

Stata Users Breakout Session

Devon Kristiansen

2022-05-18

Setup

Load `.dta` file

```
. cd "z:\pma\admin\presentations\workshop2022"  
z:\pma\admin\presentations\workshop2022
```

```
.  
. use workshop_2022.dta
```

Result of the Female Questionnaire in Phase 1 vs Phase 2

```
. tab resultfq_2 resultfq_1, miss
```

result of female questionnaire	result of female questionnaire			Total
	completed	partly co	.	
completed	12,501	8	4,506	17,015
not at home	106	0	0	106
postponed	24	0	0	24
refused	87	0	0	87
partly completed	14	0	8	22
respondent moved	18	0	0	18
incapacitated	24	0	0	24
not interviewed (fema	4	0	0	4
not interviewed (hous	197	0	0	197
niu (not in universe)	1,352	1	0	1,353
.	1,987	25	0	2,012
Total	16,314	34	4,514	20,862

Dropping women who did not complete a survey in both surveys

```
. keep if resultfq_1 == 1  
(4,548 observations deleted)
```

```
. keep if resultfq_2 == 1  
(3,813 observations deleted)
```

Dropping women who were not part of the de facto population

```
. keep if (resident_1 == 11 | resident_1 == 22) & (resident_2 == 11 | resident_2 == 22)  
(358 observations deleted)
```

We'll call our dependent variable `category`

```
> gen category = .  
(12,143 missing values generated)  
  
. replace category = 1 if cp_1 == 0 & cp_2 == 0  
(5,107 real changes made)  
  
. replace category = 2 if cp_1 == 1 & cp_2 == 1  
(3,917 real changes made)  
  
. replace category = 3 if cp_1 == 0 & cp_2 == 1  
(1,939 real changes made)  
  
. replace category = 4 if cp_1 == 1 & cp_2 == 0  
(1,178 real changes made)
```

“Non-users” were not using a method at the time of *both* of their interviews.

“Users” were using a method at the time of *both* of their interviews.

```
. label define categorical 1 "Non-user" 2 "User" 3 "Adopted FP" 4 "Discontinued FP"
```

```
. label values category categorical
```

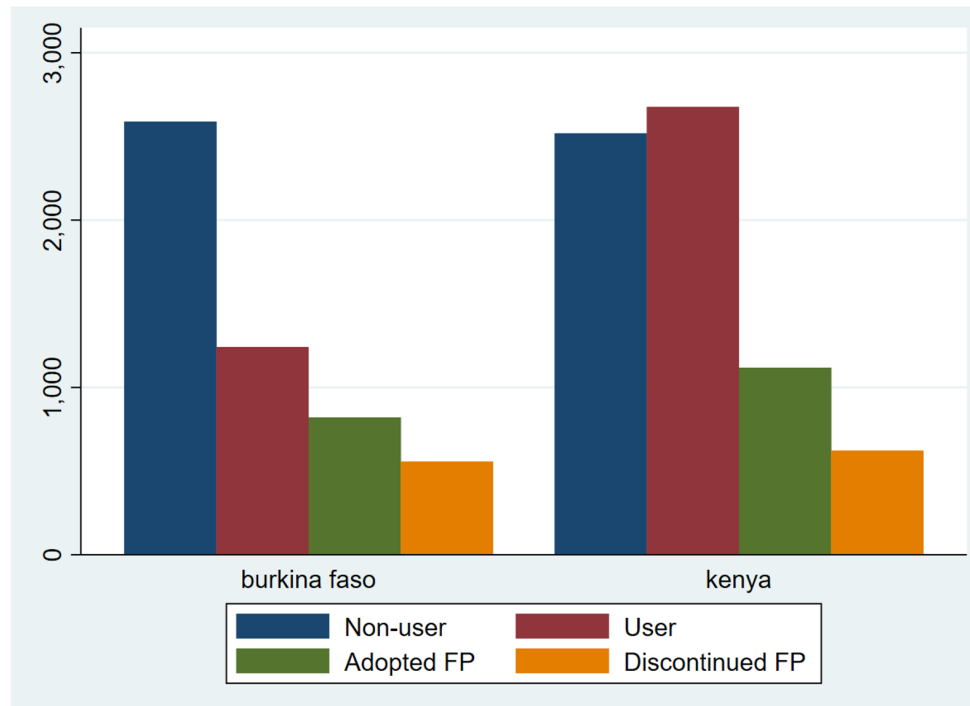
```
. tab category, gen(cat_)
```

category	Freq.	Percent	Cum.
-----+-----			
Non-user	5,107	42.06	42.06
User	3,917	32.26	74.33
Adopted FP	1,939	15.97	90.30
Discontinued FP	1,178	9.70	100.00
-----+-----			
Total	12,141	100.00	

Data Visualization

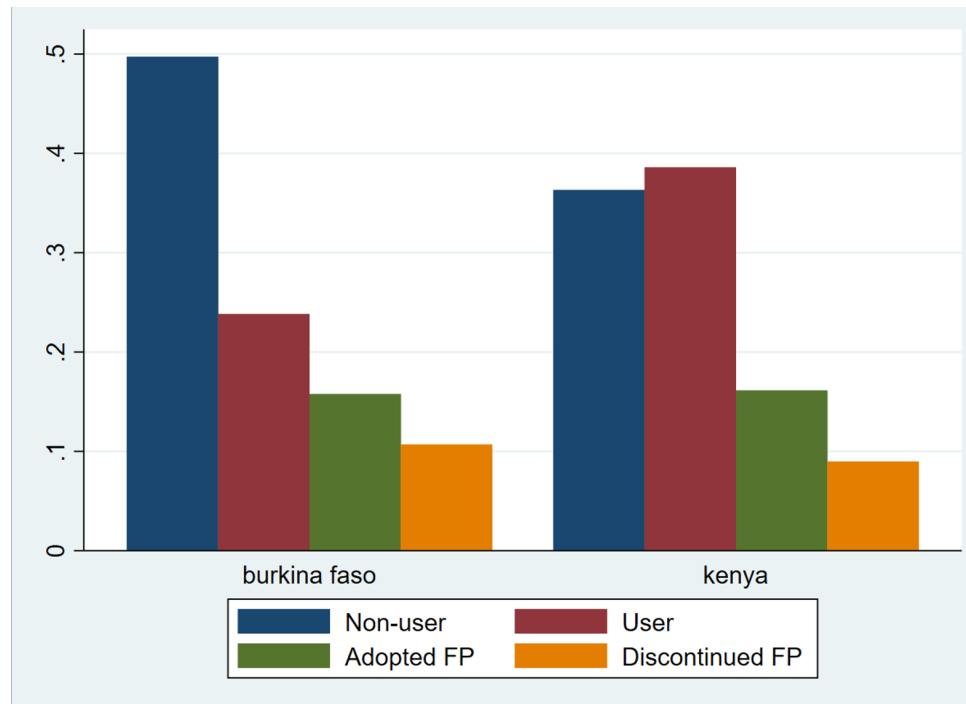
First graph uses counts of interviewed women

```
. graph bar (sum) cat_1-cat_4, over(country) legend(label(1 "Non-user")  
label(2 "User") label(3 "Adopted FP") label(4 "Discontinued FP"))
```



Second graph uses proportions, so the visualization isn't biased by a difference in sample sizes

```
. graph bar cat_1-cat_4, over(country) legend(label(1 "Non-user") label(2 "User")  
label(3 "Adopted FP") label(4 "Discontinued FP"))
```



Data Analysis

Rename outcome variable

```
. rename cat_3 adoption  
. rename cat_4 discontinue
```

Explanatory variables

```
. tab cvincomeloss_2, miss
```

income loss resulted from covid-19 restrictions	Freq.	Percent	Cum.
no	658	5.42	5.42
yes	7,566	62.31	67.73
don't know	2	0.02	67.74
niu (not in universe)	3,917	32.26	100.00
Total	12,143	100.00	

use hhincomelossamt to understand who did not lose income in cvincomeloss

```
. tab cvincomeloss_2 hhincomelossamt_2
```

income loss resulted	household income loss since covid-19				Total
	from covid-19	restrictions			
restrictions	none	partial	complete	no respon	
no	0	547	111	0	658
yes	0	5,449	2,117	0	7,566
don't know	0	2	0	0	2
niu (not in universe)	3,904	0	0	13	3,917
Total	3,904	5,998	2,228	13	12,143

```
. replace cvincomeloss_2 = 0 if hhincomelossamt_2 == 1  
(3,904 real changes made)
```

look at the other explanatory variable

```
. tab country covidconcern_2, row
```

pma country	concerned about getting infected					Total
	not conce	a little	concerned	very conc	currently	
burkina faso	212	461	955	3,576	1	5,208
	4.07	8.85	18.34	68.66	0.02	100.00
kenya	162	216	1,515	5,034	8	6,935
	2.34	3.11	21.85	72.59	0.12	100.00
Total	374	677	2,470	8,610	9	12,143
	3.08	5.58	20.34	70.91	0.07	100.00

pma country	concerned about getting infected	
	no respon	Total
burkina faso	3	5,208
	0.06	100.00
kenya	0	6,935
	0.00	100.00
Total	3	12,143
	0.02	100.00

replace NIU to missing

```
. forvalues i = 1/2 {  
  foreach var in age marstat educattgen cvincomeloss covidconcern  
  hhincomelossamt wealtht cp {  
    replace `var'`_i' = . if `var'`_i' > 90  
  }  
}  
(0 real changes made)  
(1 real change made, 1 to missing)  
(2 real changes made, 2 to missing)  
(0 real changes made)  
(0 real changes made)  
(0 real changes made)  
(2 real changes made, 2 to missing)  
(2 real changes made, 2 to missing)  
(0 real changes made)  
(0 real changes made)  
(1 real change made, 1 to missing)  
(15 real changes made, 15 to missing)  
(3 real changes made, 3 to missing)  
(13 real changes made, 13 to missing)  
(993 real changes made, 993 to missing)  
(0 real changes made)
```

Establishing the survey weight settings

```
. svyset [pw=panelweight], psu(eaid_1) strata(strata_1)
```

```
    pweight: panelweight  
      VCE: linearized  
Single unit: missing  
  Strata 1: strata_1  
    SU 1: eaid_1  
    FPC 1: <zero>
```

```
.
```

Demonstrating weighted proportions

```
. tab country adoption, row
```

pma country	category==Adopted FP		Total
	0	1	
burkina faso	4,386 84.23	821 15.77	5,207 100.00
kenya	5,816 83.88	1,118 16.12	6,934 100.00
Total	10,202 84.03	1,939 15.97	12,141 100.00


```
. svy: tab country adoption, row
(running tabulate on estimation sample)
```

```
Number of strata   =          23      Number of obs      =       12,141
Number of PSUs    =          474      Population size    =  12,134.981
Design df         =                   =          451
```

```
-----+-----
pma      |category==Adopted FP
country  |      0      1  Total
-----+-----
burkina  | .8503  .1497      1
kenya    | .8353  .1647      1
        |
Total    | .8418  .1582      1
-----+-----
```

Key: row proportion

Pearson:

```
Uncorrected  chi2(1)      =       5.0554
Design-based F(1, 451)   =       1.8652      P = 0.1727
```

Creating an age category recode

```
. recode age_2 (15/24=1) (25/34=2) (35/49=3), gen(age_rec)
(12143 differences between age_2 and age_rec)

. label define agerecode 1 "15-24" 2 "25-34" 3 "35-49"

. label values age_rec agerecode

.

. recode birthevent_2 (99=0) (0=0) (1/2=1) (else=2), gen(birth_rec)
(10389 differences between birthevent_2 and birth_rec)

. label define birthrcode 0 "No births" 1 "1 or 2 births" 2 "3+ births"

. label values birth_rec birthrcode
```

Logistic regressions

```
. svy: logit adoption i.age_rec urban i.wealtht_2 i.educattgen_2
cvincomeloss_2 i.covidconcern_2 if country == 1
```

Survey: Logistic regression

```
Number of strata =      2          Number of obs   =    4,884
Number of PSUs  =    167          Population size = 4,917.0532
                                          Design df    =     165
                                          F( 12, 154)  =     2.37
                                          Prob > F     =    0.0080
```

	adoption	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]

age_rec						
25-34	1.226821	.1934572	1.30	0.197	.898594	1.674938
35-49	.8174621	.1453411	-1.13	0.259	.5754526	1.16125
urban	1.146531	.177051	0.89	0.377	.8452188	1.555257

wealtht_2						
middle tertile	.9513569	.1448867	-0.33	0.744	.704291	1.285094
highest tertile	.7452848	.1341151	-1.63	0.104	.5224149	1.063234

educattgen_2						
primary/middle school	1.21569	.1879139	1.26	0.208	.895936	1.649561
secondary/post-primary	1.215347	.2058244	1.15	0.251	.8699221	1.697931
tertiary/post-secondary	1.613717	.4801887	1.61	0.110	.8967397	2.903944

cvincomeloss_2	1.226766	.1760785	1.42	0.156	.9240291	1.628688

covidconcern_2						
a little concerned	1.85145	.6998676	1.63	0.105	.8777511	3.905283
concerned	1.231492	.4213398	0.61	0.544	.6266885	2.419979
very concerned	1.692226	.557636	1.60	0.112	.8828585	3.243588

_cons	.0988454	.036659	-6.24	0.000	.0475266	.2055778

Note: _cons estimates baseline odds.

```
. svy: logit adoption i.age_rec urban i.wealtht_2 i.educattgen_2
cvincomeloss_2 i.covidconcern_2 if country == 7
```

Survey: Logistic regression

```
Number of strata = 21          Number of obs = 6,250
Number of PSUs  = 307        Population size = 6,154.4236
                                   Design df = 286
                                   F( 13, 274) = 2.50
                                   Prob > F = 0.0031
```

	adoption	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]
age_rec						
25-34	1.094059	.1216833	0.81	0.420	.8789546	1.361807
35-49	.8469074	.0952419	-1.48	0.141	.6787413	1.056739
urban	1.04929	.1261373	0.40	0.689	.8282016	1.329398
wealtht_2						
middle tertile	.9526526	.0966364	-0.48	0.633	.7802294	1.16318
highest tertile	.748508	.0988333	-2.19	0.029	.5771993	.97066
educattgen_2						
primary/middle school	1.110445	.2934678	0.40	0.692	.6600643	1.868135
secondary/post-primary	1.456977	.4119012	1.33	0.184	.8351936	2.541664
tertiary/post-secondary	1.70385	.5049793	1.80	0.073	.9507886	3.053364
cvincomeloss_2	1.145468	.1109547	1.40	0.162	.946633	1.386067
covidconcern_2						
a little concerned	.4658068	.1763542	-2.02	0.045	.2210914	.981386
concerned	.695045	.2078831	-1.22	0.225	.3857825	1.252228
very concerned	.7418588	.2221675	-1.00	0.320	.4114576	1.337573
currently / previously..	1.715145	1.741972	0.53	0.596	.2323332	12.66165
_cons	.2004338	.0769846	-4.18	0.000	.0941117	.4268726

Note: _cons estimates baseline odds.

```
. svy: logit discontinue i.age_rec urban i.wealtht_2 i.educattgen_2
cvincomeloss_2 i.covidconcern_2 if country == 1
```

Survey: Logistic regression

```
Number of strata =      2          Number of obs   =    4,884
Number of PSUs   =    167        Population size = 4,917.0532
                                   Design df       =     165
                                   F( 12, 154)       =     2.63
                                   Prob > F        =     0.0032
```

discontinue	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	

age_rec						
25-34	1.997721	.395008	3.50	0.001	1.352022	2.951793
35-49	1.502569	.3049388	2.01	0.046	1.006494	2.243148
urban	1.246211	.2167926	1.27	0.208	.8839372	1.75696

wealtht_2						
middle tertile	1.292635	.2078525	1.60	0.112	.9410084	1.775654
highest tertile	.8651644	.1962709	-0.64	0.524	.5528006	1.354031

educattgen_2						
primary/middle school	1.55014	.3041314	2.23	0.027	1.052288	2.283533
secondary/post-primary	1.590812	.3603386	2.05	0.042	1.017154	2.488004
tertiary/post-secondary	1.244363	.3600868	0.76	0.451	.7027667	2.203347
cvincomeloss_2	.9379963	.1561508	-0.38	0.701	.6752308	1.303017

covidconcern_2						
a little concerned	.9803776	.4661284	-0.04	0.967	.3834356	2.506653
concerned	1.198057	.5428279	0.40	0.691	.4897297	2.930881
very concerned	1.232284	.4641504	0.55	0.580	.5857734	2.592341
_cons	.0474947	.0207013	-6.99	0.000	.020086	.1123044

Note: _cons estimates baseline odds.

```
. svy: logit discontinue i.age_rec urban i.wealtht_2 i.educattgen_2
cvincomeloss_2 i.covidconcern_2 if country == 7
```

Survey: Logistic regression

```
Number of strata = 21          Number of obs = 6,244
Number of PSUs  = 307        Population size = 6,147.8281
                                   Design df = 286
                                   F( 12, 275) = 6.50
                                   Prob > F = 0.0000
```

discontinue	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
age_rec						
25-34	2.107144	.3068978	5.12	0.000	1.581951	2.806697
35-49	1.414137	.206907	2.37	0.019	1.060279	1.886092
urban	1.233563	.1806088	1.43	0.153	.9247091	1.645573
wealtht_2						
middle tertile	.9911905	.1392847	-0.06	0.950	.7516848	1.307009
highest tertile	1.003775	.1802205	0.02	0.983	.7049528	1.429266
educattgen_2						
primary/middle school	2.081409	.6423746	2.38	0.018	1.133806	3.820989
secondary/post-primary	2.098543	.7022264	2.22	0.028	1.08611	4.054733
tertiary/post-secondary	2.692685	.9507192	2.81	0.005	1.343911	5.395113
cvincomeloss_2	1.023195	.1439162	0.16	0.871	.7757561	1.349559
covidconcern_2						
a little concerned	8.217591	5.473967	3.16	0.002	2.214755	30.49041
concerned	4.33853	2.902794	2.19	0.029	1.162535	16.1912
very concerned	4.005398	2.649054	2.10	0.037	1.089666	14.72306
currently / previously..	1	(empty)				
_cons	.0069633	.0058409	-5.92	0.000	.0013359	.0362955

Note: _cons estimates baseline odds.

Logistic regressions with parity

```
. svy: logit adoption i.age_rec i.birth_rec urban i.wealtht_2 i.educattgen_2
cvincomeloss_2 i.covidconcern_2 if country == 1
```

Survey: Logistic regression

Number of strata	=	2	Number of obs	=	4,884
Number of PSUs	=	167	Population size	=	4,917.0532
			Design df	=	165
			F(14, 152)	=	3.42
			Prob > F	=	0.0001

adoption	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
age_rec						
25-34	.8558144	.1484611	-0.90	0.371	.6076139	1.205401
35-49	.5507701	.1085525	-3.03	0.003	.3732206	.8127839
birth_rec						
1 or 2 births	1.968136	.3218075	4.14	0.000	1.425108	2.718081
3+ births	2.454594	.5072408	4.35	0.000	1.632223	3.691304
urban	1.213194	.1921216	1.22	0.224	.8874379	1.658528
wealtht_2						
middle tertile	.9199756	.1427184	-0.54	0.592	.6772536	1.249687
highest tertile	.7194164	.1314295	-1.80	0.073	.5015637	1.031893
educattgen_2						
primary/middle school	1.278157	.1984294	1.58	0.116	.9407234	1.736628
secondary/post-primary	1.551715	.2709867	2.52	0.013	1.099163	2.190592
tertiary/post-secondary	2.419245	.7651474	2.79	0.006	1.295617	4.517342
cvincomeloss_2	1.209911	.1778908	1.30	0.197	.9050617	1.617441
covidconcern_2						
a little concerned	1.902249	.7444261	1.64	0.102	.8784095	4.119437
concerned	1.208323	.4226843	0.54	0.589	.6056523	2.410696
very concerned	1.652469	.5595275	1.48	0.140	.8468102	3.224634
_cons	.0613305	.0256176	-6.68	0.000	.0268846	.1399101

Note: _cons estimates baseline odds.

```
. svy: logit adoption i.age_rec i.birth_rec urban i.wealtht_2 i.educattgen_2
cvincomeloss_2 i.covidconcern_2 if country == 7
```

Survey: Logistic regression

```
Number of strata =      21          Number of obs   =      6,250
Number of PSUs   =      307        Population size = 6,154.4236
Design df        =                =      286
F( 15, 272)     =                =      6.69
Prob > F        =                =      0.0000
```

adoption	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	

age_rec						
25-34	.6956052	.0883514	-2.86	0.005	.5417376	.893175
35-49	.54574	.0830337	-3.98	0.000	.4045071	.7362841
birth_rec						
1 or 2 births	2.614024	.4046864	6.21	0.000	1.9274	3.545252
3+ births	2.379317	.4200688	4.91	0.000	1.680869	3.367989
urban	1.057397	.1325801	0.45	0.657	.826148	1.353375
wealtht_2						
middle tertile	.9646635	.0990405	-0.35	0.726	.7881573	1.180698
highest tertile	.7643795	.1040363	-1.97	0.049	.5847409	.9992051
educattgen_2						
primary/middle school	1.104853	.2896461	0.38	0.704	.6594879	1.850982
secondary/post-primary	1.565569	.4348617	1.61	0.108	.9062182	2.704654
tertiary/post-secondary	1.830081	.5345897	2.07	0.039	1.029833	3.252173
cvincomeloss_2	1.110472	.1094672	1.06	0.289	.914623	1.348259
covidconcern_2						
a little concerned	.4640682	.1750394	-2.04	0.043	.2208805	.9750038
concerned	.6900264	.2072154	-1.24	0.218	.3820877	1.246144
very concerned	.7342906	.2210459	-1.03	0.306	.4060122	1.327996
currently / previously..	2.038117	1.961597	0.74	0.460	.3065518	13.55047
_cons	.130007	.0499182	-5.31	0.000	.0610585	.2768136

Note: _cons estimates baseline odds.


```
. svy: logit discontinue i.age_rec i.birth_rec urban i.wealtht_2 i.educattgen_2
cvincomeloss_2 i.covidconcern_2 if country == 1
```

Survey: Logistic regression

```
Number of strata =      2          Number of obs   =    4,884
Number of PSUs  =    167          Population size = 4,917.0532
                                          Design df    =     165
                                          F( 14, 152)  =     6.09
                                          Prob > F     =    0.0000
```

discontinue	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	

age_rec						
25-34	1.420174	.2657685	1.87	0.063	.9814596	2.054994
35-49	1.100599	.2220481	0.48	0.635	.7389728	1.639191
birth_rec						
1 or 2 births	4.027007	.9259732	6.06	0.000	2.557468	6.340952
3+ births	3.454542	.9191819	4.66	0.000	2.042819	5.841858
urban						
urban	1.299959	.2309044	1.48	0.142	.9154154	1.846039
wealtht_2						
middle tertile	1.215252	.2034112	1.16	0.246	.8732487	1.6912
highest tertile	.7987195	.18717	-0.96	0.339	.5028657	1.268635
educattgen_2						
primary/middle school	1.575616	.2916368	2.46	0.015	1.093287	2.270736
secondary/post-primary	2.103299	.452831	3.45	0.001	1.37495	3.217474
tertiary/post-secondary	1.979558	.5905862	2.29	0.023	1.098354	3.567748
cvincomeloss_2						
cvincomeloss_2	.916188	.1532561	-0.52	0.601	.6584869	1.274741
covidconcern_2						
a little concerned	.9978152	.4683988	-0.00	0.996	.3949326	2.521026
concerned	1.163408	.5215497	0.34	0.736	.4800905	2.819296
very concerned	1.195911	.4439307	0.48	0.630	.5746353	2.48889
_cons						
_cons	.0195486	.0083054	-9.26	0.000	.008449	.0452301

Note: _cons estimates baseline odds.

```
. svy: logit discontinue i.age_rec i.birth_rec urban i.wealtht_2 i.educattgen_2
cvincomeloss_2 i.covidconcern_2 if country == 7
```

Survey: Logistic regression

```
Number of strata = 21          Number of obs = 6,244
Number of PSUs  = 307        Population size = 6,147.8281
Design df       = 286
F( 14, 273)    = 8.52
Prob > F       = 0.0000
```

discontinue	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	

age_rec						
25-34	1.34614	.2343759	1.71	0.089	.9555599	1.896367
35-49	.9021614	.1683259	-0.55	0.581	.6248727	1.302498
birth_rec						
1 or 2 births	2.617935	.5974912	4.22	0.000	1.670566	4.102552
3+ births	2.455745	.6056097	3.64	0.000	1.511391	3.990153
urban	1.247157	.1808523	1.52	0.129	.9374805	1.659127
wealtht_2						
middle tertile	1.000419	.140337	0.00	0.998	.7590481	1.318543
highest tertile	1.026956	.1840712	0.15	0.882	.7216631	1.461401
educattgen_2						
primary/middle school	2.067883	.6330899	2.37	0.018	1.131931	3.777738
secondary/post-primary	2.199786	.7279246	2.38	0.018	1.146872	4.219355
tertiary/post-secondary	2.863487	1.013747	2.97	0.003	1.426486	5.748079
cvincomeloss_2	.9949636	.1386367	-0.04	0.971	.7563068	1.30893
covidconcern_2						
a little concerned	8.394167	5.599473	3.19	0.002	2.258167	31.20321
concerned	4.393515	2.950779	2.20	0.028	1.171379	16.47885
very concerned	4.055277	2.690995	2.11	0.036	1.098452	14.97132
currently / previously..	1	(empty)				
_cons	.0043616	.0036492	-6.50	0.000	.0008403	.0226386

Note: _cons estimates baseline odds.