cat4_2
;
if wt=. then wt=1;
run;

/*Macro for Kappa and Tables*/
%macro kappa (var1,var2,var3);
/*Cohen's Kappa*/
proc freq data=raters_freqfix nlevels;
    weight wt / zeros;
    tables &var1*&var2 / nocol norow nopercent; *agree sparse;
      exact kappa;
      ods output Kappa=kapcoef_&var3 KappaTest=kappval_&var3
CrossTabFreqs=freqs_&var3;
    run;
/*Table- Kappa coefficient with CI*/
*kappa test value;
data kapcoefest_&var3;
      set kapcoef_&var3;
      if name1="_KAPPA_";
      rename nvalue1=kapcoef;
      keep nvalue1;
run;
*lower CI;
data kapcoefl_&var3;
      set kapcoef_&var3;
      if name1="L_KAPPA";
```

```
      rename nvalue1=kapcoefl;
      keep nvalue1;
run;
*upper CI;
data kapcoefu_&var3;
      set kapcoef_&var3;
      if name1="U_KAPPA";
      rename nvalue1=kapcoefu;
      keep nvalue1;
run;
*kappa test(lower, upper);
data kappa_&var3;
      length kapest $50;
      merge kapcoefest_&var3 kapcoefl_&var3 kapcoefu_&var3;
      kapest=compress(put(kapcoef,8.3))||"
("||compress(put(kapcoefl,8.3))||","||compress(put(kapcoefu,8.3))||")";
      keep kapest;
run;
/*Table- pvalue*/
data kappval_&var3;
      set kappval_&var3;
      if name1="P2_KAPPA";
      rename nvalue1=pval;
      keep nvalue1;
run;
/*Table- crosstabs*/
*remove missing;
data freqs_&var3;
      set freqs_&var3;
      if &var1 ne .;
      if &var2 ne .;
run;
*organize table var1 x var2;
proc transpose data=freqs_&var3 out=tfreqs_&var3 suffix=_&var2; *mri x gross;
      by &var1;
      id &var2;
      var Frequency;
run;
/*Table- final merge*/
data fullkap_&var3;
      merge kappa_&var3 kappval_&var3 tfreqs_&var3;
      drop _name_ _label_;
run;
%mend;
*CCERR pariwise;
%kappa (CCERR1,CCERR2,CCERR_1x2);
%kappa (CCERR1,CCERR3,CCERR_1x3);
%kappa (CCERR1,CCERR4,CCERR_1x4);
%kappa (CCERR2,CCERR3,CCERR_2x3);
%kappa (CCERR2,CCERR4,CCERR_2x4);
%kappa (CCERR3,CCERR4,CCERR_3x4);
*QI pairwise;
%kappa (QI1,QI2,QI_1x2);
%kappa (QI1,QI3,QI_1x3);
%kappa (QI1,QI4,QI_1x4);
%kappa (QI2,QI3,QI_2x3);
```

```
%kappa (QI2,QI4,QI_2x4);
%kappa (QI3,QI4,QI_3x4);
*Cat_1 pairwise;
%KAPPA (cat_1_1,cat_1_2,cat1_1x2);
%KAPPA (cat_1_1,cat_1_3,cat1_1x3);
%KAPPA (cat_1_1,cat_1_4,cat1_1x4);
%KAPPA (cat_1_2,cat_1_3,cat1_2x3);
%KAPPA (cat_1_2,cat_1_4,cat1_2x4);
%KAPPA (cat_1_3,cat_1_4,cat1_3x4);
*Cat_2 pairwise;
%KAPPA (cat_2_1,cat_2_2,cat2_1x2);
%KAPPA (cat_2_1,cat_2_3,cat2_1x3);
%KAPPA (cat_2_1,cat_2_4,cat2_1x4);
%KAPPA (cat_2_2,cat_2_3,cat2_2x3);
%KAPPA (cat_2_2,cat_2_4,cat2_2x4);
%KAPPA (cat_2_3,cat_2_4,cat2_3x4);
*Cat_3 pairwise;
%KAPPA (cat_3_1,cat_3_2,cat3_1x2);
%KAPPA (cat_3_1,cat_3_3,cat3_1x3);
%KAPPA (cat_3_1,cat_3_4,cat3_1x4);
%KAPPA (cat_3_2,cat_3_3,cat3_2x3);
%KAPPA (cat_3_2,cat_3_4,cat3_2x4);
%KAPPA (cat_3_3,cat_3_4,cat3_3x4);
*Cat_4 pairwise;
%KAPPA (cat_4_1,cat_4_2,cat4_1x2);
%KAPPA (cat_4_1,cat_4_3,cat4_1x3);
%KAPPA (cat_4_1,cat_4_4,cat4_1x4);
%KAPPA (cat_4_2,cat_4_3,cat4_2x3);
%KAPPA (cat_4_2,cat_4_4,cat4_2x4);
%KAPPA (cat_4_3,cat_4_4,cat4_3x4);

/*Output tables to rtf file*/
options nodate nonumber ;
ods rtf file="S:\ASL Current\ASL
Requests_Current\rmto225.1165\Programs\Reliability-
HB\ReliabilityTables2.rtf" startpage=no;
title "Cohen's Kappa for Testing Inter-rate Reliability";
*ccerr tables;
%macro ccerr (var1, var2,var3);
ods text="Reliability for &var1 by &var2";
proc report data=fullkap_&var3 nowindows headskip split="|";
      columns &var1 _0_&var2 _1_&var2 _2_&var2
                  kapest;
      define &var1 / display "Frequencies | &var1 Score Category" center;
      define _0_&var2 / display "Frequencies | &var2 Score Category 0 | Low"
center;
      define _1_&var2 / display "Frequencies | &var2 Score Category 1 |
Medium" center;
      define _2_&var2 / display "Frequencies | &var2 Score Category 2 | High"
center;
      define kapest / display "Kappa Estimate (95% CI)" center;
run;
%mend;
%ccerr (CCERR1,CCERR2,CCERR_1x2);
%ccerr (CCERR1,CCERR3,CCERR_1x3);
%ccerr (CCERR1,CCERR4,CCERR_1x4);
```

```
%ccerr (CCERR2,CCERR3,CCERR_2x3);
%ccerr (CCERR2,CCERR4,CCERR_2x4);
%ccerr (CCERR3,CCERR4,CCERR_3x4);
*QI tables;
%macro QI (var1, var2,var3);
ods text="Reliability for &var1 by &var2";
proc report data=fullkap_&var3 nowindows headline headskip split="|";
      columns &var1 _0_&var2 _1_&var2 _2_&var2
                  kapest;
      define &var1 / display "Frequencies | &var1 Score Category" center;
      define _0_&var2 / display "Frequencies | &var2 Score Category 0 | Low"
center;
      define _1_&var2 / display "Frequencies | &var2 Score Category 1 |
Medium" center;
      define _2_&var2 / display "Frequencies | &var2 Score Category 2 | High"
center;
      define kapest / display "Kappa Estimate (95% CI)" center;
run;
%mend;
%QI (QI1,QI2,QI_1x2);
%QI (QI1,QI3,QI_1x3);
%QI (QI1,QI4,QI_1x4);
%QI (QI2,QI3,QI_2x3);
%QI (QI2,QI4,QI_2x4);
%QI (QI3,QI4,QI_3x4);
*cat1-4 tables;
%macro cat (var1, var2,var3);
ods text="Reliability for &var1 by &var2";
proc report data=fullkap_&var3 nowindows headline headskip split="|";
      columns &var1 _0_&var2 _2_&var2
                  kapest;
      define &var1 / display "Frequencies | &var1 Score Category" center;
      define _0_&var2 / display "Frequencies | &var2 Score Category 0 | Low"
center;
      define _2_&var2 / display "Frequencies | &var2 Score Category 2 | High"
center;
      define kapest / display "Kappa Estimate (95% CI)" center;
run;
%mend;
*cat1 tables;
%cat (cat_1_1,cat_1_2,cat1_1x2);
%cat (cat_1_1,cat_1_3,cat1_1x3);
%cat (cat_1_1,cat_1_4,cat1_1x4);
%cat (cat_1_2,cat_1_3,cat1_2x3);
%cat (cat_1_2,cat_1_4,cat1_2x4);
%cat (cat_1_3,cat_1_4,cat1_3x4);
*cat2 tables;
%cat (cat_2_1,cat_2_2,cat2_1x2);
%cat (cat_2_1,cat_2_3,cat2_1x3);
%cat (cat_2_1,cat_2_4,cat2_1x4);
%cat (cat_2_2,cat_2_3,cat2_2x3);
%cat (cat_2_2,cat_2_4,cat2_2x4);
%cat (cat_2_3,cat_2_4,cat2_3x4);
*cat3 tables;
%cat (cat_3_1,cat_3_2,cat3_1x2);
%cat (cat_3_1,cat_3_3,cat3_1x3);
```

```
%cat (cat_3_1,cat_3_4,cat3_1x4);
%cat (cat_3_2,cat_3_3,cat3_2x3);
%cat (cat_3_2,cat_3_4,cat3_2x4);
%cat (cat_3_3,cat_3_4,cat3_3x4);
*cat4 tables;
%cat (cat_4_1,cat_4_2,cat4_1x2);
%cat (cat_4_1,cat_4_3,cat4_1x3);
%cat (cat_4_1,cat_4_4,cat4_1x4);
%cat (cat_4_2,cat_4_3,cat4_2x3);
%cat (cat_4_2,cat_4_4,cat4_2x4);
%cat (cat_4_3,cat_4_4,cat4_3x4);
title;
ods rtf close;
```

Project 9: Survey Results Consultation

Description

This was a follow-up from an investigator who requested data management and analysis from the Applied Statistics Laboratory. The investigator created a smartphone app for doctors to input medical student review into. She wanted to know if there was any improvement in quality of medical student review comparing before the app and after the introduction of the app. In addition to this work, she conducted surveys to gather the opinions of the doctors using the app. One survey was sent out each month of the study, August through December. Another survey was sent out before doctors began using the app, in July, and after the study was over, at the end of December. The investigator requested the distribution of responses for each survey question.

Report 1: Survey Results July vs. December

	Question	Answer Options	July Count (Percent)	December Count (Percent)
q1	How many medical student evaluations do you complete in 6 months?	0 evaluations	2 (7.14%)	1 (6.67%)
		1-2 evaluations	7 (25.00%)	7 (46.67%)
		3-4 evaluations	12 (42.86%)	5 (33.33%)
		5 or more evaluations	7 (25.00%)	2 (13.33%)
		No Response	19	32
q2	How would you rate the quantity of comments that you write on medical student evaluations?	A few phrases	16 (57.14%)	7 (46.67%)
		At least a paragraph	1 (3.57%)	2 (13.33%)
		No Response	19	32
		No comments	2 (7.14%)	1 (6.67%)
		Several sentences	9 (32.14%)	5 (33.33%)
q3	The quality of comments that I write on medical student evaluations is	Adequate	18 (64.29%)	11 (73.33%)
		Good	3 (10.71%)	2 (13.33%)

	Question	Answer Options	July Count (Percent)	December Count (Percent)
		No Response	19	32
		Poor	4 (14.29%)	1 (6.67%)
		Very good	1 (3.57%)	1 (6.67%)
		Very poor	2 (7.14%)	
q4_1	Please indicate the reasons that you have difficulty completing the comments on the medical student evaluations (you may mark more than one item): (choice=I dont have enough time.)	Checked	6 (12.77%)	2 (4.26%)
		Unchecked	41 (87.23%)	45 (95.74%)
q4_2	Please indicate the reasons that you have difficulty completing the comments on the medical student evaluations (you may mark more than one item): (choice=I forget many of my comments about the student by the time I fill out the evaluation.)	Checked	13 (27.66%)	4 (8.51%)
		Unchecked	34 (72.34%)	43 (91.49%)
q4_3	Please indicate the reasons that you have difficulty completing the comments on the medical student evaluations (you may mark more than one item): (choice=I am not sure what type of comments I should write.)	Checked	14 (29.79%)	4 (8.51%)
		Unchecked	33 (70.21%)	43 (91.49%)
q4_4	Please indicate the reasons that you have difficulty completing the comments on the medical student evaluations (you may mark more than one item): (choice=Other)	Checked	4 (8.51%)	
		Unchecked	43 (91.49%)	47 (100.00%)
q4_5	Please indicate the reasons that you have difficulty completing the comments on the medical student evaluations (you may mark more than one item): (choice=None. I have no difficulty completing the comments on student evaluations.)	Checked	2 (4.26%)	7 (14.89%)
		Unchecked	45 (95.74%)	40 (85.11%)
q5	What type of mobile device do you primarily use?	Android phone	4 (14.29%)	3 (20.00%)
		No Response	19	32
		iPhone	24 (85.71%)	12 (80.00%)
q6	How comfortable are you when using an app on your mobile device?	Comfortable	7 (25.00%)	3 (20.00%)
		Neutral	2 (7.14%)	1 (6.67%)
		No Response	19	32
		Very comfortable	13 (46.43%)	7 (46.67%)

	Question	Answer Options	July Count (Percent)	December Count (Percent)
		Very uncomfortable	6 (21.43%)	4 (26.67%)
q7	How often do you use apps on your mobile device?	Almost never	1 (3.57%)	1 (6.67%)
		Daily	9 (32.14%)	2 (13.33%)
		Monthly	1 (3.57%)	1 (6.67%)
		Multiple times per day	15 (53.57%)	11 (73.33%)
		No Response	19	32
		Weekly	2 (7.14%)	

Report 2: Survey Results August through December

	Question	Answer Options	August Count (Percent)	September Count (Percent)	October Count (Percent)	November Count (Percent)	December Count (Percent)
q1	For the most recent week that you worked with a medical student, how often did you use the mobile diary app?	1-2 days	4 (16.00%)	7 (31.82%)	2 (10.53%)	5 (31.25%)	5 (33.33%)
		3-4 days	1 (4.00%)				1 (6.67%)
		5 or more days					1 (6.67%)
		Never	20 (80.00%)	15 (68.18%)	17 (89.47%)	11 (68.75%)	8 (53.33%)
		No Response	22	25	28	31	32
q2	Over the past month, how often did you use the mobile diary app to help you complete medical student evaluations?	A few times	6 (24.00%)	5 (22.73%)	1 (5.26%)	3 (18.75%)	4 (26.67%)
		Always		1 (4.55%)	1 (5.26%)	1 (6.25%)	
		I did not complete any student evaluations in the past month.	3 (12.00%)	5 (22.73%)	3 (15.79%)	1 (6.25%)	1 (6.67%)

	Question	Answer Options	August Count (Percent)	September Count (Percent)	October Count (Percent)	November Count (Percent)	December Count (Percent)
		Most of the time				2 (12.50%)	2 (13.33%)
		Never	16 (64.00%)	11 (50.00%)	14 (73.68%)	9 (56.25%)	8 (53.33%)
		No Response	22	25	28	31	32
q3_1	Please indicate the reasons that you did not use the mobile diary app (you may mark more than one item): (choice=Lack of time)	Checked	6 (12.77%)	9 (19.15%)	2 (4.26%)	4 (8.51%)	4 (8.51%)
		Unchecked	41 (87.23%)	38 (80.85%)	45 (95.74%)	43 (91.49%)	43 (91.49%)
q3_2	Please indicate the reasons that you did not use the mobile diary app (you may mark more than one item): (choice=Forgot about the app)	Checked	6 (12.77%)	7 (14.89%)	9 (19.15%)	10 (21.28%)	5 (10.64%)
		Unchecked	41 (87.23%)	40 (85.11%)	38 (80.85%)	37 (78.72%)	42 (89.36%)
q3_3	Please indicate the reasons that you did not use the mobile diary app (you may mark more than one item): (choice=Too difficult to use the app)	Checked		1 (2.13%)			1 (2.13%)
		Unchecked	47 (100.00%)	46 (97.87%)	47 (100.00%)	47 (100.00%)	46 (97.87%)
q3_4	Please indicate the reasons that you did not use the mobile diary app (you may mark more than one item): (choice=Did not think the app was useful)	Checked	5 (10.64%)	4 (8.51%)	1 (2.13%)		2 (4.26%)
		Unchecked	42 (89.36%)	43 (91.49%)	46 (97.87%)	47 (100.00%)	45 (95.74%)

	Question	Answer Options	August Count (Percent)	September Count (Percent)	October Count (Percent)	November Count (Percent)	December Count (Percent)
q3_5	Please indicate the reasons that you did not take notes on daily student performance (you may mark more than one item): (choice=Comments on student evaluations are not a priority for me)	Checked		1 (2.13%)			
	Please indicate the reasons that you did not use the mobile diary app (you may mark more than one item): (choice=Comments on student evaluations are not a priority for me)	Unchecked	47 (100.00%)	46 (97.87%)	47 (100.00%)	47 (100.00%)	47 (100.00%)
q3_6	Please indicate the reasons that you did not use the mobile diary app (you may mark more than one item): (choice=Other)	Checked	6 (12.77%)	6 (12.77%)	7 (14.89%)	5 (10.64%)	3 (6.38%)
		Unchecked	41 (87.23%)	41 (87.23%)	40 (85.11%)	42 (89.36%)	44 (93.62%)
q3_7	Please indicate the reasons that you did not use the mobile diary app (you may mark more than one item): (choice=None)	Checked	5 (10.64%)	1 (2.13%)			3 (6.38%)
		Unchecked	42 (89.36%)	46 (97.87%)	47 (100.00%)	47 (100.00%)	44 (93.62%)
q4	How comfortable are you when using the mobile diary app?	Comfortable			2 (10.53%)	4 (25.00%)	3 (20.00%)
		Neutral			14 (73.68%)	8 (50.00%)	8 (53.33%)
		No Response			28	31	32
		Uncomfortable					1 (6.67%)
		Very comfortable			1 (5.26%)		2 (13.33%)
		Very uncomfortable			2 (10.53%)	4 (25.00%)	1 (6.67%)

	Question	Answer Options	August Count (Percent)	September Count (Percent)	October Count (Percent)	November Count (Percent)	December Count (Percent)
q5	How satisfied are you with the mobile diary app?	Neutral			15 (78.95%)	11 (68.75%)	11 (73.33%)
		No Response			28	31	32
		Satisfied			2 (10.53%)	4 (25.00%)	2 (13.33%)
		Very dissatisfied				1 (6.25%)	
		Very satisfied			2 (10.53%)		2 (13.33%)

Critical Thinking

The surveys were taken by participants on RedCap. Learning how to work with data directly from RedCap in SAS was new to me. This consultation work also involved editing and merging tables in SAS and utilizing proc report to produce tables which were easy to read. Because the investigator just wanted the distribution of answers for each question of their surveys, I organized the responses with a column for each month's responses. This allowed the investigator to visually note any trends. I thought this would be the best way to present the data in case she noticed a response seemed to greatly change over time. If requested, further analysis to determine if any responses distributions were significantly different at different time points could be conducted.

Summary

This investigator had multiple surveys across several months. I used data from their RedCap surveys to produce tables which showed the response distribution for each question. Although this is only one way to present survey results, I think it was very useful practice. Surveys seem to be a very common and it is likely I will need to present survey results again. This consultation also allowed me to practice merging tables correctly and utilizing the proc report function. The investigator was able to examine the results of her surveys from my report and utilize them in her research.

SAS Code

```
/*bring in SAS dataset from REDCap*/
options nofmterr;
libname app "S:\ASL Current\ASL Requests_Current\rmto225.1165\Data\Raw Data";

/*creating temporary dataset*/
data diary_app;
      set app.redcap;
run;
/*proc contents data=diary_app varnum;*/
/*run;*/

/*reading in Redcap formats*/
proc format;
      value my_first_instrument_complete_ 0='Incomplete' 1='Unverified'
            2='Complete';
      value evals_per_6_mo_ 1='0 evaluations' 2='1-2 evaluations'
            3='3-4 evaluations' 4='5 or more evaluations' 99='No Response';
      value self_rated_quantity_ 1='No comments' 2='A few phrases'
            3='Several sentences' 4='At least a paragraph' 99='No Response';
      value self_rated_quality_ 1='Very poor' 2='Poor'
            3='Adequate' 4='Good'
            5='Very good' 99='No Response';
      value reasons_no_comments___1_ 0='Unchecked' 1='Checked' 99='No
Response';
      value reasons_no_comments___2_ 0='Unchecked' 1='Checked' 99='No
Response';
      value reasons_no_comments___3_ 0='Unchecked' 1='Checked' 99='No
Response';
      value reasons_no_comments___4_ 0='Unchecked' 1='Checked' 99='No
Response';
      value reasons_no_comments___5_ 0='Unchecked' 1='Checked' 99='No
Response';
      value type_of_device_ 1='iPhone' 2='Android phone'
            3='Other' 99='No Response';
      value comfort_with_apps_ 1='Very uncomfortable' 2='Uncomfortable'
            3='Neutral' 4='Comfortable'
            5='Very comfortable' 99='No Response';
      value frequency_app_use_ 1='Almost never' 2='Monthly'
            3='Weekly' 4='Daily'
            5='Multiple times per day' 99='No Response';
      value july_baseline_survey_complete_ 0='Incomplete' 1='Unverified'
            2='Complete';
      value aug_notes_per_day_ 1='Never' 2='1-2 days'
            3='3-4 days' 4='5 or more days' 99='No Response';
      value aug_use_daily_notes_ 1='Never' 2='A few times'
            3='Most of the time' 4='Always'
            5='I did not complete any student evaluations in the past month.'
99='No Response';
      value aug_reasons_why_not___1_ 0='Unchecked' 1='Checked' 99='No
Response';
      value aug_reasons_why_not___2_ 0='Unchecked' 1='Checked' 99='No
Response';
      value aug_reasons_why_not___3_ 0='Unchecked' 1='Checked' 99='No
Response';
```

```
      value aug_reasons_why_not___4_ 0='Unchecked' 1='Checked' 99='No
Response';
      value aug_reasons_why_not___5_ 0='Unchecked' 1='Checked' 99='No
Response';
      value aug_reasons_why_not___6_ 0='Unchecked' 1='Checked' 99='No
Response';
      value aug_reasons_why_not___7_ 0='Unchecked' 1='Checked' 99='No
Response';
      value august_survey_complete_ 0='Incomplete' 1='Unverified'
            2='Complete' 99='No Response';
      value sept_notes_per_day_ 1='Never' 2='1-2 days'
            3='3-4 days' 4='5 or more days' 99='No Response';
      value sept_use_daily_notes_ 1='Never' 2='A few times'
            3='Most of the time' 4='Always'
            5='I did not complete any student evaluations in the past month.'
99='No Response';
      value sept_reasons_why_not___1_ 0='Unchecked' 1='Checked' 99='No
Response';
      value sept_reasons_why_not___2_ 0='Unchecked' 1='Checked' 99='No
Response';
      value sept_reasons_why_not___3_ 0='Unchecked' 1='Checked' 99='No
Response';
      value sept_reasons_why_not___4_ 0='Unchecked' 1='Checked' 99='No
Response';
      value sept_reasons_why_not___5_ 0='Unchecked' 1='Checked' 99='No
Response';
      value sept_reasons_why_not___6_ 0='Unchecked' 1='Checked' 99='No
Response';
      value sept_reasons_why_not___7_ 0='Unchecked' 1='Checked' 99='No
Response';
      value september_survey_complete_ 0='Incomplete' 1='Unverified'
            2='Complete';
      value oct_daily_app_use_ 1='Never' 2='1-2 days'
            3='3-4 days' 4='5 or more days' 99='No Response';
      value oct_use_app_for_evals_ 1='Never' 2='A few times'
            3='Most of the time' 4='Always'
            5='I did not complete any student evaluations in the past month.'
99='No Response';
      value oct_reasons_why_not___1_ 0='Unchecked' 1='Checked' 99='No
Response';
      value oct_reasons_why_not___2_ 0='Unchecked' 1='Checked' 99='No
Response';
      value oct_reasons_why_not___3_ 0='Unchecked' 1='Checked' 99='No
Response';
      value oct_reasons_why_not___4_ 0='Unchecked' 1='Checked' 99='No
Response';
      value oct_reasons_why_not___5_ 0='Unchecked' 1='Checked' 99='No
Response';
      value oct_reasons_why_not___6_ 0='Unchecked' 1='Checked' 99='No
Response';
      value oct_reasons_why_not___7_ 0='Unchecked' 1='Checked' 99='No
Response';
      value oct_app_comfort_ 1='Very uncomfortable' 2='Uncomfortable'
            3='Neutral' 4='Comfortable'
            5='Very comfortable' 99='No Response';
      value oct_app_satisfaction_ 1='Very dissatisfied' 2='Dissatisfied'
            3='Neutral' 4='Satisfied'
```

```
          5='Very satisfied' 99='No Response';
     value october_survey_complete_ 0='Incomplete' 1='Unverified'
          2='Complete';
     value nov_daily_app_use_ 1='Never' 2='1-2 days'
          3='3-4 days' 4='5 or more days' 99='No Response';
     value nov_use_app_for_evals_ 1='Never' 2='A few times'
          3='Most of the time' 4='Always'
          5='I did not complete any student evaluations in the past month.'
99='No Response';
     value nov_reasons_why_not___1_ 0='Unchecked' 1='Checked' 99='No
Response';
     value nov_reasons_why_not___2_ 0='Unchecked' 1='Checked' 99='No
Response';
     value nov_reasons_why_not___3_ 0='Unchecked' 1='Checked' 99='No
Response';
     value nov_reasons_why_not___4_ 0='Unchecked' 1='Checked' 99='No
Response';
     value nov_reasons_why_not___5_ 0='Unchecked' 1='Checked' 99='No
Response';
     value nov_reasons_why_not___6_ 0='Unchecked' 1='Checked' 99='No
Response';
     value nov_reasons_why_not___7_ 0='Unchecked' 1='Checked' 99='No
Response';
     value nov_app_comfort_ 1='Very uncomfortable' 2='Uncomfortable'
          3='Neutral' 4='Comfortable'
          5='Very comfortable' 99='No Response';
     value nov_app_satisfaction_ 1='Very dissatisfied' 2='Dissatisfied'
          3='Neutral' 4='Satisfied'
          5='Very satisfied' 99='No Response';
     value november_survey_complete_ 0='Incomplete' 1='Unverified'
          2='Complete';
     value evals_per_6_mo_v2_ 1='0 evaluations' 2='1-2 evaluations'
          3='3-4 evaluations' 4='5 or more evaluations' 99='No Response';
     value self_rated_quantity_v2_ 1='No comments' 2='A few phrases'
          3='Several sentences' 4='At least a paragraph' 99='No Response';
     value self_rated_quality_v2_ 1='Very poor' 2='Poor'
          3='Adequate' 4='Good'
          5='Very good' 99='No Response';
     value reasons_no_comments_v2___1_ 0='Unchecked' 1='Checked' 99='No
Response';
     value reasons_no_comments_v2___2_ 0='Unchecked' 1='Checked' 99='No
Response';
     value reasons_no_comments_v2___3_ 0='Unchecked' 1='Checked' 99='No
Response';
     value reasons_no_comments_v2___4_ 0='Unchecked' 1='Checked' 99='No
Response';
     value reasons_no_comments_v2___5_ 0='Unchecked' 1='Checked' 99='No
Response';
     value type_of_device_v2_ 1='iPhone' 2='Android phone'
          3='Other' 99='No Response';
     value comfort_with_apps_v2_ 1='Very uncomfortable' 2='Uncomfortable'
          3='Neutral' 4='Comfortable'
          5='Very comfortable' 99='No Response';
     value frequency_app_use_v2_ 1='Almost never' 2='Monthly'
          3='Weekly' 4='Daily'
          5='Multiple times per day' 99='No Response';
     value dec_daily_app_use_ 1='Never' 2='1-2 days'
```

```
            3='3-4 days' 4='5 or more days' 99='No Response';
      value dec_use_app_for_evals_ 1='Never' 2='A few times'
            3='Most of the time' 4='Always'
            5='I did not complete any student evaluations in the past month.'
99='No Response';
      value dec_reasons_why_not___1_ 0='Unchecked' 1='Checked' 99='No
Response';
      value dec_reasons_why_not___2_ 0='Unchecked' 1='Checked' 99='No
Response';
      value dec_reasons_why_not___3_ 0='Unchecked' 1='Checked' 99='No
Response';
      value dec_reasons_why_not___4_ 0='Unchecked' 1='Checked' 99='No
Response';
      value dec_reasons_why_not___5_ 0='Unchecked' 1='Checked' 99='No
Response';
      value dec_reasons_why_not___6_ 0='Unchecked' 1='Checked' 99='No
Response';
      value dec_reasons_why_not___7_ 0='Unchecked' 1='Checked' 99='No
Response';
      value dec_app_comfort_ 1='Very uncomfortable' 2='Uncomfortable'
            3='Neutral' 4='Comfortable'
            5='Very comfortable' 99='No Response';
      value dec_app_satisfaction_ 1='Very dissatisfied' 2='Dissatisfied'
            3='Neutral' 4='Satisfied'
            5='Very satisfied' 99='No Response';
      value december_exit_survey_complete_ 0='Incomplete' 1='Unverified'
            2='Complete';
      run;

data diary_app;
      set diary_app;

      format my_first_instrument_complete my_first_instrument_complete_.;
      format evals_per_6_mo evals_per_6_mo_.;
      format self_rated_quantity self_rated_quantity_.;
      format self_rated_quality self_rated_quality_.;
      format reasons_no_comments___1 reasons_no_comments___1_.;
      format reasons_no_comments___2 reasons_no_comments___2_.;
      format reasons_no_comments___3 reasons_no_comments___3_.;
      format reasons_no_comments___4 reasons_no_comments___4_.;
      format reasons_no_comments___5 reasons_no_comments___5_.;
      format type_of_device type_of_device_.;
      format comfort_with_apps comfort_with_apps_.;
      format frequency_app_use frequency_app_use_.;
      format july_baseline_survey_complete july_baseline_survey_complete_.;
      format aug_notes_per_day aug_notes_per_day_.;
      format aug_use_daily_notes aug_use_daily_notes_.;
      format aug_reasons_why_not___1 aug_reasons_why_not___1_.;
      format aug_reasons_why_not___2 aug_reasons_why_not___2_.;
      format aug_reasons_why_not___3 aug_reasons_why_not___3_.;
      format aug_reasons_why_not___4 aug_reasons_why_not___4_.;
      format aug_reasons_why_not___5 aug_reasons_why_not___5_.;
      format aug_reasons_why_not___6 aug_reasons_why_not___6_.;
      format aug_reasons_why_not___7 aug_reasons_why_not___7_.;
      format august_survey_complete august_survey_complete_.;
      format sept_notes_per_day sept_notes_per_day_.;
      format sept_use_daily_notes sept_use_daily_notes_.;
```

```
format sept_reasons_why_not___1 sept_reasons_why_not___1_.;
format sept_reasons_why_not___2 sept_reasons_why_not___2_.;
format sept_reasons_why_not___3 sept_reasons_why_not___3_.;
format sept_reasons_why_not___4 sept_reasons_why_not___4_.;
format sept_reasons_why_not___5 sept_reasons_why_not___5_.;
format sept_reasons_why_not___6 sept_reasons_why_not___6_.;
format sept_reasons_why_not___7 sept_reasons_why_not___7_.;
format september_survey_complete september_survey_complete_.;
format oct_daily_app_use oct_daily_app_use_.;
format oct_use_app_for_evals oct_use_app_for_evals_.;
format oct_reasons_why_not___1 oct_reasons_why_not___1_.;
format oct_reasons_why_not___2 oct_reasons_why_not___2_.;
format oct_reasons_why_not___3 oct_reasons_why_not___3_.;
format oct_reasons_why_not___4 oct_reasons_why_not___4_.;
format oct_reasons_why_not___5 oct_reasons_why_not___5_.;
format oct_reasons_why_not___6 oct_reasons_why_not___6_.;
format oct_reasons_why_not___7 oct_reasons_why_not___7_.;
format oct_app_comfort oct_app_comfort_.;
format oct_app_satisfaction oct_app_satisfaction_.;
format october_survey_complete october_survey_complete_.;
format nov_daily_app_use nov_daily_app_use_.;
format nov_use_app_for_evals nov_use_app_for_evals_.;
format nov_reasons_why_not___1 nov_reasons_why_not___1_.;
format nov_reasons_why_not___2 nov_reasons_why_not___2_.;
format nov_reasons_why_not___3 nov_reasons_why_not___3_.;
format nov_reasons_why_not___4 nov_reasons_why_not___4_.;
format nov_reasons_why_not___5 nov_reasons_why_not___5_.;
format nov_reasons_why_not___6 nov_reasons_why_not___6_.;
format nov_reasons_why_not___7 nov_reasons_why_not___7_.;
format nov_app_comfort nov_app_comfort_.;
format nov_app_satisfaction nov_app_satisfaction_.;
format november_survey_complete november_survey_complete_.;
format evals_per_6_mo_v2 evals_per_6_mo_v2_.;
format self_rated_quantity_v2 self_rated_quantity_v2_.;
format self_rated_quality_v2 self_rated_quality_v2_.;
format reasons_no_comments_v2___1 reasons_no_comments_v2___1_.;
format reasons_no_comments_v2___2 reasons_no_comments_v2___2_.;
format reasons_no_comments_v2___3 reasons_no_comments_v2___3_.;
format reasons_no_comments_v2___4 reasons_no_comments_v2___4_.;
format reasons_no_comments_v2___5 reasons_no_comments_v2___5_.;
format type_of_device_v2 type_of_device_v2_.;
format comfort_with_apps_v2 comfort_with_apps_v2_.;
format frequency_app_use_v2 frequency_app_use_v2_.;
format dec_daily_app_use dec_daily_app_use_.;
format dec_use_app_for_evals dec_use_app_for_evals_.;
format dec_reasons_why_not___1 dec_reasons_why_not___1_.;
format dec_reasons_why_not___2 dec_reasons_why_not___2_.;
format dec_reasons_why_not___3 dec_reasons_why_not___3_.;
format dec_reasons_why_not___4 dec_reasons_why_not___4_.;
format dec_reasons_why_not___5 dec_reasons_why_not___5_.;
format dec_reasons_why_not___6 dec_reasons_why_not___6_.;
format dec_reasons_why_not___7 dec_reasons_why_not___7_.;
format dec_app_comfort dec_app_comfort_.;
format dec_app_satisfaction dec_app_satisfaction_.;
format december_exit_survey_complete december_exit_survey_complete_.;
run;
```

```
proc freq data=diary_app;
      tables evals_per_6_mo;
run;

*%let varfreq=evals_per_6_mo;
*%let ques=q1;
%macro surveys(data,varfreq,ques,format);
proc transpose data=&data (obs=1 keep=&varfreq) out=test;
run;
data _null_;
      set test;
      call symput("varlabel",trim(_label_));
run;
proc freq data=&data;
      tables &varfreq / out=freq_&varfreq;
run;
proc stdize data=freq_&varfreq missing=99 out=fix_&varfreq reponly;
run;
proc sort data=fix_&varfreq;
      by &varfreq;
run;
data fix2_&varfreq;
      set fix_&varfreq;
      length varlabs $500;
      varlabs="&varlabel";
      length question $10;
      question="&ques";
      length options $500;
      options=put(&varfreq,&format);
      length nper $100;
      if &varfreq ne 99 then do;
            nper=compress(put(COUNT,8.0))||"
("||compress(put(PERCENT,8.2))||"%)";
      end;
      else do;
            nper=compress(put(COUNT,8.0));
      end;
      drop count percent &varfreq;
run;
%mend;
/*July Baseline*/
%surveys(diary_app,evals_per_6_mo,q1,evals_per_6_mo_.);
%surveys(diary_app,self_rated_quantity,q2,self_rated_quantity_.);
%surveys(diary_app,self_rated_quality,q3,self_rated_quality_.);
%surveys(diary_app,reasons_no_comments___1,q4_1,reasons_no_comments___1_.);
%surveys(diary_app,reasons_no_comments___2,q4_2,reasons_no_comments___2_.);
%surveys(diary_app,reasons_no_comments___3,q4_3,reasons_no_comments___3_.);
%surveys(diary_app,reasons_no_comments___4,q4_4,reasons_no_comments___4_.);
%surveys(diary_app,reasons_no_comments___5,q4_5,reasons_no_comments___5_.);
%surveys(diary_app,type_of_device,q5,type_of_device_.);
%surveys(diary_app,comfort_with_apps,q6,comfort_with_apps_.);
%surveys(diary_app,frequency_app_use,q7,frequency_app_use_.);

data july;
```

```
      set fix2_evals_per_6_mo fix2_self_rated_quantity
fix2_self_rated_quality fix2_reasons_no_comments___1
fix2_reasons_no_comments___2
            fix2_reasons_no_comments___3 fix2_reasons_no_comments___4
fix2_reasons_no_comments___5 fix2_type_of_device fix2_comfort_with_apps
            fix2_frequency_app_use;
run;

/*August*/
%surveys(diary_app,aug_notes_per_day,q1,aug_notes_per_day_.);
%surveys(diary_app,aug_use_daily_notes,q2,aug_use_daily_notes_.);
%surveys(diary_app,aug_reasons_why_not___1,q3_1,aug_reasons_why_not___1_.);
%surveys(diary_app,aug_reasons_why_not___2,q3_2,aug_reasons_why_not___2_.);
%surveys(diary_app,aug_reasons_why_not___3,q3_3,aug_reasons_why_not___3_.);
%surveys(diary_app,aug_reasons_why_not___4,q3_4,aug_reasons_why_not___4_.);
%surveys(diary_app,aug_reasons_why_not___5,q3_5,aug_reasons_why_not___5_.);
%surveys(diary_app,aug_reasons_why_not___6,q3_6,aug_reasons_why_not___6_.);
%surveys(diary_app,aug_reasons_why_not___7,q3_7,aug_reasons_why_not___7_.);

data august;
      set fix2_aug_notes_per_day fix2_aug_use_daily_notes
fix2_aug_reasons_why_not___1 fix2_aug_reasons_why_not___2
            fix2_aug_reasons_why_not___3 fix2_aug_reasons_why_not___4
fix2_aug_reasons_why_not___5 fix2_aug_reasons_why_not___6
            fix2_aug_reasons_why_not___7;
run;

/*September*/
%surveys(diary_app,sept_notes_per_day,q1,sept_notes_per_day_.);
%surveys(diary_app,sept_use_daily_notes,q2,sept_use_daily_notes_.);
%surveys(diary_app,sept_reasons_why_not___1,q3_1,sept_reasons_why_not___1_.);
%surveys(diary_app,sept_reasons_why_not___2,q3_2,sept_reasons_why_not___2_.);
%surveys(diary_app,sept_reasons_why_not___3,q3_3,sept_reasons_why_not___3_.);
%surveys(diary_app,sept_reasons_why_not___4,q3_4,sept_reasons_why_not___4_.);
%surveys(diary_app,sept_reasons_why_not___5,q3_5,sept_reasons_why_not___5_.);
%surveys(diary_app,sept_reasons_why_not___6,q3_6,sept_reasons_why_not___6_.);
%surveys(diary_app,sept_reasons_why_not___7,q3_7,sept_reasons_why_not___7_.);

data september;
      set fix2_sept_notes_per_day fix2_sept_use_daily_notes
fix2_sept_reasons_why_not___1 fix2_sept_reasons_why_not___2
            fix2_sept_reasons_why_not___3 fix2_sept_reasons_why_not___4
fix2_sept_reasons_why_not___5 fix2_sept_reasons_why_not___6
            fix2_sept_reasons_why_not___7;
run;

/*October*/
%surveys(diary_app,oct_daily_app_use,q1,oct_daily_app_use_.);
%surveys(diary_app,oct_use_app_for_evals,q2,oct_use_app_for_evals_.);
%surveys(diary_app,oct_reasons_why_not___1,q3_1,oct_reasons_why_not___1_.);
%surveys(diary_app,oct_reasons_why_not___2,q3_2,oct_reasons_why_not___2_.);
%surveys(diary_app,oct_reasons_why_not___3,q3_3,oct_reasons_why_not___3_.);
%surveys(diary_app,oct_reasons_why_not___4,q3_4,oct_reasons_why_not___4_.);
%surveys(diary_app,oct_reasons_why_not___5,q3_5,oct_reasons_why_not___5_.);
%surveys(diary_app,oct_reasons_why_not___6,q3_6,oct_reasons_why_not___6_.);
%surveys(diary_app,oct_reasons_why_not___7,q3_7,oct_reasons_why_not___7_.);
%surveys(diary_app,oct_app_comfort,q4,oct_app_comfort_.);
```

```
%surveys(diary_app,oct_app_satisfaction,q5,oct_app_satisfaction_.);


data october;
      set fix2_oct_daily_app_use fix2_oct_use_app_for_evals
fix2_oct_reasons_why_not___1 fix2_oct_reasons_why_not___2
            fix2_oct_reasons_why_not___3 fix2_oct_reasons_why_not___4
fix2_oct_reasons_why_not___5 fix2_oct_reasons_why_not___6
            fix2_oct_reasons_why_not___7 fix2_oct_app_comfort
fix2_oct_app_satisfaction;
run;

/*November*/
%surveys(diary_app,nov_daily_app_use,q1,nov_daily_app_use_.);
%surveys(diary_app,nov_use_app_for_evals,q2,nov_use_app_for_evals_.);
%surveys(diary_app,nov_reasons_why_not___1,q3_1,nov_reasons_why_not___1_.);
%surveys(diary_app,nov_reasons_why_not___2,q3_2,nov_reasons_why_not___2_.);
%surveys(diary_app,nov_reasons_why_not___3,q3_3,nov_reasons_why_not___3_.);
%surveys(diary_app,nov_reasons_why_not___4,q3_4,nov_reasons_why_not___4_.);
%surveys(diary_app,nov_reasons_why_not___5,q3_5,nov_reasons_why_not___5_.);
%surveys(diary_app,nov_reasons_why_not___6,q3_6,nov_reasons_why_not___6_.);
%surveys(diary_app,nov_reasons_why_not___7,q3_7,nov_reasons_why_not___7_.);
%surveys(diary_app,nov_app_comfort,q4,nov_app_comfort_.);
%surveys(diary_app,nov_app_satisfaction,q5,nov_app_satisfaction_.);


data november;
      set fix2_nov_daily_app_use fix2_nov_use_app_for_evals
fix2_nov_reasons_why_not___1 fix2_nov_reasons_why_not___2
            fix2_nov_reasons_why_not___3 fix2_nov_reasons_why_not___4
fix2_nov_reasons_why_not___5 fix2_nov_reasons_why_not___6
            fix2_nov_reasons_why_not___7 fix2_nov_app_comfort
fix2_nov_app_satisfaction;
run;

/*December*/
%surveys(diary_app,dec_daily_app_use,q1,dec_daily_app_use_.);
%surveys(diary_app,dec_use_app_for_evals,q2,dec_use_app_for_evals_.);
%surveys(diary_app,dec_reasons_why_not___1,q3_1,dec_reasons_why_not___1_.);
%surveys(diary_app,dec_reasons_why_not___2,q3_2,dec_reasons_why_not___2_.);
%surveys(diary_app,dec_reasons_why_not___3,q3_3,dec_reasons_why_not___3_.);
%surveys(diary_app,dec_reasons_why_not___4,q3_4,dec_reasons_why_not___4_.);
%surveys(diary_app,dec_reasons_why_not___5,q3_5,dec_reasons_why_not___5_.);
%surveys(diary_app,dec_reasons_why_not___6,q3_6,dec_reasons_why_not___6_.);
%surveys(diary_app,dec_reasons_why_not___7,q3_7,dec_reasons_why_not___7_.);
%surveys(diary_app,dec_app_comfort,q4,dec_app_comfort_.);
%surveys(diary_app,dec_app_satisfaction,q5,dec_app_satisfaction_.);

data december;
      set fix2_dec_daily_app_use fix2_dec_use_app_for_evals
fix2_dec_reasons_why_not___1 fix2_dec_reasons_why_not___2
            fix2_dec_reasons_why_not___3 fix2_dec_reasons_why_not___4
fix2_dec_reasons_why_not___5 fix2_dec_reasons_why_not___6
            fix2_dec_reasons_why_not___7 fix2_dec_app_comfort
fix2_dec_app_satisfaction;
run;
```

```
/*December Baseline*/
%surveys(diary_app,evals_per_6_mo_v2,q1,evals_per_6_mo_v2_.);
%surveys(diary_app,self_rated_quantity_v2,q2,self_rated_quantity_v2_.);
%surveys(diary_app,self_rated_quality_v2,q3,self_rated_quality_v2_.);
%surveys(diary_app,reasons_no_comments_v2___1,q4_1,reasons_no_comments_v2___1
_.);
%surveys(diary_app,reasons_no_comments_v2___2,q4_2,reasons_no_comments_v2___2
_.);
%surveys(diary_app,reasons_no_comments_v2___3,q4_3,reasons_no_comments_v2___3
_.);
%surveys(diary_app,reasons_no_comments_v2___4,q4_4,reasons_no_comments_v2___4
_.);
%surveys(diary_app,reasons_no_comments_v2___5,q4_5,reasons_no_comments_v2___5
_.);
%surveys(diary_app,type_of_device_v2,q5,type_of_device_v2_.);
%surveys(diary_app,comfort_with_apps_v2,q6,comfort_with_apps_v2_.);
%surveys(diary_app,frequency_app_use_v2,q7,frequency_app_use_v2_.);

data decemberbase;
      set fix2_evals_per_6_mo_v2 fix2_self_rated_quantity_v2
fix2_self_rated_quality_v2 fix2_reasons_no_comments_v2___1
fix2_reasons_no_comments_v2___2
            fix2_reasons_no_comments_v2___3 fix2_reasons_no_comments_v2___4
fix2_reasons_no_comments_v2___5 fix2_type_of_device_v2
fix2_comfort_with_apps_v2
            fix2_frequency_app_use_v2;
run;

/*Distribution table july december baseline*/
*copy nper column to rename;
data decemberbase2;
set decemberbase;
december=nper;
drop nper;
run;
data julybase2;
set july;
july=nper;
drop nper;
run;
*sort;
proc sort data=julybase2;
by question options;
run;
proc sort data=decemberbase2;
by question options;
run;
*merge;
data baseline;
merge julybase2 decemberbase2;
by question options;
run;

/*Distribution table aug-dec*/
data  august2;
set august;
```

```
august=nper;
drop nper;
run;
data september2;
set september;
september=nper;
drop nper;
run;
data october2;
set october;
october=nper;
drop nper;
run;
data november2;
set november;
november=nper;
drop nper;
run;
data december2;
set december;
december=nper;
drop nper;
run;
*sort;
proc sort data=august2;
by question options;
run;
proc sort data=september2;
by question options;
run;
proc sort data=october2;
by question options;
run;
proc sort data=november2;
by question options;
run;
proc sort data=december2;
by question options;
run;
*merge;
data survey;
merge august2 september2 october2 november2 december2;
by question options;
run;

/*reports*/
options nodate nonumber;
ods rtf file="S:\ASL Current\ASL
Requests_Current\rmto225.1165\Reports_Output\Survey Answer Distributions
Baseline July December.rtf";

title "Baseline Survey Results July and December";
proc report data=baseline split="|";
      columns question varlabs options july december;
      define question / group " ";
      define varlabs / group "Question" left;
      define options / display "Answer Options" left;
```

```
      define july / display "July Count (Percent)" center;
      define december / display "December Count (Percent)" center;
run;

ods rtf close;

options nodate nonumber orientation=landscape;
ods rtf file="S:\ASL Current\ASL
Requests_Current\rmto225.1165\Reports_Output\Survey Answer Distributions
August December.rtf";

title "Survey Results August-December";
proc report data=survey split="|";
      columns question varlabs options august september october november
december;
      define question / group " ";
      define varlabs / group "Question" left;
      define options / display "Answer Options" left;
      define august / display "August Count (Percent)" center;
      define september / display "September Count (Percent)" center;
      define october / display "October Count (Percent)" center;
      define november / display "November Count (Percent)" center;
      define december / display "December Count (Percent)" center;
run;

ods rtf close;
```

Project 10: Summary Statistics Consultation

Description

This was a follow-up from an investigator who requested data management and analysis from the Applied Statistics Laboratory. The investigator created a smartphone app for doctors to input medical student review into. She wanted to know if there was any improvement in quality of medical student review comparing before the app and after the introduction of the app. Specifically, she requested summary statistics of her data including means, medians, and boxplots. My work on this consultation included:

1. Updating the response means at each timepoint to include all her data. Preliminary results only included the first three months of her data, I updated it to include all six months of her data.
2. Producing the response medians of 3 variables broken down by survey location site and time point in the study.
3. Producing boxplots to visually display the medians by variable, site, and time point in the study.

Report 1: Updated Response Means at Each Timepoint

	Time	Mean (Standard Error)
CAT1	Baseline	0.34 (0.23)
	Post Intervention	0.38 (0.23)
CAT2	Baseline	2.02 (0.26)
	Post Intervention	2.13 (0.26)
CAT3	Baseline	0.74 (0.25)
	Post Intervention	0.70 (0.25)
CAT4	Baseline	0.34 (0.13)
	Post Intervention	0.27 (0.13)
CCERR	Baseline	14.59 (1.58)
	Post Intervention	14.92 (1.58)
QI	Baseline	2.69 (0.55)
	Post Intervention	2.42 (0.54)
WORDS	Baseline	52.12 (17.04)
	Post Intervention	50.19 (17.03)

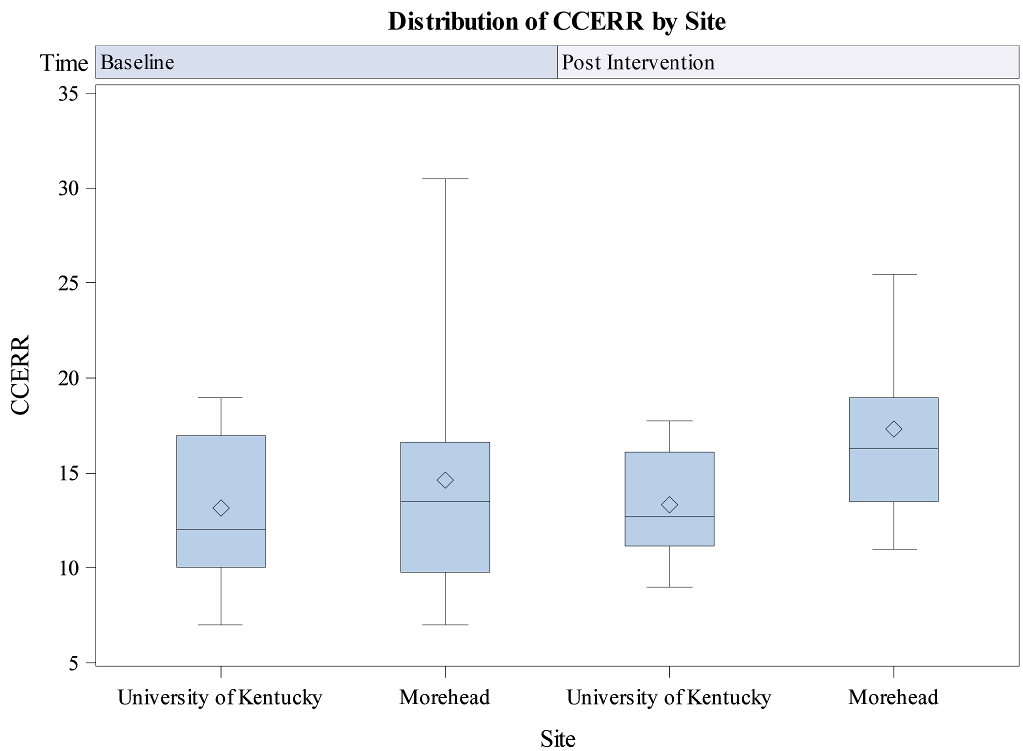
Testing Time Effect with Repeated Measures ANOVA

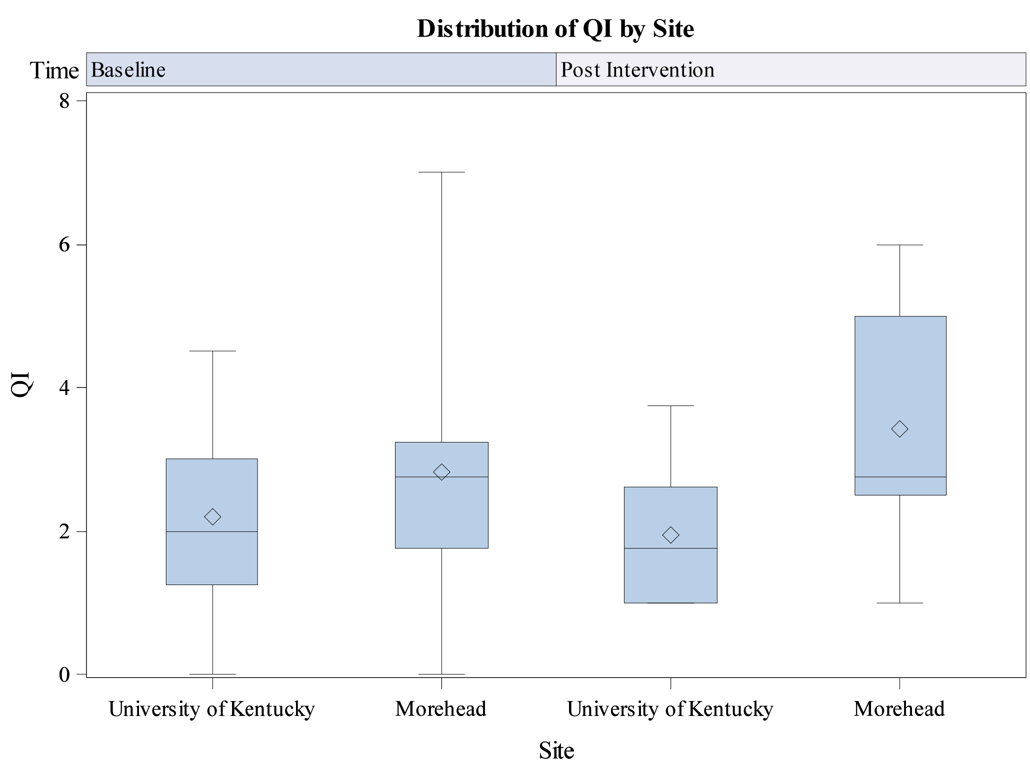
Response	Effect	Test Statistic (Degrees of Freedom), p-value
CAT1	Time	F(1,26)=0.16, p-value=0.6948
CAT2	Time	F(1,97)=1.05, p-value=0.3078
CAT3	Time	F(1,65)=0.62, p-value=0.4330
CAT4	Time	F(1,38)=1.75, p-value=0.1934
CCERR	Time	F(1,100)=0.41, p-value=0.5233
QI	Time	F(1,98)=0.34, p-value=0.5593
WORDS	Time	F(1,100)=0.07, p-value=0.7935

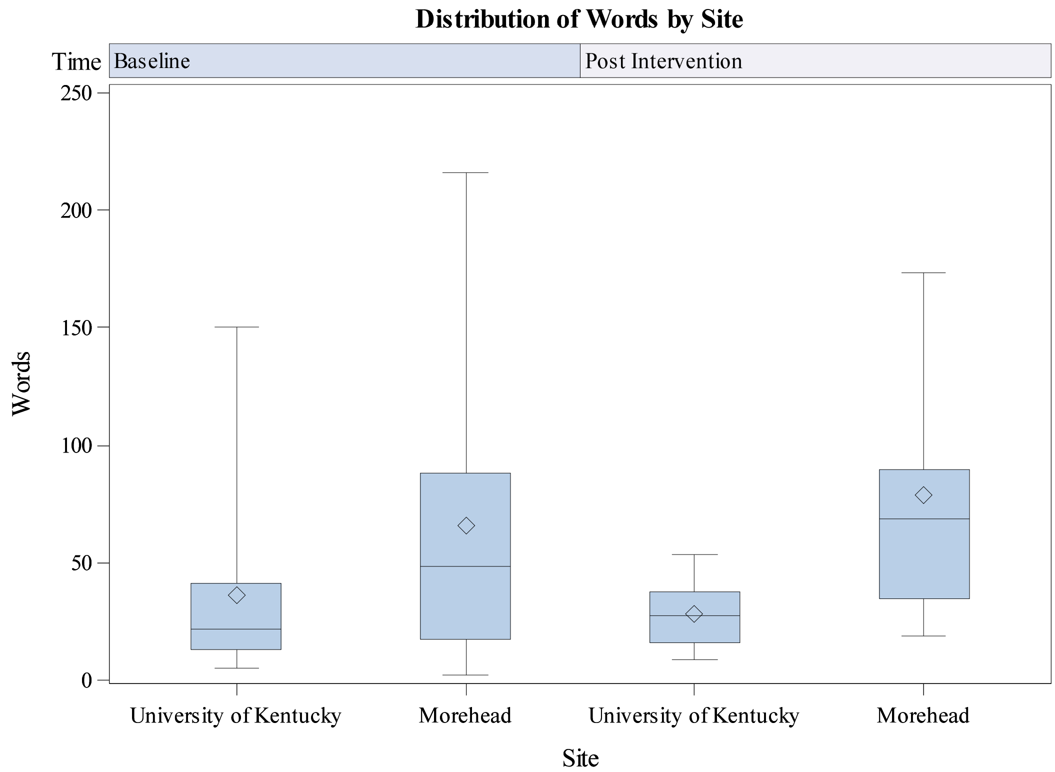
Report 2: Response Medians at Each Timepoint by Site

	Site	Time	Median (Minimum, Maximum)
CCERR	University of Kentucky	Baseline	12.00 (7.00, 19.00)
	University of Kentucky	Post Intervention	12.75 (9.00, 17.75)
	Morehead	Baseline	13.50 (7.00, 30.50)
	Morehead	Post Intervention	16.25 (11.00, 25.50)
QI	University of Kentucky	Baseline	2.00 (0.00, 4.50)
	University of Kentucky	Post Intervention	1.75 (1.00, 3.75)
	Morehead	Baseline	2.75 (0.00, 7.00)
	Morehead	Post Intervention	2.75 (1.00, 6.00)
WORDS	University of Kentucky	Baseline	22.00 (5.00, 150.00)
	University of Kentucky	Post Intervention	27.75 (8.50, 53.50)
	Morehead	Baseline	48.25 (2.00, 216.00)
	Morehead	Post Intervention	69.00 (19.00, 173.50)

Report 3: Boxplots of Medians







Critical Thinking

The first table in **Report 1** shows the means and standard errors for each response variable at each time point. When looking at the means of baseline vs. post-intervention, there appears to be little difference. The second table in **Report 1**, Testing Time Effect with Repeated Measures ANOVA, shows the results of a repeated measures ANOVA adjusting for site. Significance level is set at 0.05. Response variables had a skewed distribution, so they were log transformed so that the normality assumption of the ANOVA was met. None of the response variables were significant.

Report 2 shows the median, minimum, and maximum for each response variable at each time point by site location. In most categories, there appears to be a small increase in the median when comparing baseline vs. post intervention at each site. However, there are wide ranges between the minimum and maximums for this data.

Report 3 shows boxplots of the data in **Report 2**. There is one plot for each variable. Each plot contains the distribution of the variable at each site and timepoint. Many of the boxplots have very long whiskers, clearly displaying the large difference between the minimums and maximums.

Summary

Although this data was not statistically significant, the investigator was still able to use it in her research. The use of her smartphone application did not have a significant difference on quality of medical student review by doctors. However, the investigator found it interesting that the minimum quality and word counts of the reviews appeared to increase when looking at baseline vs. post intervention. Working on this consultation gave me the opportunity to complete multiple follow-ups with the same investigator and be a part of their research. The investigator was able to publish her results as a poster, found on the last page of this portfolio.

SAS Code: Report 1

```
/*import baseline data*/
proc import out=baseline datafile="S:\ASL Current\ASL
Requests_Current\rmto225.1165\Data\Raw Data\Evaluation011218.xlsx" dbms=xlsx
replace;
      sheet='Time X';
    getnames=yes;
run;

/*import october data*/
proc import out=october datafile="S:\ASL Current\ASL
Requests_Current\rmto225.1165\Data\Raw Data\Evaluation011218.xlsx" dbms=xlsx
replace;
      sheet='Time Y';
    getnames=yes;
run;

/*full data*/
data alldata;
      set baseline october;
run;

/*average across raters*/
data raters;
      set alldata;
      *CCERR;
      CCERR=mean(CCERR1,CCERR2,CCERR3,CCERR4);
      lnccerr=log(ccerr);

      *QI;
```

```
	QI=mean(QI1,QI2,QI3,QI4);
	lnqi=log(qi);

	*Cat1;
	Cat1=mean(Cat_1_1,Cat_1_2,Cat_1_3,Cat_1_4);
	lncat1=log(cat1);

	*Cat2;
	Cat2=mean(Cat_2_1,Cat_2_2,Cat_2_3,Cat_2_4);
	lncat2=log(cat2);

	*Cat3;
	Cat3=mean(Cat_3_1,Cat_3_2,Cat_3_3,Cat_3_4);
	lncat3=log(cat3);

	*Cat4;
	Cat4=mean(Cat_4_1,Cat_4_2,Cat_4_3,Cat_4_4);
	lncat4=log(cat4);

	*words;
	lnwords=log(words);

	keep Data_ID time site faculty words ccerr qi cat1 cat2 cat3 cat4
lnccerr lnqi lncat1 lncat2 lncat3 lncat4 lnwords;
run;

/*1 row per faculty member*/
proc sort data=raters;
	by time site faculty;
run;
proc means data=raters mean;
	by time site faculty;
	var words ccerr qi cat1 cat2 cat3 cat4 lnccerr lnqi lncat1 lncat2
lncat3 lncat4 lnwords;
	output out=perfac mean=words ccerr qi cat1 cat2 cat3 cat4 lnccerr lnqi
lncat1 lncat2 lncat3 lncat4 lnwords;
run;

/*macro for tables*/
%macro tabs(yvar,yvar2);
proc glimmix data=raters;
	class time site faculty;
	model &yvar = time;
	random site;
	random _residual_ / subject=faculty type=cs;
	lsmeans time / pdiff;
	ods output lsmeans=ls_&yvar tests3=type3_&yvar;
run;
data t3_&yvar;
	set type3_&yvar;
	length varname $100;
	var="&yvar";
	varname=upcase(var);
	length varname2 $100;
	var2="&yvar2";
	varname2=upcase(var2);
	length test_stat $100;
```

```
      test_stat="F("||compress(put(NumDF,8.0))||","||compress(put(DenDF,8.0))
||")="||compress(put(FValue,8.2))||", p-
value="||compress(put(ProbF,pvalue6.4));
run;
data means_&yvar;
      set ls_&yvar;
      length varname $100;
      var="&yvar";
      varname=upcase(var);
      length varname2 $100;
      var2="&yvar2";
      varname2=upcase(var2);
      length mse $100;
      mse=compress(put(estimate,8.2))||" ("||compress(put(stderr,8.2))||")";
      keep varname varname2 time mse;
run;
data graph_&yvar;
      set ls_&yvar;
      lower=estimate-stderr;
      upper=estimate+stderr;
run;
%mend;
%tabs(ccerr,ccerr);
%tabs(qi,qi);
%tabs(cat1,cat1);
%tabs(cat2,cat2);
%tabs(cat3,cat3);
%tabs(cat4,cat4);
%tabs(words,words);
%tabs(lnccerr,ccerr);
%tabs(lnqi,qi);
%tabs(lncat1,cat1);
%tabs(lncat2,cat2);
%tabs(lncat3,cat3);
%tabs(lncat4,cat4);
%tabs(lnwords,words);


/*anova results*/
data anova;
      set t3_ccerr t3_qi t3_cat1 t3_cat2 t3_cat3 t3_cat4 t3_words;
run;
data lnanova;
      set t3_lnccerr t3_lnqi t3_lncat1 t3_lncat2 t3_lncat3 t3_lncat4
t3_lnwords;
run;


/*means by time*/
data means;
      set means_ccerr means_qi means_cat1 means_cat2 means_cat3 means_cat4
means_words;
run;
data lnmeans;
      set means_lnccerr means_lnqi means_lncat1 means_lncat2 means_lncat3
means_lncat4 means_lnwords;
run;

proc format;
```

```
      value $time "X"="Baseline"
                  "Y"="Post Intervention";
run;

/*macro for graphs*/
%macro graph(yvar,lab);
title "Mean &lab by Time";
proc sgplot data=graph_&yvar noautolegend;
      scatter x=time y=estimate / yerrorlower=lower yerrorupper=upper
markerattrs=(color=blue symbol=circlefilled);
      series x=time y=estimate / lineattrs=(color=blue pattern=2);
      yaxis label="Mean &lab";
      format time $time.;
run;
title;
%mend;

options nodate nonumber;
ods rtf file="S:\ASL Current\ASL
Requests_Current\rmto225.1165\Reports_Output\Results Baseline and October_
Updated 011718_2.rtf";
title "Testing Time Effect with Repeated Measures ANOVA";
proc report data=lnanova split="|";
      columns varname2 effect test_stat;
      define varname2 / group "Response" left;
      define effect / display "Effect" left;
      define test_stat / display "Test Statistic (Degrees of Freedom), p-
value" center;
run;
ods text="All analyses were completed in SAS 9.4 (SAS Institute Inc., Cary,
NC, USA).";
ods text="The table above shows the results of a repeated measures ANOVA
adjusting for site.  Significance level is set at 0.05.  Response variables
had a skewed distribution, so they were log transformed so that the normality
assumption of the ANOVA was met.  Currently, none of the response variables
are significant.";
title;
title "Response Means at each Time Point";
proc report data=means split="|";
      columns varname time mse;
      define varname / group " " left;
      define time / display "Time" left format=$time.;
      define mse / display "Mean (Standard Error)" center;
run;
ods text="The table above show the means and standard errors for each
response variable at each time point.  Graphs of these values are in the
pages that follow.";
title;
ods rtf close;
```

SAS Code: Report 2 and 3

```
/*import baseline data*/
proc import out=pre datafile="S:\ASL Current\ASL
Requests_Current\rmto225.1165\Data\Raw Data\Evaluation011218.xlsx" dbms=xlsx
replace;
      sheet='Time X';
    getnames=yes;
run;
/*import post data*/
proc import out=post datafile="S:\ASL Current\ASL
Requests_Current\rmto225.1165\Data\Raw Data\Evaluation011218.xlsx" dbms=xlsx
replace;
      sheet='Time Y';
    getnames=yes;
run;
/*full data*/
data all;
      set pre post;
run;

/*median across raters*/
data raters;
      set all;
      *CCERR;
      CCERR=median(CCERR1,CCERR2,CCERR3,CCERR4);
      *QI;
      QI=median(QI1,QI2,QI3,QI4);
      *Cat1;
      Cat1=median(Cat_1_1,Cat_1_2,Cat_1_3,Cat_1_4);
      *Cat2;
      Cat2=median(Cat_2_1,Cat_2_2,Cat_2_3,Cat_2_4);
      *Cat3;
      Cat3=median(Cat_3_1,Cat_3_2,Cat_3_3,Cat_3_4);
      *Cat4;
      Cat4=median(Cat_4_1,Cat_4_2,Cat_4_3,Cat_4_4);
      *words;
      Words=words;
      keep Data_ID time site faculty CCERR QI Words;
run;

/*1 row per faculty member-median*/
proc sort data=raters;
      by time site faculty;
run;
proc means data=raters median min max;
      by time site faculty;
      var CCERR QI Words;
      output out=medminmax median=CCERR QI Words ; *min=CCERRmin QImin
max=CCERRmax QImax;
run;
data medminmax;
set medminmax;
drop _TYPE_ _FREQ_;
run;

/*median of site A time X- the median ccerr of site a at time x was (median
of these ccerr values). min was (min of these values)*/
proc sort data=medminmax;
```

```
      by time site faculty;
run;
proc means data=medminmax median min max;
by time site;
var ccerr qi Words;
output out=sitetime median=CCERR QI Words min=CCERRmin QImin Wordsmin
max=CCERRmax QImax Wordsmax;
run;
data sitetime;
set sitetime;
drop _TYPE_ _FREQ_;
run;

*transform sitetime- one variable for med min and max- mmm;
*ccerr;
data mmm_ccerr;
set sitetime;
length varname $100;
var="CCERR";
varname=upcase(var);
length varname2 $100;
var2="ccerr";
varname2=upcase(var2);
length mmm $100;
mmm=compress(put(ccerr,8.2))||" ("||compress(put(ccerrmin,8.2))||",
"||compress(put(ccerrmax,8.2))||")";
keep varname site time mmm;
run;
*qi;
data mmm_qi;
set sitetime;
length varname $100;
var="QI";
varname=upcase(var);
length varname2 $100;
var2="qi";
varname2=upcase(var2);
length mmm $100;
mmm=compress(put(qi,8.2))||" ("||compress(put(qimin,8.2))||",
"||compress(put(qimax,8.2))||")";
keep varname site time mmm;
run;
*words;
data mmm_words;
set sitetime;
length varname $100;
var="Words";
varname=upcase(var);
length varname2 $100;
var2="words";
varname2=upcase(var2);
length mmm $100;
mmm=compress(put(Words,8.2))||" ("||compress(put(Wordsmin,8.2))||",
"||compress(put(Wordsmax,8.2))||")";
keep varname site time mmm;
run;

/*med min max by time*/
data mmm;
```

```
set mmm_ccerr mmm_qi mmm_words;
run;
proc format;
      value $time "X"="Baseline"
                        "Y"="Post Intervention";
      value $site "A"="University of Kentucky"
                        "B"="Morehead";
run;
proc sort data=mmm;
by varname site time;
run;

/*report*/
*output these values and format to make table- see setup;
options nodate nonumber;
ods rtf file="S:\ASL Current\ASL
Requests_Current\rmto225.1165\Reports_Output\Descriptive Results
Medians2.rtf";
title "Response Medians at Each Timepoint by Site";
proc report data=mmm split="|";
      columns varname site time mmm ;
      define varname / group " " left;
      define site / display "Site" left format=$site. ;
      define time / display "Time" left format=$time.;
      define mmm / display "Median (Minimum, Maximum)" center;
run;
ods text="All analyses were completed in SAS 9.4 (SAS Institute Inc., Cary,
NC, USA).";
ods text="The table above show the median, minimum and maximum for each
response variable at each time point by site location. Boxplots of of this
data are below.";
title;

title "Boxplot of Medians for CCERR";
proc boxplot data=medminmax;
plot CCERR*site(time);
format site $site. time $time.;
run;
quit;

title "Boxplot of Medians for QI";
proc boxplot data=medminmax;
plot QI*site(time);
format site $site. time $time.;
run;
quit;

title "Boxplot of Medians for Word Count";
proc boxplot data=medminmax;
plot words*site(time);
format site $site. time $time.;
run;
quit;

ods rtf close;
```

Use of a Mobile Diary App to Influence Narrative Feedback on Medical Student Evaluations

Rebecca Todd, MD[1]; Jacob Martin, MD[1]; Kristen McQuerry, PhD[2]; Hannah Bellamy, BS[2]

[1]University of Kentucky College of Medicine, Department of Obstetrics and Gynecology

[2]University of Kentucky, Department of Statistics

APGO Academic Scholars and Leaders Mentors: Jody Steinauer, MD; Wendy Hansen, MD; Brady Reynolds, PhD

UK College of Medicine

Background

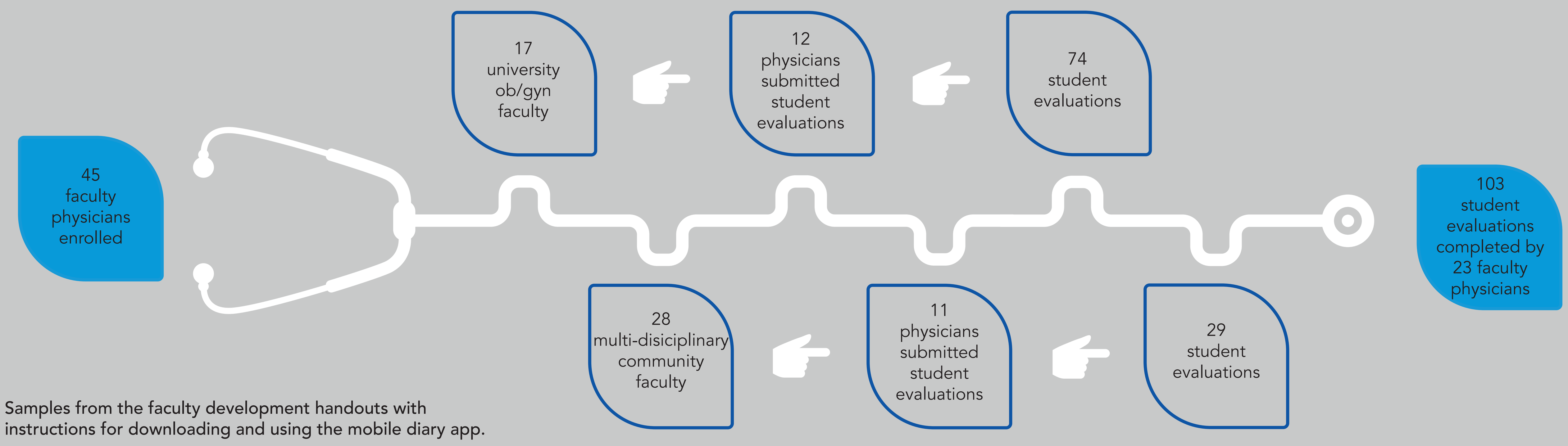
Feedback is an essential element for medical students on their path to becoming a physician. Due to the enormous impact that faculty feedback has on medical student development, it has become a focus of medical education research. Much of this research has led to improvements in the timing of feedback, the scales used, and the verbal delivery method (Bing-You R et al. Feedback for Learners in Medical Education: What is Known? A Scoping Review. Academic Medicine; 2017, DOI: 10.1097/ACM.0000000000001578). However, a common complaint from students and administration concerns the lack of consistent high quality written narrative feedback on student evaluations. This is especially problematic among volunteer community faculty who have less incentive to dedicate time and effort on the comment section of these evaluations. The knowledge of how to improve the quantity and quality of narrative feedback has been studied and shows that keeping a daily record of student performance produces ample faculty comments (Dudek N, Dojeiji S. Twelve tips for completing quality in-training evaluation reports. Medical Teacher; 2014, 1-5). Furthermore, mobile technology has had success in improving spot evaluations of student performance including the Minute Feedback System and the mini-CEX (Ferenchick, GS et al. The implementation of a mobile problem-specific electronic CEX for assessing directly observed student-patient encounters. Medical Education Online; 2010, 15: 4276 – DOI:10.3402/meo.v15i0.4276). This project aimed to consider known methods for improving narrative feedback with the use of new and popular technology.

Objective

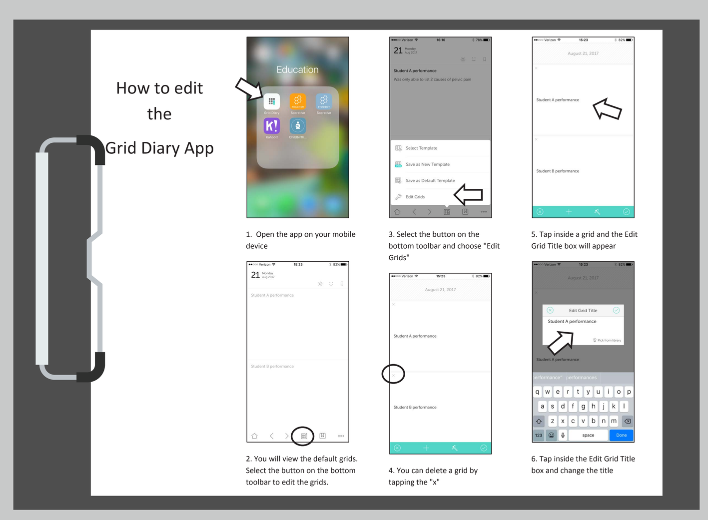
To evaluate the impact of using a mobile diary app on the quantity and quality of narrative feedback written on medical student evaluations during medicine, surgery, and ob/gyn rotations at rural community and university hospitals.

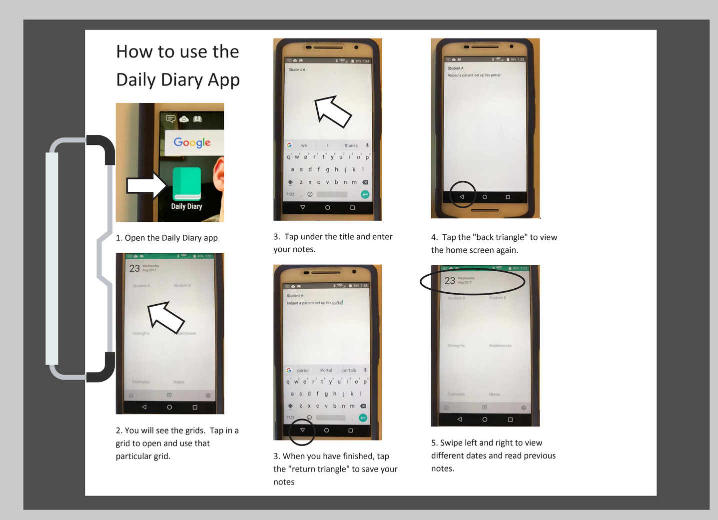
Methods

Narrative feedback written by multidisciplinary faculty members in university and rural community settings was obtained for 6 months during a cohort study. Forty-five faculty members consented to participate in the study, and 23 completed student evaluations during the project time frame. Months 1-3 served as the baseline group and months 4-6 were the comparison group. A baseline faculty development session was performed during which participants were instructed on ways to improve narrative feedback including keeping daily notes on student performance. The intervention was implemented in month 3 and consisted of a faculty development session that reiterated the recommendations given previously with the addition of training on how to use a mobile diary app (Grid Diary for iPhones and Daily Diary for Android phones) to record notes on daily student performance. De-identified evaluations written by the 23 faculty participants were obtained for months 1-6. The narrative feedback was evaluated by 4 blinded raters for quality using the Quality Improvement Instrument (QI) and the Completed Clinical Evaluation Report Rating (CCERR). Quantity was determined by a simple word count. Based on pilot data, the study had 80% power to detect a 30% increase in word count (15 words). Scores were compared using ANOVA. Anonymous monthly surveys were administered to all 45 participants via email on satisfaction with and adherence to using the mobile diary app. Descriptive statistics were used to analyze the survey data.



Samples from the faculty development handouts with instructions for downloading and using the mobile diary app.





Validated Tools for Measuring Quality

Completed Clinical Evaluation Report Rating (CCERR) developed by Dudek NL, Marks MB, Lee C & Wood T. and published as "Assessing the quality of supervisor's completed clinical evaluation reports" in Medical Education, 2008; 42: 816-822.

The Quality Improvement Instrument, designed by Maggie Bartlett et al and published in Education for Primary Care, Vol. 28, No 1.

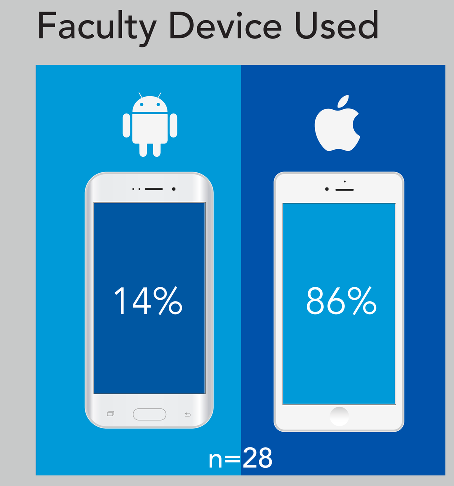
Results

Comparison of Narrative Feedback: Quality and Quantity

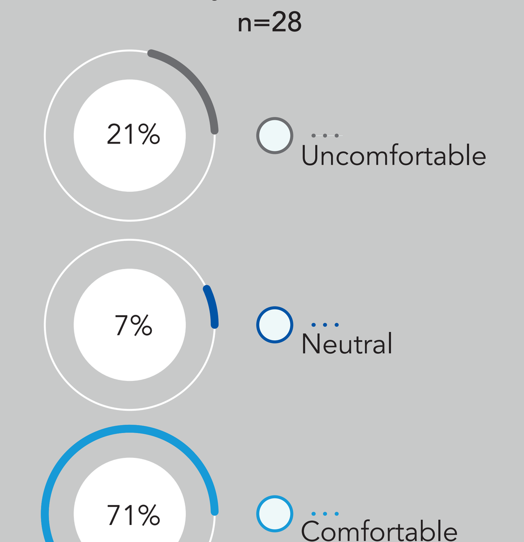
		Mean	Standard Error	P-value
QI	Baseline	2.69	0.55	
	Post Intervention	2.42	0.54	0.5593
CCERR	Baseline	14.59	1.58	
	Post Intervention	14.92	1.58	0.5233
Word Count	Baseline	52.12	17.04	
	Post Intervention	50.19	17.03	0.7935

Significance level is set at 0.05. All analyses were completed in SAS 9.4 (SAS Institute Inc., Cary, NC, USA). The table above shows the means and standard errors for each response variable at each time point. This table shows the results of a repeated measures ANOVA adjusting for site. Response variables had a skewed distribution, so they were log transformed so that the normality assumption of the ANOVA was met. The range of possible scores for QI is 0 through 7 and CCERR is 7 through 35. The ANOVA was analyzed using the average score of all four raters. Cohen's Kappa was estimated to determine the degree of agreement among two reviewers (0.01-0.20 = slight agreement, 0.21-0.40 = fair agreement, 0.41-0.60 = moderate agreement, 0.61-0.80 = substantial agreement). CCERR and QI were each grouped into categories Low, Medium, and High because these ratings had greater than six response types. For QI, all but one inter-rater agreements were moderate or substantial. For CCERR, the majority of inter-rater agreements were fair. The boxplots below show descriptive statistics including the medium, minimum, and maximum for each response variable at each time point by site location.

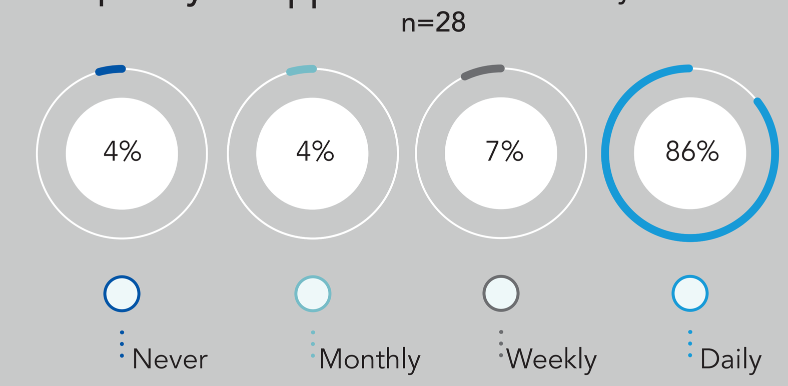
Faculty Device Used



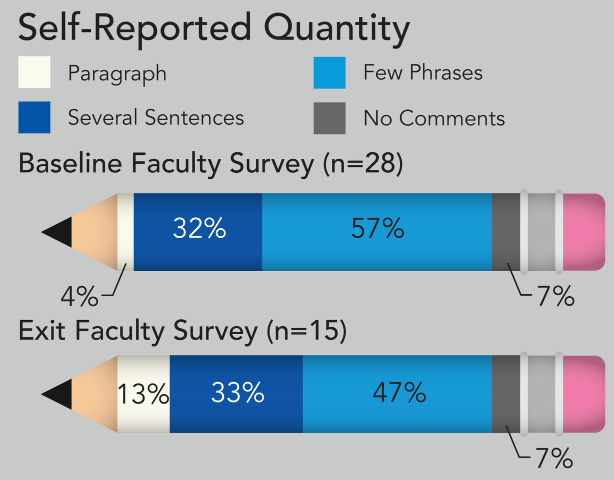
Comfort with App Use: Baseline Faculty Characteristics n=28



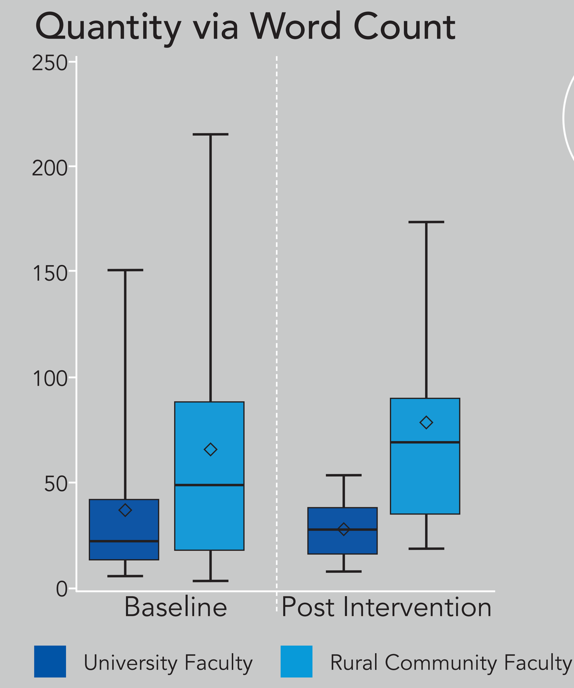
Frequency of App Use: Baseline Faculty Characteristics n=28



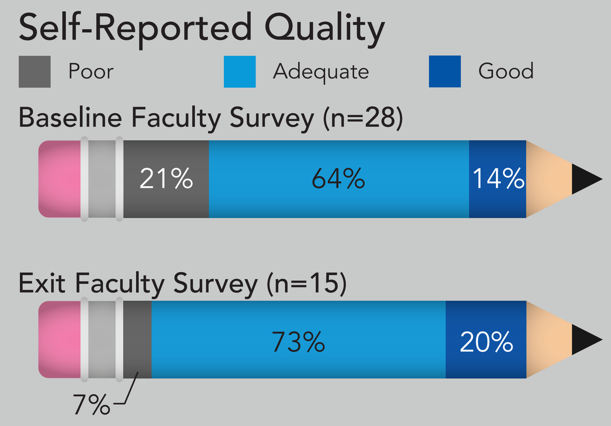
Self-Reported Quantity



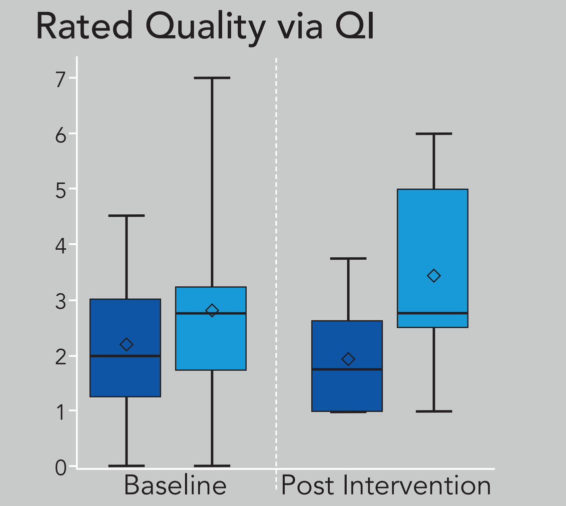
Quantity via Word Count



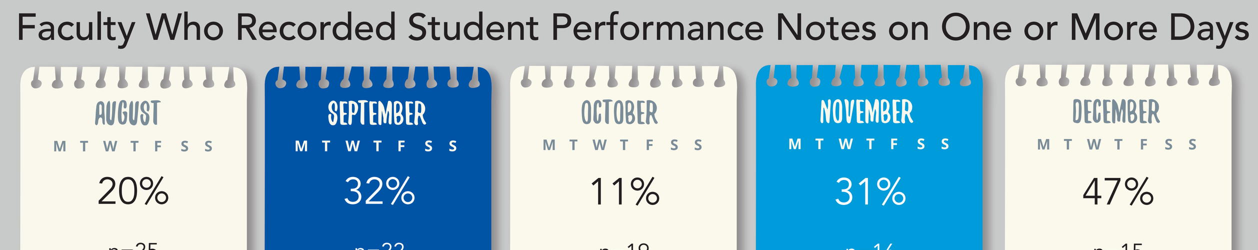
Self-Reported Quality



Rated Quality via QI



Faculty Who Recorded Student Performance Notes on One or More Days



Aug/Sept - used any method

Oct/Nov/Dec - used mobile diary app

Conclusions

Use of a mobile diary app to record daily student performance did not statistically improve quantity or rated quality of narrative feedback in medical education. A promising trend was the increase in minimum quality scores at all sites in the post intervention. Descriptive statistics showed that most faculty use apps frequently and are comfortable using apps. However, fewer than half of faculty used the mobile diary app on a regular basis. Most faculty self-assessed the quality of their narrative feedback as adequate and indicated that they wrote a few phrases to several sentences. The strengths of this project are the objective assessments of narrative feedback using blinded raters and validated tools. The faculty surveys showed positive self-assessments of quality which were not supported by the objective raters. All average quality scores were less than half of the maximum possible score. Weaknesses of the study included only fair inter-rater reliability for CCERR and poor faculty participation in the exit survey.

Special thanks to our student team members: Heston Arnold, Sara Keshtvarz, and Makayla Lewis.