

Behavioral Responses to Risk: Evidence from Personal Ads.*

Hasan Shahid†

September, 2024

[Click here for the latest version](#)

Preliminary Draft

Please do not distribute

Estimating how individuals respond to changes in the cost of engaging in risky behavior is challenging because risky behaviors are private. To study risky sexual behaviors, this study leverages Optical Character Recognition (OCR) and machine learning to construct a unique dataset of over 170,000 men's personal ads posted in the oldest and largest national LGBT publication in the United States, from 1975 to 1992. The vast majority of these personal ads represent men seeking other men for sexual or romantic relationships. This study documents significant changes in the content of personal ads well before any substantial efforts by the federal government to disseminate AIDS information. Most notably, I show that the early years of the epidemic saw an increase in the number of personal ads with a greater emphasis on 'safe-sex' practices. Thereafter, I limit analysis to the early years of the virus (before 1986). During this period, reports of AIDS represented an exogenous increase in the cost of risky sexual behavior because the virus could only be detected in the late phases of infection. This meant that a city's first AIDS case was exogenous to current patterns of risky sexual behavior. Therefore, I exploit variation in the timing of the first report of an AIDS case across U.S. cities and demonstrate that the take-up of safety language is causally related to the rising costs of risky sex. I find no evidence that the first AIDS case affected the total number of ads posted, ads seeking long-term relationships, ads that are explicitly sexual, or ads involving specific high risk sexual activities such as being the receptive anal partner. Additionally, I show that changes in the ads inform us about changes in behaviors by showing that areas with larger adoption of 'safety' language in the content of personal ads also experienced the greatest declines in rates of syphilis among men.

*I am grateful to Christopher Carpenter, Marcus Dillender, Tom Mroz, Kirsty Clark, Tara Mckay, Brian Beach, Analisa Packham, Panka Bencsik, Jesús Villero, Joshua C. Martin, Laura Nettuno, João Tampellini, Anyah Prasad, and seminar and conference participants at Vanderbilt University's LGBTQ+ Policy Lab, and the 2024 National LGBTQ Health Conference.

†Vanderbilt University, Department of Economics. Email: hasan.shahid@vanderbilt.edu

1 Introduction

Estimating the effect of a change in the perceived cost of risky behavior on such behaviors is challenging. Risky behaviors are private and not readily observed in datasets. To study risky behaviors, researchers rely either on self-reported survey data or data about certain outcomes which can proxy for those behaviors. Surveys are often limited in the depth of information they can provide because they only capture responses to the questions that are included in the survey. Additionally, self-reports often contain significant misreporting, and this misreporting is not randomly distributed (Cornu and De Paula, 2019).¹ Although using outcomes data as a proxy for risky behaviors accounts for individual misreporting, it provides limited information about the types of behaviors that are resulting in changes in the outcome variable. Outcomes data is also an imperfect proxy for risk. For example, there is significant misclassification between drug use disorder deaths, accidental deaths and suicide deaths which means that changes in drug use deaths do not necessarily inform us about changes in drug use (Schmeckenbecher et al., 2024). Additionally, there are spacial and temporal differences in the availability of sexually transmitted infection (STI) testing meaning that changes in STI rates could either represent changes in testing availability or changes in risky sex. Secondly, changes in the perceived costs of risky behaviors are rarely exogenous which makes it difficult to establish causal claims. For example, higher levels of risky sexual behavior may lead to greater STI risk, making it challenging to isolate the effect of increased STI risk on risky sexual behaviors (Oster, 2012, 2005).²

This paper addresses these problems by constructing a new dataset of over 170,000 men's personal ads, compiled from the classified section of the oldest and largest national LGBT publication in the United States, *the Advocate*, in the early years of the HIV/AIDS epidemic. The vast majority of these personal ads represent men seeking other men for sexual or romantic relationships. I use Optical Character Recognition (OCR) and machine learning to digitize and extract meaningful information from these ads about behavioral responses to HIV/AIDS in the early years of the epidemic. I first document overall trends in the number and types of

¹For example, Soulakova et al. (2012) finds significant differences in smoking misreporting rates by age and De Walque (2007) finds that married individuals are significantly more likely to misreport previous sexual behaviors compared to unmarried individuals.

²Prior research circumvents this issue by specifically exploiting shocks to information rather than risk (Kerwin, 2020; Delavande and Kohler, 2012.)

ads posted in the advocate. I document that the overall number of ads increases in the years leading up to the first reports of HIV/AIDS in mid-1981. Ads continue to rise and peak in late 1983 and are followed by a large decline thereafter. To examine these trends more closely, I employ machine learning techniques and search for specific terms to analyze how the content of ads evolves over time. Most notably, I find a significant increase in the number ads which include specific ‘safety’ language.^{3,4}

Thereafter, I limit my analysis to the early years of the virus (before 1986). During this period, reports of HIV/AIDS represented an exogenous increase in the perceived cost of risky sexual behaviors. During the early years of the virus, HIV/AIDS could only be detected in the late phases of infection.⁵ This meant that a city’s first AIDS case was exogenous to current patterns of risky sexual behaviors. This study exploits variation in the timing of a city’s first reported AIDS case to estimate the causal effect of an increase in STI risk on behavior. The early years of the virus were also characterized by limited action by the Federal Government. Limiting the analysis to these early years ensures that my estimates are not contaminated by the Federal Government’s AIDS education efforts, which began in 1986.⁶

To explore whether the first reported AIDS case represents an exogenous increase in the perceived cost of risky sexual behaviors, I construct a database of HIV/AIDS news reports published in leading local newspapers using the ProQuest Historical Newspaper database and TDM Studio. Utilizing an event study difference-in-differences design, I demonstrate that the first reported case of AIDS is associated with a significant increase in local newspaper reporting on HIV/AIDS. This increased coverage likely reflects either heightened concerns about the risk of HIV/AIDS or induces fears of contraction. In either case, it signifies an increase in the perceived cost of engaging in risky sexual behavior.

Thereafter, I use an event study difference-in-differences design to estimate the effect of the

³I group together all ads which include any of the following terms: “safe”, “condom”, or “protect”.

⁴I also document trends in the proportion of ads that are oriented towards long-term relationships, are explicitly sexual, or involve specific risky sexual acts such as receptive anal sex, insertive anal sex, and the sale of sex, as detailed in [subsubsection 3.1.3](#).

⁵An HIV infection only results in minor symptoms in the early phases of the infection. Before the development of an HIV test in 1985, the virus could only be detected after significant depletion of the immune system as the virus progressed to AIDS, the final and most severe phase of an HIV infection.

⁶In 1986, the U.S. Surgeon General released a report detailing how HIV/AIDS was transmitted and advocated for AIDS education. The report received widespread publicity and sparked a national debate about AIDS education in schools. The Centers for Disease Control (CDC) launched an AIDS education campaign in 1987 called ‘America Responds to AIDS’. This campaign included nationwide public service announcements, educational materials, and community outreach programs aimed at promoting HIV prevention, and safe sex practices.

first AIDS case on personal ads. While the first reported AIDS case appears to have little impact on the total number of ads posted, there is a noticeable shift in the content of the ads. I find significant evidence that individuals respond to the first reported AIDS case by expressing a preference for “safe-sex”. I find no significant evidence that the first AIDS case affected the number of personal ads seeking long-term relationships, explicitly sexual ads, or ads involving specific high-risk sexual activities, such as being the receptive anal partner. Taken together, this suggests that despite the increasing costs of risky sex, individuals continued to use the ads to seek out similar kinds of partnerships and encounters, but with a greater emphasis on safety measures to reduce the risk of HIV/AIDS transmission.

Changes in the content of the ad could either represent individuals who have previously posted in the magazine recognizing differences in risk and deciding to post a different ad in response, or it could represent changes in the composition of individuals posting ads. To explore which of these mechanisms is driving the changes documented in this paper, I use the phone numbers and P.O. Box numbers provided in the ads to create consistent panels of repeat posters. Although many individuals can be identified as repeat posters, my sample sizes become significantly smaller when I try to follow these individuals over longer periods of time. Although these sample sizes are too small to conduct causal analysis similar to my main specification, I am able to construct smaller subsets of consistent populations whom I follow over time. Even with consistent sample populations, I find significant evidence of the take-up of safety language. This suggests that the increase in safety ads was not solely the result of changes in the composition of individuals posting ads.

Additionally, the content of the personal ads are not necessarily indicative of changes in behaviors. Personal ads often contain vague or ambiguous language that may not necessarily inform us about changes in behavior. For example, many ads express a preference for “safe-sex” but do not mention the specific risk mitigation strategies. Therefore, I also compare changes in the content of personal ads to changes in rates of sexually transmitted infections (STIs). Syphilis is heavily concentrated among men who have sex with men (MSM). Although STI rates are only available at the state-level during this period and I am unable to use my main specification to estimate effects on STI rates, I can compare state-level changes in personal ads to state-level changes in STI rates to check whether patterns of personal ads represent changes in behavior. I find that states with larger adoption of ‘safety’ language in the content of personal ads also

experienced the greatest decline in syphilis rates among men. This indicates that personal ads can offer meaningful insights into patterns of behavior, not just preferences.

This study is part of a growing literature which uses natural language processing and machine learning tools to generate insight from unstructured text and image datasets (Dell, 2024; Davis et al., 2020; Atalay et al., 2020). These tools allow me to analyze the universe of ads posted in the classified section.⁷ Although other researchers have used personal ads in LGBTQ+ magazines and newspapers to gain insight into the sexual behaviors of men who have sex with men (MSM), their analyses are based on manually digitized ads, resulting in smaller sample sizes that do not lend themselves to causal inference. Table A.1 lists these studies and the number of ads used in each study. Notably, each of these studies utilizes less than 1% of my sample size. Nonetheless, these studies provide valuable insight about the content of men’s personal ads. Lee (1976), Lumby (1978), Bartholome et al. (2000) and Baker (2003) explore themes and preferences represented in the ads while Thorne and Coupland (1998), Laner and Kamel (1978), Gonzales and Meyers (1993), and Hatala and Prehodka (1996) compare personal ads by gender and sexual orientation. Most of these studies document that gay men’s ads place a strong emphasis on physical appearance and sexual relationships. Several papers also document themes in these ads in the context of the HIV/AIDS crisis. Davidson (1991) documents increasing health concerns during the early years of HIV/AIDS while Smith (2000) documents the prevalence of safety language. Building on these findings, my paper shows that we can use natural language processing and machine learning tools to form a significantly larger sample of personal ads to generate meaningful information about specific sexual behaviors and responses to STI risk.

I also contribute to the growing literature in economics that explores behavioral responses to changes in the cost of risky behavior. There is an extensive body of literature examining the effects of changes in the direct and indirect costs of cigarettes and e-cigarettes on smoking behaviors (Pesko et al., 2020; Nesson, 2017; Callison and Kaestner, 2014; Cotti et al., 2016; Hansen et al., 2017), the direct and indirect costs of alcohol on alcohol consumption (Hinnosaar and Liu, 2022; Carpenter and Dobkin, 2009; Miravete et al., 2020; Chalfin et al., 2023; Schilbach, 2019), and the direct and indirect costs of drugs on drug use (Ruhm, 2019; Packham, 2022; Hansen et al., 2017; Doleac and Mukherjee, 2022).

⁷The analysis provided in this paper is based on over 170,000 personal ads.

More specific to this paper, I contribute to a literature which explores behavioral responses to changes in the perceived cost of risky sex. Some of this literature exploits changes in STI risk (Oster, 2005; Auld, 2006; Shahid, 2024; Spencer, 2024) while others exploit changes in information which shape perceptions of risk (Kerwin, 2020; Delavande and Kohler, 2012). Another strand of literature exploits changes in the costs of risky sex induced by technological innovation (Shahid, 2023; Mann, 2023; Chan et al., 2016; Baranov and Kohler, 2018). My contribution is unique because of it's focus on specific behavioral responses rather than outcomes data.⁸ Understanding specific behavioral responses deepens our understanding of how individuals respond to risks and could help inform the design of targeted interventions.

2 Background

2.1 Spread of HIV/AIDS

Although an article titled “Rare Cancer Seen in 41 Homosexuals” published on July 3rd, 1981 in the *New York Times* is widely cited as the first newspaper report on HIV/AIDS, the *New York Native*, a local gay newspaper, had reported about “an exotic new disease” that was striking gay men in New York as early as May 1981 (Streitmatter, 1995a).⁹ Despite the absence of scientific evidence showing the cause of the illness, even these early reports linked the disease to “frequent sexual encounters with different (homosexual) partners” (Altman, 1981). It was not until 1983 that scientists identified the illness as being caused by a virus, later named Human Immunodeficiency Virus (HIV), which could be transmitted through sexual contact. However, even before this discovery, community responses had already emphasized reducing or abstaining from sexual activity and practicing safe sex.¹⁰

Early cases of the virus were concentrated among gay men in large urban centers. Consequently, gay men living outside these areas often perceived little to no risk of contracting HIV/AIDS (St Lawrence et al., 1989). There was a large amount of reporting about the first

⁸My findings are most similar to Ahituv et al. (1996) who uses survey data to show that higher rates of AIDS are associated with higher levels of condoms.

⁹There is now significant evidence of earlier cases of HIV/AIDS with recent retrospective studies identifying evidence of infections as early as 1969 (AIDS Foundation of Chicago, 2024).

¹⁰In June 1982, San Francisco based community organization, *The Sisters of Perpetual Indulgence*, distributed a flier titled “Play Fair” which emphasized safe sex practices (Gay in the 80s, 2013). In July 1982, the New York-based organization *Gay Men’s Health Crisis* distributed its first newsletter, which included recommendations to reduce the number of sexual partners to prevent the spread of AIDS (Terra, 2023).

local case of AIDS in local newspapers.¹¹ As reports of AIDS emerged in other cities, these perceptions likely began to shift.

In the early years, nearly all diagnoses represented late forms of infection. An HIV infection only results in minor symptoms in the early phases of the infection. In these years, AIDS was identified through the diagnosis of Kaposi's Sarcoma (KS), Pneumocystis carinii Pneumonia (PCP) or other opportunistic diseases. These only emerge after significant depletion of the immune system. Before the development of an HIV test in 1985, the virus could only be detected after the virus progressed to AIDS, the final and most severe phase of an HIV infection. The fact that most AIDS diagnoses represented late phases of infection is also evident in the life expectancies of those diagnosed. Of all individuals diagnosed with the virus by January 1983, 75% were known dead by April 1985 ([Centers for Disease Control and Prevention, 1985](#)). We know now that even untreated HIV positive individuals live approximately 10 years before their disease progresses to AIDS ([National Institutes of Health, 2024](#)).

2.2 Government Inaction

In the early years of the virus, President Ronald Reagan did not address HIV/AIDS. He briefly mentioned the virus after the death of Hollywood actor, Rock Hudson, in 1985 and provided a more detailed address in 1987. Prior to 1986, the Federal Government made few attempts to disseminate HIV/AIDS related information. In October 1986, Surgeon General Everette Koop released a report on AIDS where he explained how HIV/AIDS was spread and advocated for HIV/AIDS education in schools. The Surgeon General's report was met with both praise and criticism and prompted a national debate about HIV/AIDS education in schools. Ultimately, an abbreviated pamphlet of the report was distributed to all U.S. households titled "Understanding AIDS" between May 26 and June 15, 1988. The Centers for Disease Control (CDC) launched an AIDS education campaign in 1987 called 'America Responds to AIDS'. This campaign included nationwide public service announcements, educational materials, and community outreach programs aimed at promoting HIV prevention, and safe sex practices.

Despite relative inaction from the Federal Government prior to 1986, there were significant community responses aimed at AIDS information. Pre-existing LGBTQ+ organizations worked

¹¹See [Figure 7](#) for examples.

to disseminate information about AIDS. For example, as early as June 1982, San Francisco based organization *The Sisters of Perpetual Indulgence* were distributing fliers titled “Play Fair!” which emphasized safe sex practices ([Gay in the 80s, 2013](#)). Several new community organizations were formed to address increasing concerns of AIDS. They also contributed to early dissemination of AIDS information. In July 1982, the recently formed New York-based organization, Gay Men’s Health Crisis distributed its first newsletter, which included the latest in AIDS research and recommendations to prevent the spread of the virus ([Terra, 2023](#)). Other significant efforts at AIDS information dissemination during this period include the L.A. Cares (Los Angeles Cooperative AIDS Risk Reduction Service). L.A. Cares was launched in 1984 by the AIDS Project Los Angeles (APLA) to disseminate AIDS information among gay men in Los Angeles ([Los Angeles Times, 1985](#)).¹²

2.3 Gay Newspapers and Personal Ads

Several gay newspapers and magazines rose to prominence following the 1969 *Stonewall Riots* ([Streitmatter, 1995b](#)). Many of these publications catered to the gay populations of specific cities.¹³ *The Advocate* soon emerged as the leading national gay publication. Although consistent circulation data is unavailable, I can obtain some information by searching through old issues of the magazine for the “Statement of Ownership, Management & Circulation”. This provides me with the average number of copies of *the Advocate* distributed in the past 12 months per issue. [Figure A.2](#) presents the distribution statistics in the located statements. By 1985, the Advocate accounted for approximately 10% of the combined circulation of all lesbian and gay publications ([Streitmatter, 1995a](#)).

Many of these publications also had classified sections which provided men a convenient way to connect with other men seeking romantic or sexual relationships.¹⁴ These were particularly popular and served as an effective means for the newspapers to generate revenue. The publication would charge a small fee to allow readers to post personal ads. *The Advocate* was unique among its competitors because it had significant distribution throughout the country. Many of

¹²Efforts involved television, newspaper, and magazine advertisements featuring Hollywood actor, Zelda Rubinstein, telling younger gay men to ‘play careful’ to prevent contracting the virus ([Story, 2019](#)).

¹³For example, the *New York Native* was circulated in New York city while the *Sentential* and *Bay Area Reported* served the city of San Francisco ([Streitmatter, 1995a](#)).

¹⁴See [Table A.1](#) for details about studies using the personals section of several other leading newspapers.

the other newspapers only served specific cities.¹⁵

During the early years of the HIV/AIDS epidemic, these publications played a pivotal role in disseminating HIV/AIDS information during a period when mainstream media was often reluctant or slow to cover the crisis. Nevertheless, many have been critical of the lack of AIDS reporting even in gay publications including *the Advocate* (Terra, 2023). Streitmatter (1995a) explains that newspapers generated some of their revenue through advertisements from bathhouses.¹⁶ These interests may have deterred certain publications from reporting the seriousness of the virus.

3 Data

3.1 Personal Ads Data

The main dataset used in this paper comes from the classified section of issues of *the Advocate* published between 1975 to 1992. *The Advocate* is the largest and longest running LGBT Publication in the United States. It was first published as a local newsletter distributed by an activist group in Los Angeles in 1967.¹⁷ It was converted into a national magazine in 1974. Over the next few years, *the Advocate* became the leading LGBT magazine in the country (Kirsch, 1995).¹⁸

The Advocate has been publishing a classified section since 1967.¹⁹ Individuals pay a small fee to mail in the text for their personal ad in the classified section.²⁰ Although individuals

¹⁵Individuals in some large cities, such as New York and San Francisco, had the option to post personal ads in several different publications. Since this paper's findings are based solely on personal ads in *the Advocate*, one limitation is that I am unable to identify individuals who may have switched between their preferred publication. However, in subsection 5.1, I find qualitatively similar estimates even after excluding all cities that reported their first AIDS case in 1981. Since the initial AIDS cases were identified in cities with large gay communities, these are also the cities where individuals had multiple publication options for posting personal ads.

¹⁶During the mid-1980s, several cities had debates about whether bathhouses should be closed in order prevent the spread of HIV/AIDS. For example, San Francisco became the first city to ban bathhouses in 1984 and New York city passed similar but less stringent regulations in 1985 (Binson and Woods, 2013).

¹⁷The Personal Rights in Defence and Education (PRIDE) was an activist organization formed in 1966 in response to frequent police raids on Los Angeles' gay bars.

¹⁸By 1985, the average number of copies distributed per issue was approximately 80,000, with nearly equal shares coming from a combination of dealers, carriers, street vendors, and counter sales, and from mail subscriptions.

¹⁹The first issue of *the Advocate* that is available to me via Proquest was published on September 1st, 1967 and contains a classified section.

²⁰Individuals pay a fixed fee for a certain number of characters but this fee increases as the number of characters increases.

use these ads for a wide variety of purposes, the vast majority of these ads represent men seeking other men for sexual or romantic relationships. Following 1992, rather than being part of the magazine itself, *the Advocate Classifieds* became a separate stand-alone magazine.²¹ I do not have access to issues of the stand-along magazine.

3.1.1 Data Collection and Cleaning

The primary dataset is produced by digitizing issues of *the Advocate* which have been purchased by the Vanderbilt University library through Proquest. Personal ads are presented one after the other in columns over multiple pages of *the Advocate*. The individual ads are sorted by state as depicted in [Figure 1](#). I exploit the vertical columns dividing the ads and Optical Character Recognition (OCR) to create a dataset where each column contains the text making up each ad.²² Since the columns of personal ads also include other items such as advertisements for other businesses, I use machine learning to identify and drop all observations which do not represent personal ads. The data collection and cleaning process is described in detail in [Appendix B](#). This process results in a dataset of 176,906 person ads.

3.1.2 Data Specifics

Subsequently, I analyze the text of each advertisement to extract meaningful information about the advertiser and their preferences. Since individuals selectively disclose information in their personal ads, and the median word count is only 25 words, the information available about each advertiser is limited. Nevertheless, it is common for advertisers to reveal demographic details such as age and race. I am able to discern the state where an advertiser resides based on the format in which the ads are listed in *the Advocate*. Ads also often include phone numbers and addresses. I use this information to identify which MSA many of the ads represent.

In addition to providing valuable demographic information, I also categorize the ads in economically meaningful ways. Specifically, a significant number of ads express a preference for long-term relationships or are explicitly sexual. I use natural language processing tools to determine whether an ad represents a preference for a long-term relationship and whether it is

²¹Information obtained from personal communication with Proquest.

²²See [Appendix B](#) for details.

explicitly sexual. From the universe of ads, I randomly select 300 and manually identify whether the ads represent a preference for a long-term relationship and whether they are explicitly sexual. I use this as training data. I use a supervised machine learning model to classify the ads. The model is trained on the manually labeled subset of 300 ads, which have been identified as either expressing a preference for long-term relationships or being explicitly sexual.²³ The trained model is then applied to the entire dataset to automatically categorize the remaining ads. Additionally, a significant portion of the ads mention specific safety concerns in the ad. I search through the personal ads for the terms “safe”, “condom” and “protect”. I group together ads which mention any of these terms and henceforth, refer to these ads as ‘safety ads.’²⁴

Additionally, many of the ads also mention a preference for specific sexual acts. Given that different sexual acts represent different levels of risk, it is possible that individuals respond to HIV/AIDS risk by changing their preferred sex act.²⁵ These are somewhat harder to categorize. Searching for ads with specific words or phrases may yield misleading trends given that individuals sometimes specify their own preferred sexual position or the position of their partner.²⁶ Therefore, I identify verbs in the ads which allow me to separate text which is self-descriptive or describes characteristics of a preferred partner.²⁷ After separating the ad into a self-descriptive statement and a statement describing the preferred partner, I identify ads which represent a preference for receptive and insertive anal sex.²⁸ A significant portion of the ads also represent individuals selling sex. I search the ads for specific terms which may indicate the sale of sex.²⁹

The ads also include important geographic information. The ads are listed in columns by state in the classified section of the magazine. The vast majority of personal ads appear under

²³I use Term Frequency-Inverse Document Frequency (TF-IDF) to vectorize the ads. This involves measuring the importance of words in the ads of a specific category. It discounts words that are common throughout the document.

²⁴I do not use a supervised machine learning strategy to categorize safety ads given the diversity in the text of these ads. Many different types of ads mention safety terms.

²⁵For example, receptive anal intercourse represents a significantly larger risk than insertive anal intercourse. These differences in risk were identified relatively early in the virus. For example, in May 1983, a manual titled “How to Have Sex in an Epidemic: One Approach” was widely distributed among gay men ([Callen and Berkowitz, 1983](#)). The manual categorizes different sexual acts by their risk of transmission.

²⁶For example, an individual seeking receptive anal sex may post that they are a “bottom” or that they are seeking a “top”.

²⁷For example, the text following verbs such as ‘seeks’ and ‘looking’ is considered as characteristics of a preferred partner. Alternatively, the text preceding verbs such as ‘wanted’ or ‘desired’ represents characteristics of a preferred partner. See [Appendix B](#) for details.

²⁸For example, if the self-descriptive statement portion of the ad contains the term ‘top’ or the statement describing the preferred partner includes the term ‘bottom’, the ad is treated as representing a preference for insertive anal sex. More details provided in [Appendix B](#).

²⁹There include terms such as “escort” or “massage”. A complete list is provided in [Appendix B](#).

the title of a specific state while some ads appear under a ‘nationwide’ and ‘international’ title. I use this information to identify the state where an advertiser may reside. The ads themselves also include additional geographic information. Most ads include either a phone-number or a P.O. Box number with a zip code. Therefore, I search each add for consecutive digits which follow the form of a phone-number or a zip code. The first three digits of a phone-number represent a particular area code and can be linked to a metropolitan statistical area. Zip codes can also be used to link to MSAs.

I provide a summary of these statistics in [Table 1](#). The majority of ads are from young white men.³⁰ Approximately 73% of ads are explicitly sexual and 31% of the ads express a preference for long-term relationships. 7% of ads mention some safety term. 13% of ads indicate a preference for insertive anal intercourse while 11% of ads indicate an interest for receptive anal intercourse. 13% of ads include terms which suggest they may indicate the sale of sex. I also present statistics about the top five states and MSAs represented in the personals. I am able to identify the state and MSA of the advertiser for 70% of the ads. Interestingly, a large portion of ads represent cities with large gay populations such as Los Angeles, New York and San Francisco.

3.1.3 Trends over time

I also explore how the number of ads evolve over time. [Figure 2](#) depicts the total number of personal ads in each issue of the advocate over time. In general, there is large variation in the total number of ads posted per issue where most issues have anywhere between 350 and 750 ads. There is also a high degree of serial correlation in the number of ads posted. Overall, I observe a rise in the average number of ads posted in the years leading up to the first reports of AIDS in mid-1981. There is an additional increase in the total number of ads posted in the early years of the virus. The number of ads peak in late 1983 and are followed by a large decline after which most issues hover around 400 ads per issue. The rise in the number of personal ads prior to the first reports of AIDS and the decrease in the total number of ads in the mid 1980s follows a similar trend to the total number of issues distributed per year.³¹

³⁰[Figure A.1](#) presents the Age distribution of respondents. There is significant bunching of respondents at ages 18, 20, 30 indicating that some individuals may be misreporting their age.

³¹I only have limited information about distribution over time. I search through old issues of the newspaper for the “Statement of Ownership, Management & Circulation”. This provides me with the average number of

To explore these trends further, I break down the sample into ads which express a preference for long-term relationships and ads which do not in the first panel of [Figure 3](#). The figure suggests that the spike in personal ads in 1983 is driven entirely by ads which represent a preference for long-term relationships.³² In the second panel of [Figure 3](#), I break down the sample into ads which are explicitly sexual and ads which are not. Similarly, I find that the 1983 spike in ads is driven by ads which are not explicitly sexual. The third panel of [Figure 3](#) presents the total number of ads which mention ‘safety’ per issue. I find that there is a significant increase in the number of safety ads during the early years of HIV/AIDS. Ads with safety terms peak in early 1988 and are followed by a downwards trend.

In [Figure 4](#), I present the proportion of ads by type. The first panel presents the proportion of ads which represent a preference for long-term relationships and ads which are explicitly sexual as a share of all ads. As suggested by [Figure 3](#), I observe a spike in ads which express a preference for long term relationships in 1983 and a dip in the proportion which are explicitly sexual. The second panel of [Figure 4](#) shows that the proportion of ads which include a safety term largely follows a similar pattern to the third panel of [Figure 3](#). The third panel of [Figure 4](#) depicts the proportion of ads representing insertive and receptive anal sex. For the most part, these trends evolve in parallel. However, in the late 1980s we observe significant decline in the proportion of ads representing receptive anal sex but no decline in the proportion of ads representing insertive anal sex.³³ The last panel of [Figure 4](#) depicts the proportion of ads representing the sale of sex over time. In general, these types of ads trend downward till the end of 1983 and increase over time thereafter.

copies of the advocate distributed in the past 12 months per issue. [Figure A.2](#) presents the distribution statistics in the located statements.

³²To ensure that this spike is not a result of problems associated with the supervised machine learning model used to classify the data, I present trends of the total number of ads which mention any of the following terms: ‘partner’, ‘boyfriend’, ‘longterm’, ‘ltr’, ‘rel’ in [Figure A.3](#). I find that the terms identify a smaller sample of long-term relationship oriented ads than my preferred supervised machine learning model but I still observe a significant spike in late 1983.

³³This is interesting because receptive anal sex represents a significantly higher level of HIV/AIDS risk compared to insertive anal sex. Shifting away from receptive anal sex may represent a risk mitigation strategy.

3.2 Newspaper Reports Database

3.2.1 Data Collection and Cleaning

To explore how newspaper reports respond to the first AIDS case in an MSA, I construct a dataset of leading daily city newspapers. I use Proquest Historical Newspaper database which gives me access to many popular historical newspapers.³⁴ I include newspapers which are published continuously from 1981-1992. Thereafter, of the newspapers provided by Proquest, I identify the leading newspaper in each city using the Editor & Publisher International year book for 1981.³⁵ In [Table A.3](#), I provide a list of these newspapers. The table provides the name of the newspapers in my dataset, the city and state they are distributed in, as well their city ranking based on information from the 1981 Editor & Publisher International year book.³⁶ Since the proquest directory only provides a subset of all newspapers, I don't always observe the leading newspaper in a city (for example, I have access to the Daily News in New York which is the second most popular newspaper in the city but I do not have access to the most popular). I also do not have any newspaper from many major city (for example, I do not have any leading newspaper from Chicago in this dataset).

Despite these problems, the newspaper database provides valuable information about local reporting of HIV/AIDS during this period. Thereafter, I identify all news reports related to the HIV/AIDS crisis. This involves using Proquest's Text and data mining tool TDM Studio. I search for HIV/AIDS related terms and then use machine learning to ensure that the search result represents an HIV/AIDS related article.³⁷ This process is described in detail in [Appendix C](#).

³⁴Several other studies have used this dataset order to study effects of local newspaper reports ([Engelberg and Parsons, 2011](#)).

³⁵Editor & Publisher International year book provides distribution statistics for each newspaper by the city it is distributed in. It has been used to evaluate the economic performance of newspapers ([Angelucci et al., 2024](#)).

³⁶A rank of 1 means that of all the newspapers listed in Editor & Publisher International year book for the city, this newspaper is listed as having the greatest distribution.

³⁷For example, the search prompt would pick up an article titled "U.S. aids China in combating bird flu disease."

3.2.2 Trends over time

The first panel of [Figure 6](#) shows the monthly total of HIV/AIDS articles in the newspaper database. Overall, HIV/AIDS reports increased over the 80s with several spikes around important HIV/AIDS events. In the early years of the virus, there was only sporadic reporting about HIV/AIDS among gay men. In 1983 there were reports of HIV/AIDS among other groups as well as information about the role of blood transfusion contributing to the HIV/AIDS crisis. Thereafter, there was a large spike in HIV/AIDS reporting when American actor Rock Hudson revealed that he had AIDS in 1985 and his subsequent death. There was another spike in HIV/AIDS reports when professional basketball player Magic Johnson went public about being HIV positive. Observing significant spikes around HIV/AIDS related events lends further credibility to my strategy of identifying HIV/AIDS articles.

3.3 AIDS Public Information Dataset

This paper employs data from the AIDS Public Information Dataset (APID) in order to identify variation in the first reported AIDS case in each metropolitan statistical area (MSA) ([Department of Health and Human Services, 2005](#)). APID contains data on monthly counts of AIDS cases by city from 1981-2002. The APID contains data on MSAs with 500,000 or more population. In order to make this data compatible with the personal ads data, I aggregate to the MSA-half year level.

For my main analysis, I exploit variation in the timing of the first reported AIDS case in the data. Although HIV/AIDS was spreading among the U.S. population well before, the first reports of AIDS emerged in 1981.³⁸ [Figure 5](#) shows the number of MSA's with reports of an AIDS cases over time. By the end of 1981, cases of AIDS had been reported in 18 cities.³⁹ Thereafter, reports of AIDS emerged in other cities. By the second half of 1985, almost all MSA's in the APID had reports of AIDS.⁴⁰

³⁸Retrospective studies have found evidence of infections as early as 1969 ([AIDS Foundation of Chicago, 2024](#)).

³⁹These include Houston, Los Angeles, Cleveland, New Haven, Portland, Pittsburgh, Miami, Boston, Atlanta, Chicago, Philadelphia, Detroit, Syracuse, Tampa, Hartford, San Francisco, Newyork and Baltimore.

⁴⁰A full list of MSA's and the first half-year they reported an AIDS case can be found in [Table A.2](#).

4 Methodology & Results

Estimating the causal effect of changes in the perceived cost of sexual interactions on behaviors is complicated because changes in risk are often the result of changes in behavior. In order to isolate the causal effect of an increase in the perceived cost of risky sexual behavior on personal ads, I exploit variation in the timing of the first report of an AIDS case across U.S. cities. AIDS represents a late stage of an HIV infection. In the early years of the virus, there was no way to test for HIV. AIDS was only diagnosed through clinical observations and the presence of specific opportunistic infections and cancers which only emerged after significant depletion of the immune system.⁴¹ Since an HIV infection causes only minor symptoms in its early stages, early detection was not possible until the development of an HIV test. Following an HIV infection, HIV enters a latent phase where it results in no or only mild symptoms for several years. This is followed by a symptomatic phase and the development of AIDS. Without treatment, the average time between an HIV infection and the development of AIDS is 10 years. Current patterns of local risky sexual behaviors would not be related to when a city reports its first AIDS case. Therefore, I argue that a city's first AIDS case is exogenous to current sexual behaviors given the 10-year incubation period of the virus.⁴²

Although I present trends for the entire period that I have access to personal ads in subsection 3.2.2, I limit causal analysis to ads posted before 1986. During this period, reports of AIDS represent an exogenous shock to the cost of risky behavior.⁴³ Additionally, it ensures that my estimates are not contaminated by the Federal Government's AIDS education efforts, which began in 1986.

4.1 First AIDS case as a Proxy for Increased Perceived Risk

There is limited information about perceptions of the cost of risky sexual behaviors, particularly in the early years of the virus. Public health surveillance systems did not collect information

⁴¹Scientists discovered that AIDS was caused by a virus which was named the Human Immunodeficiency Virus in 1983. The U.S. Food and Drug Association only approved of a test in 1985.

⁴²Spencer (2024) makes a similar argument when evaluating the effect of AIDS on women's birthrates.

⁴³Since the time between infection and the development of AIDS can be as little as 5 years, AIDS cases following 1986 could be the result of infections which occurred after first reports of AIDS in the U.S. in 1981 (Poorolajal et al., 2016). Limiting analysis to ads posted before 1986, ensures that the treatment variable represents an exogenous shock.

about perceptions of HIV/AIDS risk till later years.⁴⁴ My preferred estimation strategy assumes that the first reported case of AIDS in a particular city represents an exogenous increase in the perceived cost of risky sex. Although there is no direct way to verify whether this is the case, in this section, I show that the first reported case of AIDS is associated with increased AIDS reporting in the local newspaper. Local newspapers reported extensively about the first local reports of AIDS in the city. I provide some examples of these reports in [Figure 7](#).

I then use Proquest Historical Newspaper database in order to construct a database of leading local newspapers in 73 cities. Thereafter, I employ ProQuest's TDM Studio to identify all articles related to HIV/AIDS.⁴⁵ I first present trends in AIDS reports over time in the first panel of [Figure 6](#). In the early years of the virus, there was only sporadic reporting, but this increased over time. I observe significant spikes in the number of AIDS-related articles around major AIDS-related events. For example, the number of AIDS-related articles surged when famous figures such as Rock Hudson and Magic Johnson revealed their HIV-positive status. These trends indicate that the constructed dataset credibly captures variation in HIV/AIDS reporting. In the second panel of [Figure 6](#), I show that areas with higher rates of AIDS also had a greater number of AIDS related articles in their local newspaper.⁴⁶ This suggests that HIV/AIDS risk was more salient in cities with higher rates of HIV/AIDS.

In order for the first AIDS case to represent an increase in the perceived cost of risky sexual behavior, we would expect that the first case be accompanied by an increase in AIDS reporting. In order to test whether this is the case, I estimate the following equation:

$$\# \text{ of AIDS Articles}_{mt} = \alpha + \sum_{l=-6}^9 \beta_l \text{ First AIDS Case[t=l]}_{mt} + \gamma_t + a_m + \epsilon_{mt} \quad (1)$$

where the dependant variable represents the the number of AIDS articles in the local news-

⁴⁴The National Health Interview Survey (NHIS) and Behavior Risk Factor Surveillance System (BRFSS) only began collecting information about perceptions of HIV/AIDS risk in 1987 in 1988 respectively. Others have used information from these surveys to show that higher rates of AIDS are associated with higher perceived risks of contraction ([Spencer, 2024](#)).

⁴⁵This process is described in detail in [Appendix C](#)

⁴⁶In the second panel of [Figure 6](#), in order to isolate spacial variation in HIV/AIDS reporting, I pool together rates for AIDS from 1981-1986 for all MSAs that are present in the APIDs as well as the constructed AIDS News dataset. This pooled rate is presented on the horizontal axis. The total number of AIDS articles published in the local newspaper from the years 1981-1986 is presented on the verticle axis.

paper of a particular MSA, m , in period, t . The treatment variable $\text{First AIDS Case}[t=l]_{mt}$ is an indicator variable which equals 1 when an observation represents an MSA, m , that is l months relative to the first report of an AIDS case. l ranges from 6 months before and 9 months after the first reported AIDS case. γ_t represents period fixed effects and a_m represents MSA fixed effects. Given recent developments in the difference-in-differences literature which identify significant flaws with linear regressions and fixed effects specifications with staggered treatment timing and treatment effect heterogeneity, in my preferred specification, I use the Imputation Estimator developed by [Borusyak et al. \(2024\)](#). This estimator is robust to heterogeneity in treatment timing and effects.

Estimates for [Equation 1](#) are provided in the third panel of [Figure 6](#). The figure suggests that the first report of an AIDS case in an MSA is associated with a significant and persistent increase in HIV/AIDS reporting. This suggests that the first AIDS case represented a non-trivial increase in the salience of HIV/AIDS. The increased coverage is expected to either reflect growing concerns about HIV/AIDS risk or induce fears of contraction.

4.2 Effect of First AIDS Case on Personal Ads

After showing that the first reported AIDS case represents an exogenous increase in the cost of risky sexual behavior, I then explore how personal ads respond to this change. To estimate the effect of the first reported AIDS case on personal ads, I employ an event study design similar to [Equation 1](#). I opt for a difference-in-differences event study design as my main specification because it allows me to observe time-varying treatment effects and test for differences in pre-existing trends in personal ads. To conduct this analysis, I first aggregate the personal ads data to the MSA-half-year level.⁴⁷ Thereafter, I estimate the following equation:

$$\text{Log}(\text{Number of Personals} + 1)_{mst} = \alpha + \sum_{l=-6}^8 \beta_l \text{First AIDS Case}[t=l]_{mst} + \gamma_t + a_m + \delta_{sm} + \epsilon_{mt} \quad (2)$$

where $\text{Log}(\text{Number of Personals} + 1)_{mst}$ represents the log of the average number of per-

⁴⁷Pooling over multiple issues allows for greater precision given the large amount of variation in the number of personal ads posted for each MSA per issue.

sonal ads posted in an issue of the magazine over the half-year period plus one in MSA, m , in season, s , and half-year, t .^{48,49} In my preferred specification, I use a log-like transformation of the outcome variable so that my coefficient of interest represents a percentage change.^{50,51} The treatment variable $\text{First AIDS Case}[t=l]_{mst}$ is an indicator which equals 1 when an observation represents an MSA, m , in season, s , that is l half-years relative to the first report of an AIDS case. l ranges from 6 half-years (3 years) before and 8 half-years (4 years) after the first reported AIDS case. In addition to controlling for period fixed effects (γ_t) and MSA fixed effects (a_m), I also control for season-MSA fixed effects (δ_{sm}) to account for any seasonal variation in personal ads. In order to account for heterogeneity in treatment timing and effects, I employ the imputation estimator developed by [Borusyak et al. \(2024\)](#). Standard errors are clustered at the MSA-level.

Estimates for [Equation 2](#) are presented in the first panel of [Figure 8](#). The event study suggests that there is little evidence that the first reported AIDS case affects the total number of ads posted.⁵² Thereafter, I explore effects on the content of personal ads. The second panel of [Figure 8](#) presents the effect on the number of personal ads which mention “safety”. This suggests that individuals respond to an increase in the perceived cost of risky sexual behavior brought about the first reported AIDS case by expressing a preference for “safe-sex”.⁵³

⁴⁸I use the average because there is some variation in the total number of issues per each 6-month period.

⁴⁹Season, s , refers to whether the observation represents the first six months of the year or whether the observation represents the last months of the year

⁵⁰I add one before applying the log so that observations equaling zero are not dropped.

⁵¹In my preferred specification, I do not use a per capita measure for my outcome variable. This is because [Chen and Roth \(2024\)](#) shows that log estimates are sensitive to the unit of measurement used. In my case, this means that per capita and per 1 million population measures yield significantly different coefficients. Nonetheless, I show that estimates are qualitatively similar using a per 1 million population measure for the outcome variable and weighting by population.

⁵²Although there is some evidence of differences in pre-trends, it is important to note that the interpretation of pre and post coefficients are different for the Imputation Estimate and a regular TWFE estimator. In this case, the reference period for the periods before treatment is period -7 instead of period -1 and the reference period for the post period is the average of all the pre-periods rather than period -1. This means that while the coefficients for period -3 to -1 are positive and statistically different from period -7, they are not very different from one another. It also implies that there is not a significant fall in coefficient between period 0 and period -1 as would be the case if interpretation an event study graph for a regular TWFE estimator.

⁵³Some researchers have expressed concerns with using log-like transformations to evaluate percentage effects when the outcome variable often equals zero. This is particularly relevant when evaluating effects on “safe-sex” ads where treatment may increase the outcome variable from zero to a positive number ([Chen and Roth, 2024](#)). As depicted in the third panel of [Figure 4](#), prior to the first reports of AIDS, only a small proportion of ads mention safety. [Chen and Roth \(2024\)](#) recommends using poisson regressions. Although poisson regressions are incompatible with the preferred imputation estimator specification, I present results from a poisson TWFE model in [Figure A.5](#). The estimates are largely similar to my main specification. The third panel of [Figure A.6](#) also presents the proportion of ads which mention safe-sex for MSAs who reported their first AIDS case in 1981 and MSAs who reported their first AIDS case later on. It clearly shows that MSAs which had first reports of AIDS in the early years of the virus experienced an increase in the proportion of safety ads much earlier than later treated MSAs. This lends further credibility to my main finding.

I also explore effects on ads seeking long-term relationships or ads which are explicitly sexual in the third and fourth panel of [Figure 8](#) and find limited effects. Although I find a significant increase in long-term relationship ads and a decrease in explicitly sexual ads in 1983-1984, these changes don't appear to be causally linked to the first reported case of AIDS in an MSA. This is in contrast to findings in [Spencer \(2024\)](#) and [Shahid \(2024\)](#) who find that individuals respond to HIV/AIDS risk by adopting monogamous different-sex relationships. I do not find evidence that the men posting personal ads respond to HIV/AIDS risk by seeking long term relationships. However, these findings must be interpreted with caution. Gay men may well have adopted long-term partnerships in response to HIV/AIDS risk but these changes may have not been captured by the data.⁵⁴ Additionally, seeking long-term relationships in the personals section also represents moderate levels of STI risk. It is possible that individuals who were previously seeking long-term relationships stop doing so after the first reported AIDS case. Given that my estimation strategy compares changes in personal ads in MSAs that have reported their first AIDS case to MSAs that have not yet reported their first AIDS case, we cannot rule out anticipatory effects. Individuals in MSAs which do not have a report of a positive AIDS case may wish to establish monogamous long term relationships as a risk mitigation tool in the future. To explore this more closely, the first two panels of [Figure A.6](#) present the proportion of relationship ads and explicitly sexual ads in MSAs which reported their first AIDS case in 1981 and MSAs which reported their first AIDS case after 1981. I find that a larger proportion of ads in later treated MSAs represent a preference for long term relationships.⁵⁵ I also find that there is a significant increase in the proportion of long-term relationship oriented ads between 1982 and 1984 for both early treated MSAs and later treated MSAs. More research is required in order to discern whether gay men responded to HIV/AIDS risk by adopting long-term monogamous relationships.

I then present estimates on the effects of the first reported AIDS case on anal insertive and anal receptive ads in the first two panels of [Figure 9](#). The event studies follow a similar trend and I find no evidence of a reduction in anal receptive ads despite the higher risk of HIV/AIDS transmission. The third panel of [Figure 9](#) shows the estimated effect of the first reported AIDS

⁵⁴For example, gay men may adopt long term-relationships with partners they meet outside of personal ads or their changed preference for long-term relationships may not be captured in their personal ads.

⁵⁵This indicates that there may be important differences in the ways in which individuals used these ads by MSA. In general, the earliest treated MSAs were also major hubs of gay populations and individuals may not need to use personal ads when searching for long term relationships.

case on the number of ads offering sex for sale, with no significant effects observed.

To summarize the event study estimates into a single estimate, I also estimate the following equation, which combines both pre and post periods:

$$\text{Log}(\text{Number of Personals} + 1)_{mst} = \alpha + \beta_l \text{First AIDS Case}_{mst} + \gamma_t + a_m + \delta_{sm} + \epsilon_{mt} \quad (3)$$

Now, $\text{First AIDS Case}_{mst}$ is an indicator variable that is equal to one for all periods following the first reported AIDS case in an MSA. All other features of [Equation 3](#) are identical to [Equation 2](#).

Estimates for [Equation 3](#) are provided in [Table 2](#). [Table 2](#) suggests that the first AIDS case increases the number of personal ads seeking “safe-sex” by approximately 6%. I find little evidence that the first AIDS case impacts the total number of personal ads, the number of long-term relationship oriented ads or the number of explicitly sexual ads.

Overall, I find that individuals do not significantly change the type of sexual partnerships they seek in the personal ads; however, there is an increased emphasis on safety.

4.3 The Intensive Margin

This paper argues that given the long incubation period of HIV, the first reported AIDS case in an MSA represents an exogenous increase in the perceived cost of risky sexual behavior. A similar argument could be made for AIDS incidence more generally. Although there may be non-linearities in the relationship between AIDS incidence and the perceived cost of risky sexual behavior, even after the first reported case of AIDS, individuals may exhibit greater fear of contracting the virus when there is a higher rate of AIDS in their MSA. In this section, I account for this intensive margin.

Unlike the first reported AIDS case in an MSA, AIDS incidence evolves dynamically. Therefore, I rely on a two-way fixed effects specification rather than an event-study model. Formally, I estimate the following equation:

$$\text{Log}(\text{Number of Personals} + 1)_{mst} = \alpha + \beta_l \text{ stdz(AIDS Rate)}_{mst} + \gamma_t + a_m + \delta_{sm} + \epsilon_{mt} \quad (4)$$

The treatment variable $\text{stdz(AIDS Rate)}_{mst}$ now represents the standardized form of the AIDS case rate in MSA, m , in season, s , in half-year, t . I use the standardized form of the AIDS case rate for ease of interpretation. All other features of this equation are identical to my main specification.

Estimates for [Equation 4](#) are provided in [Table 3](#). The estimates suggest that a 1 standard deviation increase in AIDS rate is associated with a 7 percentage point increase in the number of safety ads. There is weaker evidence that higher rates of AIDS results in small decreases in the overall number of personal ads, the number of long-term relationship oriented ads and explicitly sexual ads. These estimates are largely in line with my main specification.

5 Robustness

5.1 Drop Earliest Treated

To ensure that the main findings are not driven by the unique characteristics of the earliest affected cities, I re-estimate my model after excluding metropolitan statistical areas (MSAs) that reported their first AIDS case in 1981. This is important for several reasons. Firstly, earliest treated cities are systematically different from later treated cities. These early-treated cities, like San Francisco and New York, were major centers of the gay community and faced the brunt of the HIV/AIDS epidemic, leading to potentially different behavioral responses compared to later-treated cities. Secondly, being the first in the country to experience cases of the virus may have triggered a distinct response compared to cities which faced the epidemic later on. Thirdly, a large proportion of the personal ads represent earliest treated MSAs.⁵⁶ Estimates for [Equation 3](#) after dropping MSAs which experienced their first AIDS case in 1981 are presented in the first four columns of [Table 4](#). Although the estimates are similar to my main specification in direction they have significantly different magnitudes. My main specification suggests that

⁵⁶Approximately 80% of personal ads are from MSAs which experienced their first AIDS case in 1981.

individuals respond to the first reports of AIDS by increasing the number of safety ads posted by 6% but my restricted sample suggests that safety ads increase by approximately 1.5%. This suggests that behavioral shifts were much larger in the earliest treated MSAs. The earliest treated MSAs also experienced higher rates of HIV/AIDS thereafter. These findings are in line with estimates in [Table 3](#) which explores the intensive margin and finds that higher rates of AIDS are associated with larger behavioral responses.

5.2 Alternative Treatment

In my main specification, I exploit variation in the timing of the first reported AIDS case. An alternative approach would be to simply compare personal ads in MSAs that were more affected by HIV/AIDS to MSAs which were less affected.⁵⁷ Formally, I estimate the following equation:

$$\text{Log}(\text{Number of Personals} + 1)_{mst} = \alpha + \beta_l \text{ stdz(Pooled AIDS Rate)}_m + \gamma_t + a_m + \delta_{sm} + \epsilon_{mt} \quad (5)$$

All features of this equation are identical to my main specification, [Equation 3](#), but the treatment variable, $\text{stdz}(\text{Pooled AIDS Case Rate})_m$ now represents the standardized form of the pooled AIDS rate from 1981-1985 for each MSA, m .^{58,59} Estimates for [Equation 5](#) are provided in columns 5 to 7 of [Table 4](#). The estimates suggest that 1 standard deviation increase in pooled AIDS rate is associated with a 8% increase in the number of safety ads. Similar to my main specification, I find no evidence of effects on other types of ads.

6 Panel Analysis

The trends documented in the study represent behaviors of a changing sample population. Changes in the content of the ad could either represent advertisers who have previously posted in the magazine recognizing differences in risk and deciding to post a different ad in response, or

⁵⁷[Shahid \(2024\)](#) uses a similar methodology to estimate the effect of the virus race on marriage rates.

⁵⁸I use the standardized form for ease of interpretation.

⁵⁹Pooling AIDS Rate over multiple years rather than using the rate for a specific year allows for greater precision given the large amount of variation in AIDS Rate over time.

it could represent changes in the composition of individuals posting ads. The HIV/AIDS crisis was accompanied with many changes which may have impacted how men meet other men. For example, in 1984, in order to curb the spread of HIV/AIDS, the city of San Francisco closed its bathhouses ([Binson and Woods, 2013](#)). Other cities took similar measures. These changes might have led to a shift in the composition of individuals placing personal ads in the newspaper and may not be reflective of changes in behaviors.

To explore whether this is the case, I use phone numbers and P.O. box numbers shared in the ads to track individual advertisers over time. Most individuals who share this identifying information in their personal ad can be tracked over multiple issues.^{[60](#)} Although I am able to track individuals over multiple issues of the advocate, I can only follow them over short periods of time. The third panel of [Figure A.7](#) shows the time difference between the first and last time I observe an individual in my dataset. Most individuals can only be followed over less than 6 months.^{[61](#)} I have a smaller sample of individuals that I can follow over longer periods of time. Although these sample sizes are too small to conduct causal analysis similar to my main specification, I can create smaller subsets of consistent populations whom I follow over time.

Therefore, I create 9 subsets of my data of consistent groups of individuals who post multiple ads that are over 1 year apart. For example, the first group includes all individuals who post an ad on the second half of 1980 or the first half of 1981 and the second half of 1982 or the first half of 1983.^{[62](#)} I then plot changes in the content of the personal ads of these consistent groups over time in [Figure 10](#). I use a different line for each group and the size of each dot is associated with the number of individuals who represent the group. My methodology yields 9 consistent groups with each group consisting of at least 170 individuals. In these graphs, slopes of individual lines represent changing content of ads within a consistent group of individuals and jumps between lines represent across group differences in ads.

The first panel of [Figure 10](#) shows trends in the proportion of individuals mentioning safety

^{[60](#)}The first panel of [Figure A.7](#) presents the total number of ads which include no form of identification, a phone number, a P.O. Box Number, or both. The figure shows that approximately 60% of ads include some form of identification. The second panel of [Figure A.7](#) shows the number of times I observe the identification across my sample of ads. I find that most advertisers post ads multiple times.

^{[61](#)}There are several reasons why I only observe individuals for short periods of time. They may only actively use the personals section for a short period of time. Individuals also regularly move and change P.O. Box and phone numbers which makes it impossible to track future posts.

^{[62](#)}The second group includes all individuals who post an ad in the second half of 1981 or the first half of 1982 and the second half of 1983 or the first half of 1984. Other groups are constructed using a similar strategy.

terms in their ads for each group. This graph follows a similar trend to the second panel of [Figure 4](#) which presents the overall proportion of safety ads over time. This suggests that we can rule out that the increase in the proportion of personal ads mentioning safety was driven solely by different groups of individuals who started posting safety related ads. We see clear evidence of individuals who have used the personals section in the past adjusting their ads over time. The second and third panel of [Figure 10](#) present trends in the proportion of ads which are long-term relationship oriented or are explicitly sexual. Although the combination of lines in these graphs follow a similar trend to the overall proportion of long-term relationship oriented ads and explicitly sexual ads presented in the first panel of [Figure 4](#), most of the lines are flat and there are significant jumps between the different groups. This indicates that while there may be an overarching trend in long-term relationship oriented ads and explicitly sexual ads over time, much of this change is likely driven by shifts in the composition of individuals placing the ads rather than changes within a consistent group of advertisers.

7 Sexually Transmitted Infection (STI) Dataset

Although this paper documents a significant increase in the proportion of safety ads in response to the HIV/AIDS crisis, changes in the content of personal ads are not necessarily indicative of changes in behavior. This is particularly relevant given that most ads which mention ‘safety’, do not specify specific risk mitigation strategies they wish to employ.⁶³ In this section, I compare the take-up of safety language in the personals to changes in rates of sexual transmitted infections to test whether the trends in the personal ads are indicative of behavioral changes.

7.1 Data

Although publicly available STI data exists for more recent years, I acquired the STI data specific to my years of analysis through a special request from the Centers for Disease Control (CDC). This request granted me access to state-level rates of gonorrhea and syphilis for the years 1963 onwards.⁶⁴ Since I do not have access to STI rates at a more granular geographic

⁶³[Figure A.8](#) shows the proportion of safety ads with the term used to identify the ad as a safety ad. Most ads are identified as safety ads with the term ‘safe’. Very few of these ads mention the use of condoms.

⁶⁴Rates of chlamydia are only available after 1996. I am also unable to differentiate between primary or secondary syphilis during my period of analysis.

level, I am unable to use my preferred estimation strategy to estimate the causal effects of HIV/AIDS on other STI rates.⁶⁵ Therefore, I compare state-level changes in the take-up of safety language during the early years of the virus to changes in rates of STIs to test whether the changing behaviors reported in the personal ads correspond to lower rates of STIs.

I present national trends in rates of syphilis and gonorrhea per 100,000 population by sex in [Figure 11](#). The figure suggests that prior to the HIV/AIDS crisis, rates of syphilis were trending upwards for men but there was a significant decline following the HIV/AIDS crisis.⁶⁶ Gonorrhea, on the other hand, was already trending downward prior to 1981 and continued to do so after the first reports of AIDS

7.2 Analysis

Men who have sex with men (MSMs) account for a disproportionate share of syphilis cases. Although national statistics on STI rates by sexual behaviors are not available for this period, many studies document that MSMs represent a significantly larger share of Syphilis cases than other STIs. For example, [Judson et al. \(1980\)](#) documents that MSMs accounted for approximately 60% of Syphilis infections and 35% of Gonorrhea infections. Other research also consistently finds that policies affecting sexual minority men impact rates of syphilis but not other STIs ([Nikolaou, 2023a; Nikolaou, 2023b; Dee, 2008](#) [Francis et al. \(2012\)](#)).

The first panel of [Figure 12](#) compare state-level changes in rates of syphilis among men to changes in proportion of safety ads. To estimate this change, I pool together rates of STIs over three years before the start of the HIV/AIDS crisis (1978-1981) and three years after most cities have reported their first AIDS case (1984-1987). I pool together the proportion of safety ads in the same way.⁶⁷ The figure shows that states which experienced a greater take-up of safety language in the personal ads also experienced the greatest declines in rates of syphilis. This indicates that changes in the content of personal ads represent broader shifts in behavior. In the second and third panel of [Figure 12](#), I compare changes in women's syphilis rate and men's gonorrhea rate to changes in the proportion of safety ads. I find no evidence that these

⁶⁵County level rates of Syphilis do not become available till 1984 and county level rates of gonorrhea and chlamydia in 1995 and 1996.

⁶⁶Others have documented falling rates of STIs among MSMs during this period. CITE

⁶⁷Pooling over multiple time periods allows for greater precision given the large amount of variation in STI rates and safety ads in the personals.

are related. To ensure that the negative relationship between safety language in the personal ads in the early 1980s and syphilis rate is not the result of pre-existing trends, I conduct a placebo test where I change the vertical axis to represent percentage changes in men’s syphilis rate between periods before the first reports of AIDS (I compare changes in men’s syphilis rates from the period 1971-1974 to 1977-1980). I find no evidence that the take-up of safety language in personal ads during the early years of the HIV/AIDS crisis is associated with changes in rates of syphilis in the 1970s.

8 Conclusion

In this study, I explore behavioral responses to the increasing cost of risk in the context of the HIV/AIDS crisis. I use Optical Character Recognition (OCR) and machine learning to construct a unique rich dataset of over 170,000 men’s personal ads posted in the oldest and largest national LGBT publication in the United States, from 1975 to 1992. I first document trends in the overall number of ads posted overtime. I show that the number of ads increases in the years leading up to the first reports of HIV/AIDS in mid-1981. Ads continue to rise and peak in late 1983 and are followed by a large decline thereafter. Thereafter, I employ machine learning techniques and search for specific terms to analyze how the content of the ads evolves overtime. Most notably, I find a significant increase in the number ads which include specific ‘safety’ language

During the early years of the virus, the first reports of AIDS represented an exogenous increase in the cost of risky sexual behavior, as infections could only be detected in the late phases of the virus (approximately 10 years after infection). Therefore, I exploit variation in the timing of the first reports of AIDS in a Metropolitan Statistical Areas to estimate the causal effect of a change in the cost of risk on behavior. To ensure that the first AIDS case represents an increase in the perceived cost of risky sexual behavior, I demonstrate that the first reported AIDS case is associated with a significant increase in local HIV/AIDS reporting. This could either represent heightened concerns about HIV/AIDS or could induce fears of contraction. In either case, it signifies an increase in the perceived cost of engaging in risky sexual behavior. Thereafter, I use an event-study difference-in-differences design to estimate the causal effect of the first reported AIDS case on personal ads. I show that while the first reported AIDS case

appears to have a limited impact on the total number of ads posted, there is a significant increase in the number of ads which mention “safe-sex”. I also find no evidence of individuals switching away from certain high-risk sexual activities such as being the anal receptive partner. This suggests that despite the rising costs associated with risky sexual behavior, individuals continued to use the ads to seek similar partnerships and encounters, placing a greater emphasis on safety measures to mitigate the risk of HIV/AIDS transmission.

The findings of this paper are similar to other settings which show that people often do not significantly alter their behaviors in response to changes in risk, particularly when alternative risk mitigation strategies are available. For example, some studies have documented the greater effectiveness of comprehensive sex education in schools compared to abstinence-only education ([Stanger-Hall and Hall, 2011](#); [Mark and Wu, 2022](#)). Similarly, [Pesko et al. \(2020\)](#) finds that individuals are more likely to respond to increasing cigarette costs by switching to e-cigarettes, which represent lower health risks, rather than quitting smoking altogether. Understanding specific behavioral responses to risks broadens our understanding of how individuals manage risks and can guide the development of more effective, targeted interventions.

9 Figures

Figure 1 : Page From Personals

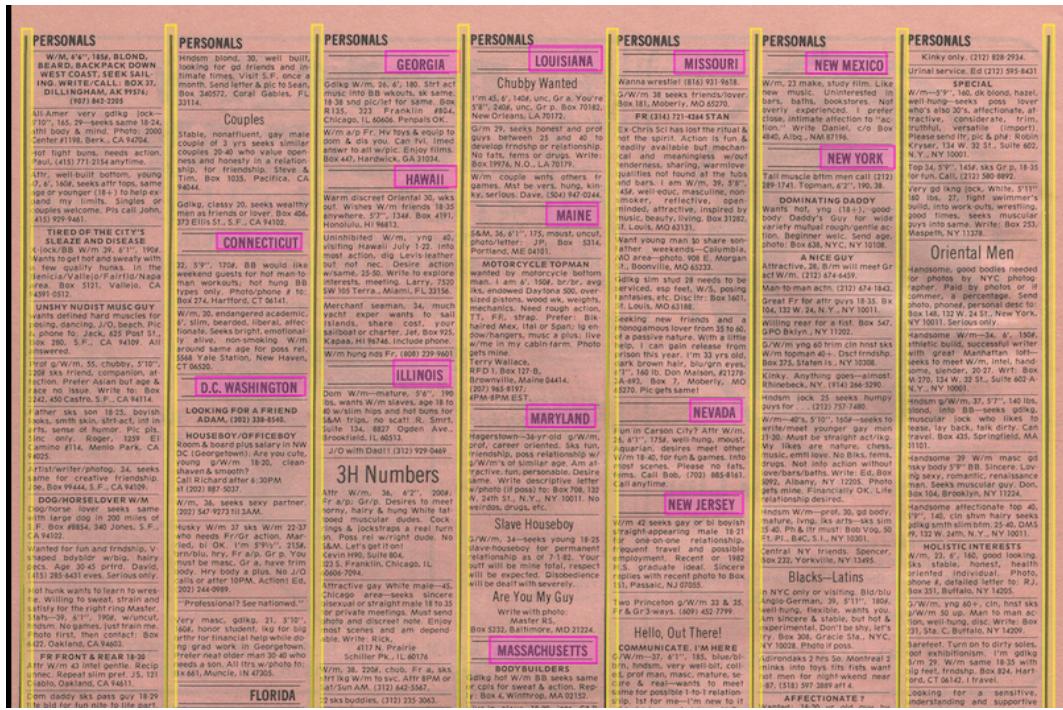


Figure 1: Source: Advocate Personal Ads

Notes: This figure presents a page from *the Advocate* personals section.

Figure 2 : Total Ads Over time

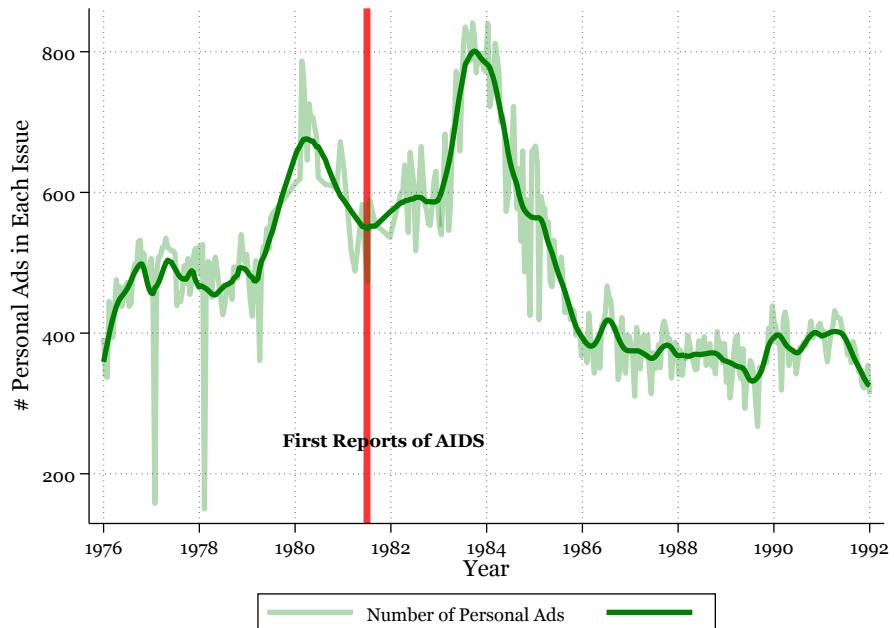


Figure 2: Source: Advocate Personal Ads

Notes: This figure represents the total number of ads posted in each issue of *the Advocate* personals section over time. The light line represents the exact number of ads in an issue while the darker line represents a smoothed version obtained from a locally weighted regression.

Figure 3 : Total Ads by Type Over time

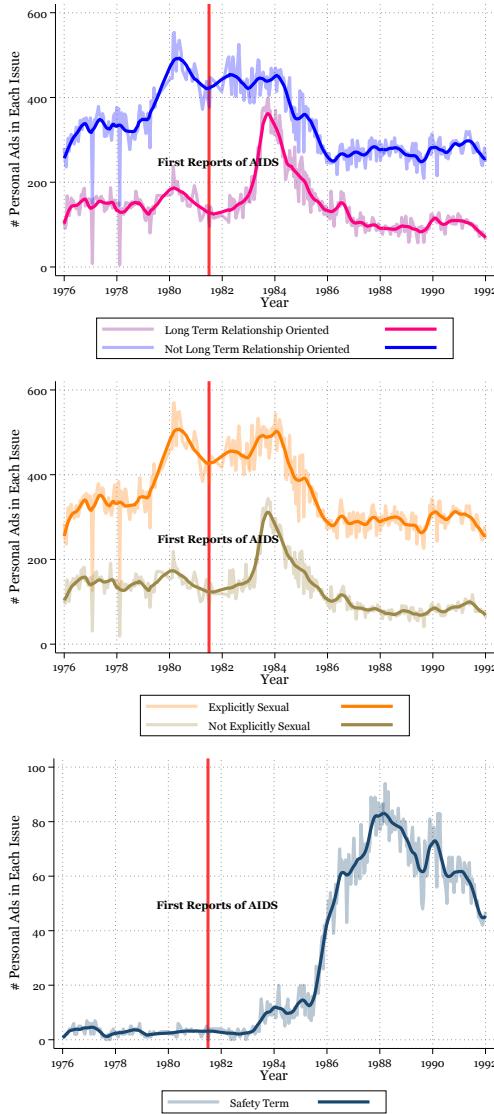


Figure 3: Source: Advocate Personal Ads

Notes: These figures represent the number of a particular type of ad posted in each issue of *the Advocate* personals section over time. The light line represents the exact number of ads in an issue while the darker line represents a smoothed version obtained from a locally weighted regression. The figure figure shows trends in the number of ads that are long term relationship oriented and the number of ads that are not long-term relationship oriented. The second figure shows trends in the number of ads that are explicitly sexual and the number of ads that are not explicitly sexual. The third figure shows trends in the number of ads that include a 'safety term' and the number of ads that do not.

Figure 4 : Proportion of Ads by Type Over time

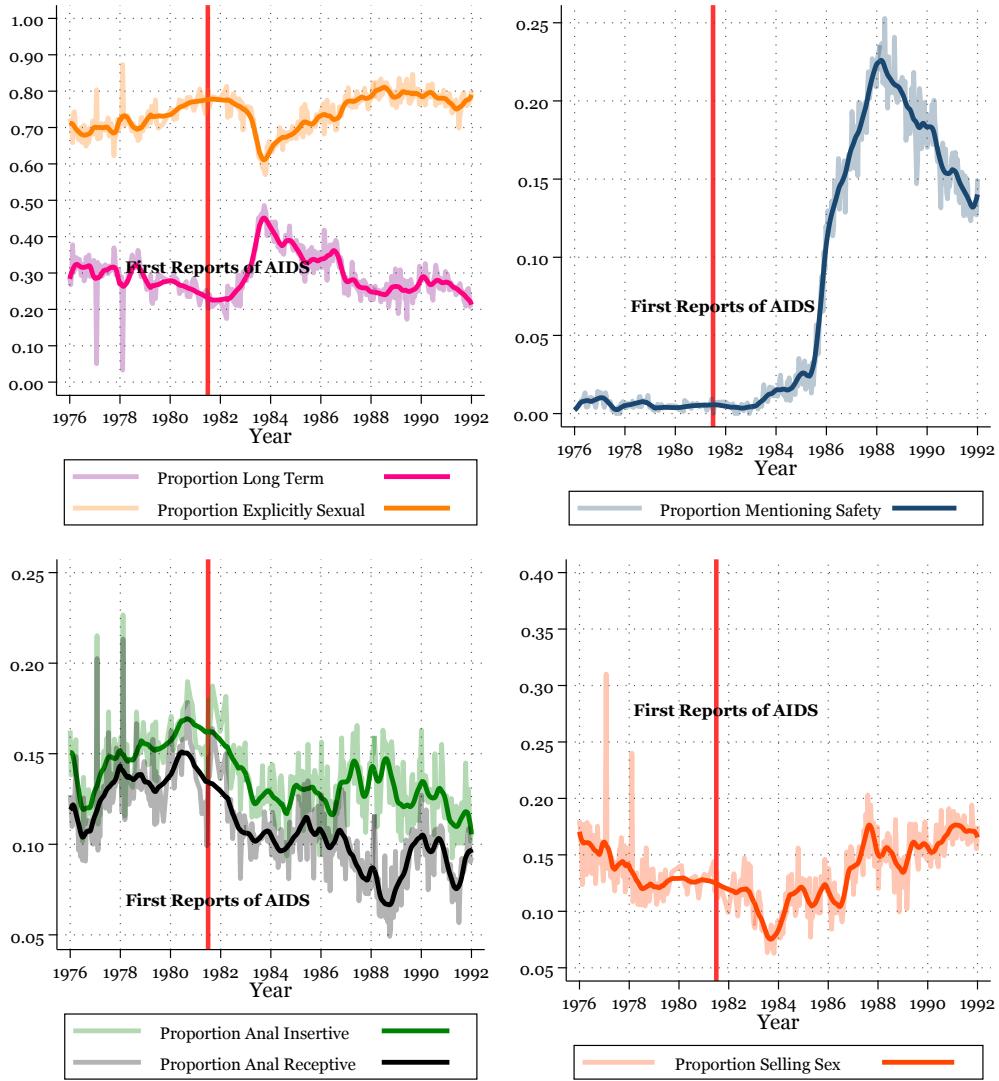


Figure 4: Source: Advocate Personal Ads

Notes: These figures represent the proportion of a particular type of ad posted in each issue of *the Advocate* personals section over time. The light line represents the exact proportion of ads in an issue while the darker line represents a smoothed version obtained from a locally weighted regression. The first figure represents the proportion of ads which are long term relationship oriented in and explicitly sexual overtime. The second figure depicts the proportion of ads which include a safety term. The third figure depicts the the proportion of ads seeking anal insertive sex and anal receptive sex. The fourth figure depicts the proportion of ads which represent the sale of sex.

Figure 5 : AIDS Public Information Dataset (APID)

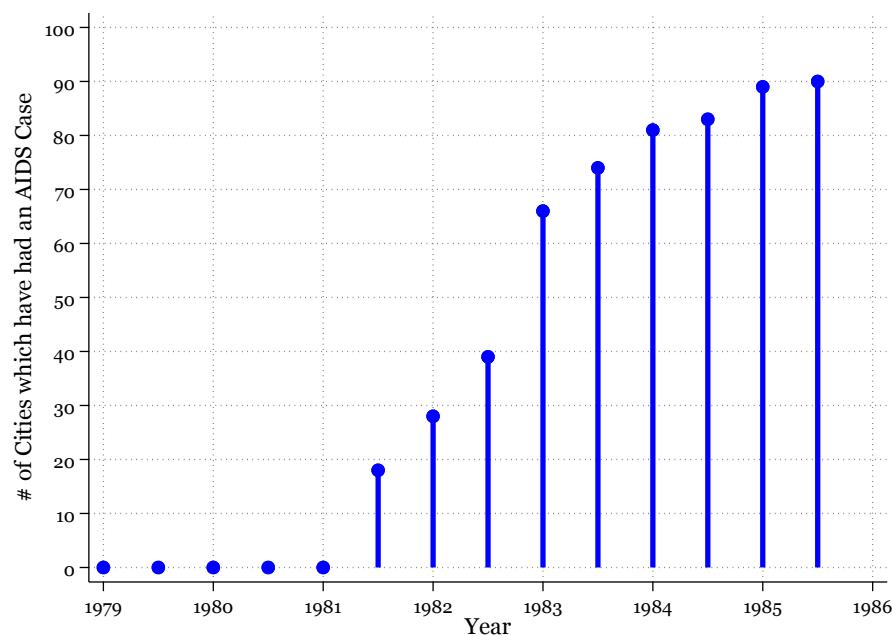


Figure 5: Source: [US Department of Health and Human Services \(US DHHS\) \(2005\)](#)

This figure depicts the total number of cities which have reported an AIDS case in the APIDs dataset by half-year.

Figure 6 : AIDS News

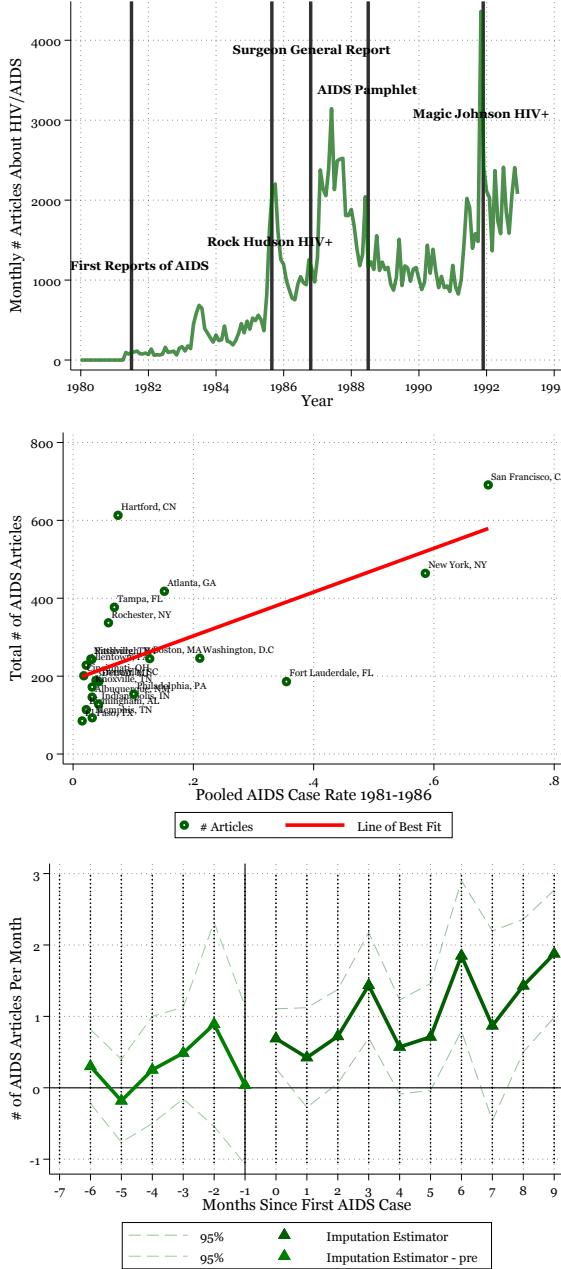


Figure 6: Source: ProQuest Historical Newspapers and AIDS Public Information Dataset

Notes: The first panel depicts the total number of HIV/AIDS related articles in the local newspapers listed in [Table A.3](#). The second panel depicts a scatter plot of the total number of HIV/AIDS articles in each MSA against the pooled AIDS case rate from 1981-1986. The third panel depicts difference-in-differences event study estimates from [Equation 1](#) measuring the effect of the first reported AIDS case on local newspaper reports about HIV/AIDS using the Imputation Estimator developed by [Borusyak et al. \(2024\)](#). Estimates control for period fixed effects and MSA fixed effects. Standard errors are clustered at the MSA level.

Figure 7 : Examples of Local News Reports

The figure consists of five separate newspaper clippings, each with a different header and some text. The first four clippings have yellow highlights around their titles.

Vermont's First AIDS Case Reported
By TED TEDFORD them to be on the watch for potential AIDS victims might seek treatment.

First AIDS Case In Del. Reported In Rehoboth
By MICHAEL MILLS Times Sussex Bureau because most of the cases have been reported in the past few months and death is usually slow

First AIDS case confirmed
MADISON, Wis. (AP) — The first confirmed case of AIDS has been reported at the Veterans Administration Hospital, associate director Robert Stott said Friday. Stott would release no further details about the case. Previously, two people with symptoms of Acquired Immune Deficiency Syndrome were being watched in the city, according to state epidemiologist Jeffrey Davis. AIDS, a mysterious disease which destroys the body's immunity system and leaves it open to a host of life-threatening conditions has primarily affects homosexual men, intravenous drug users, hemophiliacs and Haitians. More than 2,000 cases have been reported nationwide to the Centers for Disease Control during the past two years.

First Iowa victim of AIDS found
By BILL GORDON

County resident is AIDS victim
Shasta County has its first reported case of Acquired Immune Deficiency Syndrome (AIDS), officials confirmed today. said. There is no known cure to the disease, doctors say.

Figure 7: The figure presents examples of local newspaper reports about the first local reported AIDS case. First Panel: *Burlington Free Press*, 1983. Second Panel: *The Daily Times*, 1983. Third Panel: *The Post-Crescent*, 1983. Fourth panel: *Des Moines Register*, 1983. Fifth Panel: *Record Searchlight*, 1983.

Figure 8 : Event Studies

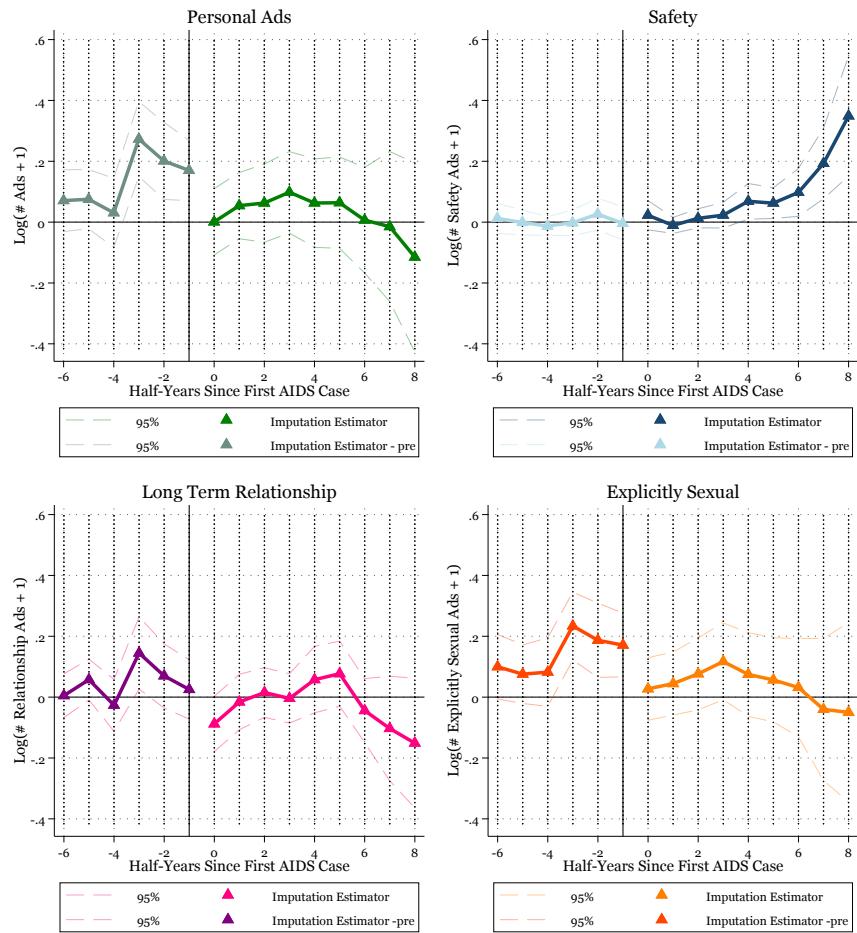


Figure 8: Source: Advocate Personal Ads and AIDS Public Information Dataset

Notes: These figures depict event study estimates of Equation 2 measuring the effect of the first reported AIDS case on personal ads using the Imputation Estimator developed by (Borusyak et al., 2024). The first figure depicts effects on the total number of personal ads. The second figure depicts effects on the number of ads which include safety terms. The third figure depicts effects on the number of ads which are long-term relationship oriented. The fourth figure depicts effects on the number of ads which are explicitly sexual. All estimates control for period fixed effects, and season-MSA fixed effects. Standard errors are clustered at the MSA-level.

Figure 9 : Additional Event Studies

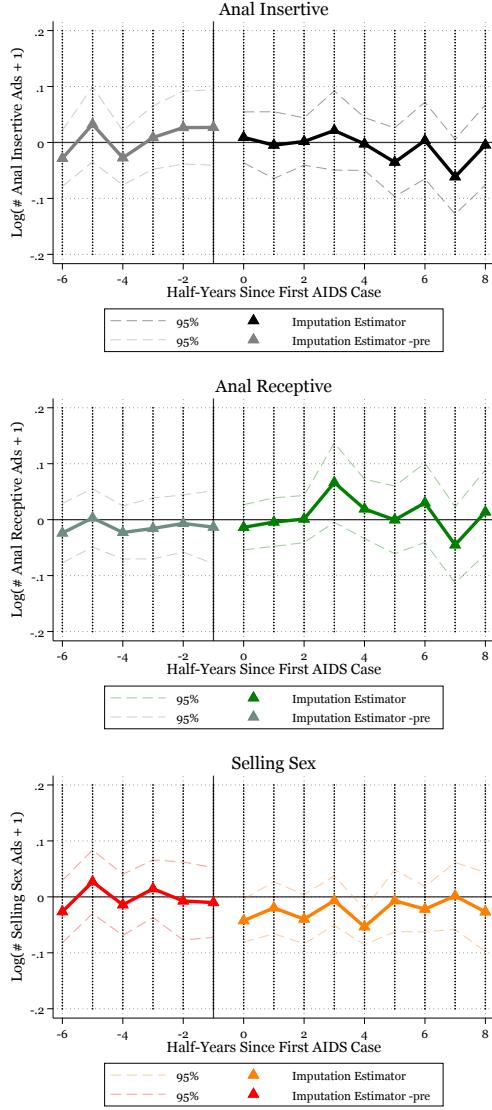


Figure 9: Source: Advocate Personal Ads and AIDS Public Information Dataset

Notes: These figures depict event study estimates of [Equation 2](#) measuring the effect of the first reported AIDS case on personal ads using the Imputation Estimator developed by ([Borusyak et al., 2024](#)). The first figure depicts effects on the number of personal ads seeking anal insertive sex. The second figure depicts effects on the number of ads which seek anal receptive sex. The third figure depicts effects on the number of ads which represent the sale of sex. All estimates control for period fixed effects, and season-MSA fixed effects. Standard errors are clustered at the MSA-level.

Figure 10 : Panel Analysis

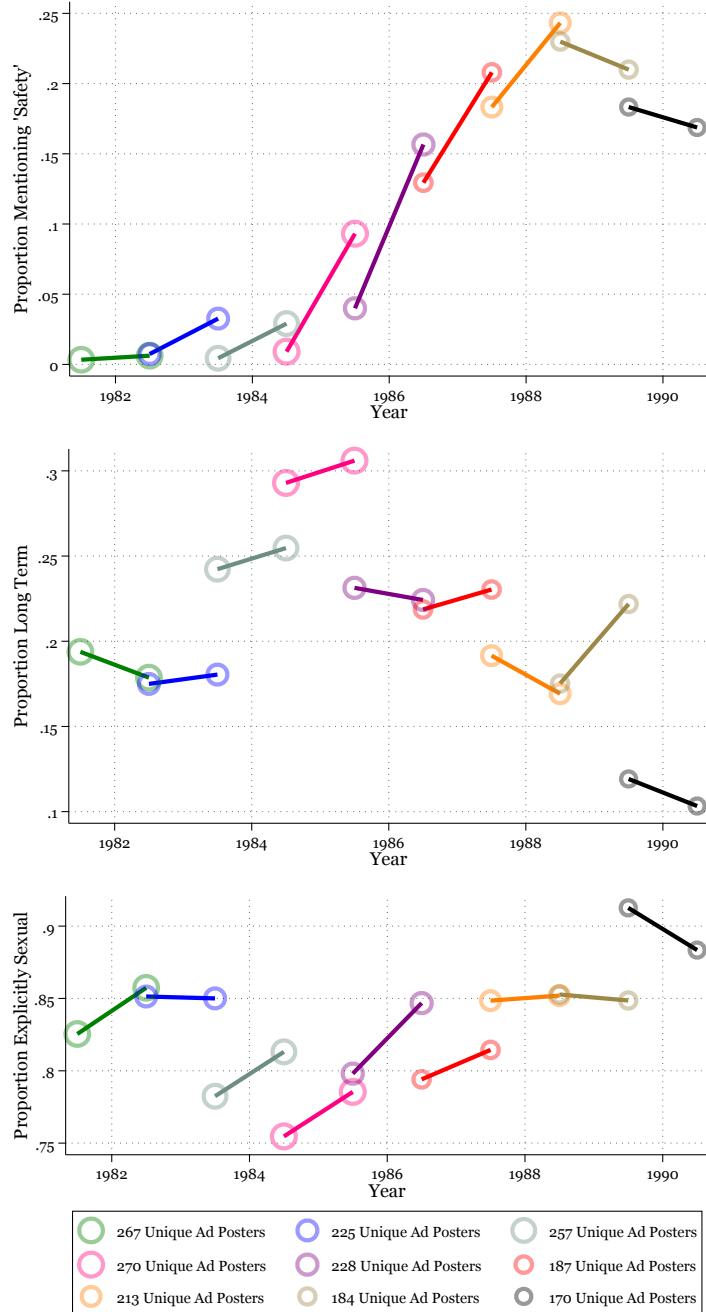


Figure 10: Source: Advocate Personal Ads

Notes: These figures depict trends in the proportion of certain types of personal ads over consistent sample populations. Each sample depicts a consistent sample of repeat posters. The first panel shows trends in the proportion of ads which include a safety term. The second panel shows trends in the proportion of ads which are long-term relationship oriented. The third panel depicts trends in the proportion of ads which are explicitly sexual.

Figure 11 : STI's over time

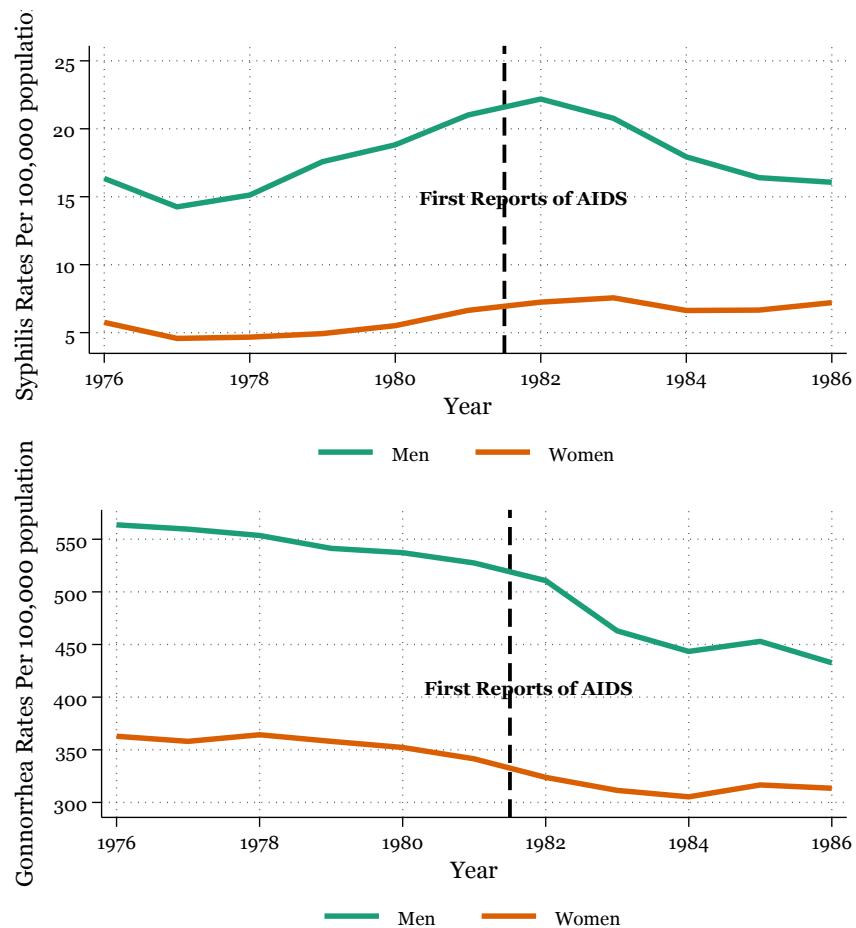


Figure 11: Source: CDC

Notes: These figures depict trends in the rates of sexually transmitted infections (STIs) by sex over time. The first panel depicts trends in rates of syphilis per 100,000 population for men and women overtime. The second panel depicts trends in the rates of Gonorrhea per 100,000 population for men and women overtime.

Figure 12 : Changes in STI rates

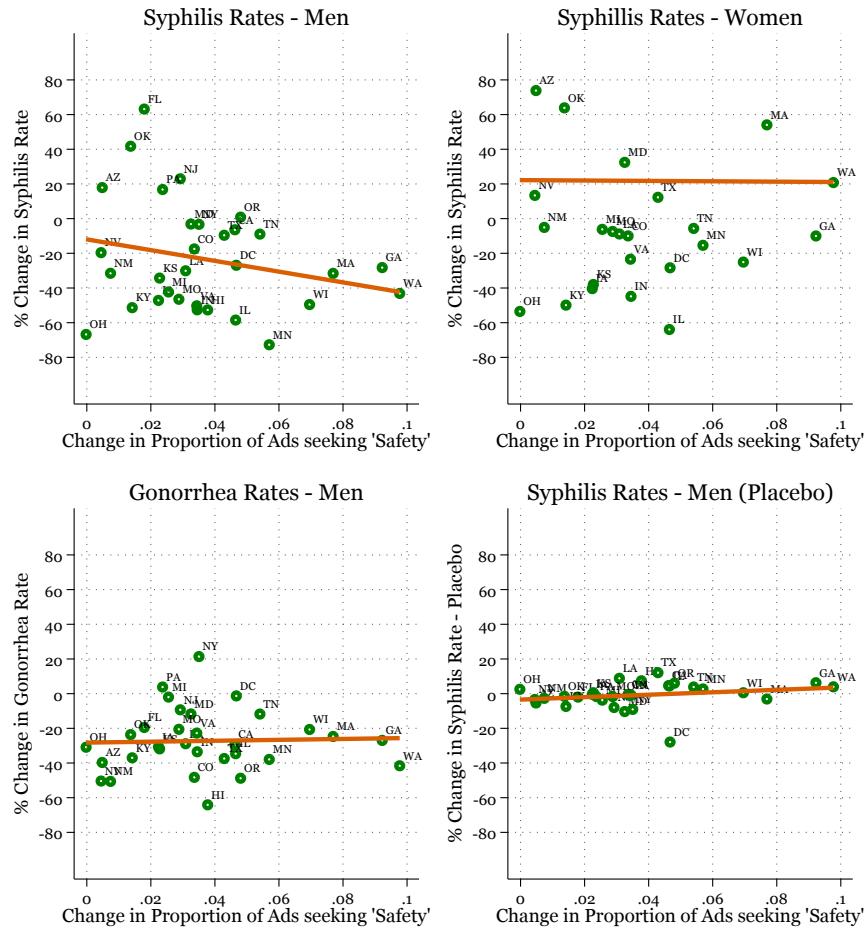


Figure 12: Source: CDC and Advocate Personal Ads

Notes: The first three figures depict scatter plots and lines of best fit for state-level percentage changes in rates of STIs to percentage changes in the proportion of ads which include safety terms from before the start of the HIV/AIDS crisis (1978-1981) to several years after the first reports of AIDS in the country (1984-1987). The first figure shows this comparison for men's syphilis rate. The second figure shows this comparison for women's syphilis rate. The third figure shows this comparison for men's gonorrhea rate. The fourth figure has the same x-axis as the previous figures but the y-axis now represents changes in rates of syphilis.

10 Tables

Table 1: Summary Statistics of Personal Ads

White	0.877 (0.328)
Black	0.063 (0.243)
Asian	0.032 (0.176)
Race Not Reported	0.375 (0.484)
Age	31.495 (10.699)
Age Not Reported	0.505 (0.500)
Long Term Relationship Oriented	0.307 (0.461)
Explicitly Sexual	0.728 (0.445)
Safety	0.069 (0.254)
Insertive Anal Intercourse	0.135 (0.341)
Receptive Anal Intercourse	0.108 (0.311)
Selling Sex	0.132 (0.338)
State- California	0.458 (0.498)
State- New York	0.246 (0.431)
State- Illinois	0.036 (0.187)
State- Florida	0.025 (0.157)
State- Texas	0.025 (0.156)
State - Not Identified	0.288 (0.453)
MSA - Los Angeles-Long Beach-Santa Ana, CA	0.357 (0.479)
MSA - New York-Newark-Jersey City, NY-NJ	0.252 (0.434)
MSA - San Francisco-Oakland-Hayward, CA	0.085 (0.279)
MSA - Chicago-Naperville-Elgin, IL-IN-WI	0.036 (0.186)
MSA - Miami-Fort Lauderdale-West Palm Beach, FL	0.023 (0.149)
MSA - Not Identified	0.288 (0.453)
Observations	176901

mean coefficients; sd in parentheses

Table 1: Source: Advocate Personal Ads

Notes: This table presents summary statistics from the all ads posted in the advocate from 1975-1992.

Table 2: Estimates from [Equation 3](#) - The Effect of First Local AIDS Case on Personal Ads

	(1) Personal Ads	(2) Safety	(3) Long Term Relationships	(4) Explicitly Sexual
First AIDS Case	0.00107 (0.062)	0.05663** (0.025)	-0.02769 (0.039)	0.00072 (0.057)
Observations	2000	2000	2000	2000
Period FE	Yes	Yes	Yes	Yes
MSA FE	Yes	Yes	Yes	Yes
Season-MSA FE	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Source: Advocate Personal Ads and AIDS Public Information Dataset

Notes: This table presents estimates from [Equation 3](#) measuring the effect of the first reported AIDS case on personal ads using the Imputation Estimator developed by ([Borusyak et al., 2024](#)). The first column depicts effects on the total number of personal ads. The second column depicts effects on the number of ads which include safety terms. The third column depicts effects on the number of ads which are long-term relationship oriented. The fourth column depicts effects on the number of ads which are explicitly sexual. All estimates control for period fixed effects, and season-MSA fixed effects. Standard errors are clustered at the MSA-level.

Table 3: Estimates from [Equation 4](#) - The Intensive Margin

	(1) Personal Ads	(2) Safety	(3) Long Term Relationships	(4) Explicitly Sexual
stdz(AIDS Case Rate)	-0.02521* (0.015)	0.06668*** (0.010)	-0.01907** (0.009)	-0.02370* (0.014)
Observations	2000	2000	2000	2000
Period FE	Yes	Yes	Yes	Yes
MSA FE	Yes	Yes	Yes	Yes
Season-MSA FE	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Source: Advocate Personal Ads and AIDS Public Information Dataset

Notes: This table presents estimates from [Equation 4](#) measuring the effect of 1 standard deviation increase in AIDS rate on personal ads using TWFE. The first column depicts effects on the total number of personal ads. The second column depicts effects on the number of ads which include safety terms. The third column depicts effects on the number of ads which are long-term relationship oriented. The fourth column depicts effects on the number of ads which are explicitly sexual. All estimates control for period fixed effects, and season-MSA fixed effects. Standard errors are clustered at the MSA-level.

Table 4: Robustness

	Drop First Treated - Imputation Estimator				Alternative Treatment - TWFE			
	(1) Personal Ads	(2) Safety	(3) Long Term Relationships	(4) Explicitly Sexual	(5) Personal Ads	(6) Safety	(7) Long Term Relationships	(8) Explicitly Sexual
First AIDS Case	0.02120 (0.080)	0.01402** (0.007)	0.00351 (0.040)	0.01429 (0.072)				
stdz(Pooled AIDS Case Rate)					0.03855 (0.039)	0.07933*** (0.026)	0.00133 (0.027)	0.03767 (0.034)
Observations	1720	1720	1720	1720	2000	2000	2000	2000
Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Season-MSA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Source: Advocate Personal Ads and AIDS Public Information Dataset

Notes: This first four column of this table presents estimates from [Equation 3](#) measuring the effect of the first reported AIDS case on personal ads using the Imputation Estimator developed by ([Borusyak et al., 2024](#)) for MSAs which reported their first AIDS case after 1981. The first column depicts effects on the total number of personal ads. The second column depicts effects on the number of ads which include safety terms. The third column depicts effects on the number of ads which are long-term relationship oriented. The fourth column depicts effects on the number of ads which are explicitly sexual. The next four columns of this table presents estimates from [Equation 5](#) measuring the effect of a 1 standard deviation increase in the pooled AIDS rate for the years 1981-1985 using a TWFE estimator. The fifth column depicts effects on the total number of personal ads. The sixth column depicts effects on the number of ads which include safety terms. The seventh column depicts effects on the number of ads which are long-term relationship oriented. The eighth column depicts effects on the number of ads which are explicitly sexual. All estimates control for period fixed effects, and season-MSA fixed effects and standard errors are clustered at the MSA-level.

References

- (1981). *Editor & Publisher International Year Book*. Editor & Publisher Co., New York, NY.
- Ahituv, A., Hotz, V. J., and Philipson, T. (1996). The responsiveness of the demand for condoms to the local prevalence of aids. *Journal of Human Resources*, pages 869–897.
- AIDS Foundation of Chicago (2024). Say his name: Robert rayford. Accessed: 2024-08-13.
- Altman, L. K. (1981). Rare cancer seen in 41 homosexuals. *The New York Times*. <https://www.nytimes.com/1981/07/03/us/rare-cancer-seen-in-41-homosexuals.html>.
- Angelucci, C., Cagé, J., and Sinkinson, M. (2024). Media competition and news diets. *American Economic Journal: Microeconomics*, 16(2):62–102.
- Atalay, E., Phongthiengham, P., Sotelo, S., and Tannenbaum, D. (2020). The evolution of work in the united states. *American Economic Journal: Applied Economics*, 12(2):1–34.
- Auld, M. C. (2006). Estimating behavioral response to the aids epidemic. *Contributions in Economic Analysis & Policy*, 5(1).
- Baker, P. (2003). No effeminate please: A corpus-based analysis of masculinity via personal adverts in gay news/times 1973–2000. *The Sociological Review*, 51(1_suppl):243–260.
- Baranov, V. and Kohler, H.-P. (2018). The impact of aids treatment on savings and human capital investment in malawi. *American Economic Journal: Applied Economics*, 10(1):266–306.
- Bartholome, A., Tewksbury, R., and Bruzzone, A. (2000). “i want a man”: Patterns of attraction in all-male personal ads. *The Journal of Men’s Studies*, 8(3):309–321.
- Binson, D. and Woods, W. J. (2013). *Gay bathhouses and public health policy*. Routledge.
- Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, page rdae007.
- Burlington Free Press (1983). May 26, 1983 (page 17 of 56). *Burlington Free Press (1923-)*. Accessed via Vanderbilt University Library.
- Callen, M. and Berkowitz, R. (1983). How to have sex in an epidemic. *News from the Front Publications, editor*.
- Callison, K. and Kaestner, R. (2014). Do higher tobacco taxes reduce adult smoking? new

- evidence of the effect of recent cigarette tax increases on adult smoking. *Economic Inquiry*, 52(1):155–172.
- Carpenter, C. and Dobkin, C. (2009). The effect of alcohol consumption on mortality: regression discontinuity evidence from the minimum drinking age. *American Economic Journal: Applied Economics*, 1(1):164–182.
- Centers for Disease Control and Prevention (1985). Current trends update: Acquired immunodeficiency syndrome – united states. *Morbidity and Mortality Weekly Report*, 34(18):245–248. Accessed: 2024-08-30.
- Chalfin, A., Hansen, B., and Ryley, R. (2023). The minimum legal drinking age and crime victimization. *Journal of Human Resources*, 58(4):1141–1177.
- Chan, T. Y., Hamilton, B. H., and Papageorge, N. W. (2016). Health, risky behaviour and the value of medical innovation for infectious disease. *The Review of Economic Studies*, 83(4):1465–1510.
- Chen, J. and Roth, J. (2024). Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics*, 139(2):891–936.
- Corno, L. and De Paula, Á. (2019). Risky sexual behaviours: Biological markers and self-reported data. *Economica*, 86(342):229–261.
- Cotti, C., Nesson, E., and Tefft, N. (2016). The effects of tobacco control policies on tobacco products, tar, and nicotine purchases among adults: Evidence from household panel data. *American Economic Journal: Economic Policy*, 8(4):103–123.
- Davidson, A. G. (1991). Looking for love in the age of aids: The language of gay personals, 1978–1988. *Journal of Sex Research*, 28(1):125–137.
- Davis, S. J., Hansen, S., and Seminario-Amez, C. (2020). Firm-level risk exposures and stock returns in the wake of covid-19. Technical report, National Bureau of Economic Research.
- De Walque, D. (2007). Sero-discordant couples in five african countries: implications for prevention strategies. *Population and development review*, 33(3):501–523.
- Dee, T. S. (2008). Forsaking all others? the effects of same-sex partnership laws on risky sex. *The Economic Journal*, 118(530):1055–1078.
- Delavande, A. and Kohler, H.-P. (2012). The impact of hiv testing on subjective expectations and risky behavior in malawi. *Demography*, 49:1011–1036.
- Dell, M. (2024). Deep learning for economists. Technical report, National Bureau of Economic Research.

Department of Health and Human Services (2005). Centers for disease control and prevention, national center for hiv, std and tb prevention, aids public information data set us surveillance data for 1981-2002. CDC WONDER On-line Database.

Des Moines Register (1983). June 10, 1983 (page 21 of 69). *Des Moines Register (1871-)*. Accessed via Vanderbilt University Library.

Doleac, J. L. and Mukherjee, A. (2022). The effects of naloxone access laws on opioid abuse, mortality, and crime. *The Journal of Law and Economics*, 65(2):211–238.

Engelberg, J. E. and Parsons, C. A. (2011). The causal impact of media in financial markets. *the Journal of Finance*, 66(1):67–97.

Francis, A. M., Mialon, H. M., and Peng, H. (2012). In sickness and in health: Same-sex marriage laws and sexually transmitted infections. *Social Science & Medicine*, 75(8):1329–1341.

Gay in the 80s (2013). 1982: Safer sex — play fair. <https://www.gayinthe80s.com/2013/08/1982-safer-sex-play-fair/>. Accessed: 2024-09-02.

Gonzales, M. H. and Meyers, S. A. (1993). "your mother would like me": Self-presentation in the personals ads of heterosexual and homosexual men and women. *SAGE Social Science Collections*.

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.

Hansen, B., Sabia, J. J., and Rees, D. I. (2017). Have cigarette taxes lost their bite? new estimates of the relationship between cigarette taxes and youth smoking. *American Journal of Health Economics*, 3(1):60–75.

Hatala, M. N., Baack, D. W., and Parmenter, R. (1998). Dating with hiv: A content analysis of gay male hiv-positive and hiv-negative personal advertisements. *Journal of Social and Personal Relationships*, 15(2):268–276.

Hatala, M. N. and Prehodka, J. (1996). Content analysis of gay male and lesbian personal advertisements. *Psychological Reports*, 78(2):371–374.

Hinnosaar, M. and Liu, E. M. (2022). Malleability of alcohol consumption: Evidence from migrants. *Journal of Health Economics*, 85:102648.

Judson, F. N., Penley, K. A., Robinson, M. E., and Smith, J. K. (1980). Comparative prevalence rates of sexually transmitted diseases in heterosexual and homosexual men. *American Journal of Epidemiology*, 112(6):836–843.

- Kerwin, J. T. (2020). Scared straight or scared to death? fatalism in response to disease risks. Technical report, mimeo.
- Kirsch, J. (1995). Dying young. *Los Angeles Times*. Accessed: 2024-08-13.
- Laner, M. R. and Kamel, G. L. (1978). Media mating i: "Newspaper" personals" ads of homosexual men. *Journal of Homosexuality*, 3(2):149–162.
- Lee, J. A. (1976). Forbidden colors of love: Patterns of gay love and gay liberation. *Journal of Homosexuality*, 1(4):401–418.
- Los Angeles Times (1985). Los angeles gays go on alert as aids epidemic widens. Accessed: 2024-09-02.
- Lumby, M. E. (1978). Men who advertise for sex. *Journal of Homosexuality*, 4(1):63–72.
- Mann, S. (2023). Pre-exposure prophylaxis and hiv incidence. *The BE Journal of Economic Analysis & Policy*, 23(4):1163–1173.
- Mark, N. D. and Wu, L. L. (2022). More comprehensive sex education reduced teen births: Quasi-experimental evidence. *Proceedings of the National Academy of Sciences*, 119(8):e2113144119.
- Miravete, E. J., Seim, K., and Thurk, J. (2020). One markup to rule them all: Taxation by liquor pricing regulation. *American Economic Journal: Microeconomics*, 12(1):1–41.
- National Institutes of Health (2024). The stages of hiv infection. <https://hivinfo.nih.gov/understanding-hiv/fact-sheets/stages-hiv-infection>. Accessed: 2024-09-02.
- Nesson, E. (2017). Heterogeneity in smokers' responses to tobacco control policies. *Health economics*, 26(2):206–225.
- Nikolaou, D. (2023a). Effects of marriage equality legislation on sexual health of the us population. *Health Economics*, 32(1):107–133.
- Nikolaou, D. (2023b). Same-sex marriage legalization and sexually transmitted infections across europe. *Oxford Bulletin of Economics and Statistics*, 85(1):35–69.
- Oster, E. (2005). Sexually transmitted infections, sexual behavior, and the hiv/aids epidemic. *The Quarterly Journal of Economics*, 120(2):467–515.
- Oster, E. (2012). Hiv and sexual behavior change: Why not africa? *Journal of health economics*, 31(1):35–49.
- Packham, A. (2022). Syringe exchange programs and harm reduction: New evidence in the wake of the opioid epidemic. *Journal of Public Economics*, 215:104733.
- Pesko, M. F., Courtemanche, C. J., and Maclean, J. C. (2020). The effects of traditional cigarette

- and e-cigarette tax rates on adult tobacco product use. *Journal of risk and uncertainty*, 60(3):229–258.
- Poorolajal, J., Hooshmand, E., Mahjub, H., Esmailnasab, N., and Jenabi, E. (2016). Survival rate of aids disease and mortality in hiv-infected patients: a meta-analysis. *Public health*, 139:3–12.
- Record Searchlight (1983). August 9, 1983 (page 12 of 44). *Record Searchlight (1975-)*. Accessed via Vanderbilt University Library.
- Ruhm, C. J. (2019). Drivers of the fatal drug epidemic. *Journal of health economics*, 64:25–42.
- Schilbach, F. (2019). Alcohol and self-control: A field experiment in india. *American economic review*, 109(4):1290–1322.
- Schmeckenbecher, J., Kapusta, N. D., Krausz, R. M., and Emilian, C. A. (2024). Autopsy rates and the misclassification of suicide and accident deaths. *European journal of epidemiology*, pages 1–18.
- Shahid, H. (2023). An antidote for despair: The effect of highly active antiretroviral therapy (haart) on suicide rates.
- Shahid, H. (2024). The impact of hiv/aids on marriage in the early years of the epidemic. *Health Economics*, 33(7):1415–1425.
- Smith, A. (2000). “safety” in gay men’s personal ads, 1985-1996. *Journal of Homosexuality*, 39(1):43–49.
- Soulakova, J. N., Hartman, A. M., Liu, B., Willis, G. B., and Augustine, S. (2012). Reliability of adult self-reported smoking history: data from the tobacco use supplement to the current population survey 2002–2003 cohort. *Nicotine & Tobacco Research*, 14(8):952–960.
- Spencer, M. K. (2024). Safer sex? the effect of aids risk on birth rates. *Journal of Health Economics*, 95:102867.
- St Lawrence, J. S., Hood, H. V., Brasfield, T., and Kelly, J. A. (1989). Differences in gay men’s aids risk knowledge and behavior patterns in high and low aids prevalence cities. *Public Health Reports*, 104(4):391.
- Stanger-Hall, K. F. and Hall, D. W. (2011). Abstinence-only education and teen pregnancy rates: Why we need comprehensive sex education in the us. *PloS one*, 6(10):e24658.
- Story, G. B. (2019). Travel through the indian ocean with the world’s last female ship captain. YouTube. Accessed: 2024-09-02.

- Streitmatter, R. (1995a). *Unspeakable: The Rise of the Gay and Lesbian Press in America*, chapter 9. Faber and Faber, Boston, MA. Chapter 9.
- Streitmatter, R. (1995b). Unspeakable: The rise of the gay and lesbian press in america. In *Unspeakable: The Rise of the Gay and Lesbian Press in America*, chapter 5, pages 116–153. Faber & Faber, Boston.
- Terra, P. (2023). The new science of sex and gender. YouTube video. Accessed: 2024-08-30.
- Tewksbury, R. (2003). Bareback sex and the quest for hiv: Assessing the relationship in internet personal advertisements of men who have sex with men. *Deviant Behavior*, 24(5):467–482.
- The Daily Times (1983). November 21, 1983 (page 1 of 22). *The Daily Times (1964-)*. Accessed via Vanderbilt University Library.
- The Post-Crescent (1983). August 20, 1983 (page 3 of 28). *The Post-Crescent (1965-)*. Accessed via Vanderbilt University Library.
- Thorne, A. and Coupland, J. (1998). Articulations of same-sex desire: Lesbian and gay male dating advertisements. *Journal of Sociolinguistics*, 2(2):233–257.
- US Department of Health and Human Services (US DHHS) (2005). Aids public information data set (apids) us surveillance data for 1981-2002. CDC WONDER On-line Database. Available from: <https://wonder.cdc.gov/>.

Appendices

A Additional Figures and Tables

Table A.1: Other Papers using Personal Ads Data

Author(s)	Newspaper/Website	Years Covered	Obs	Country
Lee, 1976	The Advocate	1970s	876	U.S.
Lumby, 1978	The Advocate	1976	1,111	U.S.
Laner and Kamel, 1978	The Advocate	1977	359	U.S.
Davidson, 1991	The Village Voice	1978, 1982, 1985, 1988	844	U.S.
Gonzales and Meyers, 1993	Several U.S. Newspapers	1988-1989	2008	U.S.
Hatala and Prehodka, 1996	Several U.S. Newspapers	1993-1994	396	U.S.
Hatala et al., 1998	San Francisco Bay Guardian, SF Weekly, Bay Times	1995-1996	100	U.S.
Thorne and Coupland, 1998	Several U.K. Newspapers	1995-1996	200	U.K.
Bartholome et al., 2000	Canadian Telephone Ad System	1997	167	Canada
Smith, 2000	Outrage Magazine	1985-1996	591	Australia
Baker, 2003	Gay Times (formerly Gay News)	1973, 1982, 1991, 2000	1,350	U.K.
Tewksbury, 2003	Unnamed website	2003	880	U.S.
Grov, 2010	Craigslist	2009	1,438	U.S.

Table A.1: Source:

Notes: This table presents a list of papers using data from personal ads from LGBT newspapers, magazine, or other sources.

Figure A.1 : Age Distribution

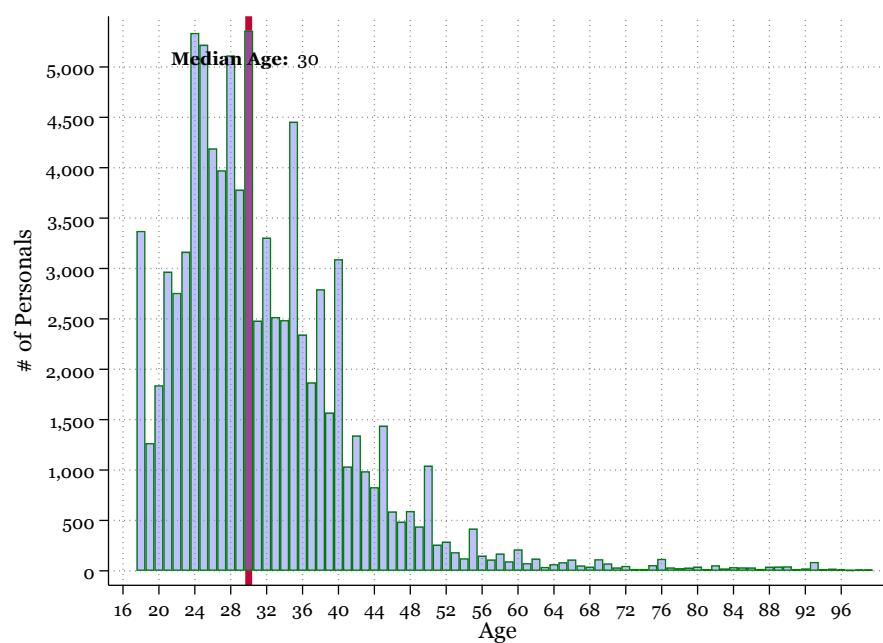


Figure A.1: Source: Advocate Personal Ads

Notes: This table presents the age distribution of individuals who post personal ads in *the Advocate*.

Figure A.2 : Copies Distributed Per Issue

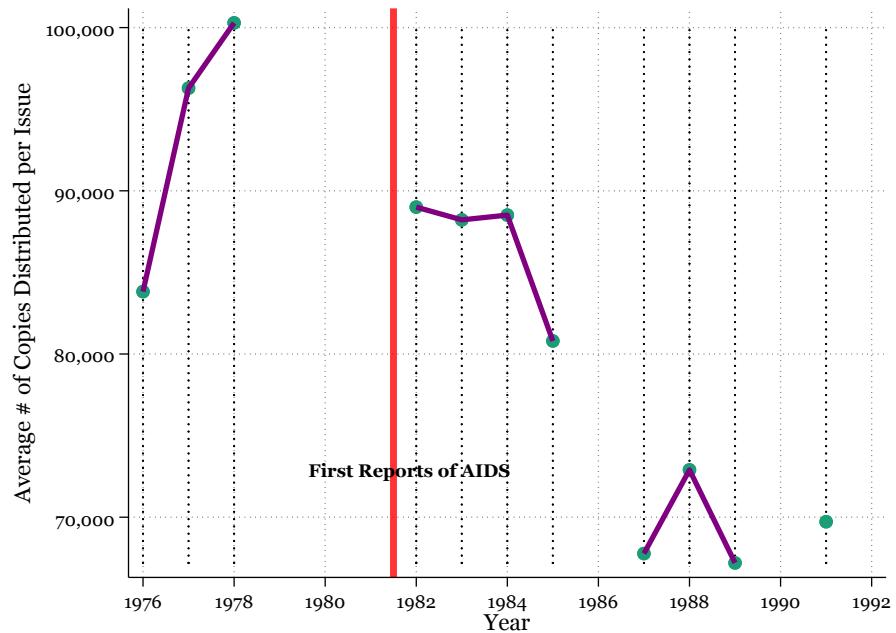


Figure A.2: Source: Advocate Personal Ads

Notes: This figure presents distribution statistics from located “Statement of Ownership, Management & Circulation” from issues of *the Advocate*. The vertical axis represents the average number of copies distributed per issue over the past 12 months.

Figure A.3 : Total Ads with Relationship Terms

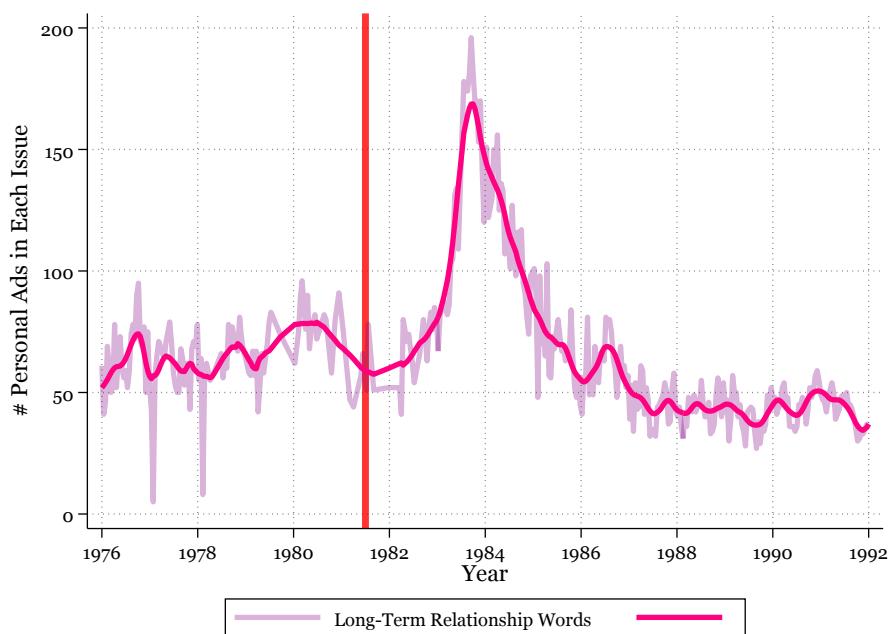


Figure A.3: Source: Advocate Personal Ads

Notes: This figure represents the total number of ads which include a relationship term posted in each issue of *the Advocate* personals section over time. The light line represents the exact number of ads with relationship terms in an issue while the darker line represents a smoothed version obtained from a locally weighted regression. These terms include 'partner', 'boyfriend', 'longterm', 'ltr', and 'rel'.

Table A.2: Year of First Reports AIDS Case

	MSA Name	First AIDS Case
1	Houston-Pasadena-The Woodlands, TX	1981.5
2	Los Angeles-Long Beach-Anaheim, CA	1981.5
3	Cleveland, OH	1981.5
4	New Haven, CT	1981.5
5	Portland-Vancouver-Hillsboro, OR-WA	1981.5
6	Pittsburgh, PA	1981.5
7	Miami-Fort Lauderdale-West Palm Beach, FL	1981.5
8	Boston-Cambridge-Newton, MA-NH	1981.5
9	Atlanta-Marietta-Sandy Springs-Roswell, GA	1981.5
10	Chicago-Naperville-Elgin, IL-IN	1981.5
11	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	1981.5
12	Detroit-Wayne-Dearborn, MI	1981.5
13	Syracuse, NY	1981.5
14	Tampa-St. Petersburg-Clearwater, FL	1981.5
15	Hartford-West Hartford-East Hartford, CT	1981.5
16	San Francisco-Oakland-Everett, CA	1981.5
17	New York-Newark-Jersey City, NY-NJ	1981.5
18	Baltimore-Columbia-Towson, MD	1981.5
19	Sacramento-Roseville-Folsom, CA	1982
20	Denver-Aurora-Centennial, CO	1982
21	Dallas-Fort Worth-Arlington, TX	1982
22	San Diego-Chula Vista-Carlsbad, CA	1982
23	Minneapolis-St. Paul-Bloomington, MN-WI	1982
24	Virginia Beach-Chesapeake-Norfolk, VA-NC	1982
25	Rochester, NY	1982
26	St. Louis, MO-IIL	1982
27	Phoenix-Mesa-Chandler, AZ	1982
28	North Port-Bradenton-Sarasota, FL	1982
29	Youngstown-Warren, OH	1982.5
30	Orlando-Kissimmee-Sanford, FL	1982.5
31	Washington-Arlington-Alexandria, DC-VA-MD-WV	1982.5
32	Afflentown-Bethlehem-Easton, PA-NJ	1982.5
33	Akron, OH	1982.5
34	Harrisburg-Carlisle, PA	1982.5
35	Scranton-Wilkes-Barre, PA	1982.5
36	Columbus, OH	1982.5
37	Birmingham, AL	1982.5
38	Seattle-Tacoma-Bellevue, WA	1982.5
39	Colorado Springs, CO	1982.5
40	San Antonio-New Braunfels, TX	1983
41	Cincinnati, OH-KY-IN	1983
42	Buffalo-Cheektowaga, NY	1983
43	Wichita, KS	1983
44	Milwaukee-Waukesha, WI	1983
45	Providence-Warwick, RI-MA	1983
46	Austin-Round Rock-San Marcos, TX	1983
47	Riverside-San Bernardino-Ontario, CA	1983
48	Raleigh-Cary, NC	1983
49	Columbia, SC	1983
50	Indianapolis-Carmel-Greenwood, IN	1983
51	Bakersfield-Delano, CA	1983
52	Urban Honolulu, HI	1983
53	Oxnard-Thousand Oaks-Ventura, CA	1983
54	Charleston-North Charleston, SC	1983
55	Oklahoma City, OK	1983
56	Springfield, MA	1983
57	Kansas City, MO-KS	1983
58	Nashville-Davidson-Murfreesboro-Franklin, TN	1983
59	New Orleans-Metairie, LA	1983
60	Jamaica, PR	1983
61	San Juan-Caguas-Yayabo, PR	1983
62	Las Vegas-Henderson-North Las Vegas, NV	1983
63	Allentown-Schneckey-Troy, NY	1983
64	Tucson, AZ	1983
65	Louisville/Jefferson County, KY-IN	1983
66	Dayton-Kettering-Beavercreek, OH	1983
67	Charlotte-Concord-Gastonia, NC-SC	1983.5
68	Jacksonville, FL	1983.5
69	Salt Lake City-Murray, UT	1983.5
70	San Jose-Sunnyvale-Santa Clara, CA	1983.5
71	Richmond, VA	1983.5
72	Mobile, AL	1983.5
73	Baton Rouge, LA	1983.5
74	Greenville-Anderson-Greer, SC	1983.5
75	McAllen-Edinburg-Mission, TX	1984
76	Vallejo, CA	1984
77	Omaha, NE-IA	1984
78	Fresno, CA	1984
79	Greensboro-High Point, NC	1984
80	El Paso, TX	1984
81	Memphis, TN-MS-AR	1984
82	Deltona-Daytona Beach-Ormond Beach, FL	1984.5
83	Albuquerque, NM	1984.5
84	Little Rock-North Little Rock-Conway, AR	1985
85	Ann Arbor, MI	1985
86	Stockton-Lodi, CA	1985
87	Toledo, OH	1985
88	Knoxville, TN	1985
89	Grand Rapids-Wyoming-Kentwood, MI	1985
90	Fort Wayne, IN	1985.5

Table A.2: Source: AIDS Public Information Dataset

Notes: This table presents the first half-year period when an MSA reports their first AIDS case.

Table A.3: Newspaper Reports

Newspaper Name	Circulation City	State	rank
Birmingham Post - Herald	Birmingham	AL	2
Los Angeles Times	Los Angeles	CA	1
Record Searchlight	Redding	CA	1
The San Francisco Examiner	San Francisco	CA	2
The Hartford Courant	Hartford	CN	1
The Washington Post	Washington	DC	1
Sun-Sentinel	Fort Lauderdale	FL	1
Fort Myers News-Press	Fort Myers	FL	1
Palm Beach Daily News	Palm Beach	FL	1
The Tampa Tribune	Tampa	FL	1
The Atlanta Journal	Atlanta	GA	1
Des Moines Register	Des Moines	IA	1
Iowa City Press-Citizen	Iowa City	IA	1
Evansville Press	Evansville	IN	1
The Indianapolis news	Indianapolis	IN	1
The Lafayette Journal and Courier	Lafayette	IN	1
Muncie Evening Press	Muncie	IN	1
Palladium - Item	Richmond	IN	1
The Daily Advertiser	Lafayette	LA	1
Daily World	Opelousas	LA	1
The Times	Shreveport	LA	1
Boston Globe	Boston	MA	1
The Daily Times	Salisbury	MD	1
The Times Herald	Port Huron	MI	1
Detroit Free Press	Detroit	MI	1
Lansing State Journal	Lansing	MI	1
Hattiesburg American	Hattiesburg	MS	1
Clarion-Ledger	Jackson	MS	1
Great Falls Tribune	Great Fall	MT	1
The Record	Hackensack	NJ	1
The Courier-News	Bridgewater-Plaifield	NJ	1
Courier - Post	Camden	NJ	1
Daily Record	Morristown	NJ	1
Albuquerque Journal	Albuquerque	NM	1
Star-Gazette	Elmira	NY	1
Ithaca Journal	Ithaca	NY	1
Daily News	New York	NY	2
Poughkeepsie Journal	Poughkeepsie	NY	1
Democrat and Chronicle	Rochester	NY	1
Chillicothe Gazette	Chillicothe	OH	1
The Cincinnati Post	Cincinnati	OH	2
The Coshocton Tribune	Coshocton	OH	1
Lancaster Eagle-Gazette	Lancaster	OH	1
News Journal	Mansfield	OH	1
The Marion Star	Marion	OH	1
The Advocate	Newark	OH	1
News Herald	Port Clinton	OH	1
The Times Recorder	Zanesville	OH	1
Statesman Journal	Salem	OR	1
The Morning Call	Allentown	PA	1
Philadelphia Daily News	Philadelphia	PA	2
Pittsburgh Press	Pittsburgh	PA	1
The Greenville News	Greenville	SC	1
Argus Leader	Sioux Falls	SD	1
The Leaf-Chronicle	Clarksville	TN	1
Jackson Sun	Jackson	TN	1
The Knoxville News-Sentinel	Knoxville	TN	1
The Commercial Appeal	Memphis	TN	1
The Daily News-Journal	Murfreesboro	TN	1
The Tennessean	Nashville	TN	1
Abilene Reporter - News	Abilene	TX	1
El Paso Herald - Post	El Paso	TX	1
Daily Press	Newport News	VA	1
Burlington Free Press	Burlington	VT	1
The Spokesman-Review	Spokane	WA	1
The Post-Crescent	Appleton	WI	1
Green Bay Press Gazette	Green Bay	WI	1
Herald Times Reporter	Manitowoc	WI	1
Marshfield News-Herald	Marshfield	WI	1
The Oshkosh Northwestern	Oshkosh	WI	1
The Sheboygan Press	Sheboygan	WI	1
The Daily Tribune	Wisconsin Rapids	WI	1

Table A.3: Source: ProQuest Historical Newspapers

Notes: This table lists all the local newspapers obtained from ProQuest Historical Newspapers database used for analysis. I only include the most popular newspaper in each MSA that is available in this dataset. This is determined from the 1981 Editor & Publisher International year book ([edi, 1981](#)).

Figure A.5 : Per 1 Million Population



Figure A.4: Source: Advocate Personal Ads and AIDS Public Information Dataset

Notes: These figures depict event study estimates of [Equation 2](#) while changing the outcome variable to represent the log of the number of ads per 1 million population plus 1. It measures the effect of the first reported AIDS case on personal ads using the Imputation Estimator developed by [Borusyak et al. \(2024\)](#). The first figure depicts effects on the total number of personal ads. The second figure depicts effects on the number of ads which include safety terms. All estimates control for period fixed effects, and season-MSA fixed effects. Standard errors are clustered at the MSA-level and estimates are weighted by population.

Figure A.5 : Poisson

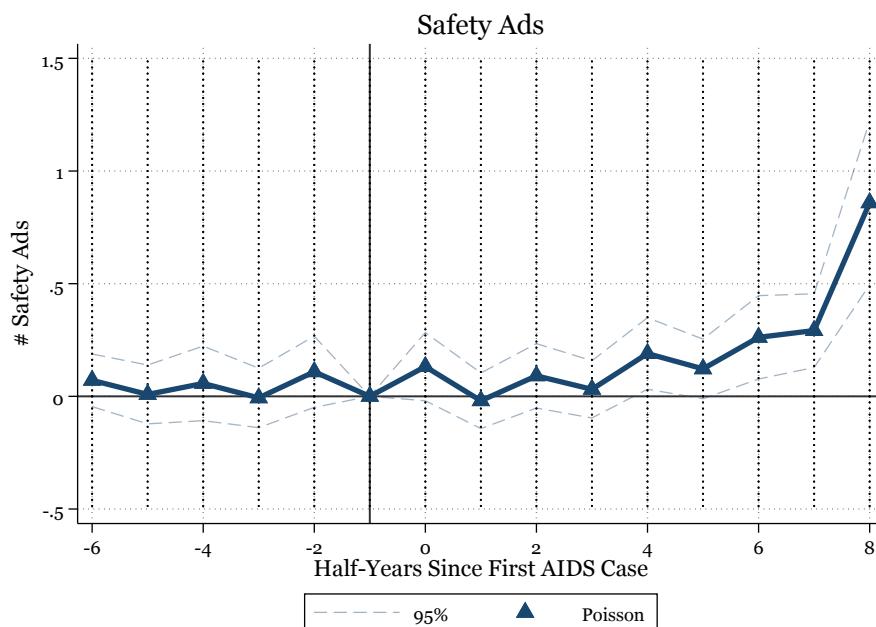


Figure A.5: Source: Advocate Personal Ads and AIDS Public Information Dataset

Notes: These figure depict event study estimates of [Equation 2](#) measuring the effect of the first reported AIDS case on personal ads using the poisson pseudo-likelihood regression. The figure depicts effects on the number of ads which include safety terms. All estimates control for period fixed effects, and season-MSA fixed effects. Standard errors are clustered at the MSA-level.

Figure A.6 : Proportion of Ads by Treatment Time

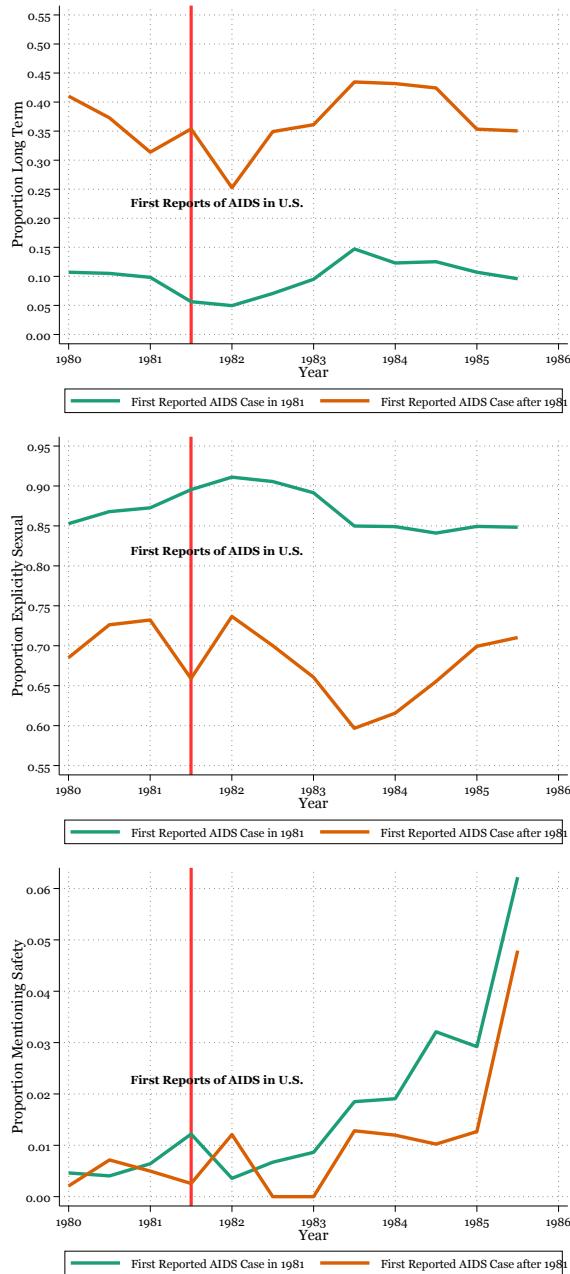


Figure A.6: Source: Advocate Personal Ads

Notes: This figure presents trends in the proportion of certain types of ads in MSA's which reported their first AIDS case 1981 and MSA's which reported their first AIDS case later on. The first figure shows these trends for the proportion of long term-relationship oriented ads. The second term shows these trends for the proportion of explicitly sexual ads. The third term shows these trends for ads which mention a safety term.

Figure A.7 : Constructing a Panel

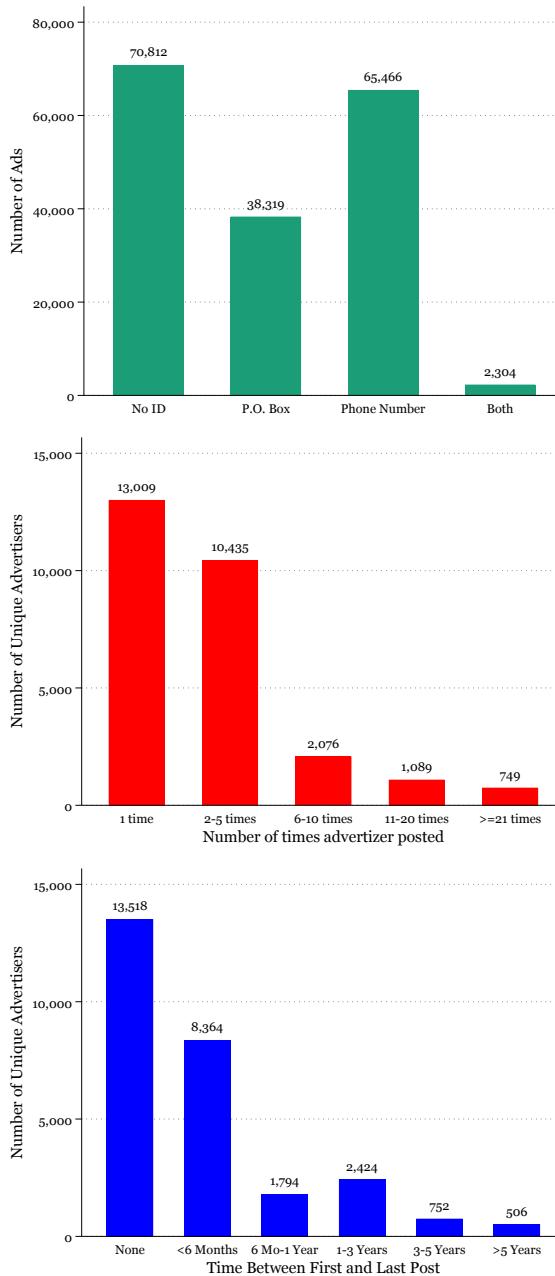


Figure A.7: Source: Advocate Personal Ads

Notes: These figure depict sample sized when using the personal ads data for panel analysis. The first figure depicts the number of ads which provide no form of identification, a P.O.Box number, a phone number or both a P.O.Box and phone number. The second figures depicts the number of unique advertisers and the number of times they post ads. The this panel shows the length of time they can be followed in the dataset.

Figure A.8 : Proportion of Safety Ads by Term

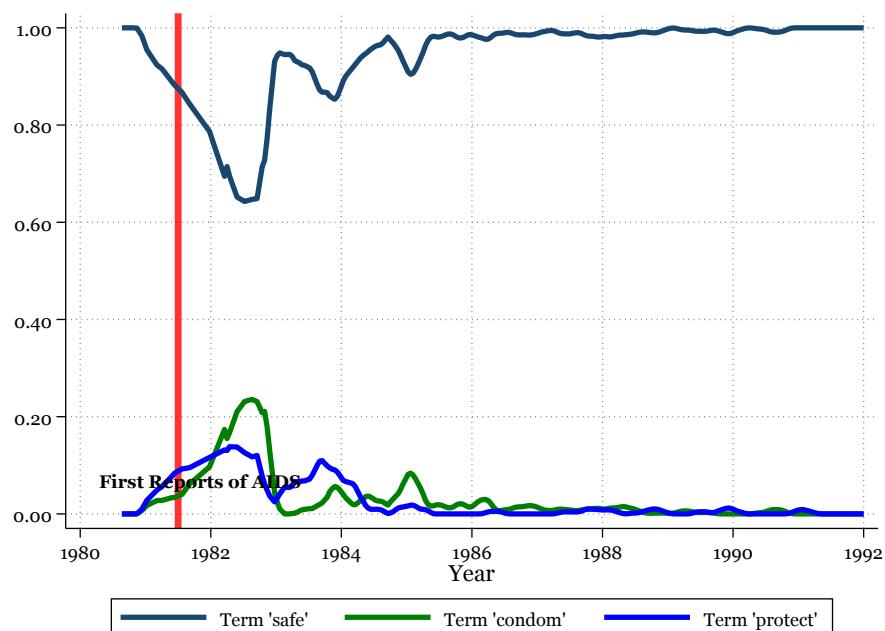


Figure A.8: Source: Advocate Personal Ads.

Notes: This paper searches for the terms ‘safe’, ‘condom’, and ‘protect’ to identify ads which mention safety-terms. In this figure I show the proportion of safety-ads that are identified using each term.

B Cleaning Personals Data

In order to construct the main dataset of personal ads from the advocate, I exploit the contours in the classified section of the newspapers. The ads are presented in vertical columns one after the other as depicted in the first panel of [Figure B](#). I go across each page and search for the proportion of dark pixels across the page. The second panel of [Figure B](#) shows the proportion of dark pixels across the X-coordinates of the page. I am able to identify the coordinates which divide the personal ads. Thereafter, I break down the ads into singular vertical strips of personal ads as depicted in the first panel of [Figure B](#). Then, I use Optical Character Recognition (OCR) to digitize the the text in each individual column as depicted in the second panel of [Figure B](#). Now the text in each ad is presented one after the other and separated by a space. I use these spaces to create a dataset where each observation represents the text in each personal ad. In this dataset, most observations represent personal ads but there is a significant number of observations which include other, non-personal ad, information. This includes advertisements for other business or the titles of columns such as the “personals” highlighted in the second panel of [Figure B](#). Therefore, I use machine learning to identify and drop all observations which do not represent personal ads. From the dataset, I randomly select 500 and manually identify whether the ads represent a personal ads. I use this as training data. I use a supervised machine learning model to classify the ads. The model is trained on the manually labeled subset of 500 ads, which have been identified as either personal ads or not.⁶⁸ The trained model is then applied to the entire dataset to automatically categorize the remaining ads.

⁶⁸I use Term Frequency-Inverse Document Frequency (TF-IDF) to vectorize the ads. This involves measuring the importance of words in the ads of a specific category. It discounts words that are common throughout the document.

Figure B : Digitizing Personal Ads 1

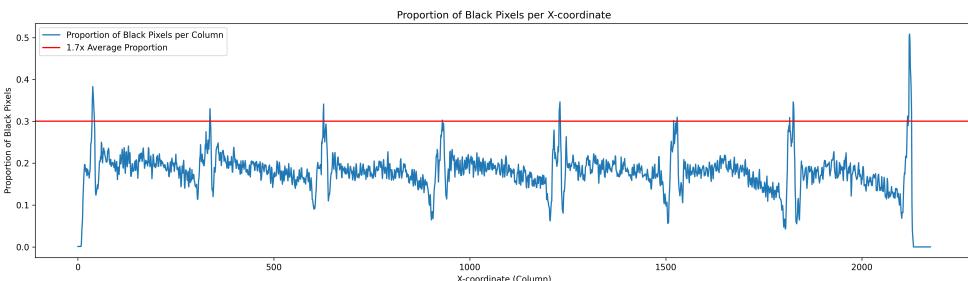
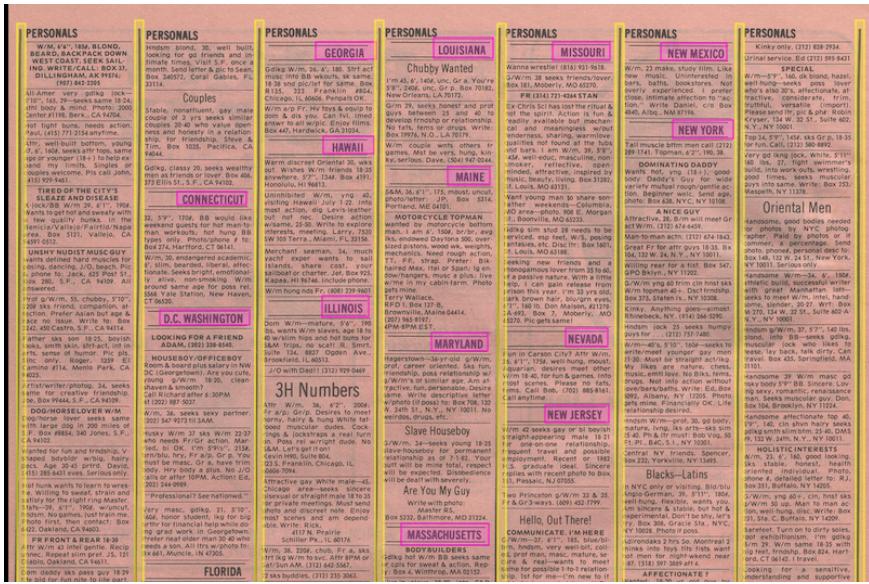


Figure B.1: Source: Advocate Personal Ads

Notes: This figure shows the methodology used to construct the personal ads dataset.

Figure B : Digitizing Personal Ads 2

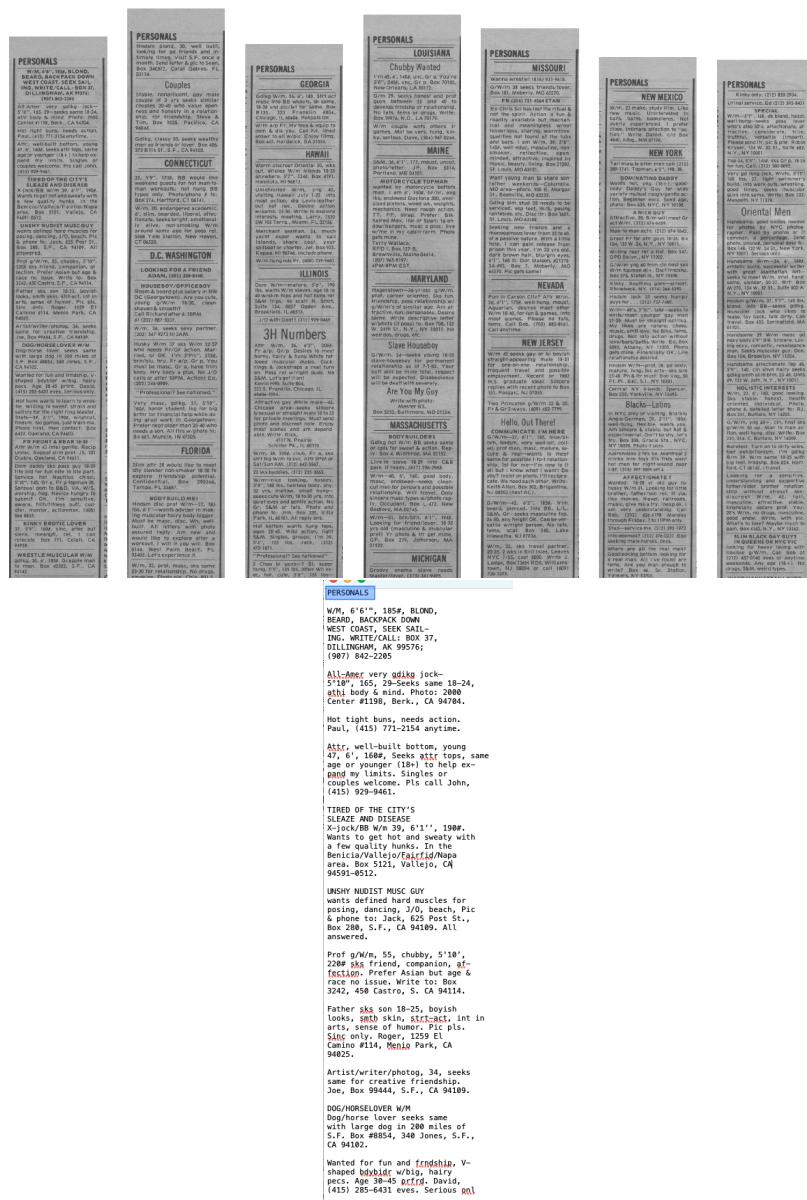


Figure B.2: Source: Advocate Personal Ads

Notes: This figure shows the methodology used to construct the personal ads dataset.

C Identifying HIV/AIDS-Related Articles

In order to identify HIV/AIDS related articles in local newspapers, I employ TDM Studio. TDM studio is a text and data mining platform developed by proquest which allows me to search through the text of each of these newspapers. To identify AIDS-related articles, I use the following search prompt:

("AIDS" AND ("disease" OR "pneumonia")) OR ("HIV" AND ("homosexual" OR "gay" OR "disease" OR "virus" OR "cancer")) OR ((("gay" OR "homosexual") AND ("disease" OR "pneumonia" OR "cancer" OR "virus")))

The prompt produces a list of articles but many articles match the search criteria but are not related to the HIV/AIDS virus.⁶⁹ To address this concern, I randomly choose 500 search results, I manually identify whether these results represent HIV/AIDS related articles or something else. I use machine learning to predict whether the remaining articles are HIV/AIDS related.

⁶⁹For example, the search prompt would pick up an article titled “U.S. aids China in combating bird flu disease.”

