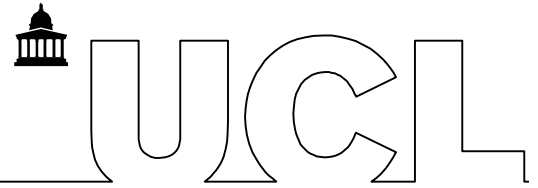


INSTITUTE OF HEALTH INFORMATICS



## Graduate Programme in Data Science for Research in Health & Biomedicine

### Assessed Coursework Submission

Student candidate number:	*****
Module:	Statistical Methods in Epidemiology
Date due:	Friday 9 February 2018, 2:00pm
Word count: (excluding references, diagrams and appendices)	1,387 words (excluding tables/figures)
Disability or medical condition for which UCL has granted special examination arrangements:	

## **Methods**

Data from a study carried out in the Western Cape, South Africa of 1,769 individuals who have tested HIV-positive were analysed to investigate the association between employment status and probability of commencing antiretroviral therapy (ART) after a one month follow-up. For this study, employment status was the exposure and ART uptake was the outcome. Age, sex and socio-economic parameters are normally considered a priori confounders or effect modifiers.

Using Stata version 14, the dataset was summarised to review the data and to look for errors. Outlying values could be replaced with the correct data if known. Continuous variable errors could be substituted with the mean of the continuous variables either side of the error. Categorical variable errors could be treated as missing data and recoded.

The  $\chi^2$  significance test was used to investigate the association between ART uptake and the following characteristics: employment status, sex, age group, previous HIV test history, previous TB treatment history, knowledge of partners HIV status, CD4 count at study enrolment, location of dwelling, education status, clinic travel arrangements, reason for the HIV test, household size, marital status, participants social capital and main household income type.

Crude odds ratios (OR) were calculated along with 95% confidence intervals (CI), p-values and test for trend with associated p-value.

Variables with p-values of  $<0.05$  were selected for further investigation using Mantel-Haenszel (MH) analysis. At this stage, missing data values were recoded as a separate category for inclusion in analysis. MH analysis were used to calculate stratified crude ORs and 95% CIs, p-values and test of homogeneity p-values for the association between employment status and ART uptake.

Using the forward stepwise logistic regression, variables for which the MH analyses yielded p-values of  $<0.05$ , were further investigated as possible confounders. Potential effect modification was investigated using test of homogeneity for variables with p-values of  $<0.05$ . The likelihood ratio test (LRT) for interaction was also presented for the relevant variables. Multivariable logistic regression was performed to adjust for confounders and stratified analysis were conducted to investigate effect modification.

## **Results**

There were 1,081 females and 688 males in the study population ( $n=1,769$ ), of which 150 (63.56%) females and 86 (36.44%) males had started ART by the one month follow-up. The age groups ranged from 295 to 465 individuals with between 9.83% to 18.15% of the individuals by age group having started ART by the one month follow-up.

The  $\chi^2$  significance tests (**Table 1.**) indicate that age, CD4 count at study enrolment, education status, marital status and participants social capital were strongly associated with the ART uptake.

With increasing **age** of participant, the odds of ART uptake increased, participants in the oldest age group (45+ years) had twice the odds of the baseline group of those  $<25$  years old. Individuals in the 35-44 (OR 1.79 [1.13-2.83]  $p=0.01$ ) and 45+ (OR 2.03 [95%CI 1.25-3.31]  $p=0.003$ ) year old age groups had a strong association with ART uptake.

P-values ( $<0.0001$ ) indicated strong association to ART uptake for participants with **CD4 counts** of  $>351$  in cells/mm<sup>3</sup>. OR decreased with increasing CD4 count, where individuals

with CD4 counts of >500 in cells/mm<sup>3</sup> being 93% less likely of ART uptake than those individuals with the baseline CD4 count of ≤100 in cells/mm<sup>3</sup>

**Education** status was weakly associated with ART uptake (OR 0.74 [95% CI 0.54-1.0], p=0.05). Participants with secondary or tertiary level education were 26% less likely of commencing ART than those with none/primary education.

Participants whose **marital status** was divorced or separated were 80% more likely to commence ART (OR 1.80 [1.14-2.85], p=0.01) than a single person (baseline).

High **social capital** was strongly associated with ART uptake (OR 1.77 [1.33-2.35], p=0.0001), where strongly socially networked participants were 77% more likely to commence ART.

Evaluation of the effects of employment on ART uptake using the MH method (**Table 2.**) found **age** to be a potential confounder (adjusted MH OR 0.69 [0.52-0.92], p=0.01) as indicated by the statistically significant p-value and the OR 0.69 at variance with the Crude OR 0.77. Age was also found to be a potential effect modifier using the test of homogeneity ( $\chi^2$ , p<0.03).

**CD4 counts** ≤350 in cells/mm<sup>3</sup> (stratified OR 0.65 [0.46-0.93], p=0.02) were recoded as a single strata due to the small number of participants in the <100 strata. Controlling for CD4 count there was 35% lower likelihood of commencing ART if in employment (adjusted MH OR 0.65 [0.49-0.88], p=0.004). The MH analysis indicates CD4 count at study enrolment as a potential confounder.

Adjusted MH analysis indicates **education** as a potential effect modifier using the test of homogeneity, ( $\chi^2$ , p<0.02). Participants with none/primary education was found to be a possible stratum specific confounder (stratified OR 0.44 [0.25-0.76], p=0.003).

**Marital status** was considered as weakly associated with ART uptake if employed (adjusted MH OR 0.76 [0.57-1.00], p=0.05) due to the OR (0.76) being close to the crude OR (0.77) and the 95% CI cross 1.00.

Adjusted MH OR analysis indicated that participants **social capital** was not a confounder (adjusted MH OR 0.77 [0.59-1.02], p=0.07) or effect modifier using the test for homogeneity ( $\chi^2$ , p=0.59).

Multivariate logistic regression analysis (**Table 3.**) adjusting for the potential confounders age and CD4 count with **ART uptake when employed was 38% less likely** (adjusted OR 0.62 [0.46-0.83], p=0.001) than for the unemployed.

Using the likelihood ratio test (LRT) for interaction, stratified analysis indicated **education** participation as a strong effect modifier (LRT 4.79, p=0.03) for the association between employment status and ART uptake. None/primary education also indicated stratum specific interaction (OR 0.44 [0.25-0.76], p=0.003), which was the same under MH analysis. Employed participants with none/primary education were 56% less likely to commence ART.

**Age** group stratified analysis indicated stratum specific interactions for <25 years (OR 0.31 [0.12-0.78], p=0.01) and 35-44 years (OR 0.58 [0.35-0.96], p=0.04). However, the LRT for age group interaction (LRT 8.33, p=0.08), indicated no strong age related effect modification overall for the association between employment status and ART uptake. The 30-34 age stratum OR 1.76 was considered a chance finding due to its wide CI distribution (95% CI 0.89-3.51) and statistically insignificant p-value (p=0.11).

## **Discussion**

After using MH OR analysis, multivariate logistic regression and stratified analysis on the dataset, it was found that being employed statistically reduces the probability of commencing antiretroviral therapy one month after testing HIV-positive.

Age and CD4 count were shown to be variables confounding the association between employment status and ART uptake. Participants age and educational attainment were found to be interaction terms on the casual pathway between employment status and ART uptake.

Interpretation of the results was based on the statistical findings in the attached **Table 3**. Irrespective of the participants age, education status or CD4 count, the HIV-positive participants were offered ART. Which may explain why age and CD4 count were considered as confounders in the analysis when controlling for employment status. The effect modification of education along the casual pathway, (particularly in the none/primary strata (OR 0.44 [0.25-0.76],  $p=0.003$ )), may be due to generational differences in educational attainment, as the study population includes individuals  $\geq 15$  to 45+ years old.

The stratified  $<25$  year age group included individuals in their teens who may be in continued education and possibly working informally, while older age groups are unlikely to include individuals in education. Apart from the 30-34 age strata, the other older age groups had a similar likelihood of ART uptake.

Participants with none/primary education were found to be 69% less likely to commence ART, possibly due to work or family commitments of less well-educated individuals or due to issues of wellbeing or transport options.

More broadly, the free mobile community based HIV testing was offered to both employed and unemployed individuals. The free service may have allowed unemployed individuals who previously had limited access to ART due to cost of treatment and transport options be more engaged with commencing ART as a result.

Employed individuals may not have been contactable by telephone during working hours or may not have been able to attend the clinic when open (Mon-Fri 07:00-14:00 and two special clinics on weekends). They may wish to avoid being associated with the HIV clinic due to their work or social networks and may opt to buy ART from alternative and more discrete sources instead of accessing a free community clinic. This may lead to under reporting of ART uptake by employed individuals at the one month follow-up.

Table 1. Crude odds analysis presenting the association of individual risk factors with ART status.

Variables	No. of individuals	No. (%) of ART cases	Crude odds ratio (95% CI)	p value (chi <sup>2</sup> )	Test for trend of odds Pr (chi <sup>2</sup> )
<b>age groups (years)</b>					
<25	295	29 (9.83)	1.0	.	.
25-29	349	33 (9.46)	0.96 (0.57-1.62)	0.87	.
30-34	357	43 (12.04)	1.26 (0.76-2.07)	0.37	.
35-44	465	76 (16.34)	1.79 (1.13-2.83)	<b>0.01</b>	.
45+	303	55 (18.15)	2.03 (1.25-3.31)	<b>0.003</b>	.
				.	<0.0001
<b>sex</b>					
female	1081	150 (13.9)	1.0	.	.
male	688	86 (12.5)	0.89 (0.67-1.18)	0.41	.
<b>prior HIV test</b>					
no	627	86 (13.72)	1.0	.	.
yes	1142	150 (13.13)	0.95 (0.72-1.27)	0.73	.
<b>previously treated for TB</b>					
no	1614	212 (13.14)	1.0	.	.
yes	65	11 (16.92)	1.35 (0.69-2.62)	0.38	.
missing	90	13 (14.44)	1.12 (0.61-2.05)	0.72	.
<b>partners HIV status</b>					
no	1586	210 (13.24)	1.0	.	.
yes	183	26 (14.21)	1.09 (0.7-1.68)	0.72	.
<b>participants CD4 count at enrolment (in cells/mm<sup>3</sup>)</b>					
≤100	59	20 (33.90)	1.0	.	.
101-350	552	155 (28.08)	0.76 (0.43-1.35)	0.35	.
351-500	459	30 (6.54)	0.14 (0.06-0.27)	<b>&lt;0.0001</b>	.
>500	654	24 (3.67)	0.07 (0.04-0.15)	<b>&lt;0.0001</b>	.
missing	45	7 (15.56)	0.36 (0.13-0.97)	<b>0.04</b>	.
				.	0.71
<b>participants live in rural area</b>					
no	1096	133 (12.14)	1.0	.	.
yes	673	103 (15.30)	1.31 (0.99-1.73)	0.06	.
<b>type of dwelling</b>					
formal	1067	155 (14.53)	1.0	.	.
informal	589	69 (11.71)	0.78 (0.58-1.06)	0.11	.
missing	113	12 (10.62)	0.70 (0.38-1.30)	0.26	.
<b>participants highest level of education</b>					
none/primary	400	65 (16.25)	1.0	.	.
secondary/tertiary	1369	171 (12.49)	0.74 (0.54-1.0)	<b>0.05</b>	.
tabulation continued on next page.					

Variables	No. of individuals	No. (%) of ART cases	Crude odds ratio (95% CI)	p value (chi <sup>2</sup> )	Test for trend of odds Pr (chi <sup>2</sup> )
<b>transport used to access clinics</b>					
taxi/other	1078	142 (13.17)	1.0	.	.
walk	691	94 (13.60)	1.04 (0.78-1.37)	0.80	.
<b>participants reason for HIV test</b>					
routine	1427	184 (12.89)	1.0	.	.
other	342	52 (15.20)	1.21 (0.87-1.69)	0.26	.
<b>household size (individuals)</b>					
0	260	34 (13.08)	1.0	.	.
1-2	596	87 (14.60)	1.14 (0.74-1.74)	0.56	.
3-5	571	68 (11.91)	0.90 (0.58-1.40)	0.63	.
>5	220	34 (15.45)	1.22 (0.73-2.03)	0.46	.
missing	122	13 (10.66)	1.12 (0.61-2.05)	0.72	.
				.	0.42
<b>marital status</b>					
single	1103	138 (12.51)	1.0	.	.
married/cohabiting	534	71 (13.30)	1.07 (0.79-1.46)	0.66	.
divorced/separated	132	27 (20.45)	1.80 (1.14-2.85)	<b>0.01</b>	.
				.	.
<b>participants social capital</b>					
low	858	86 (10.02)	1.0	.	.
high	911	150 (16.47)	1.77 (1.33-2.35)	<b>0.0001</b>	.
<b>main household income</b>					
fulltime	806	102 (12.66)	1.0	.	.
informal	632	82 (12.97)	1.03 (0.75-1.41)	0.86	.
grant/pension	331	52 (15.71)	1.29 (0.90-1.85)	0.17	.
				.	.
<b>employment status</b>					
unemployed	763	115 (15.07)	1.0	.	.
employed	1006	121 (12.03)	0.77 (0.59-1.01)	0.06	.

Note: p-values indicating an association with ART uptake highlighted in bold

Table 2. The association between employment status and ART uptake, stratified by potential confounders using Mantel-Haenszel analysis.

Variables	Stratified odds ratio (95% CI)	p value (chi <sup>2</sup> )	Test of homogeneity Pr (chi <sup>2</sup> )
<b>Crude OR (employed)</b>	0.77 (0.59-1.01)	0.06	.
<b>Stratified analysis by risk factor categories</b>			
<b>age groups (years)</b>			
<25	0.31 (0.12-0.79)	<0.01	.
25-29	0.64 (0.31-1.34)	0.23	.
30-34	1.76 (0.88-3.52)	0.10	.
35-44	0.59 (0.35-0.97)	0.03	.
45+	0.64 (0.35-1.15)	0.13	.
MH (Adjusted OR)	0.69 (0.52-0.92)	<b>0.01</b>	<b>&lt;0.03</b>
<b>participants CD4 count at enrolment (in cells/mm<sup>3</sup>)</b>			
<=350	0.65 (0.46-0.93)	0.02	.
351-500	0.68 (0.32-1.43)	0.31	.
>500	0.59 (0.26-1.37)	0.22	.
missing	0.78 (0.15-4.07)	0.77	.
MH (Adjusted OR)	0.65 (0.49-0.88)	<b>0.004</b>	0.99
<b>marital status</b>			
single	0.62 (0.43-0.89)	<0.01	.
married/cohabiting	1.10 (0.65-1.88)	0.72	.
divorced/widowed	0.87 (0.37-2.04)	0.74	.
MH (Adjusted OR)	0.76 (0.57-1.00)	0.05	0.20
<b>participants highest level of education</b>			
none/primary	0.44 (0.25-0.76)	0.003	.
secondary/tertiary	0.95 (0.69-1.31)	0.75	.
MH (Adjusted OR)	0.77 (0.59-1.02)	0.07	<b>0.02</b>
<b>participants social capital</b>			
low	0.70 (0.45-1.10)	0.12	.
high	0.83 (0.58-1.17)	0.28	.
MH (Adjusted OR)	0.77 (0.59-1.02)	0.07	0.59

Note: p-values indicating confounders and effect modifiers highlighted in bold

Table 3. Logistic regression of interaction and confounding variables associated with ART after being controlled for employment.

<b>ARTstart</b>	<b>Odds Ratio</b>	<b>95% CI</b>	<b>p value</b>	<b>Likelihood Ratio Test for interaction (Pr (chi<sup>2</sup>))</b>
<b>OR adjusted for the confounders age and CD4 count</b>				
unemployed	1.0	.	.	.
employment	0.62	0.46-0.83	0.001	.
<b>The crude association between employment status and ARTstart stratified by interaction terms</b>				
<b>age groups (years)</b>				
<25	0.31	0.12-0.78	0.01	.
25-29	0.64	0.31-1.33	0.24	.
30-34	1.76	0.89-3.51	0.11	.
35-44	0.58	0.35-0.96	0.04	.
45+	0.64	0.35-1.15	0.14	.
				8.33 (0.08)
<b>participants highest level of education</b>				
none/primary	0.44	0.25-0.76	0.003	.
secondary/tertiary	0.95	0.69-1.31	0.75	.
				4.79 (0.03)