Runjing Lu

Lu Zhang

Emory University

2013 Undergraduate Data Palooza Research Competition

26 July 2013

The Study of ADIS Trends in the United States during 1985 and 2005

1. Introduction

Since 1981, AIDS has always been a threat to people’s health. As a result, thousands of studies have been conducted on AIDS each year on an international base and much achievement has been attained up to now. In this paper, we focus on the examination of the US AIDS incidence, use the data from AIDS database provided by the competition from 1985 to 2005, and analyze these data through panel regression model. By analyzing the regression results, we arrive at the following conclusions about AIDS pattern during 1985 and 2005. (1) The numbers of aids cases are significantly greater in the years of 92-94 and 02-04 than those in other years. (2) The state aids incidence and state average aids incidence drop sharply during mid-to-late 90s and during 2004 and 2005. (3) The southern states’ aids incidence proportion is significantly higher than that of non-southern state during the years of 1992-94 and 2002-04.

This paper is organized as follows. Section 2 describes the data set. Section 3 introduces the methodology and discusses our results. Section 4 explains the regression results by relating them to the social, economic, and medical contextual
2. Data

We get the data from the AIDS database provided by the competition. The dataset contains weekly AIDS incidence for each state in the U.S. from 1984 to 2005. Due to incomplete data collection and other unknown factors, the original data fluctuates a lot with weekly difference ranging from 0 to over 100. To lower the data’s fluctuation and better estimate the AIDS incidence pattern, we sum the AIDS incidence in each year for each state together to get the annual AIDS incidence for each state. During the summation process, we discover that several states, including Wyoming, Virgin Islands, American Samoa, Guam, and Northern Mariana Islands, suffer from incomplete data, so we delete these states from our analysis. In addition, disease pattern is highly correlated to geographical factors, such as location. To lower the impact of location on the pattern, we focus our analysis on U.S. mainland and omit the states outside of U.S. mainland, including Puerto Rico and Hawaii (we do not omit Alaska because it is geographically close to the U.S. mainland than other non-mainland states). After the omission, there leave 49 states and we sum up their annual AIDS incidence to get the annual AIDS incidence for the whole nation. Since many states lack the data in 1984, we set 1985 as the beginning year and analyze the AIDS trend from 1985 to 2005.

Figure 1 shows the evolution of annual U.S. AIDS incidence. In the figure, there are two peaks of AIDS incidence from 1985 to 2005. The first peak is during 1992-1994 and the second peak is during 2002-2004. To better analyze this
chronological disease pattern and to examine whether these two peaks are statistically significant, we carry out a panel data analysis in the next section. There may be social, economic, and other factors correlated with or contributing to the chronological pattern and these factors will be analyzed in the related contextual information section.

Figure 1: The evolution of annual U.S. AIDS incidence (1985-2005)

In addition, according to previous studies, Southern states experience higher AIDS incidence than other non-Southern states. To better analyze this geographical AIDS pattern in U.S., we categorize states by whether they locate in the South or not. According to the definition of Southern states by the US Census Bureau, we define Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia,
Delaware, Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma, Texas and District of Columbia as Southern states. Figure 2 represents the evolution of AIDS incidence in different regions and figure 3 shows each region’s contribution to the national total AIDS incidence. In figure 3, the contribution of Southern states has been increasing since 1985, while other regions’ contribution has been dropping steadily. Therefore, we hypothesize that the AIDS incidence is disproportionately centered in the South, and we will carry out panel data regression to test whether this geographical AIDS pattern is statistically significant in the next section. There may be social, economic, and other factors correlated with or contributing to this pattern and the related contextual information section will analyze these factors.
Figure 2: The evolution of annual U.S. AIDS incidence for different regions (1985-2005)

Here, The West includes Washington, Montana, Oregon, Idaho, California, Nevada, Utah, Colorado, Arizona, New Mexico, and Alaska. The Midwest includes North Dakota, South Dakota, Nebraska, Kansas, Minnesota, Iowa, Missouri, Wisconsin, Illinois, Indiana, Michigan, and Ohio. The Northeast consists of Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New Jersey, New York, and Pennsylvania. The South consists of Texas, Oklahoma, Arkansas, Louisiana, Mississippi, Tennessee, Alabama, Georgia, Florida, South Carolina, North Carolina, Virginia, West Virginia, Kentucky, Delaware and District of Columbia. The areas of different colors represent the number of AIDS incidence for different regions. The outline of the highest area is the annual AIDS incidence for the whole country.

Sources: AIDS database of the competition.
Graphing Method: Use the staked area graph in the Excel; y axis is the annual AIDS incidence for four regions and x axis is the year.
Figure 3: Different regions’ contribution to total AIDS incidence (1985-2005)


Sources: AIDS database of the competition.
Graphing Method: Use the 100% stacked area graph in the Excel; y axis is the percentage each region’s AIDS incidence contribute and x axis is the year.

3. Methodologies and Regression Results

To determine whether AIDS incidence experience peak during 1992-1994 and 2002-2004 statistically and whether the AIDS incidence is disproportionally centered in the South, We use panel regression model to analyze the data from the AIDS database. The regression model is

\[ Y_{it} = a + b1 \cdot \text{Stated} + b2 \cdot \text{Year} + b3 \cdot \text{Stated} \cdot \text{Year} + e, \]
where i and t represent i state and t year, respectively. Yit is our interested dependent variables: percentage, mean, statetotal, and their corresponding logarithm. Percentage is the proportion a particular state’s annual AIDS incidence contributes to the national total. The percentage equals state AIDS incidence divided by country AIDS incidence. Lnpercentage is the log of actual percentage. Mean is each state’s annual arithmetic average of AIDS incidence. Lnmean is the log of mean. Statetotal is each state’s annual total cases of AIDS. It equals the sum of each state’s weekly data in a particular year. Lnstatetotal is the log of statetotal. Yeard is the dummy variable representing the data’s time period. Yeard is 1 if the data is during 1992-1994 and 2002-2004 and 0 otherwise. Its coefficient can tell us by what extent our interested variables during 1992-1994 and 2002-2004 is higher than those during other time periods. Stated is the dummy variable representing where the state locates. Stated is 1 if the state locates in the South and 0 otherwise. Stated’s coefficient represents how much Southern states’ AIDS incidence is greater than its non-Southern counterparts. Stated*Yeard is the interaction term of the previous two dummy variables. Stated*Yeard’s coefficient explains how much the Southern states’ AIDS incidence is greater than its non-Southern counterparts’ during 1992-1994 and 2002-2004. To account for the heteroskedasticity, we use “re” option when regressing the panel data in STATA. The regression results are shown in table 1.
Table 1: Regression results

The asterisk next to the coefficient represents at which level the coefficient is significant:
*** = 1% level, ** = 5% level, and * = 10% level. The number under the coefficient is its standard deviation. Lnpercentage, lnmean, and lnstatetotal are the logarithms of percentage, mean, and statetotal.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Explanatory Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>lnpercentage</td>
<td>percentage</td>
<td>lnmean</td>
<td>mean</td>
<td>lnstatetotal</td>
<td>statetotal</td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td>-5.202269***</td>
<td>-.0011048***</td>
<td>1.385735***</td>
<td>14.34184***</td>
<td>5.145905</td>
<td>704.1822***</td>
</tr>
<tr>
<td>Stated</td>
<td></td>
<td>.7262987</td>
<td>.0053393</td>
<td>.2395941</td>
<td>4.030327</td>
<td>.2466723</td>
<td>214.2571</td>
</tr>
<tr>
<td>Yeard</td>
<td></td>
<td>.550744</td>
<td>-.0001732</td>
<td>.1562771</td>
<td>1.161173</td>
<td>.5361789</td>
<td>-12.28222</td>
</tr>
<tr>
<td>Stated*Yeard</td>
<td></td>
<td>.0307003</td>
<td>.0155666</td>
<td>.6984807</td>
<td>11.75032</td>
<td>.7191104</td>
<td>624.6615</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.0845376***</td>
<td>.0008329</td>
<td>.6206213***</td>
<td>9.380709***</td>
<td>.5873821</td>
<td>383.3252***</td>
</tr>
<tr>
<td>R2</td>
<td></td>
<td>.0891646</td>
<td>.0008329</td>
<td>.0540553</td>
<td>1.242088</td>
<td>.0583927</td>
<td>62.93297</td>
</tr>
<tr>
<td>Total Panel Observation</td>
<td></td>
<td>1062</td>
<td>1071</td>
<td>1062</td>
<td>1071</td>
<td>1062</td>
<td>1071</td>
</tr>
</tbody>
</table>

Table 1 shows that the coefficient of Yeard is always significant at 1% level in six models, implying that the percentage of state AIDS incidence, the average of state AIDS incidence, and the annual state AIDS incidence during 1992-1994 and 2002-2004 are higher than those during other periods. On average, the state’s AIDS incidence proportion, the average of state annual AIDS incidence, and the total state annual AIDS incidence during 1992-1994 and 2002-2004 are 0.83%, 9.38, and 383.3252 higher than their counterparts during other periods. Surprisingly, the coefficients of Stated and Stated*Yeard are mostly insignificant, which indicates that the geographical difference in AIDS incidence between Southern states and
non-Southern states is not statistically significant. However, in model (1), the coefficient of Stated*Yeard is significant at 10% level. This coefficient suggests that during 1992-1994 and during 2002-2004, the AIDS incidence percentage of Southern states over national total AIDS incidence is 0.15% higher than that of non-Southern states. Thus, when it comes to the percentage a particular state’s AIDS incidence contributes to the country total, AIDS shows geographical, disproportionate distribution.

From the regression results above, we conclude the following three phenomena about AIDS. (1) The state AIDS incidence during 1992-1994 and 2002-2004 is significantly greater than that in other years. (2) The state AIDS incidence drops significantly during mid to late 90s and 2004-2005. (3) Southern state AIDS incidence proportion is significantly higher than that of non-Southern states during 1992-1994 and 2002-2004. The next session explains these AIDS phenomena by relating social and economic contextual information during that time to them.

4. Related Contextual Information

4.1 The state total AIDS incidence during 1992-1994 and 2002-2004 is significantly greater than that in other years.

4.1.1 Economic situation and Government funding for AIDS programs

Funding levels are important to many AIDS programs because the levels determine the number of patients those programs can serve, the amount of medication (including preventative vaccine) they can provide, and the anti-AIDS activities they
can carry out (Kaplan 155). Of all the funding resources, government funding is the most important because government is richer and more powerful than other private sponsors. Meanwhile, government funding is highly correlated with a country’s economic situation. Thus, there exists some correlation between economic condition, government funding and the AIDS chronological pattern. Generally speaking, more funding results in less AIDS incidence, because AIDS programs can spend more on AIDS education and AIDS preventative activities, which reduce AIDS incidence. Therefore, we hypothesize that the poor economic condition and the decreased government funding in the early 1990s and early 2000s contribute to the peak of AIDS incidence during 1992-1994 and 2002-2004.

Figure 4 depicts the federal funding growth rate and the AIDS incidence growth rate. They have the same chronological trend and both experience peaks during 1992-1994 and 2002-2004, except that the growth rate’s fluctuation is much smaller than the AIDS incidence’s. At first sight, this positive correlation between funding growth rate and AIDS incidence is surprising because common sense tells us that when funding is higher, the AIDS incidence should be lower. However, examining the trends closely, we find that government funding growth rate decreases right before 1993 and 2003 and the AIDS incidence increases right after 1993 and 2003, respectively. Therefore, there exists a lag between the decrease of government funding and the increase of AIDS incidence. The funding for AIDS programs needs time to reach the programs and the programs need time to implement and have real impacts on the local AIDS incidence. If we shift the growth rate of government
funding to the left by one year, as shown in Figure 5, it is clear that the two nadirs of funding growth rate corresponds well with the two peaks of AIDS incidence. Thus, the decrease in federal funding for AIDS programs may attribute to the increase in AIDS incidence during 1992-1994 and 2002-2004.

Figure 4: The evolution of the federal funding growth rate and the AIDS incidence growth rate (1985-2005)

The left y axis is the aids incidence (in actual number); the right y axis is the funding growth rate (in percentage point); the x axis is the year.

Graphing Method: Use the line graph in the Excel; set secondary y axis to show two y axes
Figure 5: The shifted evolution of the federal funding growth rate and the AIDS incidence growth rate (1985-2005)

Data sources: Johnston, Louis and Williamson, Samuel H. “What Was the U.S. GDP Then?” Measuring Worth (2012); AIDS database

Graphing Method: we shift the government funding growth line in figure 5 to the left by one year and get this figure.

Furthermore, the decrease in the government funding was due largely to the poor economic situation then. Figure 6 shows that the real GDP growth rate and the government funding growth rate show strong correlation with each other and the former experiences peaks and nadirs about 1 year before the latter. There exists one-year lag between the two rates because government funding plan is based on government spending in the last year, which is a main indicator of real GDP. In the early 1990s, the Federal Reserve's decision to raise interest rates, the beginning of the Gulf War, and the resulting 1990 spike of oil price slowed GDP development, increased inflation, worsened unemployment, and massived government budgetary deficits (http://en.wikipedia.org). In the beginning of the 2000s, U.S. also experienced
economic recession, which was due to a series of uncertain incidents. The over-investment in high technology in the late 1990s and the accounting scandal in 2002 served to lower the GDP development. Meanwhile, the “9’11” terrorist attack in 2001 and the beginning of the Iraq War in 1993 increased US’s military and security spending greatly. The economic downturn and the great increase in non-medical spending together impaired government’s financial support for ADIS-related programs.

In conclusion, real GDP growth rate decreases in year T, government funding for AIDS programs decreases in year T+1, and AIDS incidence increases in year T+2. AIDS incidence during 1992-1994 and 2002-2004 was significantly greater than that in other year because the U.S. was experiencing economic downturn and the federal funding for AIDS programs was lower at that time.
Figure 6: Real GDP Growth Rate and Government Funding for AIDS Growth Rate (1985-2005)

The real GDP is based on chained 2005 U.S. dollar. The left y axis is the government funding rate; the right y axis is the real GDP growth rate; the x axis is the year.

Data sources: Johnston, Louis and Williamson, Samuel H. “What Was the U.S. GDP Then?” Measuring Worth (2012); AIDS database
Graphing Method: Use the line graph in the Excel; set secondary y axis to show two y axes.

4.1.2 Immigration from third world countries

That AIDS incidence during 1992-1994 and 2002-2004 is significantly greater than that in other year may be due to higher immigration from third world countries then. Every year, huge population shifts from third world countries, like Iraq, Afghanistan, and Pakistan, to the U.S. Even though U.S. had the immigration policy that banned people infected with HIV from entering before 2012 (the Obama government lifted this ban in 2012), people from poor countries still pose risk of higher AIDS incidence on the U.S.. The reason is that people from those countries
usually lack sufficient amount of health education and are more likely to get involved in unprotected sex, prostitution, and other highly risky AIDS/HIV-related activities (Altman 43). Therefore, we hypothesize that the exceptionally high AIDS incidence during 1992-1994 and 2002-2004 may be related to the increase of third world immigration then. Here, due to the limitation of data we can get, we use refugee arrivals as the proxy for the third world immigration.

Figure 7 shows that the evolution of annual number of refugees and that of annual AIDS incidence have high correlation with each other. Refugee arrivals reach local maximum during 89-91 and 98-00 and total AIDS incidence reaches local maximum during 1992-1994 and 2002-2004 correspondingly, which shows that the peaks of annual AIDS incidence have a three-year lag after the peaks of refugee arrivals. Immigration peak does not bring about AIDS incidence peak immediately because refugees will not bring HIV into U.S. directly due to US immigration policy stated earlier. However, due to their poor financial situation and lower level of health education, these refugees are more likely to get involved in highly risky AIDS/HIV-related activities and infected with HIV in later days. Considering the time refugees need to get accustomed to the environment before participating in risky activities related to AIDS/HIV and the time before they are diagnosed with HIV, a three-year lag is reasonable.
Figure 7: The evolution of refugee arrivals and aids incidence (1985-2005)

The definition of refugees is “persons who sought residence in the United States in order to avoid persecution in their country of origin”.

4.1.3 Medical Methods to Diagnose AIDS/HIV

In addition to social and economic factors mentioned above, medical factors can also lead to the high point of AIDS incidence during 2002 and 2004. According to most previous studies, the annual AIDS incidence is interpreted as the number of people being infected with AIDS in a particular year (Hollander 3). However, these papers ignore the fact that many people, especially those in the South, had been infected with HIV for a long time before they were diagnosed with the disease (Vermund S7). Thus, the increase of AIDS incidence may not due to the increase in people’s infection, but due to new, rigorous methods to detect whether a person is infected or not.
Under the assumption that ADIS-risky people’s possibility to be infected with AIDS remains the same, easier access to and lower cost of AIDS tests can raise their possibility to take the tests and increase the ADIS incidence. The actions taken by AIDS-related government agencies and NGOs can attest this argument. In April 2003 the CDC announced an ambitious HIV/AIDS initiative to prevent new HIV infections and to identify HIV-positive people who may not be aware of their infection. The strategies of the initiative includes (1) making HIV testing a routine part of medical care and (2) implementing new models for diagnosing HIV infections outside medical settings (including the use of rapid HIV tests) (Kaplan 157). In addition, in 1994 U.S. FDA approved an oral HIV test, the first non-blood based antibody test for HIV. These initiatives can be possible cause to the increase of AIDS diagnoses and AIDS incidence during 1992-1994 because they made it easier and cheaper for people to take AIDS test and get results (http://www.pbs.org).

4.1.4 Gay Right Movement

Gay right movement can also contribute to the higher AIDS incidence during 1992-1994 because homosexual activities took up large proportion in the causes to AIDS during then. Between January 1993, and October 1995, 45% of reported cases were among men who had sex with men, 27% were among users of injection drugs, and 10% were among persons who contracted HIV through heterosexual contact (Hollander 1). Thus, we hypothesize that the gay right movement during early 1990s contributed to the apex of AIDS incidence during 1992-1994.
The gay rights movement in the early 1990s normalized the homosexuality and MSM (men who have sex with men). Growing number of people were open to the gay society and having sex with someone of the same sex was no longer a taboo. In 1993, during the presidential election, Bill Clinton promised to reverse the executive order barring gays and lesbians from serving in the U.S. military. Later, this “Don’t ask, Don’t tell” was passed into law and meant the government would no longer "ask" recruits if they were gay; so long as military personnel didn't "tell" anyone about their sexual preference - nor engage in homosexual acts - they could serve (http://www.time.com). Throughout Bill Clinton’s serve, gay right movement won growing support. Thus, the gay right movement in the early 1990s might contribute to the AIDS incidence peak from 1992 to 1994.

4.2 The state AIDS incidence drops significantly during mid to late 90s and 2004-2005

As HIV becomes increasingly threatening, specialists has been trying to discover effective therapies to cure patients as well as to prevent the spread of AIDS. The appearance of antiretroviral therapy (ART) can largely explain why AIDS incidences dropped sharply in mid to late 1990s. In 1994, tests on HIV-infected pregnant women proved that antiretroviral drugs are especially useful in preventing the “vertical transmission” of AIDS (Connor et al, 1173–1180). Moreover, with constant progress in AIDS treatment and highly active antiretroviral therapy (HAART) in 1996, the growth rate of HIV-infected population has noticeably declined and the patients live
much longer lives than before (MMWR, 1149–1153).

In addition, the U.S. government also plays an important role in controlling AIDS. Effective policies to prevent AIDS were enacted during Bush administration (2001-2009). There was a huge drop of HIV incidence around the year 2004 and afterwards. This decline is closely associated with the introduction of the U.S President’s Emergency Plan for AIDS Relief (PEPFAR). On January 28th, 2003, president George W. Bush enacted PEPFAR in the US, which intended to “prevent 7 million new AIDS infections, treat at least 2 million people with life-extending drugs, and provide humane care for 10 million people suffering from or impacted by AIDS” (Dietrich, 277-292). PEPFAR has accomplished several achievements since it was announced.

4.3 Southern state AIDS incidence proportion is significantly higher than that of non-Southern states during 1992-1994 and 2002-2004

This geographical disproportion may be due to the poor treatment in the South, as Moon and her co-workers indicate in their work (Moon et al. 279-281). In addition, cross-region migration is another reason for the higher AIDS incidence growth rate in the South. An analysis of 760 HIV-infected people in rural areas of Alabama and Mississippi found that one third of the subjects moved to the South after being diagnosed with HIV (Agee et al., 51-58). Furthermore, higher portion of African American in Southern America also accounts for the higher growth rate of AIDS incidences in the South of the States. African-Americans comprise 13% of the total US
population but 43% of the estimated numbers of persons living with AIDS at the end of 2004; 60% female cases are African American women. Southern states (22.8%) have about three times the African American population than other regions (7.4%) (Hanzhu Qian et al, 6-9).

5. Conclusion

From our data analysis and panel regression, we arrive at the following conclusions. First, the state AIDS incidence during 1992-1994 and 2002-2004 is noticeably greater than that in other years. This is possibly related to the poor economic condition and decreased amount of government funding in the early 1900s and 2000s. The unusually high AIDS incidence in these periods may also due to the high immigration rate from the third world, where people are not able to receive good health education and advanced medical care. Additionally, more advanced AIDS diagnosing methods introduced in 1994 and 2003 and the Gay right movement during 1992-1994 further explain the increasing number of people infected with AIDS. Second, the state AIDS incidence drops sharply during mid to late 90s and 2004-2005. Reasons for the large decline in these periods include the introduction of new therapy methods of HIV, such as the antiretroviral therapy (ART), and increasingly effective government policies on AIDS prevention, such as the U.S President’s Emergency Plan for AIDS Relief (PEPFAR) enacted during Bush administration. Third, southern state AIDS incidence proportion is significantly higher than that of non-Southern states during 1992-1994 and 2002-2004. This is associated with the poor treatment, the cross-region
migration, and the higher portion of African American in Southern America.

Work Cited


AIDS Database in Undergraduate Data Palooza Research Competition


Moon, T., Vermund, S., Tong, T., & Holmberg, S.. “Opportunities to Improve


