

INTERNET'S DIRTY SECRET: ASSESSING THE IMPACT OF ONLINE INTERMEDIARIES ON HIV TRANSMISSION

Jason Chan

Stern School of Business, New York University, 44 West Fourth Street, Suite 8-185,
New York, NY 10012 U.S.A. {jchan@stern.nyu.edu}

Anindya Ghose

Stern School of Business, New York University, New York, NY 10012 U.S.A. {aghose@stern.nyu.edu}
Korea University Business School, Korea University, Seoul 136-701, SOUTH KOREA

Appendix

Table A1. Craigslist Entry Timing

State	Entry City	Year	Entry Month	Avg. Population Density
New York	New York City	2000	8	406.95
Colorado	Denver	2001	4	45
Texas	Austin	2001	4	88.05
Arizona	Phoenix	2002	10	50.75
Florida	Miami	2002	10	324.3
Minnesota	Minneapolis	2002	10	64.2
Michigan	Detroit	2003	4	175.3
Louisiana	New Orleans	2003	11	104.15
Missouri	St Louis	2003	11	84.25
Nevada	Las Vegas	2003	11	21.4
North Carolina	Raleigh	2003	11	180.85
Ohio	Cleveland	2003	11	280.05
Indiana	Indianapolis	2004	2	175.35
Tennessee	Memphis	2004	2	145.95
Virginia	Norfolk	2004	2	190.9
Wisconsin	Milwaukee	2004	2	102
Alaska	Anchorage	2004	9	1.15
Idaho	Boise	2004	9	17.35
New Mexico	Albuquerque	2004	9	16
Utah	Salt Lake	2004	9	30.4
Nebraska	Omaha	2004	11	23.05
Oklahoma	Tulsa	2004	11	52.5
Iowa	Des Moines	2005	1	53.45
Kansas	Kansas City	2005	2	32.9
South Carolina	Columbia	2005	2	143.7
Alabama	Birmingham	2005	4	91.1
Arkansas	Little Rock	2005	4	53.7
New Jersey	New Jersey	2005	4	1169.85
Mississippi	Jackson	2005	6	61.9
West Virginia	Charleston	2005	6	76.15
North Dakota	North Dakota	2005	7	9.5
South Dakota	South Dakota	2005	7	10.35
Wyoming	Wyoming	2005	7	5.45

Table A2. Main Results with Additional Age Covariates

Variables	OLS					Negative Binomial				
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Binary Entry	0.145*	0.222**	0.193**	0.159*	0.240**	0.155***	0.226***	0.200***	0.151**	0.239***
	(0.08)	(0.09)	(0.07)	(0.09)	(0.09)	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)
Age 15-19 Proportion	-2.944	5.418	27.216	39.753	4.095	-6.625	8.572	30.030	53.187**	7.293
	(14.51)	(22.28)	(28.68)	(35.85)	(22.17)	(9.86)	(12.76)	(19.04)	(23.32)	(12.90)
Age 20-39 Proportion	-3.742	-6.116	-9.689	22.715	-6.680	-7.736	-6.724	-11.173	27.693**	-7.294
	(6.96)	(11.57)	(12.85)	(16.87)	(11.66)	(6.21)	(7.38)	(8.94)	(13.30)	(7.46)
Age 40-59 Proportion	-16.280*	-24.782**	-25.185*	-15.002	-23.734*	-16.795***	-24.652***	-24.161***	-17.882	-23.593***
	(8.21)	(12.01)	(13.64)	(18.71)	(11.82)	(6.19)	(7.70)	(9.14)	(14.30)	(7.65)
African American Proportion	-0.570	-0.886	-14.279	-35.436	0.079	-0.342	-0.714	-13.851*	-34.414**	0.242
	(12.45)	(15.53)	(16.00)	(28.46)	(15.06)	(6.31)	(8.27)	(7.87)	(16.03)	(8.09)
Log (Population Size)	0.718	-0.469	-1.418	-0.384	-0.455	0.733	-0.436	-1.165	-0.898	-0.425
	(0.80)	(1.05)	(1.81)	(2.10)	(1.05)	(0.57)	(0.76)	(1.05)	(1.43)	(0.76)
High School Attainment Proportion	1.617	1.765	-1.984	0.719	1.888	1.098	1.843	-2.371	0.709	1.992
	(1.49)	(1.80)	(1.77)	(1.55)	(1.74)	(1.09)	(1.29)	(2.16)	(1.37)	(1.29)
College Attainment Proportion	-0.985	-1.162	3.194*	0.051	-1.103	-0.722	-1.388	3.218**	-0.115	-1.327
	(1.06)	(1.30)	(1.73)	(1.51)	(1.26)	(0.90)	(1.05)	(1.42)	(1.12)	(1.05)
Log (Median Household Income)	-1.021**	-1.744**	-2.660**	-1.930**	-1.692**	-0.860**	-1.661***	-2.630***	-1.905***	-1.612***
	(0.50)	(0.74)	(1.03)	(0.82)	(0.73)	(0.37)	(0.47)	(0.69)	(0.47)	(0.47)
Log (No. Internet Lines)	0.095**	0.121	0.110	0.091	0.120	0.117***	0.135**	0.144	0.106	0.134**
	(0.04)	(0.09)	(0.09)	(0.09)	(0.09)	(0.04)	(0.07)	(0.09)	(0.07)	(0.07)
R-squared	0.271	0.301	0.394	0.226	0.309	-	-	-	-	-
Weighted by population		✓	✓	✓	✓		✓	✓	✓	✓
Controls × Time Trend			✓					✓		
P-Score Matched Samples				✓					✓	
Alternative Entry Label					✓					✓
Observations	306	306	306	194	306	306	306	306	194	306

The dependent variables for Models 1–5 are the log number of HIV cases, while the dependent variables for Models 6–10 are the number of HIV cases. Robust standard errors are reported in parentheses below coefficient values. Models 1 and 6 are unweighted regressions, while Models 2–5 and 7–10 are weighted regressions. An alternative definition of entry year is used for Models 5 and 10. Locations that have Craigslist entry in the final quarter of the year (i.e., October–December) are labeled as having site entry in the following year (i.e., $t + 1$). All models have binary entry regressors and include state and year fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3. Additional Alternative Model Specifications

Variables	OLS		Negative Binomial	
	Model 1	Model 2	Model 3	Model 4
Binary Entry	0.102** (0.04)	0.087** (0.04)	0.142* (0.08)	0.132** (0.05)
R-squared	0.145	-	0.540	-
State × Time Trend			✓	✓
Observations	66	66	106	106

The dependent variables in Models 1 and 2 are the number of HIV cases, averaged over the pre and post entry periods. The average values are being logged in Model 1. Models 1 and 2 are made up of two observations per state (33 states), which results in 66 observations. In Models 3 and 4, we run panel regressions with the log number of HIV cases and the number of HIV cases, respectively. The interaction terms of state dummies with time trend are included as covariates in Models 3 and 4. Robust standard errors are used for Models 3 and 4. State and year fixed effects are included for Models 3 and 4. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A4. Craigslist Entry on Syphilis

Variables	OLS		Negative Binomial	
	Binary Entry	Personal Ads	Binary Entry	Personal Ads
	Model 1	Model 2	Model 3	Model 4
Craigslist Presence	0.188** (0.09)	0.155** (0.07)	0.165** (0.08)	0.149** (0.06)
Age 15-19 Proportion	93.359*** (28.24)	102.102*** (34.28)	91.787*** (18.92)	108.598*** (28.69)
Age 20-39 Proportion	76.825*** (23.13)	86.726*** (28.50)	81.180*** (13.20)	99.523*** (17.28)
Age 40-59 Proportion	86.224*** (28.06)	77.208** (33.47)	87.608*** (18.21)	77.511*** (20.33)
African American Proportion	-1.657 (9.74)	11.145 (31.40)	-1.790 (6.84)	3.945 (15.82)
Log (Population Size)	7.234* (3.64)	5.489 (3.83)	7.073*** (2.05)	5.425** (2.32)
High School Attainment Proportion	-3.812 (2.78)	-7.325** (3.40)	-5.072** (2.44)	-7.759** (3.10)
College Attainment Proportion	-1.703 (2.60)	-2.949 (3.35)	-1.574 (2.07)	-3.247 (2.57)
Log (Median Household Income)	0.029 (0.94)	-0.494 (1.06)	0.390 (0.66)	-0.126 (0.78)
Log (No. Internet Lines)	-0.004 (0.15)	0.139 (0.18)	-0.004 (0.14)	0.170 (0.16)
R-squared	0.29	0.27	-	-
Observations	491	342	491	342

The dependent variable for Models 1 and 2 is the log number of HIV cases, while that for Models 3 and 4 is the number of HIV cases. Robust standard errors are reported in parentheses. State and year fixed effects are added for all models. *p < 0.10, **p < 0.05, ***p < 0.01.