

Internet's Dirty Secret: Assessing the Impact of Online Intermediaries on HIV Transmission*

(Accepted at *MIS Quarterly*)

Jason Chan
New York University
Stern School of Business
44 West Fourth Street, Suite 8-185
New York, NY 10012
United States
jchan@stern.nyu.edu

Anindya Ghose
New York University
Stern School of Business
44 West Fourth Street, Suite 8-94
New York, NY 10012
United States
aghose@stern.nyu.edu

Abstract

Online platforms offer access to a larger social group than is generally available through offline contacts, making the internet an emerging venue for seeking casual sex partners. The ease of seeking sex partners through classified ad sites may promote risky behaviors that increase transmission of STDs. In this paper, using a natural experiment set up, we investigate whether the entry of a major online personals ad site, Craigslist, increases the prevalence of HIV over a 10 year period from 1999 to 2008 across 33 states in the United States. After controlling for extraneous factors, our results suggest that the entry of Craigslist is related to a 15.9 percent increase in HIV cases. Our analysis suggests that the site entry produces an average of 6130 to 6455 cases of HIV infection in the U.S. each year, mapping out to \$62 million to \$65.3 million in annual treatment costs. In addition, the analyses reveal that non-market related casual sex is the primary driver of the increase in HIV cases, in contrast to paid transactions solicited on the site (e.g., escort services and prostitution) which has a negative relationship with HIV trends. These findings are essential to the understanding of the social routes through which HIV transmission takes place and the extent to which site entry can influence HIV trends. Implications for healthcare practitioners and policy makers are discussed.

Keywords: classified ad sites, HIV, internet, online intermediaries, transmission route

* The authors thank Allan Collard-Wexler, Gordon Gao, Avi Goldfarb, Shane Greenstein, Robert Seamans, Rahul Telang Catherine Tucker, and seminar participants at the 2nd Workshop of Health IT and Economics (WHITE 2011), the 2012 International Industrial Organization Conference, the 34th Annual NBER Summer Institute, the 2012 Statistical Challenges in E-Commerce Research (SCECR), and the 2013 NET Institute Conference for extremely valuable suggestions. We acknowledge the NET Institute for financial support. The usual disclaimers apply.

Internet's Dirty Secret: Assessing the Impact of Online Intermediaries on HIV Transmission

1. Introduction

Over the last decade, the internet has emerged as a popular platform for seeking sex partners. The expansive reach offered by the internet provides access to a larger social group than is generally available through offline contacts, making it easier for individuals to locate similar others who are seeking casual sex (Garofalo et al. 2007). In particular, the introduction of online classified advertising sites and hook up sites has greatly facilitated the search for casual sex partners. The availability and anonymity in individual browsing coupled with non-costly posting of online ads result in an increased access to potential sex partners and subsequently, a heightened frequency of casual sex encounters (Moskowitz and Seal 2010, Grov 2010). In addition, the convenience brought by online personal ad listings in facilitating casual hook ups produces a shift in individual's casual sex-seeking behavior. Interviews reveal that not only do users search for strangers online to have no-strings attached relationships; they also post ads to solicit more diverse sexual experiences with multiple partners.^{1,2} A major risk of having casual sex with online strangers involves the possibility of contracting sexually transmitted diseases (STDs). No matter how carefully one tries to vet one's partners, it is typically difficult to accurately know their STD status. Consequently, the increase in casual sex frequency and sexual partners are likely to induce a greater incidence of STDs within the population.

¹ For instance, a 32-year old newspaper editor posted an ad in which she declared December 2002 to be a moral free zone, where she would have sex with whoever she wants with minimum self restraint. Her ad received responses from 70 men of which she met up with ten. Report available at http://nymag.com/nymetro/news/features/n_8227.

² Personal ads are also being used by individuals to look for extramarital affairs as exemplified by the recent case of the married Congressman, Christopher Lee. In an attempt to solicit a woman found on Craigslist's personals section, the Congressman emailed his shirtless photo to her. Lee made an abrupt resignation from his Upstate New York seat when news of his correspondence was reported online. Report available at <http://www.washingtonpost.com/wp-dyn/content/article/2011/02/10/AR2011021003020.html>.

STDs are a major public health concern that imposes both physical strains and psychological stigmatization on infected patients. Apart from these consequences, STDs impose a heavy economic toll on society: it costs the U.S. healthcare system \$16.4 billion to treat 19 million new STD infections each year.³ Among various STDs, the HIV epidemic stands out as an important public health problem that involves widespread mortality and morbidity, and has crucial social and economic consequences (UNAIDS 2009). Tracking the progression of HIV infection and identifying the source of its transmission routes is necessary for effective intervention and treatment of the affected population. As online technology increasingly permeates society and personal lives, its usage warrants closer examination to understand its potential impact on disease transmission. Exogenous technology shocks via the entry of online classified ad sites can disrupt the control over HIV proliferation and undermine the effectiveness of HIV prevention programs. Though the relationship between online classified ads sites and HIV progression bears crucial importance, we have limited understanding on how the increased usage of online classified ad sites affects HIV trends over time (Chiasson et al. 2006, Grov 2010).

This study aims to investigate whether the entry of a major classified ad website, Craigslist, increases the prevalence of HIV cases. Craigslist is a website featuring free online classified ads with sections devoted to jobs, housing, personals, items and services for sale. It was started in 1995 as an email distribution list of friends in San Francisco Bay area. As the service got more popular in 1996, a web interface was established. Following that, Craigslist was incorporated as a company in 1999 and began to expand to other locations. Craigslist expanded steadily over the

³ Figures reported by Center for Disease Control and Prevention on <http://www.cdc.gov/std/stats09/trends2009.pdf>

years and is present in more than 700 local sites in 70 countries in 2011.⁴ A typical personal ad post on Craigslist consists of an open-ended text requesting for meet up, a telephone number, a randomly generated Craigslist email address, and photographs of the poster. Personal ads are classified into subcategories that denote the relationship type that users are seeking for (i.e., men seeking men, men seeking women, women seeking men, women seeking women). Upon closer examination of the content of personal ads, it is observed that numerous ads are posted with the interest of soliciting casual sex partners.

The synchronicity between the presence of Craigslist in a given region and HIV incidence in that region sets the initial motivation to suspect a relationship between the two. Figure 1 illustrates the patterns in HIV incidence and Craigslist's presence in different regions of the United States (U.S.). In the East Central area, the annual number of new HIV cases was relatively stable from 2001 to 2004. An upward trend in HIV incidence is observed in the region as more Craigslist sites were launched post-2005 and is most acute when the number of personal ads increases past the 31,200 mark. In the West Central and Mountain areas, the peaks in HIV prevalence are falling prior to 2002. As the activity levels on Craigslist sites grew after 2003, the peaks in HIV incidence began to rise accordingly.

It is posited that the increase in HIV prevalence results from the entry of Craigslist sites in the country. In particular, we argue that the introduction of Craigslist provides users with the ability to locate and socialize with specific groups of individuals who are inclined to engage in casual sex but are otherwise hard to locate in one's day-to-day life and community. By presenting

⁴ It is observed that Craigslist does not actively advertise about its entry in new locations. Traffic to the site is driven mainly by word of mouth. See <http://www.craigslist.org/about/factsheet> for more information on its entry characteristics, accessed August 2011.

participants with key information regarding sexual encounters, including (i) location and time of meet up, (ii) physical characteristics of potential partners (e.g., ethnicity, age, physical build, voluptuousness), (iii) societal characteristics (e.g., wealth, job, marital status), (iv) sexual orientation and (v) specific expectations and requests of intimate exchange, personal ads not only signal sexual availability of individuals but also facilitate the matching of casual sex partners.

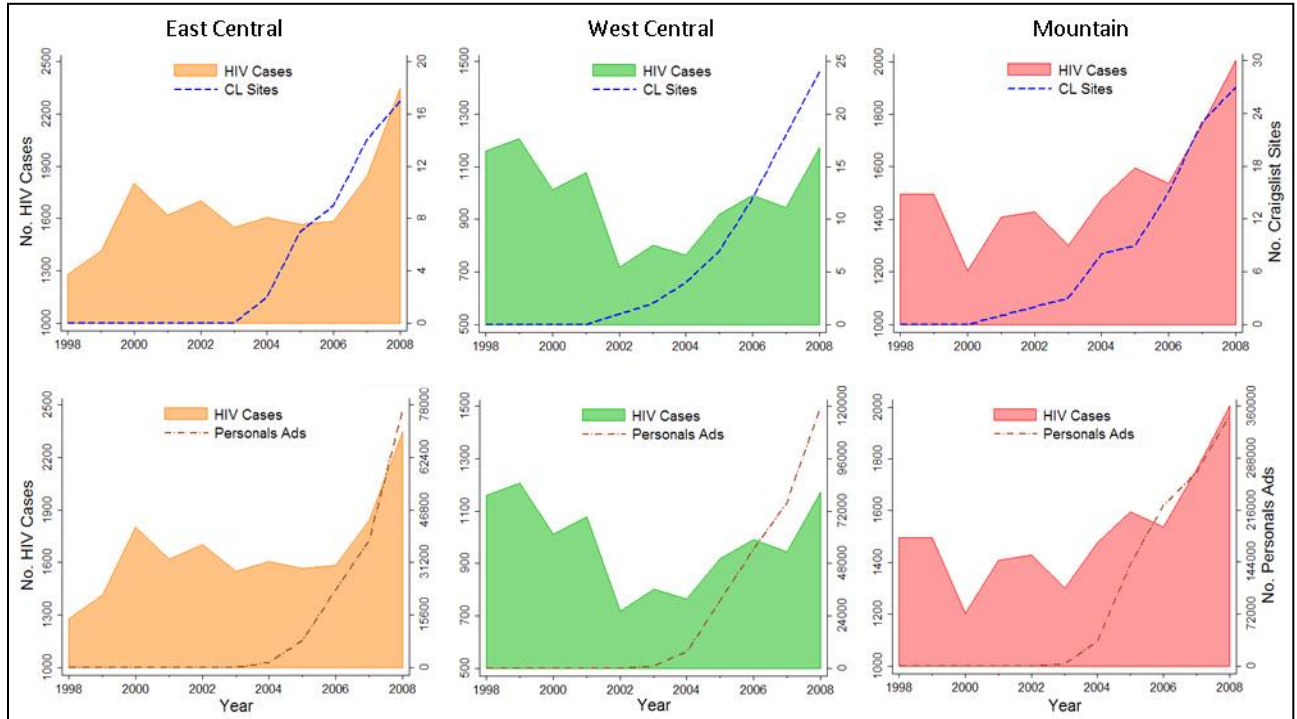


Figure 1: The relationship between HIV trends and Craigslist entry and activity level

Using a national panel data set constructed from six sources, we examine the longitudinal relationship between Craigslist’s entry and HIV trends in 33 U.S. states from 1999 to 2008. To identify the entry effects on HIV proliferation, we rely on a natural experiment setup inherent in Craigslist's expansion in the United States. During its expansion, the classified ad site was made available in certain locations at each time period, thereby providing an exogenous source of variation in site entry across states and years. Exploiting the natural experiment framework, we run panel regressions of HIV incidence on site entry with state and year fixed effects, and included multiple controls to account for demographic features, socioeconomic factors and

internet accessibility which may affect HIV transmission. A series of checks are performed to examine the robustness of the main results. In particular, we run count data models to account for the count nature of the dependent variable, included time-varying state covariates to control for effects not captured by the state and year fixed effects, relied on propensity score matching techniques to address potential confounding effects from unobservables, used the number of posted personals ads as a more precise measure of site activity, and performed falsification checks to ascertain the estimated effect was not spurious. In establishing validity of our analyses, a set of systematic checks was performed to examine whether Craigslist's entry is exogenous with respect to HIV trends. We also examine the possibility of increased HIV testing leading to a greater prevalence of HIV incidence during the period of site expansion. Finally, we conduct a set of tests to understand the underlying mechanism that drives HIV trends. In particular, we test how each ad category within Craigslist's personals section affects HIV incidence.

Our empirical analyses reveal that the entry of Craigslist leads to a 15.9 percent increase in HIV cases. Conservative estimates from our robustness checks suggests that the site entry produces between 6130 and 6455 cases of HIV infection in the U.S. each year on average, which maps out to between 62 million and 65.3 million dollars worth of annual treatment costs. We find that this result is robust under various alternative model specifications. Falsification tests show that the relationship between site entry and HIV trends did not arise spuriously, and a pre-entry trend leading to an increase in HIV cases is not present. We also note that a one year delay exists between the site entry and the increase in HIV trends. Additionally, results from our empirical checks corroborate the belief that Craigslist's entry in different locations is exogenous with respect to HIV trends. Our analyses also show that the increase in HIV incidence is unlikely to

be driven by better diagnosis as the rate of HIV testing did not increase over the study period. Finally, we find that the rise in HIV as a consequence of entry is attributed to casual hookups solicited on the site as opposed to market related sexual transactions (i.e., prostitution activities).

Our paper aims to make a few key contributions to the literature. We are among the first in the emerging literature to research the link between the diffusion of online intermediaries and disease propagation by assessing the longitudinal impact of a classified ad site entry on HIV incidence in the U.S. This research effort provides an initial assessment on whether the launch of Craigslist is related to HIV transmission, and quantifies the entry impact on societal levels of the HIV epidemic. Study results can inform academicians, policy makers, healthcare practitioners and other relevant stakeholders in various IS-related realms on the social and health impacts of online intermediaries, and provide insights on the economic significance of entry effects of classified ad sites on HIV transmission. Second, by augmenting our main data set with a health behavioral survey, we are able to address the question of whether the increase in HIV incidence after Craigslist's entry is due to an increased frequency in HIV testing or an actual increase in disease transmission arising from a greater number of casual hook ups. Given that disparate policy implications are appropriate for each of two reasonings, it is imperative to unravel the underlying cause leading to the observed study results. Third, the rich nature of our dataset allows us to unpack the casual mechanisms in the relationship between classified ad sites and HIV trends. In particular, we provide deeper insights on how the use of classified ad sites leads to greater HIV prevalence by identifying the ad category within Craigslist that is related to the observed increase in HIV trends. Identification of the user group at risk of HIV transmission

serves as a crucial input for the effective design and targeting of STD awareness programs, epidemic intervention, and healthcare treatment.

The rest of the paper is organized as follows. We discuss related literature in Section 2. In Section 3, we describe the data used in the study. We then delineate the empirical methodology and the set of empirical tests in Section 4. In Section 5, we report the results of our analyses. Finally, we conclude with a discussion of study limitations, future work and study implications in Section 6.

2. Literature Review

In this section, we draw upon various relevant literatures to delineate the relationship between the presence of classified ad sites and HIV trends. We also discuss other literature related to the current study context.

2.1 Online Classifieds and HIV Trends

We relate the link between the introduction of classified ad sites and HIV prevalence to two pertinent streams of literature – choices motivated by the immediacy in gratification and changes in market dynamics due to online intermediation. Extant research shows that people prefer immediate rewards even when the delayed rewards are considerably larger (Rachlin 2000, Romer et al. 2010). In the current study context, given the choice to have casual sex and the choice to enjoy STD/HIV free status via abstinence, individuals are inclined to choose the former because of the immediate gratification that casual sex offers. The choice of seeking for casual sex is an impulse that is automatic, fast and emotionally driven, while abstinence involves self regulation that is effortful, slow and cognitively driven (Metcalf and Mischel 1999, Strack and

Deutsch 2004). The newfound ease in soliciting casual partners via online classified ads is likely to draw one's attention towards the immediate reward of physical pleasure, making the efforts to self control one's sexual desires frustrating, tiring, and draining (Karniol and Miller 1983, Baumeister and Heatherton 1996). Researchers have also theorized that individuals devalue delayed rewards as its receipt is less certain than that of an immediate reward (Green and Myerson 2004, Wilson and Daly 2006). As such, individuals are inclined to discount the future value of staying STD/HIV free and put high value on the instant gratification that casual sex offers.

As an online intermediary that facilitates the matching of sex-seeking individuals, classified ad sites differ from physical hookup venues and traditional advertising mediums in three important ways. First, communication and exchanges facilitated by the online intermediary are not limited by spatial and temporal constraints (Bailey and Bakos 1997, Jin and Robey 1999). Unlike hooking up in a bar or lounge, individuals seeking casual partners on classified ad sites need not be physically present at a specific location at the same time. Also, the duration from posting an ad to receiving responses is considerably shorter on classified ad sites compared to that on newspapers. Second, the search costs associated with seeking casual partners on an online intermediary is greatly reduced (Bakos 1997, Il-Hann and Terwiesch 2003). Classified ad sites aggregate sexually available individuals and facilitate explicit communication about sexual interests and desired behaviors through written text and photos, thus making the travelling costs and screening costs inherent in physical hookup situations redundant. In addition, compared to ad postings in newspapers, the posting of personal ads on Craigslist is feeless, contributing to another source of search cost reduction. Third, the variety of offerings that can be solicited from

an online intermediary is broader than its counterparts (Brynjolfsson et al. 2003, Jin and Robey 1999). The wide variety of casual partners available on classified ad sites is likely to arise as barriers of spatial limitation and search costs are removed. Taken jointly, these features of online intermediation improve the efficiency of the market for casual sex by allowing individuals to identify and meet with large number of anonymous sex partners they might not meet otherwise, increasing the risk for HIV transmission for locations with Craigslist.

Choosing to have casual sex despite infection risks, together with the ease in seeking casual partners from classified ad sites can create an upward shift in HIV prevalence in two possible fashions. On the one hand, HIV transmissions can increase due to the heightened willingness to have sex with online strangers, with little concern over HIV infection. Surveys reveal that substantial proportions of individuals who met partners online for sexual encounters had multiple sex partners and did not use condoms (Benotsch et al. 2002, Padgett 2003). Compared to offline sex seekers, significantly more online seekers, including HIV negative individuals, are engaged in unprotected anal intercourse, have sexual exposure to persons known to be HIV positive, and have reported less worry about HIV infection due to improved HIV treatment (Elford et al. 2001, Kim et al. 2001, McFarlane et al. 2000, Liao et al. 2006). On the other hand, HIV prevalence may also be increasing as a result of uninfected users having sexual acts with infected individuals without knowledge of their partners' true HIV status. Though a sizable population of HIV positive individuals is seeking casual partners online (Liao et al. 2006), infected individuals have either withheld their HIV status or misrepresented their HIV status to prospective sex partners met online (Carballo-Diéguez et al. 2006, Moskowitz and Seal 2010). Irresponsible and risky behavior of having sexual contact with online partners prior to HIV testing is also observed

(McFarlane et al. 2000). Furthermore, a large proportion of online sex seeking users who are HIV positive do not practice serosorting⁵ and have unprotected sex with online partners, thereby raising the risk of infection rate among HIV negative users (Halkitis and Parsons 2003, Hirshfield et al. 2004).

2.2 Other Related Literature

Our study is related to two other sets of research. First, this paper contributes to the emerging stream of literature that investigates the impact of the internet and information technologies on medical related outcomes. By studying this relationship, we examine an increasingly important area of research that has managerial implications for site owners and healthcare practitioners. Instances of such work include Gao et al. (2012) on whether patients' online ratings reflect physician quality, Yan and Tan (2010) on the social supports from online communities enhancing patients' health outcomes, Miller and Tucker (2011) on electronic medical records improving neonatal outcomes, Ayyagari et al. (2011) on how various components of information technology influence stress and well-being, and Eysenbach (2008) on the potential link between internet use and cancer outcomes.

Second, our study builds on and relates to the literature that examines the impact of matching platforms on market outcomes. This stream of work investigates the economic and social consequences that arise due to the matching role facilitated by social websites that have emerged over the last decade. For instance, Hitsch et al. (2010) examine the economic mechanisms that underlie mate selection on online dating sites and show that these sites facilitate the efficient matching of market participants based on individual preferences. Seamans and Zhu (2013)

⁵ Serosorting is the practice of choosing sexual partners based on their HIV status.

quantify the entry impact of matching platforms on strategic choices in the newspaper market. Through their empirical study, it is found that affected newspaper owners raised the subscription prices and increased its level of differentiation upon Craigslist's entry into the local market. In addition to this result, Gurun and Butler (2012) find that the local new newspapers increased their usage of slants in response to Craigslist's entry to their market. Kroft and Pope (2013) find a reduction in real estate vacancy rates as a result of increased matching efficiency brought by Craigslist. Logan and Shah (2012) examine the information dynamics of an online matching platform that features male sex workers. They demonstrate that the pictures and text reviews of sex workers on the website affect the prices of sex workers positively.

3. Data

To fulfill the goal of investigating the impact of Craigslist's entry on HIV trends, we constructed a nationwide panel data for 33 U.S. states across a period of 10 years using six disparate datasets. Summary statistics of the dataset is provided in Table 1. Spanning from 1999 to 2008, we consolidated the annual number of HIV cases from the HIV Surveillance Reports, maintained by the Center of Disease Control and Prevention (CDC).⁶ The CDC constructs data on HIV incidence using confidential name-based HIV infection reports provided by various states. This method of HIV reporting achieves high levels of accuracy and reliability. In our sample, we only included states that have initiated HIV infection reporting before Craigslist's launch to ensure that we observe HIV trends at each location before and after entry. Using pre-entry HIV figures as a counterfactual against the post-entry HIV trends, we are able to infer whether site entry results in an increase in HIV cases.

⁶ Collected from <http://www.cdc.gov/hiv/topics/surveillance/resources/reports/past.htm>.

Table 1: Summary Statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
Log No. HIV Cases	306	5.627	1.565	1.099	9.347
Craigslist Entry	306	0.533	-	-	-
Log Avg. Personal Ads Posted Daily	306	1.429	1.575	0	5.669
Log Avg. Erotic Ads Posted Daily	306	1.638	1.803	0	5.720
Age 20-39 Proportion	306	0.277	0.014	0.254	0.316
High School Attainment Proportion	306	0.856	0.040	0.771	0.930
College Attainment Proportion	306	0.248	0.044	0.151	0.387
African American Proportion	306	0.117	0.097	0.006	0.375
Log Median Household Income	306	10.813	0.147	10.480	11.206
Log Population Size	306	15.161	0.907	13.108	17.006
Log No. High Speed Lines	306	12.675	1.655	6.841	16.025
Log No. Rabies Cases	300	3.441	1.793	0	6.810
Log No. Legionellosis Cases	306	2.650	1.404	0	6.129
HIV Testing Proportion	306	0.062	0.030	0.019	0.179

There are twenty four missing observations from number of personal ads and number of high speed lines, which makes the total count of observations to be 306, instead of the full sample size of 330.

To examine the entry timing of Craigslist into a location, we collected data on the years in which new sites are launched through Craigslist's website.⁷ Appendix Table 1 documents the entry timing of Craigslist for the states in our study sample. We note Craigslist enters large urban states first, like New York, Texas and Florida, followed by smaller rural states. In particular, states above the median population density have a mean entry timing of 39.3 months, while states below the median population density have a mean entry timing of 47.7 months. With this data, we constructed a binary entry indicator for a state for a given year. On top of the binary entry variable, we also constructed a measure of the activity level of Craigslist's personals section for each location. Using a scraping program, we collected individual personal ads retrospectively through Archive.org, a site which stores digitized snapshots of websites over time. The contents of Craigslist were archived from December 1998 to recent years, allowing us to observe the activity level on the site in our entire study period. Through the data collection

⁷ Collected from <http://www.craigslist.org/about/expansion>, accessed on 18 June 2011.

exercise, we have records of 1,160,663 distinct personal ads over 24,147 state-days. In this sample of personal ads, we observe all personal ads posted in each ad category for the days in which snapshots are taken. We calculate a conservative proxy for the daily site activity level by averaging the number of personal ads over all observed days within a location.

To account for potential confounding effects that influence HIV trends at each state, we included demographic characteristics, socioeconomic factors, and internet availability as control variables. These control variables are collected from several sources. Age, ethnicity, and population size data are retrieved from the Surveillance Epidemiology and End Results (SEER) database.⁸ The SEER database collects demographic data and uses them for supporting cancer-related research for the population within the United States. The SEER database has extensive breakdown of the age group proportion and ethnicity proportion for all U.S. states over the study period, making it an ideal source for constructing demographic variables for the current study. For analysis purposes, we created proportion figures for the age group 20-39.⁹ The inclusion of age group variables in our analysis helps to control for the proportion of sexually active individuals which can influence the frequency of casual sex among the populations (Reece et al. 2010). We also computed the proportion of African Americans to serve as controls. Ethnic groups are added to control for patterns in sexual networks within and between racial groups that may potentially affect HIV transmission (Laumann and Youm 1999). Population size is included as a covariate to account for the level of urbanization of each location, which may be related to casual sex behavior.

⁸ Provided by the U.S. National Institutes of Health, available at <http://seer.cancer.gov/popdata/download.html>.

⁹ We find that main results are qualitatively similar when additional age brackets are included. See Appendix Table 2 for details.

Data on education attainment and household income is obtained from the U.S. Census Bureau. The Census Bureau constructs these two data sets via yearly surveys on the level of education attained and household income at each state.¹⁰ Educational attainment refers to the highest level of education that an individual has completed and we used high school and college attainment figures in our analyses to control for effects of schooling and education. The yearly household income data provided by the Census is available in median form for each state. Both the education attainment and the yearly household income are used to control for socioeconomic differences that could result in differential access to healthcare and information needed to avoid STD/HIV infections (Wasserheit and Aral 1996). In addition, we included the number of high speed lines connecting residential homes to the internet as a control for accessibility to alternative hook-up sites. This data is reported on FCC Form 477 from 1999 to 2008.¹¹

We augment the main analysis data set with survey data on HIV testing behavior. The CDC tracks the health conditions and risk behaviors in the U.S. annually using the Behavioral Risk Factor Surveillance System (BRFSS).¹² The BRFSS represents the largest on-going telephone health survey system that collects timely and accurate data on health-related behavior at the state level. Using survey responses from the BRFSS, we calculate the proportion of respondents who has had HIV testing in the past year during the study period. We incorporate this variable in our analysis to check for changes in HIV testing behavior across the study period and to assess whether the changes in testing behavior influence the trends observed in HIV incidence.

4. Empirical Methodology

¹⁰ Retrieved from <http://www.census.gov/hhes/socdemo/education/data/cps/index.html> and <http://www.census.gov/hhes/www/income/data/statemedian/index.html> respectively.

¹¹ Retrieved from <http://transition.fcc.gov/wcb/iatd/comp.html>.

¹² Available at http://www.cdc.gov/brfss/technical_infodata/surveydata.htm.

4.1 Main Analysis

The expansion of Craigslist into different locations over various time periods creates a natural experiment setting that allows the comparison of the difference in HIV incidence after and before site entry for states with Craigslist to the same difference for states that has yet to have Craigslist sites. We exploit the exogenous variation in Craigslist's entry across states and years in the natural experiment as the basis for identifying entry effects on HIV trends. This identification strategy has been implemented in several extant studies including Dranove et al. (2003) and Jin and Leslie (2003). To test for site entry effects, we estimate regressions in the form

$$\ln(y_{st}) = \mathbf{A}_s + \mathbf{B}_t + g \cdot \mathbf{Z}_{st} + p \cdot \mathbf{C}_{st} + e_{st}, \quad (1)$$

where s indexes states and t indexes time, $t = 1999, \dots, 2008$; y_{st} is the number of HIV cases for state s at time t ; \mathbf{A}_s is a vector of 33 state fixed effects; \mathbf{B}_t is a vector of time fixed effects; \mathbf{Z}_{st} is a vector of state demographics features, socioeconomic indicators and internet availability, which includes age proportion, ethnicity proportion, population size, education attainment proportion, median income levels, and number of high speed Internet lines; \mathbf{C}_{st} is the binary indicator for Craigslist entry, that is, $\mathbf{C}_{st} = 1$ if the state has Craigslist in a particular year, zero otherwise; and e_{st} is an error term. The coefficient p is the difference-in-difference estimate of the effect of Craigslist's entry on the incidence of HIV. If $p > 0$, then site entry has caused an increase in HIV prevalence.

In the above specification, the state level fixed effects controls for time-invariant differences across states and the year fixed effects control for common macroeconomic shocks across time. The inclusion of these fixed effects makes each state in a given year comparable to any other state at other time periods. In addition to the state and year fixed effects, certain demographic

and socioeconomic trends may still drive the incidence of HIV. To account for such effects, several control variables \mathbf{Z}_{st} are included in the model specification to account for factors that vary within each state over time. Details of these control variables have been delineated earlier in the data section. For all model specifications, the error terms are clustered at the state level to account for autocorrelation in the data (Bertrand et al. 2004). Following the practice in regressions involving STD trends, we weight our regressions by the relevant state populations (Carpenter 2005).¹³

The fixed effects framework together with covariates may not be able to account for potential time-varying effects that influence HIV trends. To assess the robustness in our results, we further run regression models with time-varying controls as per extant studies (Athey and Stern 2002). We execute this check by including interaction terms of the state covariates with the linear time trend as follows

$$\ln(y_{st}) = \mathbf{A}_s + \mathbf{B}_t + g \cdot \mathbf{Z}_{st} + p \cdot \mathbf{C}_{st} + r \cdot \mathbf{Z}_{st} \cdot \mathbf{T}_t + e_{st}. \quad (2)$$

We further assess the robustness of the results with respect to confounding effects from unobservables by using a matched sample of observations derived from propensity score matching. To account for differences in entry times over the year, we run separate regressions using an alternative measure of site entry which labels locations experiencing entry in the late part of the year as having entry in the subsequent year. In addition, analyses using negative binomial regressions are performed to account for the count nature of HIV incidents. Negative binomial regressions implemented with dummy variables to represent fixed effects are used so that incidental parameter bias can be avoided (Allison and Waterman 2002).

¹³ We note that un-weighted regressions produce estimates that are similar in sign and statistical significance in our main specifications.

Since Craigslist enters into metro areas within states, it might be more appropriate to examine the entry effect at finer geographical levels. Using data from counties in Florida, Michigan, Oregon, Washington, and New York City, we construct a separate panel dataset consisting of HIV infection, site entry and covariates matched at the corresponding levels. We find that site entry occurs only once for each county, which meant that analysis at the county level avoids the issue of capturing the effect of multiple sites located in one geographical area. In addition, we exclude Craigslist sites that extend across multiple counties in our analysis so as to match the county-level counts of HIV infection to specific site entries.¹⁴ This simplifies the analysis in that we do not have to deal with the mechanics of attributing the rise in HIV at different counties that are linked to the same Craigslist site. Based on this subsample of locations, we rerun our main analyses.

Given that site entry can take place in any month of a year, a binary year indicator for site entry may not capture the presence of Craigslist adequately, especially for locations that experience site entry in the later months of a year. Furthermore, heterogeneity in population's online sex seeking behavior and the ease of access to the Internet can result in disparate usage levels of Craigslist personal ads across locations. Sex seeking behavior via online classified ads is likely to be more numerous in urban and built-up areas compared to rural locations. Recognizing these issues, we conduct a robustness check using a measure of site activity, the average daily number of personal posts, as a regressor in place of the binary entry variable in our models. By capturing

¹⁴ For instance, the Craigslist site for the city of Portland (Oregon) services the population in the counties of Multnomah, Washington and Clackamas. Such instances are dropped from our analysis. We check the counties that are linked to a city using the Census' website (<http://quickfacts.census.gov/qfd/index.html>).

the activity level on Craigslist's personals section directly, the number of personal ads has the added benefit of reducing the threat of contemporaneous effects of unobserved factors associated with Craigslist's entry.

4.2 Falsification Checks

It is plausible that the previous set of regressions is picking up spurious entry effects as a result of coincidence. To assess the possibility of the entry variable picking up significant effects by chance, we run falsification tests using regressions to predict the incidence of non-sexually transmitted diseases based on Craigslist's entry. Given that site entry bears no relationship towards the proliferation of diseases that are not transmitted through physical human contact, we expect the entry coefficients not to pick up any effect in the falsification regressions involving these diseases. In executing these falsification tests, we choose rabies and *Legionellosis* to be the dependent variables as these diseases are transmitted mainly via infected animals and contaminated water sources, respectively.¹⁵

The relationship between the entry variables and HIV trends may also be driven by unobserved confounding factors. A potential confounding effect can arise from an increase in risky sexual behavior in response to medical breakthroughs in HIV treatment. In particular, the introduction of the highly active antiretroviral therapy (HAART) in the mid-1990s greatly lowered the mortality rate of HIV positive patients, boosted individuals' sexual activity, and thus facilitated the spread in HIV (Lakdawalla et al. 2006). However, a temporal gap exists between HAART availability and Craigslist's entry in locations. We assess whether the increase in HIV trends due to pre-entry events overlaps with the period of Craigslist entry at various locations. To perform

¹⁵ See Belotto et al. (2005) and Muder et al. (1986) for detailed reviews of the transmission modes of the two diseases.

this check, we conduct a falsification test through the use of placebo dummy variables in our regressions. We include two years of pre-entry dummies as placebos along with four years of post-entry dummies to capture potential inter-temporal entry effects as follows

$$\ln(y_{st}) = \mathbf{A}_s + \mathbf{B}_t + g \cdot \mathbf{Z}_{st} + \sum_j p_j \cdot \mathbf{C}_{st}^j + e_{st}, \quad (3)$$

where $j \in \{-2, -1, 1, 2, 3, 4\}$, indicating whether year t is the j th year since Craigslist's entry in state s . In these regressions, the omitted category is the year of Craigslist's entry (\mathbf{C}_{st}^0). In the presence of an overlapping trend of increasing HIV incidence prior to site entry, the placebo indicators would produce positive and significant coefficients. In addition, the coefficients of the post-entry indicators would reflect the immediacy of entry impacts on HIV trends.

4.3 Entry Exogeneity

There is a potential concern that Craigslist's entry decision into a state is endogenous and may be influenced by unobserved casual sex behavior prevalent in a location. For instance, an increase in the proportion of individuals actively seeking out casual sex partners would simultaneously lead to an increase in HIV cases and more requests to launch Craigslist in their locations. To interpret the main results causally, we require the entry of Craigslist to be exogenous with respect to HIV trends. We believe that the variation in entry timing and location is exogenous for the following two reasons.

First, Craigslist is a classified ads website made up of multiple sections. Other than personals, it has sections devoted to jobs, housing, for sale items, services, community, gigs, resumes, and discussion forums. Given that site usage does not reside primarily in the look-up of personals, entry decisions are unlikely to be driven solely by user demand for casual sex partners. Second, to sustain the costs of its operations, Craigslist defrays its costs by charging posting fees for ads

under the jobs, brokered apartments and therapeutic services categories.¹⁶ Thus, the potential revenues received from these ad postings are more likely to influence entry decisions than the demand for hook ups at a location. Under this motivation, entry decisions are likely to be based on the local demographics and internet availability which are indicators for potential site activity level.

Though there are no practical reasons for entry to be driven by casual sex trends, there may be unobserved factors that induce site entry. To systematically assess whether unobservable variables related to HIV trends are driving entry decisions, we use hierarchical duration models to predict the entry of Craigslist using demographic variables, socioeconomic factors, internet availability, and HIV incidence. In the duration models, the dependent variable is made up of a series of 0's and ends with a 1 in the year at which entry takes place. By assessing the impact of HIV incidence on site entry, we are effectively testing whether trends in casual sex prevalence and other related risky behaviors correlated with HIV incidence are inducing the entry patterns. The coefficients on the HIV incidence of this test provide insights on whether omitted variables bias and reverse causality create issues in our main analyses.

4.4 HIV-Inducing Mechanisms

Though the set of analyses described earlier can be used to assess whether the HIV trends are influenced by the entry of Craigslist, they are unable to inform us on how site usage leads to more HIV cases. Two key questions related to the underlying mechanisms of Craigslist usage are crucial to our understanding. First, the rise in HIV incidence with respect to Craigslist entry may be induced by greater levels of HIV testing during the study period. After having sexual relations

¹⁶ See Craigslist factsheet, available at <http://www.craigslist.org/about/factsheet>.

with casual partners solicited online, individuals may be more conscious about their HIV status and are going for HIV screenings. Alternatively, sex seeking individuals may be requesting online partners to provide evidence of their HIV status before consenting to having sex, leading to more testing and the discovery of previously unknown HIV positive status. Under these two possibilities, the increase in HIV testing behavior consequent to Craigslist's entry, is ultimately responsible for the observed increase in HIV cases. This alternative explanation, if found true, would mean that there may not be an actual increase in HIV transmission arising from sexual acts facilitated by Craigslist.

To understand the nature of the relationship between HIV incidence and Craigslist, we begin by assessing the trend of HIV testing from 1999 to 2008, using survey responses from the BRFSS. We regress HIV testing proportion on a linear time trend representing the study years (i.e., 1998 = 1, 1999 = 2, ..., 2008 = 10) and state covariates. Coefficients on the time trend variable provide insights on the potential statistical changes in HIV testing behaviors across the years. Next, we run a regression to understand the relationship between site entry and HIV testing behavior. In particular, we seek to understand whether the site usage has brought about safer attitudes towards HIV testing. Finally, we include the HIV testing proportion as a covariate in our main regressions models (i.e., Equation 1) to see if our results change qualitatively after controlling for potential shifts in HIV testing behavior.

Second, from a policy point of view, it is imperative to understand the underlying mechanism that drives HIV trends by identifying the type of casual sex solicited on Craigslist. Broadly speaking, there are two main types of soliciting activities that take place via Craigslist personal

ads: (i) non-market related casual sex and (ii) market related casual sex. The former type of soliciting involves the search of strangers on the site who are willing to have casual hookups without monetary payments. This type of soliciting includes mutually willing individuals who are looking for ‘no-string attached’ sex and/or ‘novelty sex’. The latter type of soliciting is made up by individuals who participate in sexual activities that involve monetary transactions. Individuals posting ads with such intent include prostitutes, escorts and their clients.

To understand which type of soliciting on Craigslist results in the increase in HIV trends, we examine the effects of ads posted at different sections on Craigslist. In addition to the “personals” section, Craigslist had another ad section, called “erotic services”. This section was largely filled with ads that solicit market-related casual sex. As Craigslist exerts little regulation over its ad postings, many ads in the erotic services section were explicit in their display of pornographic photos, the listing of prices, and the descriptions of sexual services offered. The erotic section was ultimately shut down in 2010 when the state attorneys general placed increasing pressure on Craigslist to limit prostitution activity on the site. We exploit the variation in number of posts from each section across states and years to understand their respective impact on HIV incidence. To execute the statistical test, we included both the average count of erotic service ads and the average number of personal ads in place of the entry variable as regressors in Equation 1 and assess the statistical significance of each of these variables.

5. Results

Table 2 presents the main results for our empirical analysis. Under an unweighted regression in Model 1, we see that the binary entry variable yields a positive and significant coefficient. This estimate represents a 17.7 percent increase in HIV cases attributed to site entry. Under a

weighted regression in Model 2, we observe that the coefficients for site entry are similar to Model 1 in sign and statistical significance. To assess the robustness of the main results with respect to time-varying state factors, we estimate the baseline model with the inclusion of interaction terms of covariates with a linear time trend. The result of this test is depicted in Model 3, wherein the inclusion of time-varying state covariates did not change the main results qualitatively. To further account for potential unobservables that may affect the estimates by making entry decision endogenous, we re-estimate the baseline model using states that are matched by demographic factors, socioeconomic characteristics and internet accessibility under a propensity score matching scheme.¹⁷ Entry estimate in Model 4 remains positive and statistically significant. In particular, this estimate indicates that a 16.8 percent increase in HIV cases follows Craigslist's entry. To understand this more concretely, we translate the coefficient into actual number of HIV cases. Using the mean number of HIV cases in the matched sample as a baseline, we calculate that 99 HIV cases are introduced on average in a state annually as a result of Craigslist's entry.

5.1 Robustness Checks

In Model 5, we assess the robustness of the main results with respect to an alternative definition of site entry. As site entry may occur in the later months of the year, a binary indicator for the year of initial entry may not accurately denote Craigslist presence. In our alternative entry measure, locations that experience site entry in the last quarter of the year (i.e., October to December) are labeled as having site entry in the following year. For instance, the site entry for Florida, which was originally in October 2002, is labeled as 2003 under the new variable definition. Regression results in Model 5 remain qualitatively similar under this alternative

¹⁷ We used age proportions, ethnic proportion, population size, education attainment proportion, median income level, number of high speed Internet lines as attributes to be matched upon. Samples are matched based on the nearest neighbor algorithm within a caliper size of 0.05, with replacement.

specification. We further coded a stricter entry variable which lists site entry in the following year for locations with Craigslist entry in the last six months of the year (i.e., July to December). The resultant estimates are largely similar. The fixed effects negative binomial regressions in Models 6 to 10 produce positive and significant coefficients for the entry variable, thus providing further confidence in the results from the OLS models. These models reveal that the sign and significance of the entry variables are robust under non-linear specifications. The estimated effects in the count data models are largely similar to that in the OLS models. The most conservative coefficient under the negative binomial specification (Model 8) represents a 15.9 percent increase in HIV upon site entry, which equates to 94 cases of HIV infection annually in a state.¹⁸

Next, to understand whether similar results are derived at finer levels of analysis, we run our baseline analyses on a subsample of county-level observations. Table 3 reports the results of this analysis. In Models 1 to 4, our baseline models indicate positive and significant coefficients for site entry, providing further assurance that the main results are robust. In addition, we estimated the baseline models using a direct measure of the activity level on Craigslist by using the number of personal ads posted on Craigslist. In Table 4, we observed that the coefficients for the log average daily number of personal ads posted are positive and significant, suggesting that a positive relationship between Craigslist's presence and HIV incidence is present after accounting for usage heterogeneity. In particular, a ten percent increase in the daily number of personal ads leads to a 0.7 percent increase in HIV cases annually in a state (Model 1). The weighted

¹⁸ We also ran regressions with the additional covariates for age 15-19 proportion and age 40-59 proportion to allow for a wider coverage of sexually active individuals (see Appendix Table 2). Coefficients for site entry remain largely unchanged.

counterpart to this estimate in Model 2 is slightly larger, indicating a 0.9 percent increase in HIV cases for a ten percent increase in ad posts. In Model 3, we restrict our analysis to a sample of locations which have at least one posted personal ad. By doing so, we can check whether the previous results are robust under a more homogenous sample that exclude state-year observations with zero site usage. Results in Model 3 show that estimates remain positive and significant. We run the same set of models under negative binomial regressions and find the results to be robust towards count data specifications.

Table 2: Impact of Craigslist Presence on HIV Trends, with Robustness Checks

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	0.163** (0.08)	0.252** (0.11)	0.212** (0.08)	0.156* (0.09)	0.271** (0.11)	0.173*** (0.06)	0.254*** (0.07)	0.214*** (0.06)	0.148** (0.07)	0.268*** (0.07)
Age 20-39 Proportion	4.586 (4.67)	-1.153 (6.62)	-9.799 (10.04)	7.088 (10.28)	-1.436 (6.67)	2.012 (4.00)	-4.050 (4.90)	-12.360* (6.80)	5.212 (6.72)	-4.339 (4.87)
African American Proportion	-5.394 (12.27)	-7.550 (13.70)	-13.120 (18.26)	-38.701 (29.39)	-6.346 (13.50)	-5.779 (5.99)	-7.064 (7.44)	-12.570 (8.91)	-37.168** (17.40)	-5.897 (7.35)
Log (Population Size)	1.695** (0.74)	0.991 (0.96)	1.152 (1.13)	0.696 (1.55)	0.939 (0.95)	1.737*** (0.48)	1.070* (0.61)	1.489** (0.64)	0.508 (1.01)	1.008* (0.61)
High School Attainment Proportion	1.834 (1.48)	1.875 (1.93)	-1.428 (1.69)	0.899 (1.51)	2.011 (1.88)	1.291 (1.08)	1.871 (1.31)	-1.966 (2.09)	0.758 (1.33)	2.035 (1.30)
College Attainment Proportion	-1.041 (1.09)	-1.615 (1.47)	3.946* (2.17)	0.125 (1.44)	-1.547 (1.43)	-0.885 (0.91)	-1.862* (1.05)	3.821** (1.51)	0.011 (1.06)	-1.785* (1.05)
Log (Median Household Income)	-0.924* (0.49)	-1.586** (0.70)	-2.411** (1.09)	-1.864** (0.85)	-1.541** (0.69)	-0.767** (0.36)	-1.507*** (0.45)	-2.355*** (0.71)	-1.801*** (0.49)	-1.466*** (0.45)
Log (No. Internet Lines)	0.098** (0.04)	0.143 (0.09)	0.048 (0.10)	0.087 (0.09)	0.142 (0.09)	0.125*** (0.05)	0.155** (0.07)	0.084 (0.10)	0.094 (0.07)	0.155** (0.07)
R-squared	0.261	0.275	0.364	0.205	0.285	-	-	-	-	-
Weighted by population		✓	✓	✓	✓		✓	✓	✓	✓
Controls × Time Trend			✓					✓		
P-Score Matched Samples				✓					✓	
Alternative Entry Labels					✓					✓
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 un-weighted 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 3: County level Analysis of Entry Impacts on HIV

	OLS		Negative Binomial	
	Model 1	Model 2	Model 3	Model 4
Binary Entry	0.115*	0.083**	0.069*	0.126***
	(0.07)	(0.04)	(0.04)	(0.04)
Age 20-39 Proportion	0.652	-0.893	4.765	-4.845
	(4.41)	(4.75)	(3.66)	(4.70)
African American Proportion	-0.016	-0.020***	-0.019*	-0.021**
	(0.01)	(0.01)	(0.01)	(0.01)
Log (Population Size)	-0.008	-0.021	-0.019	-0.003
	(0.03)	(0.02)	(0.01)	(0.01)
High School Attainment Proportion	-6.187	-4.597	-4.654	-1.760
	(3.97)	(4.05)	(3.02)	(3.29)
College Attainment Proportion	1.722**	1.713**	1.456***	1.399**
	(0.65)	(0.66)	(0.51)	(0.58)
Log (Median Household Income)	0.055	-0.289	-0.461	-0.112
	(0.49)	(0.29)	(0.33)	(0.29)
Log (No. Broadband Providers)	0.016	0.178	0.260**	0.292**
	(0.18)	(0.15)	(0.12)	(0.13)
R-squared	0.177	0.246	-	-
Weighted by population		✓		✓
Observations	239	239	239	239

The dependent variables for Models 1 and 2 are the log number of HIV cases, while the dependent variables for Models 3 and 4 are the number of HIV cases. Robust standard errors are reported in parentheses below coefficient values. All models have binary entry regressors and include state and year fixed effects. * p < 0.10, ** p < 0.05, *** p < 0.01.

Finally, we performed two further tests to assess the robustness of the results. First, we adopt a more conservative approach of implementing the difference-in-difference estimator. We do so by averaging the data points before and after the Craigslist entry for each state, and regressing this dependent variable on a dummy indicator for post-entry. Under this specification, we run a basic model with sixty-six observations that comprise of two observations for each state. In Appendix Table 3, we report the results of this specification in Models 1 and 2. Entry coefficients remain positive and significant under this modeling approach. Second, we further performed a test to account for the potential nonlinear increases in HIV trends in large urban states. In this test, we added the interaction terms of state dummies with linear time trend in the regressions. After

accounting for nonlinearities across states, it is assuring to see that entry coefficients remain positive and significant (Models 3 and 4).

Table 4: Impact of Craigslist Personal Ads on HIV

Variables	OLS			Negative Binomial		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Log (Avg. Personal Ads)	0.075* (0.04)	0.096** (0.04)	0.090** (0.04)	0.063** (0.03)	0.093*** (0.03)	0.095*** (0.03)
Age 20-39 Proportion	8.650* (4.75)	3.448 (7.06)	5.044 (15.90)	5.691 (4.12)	-0.143 (4.80)	-0.968 (12.58)
African American Proportion	-7.166 (11.78)	-11.418 (13.03)	4.392 (10.48)	-7.095 (5.91)	-10.464 (7.25)	4.685 (8.29)
Log (Population Size)	1.280 (0.81)	0.317 (1.06)	0.358 (1.71)	1.401*** (0.50)	0.440 (0.63)	0.617 (0.99)
High School Attainment Proportion	1.581 (1.41)	1.587 (1.71)	0.190 (1.60)	1.144 (1.10)	1.642 (1.33)	-0.017 (1.72)
College Attainment Proportion	-0.789 (0.92)	-1.581 (1.26)	-1.192 (1.57)	-0.813 (0.91)	-2.022* (1.06)	-1.588 (1.41)
Log (Median Household Income)	-0.981* (0.49)	-1.736** (0.70)	-0.716 (0.61)	-0.825** (0.36)	-1.676*** (0.44)	-0.906** (0.43)
Log (No. Internet Lines)	0.113** (0.05)	0.177* (0.10)	0.256 (0.28)	0.137*** (0.05)	0.189** (0.08)	0.136 (0.21)
R-squared	0.259	0.270	0.533	-	-	-
Weighted by population		✓	✓		✓	✓
At least one ad sample			✓			✓
Observations	306	306	165	306	306	165

The dependent variables for Models 1-3 are the log number of HIV cases, while the dependent variables for Models 4-6 are the number of HIV cases. Robust standard errors are reported in parentheses below coefficient values. Models 3 and 6 include only observations that have at least one personal ad posted. All models include state and year fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.2 Falsification Tests

Our first falsification exercise involves the testing of whether the Craigslist entry variables are picking up spurious effects. Significant entry coefficients in the regressions involving diseases that are not transmitted through physical human contact would signal issues with the regression results involving HIV incidence. Results of the falsification tests are depicted in Table 5. It is observed that the entry variables (i.e., binary entry variable and average daily number of personal ads) do not hold statistically significant relationships with the incidence of rabies and

legionellosis. This provides supporting evidence in favor of a non-coincidental effect captured by the entry variables in the regressions in Table 2.¹⁹

In our second falsification test, we examine whether the increase in HIV trends propagated from earlier time periods. Results of this falsification test are reported in Table 6. Across all models, we observe that the two year pre-entry placebo variables did not pick up any pre-entry effect. This suggests that the positive relationship between HIV trends and Craigslist entry observed in previous analyses is unlikely to be an artifact effect that propagated from periods prior to Craigslist's entry. In addition, the coefficient for first year of entry (i.e. Craigslist Entry₁) is not significant, implying that the effect of Craigslist on HIV did not arise immediately after site entry. The effect on HIV only shows up in the second year following Craigslist's entry, indicating that there is at least a one year delay between site entry and observing its effect on HIV trends. One would expect the site traffic on Craigslist to ramp up only a few months after the site launch as it takes time for users to be aware of Craigslist's presence in a new location. On top of that, it can take up to six weeks from the time a user is first infected with HIV to the time acute retroviral syndrome develops within the individual (Kahn and Walker 1998). More importantly, an infected individual may take as long as six months from initial exposure to build up enough antibodies to be detectable as HIV positive in tests.²⁰ Given these delays, an entry effect in the first year is suspect and may be a sign of spurious effects in the model. The absence of first year entry effects in Table 6 serves to further validate the authenticity of the effect picked by the entry variables.

¹⁹ We also conducted a separate falsification test using syphilis as the dependent variable. Using the full sample of U.S. states, we find that variables denoting Craigslist presence hold positive and significant coefficients. Results are shown in Appendix Table 4.

²⁰ The period in which HIV infected people are not tested positive is termed as the seroconversion period. More information on seroconversion can be found at <http://www.webmd.com/hw-popup/seroconversion-period>.

Table 5: Falsification Test Using Non-Sexually Transmitted Diseases

Variables	OLS				Negative Binomial			
	Binary Entry		Personal Ads		Binary Entry		Personal Ads	
	Rabies Model 1	Legionellosis Model 2	Rabies Model 3	Legionellosis Model 4	Rabies Model 5	Legionellosis Model 6	Rabies Model 7	Legionellosis Model 8
Craigslist Presence	-0.485 (0.37)	0.066 (0.14)	0.448 (0.51)	0.062 (0.04)	-0.231 (0.28)	0.088 (0.11)	0.350 (0.25)	0.053 (0.05)
Age 20-39 Proportion	34.592 (50.89)	-25.217** (9.49)	40.777 (55.38)	-21.655** (9.79)	13.241 (25.10)	-23.383** (10.63)	26.365 (30.84)	-21.366* (10.94)
African American Proportion	-31.204 (79.63)	-5.411 (20.12)	-54.485 (91.20)	-6.060 (18.19)	-44.646 (50.88)	-12.480 (16.36)	-86.025 (72.15)	-12.726 (16.02)
Log (Population Size)	5.792 (5.60)	2.482 (1.58)	1.054 (4.25)	2.122 (1.48)	3.970 (3.91)	2.186* (1.15)	1.015 (3.31)	1.978* (1.13)
High School Attainment Proportion	-30.715 (22.04)	-0.081 (3.19)	-32.777 (22.06)	-0.531 (3.15)	-15.574* (9.46)	-1.901 (3.58)	-18.733* (10.47)	-2.604 (3.66)
College Attainment Proportion	-7.703 (7.60)	3.991 (3.18)	-2.423 (8.99)	4.112 (3.33)	-5.893 (5.07)	3.016 (2.35)	-2.906 (5.19)	3.235 (2.39)
Log (Median Household Income)	-6.324** (2.72)	1.883* (0.97)	-5.532* (2.92)	1.734* (0.92)	-3.998** (1.77)	2.185*** (0.73)	-3.374** (1.72)	1.964*** (0.73)
Log (No. Internet Lines)	-0.032 (0.25)	-0.284* (0.17)	0.081 (0.26)	-0.266* (0.15)	-0.070 (0.25)	-0.508*** (0.20)	-0.016 (0.25)	-0.477** (0.19)
R-squared	0.148	0.519	0.166	0.522	-	-	-	-
Observations	300	306	300	306	300	306	300	306

The dependent variables for Models 1 and 3 are the log number of rabies cases and that for Model 2 and 4 are the log number of legionellosis cases respectively. The dependent variables for Models 5 and 7 are the number of rabies cases and that for Model 6 and 8 are the number of legionellosis cases. Models under the heading, "Personal Ads", have the logged average number of personal ads as its regressor. Robust standard errors are reported in parentheses below coefficient values. All regressions have state and year fixed effects. Regressions are weighted by the state population. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 6: Falsification Test Using Pre and Post Entry Indicators

Variables	OLS		Negative Binomial	
	Model 1	Model 2	Model 3	Model 4
Craigslist Entry ₋₂	-0.084 (0.06)	-0.069 (0.06)	-0.051 (0.05)	-0.058 (0.06)
Craigslist Entry ₋₁	-0.026 (0.05)	-0.002 (0.05)	-0.023 (0.04)	-0.002 (0.04)
Craigslist Entry ₁	-0.019 (0.05)	0.025 (0.05)	0.002 (0.05)	0.036 (0.05)
Craigslist Entry ₂	0.128** (0.05)	0.168*** (0.06)	0.116*** (0.04)	0.169*** (0.05)
Craigslist Entry ₃	0.114** (0.05)	0.136** (0.06)	0.096** (0.04)	0.136*** (0.04)
Craigslist Entry ₄	0.100*** (0.04)	0.111** (0.04)	0.093** (0.04)	0.106*** (0.04)
Age 20-39 Proportion	3.883 (4.54)	-2.435 (6.42)	1.197 (4.09)	-5.598 (5.13)
African American Proportion	-3.835 (12.06)	-4.477 (13.64)	-4.325 (5.99)	-4.033 (7.51)
Log (Population Size)	1.603** (0.77)	1.071 (1.02)	1.720*** (0.50)	1.204** (0.60)
High School Attainment Proportion	1.732 (1.47)	1.672 (2.00)	1.236 (1.12)	1.591 (1.30)
College Attainment Proportion	-0.764 (0.98)	-1.316 (1.33)	-0.734 (0.91)	-1.576 (1.06)
Log (Median Household Income)	-0.917* (0.50)	-1.611** (0.71)	-0.737* (0.38)	-1.484*** (0.47)
Log (No. Internet Lines)	0.105** (0.04)	0.128 (0.09)	0.126*** (0.05)	0.133* (0.07)
R-squared	0.281	0.296	-	-
Weighted by population		✓		✓
Observations	306	306	306	306

The dependent variables for Models 1 and 2 are the log number of HIV cases and that for Model 3 and 4 are the number of HIV cases. Robust standard errors are reported in parentheses below coefficient values. All regressions include state and year fixed effects. * p < 0.10, ** p < 0.05, *** p < 0.01.

5.3 Exogeneity of Entry Decisions

To systematically assess whether unobservable variables related to HIV trends are driving entry decisions, we ran hierarchical duration models to predict the entry of Craigslist using demographic variables, socioeconomic factors and HIV trends. The results of this test provide insights on whether omitted variables bias and reverse causality create issues in our main analyses. In Table 7, we observe that several demographic factors and internet availability are

significant predictors of entry. In particular, the number of high speed Internet lines holds a positive and significant relationship with site entry. With more high speed lines, access and usage of Craigslist sites would be enhanced. In addition, the duration model shows that proportion of internet savvy population (aged 20-39) is a positive determinant of site entry. These observations provide evidence in favor of our reasoning, wherein entry decisions are contingent on the potential site activity level that drives future earnings from paid ads.

Table 7: Hierarchical Duration Models Predicting Entry into States

Variables	Logistic Regression		
	Model 1	Model 2	Model 3
Age 20-39 Proportion	70.167*** (27.06)	84.973*** (31.11)	87.455*** (33.74)
African American Proportion	-11.153** (4.94)	-12.758** (6.43)	-16.929** (6.96)
Log (Population Size)	2.076*** (0.61)	-1.239 (1.91)	-1.097 (1.84)
High School Attainment Proportion		7.577 (17.06)	26.091 (21.01)
College Attainment Proportion		-6.955 (12.94)	-12.835 (13.91)
Log (Median Household Income)		-3.964 (4.84)	-4.590 (4.95)
Log (No. Internet Lines)		3.374* (1.84)	1.611 (2.02)
Log (No. HIV Cases)			1.547 (0.98)
Log likelihood	-32.957	-30.862	-29.534
Observations	109	109	109

The dependent variable is made up of a series of 0's and has a 1 in the year at which entry takes place at a state, after which the unit is dropped entirely from the analysis. Robust standard errors are reported in parentheses below coefficient values. Likelihood ratio test shows that the unrestrictive model (Model 3) does not have better model fit than the restrictive model (Model 2). * p < 0.10, ** p < 0.05, ***p < 0.01.

In the final step of the hierarchical regression, the addition of the HIV incidence variable is expected to pick up effects from unobserved factors that influence HIV trends and entry simultaneously. The estimated coefficients on HIV incidence are not significantly different from zero and the likelihood ratio tests show that the less restrictive model (Model 3) does not perform

significantly better than the restrictive model (Model 2). These results imply that entry decisions are not related to HIV dynamics, and more importantly, it shows that there is no evidence of unobserved effects driving HIV trends and entry concurrently after demographic, socioeconomic factors and internet availability are accounted for. These empirical checks support the intuition in favor of exogenous entry decisions and rules out alternative explanations that suggest that unobserved factors are driving the relationship between entry and HIV trends. In particular, increasing rates of casual sex behavior as a result of the introduction of HIV treatment are unlikely to create systematic increases on HIV trends in this study.

5.4 HIV Causing Mechanisms

From the analyses thus far, results suggest that the entry of Craigslist has the effect of increasing the prevalence of HIV. However, it is less clear what underlying mechanisms are driving these trends. To understand the mechanisms behind the rise in reported HIV cases after Craigslist's entry, we examine (i) whether the increase in HIV prevalence is due to increased HIV testing, and (ii) which type(s) of solicitation on Craigslist is influencing HIV incidence.

To understand how HIV testing behavior has changed over time, we plot a graph depicting the proportion of HIV testing over the study period. Figure 2 shows a general decreasing trend in the HIV testing behavior. We further run a set of regressions to affirm the statistical relationship between HIV testing trends and time. In Table 8, the negative and significant time variables in both Models 1 and 2 provide statistical evidence for the decrease in HIV prevalence over time. This result provides initial evidence to suggest that the increase in HIV prevalence during the periods of Craigslist presence is unlikely to arise from increased HIV testing.

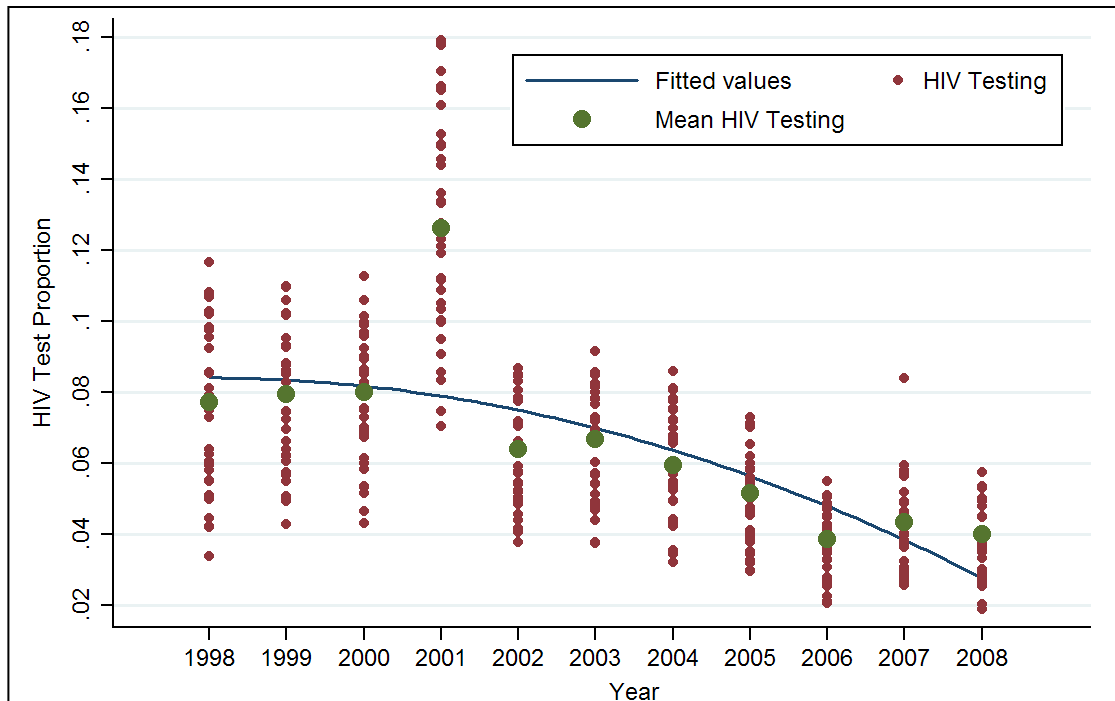


Figure 2: The Decreasing Trend of HIV Testing over Time

Next, to understand whether site entry results in a shift towards safer sexual behavior in terms of HIV testing, the proportion of HIV testing is regressed on Craigslist’s entry and the logged number of personal ads. Models 3 and 4 in Table 8 show the results of these regressions. Across these two models, the variables representing Craigslist’s entry do not hold statistical relationships with HIV testing behavior. This test provides further evidence that disputes the possibility of Craigslist’s usage leads to a shift in HIV testing attitudes.

Finally, we assess whether HIV testing trends impact HIV prevalence by including the proportion of HIV testing as a regressor in the main regressions. Table 9 depicts the results of these regressions. In particular, all models consistently show that the HIV testing variable does not bear statistically significant relationships with HIV incidence. After the inclusion of the proportion of HIV testing, the binary entry variable and number of personal ads remains positive and significant in both the OLS and negative binomial specifications. Taken as a whole, the

results in Tables 8 and 9 provide evidence against the possibility of rising HIV trends as a result of increased level of HIV testing during Craigslist's entry. The rise in HIV prevalence is more likely to arise due to the increase in disease transmission over the change in HIV testing behavior.

Table 8: Time and Entry Effects on Testing Behavior

Variables	Model 1	Model 2	Binary Entry	Personal Ads
			Model 3	Model 4
Time	-0.007*** (0.00)	-0.012*** (0.00)		
Craigslist Presence			0.002 (0.00)	0.001 (0.00)
Age 20-39 Proportion		0.425* (0.25)	0.248 (0.24)	-0.843 (0.58)
African American Proportion		0.626 (0.44)	1.214*** (0.41)	0.430 (0.60)
Log (Population Size)		-0.053* (0.03)	-0.072** (0.03)	-0.023 (0.07)
High School Attainment Proportion		-0.125 (0.12)	-0.031 (0.05)	0.009 (0.04)
College Attainment Proportion		-0.077 (0.08)	-0.061* (0.03)	-0.067 (0.06)
Log (Median Household Income)		-0.029 (0.03)	0.005 (0.01)	0.025 (0.01)
Log (No. Internet Lines)		0.016*** (0.00)	0.004 (0.00)	0.004 (0.01)
R-squared	0.578	0.618	0.933	0.902
Observations	309	309	309	166

The dependent variables for both models are the proportion of HIV testing in a state for a given year. Time represents a linear time trend for the years 1999-2008. Robust standard errors are reported in parentheses below coefficient values. Regressions are weighted by state population. * p < 0.10, ** p < 0.05, ***p < 0.01.

In the next set of tests, we attempt to understand whether the rise in HIV incidence is due to casual sex from non-market transactions (i.e., no-strings attached sex, novelty sex) or from market-related transactions (i.e., prostitutes, escorts). The analysis strategy involves examining the relative magnitudes and significance of the variables representing counts of personal ads and erotic service ads.

The results in Table 10 revealed positive and significant coefficients for the number of personals ads and negatively significant coefficients for erotic service ads. This result implies that the upward shift in HIV trends is largely influenced by the ads hosted in the personals section, that is, the non-market casual hookups. Interestingly, the negative coefficient for number of erotic service ads suggests that the escort and prostitution services advertised in Craigslist are shifting users to sexual behaviors that reduce HIV transmission. This finding is in line with existing research which shows that internet-facilitated sex workers represent a population that is dissimilar from street-based sex workers in earlier eras (Cunningham and Kendall 2010). Relative to the street workers, sex workers who solicit customers online are more likely to be well-educated, hold health insurance, work only part time, and are less likely to participate in risky sexual practices with clients. We discuss the managerial implications for the disparate roles of personal ads and erotic service ads on HIV prevalence next.

Table 9: Assessing the Effect of Testing Behaviors on HIV

Variables	OLS		Negative Binomial	
	Binary Entry	Personal Ads	Binary Entry	Personal Ads
	Model 1	Model 2	Model 3	Model 4
HIV Testing Proportion	-1.269 (1.56)	0.140 (2.51)	-0.935 (1.51)	0.126 (2.14)
Craigslist Presence	0.253** (0.11)	0.090** (0.04)	0.255*** (0.07)	0.095*** (0.03)
Age 20-39 Proportion	-0.916 (6.77)	5.163 (16.73)	-3.927 (4.95)	-0.828 (13.37)
African American Proportion	-6.237 (12.97)	4.332 (10.31)	-6.068 (7.33)	4.630 (8.40)
Log (Population Size)	0.889 (0.99)	0.361 (1.70)	0.991 (0.62)	0.619 (0.98)
High School Attainment Proportion	1.870 (1.93)	0.189 (1.61)	1.866 (1.31)	-0.019 (1.72)
College Attainment Proportion	-1.685 (1.48)	-1.183 (1.55)	-1.920* (1.06)	-1.578 (1.39)
Log (Median Household Income)	-1.571** (0.68)	-0.719 (0.62)	-1.499*** (0.44)	-0.910** (0.44)
Log (No. Internet Lines)	0.149 (0.09)	0.255 (0.27)	0.159** (0.07)	0.135 (0.20)
R-squared	0.277	0.533	-	-
Observations	306	165	306	165

The dependent variables for Models 1 and 2 are the log number of HIV cases, while the dependent variables for Models 3 and 4 are the number of HIV cases. Robust standard errors are reported in parentheses below coefficient values. Models 1 and 3 have binary entry regressors and Models 2 and 4 have the logged average number of personal ads posted as regressors. All models include state and year fixed effects. Regressions are weighted by the state population. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Impact of Personal Ad and Erotic Ads on HIV

Variables	OLS	Negative Binomial
	Model 1	Model 2
Log (No. Erotic Ads)	-0.074*** (0.03)	-0.073* (0.04)
Log (No. Personal Ads)	0.171*** (0.06)	0.169*** (0.05)
Age 20-39 Proportion	2.804 (6.98)	-0.619 (4.80)
African American Proportion	-12.270 (12.90)	-11.374 (7.28)
Log (Population Size)	0.412 (1.04)	0.491 (0.63)
High School Attainment Proportion	1.367 (1.68)	1.457 (1.33)
College Attainment Proportion	-1.382 (1.23)	-1.808* (1.02)
Log (Median Household Income)	-1.794** (0.71)	-1.707*** (0.44)
Log (No. Internet Lines)	0.156 (0.10)	0.163** (0.07)
R-squared	0.293	-
Observations	306	306

The dependent variable for Model 1 is the log number of HIV cases, while the dependent variable for Model 2 is the number of HIV cases. Robust standard errors are reported in parentheses below coefficient values. All models include state and year fixed effects. Regressions are weighted by the state population. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6. Discussion and Implications

In this study, we assess and quantify the impact of Craigslist's entry on increasing the reported number of HIV cases in the United States. To our best knowledge, this study represents the first empirical effort that quantifies the impact of site entry on increasing reported HIV cases in the U.S. Utilizing Craigslist's varying entry times into locations, we are able to identify the effect of its personal ads on reported HIV trends. The current research goes beyond extant survey and interview studies in establishing the casual relationship between online intermediaries on STD trends.

This study provides several important managerial insights to the IS community, policy makers and healthcare practitioners. First, the estimates on entry-induced HIV incidence can be used to derive the economic significance of Craigslist's entry on healthcare costs in the country. Using conservative estimates derived from the models estimated with matched samples derived from propensity score matching (Models 4 and 9 of Table 2), we estimate that the healthcare cost of treating patients who are infected with HIV from casual sex solicited at sites like Craigslist to be between \$62 million and \$65.3 million annually.²¹ The magnitude of this figure signifies that the rise in reported HIV cases as a result of site entry has a non-trivial economic impact on healthcare expenses in the U.S. This result serves as an useful input towards the formulation of policies on the governance over website content and towards regimes for tracking of HIV transmission in the country.

Second, by identifying the group of users that is at risk of HIV infection, our result provides healthcare practitioners with insights on how they can effectively target their efforts to reduce disease transmission facilitated through classified ad sites. Though we demonstrate that individuals looking for non-market casual hookups are on Craigslist are at risk of contracting STDs, similar trends may potentially be observed at other hookup websites (e.g., adultfriendfinder.com, backpage.com, manhunt.net), given the similarity in the roles these sites play in terms of facilitating casual sex encounters. For an effective management of the emerging

²¹ First, we extrapolate the number of entry-induced HIV cases from state-level to nation-level via population size. This calculation yields an estimated range of 6130 to 6455 HIV cases arising in the U.S. as a result of Craigslist's presence (using 2008 figures). Next, we compute the average annual treatment cost in 2008 dollars, using figures from Chesson et al. (2004). The annual treatment cost works out to be \$10,121.60 per HIV patient. The final step involves the product of the nationwide number of HIV cases with the annual treatment cost for all infected patient, which provides an estimate of \$62 to \$65.3 million spent on annual treatment costs.

online transmission route, awareness programs and intervention campaigns could be designed to target users who frequent hookup sites with the intentions to solicit casual hookups with online strangers.

Finally, the finding on erotic service ads negating the prevalence of HIV provides important inputs to policy makers and public health professionals on the relationship between internet-facilitated sex work and the dynamics of HIV and other STDs. The technology driven marketplace augments the sex work industry with providers who are better-educated and well-informed about risks of STD transmission (Cunningham and Kendall 2010). Serving as a substitute for non-market casual hookups, market-related sexual transactions solicited on classified ad sites may shift online sex seekers towards safer sexual practices, thereby countervailing the onset of HIV transmission arising from this new transmission route. This finding implies that proposals for interventions should carefully consider the more subtle effects of new technologies, which may reduce some of the new challenges arising from technology introduction and potentially mitigate existing infection risks associated with the sex work industry.

Our paper has a few limitations, some of which can pave the way for future research in this area. First, we do not have complete data on HIV incidence at finer geographical levels. Craigslist was introduced into metro areas within states and it will be ideal to examine its impact on HIV trends at the metropolitan area level. A consolidated and standardized data set on HIV prevalence at the metropolitan areas over the study period is not readily available. Future efforts in making such data available would help to improve the estimates derived from this study. Second, another

study limitation pertains to the lack of data on the usage levels of other hookup or dating websites. However, we are less worried about these variables causing the study results to change qualitatively, as our checks reveal that Craigslist entry is exogenous with respect to HIV prevalence and we have controlled for the effects from other sites using internet availability as a covariate. Furthermore, Craigslist is the first and the largest free website that facilitated casual hookups during the study period. Other free hookup or dating sites only entered the space after 2003 and received lower web traffic than Craigslist. A related limitation is that study findings for Craigslist may not generalize to the market of online dating sites, as the users of these sites can differ greatly from those in hookup sites. Users on dating sites such as Match.com, OkCupid and PlentyofFish are required to invest time to create profile pages and even spend money to increase the chances of interacting and meeting with other users. Given these investments, users on dating sites are more likely to be seeking serious relationships over casual hookups. Finally, though the results are indicative of a causal relationship between Craigslist and HIV prevalence, this interpretation is only possible through the ruling out of alternative explanations via a battery of empirical tests. While the effect on HIV prevalence shows up in the early years of Craigslist entry, one should note that these effects may flatten out over time. Furthermore, the scope of this paper is unable to inform policy on the effectiveness in curbing STD propagation via the shutdown of Craigslist's personals section, given that we do not have the appropriate counterfactuals to test this. Notwithstanding these limitations, this work represents the first empirical attempt in assessing and quantifying the relationship between Craigslist's entry and reported HIV prevalence via with a set of robust empirical tests performed under a natural experiment setting.

As usage of sites with similar features to Craigslist's personal ads increases in recent years, the understanding of the impact of these online intermediaries on STD transmission becomes increasingly crucial for the development of effective social policy, public health actions and online regulations. Though there is increasing documentation of casual partners who meet through the Internet over the last decade, little is known about how this shift in sexual behavior affects STD transmission. Our results fill this gap in the literature by quantifying the magnitude of website induced HIV incidence and identifying the mechanisms of HIV transmission on Craigslist. The implications for future research and policy are important, because the discovery of a new social transmission route is critical for managing disease progression in large populations, targeting safe sex behavior effectively, and designing policies for online platforms operating in this space.

References

- Allison, P.D., and Waterman, R.P. 2002. "Fixed-Effects Negative Binomial Regression Models," *Sociological Methodology* (32:1), pp 247-265.
- Athey, S., and Stern, S. 2002. "The Impact of Information Technology on Emergency Health Care Outcomes," *RAND Journal of Economics* (33:3), pp 399-432.
- Ayyagari, R., Grover, V., and Purvis, R. 2011. "Technostress: Technological Antecedents and Implications," *MIS Quarterly* (35:4), pp 831-858.
- Bailey, J.P., and Bakos, Y. 1997. "An Exploratory Study of the Emerging Role of Electronic Intermediaries," *International Journal of Electronic Commerce* (1:3), pp 7-20.
- Bakos, J.Y. 1997. "Reducing Buyer Search Costs: Implications for Electronic Marketplaces," *Management Science* (43:12), December 1, 1997, pp 1676-1692.
- Baumeister, R.F., and Heatherton, T.F. 1996. "Self-Regulation Failure: An Overview," *Psychological Inquiry* (7:1), 2013/01/02, pp 1-15.
- Belotto, A., Leanes, L.F., Schneider, M.C., Tamayo, H., and Correa, E. 2005. "Overview of Rabies in the Americas," *Virus Research* (111:1), pp 5-12.
- Benotsch, E., Kalichman, S., and Cage, M. 2002. "Men Who Have Met Sex Partners Via the Internet: Prevalence, Predictors, and Implications for Hiv Prevention," *Archives of Sexual Behavior* (31:2), pp 177-183.
- Bertrand, M., Duflo, E., and Mullainathan, S. 2004. "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics* (119:1), February 1, 2004, pp 249-275.
- Brynjolfsson, E., Hu, Y., and Smith, M.D. 2003. "Consumer Surplus in the Digital Economy:

- Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science* (49:11), pp 1580-1596.
- Carballo-Diequez, A., Miner, M., Dolezal, C., Rosser, B.R.S., and Jacoby, S. 2006. "Sexual Negotiation, HIV-Status Disclosure, and Sexual Risk Behavior among Latino Men Who Use the Internet to Seek Sex with Other Men," *Archives of Sexual Behavior* (35:4), pp 473-481.
- Carpenter, C. 2005. "Youth Alcohol Use and Risky Sexual Behavior: Evidence from Underage Drunk Driving Laws," *Journal of Health Economics* (24:3), pp 613-628.
- Chesson, H.W., Blandford, J.M., Gift, T.L., Tao, G., and Irwin, K.L. 2004. "The Estimated Direct Medical Cost of Sexually Transmitted Diseases among American Youth, 2000," *Perspectives on Sexual and Reproductive Health* (36:1), pp 11-19.
- Chiasson, M., Parsons, J., Tesoriero, J., Carballo-Diequez, A., Hirshfield, S., and Remien, R. 2006. "Hiv Behavioral Research Online," *Journal of Urban Health* (83:1), pp 73-85.
- Cunningham, S., and Kendall, T.D. 2010. "Risk Behaviours among Internet-Facilitated Sex Workers: Evidence from Two New Datasets," *Sexually Transmitted Infections* (86:Suppl 3), pp iii100-iii105.
- Cunningham, S., and Kendall, T.D. 2011. "Prostitution 2.0: The Changing Face of Sex Work," *Journal of Urban Economics* (69:3), pp 273-287.
- Dranove, D., Kessler, D., McClellan, M., and Satterthwaite, M. 2003. "Is More Information Better? The Effects of "Report Cards" on Health Care Providers," *Journal of Political Economy* (111:3), pp 555-588.
- Elford, J., Bolding, G., and Sherr, L. 2001. "Seeking Sex on the Internet and Sexual Risk Behaviour among Gay Men Using London Gyms," *AIDS* (15:11), p 1409.
- Eysenbach, G. 2003. "The Impact of the Internet on Cancer Outcomes," *CA: A Cancer Journal for Clinicians* (53:6), pp 356-371.
- Gao, G.G., McCullough, J.S., Agarwal, R., and Jha, A.K. 2012. "A Changing Landscape of Physician Quality Reporting: Analysis of Patients' Online Ratings of Their Physicians over a 5-Year Period," *Journal of Medical Internet Research* (14:1).
- Garofalo, R., Herrick, A., Mustanski, B.S., and Donenberg, G.R. 2007. "Tip of the Iceberg: Young Men Who Have Sex with Men, the Internet, and Hiv Risk," *American Journal of Public Health* (97:6), pp 1113-1117.
- Green, L., and Myerson, J. 2004. "A Discounting Framework for Choice with Delayed and Probabilistic Rewards," *Psychological Bulletin* (130:5), pp 769-792.
- Grov, C. 2010. "Risky Sex- and Drug-Seeking in a Probability Sample of Men-for-Men Online Bulletin Board Postings," *AIDS and Behavior* (14:6), pp 1387-1392.
- Gurun, U.G., and Butler, A.W. 2012. "Don't Believe the Hype: Local Media Slant, Local Advertising, and Firm Value," *The Journal of Finance* (67:2), pp 561-598.
- Halkitis, P.N., and Parsons, J.T. 2003. "Intentional Unsafe Sex (Barebacking) among HIV-Positive Gay Men Who Seek Sexual Partners on the Internet," *AIDS Care* (15:3), pp 367-378.
- Hirshfield, S., Remien, R.H., Humberstone, M., Walavalkar, I., and Chiasson, M.A. 2004. "Substance Use and High-Risk Sex among Men Who Have Sex with Men: A National Online Study in the USA," *AIDS Care* (16:8), pp 1036-1047.
- Hitsch, G., J., Hortacısu, A., and Ariely, D. 2010. "Matching and Sorting in Online Dating," *American Economic Review* (100:1), pp 130-163.
- Il-Horn, H., and Terwiesch, C. 2003. "Measuring the Frictional Costs of Online Transactions: The Case of a Name-Your-Own-Price Channel," *Management Science* (49:11), pp 1563-

1579.

- Jin, G.Z., and Leslie, P. 2003. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards," *Quarterly Journal of Economics* (118:2), pp 409-451.
- Jin, L., and Robey, D. 1999. "Explaining Cybermediation: An Organizational Analysis of Electronic Retailing," *International Journal of Electronic Commerce* (3:4), pp 47-65.
- Kahn, J.O., and Walker, B.D. 1998. "Acute Human Immunodeficiency Virus Type 1 Infection," *The New England Journal of Medicine* (339:1), pp. 33-39.
- Karniol, R., and Miller, D.T. 1983. "Why Not Wait?: A Cognitive Model of Self-Imposed Delay Termination," *Journal of Personality and Social Psychology* (45:4), pp 935-942.
- Katz, M.H., Schwarcz, S.K., Kellogg, T.A., Klausner, J.D., Dilley, J.W., Gibson, S., and McFarland, W. 2002. "Impact of Highly Active Antiretroviral Treatment on HIV Seroincidence among Men Who Have Sex with Men: San Francisco," *American Journal of Public Health* (92:3).
- Kim, A.A., Kent, C., McFarland, W., and Klausner, J.D. 2001. "Cruising on the Internet Highway," *Journal of Acquired Immune Deficiency Syndromes* (28:1), pp 89-93.
- Kroft, K., and Pope, D.G. 2013. "Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist", *Journal of Labor Economics* (forthcoming).
- Lakdawalla, D., Sood, N., and Goldman, D. 2006. "Hiv Breakthroughs and Risky Sexual Behavior," *Quarterly Journal of Economics* (121:3), pp 1063-1102.
- Laumann, E.O., and Youm, Y. 1999. "Racial/Ethnic Group Differences in the Prevalence of Sexually Transmitted Diseases in the United States: A Network Explanation," *Sexually Transmitted Diseases* (26:5), pp 250-261.
- Liau, A., Millett, G., and Marks, G. 2006. "Meta-Analytic Examination of Online Sex-Seeking and Sexual Risk Behavior among Men Who Have Sex with Men," *Sexually Transmitted Diseases* (33:9), pp 576-584.
- Logan, T., and Shah, M. 2009. "Face Value: Information and Signaling in an Illegal Market." NBER Working paper.
- McFarlane, M., Bull, S., and Rietmeijer, C. 2000. "The Internet as a Newly Emerging Risk Environment for Sexually Transmitted Diseases," *Journal of the American Medical Association* (284:4), pp 443-446.
- Metcalf, J., and Mischel, W. 1999. "A Hot/Cool-System Analysis of Delay of Gratification: Dynamics of Willpower," *Psychological Review* (106:1), pp 3-19.
- Miller, A.R., and Tucker, C.E. 2011. "Can Health Care Information Technology Save Babies?," *Journal of Political Economy* (119:2), pp 289-324.
- Moskowitz, D.A., and Seal, D.W. 2010. " 'GWM Looking for Sex - Serious Only': The Interplay of Sexual Ad Placement Frequency and Success on the Sexual Health Of "Men Seeking Men" On Craigslist," *Journal of Gay & Lesbian Social Services* (22:4), 2013/01/02, pp 399-412.
- Muder, R.R., Yu, V.L., and Woo, A.H. 1986. "Mode of Transmission of Legionella Pneumophila: A Critical Review," *Archives of Internal Medicine* (146:8), p 1607.
- Padgett, P. 2003. "The Effects of the Internet on Women's Sexual Health and Sexuality.," in: *STD/HIV Prevention and the Internet*. Washington, DC.
- Rachlin, H. 2000. *The Science of Self-Control*. Cambridge, MA: Harvard University Press.
- Reece, M., Herbenick, D., Schick, V., Sanders, S.A., Dodge, B., and Fortenberry, J.D. 2010. "Sexual Behaviors, Relationships, and Perceived Health among Adult Men in the United States: Results from a National Probability Sample," *Journal of Sexual Medicine* (7), pp 291-304.

- Romer, D., Duckworth, A., Sznitman, S., and Park, S. 2010. "Can Adolescents Learn Self-Control? Delay of Gratification in the Development of Control over Risk Taking," *Prevention Science* (11:3), pp 319-330.
- Seamans, R., and Zhu, F. 2013. "Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers," *Management Science* (forthcoming).
- Strack, F., and Deutsch, R. 2004. "Reflective and Impulsive Determinants of Social Behavior," *Personality and Social Psychology Review* (8:3), August 1, 2004, pp 220-247.
- UNAIDS. 2009. "Aids Epidemic Update 2009." from http://data.unaids.org/pub/report/2009/jc1700_epi_update_2009_en.pdf
- Wasserheit, J.N., and Aral, S.O. 1996. "The Dynamic Topology of Sexually Transmitted Disease Epidemics: Implications for Prevention Strategies," *Journal of Infectious Diseases* (174:Supplement 2), pp S201-S213.
- Wilson, M., and Daly, M. 2006. "Are Juvenile Offenders Extreme Future Discounters?," *Psychological Science* (17:11), pp 989-994.
- Yan, L., and Tan, Y. 2010. "An Empirical Study of Online Supports among Patients." Working paper.

Appendix

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

Author Biographies

Jason Chan is a Ph.D. candidate in the Department of Information, Operations and Management Sciences at New York University's Leonard N. Stern School of Business. His research interests involve examining and quantifying the impact of online platforms on social, health, and economic outcomes. His work has received media mentions in Forbes, Market Watch, NBC, and Washington Post, and has appeared in IEEE Transactions of Engineering Management, Journal of Association of Information Systems, National Bureau of Economic Research Summer Meetings, International Industrial Organization Conference, among others. His research received a best paper award at the Workshop of Health IT and Economics. He received his B.Sc. degree from National University of Singapore.

Anindya Ghose is a Professor of Information, Operations and Management Sciences and a Professor of Marketing at New York University's Leonard N. Stern School of Business. He is the co-Director of the Center for Business Analytics at NYU Stern. He is the Robert L. & Dale Atkins Rosen Faculty Fellow and a Daniel P. Paduano Fellow of Business Ethics at NYU Stern. He has been a visiting Professor at the Wharton School of Business and is an affiliated faculty at Carnegie Mellon University. He had worked with leading firms in the information technology, retail, financial services, telecommunications, and travel industries on projects related to internet marketing, social media analytics, mobile marketing and digital advertising analytics across the US, India, China, Europe and South Korea. He also plays a senior advisory role to start-ups in the Internet and mobile space. His research has been profiled numerous times in the BBC, New York Times, Financial Times, Forbes, NBC, CNBC, USA Today, Xinhua, Reuters, Washington Post, New York Daily, National Public Radio, Wall Street Journal, MSNBC, China Daily, Knowledge@Wharton, and elsewhere. He teaches courses on social media, digital marketing, and business analytics at the undergraduate, MBA, Executive MBA, and Executive Education level in various parts of the world. His work has been published in leading journals in Marketing, Economics, IS, and Computer Science. His research has received nine best paper awards and nominations. He is a Marketing Science Institute Young Scholar and a recipient of the highly prestigious National Science Foundation CAREER award. His research has also been recognized by five globally competitive awards from Google and Microsoft. He serves as an Associate Editor of Management Science and a Senior Editor of Information Systems Research. He received his M.S. and Ph.D. from Carnegie Mellon University's Tepper School of Business and an MBA from IIM Calcutta.