EP 755: Infectious Disease Epidemiology
Final Project
Due for Grading on Thursday, April 25th 2024 by 10am

Mingshan Guo

The following project is based on the following manuscript - The following questions pertain to the following article - Brennan AT, Vetter B, Masuku SD, et al. Integration of point-of-care screening for type 2 diabetes mellitus and hypertension into the COVID-19 vaccine programme in Johannesburg, South Africa. BMC Public Health. 2023;23(1):2291. Published 2023 Nov 20. doi:10.1186/s12889-023-17190-6

NOTE: You will need to cut and paste your R or SAS code to the end of the report for this project. Do not attached your output to the report. The results need to be displayed within the assignment below.

Part I: Basic coding to create new variables:

1. The file "EP755_final.xlsx." is de-identified data from the original analysis. The variables and corresponding description are in Table 1 below. You will need to import the dataset into SAS or R and create new variables from existing ones highlighted in grey in the table –age and bmi.

Variable	Description	Original format	New variable	New format
id	patient ID	continuous		
ref	patient was referred for follow-up care	yes=1; no=0		
fasted	fasted 8 hours	yes=1; no=0		
linked	linked to care	yes=1; no=0		
syt_bp	systolic blood pressure	continuous		
dyst_bp	diastolic blood pressure	continuous		
glucose_val	plasma glucose level	continuous		
*HIV	HIV status	PLWH=1; PLWOH=0		
*site	study site	1, 2, 3, 4		
*employ	employment status	1=employed; 0=unemployed		
*smk	smoking status	1=ever; 0=never		
*prevdm	previous diabetes diagnosis	yes=1; no=0		
*prevhtn	previous hypertension diagnosis	yes=1; no=0		
*mh	mental health condition	yes=1; no=0		
*resp	respiratory condition	yes=1; no=0		
*cvd	cardiovascular disease	yes=1; no=0		
*bio_sex	biological sex	female =0; male = 1		
*age	age at enrollment	continuous	*agelt40	1=<40; 0=>=40

***bmi**	body mass index	continuous	*obese	1=>=30; 0=<30

2. Create a baseline table (Table 1), stratified by HIV status (people living with HIV – PLWH and people living without HIV – PLWOH). The variables with an asterisk (*) in the table above should be included in your Table 1 (you can reference the supplemental table in the paper to see how you can format your table. You need to present the n and corresponding column % for each cell. For age and bmi, in addition to the categorical variables being displayed in the table, you should have a row displaying the median and corresponding interquartile range (IQR). For variables categorical variables that have missing values you need to have a row to display the n and % missing. (*NOTE: the data in this table will not match the table in the published paper.)*

Table 1. Characteristics of participants stratified by HIV status and in total, with n(%) for categorical variables, median(IQR) for continuous variables, and n(%) of missing values in Johannesburg, South Africa (N=1376)

	With HIV n=341(24.78)		Without HIV n=1035(75.22)		Total N=1376	
	n(%) for categorical variables; median(IQR) for continuous variables	n(%) of missing values	n(%) for categorical variables; median(IQR) for continuous variables	n(%) of missing values	n(%) for categorical variables; median(IQR) for continuous variables	n(%) of missing values
Study site						
1	61(17.89)	0(0)	219(21.16)	0(0)	280(20.35)	0(0)
2	213(62.46)		596(57.58)		809(58.79)	
3	22(6.45)		62(5.99)		84(6.10)	
4	45(13.20)		158(15.27)		203(14.75)	
Employment Status						
employed	178(52.20)	0(0)	509(49.18)	0(0)	687(49.93)	0(0)
unemployed	163(47.80)		526(50.82)		689(50.07)	
Smoking status						
ever	109(31.96)	0(0)	338(32.66)	0(0)	447(32.49)	0(0)
never	232(68.04)		697(67.34)		929(67.51)	
Previous diabetes diagnosis						
yes	6(1.78)	3(0.88)	28(2.72)	7(0.68)	34(2.49)	10(0.73)
no	332(98.22)		1000(97.28)		1332(97.51)	
Previous hypertension diagnosis						
yes	27(8.36)	18(5.28)	72(7.35)	55(5.31)	99(7.60)	73(5.31)
no	296(91.64)		908(92.65)		1204(92.40)	
Mental health condition						
yes	3(0.88)	0(0)	3(0.29)	0(0)	6(0.44)	0(0)
no	338(99.12)		1032(99.71)		1370(99.56)	
Respiratory condition						
yes	1(0.29)	0(0)	11(1.06)	0(0)	12(0.87)	0(0)
no	340(99.71)		1024(98.94)		1364(99.13)	
Cardiovascular disease						
yes	6(1.76)	0(0)	15(1.45)	0(0)	21(1.53)	0(0)
no	335(98.24)		1020(98.55)		1355(98.47)	

Biological sex						
female	179(52.49)	0(0)	559(54.01)	0(0)	738(53.63)	0(0)
male	162(47.51)		476(45.99)		638(46.37)	
Age category						
age<40	189(55.43)	0(0)	570(55.07)	0(0)	750(54.51)	0(0)
age>=40	152(44.57)		465(44.93)		626(45.49)	
Obese						
bmi>=30	99(29.03)	0(0)	289(27.92)	0(0)	388(28.20)	0(0)
Bmi<30	242(70.97)		746(72.08)		988(71.80)	
Age	39(30, 47)	0(0)	38(30, 47)	0(0)	38(30, 47)	0(0)
bmi	25.66(21.78, 30.86)		25.64 (21.83, 30.78)		25.65(21.83, 30.81)	

3. Create 4 new binary (1=yes; 0=no) outcome variables and display them in a table stratified by HIV status. Just like table 1, you need to present the n and corresponding column % for each cell.

   a) elevated blood glucose (<11.1 mmol/L =0, ≥11.1 mmol/L=1 if not fasted; <7.0 mmol/L =0, ≥7.0 mmol/L=1 if fasted);
   b) elevated blood pressure (systolic blood pressure ≥140 mmHG AND diastolic blood pressure ≥90 mmHG);
   c) elevated blood glucose (<11.1 mmol/L =0, ≥11.1 mmol/L=1 if not fasted; <7.0 mmol/L =0, ≥7.0 mmol/L=1 if fasted) amongst those with no previous diagnosis of diabetes;
   d) elevated blood pressure (systolic blood pressure ≥140 mmHG AND diastolic blood pressure ≥90 mmHG) amongst those with no previous diagnosis of hypertension);
   e) linked to care amongst those who were referred (variable ref in dataset) for follow-up (yes=1; no=0). *NOTE: the variable linked is already created for you. You just need to add it to your table.*

Table 2. n(%) on elevated blood glucose, elevated blood pressure, linked to care, stratified by HIV status and in total, with n(%) of missing values in Johannesburg, South Africa (N=1376)

	With HIV n=341(24.78)		**Without HIV n=1035(75.22)**		**Total N=1376**	
	n(%)	**n(%) of missing values**	**n(%)**	**n(%) of missing values**	**n(%)**	**n(%) of missing values**
Elevated blood glucose						
yes	5(1.47)	0(0)	17(1.64)	0(0)	22(1.60)	0(0)
no	336(98.53)		1018(98.36)		1354(98.40)	
Elevated blood pressure						
yes	52(15.25)	0(0)	161(15.56)	0(0)	213(15.48)	0(0)
no	289(84.75)		874(84.44)		1163(84.52)	
Elevated blood glucose among those with no previous diabetes diagnosis						
yes	2(0.60)	9(2.64)	10(1.00)	35(3.38)	12(0.90)	44(3.20)
no	330(99.40)		990(99.00)		1320(99.10)	
Elevated blood pressure among those with no previous hypertension diagnosis						
yes	34(11.49)	45(13.20)	106(11.67)	127(12.27)	140(11.63)	172(12.5)
no	262(88.51)		802(88.33)		1064(88.37)	
Linked to care						

yes	27(56.25)	293(85.92)	64(55.17)	919(88.79)	91(55.49)	1212(88.08)
no	21(43.75)		52(44.83)		73(44.51)	

Part 2: Confounding and effect measure modification:

1. You need to assess if the following 6 clinical and demographic characteristics are confounders in your analysis for the outcome 'd' (elevated blood pressure amongst those with no previous diagnosis of hypertension) above: sex, age <40, clinically obese, previous diabetes, smoking and employed. 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 6 variables.

	Sex = female	
	PLWH	PLWOH
Elevated BP+	10	39
Elevated BP-	136	436
Total	146	475
Risk	0.068	0.082
Risk Ratio	0.83	reference
% Exposed	23.51	-

	Sex = male	
	PLWH	PLWOH
Elevated BP+	24	67
Elevated BP-	126	366
Total	150	433
Risk	0.16	0.15
Risk Ratio	1.07	reference
% Exposed	25.73	-

Sex=female:
10/146=0.068
39/475=0.082
0.068/0.082=0.83
146/(146+475)=23.51%
Sex=male:
24/150=0.16
67/433=0.15
0.16/0.15=1.07
150/(150+433)=25.73%

- Is the confounder related to the exposure?
  Since 23.51% is not equal to 25.73%, the percentage of exposed individuals differ between strata. Confounder sex is related to the exposure HIV.
- Is the confounder related to the outcome independent of exposure?
  Since 0.082 is not equal to 0.15, the risk of outcome among the unexposed differ between strata. Confounder sex is related to elevated blood pressure amongst those with no previous diagnosis of hypertension, independent of exposure HIV.
- Is the confounder also an effect measure modifier?
  Since 0.83 is not equal to 1.07, confounder sex is an effect measure modifier.

	Age>=40

	Age<40

	PLWH	PLWOH
Elevated BP+	19	60
Elevated BP-	102	290
Total	121	350
Risk	0.16	0.17
Risk Ratio	0.94	reference
% Exposed	25.69	-

	PLWH	PLWOH
Elevated BP+	15	46
Elevated BP-	160	512
Total	175	558
Risk	0.086	0.082
Risk Ratio	1.05	reference
% Exposed	23.87	-

Age>=40:
19/121=0.16
60/350=0.17
0.16/0.17=0.94
121/(121+350)=25.69%
Age<40:
15/175=0.086
46/558=0.082
0.086/0.082=1.05
175/(175+558)=23.87%

- Is the confounder related to the exposure?
  Since 25.69% is not equal to 23.87%, the percentage of exposed individuals differ between strata. Confounder age<40 is related to the exposure HIV.
- Is the confounder related to the outcome independent of exposure?
  Since 0.17 is not equal to 0.082, the risk of outcome among the unexposed differ between strata. Confounder age<40 is related to elevated blood pressure amongst those with no previous diagnosis of hypertension, independent of exposure HIV.
- Is the confounder also an effect measure modifier?
  Since 0.94 is not equal to 1.05, confounder age<40 is an effect measure modifier.

	Obese = no	
	PLWH	PLWOH
Elevated BP+	26	66
Elevated BP-	190	617
Total	216	683
Risk	0.12	0.097
Risk Ratio	1.24	reference
% Exposed	24.03	-

	Obese = yes	
	PLWH	PLWOH
Elevated BP+	8	40
Elevated BP-	72	185
Total	80	225
Risk	0.1	0.18
Risk Ratio	0.56	reference
% Exposed	26.23	-

Obese=no:
26/216=0.12
66/683=0.097
0.12/0.097=1.24
216/(216+683)=24.03%
Obese=yes:
8/80=0.1
40/225=0.18
0.1/0.18=0.56
80/(80+225)=26.23%

- Is the confounder related to the exposure?
  Since 24.03% is not equal to 26.23%, the percentage of exposed individuals differ between strata. Confounder clinical obese is related to the exposure HIV.
- Is the confounder related to the outcome independent of exposure?
  Since 0.097 is not equal to 0.18, the risk of outcome among the unexposed differ between strata. Confounder clinical obese is related to elevated blood pressure amongst those with no previous diagnosis of hypertension, independent of exposure HIV.
- Is the confounder also an effect measure modifier?
  Since 1.24 is not equal to 0.56, confounder clinical obese is an effect measure modifier.

	Previous diabetes = no	
	PLWH	PLWOH
Elevated BP+	34	103
Elevated BP-	260	791
Total	294	894
Risk	0.1156	0.1152
Risk Ratio	1.0035	reference
% Exposed	24.75	-

	Previous diabetes = yes	
	PLWH	PLWOH
Elevated BP+	0	2
Elevated BP-	2	8
Total	2	10
Risk	0	0.2
Risk Ratio	0	reference
% Exposed	16.67	-

Previous Diabetes=no:
34/294=0.1156
103/894=0.1152
0.1156/0.1152=1.0035
294/(294+894)=24.75%
Previous Diabetes=yes:
0/2=0
2/10=0.2
0/0.2=0
2/(2+10)=16.67%

- Is the confounder related to the exposure?
  Since 24.75% is not equal to 16.67%, the percentage of exposed individuals differ between strata. Confounder previous diabetes is related to the exposure HIV.
- Is the confounder related to the outcome independent of exposure?
  Since 0.1152 is not equal to 0.2, the risk of outcome among the unexposed differ between strata. Confounder previous diabetes is related to elevated blood pressure amongst those with no previous diagnosis of hypertension, independent of exposure HIV.
- Is the confounder also an effect measure modifier?
  The data is not sufficient to make a conclusion because there are no people in the exposure+ * outcome+ cell. Merely according to the data, since 1.0035 is not equal to 0, confounder previous diabetes is an effect measure modifier. However, this conclusion is not valuable.

	Smoke = no

	Smoke = yes

	PLWH	PLWOH
Elevated BP+	19	71
Elevated BP-	180	527
Total	199	598
Risk	0.095	0.12
Risk Ratio	0.79	reference
% Exposed	24.97	-

	PLWH	PLWOH
Elevated BP+	15	35
Elevated BP-	82	275
Total	97	310
Risk	0.15	0.11
Risk Ratio	1.36	reference
% Exposed	23.83	-

Smoke=no:
19/199=0.095
71/598=0.12
0.095/0.12=0.79
199/(199+598)=24.97%
Smoke=yes:
15/97=0.15
35/310=0.11
0.15/0.11=1.36
97/(97+310)=23.83%

- Is the confounder related to the exposure?
  Since 24.97% is not equal to 23.83%, the percentage of exposed individuals differ between strata. Confounder smoking is related to the exposure HIV.
- Is the confounder related to the outcome independent of exposure?
  Since 0.12 is approximately equal to 0.11, the risk of outcome among the unexposed do not differ between strata. Confounder smoking is not related to elevated blood pressure amongst those with no previous diagnosis of hypertension, independent of exposure HIV.
- Is the confounder also an effect measure modifier?
  Since 0.79 is not equal to 1.36, confounder smoking is an effect measure modifier.

	Employ = no	
	PLWH	PLWOH
Elevated BP+	15	50
Elevated BP-	129	418
Total	144	468
Risk	0.10	0.11
Risk Ratio	0.91	reference
% Exposed	23.53	-

	Employ = yes	
	PLWH	PLWOH
Elevated BP+	19	56
Elevated BP-	133	384
Total	152	440
Risk	0.125	0.13
Risk Ratio	0.96	reference
% Exposed	25.68	-

Employ=no:
15/144=0.10
50/468=0.11
0.10/0.11=0.91
144/(144+468)=23.53%
Employ=yes:
19/152=0.125
56/440=0.13
0.125/0.13=0.96
152/(152+440)=25.68%

- Is the confounder related to the exposure?
  Since 23.53% is not equal to 25.68%, the percentage of exposed individuals differ between strata. Confounder employ is related to the exposure HIV.
- Is the confounder related to the outcome independent of exposure?
  Since 0.11 is approximately equal to 0.13, the risk of outcome among the unexposed do not differ between strata. Confounder employ is not related to elevated blood pressure amongst those with no previous diagnosis of hypertension, independent of exposure HIV.
- Is the confounder also an effect measure modifier?
  Since 0.91 is approximately equal to 0.96, confounder employ is not an effect measure modifier.

Part 3: Modeling:

1. Run a crude and adjusted regression model to estimate the risk ratio assessing the association of HIV status (PLWH/PLWOH) with the outcome of elevated blood pressure (hypertension). Controlling for the HIV status in addition to any of the 6 variables that you identified as confounders in Part 2 above.

2. Complete the table below based on the results of the model you ran model. *NOTE: The variables for this project are categorized differently from many variables in the original analysis.*

Characteristic		Crude Risk Ratio (95% CI)	Adjusted Risk Ratio (95% CI)
HIV status	PLWOH	Reference	Reference
	PLWH	0.98(0.74, 1.31)	0.93(0.70, 1.23)

Confounders identified in Part 2 above: bio_sex, agelt40, obese, prevdm (which are related to both exposure and outcome)

3. Interpret your adjusted measure of association between HIV infection and the outcome of elevated blood pressure.

   Participants with HIV infection had 0.93 times the risk of having elevated blood pressure with a 95% CI of (0.70, 1.23), controlling for biological sex, whether age is less than 40, clinical obese, and previous diagnosis of diabetes.

4. Calculate the 10% change in estimate for the association of HIV infection and the outcome of elevated blood pressure.

   10% change in estimate = (0.98-0.93)/0.93 = 0.0538 = 5.38%

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

   The association between HIV infection and elevated blood pressure was confounded by biological sex, whether age is less than 40, clinical obese, and previous diagnosis of diabetes, by 5.38%. Since 10% change in estimate was 5.38% < 10%, there was no bias.

Code:

```
proc import out=final_project
        datafile="/home/u63582559/EP755/EP755_final.xlsx"
        DBMS=XLSX replace;
        getnames=yes;
run;
/*Part 1_1*/
data final_project_1;
        set final_project;
        if age=. then agelt40=.;
        else if age<40 then agelt40=1;
        else if age>=40 then agelt40=0;
        if bmi=. then obese=.;
        else if bmi>=30 then obese=1;
        else if bmi<30 then obese=0;
/*Part 1_2*/
proc sort;
        by HIV;
run;
proc freq order=data;
        tables HIV;
run;
proc freq;
        tables HIV site employ smk prevdm prevhtn mh resp cvd bio_sex agelt40 obese;
run;
proc freq;
        tables HIV site employ smk prevdm prevhtn mh resp cvd bio_sex agelt40 obese;
        by HIV;
run;
proc univariate data=final_project_1;
        var age bmi;
run;
proc freq;
        tables HIV site employ smk prevdm prevhtn mh resp cvd bio_sex agelt40 obese;
        by HIV;
run;
proc univariate data=final_project_1;
        var age bmi;
run;
proc means data=final_project_1 n nmiss;
        var age bmi;
run;
proc univariate data=final_project_1;
        var age bmi;
        by HIV;
run;
/*Part 1_3*/
data final_project_2;
        set final_project_1;
        if (fasted=. or glucose_val=.) then elevated_gl=.;
```

```
        else if (fasted=1 and glucose_val>=7) or (fasted=0 and glucose_val>=11.1) then elevated_gl=1;
        else if (fasted=1 and glucose_val<7) or (fasted=0 and glucose_val<11.1) then elevated_gl=0;
        if (syt_bp=. or dyst_bp=.) then elevated_bp=.;
        else if (syt_bp>=140 and dyst_bp>=90) then elevated_bp=1;
        else if (syt_bp<140 or dyst_bp<90) then elevated_bp=0;
        if (elevated_gl=. or prevdm=.) then elevated_glnpd=.;
        else if (elevated_gl=1 and prevdm=0) then elevated_glnpd=1;
        else if (elevated_gl=0 and prevdm=0) then elevated_glnpd=0;
        if (elevated_bp=. or prevhtn=.) then elevated_bpnph=.;
        else if (elevated_bp=1 and prevhtn=0) then elevated_bpnph=1;
        else if (elevated_bp=0 and prevhtn=0) then elevated_bpnph=0;
run;
proc sort;
        by HIV;
run;
proc freq order=data;
        tables elevated_gl elevated_bp elevated_glnpd elevated_bpnph linked;
run;
proc freq order=data;
        tables elevated_gl elevated_bp elevated_glnpd elevated_bpnph linked;
        by HIV;
run;

/*Part 2*/
proc sort data=final_project_2;;
        by bio_sex;
run;
proc freq data=final_project_2;
        tables HIV*elevated_bpnph;
        by bio_sex;
run;
proc sort data=final_project_2;;
        by agelt40;
run;
proc freq data=final_project_2;
        tables HIV*elevated_bpnph;
        by agelt40;
run;
proc sort data=final_project_2;;
        by obese;
run;
proc freq data=final_project_2;
        tables HIV*elevated_bpnph;
        by obese;
run;
proc sort data=final_project_2;;
        by prevdm;
run;
proc freq data=final_project_2;
        tables HIV*elevated_bpnph;
        by prevdm;
run;
proc sort data=final_project_2;;
```

```
		by smk;
	run;
	proc freq data=final_project_2;
		tables HIV*elevated_bpnph;
		by smk;
	run;
	proc sort data=final_project_2;;
		by employ;
	run;
	proc freq data=final_project_2;
		tables HIV*elevated_bpnph;
		by employ;
	run;

	/*Part 3*/
	proc genmod data=final_project_2 desc;
		model elevated_bp = HIV / link=log dist=bin;
		estimate "Risk Ratio" HIV 1 / exp;
	run;
	proc genmod data=final_project_2 desc;
		model elevated_bp = HIV bio_sex agelt40 obese prevdm / link=log dist=bin;
		estimate "Risk Ratio for PLWH vs PLWOH" HIV 1 / exp;
		estimate "Risk Ratio for male vs female" bio_sex 1 / exp;
		estimate "Risk Ratio for agelet40 vs agelaeqt40" agelt40 1 / exp;
		estimate "Risk Ratio for obese vs non_obese" obese 1 / exp;
		estimate "Risk Ratio for prediabetes vs no_prediabetes" prevdm 1 / exp;
run;
```