Name: Matthew Zuniga

EP 755: Infectious Disease Epidemiology
Final Project

The following project is based on the following manuscript - The following questions pertain to the following article - Rosen S, Maskew M, Larson BA, Brennan AT, Tsikhutsu I, Fox MP, Vezi L, Bii M, Venter WDF. Simplified clinical algorithm for identifying patients eligible for same-day HIV treatment initiation (SLATE): Results from an individually randomized trial in South Africa and Kenya. PLoS Med. 2019 Sep 16;16(9):e1002912. doi: 10.1371/journal.pmed.1002912. PMID: 31525187; PMCID: PMC6746347.

NOTE: You will need to cut and paste your R or SAS code to the end of the report for this project. The results need to be displayed within the assignment below.

1. The file "slate.xls." is de-identified data from the original analysis. The variables and corresponding description are in Table 1 below. You will need to create new variables from existing ones highlighted in grey in the table - sex, age, TB_symptoms_number and first_cd4 (4 points).

Variable	Description	Original format	New variable	New format
ID	patient ID	Continuous		
***arm**	treatment arm	1=yes; 0=no		
***site**	treatment site	1, 2 and 3		
***house type**	type of home	Town (urban); Center (peri-urban); Rural home or village		
***employment**	employment status	1=employed; 0=unemployed		
***married**	marital status	1=married 0=single/divorced		
***hiv_test_previous**	previous HIV test	yes; no		
***previous_hiv_care**	previous HIV care	yes; no		
***tb_positive**	TB positive at enrollment	0=no; yes=1		
***retained**	outcome retained in care	0=no; yes=1		
days_to_art	days to ART initiation from enrollment	continuous		
sex	biological sex	male; female	*male	1=yes; 0=no
age	age at enrollment	continuous	*agelt40	1='<40'; 0='>=40'
TB_symptoms_number	number TB symptoms	continuous	*tb_symp	0; 1 and >=2
first_cd4	CD4 count at enrollment	continuous	*low_cd4	1='<200'; 0='>=200'

2. Create a baseline table (Table 1), stratified by study arm. The variables with an asterisk (*) in the table above should be included in Table 1 below. (28 points)

    NOTE: The variables for this project are categorized differently from many variables in the original analysis. You should have a missing category for all variables that having missingness.

Characteristic	Standard arm (n=237)	SLATE arm (n=240)	Total (N=477)
OUTCOMES			
Site:			
1	90 (37.97%)	90 (37.50%)	
2	78 (32.91%)	81 (33.75%)	
3	69 (29.11%)	69 (28.75%)	
House type:			
Center (peri-urban)	73 (30.80%)	66 (27.50%)	
Rural home or village	134 (56.54%)	139 (57.92%)	
Town (urban)	30 (12.66%)	35 (14.58%)	
Employment:			
Unemployed	47 (19.83%)	61 (25.42%)	
Employed	190 (80.17%)	179 (74.58%)	
Married:			
Single/divorced	111 (46.84%)	122 (50.83%)	
Married	126 (53.16%)	118 (49.17%)	
Previous HIV test:			
No	144 (60.76%)	148 (61.67%)	
Yes	93 (39.24%)	92 (38.33%)	
Previous HIV care:			
No	192 (81.01%)	195 (81.25%)	
Yes	45 (18.99%)	45 (18.75%)	
TB positive at enrollment:			
No	229 (96.62%)	224 (93.33%)	
Yes	8 (3.38%)	16 (6.67%)	
Male:			
No	134 (56.54%)	142 (59.17%)	

Yes	103 (43.46%)	98 (40.83)
Age at enrollment:		
<40	160 (67.51%)	142 (59.17%)
≥40	77 (32.49%)	98 (40.83)
Number TB symptoms:		
No symptoms	N/A	147 (61.25%)
1 symptom	N/A	13 (5.42%)
≥2 symptoms	N/A	80 (33.33%)
CD4 count:		
≥200	89 (37.55%)	139 (57.92%)
<200	148 (62.45%)	101 (42.08%)
Initiating ART 0 days:		
No	97 (43.30%)	65 (28.02%)
Yes	127 (56.70%)	167 (71.98%)
Initiating ART <= 7 days:		
No	51 (21.52%)	25 (10.42%)
Yes	186 (78.48%)	215 (89.58%)
Initiating ART <= 14 days:		
No	23 (9.70%)	15 (6.25%)
Yes	214 (90.30%)	225 (93.75%)
Initiating ART <= 28 days:		
No	14 (5.91%)	6 (2.50%)
Yes	223 (94.09%)	234 (97.50%)
Initiating ART <= 90 days:		
No	2 (0.84%)	1 (0.42%)
Yes	235 (99.16%)	239 (99.58%)
Outcome retained in care:		
No	92 (38.82%)	100 (41.67%)
Yes	145 (61.18%)	140 (58.33%)
Missing:		
Number TB symptoms	237	
Initiating ART 0 days	13	8

*Did not collect TB symptom data at screening, which is why it is missing for standard arm patients.

3. Create 5 new binary (1=yes; 0=no) variables for initiating ART in a) 0 days, b) <=7 days, c) <=14 days, d) <=28 days and e) <=90 days using the variable days_to_art in the dataset. Add these 5 new variables plus the outcome of retention (variable retained) (n (%) that achieved the outcome) to the end of Table 1 above stratified by study arm. (6 points)

Name: Matthew Zuniga

1. You need to assess if the following five (5) clinical and demographic characteristics are confounders in your analysis for the outcome of **retained**: male, agelt40, employed, married and lowcd4. You need to calculate the risk and the risk ratios for all strata. For confounding, you only need to show calculations for: 1) is the confounder related to the exposure and 2) is the confounder related to the outcome INDEPENDENT of exposure. You do not have to calculate the 10% change in estimate.

   NOTE: Use the same structure of the 2x2 tables below for the other variables. Also, be sure to state what calculations you are basing your conclusions on for the 3 questions below for all five (5) variables.

	Male=no	
	SLATE arm	Standard arm
Retained	84	80
Not retained	58	54
Total	142	134
Risk	0.59	0.60
Risk Ratio	0.99	reference
% Exposed	59.1%	-

	Male=yes	
	SLATE arm	Standard arm
Retained	56	65
Not retained	42	38
Total	98	103
Risk	0.57	0.63
Risk Ratio	0.91	reference
% Exposed	57.1%	-

Male = no

Risk SLATE = (84/142)
           = 0.59

Risk Standard = (80/134)
              = 0.60
Risk Ratio = (84/142) / (80/134)
           = 0.99

% Exposed = (84/142)*100% = 59.1%

Male = Yes

Risk SLATE = (56/98)
           = 0.57

Risk Standard = (65/103)
              = 0.63

Risk Ratio = (56/98) / (65/103)
           = 0.91

% Exposed = (56/98)*100% = 57.1%

Confounding and EMM for Male variable:

Name: Matthew Zuniga

The male confounder is not related to the exposure as the estimated % of exposure in the SLATE arm category do not differ across strata (59.1 % = 57.1%). Furthermore, the male confounder is not related to the outcome independent of exposure as the calculated risk amongst the unexposed (standard arm) do not differ cross strata (0.60 = 0.63). Male-sex status is not an effect measure modifier as the observed risk ratios are the same across strata (0.99 = 0.91).

	Agelt40 = '≥40'	
	SLATE arm	Standard arm
Retained	28	52
Not retained	19	25
Total	47	77
Risk	0.60	0.68
Risk Ratio	0.88	reference
% Exposed	60%	-

	Agelt40 = '<40'	
	SLATE arm	Standard arm
Retained	112	93
Not retained	81	67
Total	193	160
Risk	0.58	0.58
Risk Ratio	1.00	reference
% Exposed	58%	-

Agelt40 = '**≥40'**

Risk SLATE = (28/47)
      = 0.60

Risk Standard = (52/77)
       = 0.68

Risk Ratio = (28/47) / (52/77)
      = 0.88

% Exposed = **(28/47)*100% = 60%**

Agelt40 = '<40'

Risk SLATE = (112/193)
      = 0.58

Risk Standard = (93/160)
       = 0.58

Risk Ratio = (112/193) / (93/160)
      = 1.00

% Exposed = (112/193)*100% = 58%

The Agelt40 confounder is not related to the exposure as the observed % exposed in the SLATE arm category is the same amongst strata (60% = 58%). Furthermore, the Agelt40 confounder is related to the outcome independent of exposure as the observed risk amongst the unexposed (standard arm) differ across strata (0.68 ≠ 0.58). Agelt40-status is an effect measure modifier as the observed risk ratios differ across strata (0.88 ≠ 1.00).

Name: Matthew Zuniga

	Employment = Unemployed	
	SLATE arm	Standard arm
Retained	38	28
Not retained	23	19
Total	61	47
Risk	0.62	0.60
Risk Ratio	1.05	reference
% Exposed	62.3%	-

	Employment = Employed	
	SLATE arm	Standard arm
Retained	102	117
Not retained	77	73
Total	179	190
Risk	0.57	0.62
Risk Ratio	0.93	reference
% Exposed	57%	-

Employment = Unemployed

Risk SLATE = (38/61)
        = 0.62

Risk Standard = (28/47)
            = 0.60

Risk Ratio = (38/61) / (28/47)
        = 1.05

% Exposed = (38/61)*100%
        = 62.3%

Employment = Employed

Risk SLATE = (102/179)
        = 0.57

Risk Standard = (117/190)
            = 0.62

Risk Ratio = (102/179) / (117/190)
        = 0.93

% Exposed = (102/179)*100%
        = 57%

Employment-status is related to the exposure as the observed % exposed in the SLATE arm category differ across strata (62.3% ≠ 57%). Furthermore, the employment-status not related to the outcome independent of exposure as the observed risk in the unexposed (standard arm) are the same across strata (0.60 = 0.62). Employment status is not an effect measure modifier as the observed risk ratios are relatively the same across strata (1.05 = 0.93).

Name: Matthew Zuniga

	Married = Single/Divorced	
	SLATE arm	Standard arm
Retained	58	67
Not retained	64	44
Total	122	111
Risk	0.48	0.60
Risk Ratio	0.79	reference
% Exposed	48%	-

	Married = Married	
	SLATE arm	Standard arm
Retained	82	78
Not retained	36	48
Total	118	126
Risk	0.69	0.62
Risk Ratio	1.12	reference
% Exposed	69%	-

Married = Single/Divorced:

Risk SLATE = (58/122)
= 0.48

Risk Standard = (67/111)
= 0.60

Risk Ratio = (58/122) / (67/111)
= 0.79

% Exposed = (58/122)*100%
= 48%

Married = Married

Risk SLATE = (82/118)
= 0.69

Risk Standard = (78/126)
= 0.62

Risk Ratio = (82/118) / (78/126)
= 1.12

% Exposed = (82/118)*100% = 69%

Marriage status is related the exposure as the observed % exposed in the SLATE arm category differ across strata (48% ≠ 69%). Furthermore, marriage status is not related to the outcome independent of exposure as the observed risk among the unexposed (standard arm) are the same across strata (0.60 = 0.62). Marriage status is an effect measure modifier as the observed risk ratios differ across strata (0.79 ≠ 1.12).

Name: Matthew Zuniga

	Lowcd4 = '≥200'	
	SLATE arm	Standard arm
Retained	78	52
Not retained	61	37
Total	139	89
Risk	0.56	0.58
Risk Ratio	0.96	reference
% Exposed	56%	-

	Lowcd4 = '<200'	
	SLATE arm	Standard arm
Retained	62	93
Not retained	39	55
Total	101	148
Risk	0.61	0.63
Risk Ratio	0.98	reference
% Exposed	61%	-

Lowcd4 = '≥200'

Risk SLATE = (78/139)
         = 0.56

Risk Standard = (52/89)
              = 0.58

Risk Ratio = **(78/139) /** (52/89)
         = 0.96

% Exposed = **(78/139)*100%**
         = 56%

Lowcd4 = '<200'

Risk SLATE = (62/101)
         = 0.61

Risk Standard = (93/148)
              = 0.63

Risk Ratio = (62/101) / (93/148)
         = 0.98

% Exposed = (62/101)*100%
         = 61%

Lowcd4 status is related to the exposure as the observed % exposed in the SLATE arm category differ across strata (56% ≠ 61%). Furthermore, Lowcd4 status is related to the outcome independent of exposure as the observed risk in the unexposed (standard arm) differ across strata (0.58 ≠ 0.63). Lowcd4 status, therefore, confounds the association between arm treatment and retention outcome. Also, Lowcd4 status is not an effect measure modifier as the observed risk ratios are the same across strata (0.96 = 0.98).

Name: Matthew Zuniga

1. Is the confounder related to the exposure?
2. Is the confounder related to the outcome independent of exposure?
3. Is the confounder also and effect measure modifier?

Name: Matthew Zuniga

1. Run a crude and adjusted regression model to estimate the risk ratio assessing the association of SLATE arm (treatment/exposure) with the outcome of retention. Controlling for the exposure (arm) in addition to any of the five (5) variables that you identified as confounders in Part 2 above. (3 points)

2. Complete the table below based on the results of you model. (4 points)

NOTE: The variables for this project are categorized differently from many variables in the original analysis.

R modeling - http://rstudio-pubs-static.s3.amazonaws.com/5752_fc41dca85dd24539bc99868697de83d0.html
SAS modeling - https://support.sas.com/resources/papers/proceedings11/345-2011.pdf

Characteristic		Crude Risk Ratio (95% CI)	Adjusted Risk Ratio (95% CI)
Arm	Standard of care	Reference	Reference
	SLATE	0.9534 (0.8228 - 1.1048)	0.9698 (0.8345 - 1.1271)

3. Interpret your adjusted measure of association between treatment and the outcome of retention. (6 points)

   Among those who received SLATE arm treatment, patients were 0.97 times the risk of being retained in the treatment program, in comparison to those those who did not receive the SLATE arm treatment, adjusted for low CD4 count.

4. Calculate the 10% change in estimate for the association of SLATE and retention. (2 point)

   (0.97 – 0.95)/0.97= 0.0206*100% = 2.06%

5. Interpret the 10% change in estimate and the direction of the bias. If there was no bias then be sure to state that. (2 points)

   The association between SLATE arm treatment and retention was not confounded by low CD4 count by >10% (2.06%). There was no bias observed.

SAS-code:
```
proc import out = Patient
datafile = '/home/u63114430/Infectious Epi Project/SLATE.xls'
DBMS = XLS replace;
getnames = yes;
run;
libname InfEpi '/home/u63114430/Infectious Epi Project';
run;
Data InfEpi.Patient_new;
set Patient;
run;
proc format;
value armf 1 = 'yes' 0 = 'no';
```

Name: Matthew Zuniga

```
value employmentf 1 = 'employed' 0 = 'unemployed';
value marriedf 1 = 'married' 0 = 'single/divorced';
value tb_positivef 0 = 'no' 1 = 'yes';
value retainedf 0 = 'no' 1 = 'yes';
value malef 1 = 'yes' 0 = 'no';
value tb_sympf 0 = 'No TB Symptoms' 1 = '1 TB Symptoms' 2 = '>=2 TB Symptoms';
value agelt40f 1 = '<40' 0 = '>=40';
value low_cd4f 1 = '<200' 0 = '>=200';
value days_art0f 1 = 'Yes' 0 = 'No';
value days_art7f 1 = 'Yes' 0 ='No';
value days_art14f 1 = 'Yes' 0 = 'No';
value days_art28f 1 = 'Yes' 0 = 'No';
value days_art90f 1 = 'Yes' 0 = 'No';
run;
Data Patient_new;
set Patient;
if sex = 'Male' then male = 1;
else if sex = 'Female' then male = 0;
if age <40 then agelt40 = 1;
else if age >= 40 then agelt40 = 0;
if TB_symptoms_number = 0 then tb_symp = 0;
else if TB_symptoms_number = 1 then tb_symp = 1;
else if TB_symptoms_number >=2 then tb_symp = 2;
if first_cd4 <200 then low_cd4 = 1;
else if first_cd4 >=200 then low_cd4 = 0;
if days_to_art = 0 then days_art0 = 1;
else if days_to_art > 0 then days_art0 = 0;
if days_to_art <=7 then days_art7 = 1;
else if days_to_art > 7 then days_art7 = 0;
if days_to_art <= 14 then days_art14 = 1;
else if days_to_art >14 then days_art14 = 0;
if days_to_art <=28 then days_art28 = 1;
else if days_to_art >28 then days_art28 = 0;
if days_to_art <= 90 then days_art90 = 1;
else if days_to_art > 90 then days_art90 = 0;
format arm armf. employment employmentf. married marriedf.
tb_positive tb_positivef. retained retainedf. male malef. tb_symp tb_sympf.
agelt40 agelt40f. low_cd4 low_cd4f. days_art0 days_art0f. days_art7 days_art7f. days_art14
days_art14f. days_art28 days_art28f. days_art90 days_art90f.;
run;
proc sort data = patient_new;
by arm;
run;
proc freq data = patient_new;
by arm;
tables site house_type employment married hiv_test_previous previous_hiv_care tb_positive retained male
agelt40 tb_symp low_cd4
days_art0 days_art7 days_art14 days_art28 days_art90;
run;
proc sort data = patient_new;
```

```
        by male;
        run;
        proc freq data = patient_new;
        by male;
        table retained*arm;
        run;
        proc sort data =patient_new;
        by agelt40;
        run;
        proc freq data = patient_new;
        by agelt40;
        table retained*arm;
        run;
        proc sort data = patient_new;
        by employment;
        run;
        proc freq data = patient_new;
        by employment;
        table retained*arm;
        run;
        proc sort data = patient_new;
        by married;
        run;
        proc freq data = patient_new;
        by married;
        table retained*arm;
        run;
        proc sort data = patient_new;
        by low_cd4;
        run;
        proc freq data = patient_new;
        by low_cd4;
        table retained*arm;
        run;
        proc genmod data = patient_new desc;
        model retained = arm / link=log dist=bin;
        estimate 'Risk Ratio' arm 1 / exp;
        run;
        proc genmod data = patient_new desc;
        model retained = arm low_cd4/ link=log dist=bin;
        estimate 'Risk Ratio' arm 1 / exp;
        estimate 'Risk Ratio of <200 vs >=200' low_cd4 1 / exp;
run;
```