Analysis plan – Information, Education and Social Networks

Laura Derksen (Frisch Center) Catherine Michaud-Leclerc (ULaval) Ardyn Nordstrom (Carleton)

This document is our analysis plan for a separate paper investigating health knowledge and sexual health outcomes using data from the project "Information, Education and Social Networks". We are currently in the process of collecting follow-up data. We first describe how we analysed the short-term data and then discuss how we used this to inform our plan for analyzing the long-term data.

Short-term data. In 2018, we collected endline data on health knowledge, family preferences and sexual behaviour. We have already analysed this data following the two regression specifications below. The main regression specification follows our initial pre-analysis plan and the specification used in Derksen et al. (2022). For this project, instead of showing heterogeneity by median ability as in Derksen et al. (2022), we study heterogeneity by sex. Some of the questions asked are particularly important for girls and the gendered differences have more important policy implications than differences across other dimensions.

Main regression specification:

$$y_i = \alpha_0 + \beta T_i + t_i \ + \sigma_s + \varepsilon_i \quad (1)$$

Regression specification for heterogeneity by gender:

$$y_i = \beta^F T_i \times Female_i + \beta^M T_i \times Male_i + \ \delta \times Male_i + t_i \ + \sigma_s + \varepsilon_i \quad (2)$$

where:

- T_i : dummy variable that equals 1 if the individual i is in the treatment group and equals 0 if the individual is in the control group (with surveys).
- $Female_i (Male_i)$: dummy variables that equal 1 if the individual is female (male).
- σ_s: stratum fixed effects.
- t_i : time trend indicating the month-year of the long-run follow-up survey.
- Robust standard errors similar to Derksen et al. (2022).
- The outcome variables y_i are described below.
- We will test individual variables (denoted by V*) as well as indices (denoted I_*). Outcomes denoted by U* will be reported as indices only and unreported individually.
- We will report individual p-values as well as false-discovery-rate (FDR) q-values, calculated across the full set of hypothesis tests in the paper.

Regression specification for correlation between LGBT attitudes and social isolation:

$$y_i = \alpha + \beta^I Isolated_i + t_i \ + \sigma_s + \varepsilon_i \quad (3)$$

where:

- $Isolated_i$: dummy variable that equals 1 if the individual is socially isolated based on whether a student has been identified as having "few friends" (as opposed to "many friends").
- The outcomes y_i are indices I_7 and I_8 as defined below.

Long-term data. We follow a two-phase approach for analysing the long-term data. For many important outcomes, we are primarily interested in the impact on young women, as they are more vulnerable to HIV infection and pregnancy than young men. We are also unsure about the possibility to attain a reasonable attrition rate for young men. The first phase will analyse data for young women only, and if attrition for young men becomes more reasonable, we will follow the second phase.

Phase 1. Analyse data for young women only using equation (1). Most outcomes are particularly relevant for young women, making phase 1 a relevant approach. We will analyse all outcomes. For questions related to LGBT attitudes (RQ6a and RQ6b) we will test for correlation between LGBT attitudes and social isolation following equation (3).

Phase 2. If we can attain a level of attrition for young men below 33%, we will then switch to our usual approach to study the average treatment effect following equation (1) and heterogeneous treatment effects by gender following equation (2).

The outcome variables y_i are questions from the short- and long-term surveys and transformed in the way described below. The outcomes from research questions (RQs) 1a, 1b, and 1d are from the short-term survey, and all remaining outcomes (including 1d) are from the long-term survey. Indices will be constructed using Anderson score (Anderson, 2008).

#	Questions	Answers	Outcome	Index
RQ1: Does access to high quality online information impact long-run knowledge around key sexual health misconceptions (fertility and HIV transmission)?				
RQ1a. Short-run quiz questions - First Endline				
V1a.1	Imagine a woman who has a menstrual cycle every month. She will be fertile on certain days every month. This is called the fertile window. That means if she has sex on those days, she might get pregnant. How many fertile days does an average woman have every cycle?	(a) 1-4 days (b) 5-7 days (c) 7-12 days (d) more than 12 days (e) Don't know (f) Refuse to answer	= 1 if (b); 0 otherwise. Missing if refused.	I_1a

V1a.2	I want to ask you about the HIV transmission rate. Imagine an HIV-positive man who is NOT on treatment (ARVs) has unprotected sex with an HIV-negative woman one time. What is the approximate chance that she gets HIV?	(a) 100% chance -- she will definitely get HIV (b) 50% chance -- she has a 50/50 chance of getting HIV (c) 10% chance -- in a single time she will probably not get HIV (d) Less than 1% chance -- in a single time she will probably not get HIV (e) Don't know (f) Refuse to answer	= 1 if (d); 0 otherwise. Missing if refused.	I_1a
V1a.3	Do you think ARVs can stop a person living with HIV from spreading the virus?	(a) No (b) Yes, ARVs can reduce the chance of spreading the virus but most of the time the virus will still spread (c) Yes, as soon as a person starts taking ARVs they cannot spread the virus at all (d) Yes, a person who has been taking ARVs properly for a long time will almost never spread the virus (e) Don't know (f) Refuse to answer	= 1 if (d); 0 otherwise. Missing if refused.	I_1a
RQ1b. Short-run quiz questions - Second Endline				
V1b.4	I want to ask you about the HIV transmission rate. Imagine an HIV-positive man who is NOT on treatment (ARVs) has unprotected	(a) 100% chance -- she will definitely get HIV	= 1 if (d); 0 otherwise. Missing if refused.	I_1b

	sex with an HIV-negative woman one time. What is the approximate chance that she gets HIV?	(b) 50% chance -- she has a 50/50 chance of getting HIV (c) 10% chance -- in a single time she will probably not get HIV (d) Less than 1% chance -- in a single time she will probably not get HIV (e) Don't know (f) Refuse to answer		
V1b.1 to V1b.3	Same three questions V1a.1 to V1a.3 as above.			I_1b
RQ1c. Long-run quiz questions				
V1c.1 to V1c.4	Same four questions V1a.1 to V1a.3 and V1b.4 as above.			I_2
RQ1d. Misinformation index – are students wrong and certain?				
V1.5 V1d.1 to V1d.4	a. For each outcome V1.1 to V1.4: On a scale of 1-5, where 1 is not certain at all, and 5 is very certain, how certain are you of your answer? b. Same four questions as above.	1=Not certain at all 2=Not very certain 3=Neutral 4=Certain 5=Very Certain	=1 if wrong and certain (=4 or 5)	I_3
RQ2. Does access to high quality online information impact age of sex, marriage and pregnancy?				
V2.1	a. Have you ever had sexual intercourse? b. At what age did you have sexual intercourse for the first time?	a. Yes, No, Prefer not to say (pns), Don't know(dk). b. Age between 0 and current age.	=1 if first sex at age<18; 0 if older or never had sex. Missing if pns or dk.	I_4

V2.2	a. Are you currently married? b. Have you had any previous marriages? c. Year(s) of marriage. d. How old are you?	a. Yes, No. b. Yes, No. c. Number d. Number	=1 if first married before age 20; = 0 if older or not married	I_4
V2.3	a. Have you ever been pregnant? (female) b. Have you ever impregnated someone? (male) c. How old were you when the first pregnancy took place?	a. b. Yes, No, Prefer not to say, Don't know. c. Age	=1 if first child at age<20; = 0 if older or no children. Missing if pns or dk.	I_4
V2.4	a. These days, what contraception do you use? Select all that apply. b. In the past month, how often have you been using contraception?	a. Pull-out method, Monthly timing, Condom, Pills, Injectable, Vaginal ring, implant. b. 1. Never; 2. Rarely; 3. Sometimes; 4. Most of the time; 5. Always.	= 1 if unprotected sex (contraception) in past month	I_4
RQ3. Does access to high quality online information impact HIV risk taking?				
U3.1	How many sexual partners have you had in your life?	Number	= Number and 0 if never had sex (V2.1a)	I_5 (index only)
U3.2	How many sexual partners have you had in the past month?	Number	= 1 if Number>1	I_5 (index only)
U3.3	In the past month, how often have you been using a method to prevent HIV infection?	1. Never; 2. Rarely; 3. Sometimes; 4. Most of the time; 5. Always.	= 1 if unprotected sex (HIV) in past month	I_5 (index only)
V3.4	For each sexual partner (past month). How old is this sexual partner?	Age	= Average age of partners	

RQ4. Does access to high quality online information impact demand for HIV-related health care services?				
V4.1	Have you been tested for HIV in the past year? In the past year have you ever taken antiretroviral therapy (ARVs or ART)?	Yes, No, Prefer Not to Say, Don't know.	=1 if tested for HIV and/or on ART past year. Missing if pns or dk.	
RQ5. Does access to high quality online information impact knowledge about sexual and gender diversity?				
U5.1	Do you know what LGBT stands for?	After recording their definition of LGBT, we informed that "LGBT stands for lesbian, gay, bisexual, or transgender".	=1 if correct acronym; 0 = otherwise	I_6
U5.1	Do you know what it means for a person to be transgender?	After recording their definition of transgender, we informed them that "A transgender person is a person whose gender does not match their sex at birth. For example, a person who was born male but identifies as a woman, or a person who was born female but identifies as a man."	=1 if correct definition; 0 = otherwise	I_6
RQ6. Does access to high quality online information impact attitudes toward sexual and reproductive rights, and gender norms?				
RQ6a. Does access to high quality online information impact acceptance of homosexuality index?				
	Please tell me for each of the following actions whether you think it can always be justified, never be			

	justified, or something in between. We will use a scale from 1-10 where 1 is never justifiable and 10 is always justifiable			
U6.1a	Homosexuality	1-10	=score	I_7 (index only)
U6.1b	Homosexuality between two adult men (gay men)	1-10	=score	I_7 (index only)
U6.1c	Homosexuality between two adult women (lesbians)	1-10	=score	I_7 (index only)
U6.2	Do you agree with this statement: all homosexual activity should be criminalized and punished by law.	1=Strongly disagree 2=Disagree 3=Neither agree or disagree 4=Agree 5=Strongly agree	=score (reversed)	I_7 (index only)
U6.3	Do you agree with this statement: same-sex marriage should be legalized in Malawi.	1=Strongly disagree 2=Disagree 3=Neither agree or disagree 4=Agree 5=Strongly agree	=score	I_7 (index only)
RQ6b. Does access to high quality online information impact acceptance of transgender index?				
	Please tell me for each of the following actions whether you think it can always be justified, never be justified, or something in between. We will use a scale from 1-10 where 1 is never justifiable and 10 is always justifiable.			
U6.4a	For a person who was born male to live and dress like a woman.	1-10	=score (index only)	I_8 (index only)
U6.4b	For a person who was born female to live and dress like a man.	1-10	=score (index only)	I_8

				(index only)
U6.5	Do you agree with this statement: a person who was born male but lives life as a woman should be free to dress as a woman at work without facing discrimination.	1=Strongly disagree 2=Disagree 3=Neither agree or disagree 4=Agree 5=Strongly agree	=score (index only)	I_8 (index only)
RQ6c. Does access to high quality online information impact feminism index?				
	Please tell me for each of the following actions whether you think it can always be justified, never be justified, or something in between. We will use a scale from 1-10 where 1 is never justifiable and 10 is always justifiable.			
U6.6	Abortion	1-10	=score (index only)	I_9 (index only)
U6.7	Do you agree with the following statement? On the whole, men make better business executives than women do.	1=Strongly disagree 2=Disagree 3=Agree 4=Strongly agree	=reversed score (index only)	I_9 (index only)
U6.8	Do you agree with the following statement? A husband should help his wife with household chores such as cooking and cleaning.	1=Strongly disagree 2=Disagree 3=Agree 4=Strongly agree	=score (index only)	I_9 (index only)
RQ6d. Does access to high quality online information impact justifiability for sex before marriage?				
	Please tell me for each of the following actions whether you think it can always be justified, never be justified, or something in between. We will use a scale from 1-10 where 1 is never justifiable and 10 is always justifiable.			
V6.9	Sex before marriage	1-10	=score	
RQ7. Does access to high quality online information impact internet use patterns?				

V7.1	At what age did you get your first smart phone?	Age	=1 if smart phone before age 20	
V7.2	Last [Wednesday/Saturday], how many hours and minutes do you think you were personally using the phone?	Time (in minutes)	= total minutes using phone on Wed. and Saturday	
V7.3	Did you pay for full internet use in any other ways over the past month? Enter the full amount in MK.	Amount (in MK)	=amount of money spent on full internet access in past month	
V7.4	How many days did you use the internet over the past week?	Number	=number of days of internet use in the past week	
U7.5	a.What was the price of the bundle in MK? b.Which services are included in the bundle?	a.Value (MK) c. Internet, WhatsApp, Facebook, Other	= total of all bundle price values of WhatsApp, Facebook or other social media bundles ; 0 if no bundles purchased	I_10 (index only)
U7.6	I want to ask you about the applications you used last Wednesday/Saturday. Only include those that you personally used for at least ten minutes. Please refer to the numbers in the digital wellbeing dashboard. a. Which apps did you use last Wednesday?	a. WhatsApp, Facebook, Instagram, Chat GPT, TikTok, Google, Chrome, Browser, Email app, Wikipedia, Twitter, Vidmate, YouTube, Other.	=minutes using social media Wed/Sat.	I_10 (index only)

	b. What type of app is this? c. For how many minutes?	b. Email, browser, Messaging, Social media, Videos, Podcasts, Music, Games, Entertainment, News. c. Number		
U7.7	How often have you used [facebook/twitter/tiktok/whatsapp] in the past month?	1 = never 2 = rarely 3 = sometimes 4 = most of the time 5 = always	= score	I_10 (index only)
U7.8	Do you have a [facebook/twitter/tiktok/whatsapp] account?	Yes or no	= 1 if yes	I_10 (index only)
U7.9	Which internet or social media sites do you use for information about health topics?	Google Wikipedia The Nation Malawi Times Malawi Zodiak BBC WhatsApp Facebook Twitter TikTok Chat GPT Instagram Telegram	=1 if social media	I_11 (index only)
U7.10	Now, I am going to read a list of options. I want you to tell me what is the best place to find information about health.	Ranking 1-7: Books; Wikipedia; Google; Chat GPT; Social Media (SM); Health Facility; Friend or family member.	=reverse rank of SM (index only)	I_11 (index only)
V7.11	Now, can you open google chrome or whatever browser you usually use. I want you to look at your browsing history, and tell me how many different Wikipedia pages did you look at in the past week.	Number of different pages	=number of pages	

V7.12	Have you ever used Wikipedia?	Yes or no	=1 if ever used Wikipedia	
U7.13	Which internet or social media sites do you use for information about health topics?	Google Wikipedia The Nation Malawi Times Malawi Zodiak BBC WhatsApp Facebook Twitter TikTok Chat GPT Instagram Telegram	=1 if Wikipedia; 0 otherwise	I_12 (index only)
U7.14	Now, I am going to read a list of options. I want you to tell me what is the best place to find information about health.	Ranking 1-7: Books; Wikipedia; Google; Chat GPT; Social Media; Health Facility; Friend or family member.	=reverse rank of Wikipedia (index only)	I_12 (index only)

References

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, *103*(484), 1481-1495.

Derksen, L., Michaud-Leclerc, C., & Souza, P. C. (2022). Restricted access: How the internet can be used to promote reading and learning. *Journal of Development Economics*, *155*, 102810.