Bayesian Modeling of HIV Risk Factors:

Confronting HIV as a Stereotypical Infection of Sexual Minorities

Nikodem Lewandowski

Abstract: In this paper I conduct an analysis of PKD survey dataset, which is a survey that collects informations of people in HIV testing centers in Poland.

Introduction

The year 1980 is proclaimed as the start of the HIV epidemic. Since then, scientists around the world have been diligently working to find new ways to protect the global public from infection and treat those who have been infected (Sharp and Hahn 2011). Thanks to increased awareness and breakthroughs in medical achievements, it is now easier to detect the HIV virus in an individual's blood. We also know that using condoms and PrEP[1] is an effective means of preventing infection.

As a result, the global number of new infections has significantly decreased, thanks to advanced treatment methods, infected individuals can now enjoy a lifespan similar to uninfected people. The global infection rate continues to decline, as reported by "Global HIV & AIDS Statistics — Fact Sheet — Unaids.org" (2023). However, in some well-developed countries, there is a concerning rise in new HIV infections, particularly noticeable in the European Union (EU) where the number of new infections has been on the rise (European Centre for Disease Prevention and Control and World Health Organization 2022). One contributing factor is the global COVID-19 epidemic, which has discouraged people from getting tested for HIV, as evidenced by our dataset. Poland is one of the countries experiencing a worsening situation.

[1] PrEP, which stands for pre-exposure prophylaxis, is a medication that significantly reduces the risk of HIV infection.

From the beginning of the HIV epidemic, this virus has been associated with non-heteronormative individuals, predominantly males. During those years, gay men faced less acceptance compared to other sexual identities. Gay clubs provided a safe and accepting environment where anonymity was valued, and engaging in anonymous sexual encounters was not uncommon. Before the awareness of HIV, the use of condoms was less prevalent, making such sexual behavior highly risky.

The stereotype of HIV as "the disease of gays" still persists. In this study, I aim to challenge this stereotype by analyzing the causes of HIV infections, focusing on causal relationships. The main questions addressed in this work are:

- Do males in general have a higher probability of contracting HIV?
- Are homosexual individuals at a higher risk than heterosexuals?
- Is HIV primarily an infection among gay males, and does the stereotype hold true that there is something 'special' about this group explaining this phenomenon?

I will attempt to answer these questions by analyzing surveys conducted in Polish HIV testing centers known as PKDs (Pol. Punkt Konsultacyjno Diagnostyczny - Diagnostic Consultation Center). At least one such center exists in every major administrative region in Poland (Voivodeship). Testing at these centers is always anonymous, free of charge, and safe. Medical professionals at these centers interview clients and qualify them for testing, including a section of the survey focused on their sexual history.

The data were collected from 2015 to 2022[2], comprising responses from over 250,000 individuals. It is important to note that this dataset does not provide accurate information on the number of new infections, as there are various other places in Poland where HIV infection can be diagnosed. Therefore, our focus is on the survey questions. The visualizations below depict the demography of the survey (Figure 1). The left plot illustrates the gender distribution[3]. Notably, the majority of clients are males, and non-heteronormative males constitute half of all male clients, a higher proportion than in the general population. Females, on the other hand, are represented more proportionally.

The plot on the right illustrates the sexual identity and gender among the individual that received HIV-positive test result[4]. Approximately 70% of HIV-positive

[2]The dataset was obtained with the cooperation of PKD officials, the files are available on my GitHub project.

[3]22 individuals identified as 'Different' gender; however, their number is too small to be considered significant and is therefore excluded from the analysis

[4]Individuals considered HIV-positive in this analysis underwent two tests: the initial screening test, which yielded a positive result, and a second test that measured the amount of HIV virus in the blood, confirming the initial result.

individuals identify as non-heteronormative males. This result raises the need for further analysis, as it indicates that non-heteronormative males are more susceptible to infection. It is worth noting that this group seems to be aware of the risk, as they are overrepresented among those who regularly get tested.

One thing to note, due to the survey's anonymity, individuals who test themselves regularly will be counted as separate entries, potentially inflating the representation of this group. This might be one of the reasons for their overrepresentation.

The question arises: Is there indeed something distinctive about non-heteronormative males that makes them more susceptible to HIV infection? I will attempt to answer this question by examining causal relationships among various predictors, with a focus on identifying variables that best predict HIV positivity.

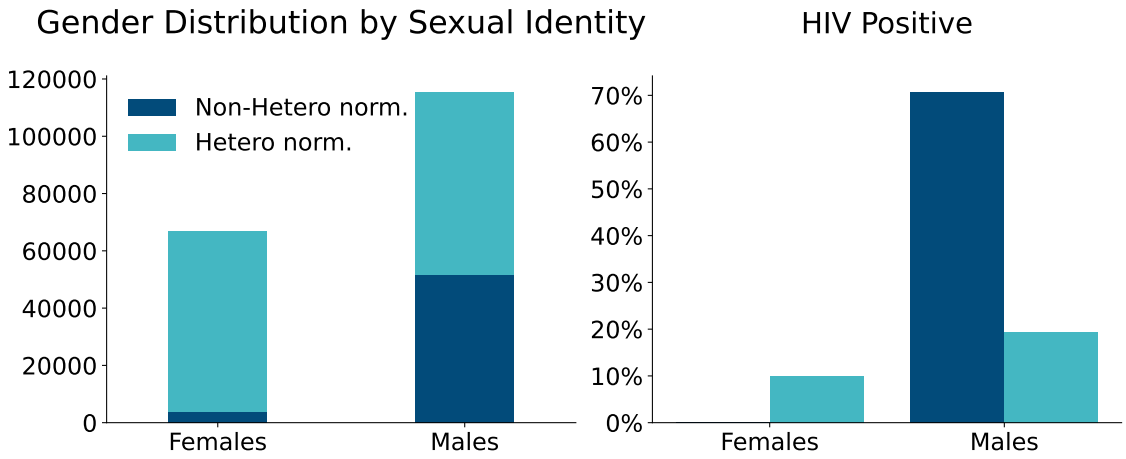

Figure 1: hello

First section

Let's first try to build a model that will trying to predict being HIV infected based on variables that we already know, so sex and sexual identity. All the statistical models in this project where build using 'numpyro' pacakage in python [link to package documentation]. It is a simplified version of 'pyro' package, both are focused on delivering tools for probabilistc data analyysis, including building advanced hiercharchical models. I used bayesian style logistic regression, the main difference in this kind of models is the possibility to choose how a prior distribution of the predictors ought to look like. The models use quadratic aproximation...

Let's see how the relation between our demographical predictors and being HIV positive look like. In the Figure Figure 2 you can see the prediction made by the model that as predictors took Gender and sexual identity of the clients. The violins

represent the distriubtion of predictions, they are not singel point estimates. It is characteristic for bayesian aproach to always interpreteae distributions as they convey more information than e.g. mean, or statistical tests.

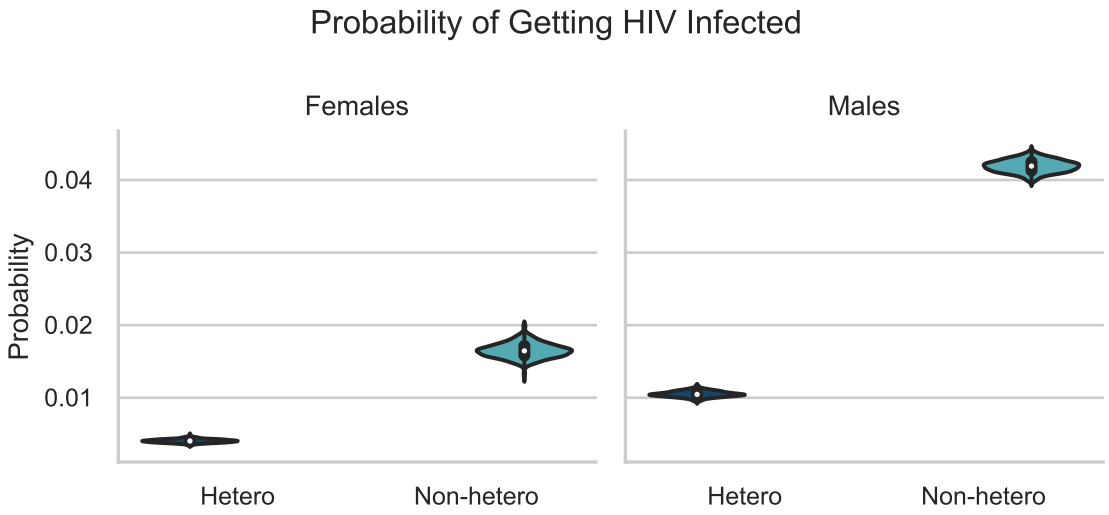

Figure 2: This visualization shows the prediction made by the logistic model that learned upon the dataset. Violins represent the distriubtion of probability preditction of getting HIV relative to gender and sexual identity.

Comparitivly the difference is significant, and non-heterosexual males seems to be at the highest risk of getting infected (when no other causes are considered). But there is also difference between genders, males again are at grater risk. In fact there are many publications that justifies the claim that woman are less susceptible to infection then man Beyrer et al. (2012). The risk for a woman to get infected during a intercourse with a infected person is smaller then for a man[5]. THIS REFERENCE TEST Figure 2

- Why non hetero woman are getting sick so common? It should be safer, mayybe confounders

But does it tell us something about the casual relationship? It certainly tell us something, but we should consider a whole range of underlying causes that might be responsible for such a result. Let's look at a DAG, represented as a Figure 3, that represents the reasoning that is along the lines of the first model.

The reasoning looks preatty poor, it sugggest that there is somehing 'special' about males and non-hetero sexuals that makes them to be more prone to HIV infection. Only thing that it does well is to support stereotypical view that HIV is an infection of homosexual males. So let's turn into the main goal of this work. That it is, let's

[5]Historically at first it was thought that only males can get AIDS deases which is caused by HIV virus.

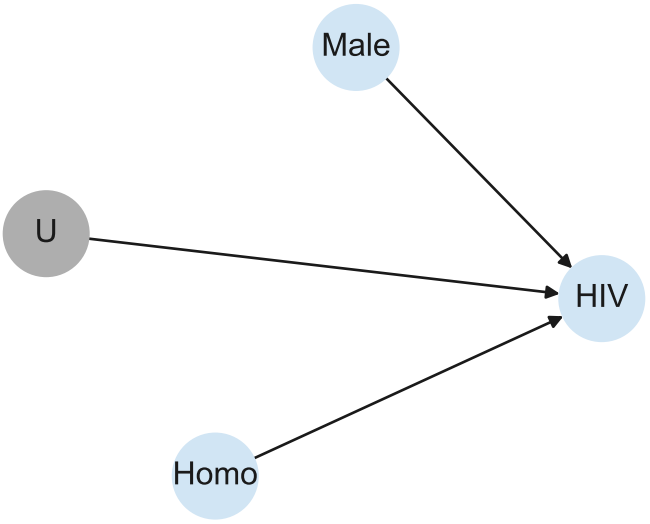

Figure 3: DAG representing naive relation between variables resembling the first model. 'U' stands for unobserved, the rest of the labels is selfexploanatory.

search for predictors that will be better at predicting HIV infection, that perhaps underlies this simple division between hetero and non-hetero people.

What predictors should we choose? The risk factors of HIV are a subject of many studies, we can use their findings and test on our dataset. According to WHO the list of the most common risk factors is "HIV and AIDS — Who.int" (2023):

"

- having condomless anal or vaginal sex;
- having another sexually transmitted infection (STI) such as syphilis, herpes, chlamydia, gonorrhoea and bacterial vaginosis;
- engaging in harmful use of alcohol and drugs in the context of sexual behaviour;
- sharing contaminated needles, syringes and other injecting equipment and drug solutions when injecting drugs;
- receiving unsafe injections, blood transfusions and tissue transplantation, and medical procedures that involve unsterile cutting or piercing; and
- experiencing accidental needle stick injuries, including among health workers.

"

As we are intrested in testing transmission based on sexual activities we will omit injective drug users and marginal cases of medical accidents when one can got infected from a patient's blood. But what can we add is the amount of sexual partners (in a year) that a person had, logic behind that is simple, more partners means more opportunities to get infected. Also, we will focus on anal sex as choosing

a risk factor, as it is the most dangerous sexual activitity whent it comes to HIV infection[6].

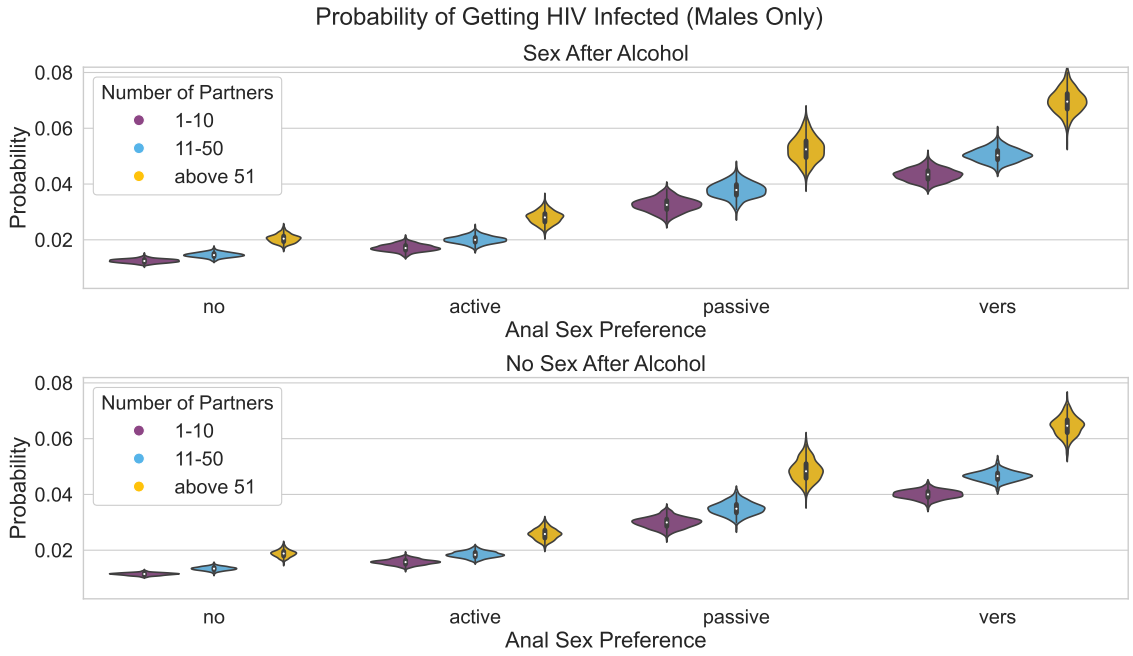

Figure 4: This is vis of males.

In the Figure 4 and Figure 5 you can see visualization of predictions created by the logistic regression model for males and females respectively. The model used the following predictors: gender, number of sexual partners (in a year), anal sex preference, having sex under the influence of alcochol. There is a striking difference between males and females, even when woman are in the most risky profile, the danger of infection is still mcuh lower than for the males. As I stated before, there are studies that are along this result caliming that woman are much less susceptible to HIV infection for bio-chemical reasons.

All that factors account to the most risky profile, a male that had above 51 sexual partners, prefers verastile anal intercourse, and does alcochol induced sex. It can be observe that any of those categories increases the probability of getting infected individually.

[6]I could have tested oral sex and danger of not using protection during oral sex, but because of troubling division between passive and active oral sex, the data variable seems to be not trustworthy. It is very common to mix up what passive and active oral sex means. By definition being on a receptive end is called passive, and a person using their mouth is called active. But in the gay community division between tops and bottoms (actives and passives) understand being receptive in oral sex as an active activity.

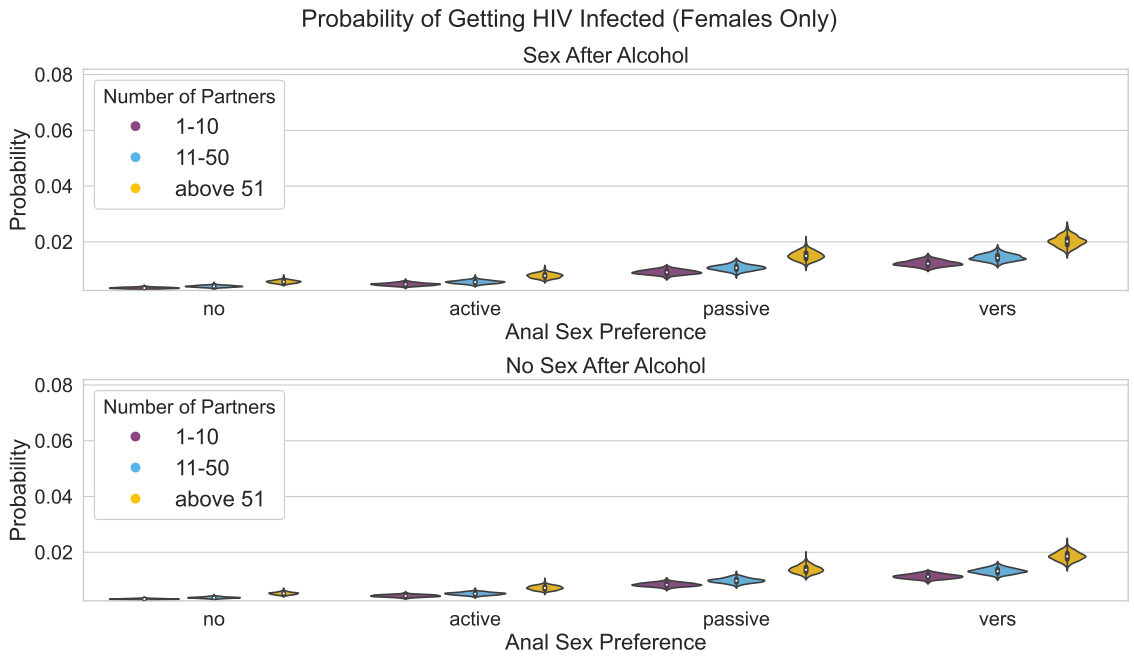

Figure 5: Females only.

Next section

The next step of my reaserch is to define a Risk Profile (RP) that maximizes the chances of getting HIV infected, the profile will not include sexual identity and gender. I will define the profile as a new binary variable in the dataset, this way will be much simpler than building a model that takes all these variables and creates a different category for all of them.

The variables that I will use to create the RP are:

- Alcochol induced sex,
- The amount of sexual partners a year > 10,
- Anal sex preference: passive or versatile,
- Anal sex protection use: sometimes or never.

In the Figure 6, that represents the probability of getting infected for the RP, we can observe that the probability for males that are in the RP group is bigger then a probability in our first model (represented in the Figure 2), that showed high probability of infection for non-hetero males. It suggests that RP profile holders are at greater risk of infection then non-hetero males. But what is the relation between those two variables? We will test with our next model.

In this pair of visualizations at Figure 7 we cann see the models that were give as predictors respectivaly a gender and sexual identity to predict being in the most risky group. As it can be seen non-hetero people and males has the highest probability in being in that group.

Let's maybe run WAIC test to compare those two models

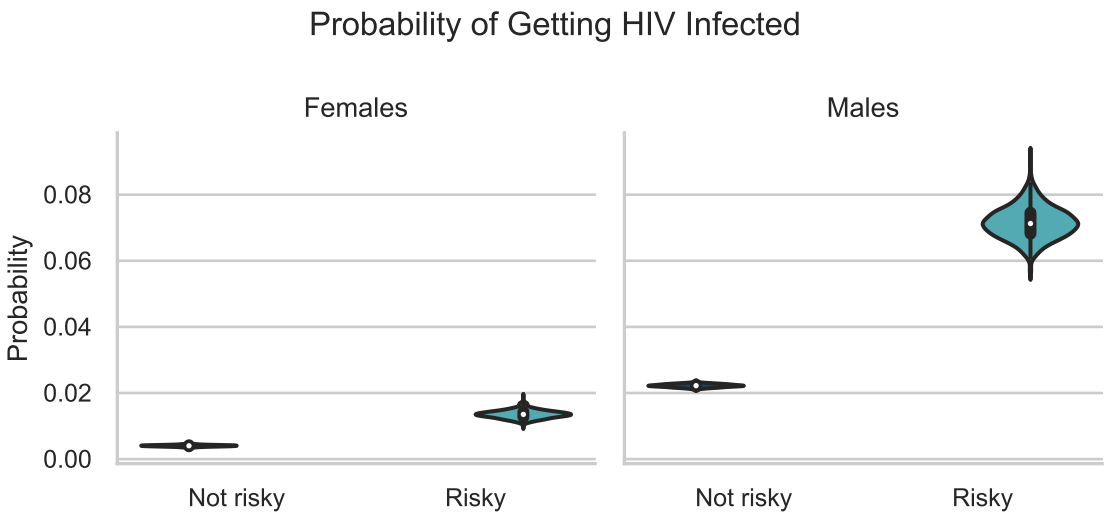

Figure 6: the caption of the plot

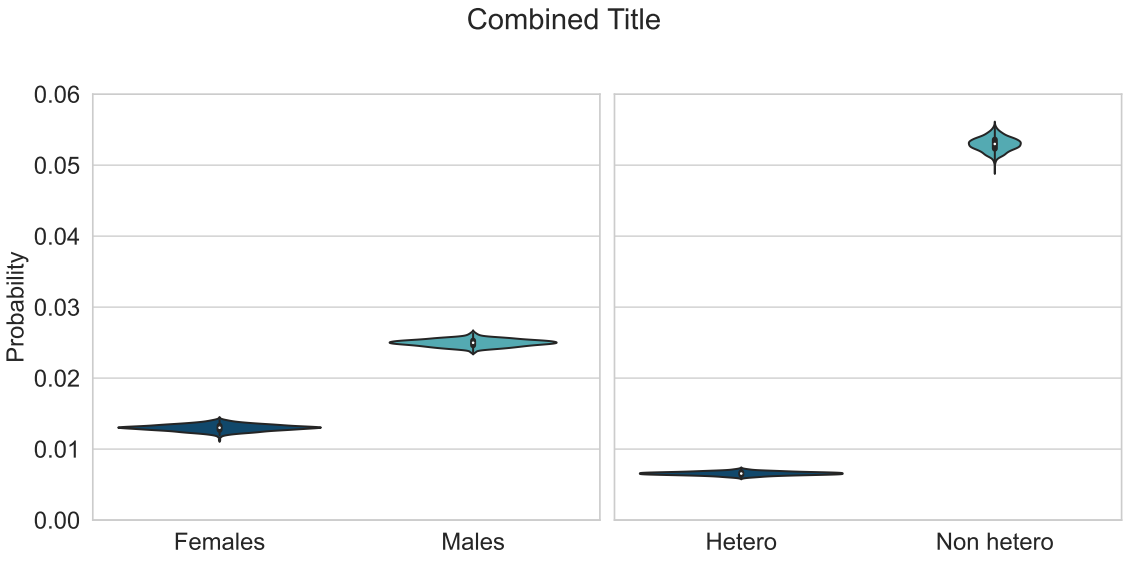

Figure 7: twin visualization

Getting back to representing the causes of getting HIV infected. The variables that increases the probability of getting HIV infected are better at getting you infected, therefore taking them as causes is more reasonable. The DAG that will represent the relation between the consdiered variables, after our analysis will look somewhere like this one Figure 8.

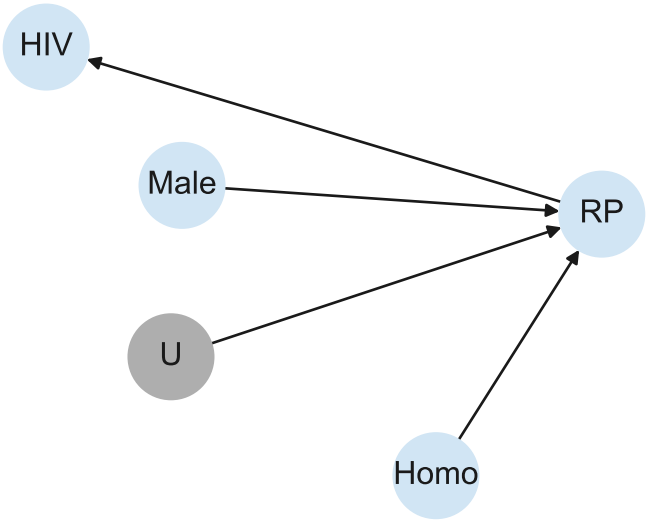

Figure 8: This DAG represents the relationship between demographic variables, RP - risk profile and getting HIV infected.

Conclusions

In this Figure 8 the DAG represents the relationshiop between predictors that we analyzed. It seems to represent the cause and effect more accurately, then just relying on a fact that non-hetero males are 'special' in some way and it makes them to get HIV infected. In fact non-hetero males has high probability of being in a RP group, risky behaviours assocaited with RP status are the causes of getting HIV infected. It includes having anal intercourse and high amount of sexual partners which is associated with non-hetero males. The reason why non-hetero males choses to be in that group is not-known, this fact is represented with Unobserved node.

Is the stereotype tru then? Well non-hetero males are in the highest risk of getting HIV infected, but their sexual identity is not a direct cause of that. Hetero sexual male that enjoys simmilar activities, so is in a RP also has higher chances of getting infected. The more direct cause is risky behaviour, and non-hetero males tend to engage in more risky behaviours. Therefore the story is more complicated. One thing is certain, the sole fact of being non-hetero person does not make you more susceptible to getting HIV infected.

Interesting observation that is in a field of interest of specialist from different science branches is the fact that woman gets infected far less often than males. My analyzis has confirmed that fact.

One thing worth attention is a worrying growth of new HIV infections in Europe. The mentioned stereotype might have a negative effect on public opinion, as hetero people might feel too confident about not getting infected, as this infection is associate mainly in non-hetero males. This analyzis showed that there is a risky behaviour pattern that all the people should know about that it can lead to HIV infection. In fact, as we have seen at the beginning, hetero males tend to test themselves much less often that non-hetero males.

References

Beyrer, Chris, Stefan D Baral, Frits van Griensven, Steven M Goodreau, Suwat Chariyalertsak, Andrea L Wirtz, and Ron Brookmeyer. 2012. "Global Epidemiology of HIV Infection in Men Who Have Sex with Men." The Lancet 380 (9839): 367–77. https://doi.org/10.1016/s0140-6736(12)60821-6.

European Centre for Disease Prevention and Control, and World Health Organization. 2022. HIV/AIDS Surveillance in Europe 2022 : 2021 Data. European Centre for Disease Prevention; Control. https://doi.org/10.2900/818446.

"Global HIV & AIDS Statistics — Fact Sheet — Unaids.org." 2023. https://www.unaids.org/en/resources/fact-sheet.

"HIV and AIDS — Who.int." 2023. https://www.who.int/news-room/fact-sheets/detail/hiv-aids.

Sharp, P. M., and B. H. Hahn. 2011. "Origins of HIV and the AIDS Pandemic." Cold Spring Harbor Perspectives in Medicine 1 (1): a006841–41. https://doi.org/10.1101/cshperspect.a006841.