

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 HIV epidemic, from that time many scientist around the world are trying their best to find new ways of prtecting global public from the infection and treating the people that got infected. Because of the aweranes that grew, and breakthrough medical achievemnts it is easy now to detect HIV virus in the blood of an individual, and we know that using condoms and PrEP is an effective way of not getting infected.

Thanks to that the global amount of new infections get very low, and in the developed countries, thanks to the new ways of treatment, people that are infected can live as long as not infected people. The global rate of infections is thankfully still getting lower. But from some time in some well developed countries we can see worrying spikes of new infections, HIV infections are growing. It is visible in the European Union (EU) where the common new number of infections is rising from some time. One of the factors is the globall Covid-19 epidemic that caused people to not getting tested for HIV (which is visible in our dataset). Poland is one of the countries where to problem is getting worse.

From the beggining of the HIV epidemic, this various was assosiated with non-heteronormative people, mostly with males. In those years gay man were much less accepted then other sexual identities. Gay clubs were one of the types of places where they can feel free, safe and accepted. Being anonymous was a virtue, and a behaviour of having anonymous sex with random people was an occurence for some. Before the knowledge about HIV using condoms was much less populat, therfore this kind of sexual behaviour was very risky.

This stereotype of HIV being “the deases of gays” is still strong. In this work I want to confront it, by analyzing the causes of getting infected. I want it this work to be centred around causal relation. The main questions of my work are:

- does males in general have higher probability of contracting HIV?
- does homosexual are at the higher risk than heterosexuals?

Main goal: Let's confront a common stereotype that HIV is mainly male gays disease. What so special about gay man that makes them most HIV sensitive?

I will try to answer these question by the analyzis of surveys conducteed in Polish HIV testing centres. The testing centres are called PKDs (pol. Punkt Konsultacyjno Diagnostyczny - Diagnostic Consultation Center), there is at least one in every major administratory region in Poland (Voivodeship). Testing there is always: anonymous, free of charge and save. The clients of this centre are always surveyed by the medical professional that qualifies them for a test. The survey has many interesting varioubles, some we will put to the test as the one of the survey's section is centred around sexual life.

The data were collected from 2015 to 2022, and have more than 250000 respondents. One thing to remember, this dataset is not an accurate source of information on the number of new infections, as there are various other places in Poland that can diagnose HIV, therefore we are focusing on the survey questions. On the visualizations below you can see the demography of the survey. PLOT on the left represents the share of gender¹ and sexual identity among the PKD clients. Majority of the clients are males, interestingly non-hetero normative males are the half of all the male clients, therefore much more than there are non-hetero normative man in the population. Females however are represented far more accurately.

The plot on the right represents sexual identity and gender among HIV positive test results². Around 70% of HIV positive people are non-hetero normative males. The result motivates further analysis, as it turned out that non-hetero norm. males are far more susiceptable to getting infected. The one positive thing about thatm is that this group of people seems to be aware of the danger, since there are over represented as a group of people that test themselves. One thing to note, in the dataset every client has their unique ID number, but as the survey is fully anonymous, people that test themselves regurarly will not be represented as one person, and some people that regurarly has MSM (sex between two males) justify that test themselves e.g. monthly. This might be a reason (or one of the reasons) why this group is overrepresented.

The question arises, is it true then that there is something special about non-hetero normative males that they are getting infected much more often? I will try to answer that by looking for causal links between various predictors, mainly searching for variables that are best at predicting being HIV positive.

¹22 people identified as 'Different' gender their number is to small to be considered significant, therefore there are excluded from the analysis.

²Those people that will be considered HIV positive in this analysis had two tests, one the usual screening test that yieldded positive result, and the second that measured the amount of HIV virus in the blood, that confirmed the first result

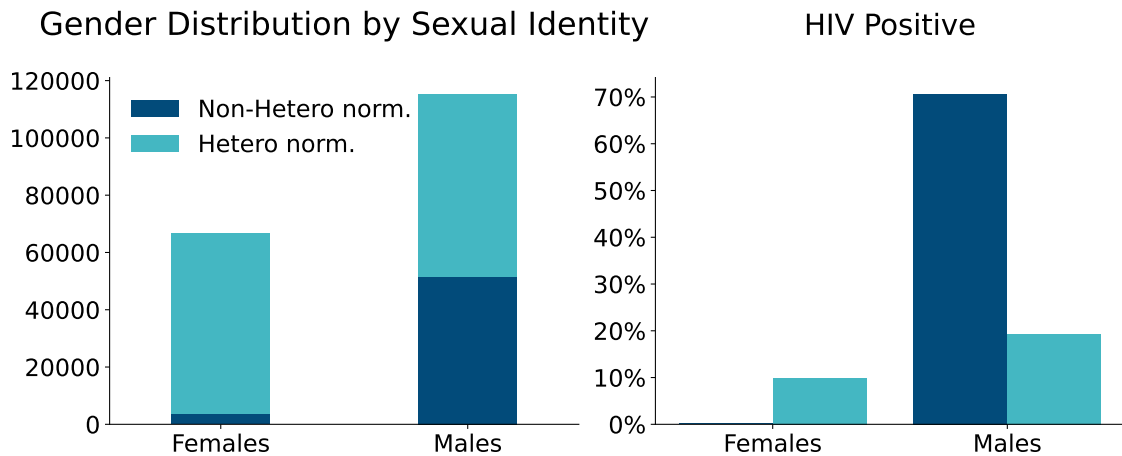


Figure 1: hello

First section

Let's first try to build a model that will try to predict being HIV infected based on variables that we already know, so sex and sexual identity. All the statistical models in this project were built using 'numpyro' package in python [link to package documentation]. It is a simplified version of 'pyro' package, both are focused on delivering tools for probabilistic data analysis, including building advanced hierarchical 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 approximation...

Let's see how the relation between our demographical predictors and being HIV positive look like. In the Figure @FigModelSexHetero you can see the prediction made by the model that as predictors took Gender and sexual identity of the clients. The violins represent the distribution of predictions, they are not single point estimates. It is characteristic for bayesian approach to always interpret distributions as they convey more information than e.g. mean, or statistical tests.

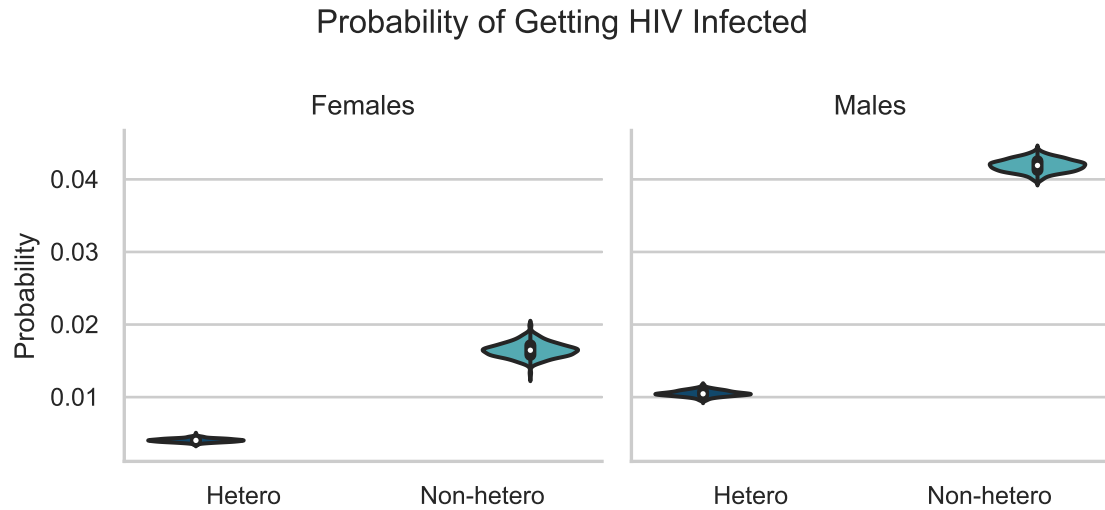


Figure 2: This visualization shows the prediction made by the logistic model that learned upon the dataset. Violins represent the distribution of probability prediction of getting HIV relative to gender and sexual identity.

Comparatively 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 greater risk. In fact there are many publications that justify the claim that women are less susceptible to infection than men. [REF LANCET and Others] The risk for a woman to get infected during an intercourse with an infected person is smaller than for a man³. THIS REFERENCE TEST Figure 2

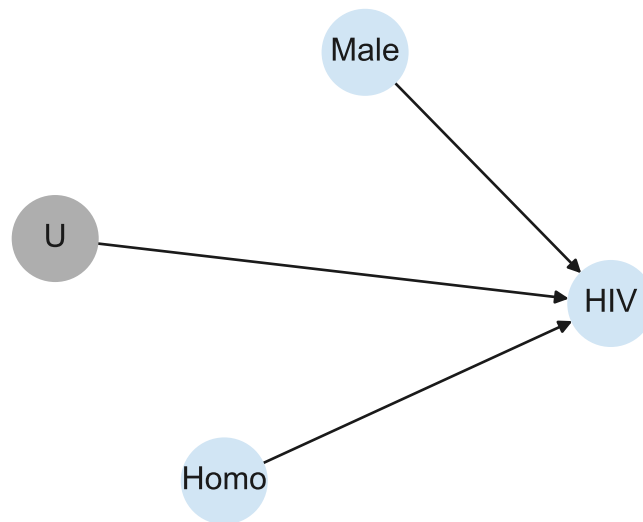


Figure 3: DAG representing naive relation between variables resembling the first model. ‘U’ stands for unobserved, the rest of the labels is self-explanatory.

³Historically at first it was thought that only males can get AIDS diseases which is caused by HIV virus.

- Why non hetero woman are getting sick so common? It should be safer, maybe 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 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 [<https://www.who.int/news-room/fact-sheets/detail/hiv-aids>]:

“

- 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 activity whent it comes to HIV infection⁴.

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

⁴I 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.

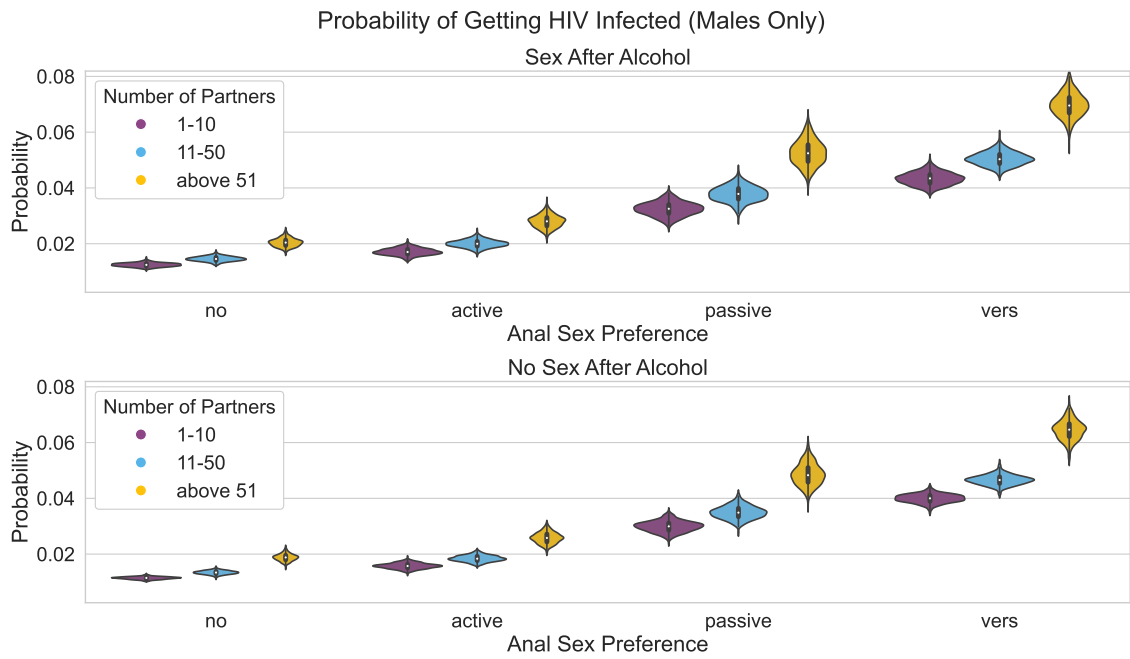


Figure 4: This is vis of males.

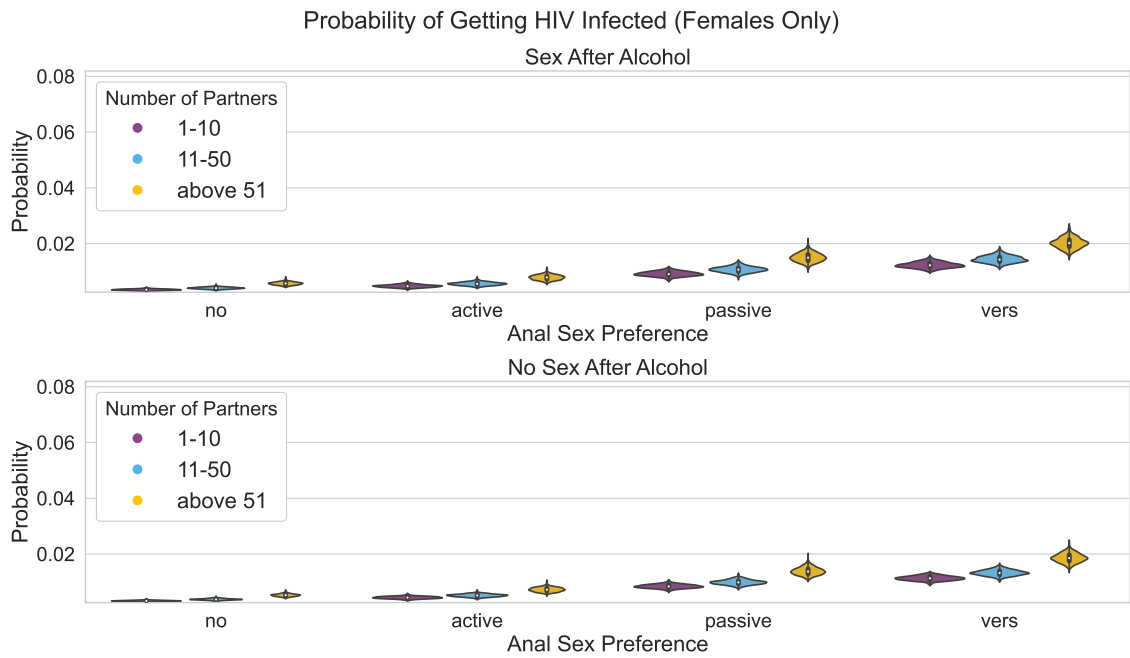


Figure 5: Females only.

danger of infection is still much lower than for the males. As I stated before, there are studies that are along this result claiming that women 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 versatile anal intercourse, and does alcohol induced sex. It can be observe that any of those categories increases the probability of getting infected individually.

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:

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

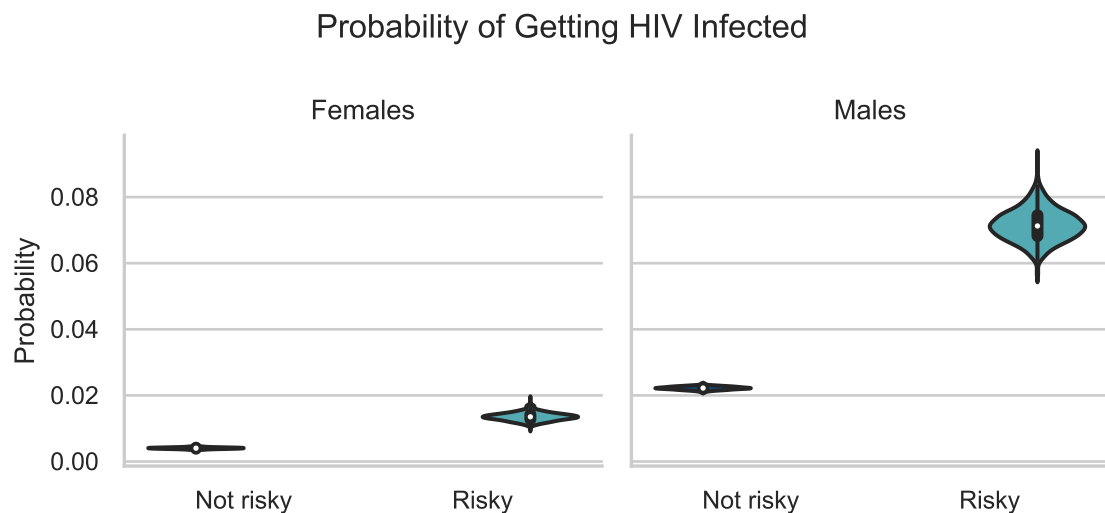


Figure 6: the caption of the plot

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 can 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

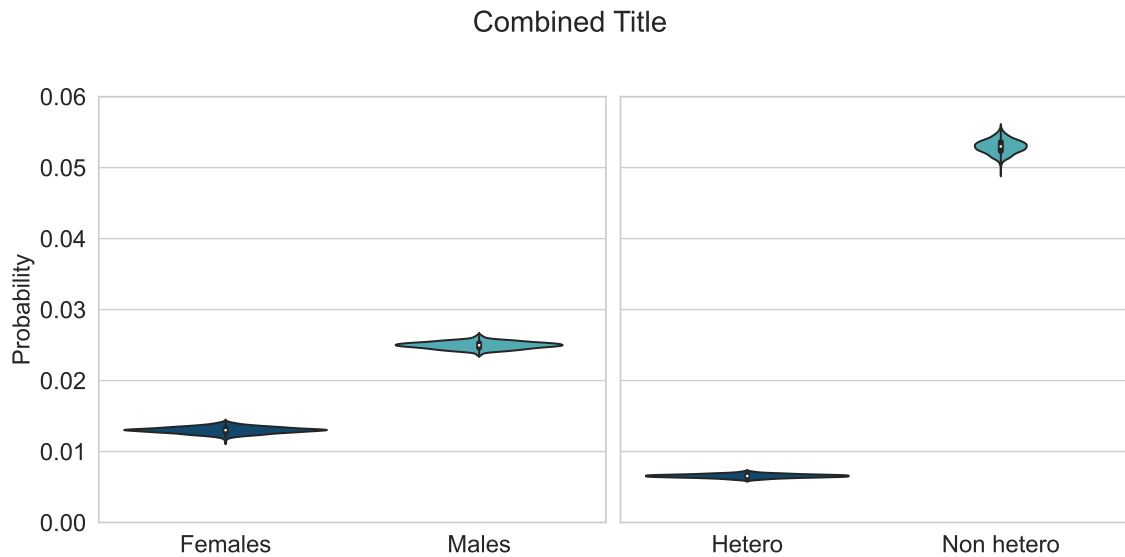


Figure 7: twin visualization

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.

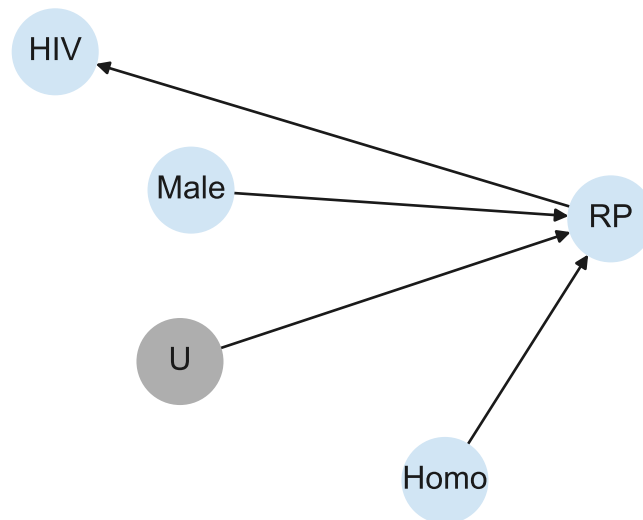


Figure 8: This DAG represents the relationship between demographic variables, RP - risk profile and getting HIV infected.

Conclusions

In this Figure 8 the DAG represents the relationship 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 associated 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 chooses to be in that group is not-known, this fact is represented with Unobserved node.

Is the stereotype true 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 similar 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 analysis 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 associated mainly in non-hetero males. This analysis 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 than non-hetero males.