A Double Digital Divide?
Matching Platforms and HIV Incidence among the Digitally Disadvantaged

Abstract

Although recent work has begun to examine the negative implications of digitized matching platforms enabled by the Internet, limited attention has been paid to whom the negative externalities of these platforms accrue. In this paper, we examine how the entry of platforms for the solicitation of casual sex influences the incidence rate of HIV by race, gender, and socio-economic status. Using a census of 12 million patients who are subjected to a natural experiment in Florida between 2002 and 2006, we find that the largest negative effect accrues to historically at risk populations (i.e. African Americans and the socio-economic lower class) that, ironically, are also disadvantaged with respect to digital inequalities. Economically, this translates into 958 additional patients contracting HIV during the 5 year period of our study (74.59% of them *African American*), and an additional financial burden of $592 million in the State of Florida alone.

Key Words: *public health, two sided matching, platforms, natural experiment, HIV, digital divide*

Introduction

The promise of public health benefits arising from the connectivity enabled by the Internet has recently captured the attention of scholars and policy makers. Frequently discussed advantages of digital connectivity for health range from support groups for chronic health conditions (Goh et al. 2009), to the use of electronic health information exchange (Agarwal et al. 2010), to applications such as Google Flu Trends that assist public health officials in monitoring epidemics (Dugas et al. 2012). However, while substantial research documents the benefits of connectivity, a growing literature in both information systems (Chan and Ghose 2013) and public health (Benotsch et al. 2002, Elford et al. 2001, Parsons et al. 2006) has begun to examine the pitfalls of connectivity as well. One concern is how the very same platforms that increase market efficiencies, decrease frictions, and enable the two sided matching process for the exchange of goods and service (e.g. eBay and Amazon.com) (Bakos and Bailey 1997, Brynjolfsson and Smith 2000, Parker and Van Alstyne 2005), are increasingly facilitating risky behaviors, such as casual sex (Chan and Ghose 2013) and drug use (Grov 2010). Beyond the social cost of increased morbidity and mortality that these behaviors create, there is also an economic burden often borne by governments and taxpayers. To the extent that treatment for sexually transmitted diseases (STDs) and rehabilitation for drugs users is expensive (Schackman et al. 2006), the externalities of these sites pose significant challenges for policy makers.

While the fact that Internet-enabled matching platforms are increasing risky behavior is recognized and documented in extant literature (Benotsch et al. 2002, Chan and Ghose 2013, Kim et al. 2001, Parsons et al. 2006), and initial aggregate estimates of the effects have been computed, one critical issue remains under studied: an understanding of the subpopulations to whom the negative externalities accrue. To the extent that there may be variation in the effects experienced across socio-demographic groups, and to the degree that health disparities are a significant policy concern (HealthyPeople 2013), understanding who is vulnerable to platform entry, and isolating the magnitude of this effect is important.

Furthermore, much of the recent work in public health related to the use of online platforms for the solicitation of casual partners (Garofalo et al. 2007, Grov 2010) is based on small scale surveys. As a result, many researchers have been unable to isolate whether the existence of the platform is increasing the amount of sexual behavior which is occurring or, simply displacing the previous offline matching process. A notable exception to small sample survey research is Chan and Ghose (2013), in which the authors examine the state-level spread of the human immunodeficiency virus (HIV) after the introduction of Craigslist; finding a nearly 16% rise in case rate at an annual cost of approximately $65 million. However, as their analysis is conducted at the state-year level with archival incidence reports from the Centers for Disease Control (CDC), the authors are unable to isolate the socio-economic and demographic characteristics of individuals who are penalized. Additionally, the archival CDC reports utilized for HIV estimates do not distinguish between different stages of disease progression for the diagnosed carriers[1]. As the clinical latency period between infection, symptomatic HIV, and then AIDS lasts, on average, 10 years in the complete absence of treatment (DHHS 2012), relating the introduction of the platform to the CDC data may yield biased estimates of the effect.

In this study we estimate the impact of Internet-enabled matching platforms on the incidence rate of asymptomatic HIV. We further examine which subpopulations, based on race, gender, and socio-economic status (SES), experience greater negative externalities as a result of platform availability. The answer to this question is not clear, *ex ante*, and poses an interesting theoretical puzzle. On one hand, epidemiological accounts of HIV suggest that infection is far more prevalent among ethnic minorities and those of lower SES (CDC 2011). However, studies relating to digital inequality underscore that these groups are less likely to utilize online resources as a result of decreased access, literacy, and training (DiMaggio et al. 2004, Hargittai 2010, Warschauer 2004).

We quantify the effect of these platforms on the HIV incidence rate[2] by exploiting a natural experiment: the introduction of Craigslist in the state of Florida between January of 2002 and December of 2006. We match the introduction of Craigslist to a census of patients in Florida hospitals. These data

---

[1] "This report presents estimated numbers of cases of HIV/AIDS (cases of HIV infection, regardless whether they have progressed to AIDS)..." (CDC, 2006)

[2] Incidence rate is defined medically as the number of new cases per population in a given time period (Dorland 2011)

contain detailed information about HIV, as well as patient characteristics. Our research, therefore, responds to calls to study the effects of variation in Internet availability and use across populations characterized by specific demographic and socio-economic factors (DiMaggio et al. 2004) and, more generally, contributes to research seeking to unpack the social and economic effects of the Internet.

Econometrically, as the entry of Craigslist is staggered temporally and geospatially, search for sexual partners is highly localized (Zenilman et al. 1999), and our data contain diagnoses of asymptomatic HIV, we are able to quantify the effect of platform introduction with more precision than previous investigations. Our identification strategy exploits the exogeneity of platform introduction into different areas at different times. To the extent that the broader objective of Craigslist is not the facilitation of the solicitation of sexual partners (this section of the platform is non-revenue generating and is one of the many functionalities that Craigslist offers), we can reasonably assume that the entry of the Craigslist forum is driven by its primary revenue stream (the posting of classified ads for employment), and not by the sexual proclivities of individuals in the area.

Empirical analysis yields three important findings robust to multiple specifications and falsification tests. First, consistent with Chan and Ghose (2013), the introduction of Craigslist significantly increases the incidence rate of HIV for patients admitted to local area hospitals. Our aggregate estimate of the effect of Craigslist on asymptomatic HIV, arguably the most proximal outcome to platform entry, suggests, however, that previous estimations of the size of this effect may have been overstated[3]. Estimates indicate the rise in asymptomatic HIV is roughly 13.5% (compared with Chan and Ghose's (2013) 15.9%)[4], resulting in an annual public welfare loss of $5.749 million in Florida as of 2006. In the aggregate, this translates to a cumulative public welfare loss of $592 million over the lifetime of all patients contracting HIV as a result of Craigslist entry during the time of the study[5].

Second, the incidence rate increase in HIV after platform entry is higher for subpopulations that

---

[3] It is important to note that Florida is a state which is characterized by higher than average HIV infection (CDC, 2006)
[4] Without knowledge of the number of Craigslist treated hospitals by state a further comparison in infeasible.
[5] The annual public welfare loss is calculated using Chesson et al's (2004) estimate of $10,121.60 annual treatment costs for HIV. The cumulative public welfare loss is calculated Schackman et al.'s (2006) estimated lifetime treatment cost of $618k.

are traditionally considered at risk for the disease, African Americans and those of lower SES, despite the documented lower level of access and use of online services among these groups. Third, and a cause for some policy concern, results indicate that despite the fact that men are considered as greater risk for HIV than women (CDC 2011) and women are modestly less likely to use the Internet (Hargittai 2010), there is no significant difference in the effects of the platform across genders: both men and women experience equal penalties with respect to the proliferation of the disease.

Background

Two sided matching platforms that provide the digital infrastructure for distinct communities of users to transact with each other have been investigated in both the economics and information systems literatures (Bakos and Bailey 1997, Brynjolfsson and Smith 2000, Parker and Van Alstyne 2005). Research suggests that these platforms reduce market frictions by serving as digital intermediaries between buyers and sellers (Bakos and Bailey 1997, Brynjolfsson and Smith 2000). From a transaction cost perspective, digital platforms reduce bargaining asymmetry and protect transacting agents from opportunism by increasing price transparency (Williamson 1981). From a markets and hierarchies perspective, the reduced search cost facilitated by the platforms decreases operating cost and can accelerate the buyer-seller match (Malone et al. 1987).

Much like early manifestations of digital two sided matching platforms (e.g. eBay and Amazon.com), online dating services such as eHarmony, Match.com, and Zoosk, have experienced wide success as a result of their ability to resolve friction in offline markets for social transactions (Bapna et al. 2012). Not unlike digital commerce platforms, dating websites facilitate the partner matching process through two mechanisms: self-selection and decreased search costs (Brynjolfsson and Smith 2000). To the extent that these platforms position themselves as satisfying the needs of specific demographics, ranging from eHarmony's general goal of facilitating long term relationships to ChristianMingle's segmented market approach based on theological belief, they are able to efficiently subdivide the market into prospective partners who meet each other's matching criteria. Further, the platforms typically gather extensive data about users to facilitate matching, resulting in drastically reduced search costs. Simply put, by concentrating users with particular tastes or objectives on an easily searchable platform these services reduce the search cost for acceptable matches, and increase the ease of sampling (i.e. dating).

Together with sites that purportedly facilitate more enduring relationships, the Internet has also spawned platforms that enable the solicitation of casual sexual encounters with no long term relationship goal attached to them (e.g. AdultFriendFinder and AshleyMadison.com). While such platforms provide equivalent benefits of market segmentation and reduced search cost, they offer the added advantage of anonymity. Because users can utilize their services without fear of social stigma; anonymity decreases another significant cost: the social damage of engaging in risky activity (Parsons et al. 2006). Although not critical for the market to function, anonymity reduces the effect of social frictions that may constrain the offline two sided sexual matching process (Kim et al. 2001). Such frictions, often a result of prevailing societal norms, include social castigation for promiscuity (Brown 1988), closeted sexual orientation (Floyd and Stein 2002), or even enforcement of a social contract (i.e. marriage) (MacDonald 1995). Research in psychology suggests that the security of anonymity (i.e. a lower likelihood of being discovered engaging in deviant behavior) is associated with more risky behaviors (Rogers 2010). With a guarantee of anonymity, to the extent that individuals no longer fear social reprimand or even criminal prosecution as a result of their actions, there is likely a decreased propensity to adhere to social norms and an increased tendency to engage in risky behavior (Padgett 2007).

Unsurprisingly, empirical work in public health (Benotsch et al. 2002, Grov 2010) has highlighted the risky behavior which these sites facilitate. Not only are users more likely to engage in unprotected sex (regardless of race (Bingham et al. 2003), gender (Padgett 2007), or sexual orientation (Garofalo et al. 2007)) with an increased number of partners (Benotsch et al. 2002, Elford et al. 2001), users of these websites are more likely to carry an STD (Elford et al. 2001) and engage in drug use (Grov 2010). Despite the growing research in this area, however, extant work has yet to address the question of *who* might be more vulnerable to the adverse consequence of contracting HIV as a result of the availability of platforms that enable the solicitation of casual sex. To answer this question, we juxtapose the characteristics of two populations: those identified as being "at risk" for HIV and those with limited access to and training on the use of online resources.

Digital Disparities and Inequality

With more than 93% of the American population having access to faster than dial-up Internet (NTIA and ESA 2013), and a large and growing segment of the population having grown up in the digital age, two increasingly common, though not entirely accurate, perceptions have emerged. First, to the extent that high speed Internet is becoming ubiquitous, it has been suggested that the domestic digital divide and digital inequalities are relics of the past that further attenuate in prevalence with each passing year (Barry 2013). Second, by virtue of early socialization with technology, young people are adept technology users; across the spectrum race, gender, and SES (Hargittai 2010). Ironically, despite the pervasiveness of these beliefs, limited empirical evidence exists to supports them (NTIA and ESA 2013). Research suggests not only that large racial disparities still exist in Internet utilization (Hargittai 2010, NTIA and ESA 2013, Zickuhr and Smith 2012), but that differences are largely driven by income and education inequalities (Hsieh et al. 2011). Moreover, even when access to online resources is equal, there exist striking differences in *how* online resources are utilized by these groups (DiMaggio et al. 2004, Hargittai 2010). Not only are the socio-economically advantaged far more skillful in their exploitation of online resources, but access and use by the socio-economically disadvantaged is rarely for capital or welfare enhancing activities (Hargittai 2010, Zillien and Hargittai 2009).

In light of this work, it is not clear to whom the negative externalities of platforms for the solicitation of casual sex will accrue. As Internet utilization rates vary across the population (DiMaggio et al. 2004, NTIA and ESA 2013, Zickuhr and Smith 2012), the likelihood of exploiting online resources is lower for individuals with differential familiarity and computer literacy (Hargittai 2010, Warschauer 2004, Zillien and Hargittai 2009). Because this utilization disparity usually falls along socio-economic lines, with those of lower SES[6] and ethnic minorities suffering a disproportionate penalty (NTIA and ESA 2013, Zickuhr and Smith 2012), it is plausible that these groups will not leverage online platforms to

[6] "Socioeconomic status is commonly conceptualized as the social standing or class of an individual or group. It is often measured as a combination of education, income, and occupation" (APA 2013)

solicit sexual partners.

However, national estimates of HIV infection indicate that both people living below the poverty line as well as ethnic minorities (specifically, Latinos and African Americans) are far more likely to be carriers of the HIV virus (CDC 2011). Further, epidemiological research suggests that HIV infection has an unequal effect on men (specifically bi- or homosexual men) as well as individuals who engage in risky sexual behavior (CDC 2011). Comparing the two groups, individuals who are digitally disadvantaged and individuals who are at higher risk for HIV, we see a significant overlap that presents an interesting theoretical puzzle. On the one hand, the digitally disadvantaged are more likely, *ex ante*, to be HIV carriers. On the other, they are also less likely to exploit online resources. As the use of the platform requires digital literacy and access, the propensity of populations not considered at risk for contracting the disease to use the platform will be higher. Yet, it is also likely that the partners these users encounter online are also, relatively, digitally advantaged thereby decreasing their risk of being an HIV carrier. To the extent that there are significant barriers to disease infection that characterize both groups and in the absence of a compelling *a priori* expectation of to whom the negative externalities will accrue, we address this puzzle empirically.

Data and Methodology

Context: Craigslist

To quantify the effect of the introduction of online platforms for the solicitation of casual sexual partners on different sub-segments of the population we exploit a natural experiment: the introduction of Craigslist into major cities in the State of Florida between January of 2002 and December of 2006[7]. Craigslist is a community based online forum for classified advertisements which was launched in 1995 in San Francisco. As of May 2014 the site, by far the largest to offer these services, received nearly 50 million visitors per month[8]. In addition to forums for the posting of resumes, items wanted, employment opportunities, housing, and music shows, Craigslist also offers a personal ads section. The personal ads

[7] Others (Chan and Ghose 2013; Seamans and Zhu 2011) have similarly exploited such natural experiments.
[8] Craigslist currently has an Alexa ranking of 10 in the United States, thereby making it the platform which facilitates the solicitation of partners. Ashley Madison and AdultFriendFinder, via comparison, have Alexa ratings of 506 and 384 respectively.

section, which hosts diverse opportunities for couples and groups to meet and interact, contains a forum called "casual encounters" which can be used for the solicitation of sexual partners. Individuals post advertisements on the casual encounters forum indicating their gender and age, preferred gender of respondents, a description of other preferences, and a method for response[9]. These advertisements, and the platform itself, satisfy the criteria for two sided matching platforms: the community is populated with likeminded individuals seeking casual sexual encounters (as the title of the forum would imply), the platform is easily searchable by user preferences (i.e. age, race, and sexual orientation), and users can operate anonymously[10].

Data

We construct a longitudinal dataset that contains a census of patients admitted to hospitals in the State of Florida from January, 2002 to December, 2006. The source of these data is the Florida Agency for Healthcare Administration (AHCA) which provides us with bed level information about every patient admitted into a hospital in the state during that time. We constrain our analysis to these dates as October of 2002 is the date of the first local implementation of Craigslist (Miami-Dade) and detailed information regarding the implementation of Craigslist is unavailable after 2006[11]. Table 1 summarizes the areas where Craigslist was introduced during the study time frame together with a comprehensive list of the locations in Florida where Craigslist was eventually installed. The AHCA dataset, used widely in prior research (Burke et al. 2003, Burke et al. 2007), offers the benefit of both observing when the patient is admitted to the hospital, i.e. before or after Craigslist implementation, as well as detailed information about the patient and the medical conditions the patient has been diagnosed with through International Classification of Diseases Revision 9 (ICD-9) codes. Due to privacy concerns, these data are aggregated to the quarter level by the AHCA (discussed further below).

Variable Definitions

Dependent Variable: The dependent variable for the main analysis is the number, i.e. count, of

---

[9] http://miami.craigslist.org/i/personals?category=cas

[10] While current versions of the website require posters to register a telephone number in order to submit an ad, these user protections had not been implemented during the time of our study.

[11] http://www.craigslist.org/about/expansion

asymptomatic HIV carriers (ICD-9 diagnosis code 'V08') admitted to a hospital in the focal quarter. As noted earlier, we use asymptomatic, as opposed to symptomatic, HIV because the latter can take years or even decades to manifest after initial infection. The delay occurs because the body initially is capable of fighting the virus, resulting in the patient entering a clinical "latency period" 3-12 weeks after infection. A detailed explanation of disease progression from Pantaleo et al. (1993) is available in Figure 1. This latency period, as well as the initial period of acute HIV syndrome where the patient will often experience nausea, weight loss, and swelling of the lymph nodes, is medically defined as asymptomatic HIV. Given the length of time symptomatic HIV and AIDS take to be observed, the asymptomatic stage of the disease is a more appropriate proximate effect of the "treatment" in this study. Further, since 1997 the standard treatment for HIV management has been highly active antiretroviral therapy (HAART), which can be administered in pill format and does not require hospitalization (Finzi et al. 1997). Econometrically, this mitigates significant identification problems by limiting re-hospitalization (because chronic management of the disease can be conducted in an outpatient setting).

Before describing the independent variables we make two qualifying observations. First, medically, once a patient has been classified as a symptomatic carrier they can never be reclassified as an asymptomatic carrier[12]. Statistically, this ensures that the population of patients seeking treatment does not intermittently swap between HIV classes. Second, our dependent variable only captures patients who seek treatment for the HIV virus. Although the patient may be diagnosed at any number of locations (e.g. blood drives, annual visits to their primary care provider, mobile testing units, etc.) accounts from active practitioners suggest that an individual (upon receiving an initial diagnosis at a blood drive or clinic) will be referred to a hospital for more thorough antibody/antigen testing. This is done for two reasons: first, to ensure that the diagnosis is not a false positive and second, to identify the type of HIV affliction (in order for proper construction of the HAART therapy "cocktail"). As of early 2014, there were 59 identified Circulating Recombinant Forms of the HIV virus across both strains of HIV.

[12] The stages of HIV are defined by the CDC based on CD4+ T-cell count (CDC, 2008) which will only decrease after infection

Independent Variables: The central independent variable of interest is the dichotomous indicator *Craigslist*, which indicates that the focal patient has been admitted to a hospital which is in the same county as a city in which *Craigslist* has been released (as determined by Table 1)[13]. As our data from AHCA is aggregated at the quarter level, this variable is set to 1 in the first *full* quarter during which the county has had access to Craigslist, as well as all subsequent quarters.

To investigate the effect of Craigslist introduction across various different subpopulations we disaggregate our dependent variable (*Total Count*) based on the various indicators which our data capture. The first, *Race/Ethnicity,* henceforth referred to as *Race,* uses three groups: the first two, *African American* and *Latino*, are indicative of populations typically characterized as more likely to be digitally disadvantaged (DiMaggio et al. 2004, Hargittai 2010), while the third group, *Caucasian*, represents a subpopulation which is not categorized as digitally disadvantaged. We next subdivide our data based on the *Gender* of the patient; as casual sex seeking behavior is observed more often in men (Elford et al. 2001). Finally, as a proxy for the SES of the patient we classify patients based on whether or not they are recipients of *Medicaid*. Because *Medicaid* is available only to people with an income of $1,869 a month or less ($1,439 in 2002 dollars) in the state of Florida[14], it is a reasonable indicator of low SES. Other patients are aggregated to the group *non-Medicaid*.

To control for time invariant unobserved heterogeneity we include fixed effects for the hospitals[15] in Florida. To account for potential effects of unmeasured environmental shocks, fixed effects for the 20 time quarters of the study period are also included. Hospitals which are absent from more than half of the study, either because they open after halfway through the sample, close before the halfway point of the sample[16], or have no record of admitting an HIV positive patient (in which case there is no variation in the

---

[13] We assume that the "treatment" effect of Craigslist entry into an area is experienced throughout the county within which the city is located.

[14] http://www.floridamedicaideligibility.com/asset.html

[15] Although the treatment of our study is applied at the county level we include hospital fixed effects to account for the fact that patients of different backgrounds may select into seeking treatment at different hospitals within the same county. Moreover, this increases the precision of the control for unobserved heterogeneity in the empirical estimations

[16] Robustness checks have been conducted eliminating all hospitals not present for the study's duration. Results are consistent and available upon request

dependent variable), are dropped from the sample (37 of 260). Hospitals that are present for more than half the study period are retained, even if they are not present in every time period. With these restrictions, the final data set used for analysis consists of 11,793,617 patients across 223 hospitals over 20 quarter time periods, yielding an effective N for hospitals of 4349, after accounting for hospitals with are partially observed. Summary statistics and correlations are reported in Table 2.

Empirical Strategy

As our dependent variable is a count we use a fixed effect negative binomial estimation with robust standard errors clustered on the county, resulting in an estimation of the aggregate difference in the differences between treated and untreated hospitals. We cluster our standard errors at the county level because the treatment, *Craigslist,* is applied at the county level[17]. We model the number of asymptomatic HIV patients which are admitted into the focal hospital *j* at time *t*, (y_{jt}) using the following specification:

$$y_{jt} = \beta_1 s_1 + M'\theta_1 + X'\delta_1 + \nu + \varepsilon$$

The variable s_1 is the dichotomous indicator of *Craigslist* treatment. M is the vector of hospital fixed effects and X is the vector of time fixed effects. The term $\{\beta_1\}$ is the parameter to be estimated and ν represents the constant. As discussed previously, after estimating the cumulative effect for the entire population, i.e. the aggregate difference in difference, we estimate the effect on the population based on *Race* [*Caucasian, African American,* and *Latino*], *SES* [*Medicaid* and *Non Medicaid*], and *Gender* [*Male* and *Female*]. Results are available in Table 3.

Results

In Table 3 we first note a strong and significant increase in the number of asymptomatic HIV patients admitted after the entry of *Craigslist*, thereby corroborating the effects found in Chan and Ghose (2013). Economically, this translates to an absolute increase of 1.136 patients per hospital quarter being admitted with asymptomatic HIV (or a relative increase of 13.5% over the average 10.13 patients per hospital quarter admitted in the absence of *Craigslist*)[18]. Interestingly, while we obtain a comparatively consistent

---

[17] We validate the assumption of spatial reach of treatment at the county level using alternative radii in robustness checks described later.

[18] All marginal effects are calculated using the `margins` command of Stata 12.1

effect across *Gender* (12.73% increase for men/14.35 % increase for women)[19], the racial effect appears to be concentrated entirely in the *African American* community, i.e. there is no statistically significant effect for *Caucasians* (p-value of 0.125) or *Latinos* (p-value of 0.103). To the extent that these effects could be described as marginally significant, it is important to note that they are positive (as a negative effect on the asymptomatic HIV rate would be counter intuitive). Furthermore, results show that there is a significantly different effect across SES. Results from a seemingly unrelated regression indicate that the increases of 12.74% for non-*Medicaid* patients and 13.92% for *Medicaid* patients are significantly different at the ($p<0.01$) level.

Graphical depictions of the results support these findings (Figures 2-9). As seen in Figure 4 the effect of *Craigslist* entry on the *African American* community is significantly larger (both in terms of the relative increase and the absolute increase) in the number of patients (compared with *Caucasians* in Figure 3 and *Latinos* in Figure 5). It is noteworthy that although the highest prevalence of HIV is in the *African American* community, these individuals constitute only 15.8% of patients admitted to hospitals in the sample (as opposed to 68.6% for *Caucasians*), further underscoring the disproportionate effect HIV is having on this subpopulation. Furthermore, while men are more likely to carry the HIV virus, we see in Figure 6 and 7 that the change in effect as a result of *Craigslist* entry is similar across the *Genders.* Finally, we see in Figures 8 and 9 that although there is a higher *ex ante* concentration of HIV in the *non-Medicaid* population, the increase post treatment for the *Medicaid* population is larger.

Unpacking the Effect on the African American Community

Motivated by the fact that the *African American* subpopulation is typically considered to be disadvantaged with respect to technology access and use, juxtaposed with the finding that it is experiencing a disproportionate penalty, we explore this effect further. To the extent that our aggregation of the data on each *Race* may mask underlying heterogeneities in the effect (i.e. a greater effect on *African American Medicaid* patients or men), we split the sample for each of the *African American* subpopulations and replicate our analysis (Table 4). Interestingly and consistent with the effect obtained at the aggregate

[19] Results from a seemingly unrelated regression indicate that there is no statistical difference across Gender

level, we see that the effect of *Craigslist* is equivalent across *Gender*. *African American* women experience a 14.39% increase in the asymptomatic HIV rate after *Craigslist* is introduced (compared with 14.95% for men). As was also found at the aggregate level, *SES* is a strong moderator. While *African American non-Medicaid* patients experience an increase of only 13.03% in the asymptomatic HIV incidence rate post-*Craigslist* implementation, *Medicaid* patients experience a 16.08% increase. Results from a seemingly unrelated post-estimation of the results substantiate that these effects are different at the ($p<0.05$) level. Graphical depictions (Figures 10-13) confirm the results[20].

Relative Time Model

Our core empirical strategy has been an aggregate difference in the differences approach between treated and untreated hospitals, i.e. a between subjects estimate. For this estimate to be valid, one concern that must be addressed is whether or not there is homogeneity in the pre-treatment trend between treated and non-treated hospitals. The concern arises because it possible that unobservable and randomly distributed environmental factors, which are native to the individual metropolitan areas, cause heterogeneity in the pre-treatment asymptomatic HIV incidence rate. For example, because *Craigslist* enters larger metropolitan areas first (e.g. Miami and Tampa) it is possible that there is a different trend in the HIV rate between these areas and those which do not receive the treatment. To rule out this potential confound and further substantiate the claim that the entry of *Craigslist* can be treated as an exogenous event at both the aggregate and subpopulation levels, we execute the following estimation:

$$y_{jt} = \rho'[s_2 * \varphi] + M'\theta_2 + X'\delta_2 + \nu + \varepsilon$$

where y_{jt} is number of asymptomatic carriers admitted to hospital j during time t, s_2 is a dichotomous variable which indicates whether or not the *Craigslist* treatment will ever affect hospital j during the study, and φ is a series of time dummies which indicate the relative chronological distance between time t and *Craigslist* implementation at hospital j, i.e. the number of quarters preceding, or following, the

[20] Similar analyses have been conducted on the *Caucasian* and *Latino* subpopulations. *Caucasian* analysis yielded no significant effects for any subpopulation. *Latino* analysis yielded a modest effect for *Medicaid* patients and women. Results of a seemingly unrelated regression indicate no statistically significant differences in the effect across men and women or across *Medicaid* and *non-Medicaid* patients

implementation of *Craigslist*. M represents the vector of hospital fixed effects while X is the vector of time fixed effects. The vector $\{\rho\}$ contains the parameters to be estimated and ν represents the constant. Intuitively, what this model allows us to do is determine whether or not there is a pre-treatment trend that is disproportionately affecting *Craigslist* hospitals, as opposed to non-*Craigslist* hospitals. As in our previous analyses, after estimating the cumulative effect for the entire population, we estimate the effect for each subpopulation based on *Race* [*Caucasian, African American,* and *Latino*], *SES* [*Medicaid* and *Non Medicaid*] and *Gender* [*Male* and *Female*]. The estimator is a negative binomial with robust standard errors clustered on the county and the omitted relative time variable is $Craigslist_{t0}$.

Results (Table 5) confirm previous findings and rule out several possible econometric concerns related to the assumptions of the difference in difference strategy. First, we see no effect of the *Craigslist* treatment on the asymptomatic HIV incidence rate before it is implemented (both in the aggregate and across subpopulations), with one exception. Of the thirty two observed pre-treatment quarters, one is significant (*Male* in $Craigslist_{t-3}$)[21]. However, the coefficient is negative and nine months prior to implementation with no discernable trend in the remainder of the pre-treatment dummies. From this we conclude that the assumption of the difference in difference models, i.e. that there is no significant and detectable dissimilarity pre-treatment, is not being violated (Banerjee and Duflo 2000). Second, we see (Column 1) that the effect of *Craigslist* is insignificant initially but grows in magnitude over time, as we might expect, becoming significant roughly nine months after implementation. Third, as seen in Column 3, the largest and fastest effect to manifest is within the *African American* population, becoming significant six months after implementation, with a net effect of 0.97 patients per hospital quarter (a 16.9% increase) one year after implementation. Interestingly, we see a small but significant effect within the *Caucasian* and *Latino* populations as well, albeit taking substantially longer (six months) to become observable. Finally, corroborating our previous findings, we see a significantly faster effect for *Medicaid*

[21] We also note that the pre-treatment dummy (not displayed) for Latinos in $Craigslist_{t-5}$ is negative and significant. Given the number of coefficients which have been estimated the probability of two out of forty being correlated is 5%, our significance cutoff level.

patients.

Robustness Checks

Thus far, the main results in the paper have been stable across different specifications, consistently pointing to significantly varying effects across socio-economic groups and *Race* (with the *Medicaid* population and *African Americans* being affected the most). Furthermore, the incidence rate for HIV increases significantly for both *Genders* after the introduction of the *Craigslist*, with the effects for men being indistinguishable from those for women. While the relative time model indicates the absence of a discernible pre-treatment difference across treated and untreated counties, both in the aggregate and at the level of socio-demographic subpopulations, and the inclusion of hospital and time fixed effects mitigates many possible confounding factors, such as the effect of the introduction of other sites for casual solicitation (which are usually implemented without geographic restrictions), other alternate explanations for our findings exist. We therefore conduct an extensive set of falsification tests to further eliminate competing explanations. With the exception of the varying treatment radius, where the subpopulation is not relevant, and changes in the propensity to be tested for HIV, where data are unavailable, we examine the robustness of our findings at both aggregate and subpopulation levels.

Varying Treatment Radius

Our first empirical concern is validating the size of the treatment radius, i.e. the geospatial effect of *Craigslist*. Due to the relative mobility of Floridians, a state with 15.69 million registered automobiles[22] for a population of 18.09 million people, we coded the treatment at the county level for the main analysis. To ensure robustness of the assumption we further investigate if contracting or enlarging the treatment radius, i.e. the geographic area assumed to be affected when *Craigslist* is implemented, changes the size of the net effect. From a policy perspective, a deeper understanding of spatial reach is important. If, for example, hospitals within the city limits of Miami and Tampa Bay were unaffected by the *Craigslist* implementation, but neighboring suburban hospitals were affected this would change the policy implications of the result drastically. To conduct this analysis we introduce two new dummy variables:

[22] http://www.census.gov/

Craigslist (City) and *Craigslist (Greater County)*. *Craigslist (City)* indicates that the hospital resides within the limits of a city which has received the *Craigslist* treatment. *Craigslist (Greater County)* indicates that the hospital is within a contiguously adjacent county to the county which received the *Craigslist* treatment. Results are available in Table 6.

As expected, a smaller treatment radius (at the city level, Column 1 of Table 6) results in a significant change in the asymptomatic HIV rate. However, consistent with research showing that search patterns for sexual partners are highly localized (Zenilman et al. 1999), the effect becomes insignificant as the radius increases beyond the county border (Column 3 of Table 6). These results support the validity of the assumptions that the geographic dispersion of *Craigslist* implementations is sufficient to be considered a natural experiment, that the implementation of *Craigslist* has a significant impact on the asymptomatic HIV rate within cities and local areas that are treated, and that it is appropriate to cluster estimates at the county level.

Other Diagnoses

One further concern with our initial estimations is that patients are not necessarily being admitted into hospitals because of a suspected infection with the HIV virus but, rather, for some other condition. For example, it is possible that a patient who suffers a heart attack will be admitted to the hospital and, during the course of her treatment, be diagnosed with HIV. Because the Florida AHCA data do not allow us to track patients over time (due to privacy concerns), this raises the possibility that the dependent variable may be counting patients multiple times. While it is unlikely that patients who have been diagnosed with asymptomatic HIV will be readmitted to the hospital due to HIV related illness (unless they have been infected with an opportunistic disease, in which case they will subsequently be diagnosed with symptomatic HIV and not be observed in the sample) [23] it is possible that a patient who has been previously diagnosed is readmitted for an emergent condition (such as a heart attack).

To mitigate this alternate explanation we re-execute our estimations using the number of patients

[23] As is well established in the virology literature (Palella et al. 2006), treatment of asymptomatic HIV in the long term typically occurs in an outpatient setting. Indeed, the definitive source for tracking HIV is the HIV Outpatient Study (Bozzette et al. 1998, Holmberg et al. 2004).

afflicted with other non-communicable medical conditions, which should be uncorrelated with *Craigslist* introduction, as the dependent variable. The rationale for this approach is two-fold. First, if indeed patients are being admitted randomly because of other conditions, i.e. the admittance is uncorrelated with the *Craigslist* treatment, then the effect of additional admittances will be wholly contained within the error term and not bias the estimation of the asymptomatic HIV effect. Second, we use non-communicable diseases because if, as we argue, *Craigslist* is increasing the density of social connections, then it may influence the incidence rate of communicable diseases (e.g. influenza, pneumonia). Empirically, we use the four leading causes of death in the United States (Heart Attack (AMI), Stroke, Melanoma, and Lung Cancer (CDC 2013)). In addition, we test the effect of the *Craigslist* treatment on symptomatic HIV to ensure that *Craigslist* is not positively influencing the incidence rate during the time of our study. Results are available in Table 7.

In each of these estimations we see that the entry of *Craigslist* into a county is not significantly influencing the incidence rate of the measured conditions. Econometrically, this suggests that the effect of patient readmission is not biasing the estimate of the effect of *Craigslist* on the asymptomatic HIV rate. Furthermore, it eliminates the possibility that the increased admittance rate of HIV patients is a result of structural changes occurring at *Craigslist* hospitals. For instance, if local or state governments had been making large scale capital investments in their local area hospitals then patients with more severe conditions, such as HIV, may travel to these hospitals (because of their increased ability to treat severely ill patients). Findings suggest that this alternative explanation is not driving the effect. Finally, our results corroborate that there is no effect on the symptomatic HIV rate[24].

Exposure Model

While our estimates of the effect of *Craigslist* on non-communicable diseases suggest that there is no aggregate change in the population after the entry of *Craigslist*, and our estimation of the pre-treatment trend in the relative time model suggest there is no change in HIV incidence pre-treatment (either in the

[24] These estimations have also been performed on the number of patients with Heart Attacks, Strokes, Lung Cancer, Melanoma, and Symptomatic HIV for each of the individual subpopulations (e.g. *Caucasian, Medicaid*). The effect of *Craigslist* introduction is insignificant for each condition in each of the subpopulations as well. Results are available upon request.

aggregate or in the individual subpopulations), it is plausible that there are changes in the *relative* composition of the population between treated and untreated counties, after treatment, which may be driving the effect. In other words, there are systematic changes in the socio-demographic characteristics of the population in each county, such as widespread migration of *African Americans* to larger cities. These changes may be the result of any number of social factors: ranging from gentrification (Glick 2008) to the economic depression and bottoming out of financial markets in 2002 (as a delayed effect of the dot com bubble burst).

To eliminate this potential confound we estimate an exposure model where we control for the population of the county in which the hospital is located, in addition to our hospital and time fixed effects. In effect, this measure captures the change in the population, as well as the individual subpopulations, over time. Prior to discussing results, we point out several potential complications with this analysis. First, because estimates of county population (both aggregate and by subpopulation) are not available at the quarter level we use the annual population level for each estimate. Second, because estimates of the population of each county which is receiving *Medicaid* at a specific point in time are unavailable (this information is aggregated to the state-year level by the Centers for Medicare & Medicaid Services) we proxy the population of the county on *Medicaid* with the population of the county living below the poverty line. As the primary exclusion criterion for *Medicaid* is income, this assumption is reasonable and the measures should be strongly correlated. The source of these data is the Area Resource File provided by the Health Resource and Services Administration (HRSA), and the Florida Department of Health.

Table 8 shows that results remain consistent in this analysis: there is a strong and significant effect of *Craigslist* entry on the asymptomatic HIV rate. Both *Genders* are affected (men and women) and the effect exists across the socio-economic spectrum (both *Medicaid* and *non-Medicaid* patients are affected, with the effect on *Medicaid* being significantly larger). Moreover, the primary *racial* effect occurs within the *African American* community (with positive but insignificant coefficients within the *Caucasian* and *Latino* communities).

Coarsened Exact Match

An implicit assumption in the main analysis is that the control group, i.e., the population of untreated hospitals in the state of Florida, is a representative counterfactual for the *Craigslist* treated hospitals. To the extent that this assumption may be challenged because different subpopulations may select into treatment at different hospitals (Shahian et al. 2012), we conduct a coarsened exact matching procedure (Iacus et al. 2012) at the hospital level. Following Marx et al. (2009) we construct a synthetic control group, i.e. one that is more homogenous based on observable characteristics, for the treated *Craigslist* hospitals. This procedure mitigates bias introduced into the estimates by limiting *ex-ante* differences between treated and untreated groups (Iacus et al. 2012, Marx et al. 2009). The control group is constructed based on a standard set of hospital-level variables (Chen et al. 1999): the number of licensed beds within the hospital (size), the number of patients the hospital treated during the period (throughput), the for profit status of the hospital, and the period of observation.

Results from the coarsened exact match (Table 9) corroborate previous estimations. As seen in Column 1, the implementation of *Craigslist* into a county results in an increase in the total number of asymptomatic HIV patients being admitted. Furthermore, there are consistent effects across *Gender* (significant but indistinguishable effects on men and women) and SES (with a larger effect accruing to the *Medicaid* population). Finally, we see that the largest effect is experienced within the *African American* community (although we see an effect within the *Latino* community, as also in the relative time model).

Exclusion of Untreated Hospitals

While the coarsened exact matching procedure reduces the possibility that there is a difference in the selection of individuals into hospitals based on observable characteristics, and the exposure model suggests that the size of the relevant subpopulations is not driving the effect, it is possible that there is variation in the relative level of health between treated and untreated counties. Although our analysis excludes all hospitals who never treat an HIV patient, thereby plausibly eliminating communities where there is no HIV to spread, it is possible that untreated counties lack the necessary infection rate of HIV to experience an effect post *Craigslist* entry. To the extent that this may introduce heterogeneity into treated and untreated groups we next execute our estimations using only hospitals which eventually receive

Craigslist treatment. In effect, this model allows us to compare the relative time trends in HIV between hospitals which receive *Craigslist* treatment earlier and those which receive treatment later. Results from this analysis (Table 10) corroborate previous estimations and, interestingly, also show a significant effect within the *Latino* community.

Heterogeneity in Treated and Untreated Time Trends

Thus far findings are stable across multiple specifications and falsification tests: we see that that neither changes in the population of treated counties nor bias in the selection of hospital are influencing the main results. Nevertheless, it is still possible that there is unobserved heterogeneity in the time trends between treated and untreated counties which our previous analyses have not revealed. To mitigate this concern, we replicate our estimations using independent time fixed effects for treated and untreated counties. To the extent that this analysis allows for differences in the asymptomatic HIV trend across the treated and untreated groups, it should reinforce the claim that the hospital fixed effects have effectively controlled for *ex ante* heterogeneity in the groups. Results are available in Table 11 and remain consistent with the largest effect accruing to *Medicaid* patients and *African Americans*. Furthermore, consistent with the relative time model and the coarsened exact match, we see a small increase of 0.29 patients per hospital quarter in the *Latino* community.

Marginal Probability of Diagnosis

One further plausible explanation of the change in incidence rate is that although there is a net increase in the number of diagnoses, the *marginal* effect, per patient, is not changing. As a result, very small changes in the population of each county, which are statistically unidentifiable, may be driving the effect. To the degree that HIV affects a small percentage of the American population (1.1mm of the 295.5mm population in 2005[25]), we must ensure that the entry of *Craigslist* is not simply influencing the number of diagnoses, but the marginal likelihood that the patient is a carrier as well; which should be independent of the population changes pre and post-*Craigslist* entry. We therefore conduct a bed level investigation to determine the increase in the marginal probability that a patient who is admitted to a *Craigslist*-treated

[25] http://www.cdc.gov/mmwr/preview/mmwrhtml/su6001a19.htm

hospital is a carrier of the HIV virus. The dependent variable for this analysis is dichotomous and indicates whether or not the focal patient is an asymptomatic HIV carrier.

We first conduct the bed level analysis using a logit model. However, because of two well established concerns with logit models, we further use a linear probability model to ensure robustness of the results. First, as noted by King and Zeng (2001), the rarity of HIV infection (less than 0.5% of the patients in our sample) can lead to a biased estimation of the standard errors. To resolve this issue we utilize robust Huber-White standard errors clustered at the county level. Second, as noted by many researchers (Ai and Norton 2003, Hoetker 2007, Zelner 2009), the interpretation of interaction terms in logit models, i.e. the interaction between the *Craigslist* treatment variable and the patient specific subpopulation indicators, is complex and requires the simulation of the marginal effects post estimation because the effect of changes in the independent variable of interest is dependent upon the values of other covariates in the model. As the statistical tools which have been designed both by Zelner (2009) and Ai and Norton (2003) have been developed for only a single interaction term to be analyzed (thereby ignoring concomitant changes in the other interaction terms), the use of interaction terms for a bed level analysis will lead to a biased estimate of the marginal effect. We therefore conduct our initial bed level analysis on the aggregate population, and repeat it for each of the subpopulations individually. Results are in Table 12. As seen in the main results, we find a significant effect across *SES* and *Gender*[26] and for African Americans, with non-significant coefficients for the *Latino* and *Caucasian* populations.

As a result of challenges with interaction terms logit models, we further estimate the marginal increase in the likelihood of HIV diagnosis after treatment with *Craigslist* using a linear probability model (LPM). Although the LPM allows for a meaningful interpretation of the interaction terms between *Craigslist* and each of the different indicators of patient characteristics (i.e. which subpopulations the focal patient belongs to), it is not without flaws. First, it can introduce heteroscedasticity into the estimates; a concern that we address by using heteroscedastic-consistent Huber-White standard errors

---

[26] Because the marginal effect of the coefficients will depend on the values of the other covariates in the model directly comparing coefficients across the model is not appropriate.

clustered by county. Second, the LPM can produce estimates which exist outside the [0..1] bounds of the model. A post estimation inspection shows that predicted values remain with the interval. *Caucasians* serve as the base case for *Race*, *non-Medicaid* patients serve as the base case for *SES*, and women as the base case for *Gender*.

Results of the LPM (Table 13) add interesting nuance to the main findings. While our previous analyses had suggested that there is no discernible difference in the effect across *Gender*, these results indicate there is a slightly larger effect for men than for women, However, to the extent that this coefficient is modestly significant ($p<0.05$) in a large sample (nearly 12 million observations (Lin et al. 2013)), the results should be interpreted with caution. Furthermore, results suggest that despite the finding that *Medicaid* patients suffer a disproportionate penalty in terms of number of admittances, the increase in the marginal probability of diagnosis across SES is negligible once other factors are simultaneously accounted for. Finally, the strikingly consistent finding across all estimations is also evident in the LPM: results indicate that a large penalty accrues to the *African American* population[27].

Change in Testing Behavior

Thus far our empirical strategy in the robustness checks has focused on eliminating unobserved heterogeneity between the treated and untreated groups. However, one further plausible alternate explanation for the change in the asymptomatic HIV incidence rate, post-*Craigslist* implementation, is that there is a change in the propensity for individuals to get tested for sexually transmitted diseases. For example, it is possible that individuals who habitually engage in risky behavior are using the website to solicit casual partners who are outside their standard pool of casual partners. If this is the case, they may worry about the possibility of contracting disease and choose to get tested for STDs after implementation at a higher rate, thereby increasing the number of HIV diagnoses without actually influencing the underlying incidence rate. To mitigate this alternate explanation we use data from the Behavioral Risk Factor Surveillance System (BRFSS) survey conducted by the CDC[28].

[27] Note that the low model fit (R^2) for the LPM is consistent with extant epidemiological research (Heinzl et al 2005)
[28] http://www.cdc.gov/brfss/

The Quinquennial BRFSS survey contains county level information on the percent of the population that reports having been tested for HIV in the previous year. We compare the propensity for individuals to be tested for HIV in 2002 and 2007 (the first year of our sample and the year immediately following our sample period) across the regions represented in our data. Results are available in Table 14. *Craigslist Treated* indicates counties which receive the *Craigslist* treatment during the course of our study and *Untreated* indicates counties which do not. We find no significant difference in the *ex-ante* or *ex-post* propensity for individuals to get tested in treated or untreated counties. Furthermore, there is no significant difference in the change in propensity to get tested for HIV between 2002 and 2007. These results further affirm the assertion that it is *Craigslist* that is causing the observed increase in HIV.

Entry Model

One further concern which is present in our data is exogeneity of the entry decision by *Craigslist*. Although our models, thus far, have indicated that there is no heterogeneity in the pre-treatment asymptomatic HIV trend it is possible that there is stable variation in the number of HIV patients across treated and untreated hospitals. To ensure that the entry of Craigslist is not systematically related to the incidence of HIV within the local area, we estimate an entry model to test if the number of HIV patients in a local area is positively correlated with entry. To the extent that a larger population of HIV positive carriers will magnify the negative effect of *Craigslist* entry (by increasing the likelihood of an encounter with an HIV carrier), it is important to rule out this threat. To conduct this analysis we execute a logit hazard specification (Singer and Willett 1993) with *Craigslist* entry as the dependent variable. Independent variables include a broad range of local socio-economic conditions such as population, age, income, poverty level, and education, as well number of HIV cases diagnosed since 1990 (the earliest year of data availability), both in the aggregate and for each subpopulation examined in the main model. As with the exposure model, the additional data required for these estimations are retrieved from the Area Resource File. Results are available in Table 15.

As would be expected, we see that an increase in the number of college graduates in a county is correlated with an increased likelihood of *Craigslist* entry. Furthermore, an increase in the number of younger users (ages 20-44) is correlated with an increased likelihood of entry. However, results indicate that the number of previously diagnosed HIV cases at the focal hospital, both in the aggregate as well as in the individual subpopulations, is not positively influencing the likelihood of entry by *Craigslist*.

Random Implementation Model

As a final robustness check we estimate a random implementation model. Although our results have been consistent across a wide variety of specifications it is still possible that the increase in the HIV rate is related to an idiosyncrasy associated with *Craigslist* treated hospitals, and not as a result of *Craigslist* implementation. If this was true, the effect would be significant with any ordering of *Craigslist* introduction in the hospitals that were eventually treated. We conduct further analysis to determine how likely it is that a *random* implementation of *Craigslist* would yield an aggregate effect size which is comparable to our estimates[29]. We estimate the probability of randomly finding the aggregate effect in three ways. First, we randomly treat 843 hospital quarters (the number of treated hospital quarters in our data) and regress the total number of HIV patients, by hospital quarter, on this dummy variable (*pseudo-treated*) along with hospital and time fixed effects. The coefficient of the estimates for *pseudo-treated* is then stored. After replicating this random treatment 500 times we calculate the mean and standard deviation of the *pseudo-treated* coefficients. From this we calculate the Z-Score of the difference between our estimated coefficient (our original β estimate) and the mean of the randomly calculated coefficients.

After completing the purely random implementation model we replicate the procedure treating only the hospitals that eventually receive *Craigslist*. Finally, we randomly swap the vectors of time dummies between treated hospitals. To conduct this analysis we change the time of implementation between the hospitals in different counties at random. For example, in this analysis a hospital in Tampa Bay (originally treated in November of 2003), might instead receive a treatment date of October of 2002 (the time of treatment in Miami). The hospital in Miami then receives the treatment time of a hospital in Orlando (February of 2004). As with our previous random implementation models the model is replicated 500 times. Results are available in Table 16 and indicate, in each of the three models, that our estimated

[29] We thank an anonymous reviewer for this suggestion.

effect size is significantly larger than what would be expected purely by chance. To the extent that our randomly determined treatment β is statistically indistinguishable from zero this further supports the conclusion that there is no significant unaccounted for difference between treated and untreated hospitals; thereby strengthening our causal claim[30].

Implications and Conclusion

Our work was motivated by the acknowledged presence of digital inequalities among different populations, policy concerns about various types of digital disparities, and the growing role of the Internet in individual and public health. While the digital two-sided matching platforms that are now widely available on the Internet offer robust benefits in regard to facilitating social contact, they may also yield connections that pose risks. We examined the effects of a platform for the solicitation of casual sexual partners on the incidence rate of asymptomatic HIV in the local area. We asked: to whom do the negative effects of platform use accrue based on the socio-demographic characteristics of race/ethnicity, gender, and socio-economic status? Although the documented HIV incidence rate is highest among ethnic minorities and the socio-economic lower class (CDC 2011), the nature and extent of Internet usage among these groups is markedly different as a result of limited access, training, and computer literacy (Barry 2013, DiMaggio et al. 2004, Hargittai 2010, Warschauer 2004). Our empirical analysis on a census of nearly 12 million patients who are subjected to a natural experiment yields three main results that are robust to multiple specifications. First, the entry of Craigslist significantly increases the asymptomatic HIV incidence rate for residents of treated areas (both the focal cities and the surrounding suburban areas). Second, the absolute increase in the HIV incidence rate is significantly larger for the digitally disadvantaged (African Americans and the socio-economic lower class). Third, we find that both men and women experience equivalent penalties in regard to the increase in HIV after the entry of Craigslist, even though men have been identified as being at greater risk for HIV infection.

We compute the economic implications of the availability of sites which facilitate such behavior.

---

[30] This analysis has also been conducted on the subpopulations. Results (available upon request) from the random implementation of the treatment time are significant for all subpopulations except *Caucasians* and *Latinos* (thereby corroborating our previous estimates).

Annually, results indicate a public welfare loss of \$5.749 million per year as of 2006 in the State of Florida[31]. In the aggregate, results suggest that roughly 958 patients have been admitted to Florida hospitals over the 843 treated hospital quarters which otherwise would not have been admitted during the 5 year period of our study (74.59% of them *African American*). At a cost of \$618,000 over the patient's lifetime (Schackman et al. 2006), this translates to an additional financial burden of \$592 million in the State of Florida alone. We note that these estimates are conservative because the diagnosis rate during the asymptomatic phase of the disease is not 100% (Janssen et al. 1992). When compared with the estimates reported in Chan and Ghose (2013), ours are smaller (a 13.5% rise relative to their 15.9%). One possible explanation for the difference in the estimates is the lack of differentiation by Chan and Ghose (2013), empirically, between symptomatic and asymptomatic HIV.

This study further contributes to research on digital inequality (DiMaggio et al. 2004, Warschauer 2004) that has historically focused on the negative downstream implications of decreased information access and connectivity, with limited attention to the negative implications of Internet access and literacy. Even within the health policy literature, discussions have been confined to the undesirable effect of limited Internet access on the dissemination of health information (Brodie et al. 2000). Our results suggest that despite the decreased access which traditionally at risk groups have to online resources, penalties continue to accrue to them disproportionately. To the extent that *African Americans* and *lower SES* subpopulations have also been widely documented as experiencing health disparities (CDC 2005), we find that Craigslist exacerbates these differences. While limited access to the resources of the Internet may reduce social welfare in most instances, there are situations where increased access can diminish welfare by promoting risky behavior. Our study highlights the importance of further investigation into the differential effects of digital inequality for both the digitally disadvantaged and the digitally advantaged; and when the beneficial and punitive effects of increased Internet access accrue.

A further implication of results relates to policies to address digital inequalities. A common

[31] This calculation is derived from the 125 treated hospitals at the conclusion of the sample using the aggregate increase of 1.136 patients per treated hospital quarter and an annual cost of \$10,121.60 (Chesson et al, 2004).

belief among policy makers is that investment in IT hardware and widespread, affordable broadband access is a solution to the digital divide (Hogg 2012). Our results provide cautionary evidence against this approach. While investments in hardware and connectivity are necessary to resolve these persistent inequalities, our results suggest that investment in training is equally important. There may be reason to worry that digitally disadvantaged groups are likely to intensify the disproportionate penalty they suffer as a result of leveraging IT for welfare diminishing activities, as opposed to the welfare enhancing activities which might mitigate inequality.

Findings from this study raise additional intriguing questions that offer rich opportunities for future work. First, given the relative digital disadvantages African Americans and the socio-economic lower class face, why is the incidence rate increase for these groups so large? Three likely reasons for the large effect can be identified. One, a common misconception of the digital divide is that it is binary (Warschauer 2004), with individuals either having, or not having, access to online resources. In contrast to this view, digital disparity is more often a continuum; ranging from no access, to usage of public access points, to unskilled and then skilled exploitation of online resources in the home (DiMaggio et al. 2004, Hargittai 2010). Indeed, the popular media documents examples of even homeless individuals using online platforms like Craigslist through public access points (Abel 2009, Redmond 2013). Two, although research suggests that the digitally disadvantaged rarely utilize online resources for welfare enhancing activities (Hargittai 2010, Zillien and Hargittai 2009), this does not imply that there is no utilization. Because the use of platforms like Craigslist requires only rudimentary digital ability (i.e. search and email), it is possible that the disadvantaged are not precluded on the basis of technical competence. Finally, research suggests that, in contrast to differences in tethered access, there is parity in the utilization of mobile broadband across socio-economic classes (Prieger 2013) and that disadvantaged users primarily access online resources using mobile technology (Jacobs 2012).

A second interesting question raised by our results relates to the small and intermittent effect observed in the Latino population. Recall that while our main estimation did not show a significant effect for *Latinos*, we found a small effect in the relative time model and in a few of the robustness checks. To the extent that Latinos are considered to be a disadvantaged population which is disproportionately affected by HIV (CDC 2011), this begs the question why African Americans and the Medicaid population are affected to such a degree, and Latinos are not. While our methodology does not allow us to determine the exact mechanism, several possibilities exist. The first is language barriers (Barry 2013). As 34.8 million Hispanic households (65% of the population) speak Spanish as the primary language in the home (Gonzalez-Barrera and Lopez 2013), and the American Craigslist interface is written in English, it is possible that language constraints prevent widespread use of the platform among the Latino community. A second possibility is limited detection and seeking of treatment. As Florida is home to more than 1 million undocumented Latinos during the time of our investigation[32], it may be the case that these individuals are not seeking treatment out of fear of deportation.

Our finding that men and women are equally affected by Craigslist is a cause for concern. Studies on gender differences in Internet use show that men are more likely to use the Internet (Hargittai 2010), although this gap is narrowing with the increased availability and use of technology among youth. Research also suggests that women tend to use the Internet to a greater extent for social relationships than instrumental transactions such as banking as compared to men (Weiser 2000). The fact that both genders significantly increase their probability of contracting HIV upon the availability of a matching platform for casual encounters is another instance of a potentially undesirable "equalizing" effect of the Internet.

We acknowledge the limitations of this work which future research can address. First, we cannot observe the Craigslist utilization rates for visitors to treated areas (from either inside or outside the state of Florida). Econometrically, this is of limited concern as it will bias results downward, making estimates more conservative (because these visitors will likely be diagnosed at the untreated hospitals near their homes thereby introducing parity into the increase in the HIV rate across the two groups). However, from the perspective of epidemiological research, further work investigating the effect of platforms which facilitate the transmission of disease across geographical regions is needed. Second, our data does not

---

[32] http://www.pewhispanic.org/2011/02/01/unauthorized-immigrant-population-brnational-and-state-trends-2010/

allow us to observe patient re-admittance. Although Florida is experiencing a decreased HIV incidence rate during the study period, and advances in medical treatment are causing patient re-admittance to slow, this is clearly a data limitation. Third, the use of *Medicaid* as a proxy for low SES can be improved upon. However, we note that because of privacy and confidentiality concerns, detailed income data at the individual level that is tied to medical records is difficult to obtain. Fourth, as Craigslist offers many opportunities for couples to meet and interact, our results cannot ensure that the increase in HIV incidence is exclusively a result of casual sexual encounters. It is plausible, for example, that long term dating enabled by the site also plays a role in the increase in HIV prevalence. Fifth, our data do not allow us to perfectly track population migration over time, as the composition of the population within subpopulations may be changing. Although our empirical results from the exposure model suggest that changes in each of the subpopulation levels is not influencing the effect and there is no concomitant spike in other conditions, we are unable to completely rule this explanation out. Sixth, the data do not allow us to observe the socio-demographic characteristics of the individual who passed the HIV virus to the patient. One future, important, extension of this work will be to investigate how the focal individual's selection of sexual partner on matching platforms influences the spread of disease. Finally, although the economic implications of the HIV externality are large, it is important to note that we have made no attempt in this work to quantify the positive effects of Craigslist. To the extent that significant public welfare may be generated by these platforms, it would be inappropriate to draw any conclusions regarding the net effect of Craigslist as a whole, whether positive or negative, on public welfare from our results.

References

Abel, D. 2009. *Panhandlers Move from Street to Internet*. The Boston Globe, Boston, MA.

Agarwal, R., Gao, G., DesRoches, C., Jha, A.K. 2010. Research Commentary—the Digital Transformation of Healthcare: Current Status and the Road Ahead. *INFORMATION SYSTEMS RESEARCH*. 21(4) 796-809.

Ai, C., Norton, E.C. 2003. Interaction Terms in Logit and Probit Models. *Economics Letters*. 80(1) 123-129.

APA. 2013. Socioeconomic Status, August 8, 2013, http://www.apa.org/topics/socioeconomic-status/.

Bakos, Y., Bailey, J. 1997. An Exploratory Study of the Emerging Role of Electronic Intermediaries. *International Journal of Electronic Commerce*. 1(3) 7-20.

Banerjee, A.V., Duflo, E. 2000. Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry. *The Quarterly Journal of Economics*. 115(3) 989-1017.

Bapna, R., Ramaprasad, J., Shmueli, G., Umyarov, A. 2012. *One-Way Mirrors in Online Dating: A Randomized Field Experiment*. Workshop on Information Systems Economics, Orlando, FL.

Barry, J.J. 2013. Crossing the Digital Divide: Is Access to the Internet an Economic Right? *Available at SSRN 2315307*.

Benotsch, E.G., Kalichman, S., 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) 177-183.

Bingham, T.A., Harawa, N.T., Johnson, D.F., Secura, G.M., MacKellar, D.A., Valleroy, L.A. 2003. The Effect of Partner Characteristics on Hiv Infection among African American Men Who Have Sex with Men in the Young Men's Survey, Los Angeles, 1999–2000. *AIDS Education and Prevention*. 15(1 Supplement) 39-52.

Bozzette, S.A., Berry, S.H., Duan, N., Frankel, M.R., Leibowitz, A.A., Lefkowitz, D., Emmons, C.-A., Senterfitt, J.W., Berk, M.L., Morton, S.C. 1998. The Care of Hiv-Infected Adults in the United States. *New England Journal of Medicine*. 339(26) 1897-1904.

Brodie, M., Flournoy, R.E., Altman, D.E., Blendon, R.J., Benson, J.M., Rosenbaum, M.D. 2000. Health Information, the Internet, and the Digital Divide. *Health Affairs*. 19(6) 255-265.

Brown, P. 1988. *The Body and Society: Men, Women, and Sexual Renunciation in Early Christianity*. Columbia University Press, New York, NY.

Brynjolfsson, E., Smith, M.D. 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science*. 46(4) 563-585.

Burke, M.A., Fournier, G.M., Prasad, K. 2003. Physician Social Networks and Geographical Variation in Medical Care.

Burke, M.A., Fournier, G.M., Prasad, K. 2007. The Diffusion of a Medical Innovation: Is Success in the Stars? *Southern Economic Journal*. 73(3) 588-603.

CDC. 2005. *Health Disparities Experienced by Black or African Americans--United States*.U.S. Department of Health and Human Services Centers for Disease Control and Prevention

CDC. 2006. *Hiv/Aids Surveillance Report, 2006*.U.S. Department of Health and Human Services Centers for Disease Control and Prevention

CDC. 2008. *Revised Surveillance Case Definitions for Hiv Infection among Adults, Adolescents, and Children Aged <18 Months and for Hiv Infection and Aids among Children Aged 18 Months to <13 Years - United States, 2008*.U.S. Department of Health and Human Services Centers for Disease Control and Prevention

CDC. 2011. *Hiv Surveillance - United States, 1981--2008*.U.S. Department of Health and Human Services Centers for Disease Control and Prevention

CDC. 2013. *Leading Causes of Death*.U.S. Department of Health and Human Services Centers for Disease Control and Prevention

Chan, J., Ghose, A. 2013. Internet's Dirty Secret: Assessing the Impact of Online Intermediaries on the Outbreak of Sexually Transmitted Diseases. *MIS Quarterly (Forthcoming)*.

Chen, J., Radford, M.J., Wang, Y., Marciniak, T.A., Krumholz, H.M. 1999. Do “America's Best Hospitals” Perform Better for Acute Myocardial Infarction? *New England Journal of Medicine*. 340(4) 286-292.

Chesson, H.W., Blandford, J.M., Gift, T.L., Tao, G., 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) 11-19.

DHHS. 2012. Aids. *Medical Care*. 916 874-7720.

DiMaggio, P., Hargittai, E., Celeste, C., Shafer, S. 2004. Digital Inequality: From Unequal Access to Differentiated Use. *Social inequality* 355-400.

Dorland. 2011. *Dorland's Illustrated Medical Dictionary*. Elsevier - Health Sciences Division, Philadelphia, PA.

Dugas, A.F., Hsieh, Y.-H., Levin, S.R., Pines, J.M., Mareiniss, D.P., Mohareb, A., Gaydos, C.A., Perl, T.M., Rothman, R.E. 2012. Google Flu Trends: Correlation with Emergency Department Influenza Rates and Crowding Metrics. *Clinical infectious diseases*. 54(4) 463-469.
Elford, J., Bolding, G., Sherr, L. 2001. Seeking Sex on the Internet and Sexual Risk Behaviour among Gay Men Using London Gyms. *AIDS*. 15(11) 1409-1415.
Finzi, D., Hermankova, M., Pierson, T., Carruth, L.M., Buck, C., Chaisson, R.E., Quinn, T.C., Chadwick, K., Margolick, J., Brookmeyer, R., Gallant, J., Markowitz, M., Ho, D.D., Richman, D.D., Siliciano, R.F. 1997. Identification of a Reservoir for Hiv-1 in Patients on Highly Active Antiretroviral Therapy. *Science*. 278(5341) 1295-1300.
Floyd, F.J., Stein, T.S. 2002. Sexual Orientation Identity Formation among Gay, Lesbian, and Bisexual Youths: Multiple Patterns of Milestone Experiences. *Journal of Research on Adolescence*. 12(2) 167-191.
Garofalo, R., Herrick, A., Mustanski, B.S., 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) 1113-1117.
Glick, J. 2008. Gentrification and the Racialized Geography of Home Equity. *Urban Affairs Review*. 44(2) 280-295.
Goh, J., Gao, G., Agarwal, R. 2009. *Impact of on-Line Health Communities on Patient's with Rare Disease*. Annual Conference on Statistical Challenges in E-Commerce, Carnegie-Mellon University.
Gonzalez-Barrera, A., Lopez, H. 2013. *Spanish Is the Most Spoken Non-English Language in U.S. Homes, Even among Non-Hispanics*Pew Research Center.
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) 1387-1392.
Hargittai, E. 2010. Digital Na (T) Ives? Variation in Internet Skills and Uses among Members of the "Net Generation". *Sociological Inquiry*. 80(1) 92-113.
HealthyPeople. 2013. *Recommendations for the Framework and Format of Healthy People 2020*.U.S. Department of Health and Human Services
Heinzl, H., Waldhör, T., Mittlböck, M. 2005. Careful Use of Pseudo R - Squared Measures in Epidemiological Studies. *Statistics in medicine*. 24(18) 2867-2872.
Hoetker, G. 2007. The Use of Logit and Probit Models in Strategic Management Research: Critical Issues. *Strategic Management Journal*. 28(4) 331-343.
Hogg, C. 2012. Bridging the Digital Divide
Holmberg, S.D., Moorman, A.C., Greenberg, A.E., Committee, D.S. 2004. Trends in Rates of Myocardial Infarction among Patients with Hiv. *New England Journal of Medicine*. 350(7) 730-732.
Hsieh, J.P.-A., Rai, A., Keil, M. 2011. Addressing Digital Inequality for the Socioeconomically Disadvantaged through Government Initiatives: Forms of Capital That Affect Ict Utilization. *INFORMATION SYSTEMS RESEARCH*. 22(2) 233-253.
Iacus, S.M., King, G., Porro, G. 2012. Causal Inference without Balance Checking: Coarsened Exact Matching. *Political analysis*. 20(1) 1-24.
Jacobs, K. 2012. *The People's Pornography: Sex and Surveillance on the Chinese Internet*. Intellect, Chicago, IL.
Janssen, R.S., St. Louis, M.E., Satten, G.A., Critchley, S.E., Petersen, L.R., Stafford, R.S., Ward, J.W., Hanson, D.L., Olivo, N., Schable, C.A., Dondero, T.J. 1992. Hiv Infection among Patients in U.S. Acute Care Hospitals. *New England Journal of Medicine*. 327(7) 445-452.
Kim, A.A., Kent, C., McFarland, W., Klausner, J.D. 2001. Cruising on the Internet Highway. *JAIDS Journal of Acquired Immune Deficiency Syndromes*. 28(1) 89-93.
King, G., Zeng, L. 2001. Explaining Rare Events in International Relations. *International Organization*. 55(3) 693.
Lin, M., Lucas Jr, H.C., Shmueli, G. 2013. Research Commentary-Too Big to Fail: Large Samples and the P-Value Problem. *INFORMATION SYSTEMS RESEARCH*. 24(4) 906-917.

MacDonald, K. 1995. The Establishment and Maintenance of Socially Imposed Monogamy in Western Europe. *Politics and the Life Sciences* 3-23.
Malone, T., Yates, J., Benjamin, R. 1987. Electronic Markets and Electronic Hierarchies. *Communications of the ACM*. 30(6) 13.
Marx, M., Strumsky, D., Fleming, L. 2009. Mobility, Skills, and the Michigan Non-Compete Experiment. *Management Science*. 55(6) 875-889.
NTIA, ESA. 2013. *Exploring the Digital Nation: America's Emerging Online Experience*.U.D.o. Commerce
Padgett, P.M. 2007. Personal Safety and Sexual Safety for Women Using Online Personal Ads. *Sexuality Research & Social Policy*. 4(2) 27-37.
Palella, F.J.J., Baker, R.K., Moorman, A.C., Chmiel, J.S., Wood, K.C., Brooks, J.T., Holmberg, S.D., Investigators, H.O.S. 2006. Mortality in the Highly Active Antiretroviral Therapy Era: Changing Causes of Death and Disease in the Hiv Outpatient Study. *JAIDS Journal of Acquired Immune Deficiency Syndromes*. 43(1).
Pantaleo, G., Graziosi, C., Fauci, A.S. 1993. The Immunopathogenesis of Human Immunodeficiency Virus Infection. *New England Journal of Medicine*. 328(5) 327-335.
Parker, G.G., Van Alstyne, M.W. 2005. Two-Sided Network Effects: A Theory of Information Product Design. *Management Science*. 51(10) 1494-1504.
Parsons, J.T., Tesoriero, J.M., Carballo-Dieguez, A., Hirshfield, S., Remien, R.H. 2006. Hiv Behavioral Research Online. *Journal of Urban Health*. 83(1) 73-85.
Prieger, J. 2013. *The Broadband Digital Divide and the Benefits of Mobile Broadband for Minorities*. Pepperdine Digital Commons.
Redmond, L. 2013. *Man Booked on Charge of Offering Sex at Library in Tewksbury*. The Lowell Sun, Lowell, MA.
Rogers, M.K. 2010. *The Psyche of Cybercriminals: A Psycho-Social Perspective*. Springer.
Schackman, B.R., Gebo, K.A., Walensky, R.P., Losina, E., Muccio, T., Sax, P.E., Weinstein, M.C., Seage III, G.R., Moore, R.D., Freedberg, K.A. 2006. The Lifetime Cost of Current Hiv Care in the United States. *Medical Care*. 44(11) 990-997.
Seamans, R., Zhu, F. 2011. *Technology Shocks in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers*. NYU Working paper.
Shahian, D.M., Nordberg, P., Meyer, G.S., Blanchfield, B.B., Mort, E.A., Torchiana, D.F., Normand, S.-L.T. 2012. Contemporary Performance of Us Teaching and Nonteaching Hospitals. *Academic Medicine*. 87(6) 701-708.
Singer, J.D., Willett, J.B. 1993. It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events. *Journal of Educational and Behavioral Statistics*. 18(2) 155-195.
Warschauer, M. 2004. *Technology and Social Inclusion: Rethinking the Digital Divide*. MIT press.
Weiser, E.B. 2000. Gender Differences in Internet Use Patterns and Internet Application Preferences: A Two-Sample Comparison. *CyberPsychology and Behavior*. 3(2) 167-178.
Williamson, O.E. 1981. The Economics of Organization: The Transaction Cost Approach. *American Journal of Sociology*. 87(3) 548.
Zelner, B.A. 2009. Using Simulation to Interpret Results from Logit, Probit, and Other Nonlinear Models. *Strategic Management Journal*. 30(12) 1335-1348.
Zenilman, J.M., Ellish, N., Fresia, A., Glass, G. 1999. The Geography of Sexual Partnerships in Baltimore: Applications of Core Theory Dynamics Using a Geographic Information System. *Sexually Transmitted Diseases*. 26(2) 75-81.
Zickuhr, K., Smith, A. 2012. Digital Differences. *Pew Internet & American Life Project*.
Zillien, N., Hargittai, E. 2009. Digital Distinction: Status ‐ Specific Types of Internet Usage*. *Social Science Quarterly*. 90(2) 274-291.

Table 1: Florida Cities (Counties) and Craigslist Implementation Date

City	County	Implementation Date
Fort Lauderdale	Broward	Jun-06
Daytona Beach	Volusia	Jun-06
Florida Keys	Monroe	-
Fort Meyers	Lee	Jun-05
Gainesville	Alachua	Jan-06
Jacksonville	Duval	Jan-05
Lakeland	Polk	-
Miami	Miami-Dade	Oct-02
Ocala	Marion	-
Okaloosa / Walton	Okaloosa / Walton	-
Orlando	Orange	Feb-04
Palm Beach	Palm Beach	Apr-05
Panama City	Bay	-
Pensacola	Escambia	Sep-05
Sarasota	Sarasota	Jun-06
Space Coast	Brevard	-
St Augustine	St Johns	-
Tallahassee	Leon	Jun-05
Tampa Bay	Hillsborough	Nov-03
Treasure Coast	Indian River	-

Table 2: Summary Statistics
N - 4349

	Variable	Mean	Std. Dev.	1	2	3	4	5	6	7	8
1	Total Count	10.58151	22.76154								
2	Craigslist	0.1938377	0.3953487	0.2083							
3	Caucasian Count	3.313405	5.636037	0.8170	0.1237						
4	African American Count	5.757186	15.31777	0.9704	0.1774	0.6950					
5	Latino Count	1.2596	4.073043	0.7370	0.3051	0.5030	0.6367				
6	Medicaid Count	3.63141	9.759341	0.9567	0.1849	0.6878	0.9609	0.7170			
7	Non Medicaid Count	6.950103	13.72202	0.9783	0.2140	0.8660	0.9263	0.7126	0.8757		
8	Male Count	6.078639	12.50318	0.9710	0.2226	0.8603	0.9063	0.7572	0.8865	0.9802	
9	Female Count	4.502874	11.03327	0.9626	0.1775	0.7105	0.9750	0.6624	0.9691	0.9075	0.8700

Table 3: Negative Binomial Estimates of Total Number of Asymptomatic HIV Cases By Group
Asymptomatic HIV Incidence – 2002 - 2006

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	Total	Total Caucasian	Total African American	Total Latino	Total Non-Medicaid	Total Medicaid	Total Male	Total Female
Craigslist	0.126***	0.0781	0.146***	0.158	0.120**	0.130***	0.120*	0.134***
	(0.0349)	(0.0509)	(0.0327)	(0.0968)	(0.0427)	(0.0300)	(0.0477)	(0.0320)
Constant	4.012***	2.263***	3.793***	-0.241*	3.401***	3.229***	3.163***	3.449***
	(0.0278)	(0.0426)	(0.0476)	(0.116)	(0.0400)	(0.0434)	(0.0438)	(0.0303)
Hospital Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log Pseudolikelihood	-9085.13	-6598.22	-6624.08	-3480.73	-8080.41	-5819.21	-7621.38	-6562.2
Observations	4,349	4,349	4,349	4,349	4,349	4,349	4,349	4,349

Robust standard errors in parentheses (Clustered on County)
*** p<0.001, ** p<0.01, * p<0.05

Table 4: Negative Binomial Estimates of Total Number of Asymptomatic HIV Cases By Ethnicity Grouped by Gender and SES

	(1)	(2)	(3)	(4)
Dependent Variable	African American Female	African American Male	African American Medicaid	African American Non Medicaid
Craigslist	0.134***	0.139**	0.149***	0.122**
	(0.0299)	(0.0472)	(0.0418)	(0.0416)
Constant	3.260***	2.912***	3.108***	3.099***
	(0.0413)	(0.0695)	(0.0666)	(0.0487)
Hospital Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Log Pseudolikelihood	-5055.61	-5052.63	-5640.31	-4335.66
Observations	4,349	4,349	4,349	4,349

Robust standard errors in parentheses (Clustered on County)
*** p<0.001, ** p<0.01, * p<0.05

Table 5: Relative Time Negative Binomial Estimates of Number of Patients Admitted with Asymptomatic HIV Quarter Level Relative Time Fixed Effects Further Than One Year from Implementation Date Omitted

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	Total	Total Caucasian	Total African American	Total Latino	Total Non-Medicaid	Total Medicaid	Total Male	Total Female
Craigslist $_{t-4}$	-0.0164	0.0229	-0.0591	0.0498	-0.0240	-0.0123	-0.0305	-0.00849
	(0.0415)	(0.0395)	(0.0682)	(0.0777)	(0.0488)	(0.0450)	(0.0538)	(0.0335)
Craigslist $_{t-3}$	-0.0507	-0.0338	-0.0547	-0.0627	-0.0397	-0.0606	-0.0762**	-0.0175
	(0.0316)	(0.0567)	(0.0369)	(0.127)	(0.0362)	(0.0483)	(0.0254)	(0.0588)
Craigslist $_{t-2}$	-0.0318	-0.0758	0.0134	0.0163	-0.0287	-0.00499	-0.0543	0.0250
	(0.0477)	(0.0600)	(0.0614)	(0.101)	(0.0456)	(0.0500)	(0.0477)	(0.0510)
Craigslist $_{t-1}$	0.0364	0.0502	0.0283	0.0410	0.0392	0.0359	0.0590	0.00845
	(0.0381)	(0.0460)	(0.0341)	(0.124)	(0.0465)	(0.0454)	(0.0407)	(0.0496)
Craigslist $_{t0}$				Omitted Group				
Craigslist $_{t+1}$	-0.0121	-0.0610	-0.0194	0.172	0.0132	-0.0265	0.00181	0.00392
	(0.0364)	(0.0422)	(0.0536)	(0.0880)	(0.0447)	(0.0299)	(0.0498)	(0.0494)
Craigslist $_{t+2}$	0.104	0.0857	0.119*	-0.00876	0.0547	0.183**	0.0692	0.151**
	(0.0551)	(0.0707)	(0.0560)	(0.105)	(0.0587)	(0.0618)	(0.0684)	(0.0487)
Craigslist $_{t+3}$	0.114***	0.0275	0.121***	0.114	0.112**	0.106*	0.113***	0.0911*
	(0.0299)	(0.0396)	(0.0269)	(0.150)	(0.0347)	(0.0465)	(0.0335)	(0.0416)
Craigslist $_{t+4}$	0.163***	0.146*	0.150***	0.296***	0.118**	0.238***	0.137**	0.187***
	(0.0286)	(0.0569)	(0.0395)	(0.0560)	(0.0374)	(0.0305)	(0.0519)	(0.0447)
Constant	4.015***	2.261***	3.801***	-0.293**	3.405***	3.225***	3.168***	3.446***
	(0.0311)	(0.0406)	(0.0425)	(0.0967)	(0.0440)	(0.0488)	(0.0472)	(0.0350)
Hospital Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log Pseudolikelihood	-9082.06	-6594.32	-6622.78	-3476.80	-8080.30	-5811.76	-7618.18	-6560.03
Observations	4,349	4,349	4,349	4,349	4,349	4,349	4,349	4,349

Robust standard errors in parentheses (Clustered on County)
*** p<0.001, ** p<0.01, * p<0.05

Table 6: Negative Binomial Estimates of Introduction of Craigslist at Different Geographic Radii
Craigslist (City) indicates introduction of platform to the focal city
Craigslist (Greater County) indicates introduction of platform to the county and all bordering counties

Dependent Variable	(1) Total	(2) Total	(3) Total
Treated City	0.143***		
	(0.0295)		
Craigslist		0.126***	
		(0.0349)	
Treated Greater County			-0.0484
			(0.0449)
Constant	3.986***	4.012***	3.998***
	(0.0283)	(0.0278)	(0.0349)
Hospital Fixed Effects	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes
Log Pseudolikelihood	-9087.41	-9085.13	-9095.56
Observations	4,349	4,349	4,349

Robust standard errors in parentheses (Clustered on County)
*** p<0.001, ** p<0.01, * p<0.05

Table 7: Negative Binomial Estimates of Effect of Craigslist Introduction Patients Being Affected by Other Conditions

Dependent Variable	(1) AMI	(2) Stroke	(3) Melanoma	(4) Lung Cancer	(5) Symptomatic HIV
Craigslist	0.00848	-0.0145	0.0416	0.0188	-0.0388
	(0.0266)	(0.0386)	(0.0712)	(0.0217)	(0.0332)
Constant	4.979***	4.753***	0.893***	4.352***	4.749***
	(0.0185)	(0.0219)	(0.118)	(0.0169)	(0.0196)
Hospital Fixed Effects	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes
Log Pseudolikelihood	-16282.17	-14047.86	-4521.84	-13438.72	-12364.11
Observations	4,349	4,349	4,349	4,349	4,349

Robust standard errors in parentheses (Clustered on County)
*** p<0.001, ** p<0.01, * p<0.05

Table 8: Exposure Model of Effect of Craigslist Introduction on Asymptomatic HIV (Negative Binomial)

Dependent Variable	(1) Total	(2) Total Caucasian	(3) Total African American	(4) Total Latino	(5) Total Medicaid	(6) Total Non-Medicaid	(7) Total Male	(8) Total Female
Craigslist	0.127***	0.0815	0.146***	0.155	0.131***	0.123**	0.128**	0.126***
	(0.0332)	(0.0495)	(0.0301)	(0.0982)	(0.0290)	(0.0397)	(0.0466)	(0.0261)
County Population	-2.81e-08							
	(6.59e-07)							
Caucasian Population		-1.83e-07						
		(8.55e-07)						
African American Population			1.25e-06					
			(1.29e-06)					
Latino Population				-6.80e-06**				
				(2.27e-06)				
Population in Poverty					5.40e-07			
					(9.16e-07)			
Population not in Poverty						-1.62e-07		
						(4.69e-07)		
Male Population							-7.86e-07	
							(1.33e-06)	
Female Population								8.50e-07
								(1.69e-06)
Constant	4.034***	2.357***	3.511***	-0.0253	3.173***	3.511***	3.460***	3.106***
	(0.521)	(0.440)	(0.309)	(0.116)	(0.104)	(0.324)	(0.509)	(0.685)
Hospital Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log Pseudolikelihood	-9085.12	-6598.18	-6623.45	-3473.41	-5819.06	-8080.3	-7621.05	-6561.87
Observations	4,349	4,349	4,349	4,349	4,349	4,349	4,349	4,349

Robust standard errors in parentheses (Clustered on County)
*** p<0.001, ** p<0.01, * p<0.05

Table 9: Negative Binomial Estimates of Total Number of Asymptomatic HIV Cases Using Coarsened Exact Matching of Hospitals

Dependent Variable	(1) Total	(2) Total Caucasian	(3) Total African American	(4) Total Latino	(5) Total Non-Medicaid	(6) Total Medicaid	(7) Total Male	(8) Total Female
Craigslist	0.189***	0.0768	0.183**	0.456*	0.157***	0.254***	0.179**	0.198***
	(0.0440)	(0.0741)	(0.0635)	(0.193)	(0.0460)	(0.0676)	(0.0549)	(0.0443)
Constant	3.828***	2.184***	3.596***	-0.158	3.235***	3.038***	3.030***	3.264***
	(0.0615)	(0.0951)	(0.0537)	(0.175)	(0.0691)	(0.0838)	(0.0849)	(0.0622)
Hospital Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log Pseudolikelihood	-4589.45	-3265.86	-3130.38	-1639.87	-4043.09	-2783.66	-3822.47	-3182.44
Observations	2,452	2,452	2,452	2,452	2,452	2,452	2,452	2,452

Robust standard errors in parentheses (Clustered on County)
*** p<0.001, ** p<0.01, * p<0.05

Table 10: Negative Binomial Estimates of Total Number of Asymptomatic HIV Cases Using Only Treated Counties

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	Total	Total Caucasian	Total African American	Total Latino	Total Non-Medicaid	Total Medicaid	Total Male	Total Female
Craigslist	0.123***	0.0765	0.131***	0.230**	0.114**	0.135***	0.116**	0.139***
	(0.0332)	(0.0596)	(0.0282)	(0.0810)	(0.0398)	(0.0259)	(0.0423)	(0.0271)
Constant	3.999***	2.278***	3.758***	-0.182	3.386***	3.233***	3.158***	3.438***
	(0.0292)	(0.0451)	(0.0518)	(0.121)	(0.0495)	(0.0488)	(0.0527)	(0.0339)
Hospital Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log Pseudolikelihood	-5237.69	-3679.29	-4173.53	-2393.83	-4710.79	-3539.87	-4456.15	-3930.49
Observations	2,162	2,162	2,162	2,162	2,162	2,162	2,162	2,162

Robust standard errors in parentheses (Clustered on County)
*** p<0.001, ** p<0.01, * p<0.05

Table 11: Negative Binomial Estimates of Total Number of Asymptomatic HIV Cases Using Independent Time Fixed Effects for Hospitals Which Receive Treatment and Those Which Never Receive Treatment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	Total	Total Caucasian	Total African American	Total Latino	Total Non-Medicaid	Total Medicaid	Total Male	Total Female
Craigslist	0.122***	0.0742	0.133***	0.230**	0.112**	0.138***	0.116**	0.140***
	(0.0328)	(0.0574)	(0.0272)	(0.0805)	(0.0390)	(0.0263)	(0.0418)	(0.0277)
Constant	4.276***	2.510***	4.029***	-0.0102	3.707***	3.347***	3.423***	3.653***
	(0.0668)	(0.0840)	(0.0633)	(0.0744)	(0.0703)	(0.0512)	(0.0664)	(0.0452)
Hospital Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Treated Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Untreated Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log Pseudolikelihood	-9075.79	-6589.29	-6615.55	-3461.24	-8069.23	-5812.54	-7611.61	-6555.67
Observations	2,452	2,452	2,452	2,452	2,452	2,452	2,452	2,452

Robust standard errors in parentheses (Clustered on County)
*** p<0.001, ** p<0.01, * p<0.05

Table 12: Logit Estimates of the Marginal Increase in Likelihood of Patient Infection with Asymptomatic HIV

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	0/1 HIV Carrier	0/1 HIV Carrier	0/1 HIV Carrier	0/1 HIV Carrier	0/1 HIV Carrier	0/1 HIV Carrier	0/1 HIV Carrier	0/1 HIV Carrier
Population Sample	Total	Total Caucasian	Total African American	Total Latino	Total Medicaid	Total Non-Medicaid	Total Male	Total Female
Craigslist	0.0988**	0.0816	0.121**	0.0962	0.0976***	0.1023**	0.0823*	0.123**
	(0.0315)	(0.0476)	(0.0392)	(0.0774)	(0.0241)	0.0325802	(0.0354)	(0.0403)
Constant	-4.794***	-5.818***	-4.256***	-5.499***	-4.3589***	-5.0324***	-4.811***	-4.789***
	(0.0342)	(0.0476)	(0.0429)	(0.0757)	(0.0406)	(0.0382)	(0.0385)	(0.0449)
Hospital Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log Pseudolikelihood	-272783.24	-98755.98	-126356.59	-35095.26	-83073.007	-186044.67	-150102.52	-119586.56
Observations	11,793,617	8,090,330	1,867,208	1,747,153	1,967,303	9,803,373	5,075,571	6,646,988

Robust standard errors in parentheses (Clustered on County)
*** p<0.001, ** p<0.01, * p<0.05

Table 13: LPM Estimates of the Marginal Increase in Likelihood of Patient Infection with Asymptomatic HIV

Dependent Variable	(1) 0/1 HIV Carrier	(2) 0/1 HIV Carrier	(3) 0/1 HIV Carrier
Craigslist	0.000533***	0.000527***	-0.000542
	(0.000103)	(9.14e-05)	(0.000498)
African American		0.00920***	0.00836***
		(0.000959)	(0.000915)
Latino		-0.000638	-0.000212
		(0.000652)	(0.000420)
Gender		0.00240***	0.00205***
		(0.000387)	(0.000338)
Medicaid		0.00272***	0.00272***
		(0.000477)	(0.000560)
African American			0.00304*
* Craigslist			(0.00126)
Latino			-0.000571
* Craigslist			(0.000326)
Gender			0.00139*
* Craigslist			(0.000583)
Medicaid			-1.09e-07
* Craigslist			(0.000718)
Constant	0.00921***	0.00271**	0.00290***
	(0.000116)	(0.000799)	(0.000752)
Hospital Fixed Effects	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes
Observations	11,793,617	11,793,617	11,793,617
R-squared	0.005	0.009	0.009

Robust standard errors in parentheses (Clustered on County)
*** p<0.001, ** p<0.01, * p<0.05

Table 14: Two Sample T-Test of Population Propensity to be Tested for HIV Before and After Treatment
T-Value indicates the outcome of the two sample T-Test comparing treated and untreated counties

Period		2002	2007	Δ From 2002 - 2007
Craigslist Treated	μ	21.3583333	21.275	-0.083
	σ	5.51979221	5.77252505	2.2993
	N	12	12	12
Untreated	μ	20.6714286	19.0303571	-1.641
	σ	4.82462299	6.22893307	6.2397
	N	56	56	56
	t-value	0.4015	1.2083	1.46

Table 15: Outcome of Logit Hazard Model for the Entry of Craigslist
Diagnosed Cases to Date Indicate the Number of Asymptomatic Cases by Subpopulation Diagnosed at Hospital *j* since 1990 (First Date of Data Availability)

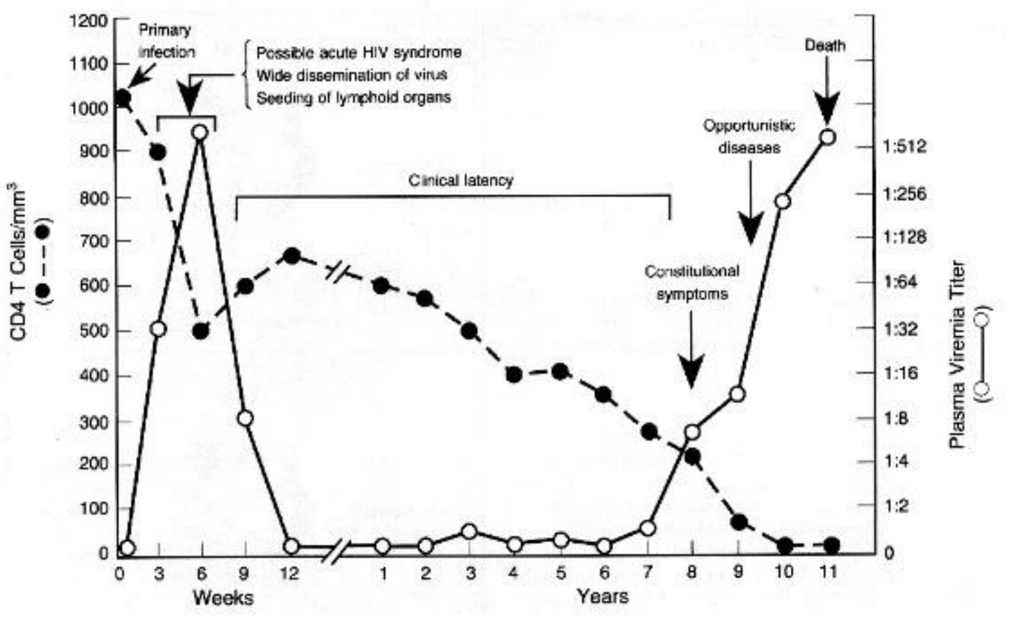
Dependent Variable	(1) Craigslist Entry	(2) Craigslist Entry	(3) Craigslist Entry	(4) Craigslist Entry	(5) Craigslist Entry	(6) Craigslist Entry
Population	-4.61e-06	-4.48e-06	-5.42e-06	-4.55e-06	-4.43e-06	-5.44e-06
	(5.96e-06)	(5.96e-06)	(6.04e-06)	(5.97e-06)	(5.96e-06)	(6.04e-06)
African American Population	-1.64e-05**	-1.70e-05**	-1.78e-05**	-1.70e-05**	-1.68e-05**	-1.85e-05**
	(6.07e-06)	(6.14e-06)	(6.31e-06)	(6.14e-06)	(6.12e-06)	(6.52e-06)
Population in Poverty	-8.75e-06	-8.35e-06	-9.15e-06	-8.24e-06	-8.49e-06	-9.04e-06
	(6.78e-06)	(6.82e-06)	(6.95e-06)	(6.83e-06)	(6.81e-06)	(7.01e-06)
Population Age 20-44 (0000s)	0.455***	0.456***	0.472***	0.458***	0.454***	0.476***
	(0.128)	(0.128)	(0.131)	(0.128)	(0.128)	(0.132)
Per Capita Income	0.884**	0.878*	0.915**	0.882*	0.877*	0.916**
	(0.342)	(0.343)	(0.346)	(0.343)	(0.343)	(0.345)
Population w/ High School Diploma (0000s)	-0.287***	-0.290***	-0.266**	-0.289***	-0.290***	-0.267**
	(0.0858)	(0.0863)	(0.0867)	(0.0864)	(0.0863)	(0.0864)
Population w/ College Diploma (0000s)	0.622*	0.621*	0.598*	0.618*	0.624*	0.600*
	(0.253)	(0.253)	(0.256)	(0.253)	(0.253)	(0.254)
Diagnosed HIV Cases to Date		0.000197				
		(0.000204)				
Diagnosed Caucasian HIV Cases to Date			-0.00443			-0.00265
			(0.00245)			(0.00700)
Diagnosed AA HIV Cases to Date			0.00166			0.00396
			(0.000919)			(0.00537)
Diagnosed Latino HIV Cases to Date			0.00575			0.00813
			(0.00492)			(0.00830)
Diagnosed Medicaid HIV Cases to Date				0.000563		-0.00299
				(0.000528)		(0.00622)
Diagnosed Male HIV Cases to Date					0.000295	-0.00177
					(0.000380)	(0.00620)
Constant	-6.908***	-6.912***	-6.932***	-6.923***	-6.905***	-6.955***
	(1.302)	(1.302)	(1.309)	(1.303)	(1.302)	(1.309)
Log Pseudolikelihood	-171.07	-170.64	-168.54	-170.53	-170.79	-168.41
Observations	1,067	1,067	1,067	1,067	1,067	1,067

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Table 16: Outcome of Random Implementation Model

	Random Implementation	Random Implementation In Treated	Random Implementation of Treatment Time
μ of Random β	0.00004	-0.00050	0.00030
σ Random β	0.01819	0.01647	0.05020
Estimated β	0.12600	0.12600	0.12600
Replications	500	500	500
Z-Score	6.92599	7.67856	2.50423
P-Value	P<0.0001	P<0.0001	P<0.01

Figure 1: Detailed HIV Progression Timeline



"During the early period after primary infection there is widespread dissemination of virus and a sharp decrease in the number of CD4 T cells in peripheral blood. An immune response to HIV ensues, with a decrease in detectable viremia followed by a prolonged period of clinical latency. The CD4 T-cell count continues to decrease during the following years, until it reaches a critical level below which there is a substantial risk of opportunistic diseases." (Pantaleo et al. 1993)

Figure 2: Aggregate Effect of Craigslist Introduction
Y-Axis: HIV Patients Admitted per Hospital Quarter
X-Axis: *Craigslist* Treatment

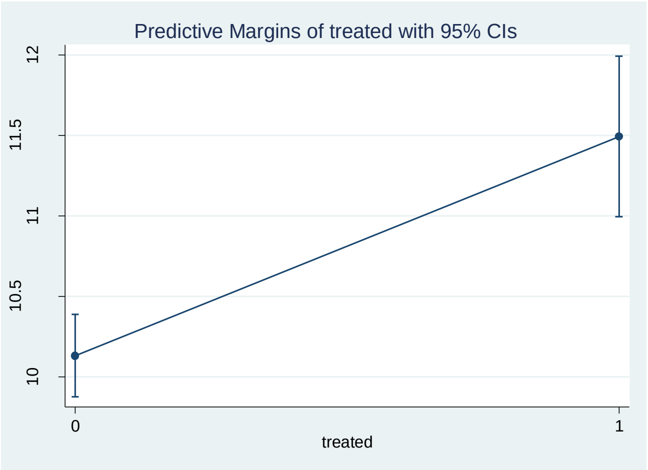


Figure 3: Aggregate Effect of Craigslist Introduction on Caucasians
Y-Axis: HIV Patients Admitted per Hospital Quarter
X-Axis: *Craigslist* Treatment

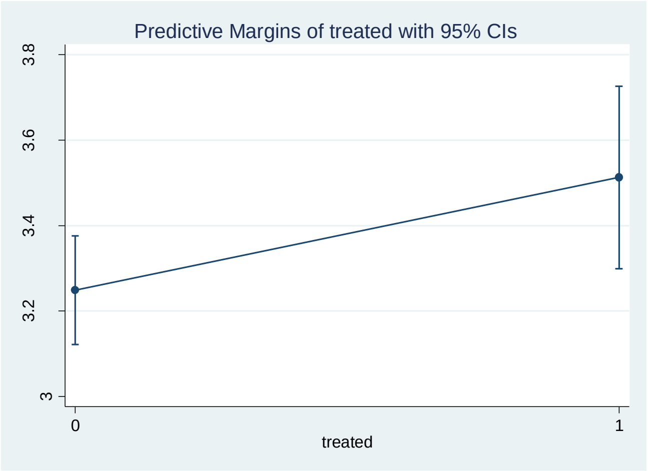


Figure 4: Aggregate Effect of Craigslist Introduction on African Americans
Y-Axis: HIV Patients Admitted per Hospital Quarter
X-Axis: *Craigslist* Treatment

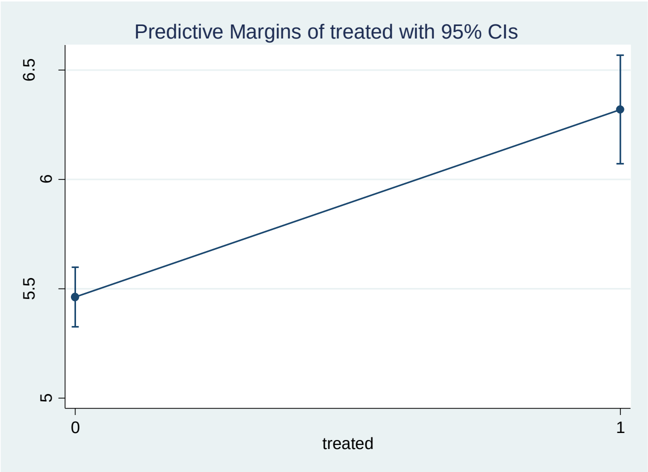


Figure 6: Aggregate Effect of Craigslist on non-Medicaid Patients
Y-Axis: HIV Patients Admitted per Hospital Quarter
X-Axis: *Craigslist* Treatment

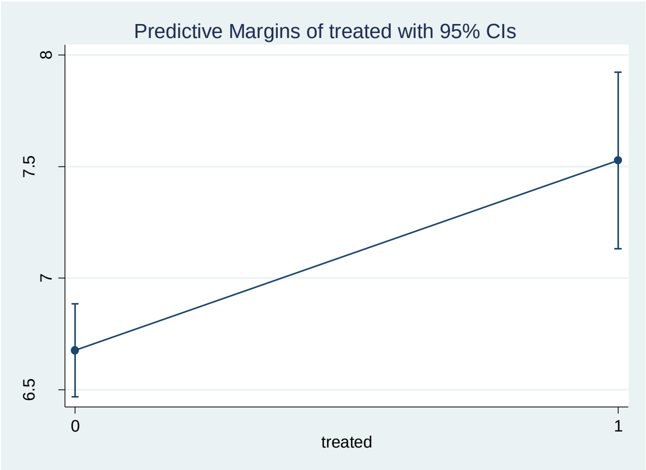


Figure 5: Aggregate Effect of Craigslist Introduction on Latinos
Y-Axis: HIV Patients Admitted per Hospital Quarter
X-Axis: *Craigslist* Treatment

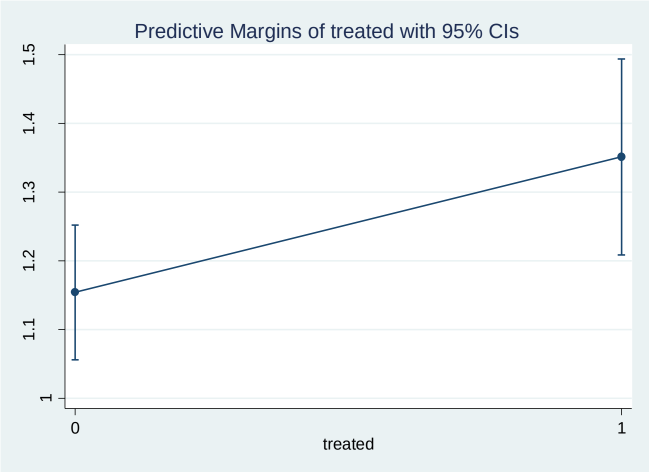


Figure 7: Aggregate Effect of Craigslist Introduction on Medicaid Patients
Y-Axis: HIV Patients Admitted per Hospital Quarter
X-Axis: *Craigslist* Treatment

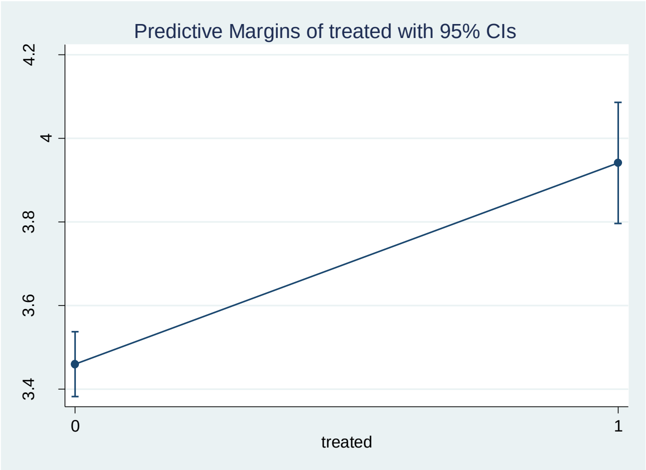


Figure 8: Aggregate Effect of Craigslist Introduction on Men
Y-Axis: HIV Patients Admitted per Hospital Quarter
X-Axis: *Craigslist* Treatment

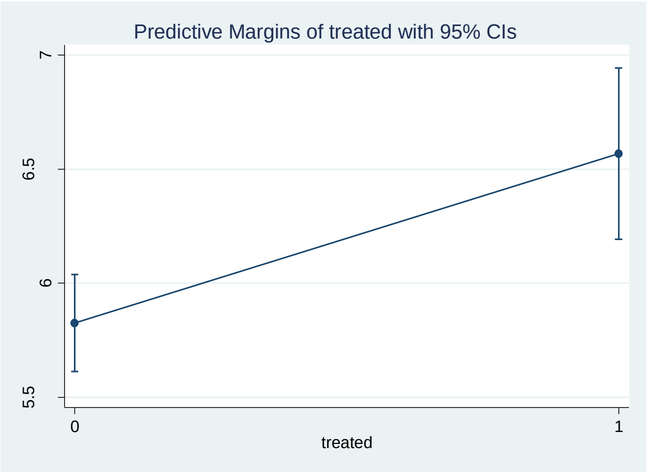


Figure 10: Aggregate Effect of Craigslist on African American Women
Y-Axis: HIV Patients Admitted per Hospital Quarter
X-Axis: *Craigslist* Treatment

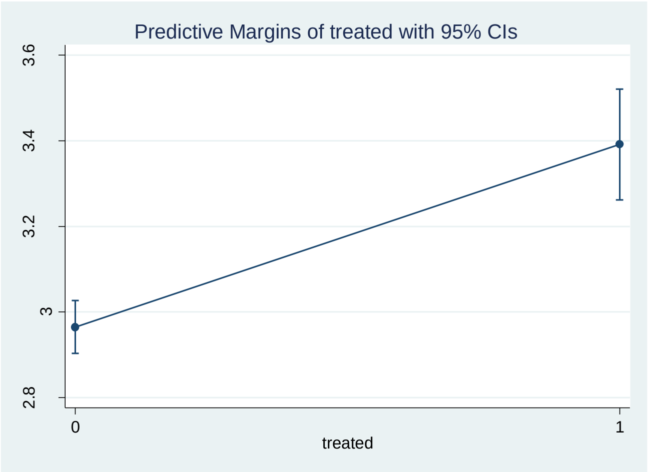


Figure 9: Aggregate Effect of Craigslist Introduction on Women
Y-Axis: HIV Patients Admitted per Hospital Quarter
X-Axis: *Craigslist* Treatment

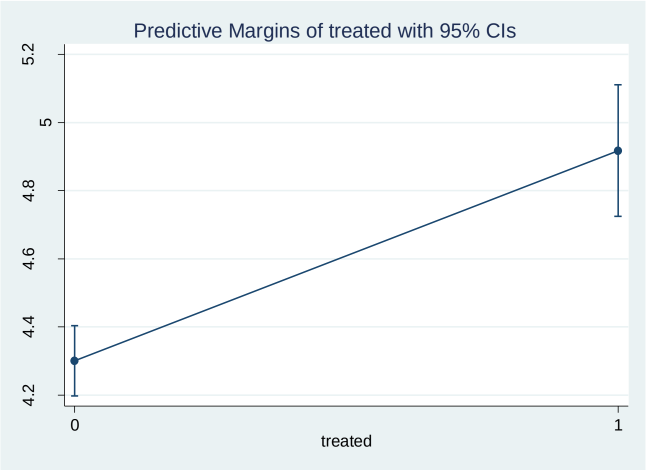


Figure 11: Aggregate Effect of Craigslist on African American Men
Y-Axis: HIV Patients Admitted per Hospital Quarter
X-Axis: *Craigslist* Treatment

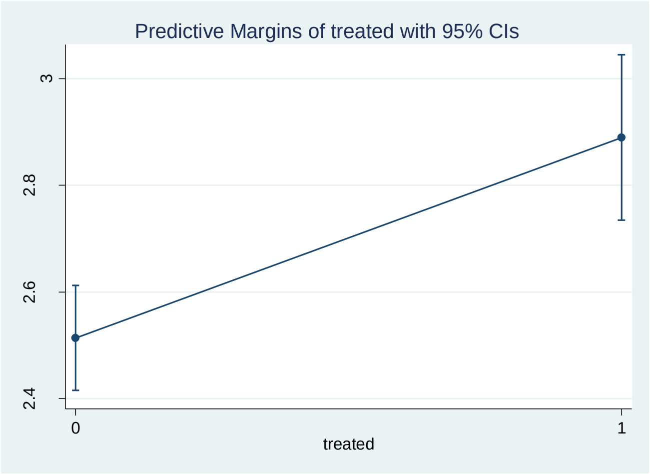


Figure 12: Aggregate Effect of Craigslist on African American Medicaid
Y-Axis: HIV Patients Admitted per Hospital Quarter
X-Axis: *Craigslist* Treatment

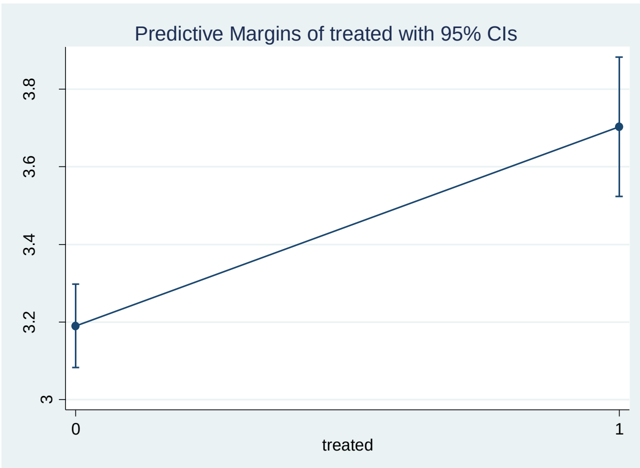


Figure 13: Effect of Craigslist on African American non-Medicaid
Y-Axis: HIV Patients Admitted per Hospital Quarter
X-Axis: *Craigslist* Treatment