Bayesian Modeling of HIV Risk Factors:

Confronting HIV as a Stereotypical Infection of Sexual Minorities

Nikodem Lewandowski

Abstract: My goal is to challenge the stereotype associating HIV infection primarily with non-heteronormative males. Using Bayesian logistic regression models and risk profile analysis, I argue that risky behaviors, rather than sexual identity or gender, are the primary drivers of HIV infection.

The research employs a dataset created from a survey conducted in Polish HIV diagnostic centers. Poland is currently experiencing a rise in new HIV infections, with non-heteronormative males constituting the majority of new cases. However, the findings reveal that high-risk behavior serves as a better predictor of infection than gender or sexual identity. This highlights the complexity of the factors driving new infections among non-heteronormative males.

This research underscores the critical need for comprehensive awareness and testing to combat the ongoing HIV epidemic. It prompts a reevaluation of the assumptions surrounding the causes of the domination of non-heteronormative males among newly infected new infections.

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.

---

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

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.

PROBLEMATIC PASSAGE 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[2] 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[3], 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

---

[2]Link to the github repository of this project can be found here

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

gender distribution[4]. 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[5]. Approximately 70% of HIV-positive 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.

In the first section, I will begin by establishing a basic logistic regression model that demonstrates the relationship between gender, sexual identity, and the probability of HIV infection. The subsequent step involves an exploration of whether a causal connection between these variables is plausible. Following this, I delve into the quest for more robust predictors. During this investigation, I identify superior predictors and formulate a risk profile (RP), which exhibits a higher probability of infection.

Remarkably, the group most likely to be within the RP comprises non-heteronormative males, shedding further light on the causal relationships among the aforementioned predictors. This relationship is visually represented with a Directed Acyclic Graph (DAG) at the conclusion of the study

These findings challenge the stereotype, asserting that the primary cause of HIV infection is not one's sexual identity but rather risky behavior.

```
C:\Users\nikod\AppData\Local\Temp\ipykernel_15480\2214647407.py:35: UserWarning: se
  ax2.set_xticklabels(['Females', 'Males'], rotation=0)
```

---

[4]22 individuals identified as 'Different' gender; however, their number is too small to be considered significant and is therefore excluded from the analysis

[5]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.
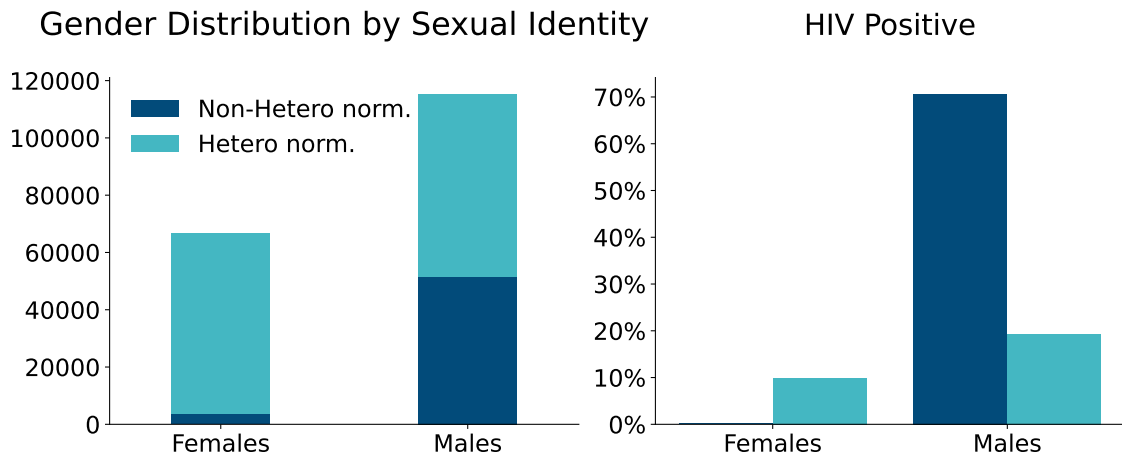
Figure 1: The left plot shows the dataset's gender and sexual identity distribution, while the right plot displays the ratio among those who tested positive.

Exploring HIV Risk Factors and Predictors

Let's begin by constructing a model aimed at predicting HIV infection based on known variables: gender and sexual identity. All statistical models in this project were built using the 'numpyro' package in Python[6]. It's a simplified version of the 'pyro' package, both of which focus on providing tools for probabilistic data analysis, including the creation of advanced hierarchical models. Specifically, I employed Bayesian-style logistic regression. The primary distinction in these models lies in the ability to specify the prior distribution of the predictors. The model is run using the No-U-Turn Sampler (NUTS), a variant of Markov Chain Monte Carlo (MCMC). Here's a template of how the model looks like:

```python
dat_listGender = {
    'HIV': PKD_model_DF.HIV.values,
    'Hetero_normative': PKD_model_DF.Hetero_normative.values,
    'Gender_encoded': PKD_model_DF.Gender_encoded.values,
    }


def model_hetero_Gender(Hetero_normative, Gender_encoded, HIV=None):
    a = numpyro.sample("a", dist.Normal(0, 1.5))
    b = numpyro.sample("b", dist.Normal(0, 0.5).expand([2]))
    c = numpyro.sample("c", dist.Normal(0, 0.5).expand([2]))

    logit_p = a + b[Hetero_normative] + c[Gender_encoded]
    numpyro.sample("HIV", dist.Binomial(logits=logit_p), obs=HIV)
```

---

[6]The package documentation can be found here.

```
HIV_HetreoNorm_Gender = MCMC(NUTS(model_hetero_Gender),
    num_warmup=500, num_samples=500)
HIV_HetreoNorm_Gender.run(random.PRNGKey(0), **dat_listGender)
```

Now, let's explore the relationship between our demographic predictors and the likelihood of being HIV positive. In Figure 2, you can observe predictions made by the model, which utilized gender and sexual identity as predictors. The violin plots represent the distribution of predictions, rather than single point estimates. This approach aligns with the Bayesian methodology, as it conveys more information than, for instance, just a mean or statistical tests.
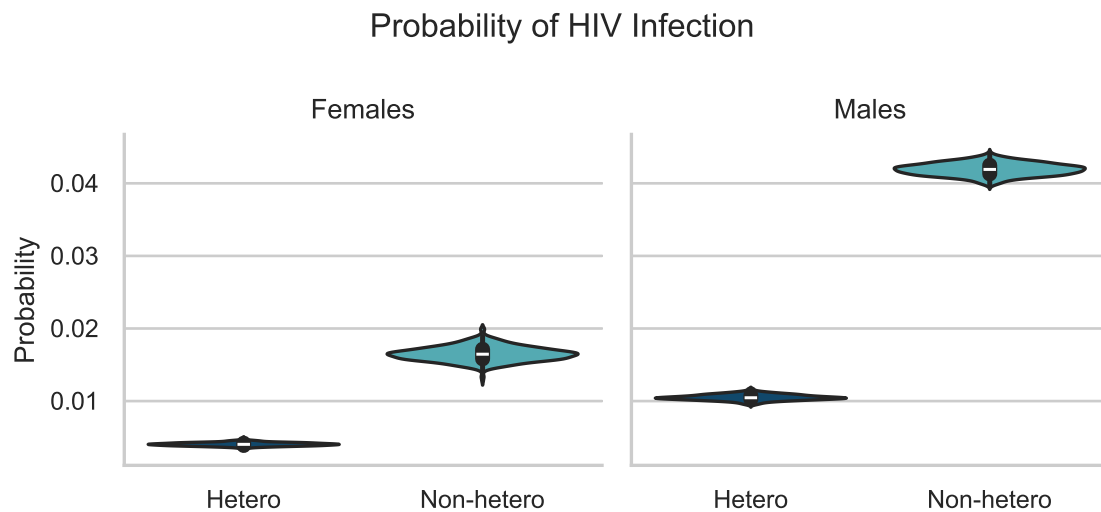

Figure 2: This visualization shows the predictions made by the logistic model trained on the dataset. Violins represent the distribution of probability predictions of acquiring HIV relative to gender and sexual identity.

Comparatively, the differences are significant, with non-heterosexual males appearing to be at the highest risk of infection when considering no other factors. However, there are also distinctions between genders, with males once again being at greater risk. In fact, there are many publications supporting the claim that women are less susceptible to infection compared to men (Beyrer et al. (2012)). The risk of a woman contracting HIV during intercourse with an infected person is lower than that for a man.[7]

But does this information provide us with insights into causal relationships? It certainly provides some insights, but we should consider a wide range of underlying causes that might be responsible for such results. Let's examine a Directed Acyclic Graph (DAG), represented as Figure 3, which reflects the reasoning of the initial model

---

[7]Historically, it was initially believed that only males could contract the AIDS disease caused by the HIV virus.
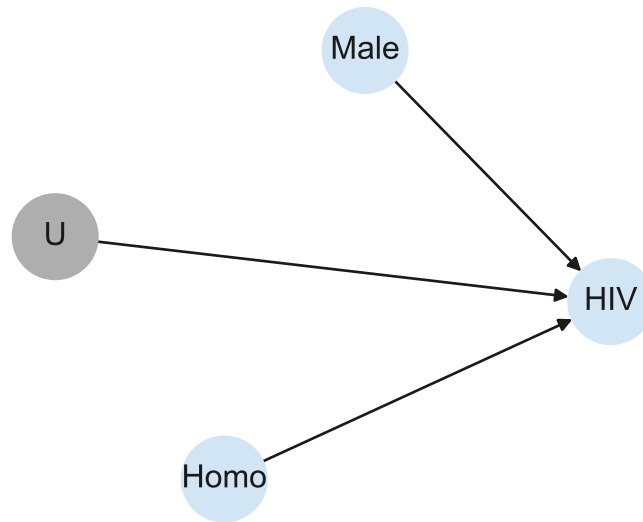
Figure 3: DAG representing the naive relation between variables resembling the first model. 'U' stands for unobserved, while the rest of the labels are self-explanatory.

The reasoning appears to be somewhat limited, suggesting that there is something 'special' about males and non-heterosexual individuals that makes them more prone to HIV infection. The primary achievement of this reasoning is to reinforce the stereotype that HIV is primarily an infection of homosexual males. Let's now shift our focus to the main objective of this work. We aim to identify predictors that may offer a better explanation for HIV infection, potentially going beyond the simplistic division between heterosexual and non-heterosexual individuals.

What predictors should we consider? The risk factors for HIV have been the subject of numerous studies, and we can leverage their findings in our dataset. According to the World Health Organization (WHO), the most common risk factors include:
"

- 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. ("HIV and AIDS — Who.int" 2023)

"

Since we are specifically interested in examining transmission based on sexual activities, we will exclude injective drug users and marginal cases of medical accidents where infection might occur through contact with a patient's blood. However, we can add the variable representing the number of sexual partners (per year) a person has. The rationale here is simple: more partners mean more opportunities for infection. Additionally, we will focus on anal sex as a chosen risk factor, given that it is the most high-risk sexual activity[8].
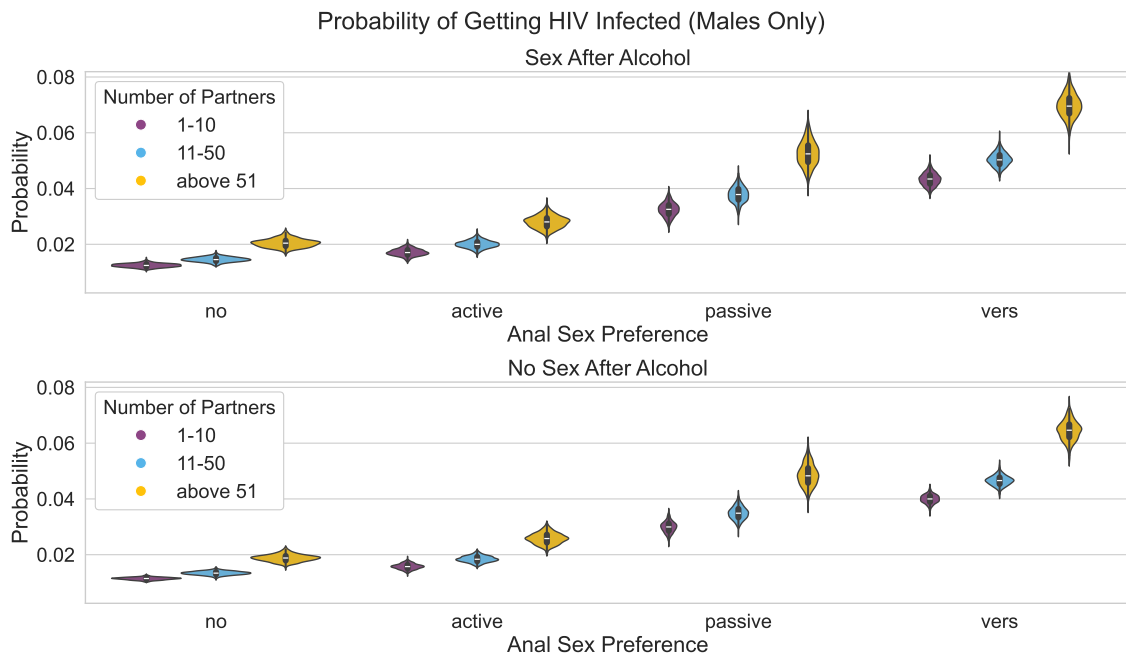

Figure 4: In this visualization, you can observe various categories considered as contributors to a risky profile. This plot focuses solely on males.

In figures: Figure 4 and Figure 5, you can observe visualizations of predictions created by the logistic regression model for males and females, respectively. The model utilized the following predictors: gender, number of sexual partners (per year), preference for anal intercourse, and engaging in sexual activity under the influence of alcohol. There is a striking contrast between males and females. Even when women fall into the most high-risk profile category, the risk of infection remains considerably lower than for males. As mentioned earlier, studies support these findings, suggesting that women are significantly less susceptible to HIV infection for biochemical reasons.

---

[8]We could have tested oral sex and the risk of not using protection during oral sex, but due to the ambiguous distinction between passive and active oral sex, the data variable seems unreliable. There is common confusion between what constitutes passive and active oral sex. By definition, being on the receptive end is called passive, while using one's mouth is called active. However, in the gay community, the division between tops and bottoms (actives and passives) often perceives receptive oral sex as an active activity.
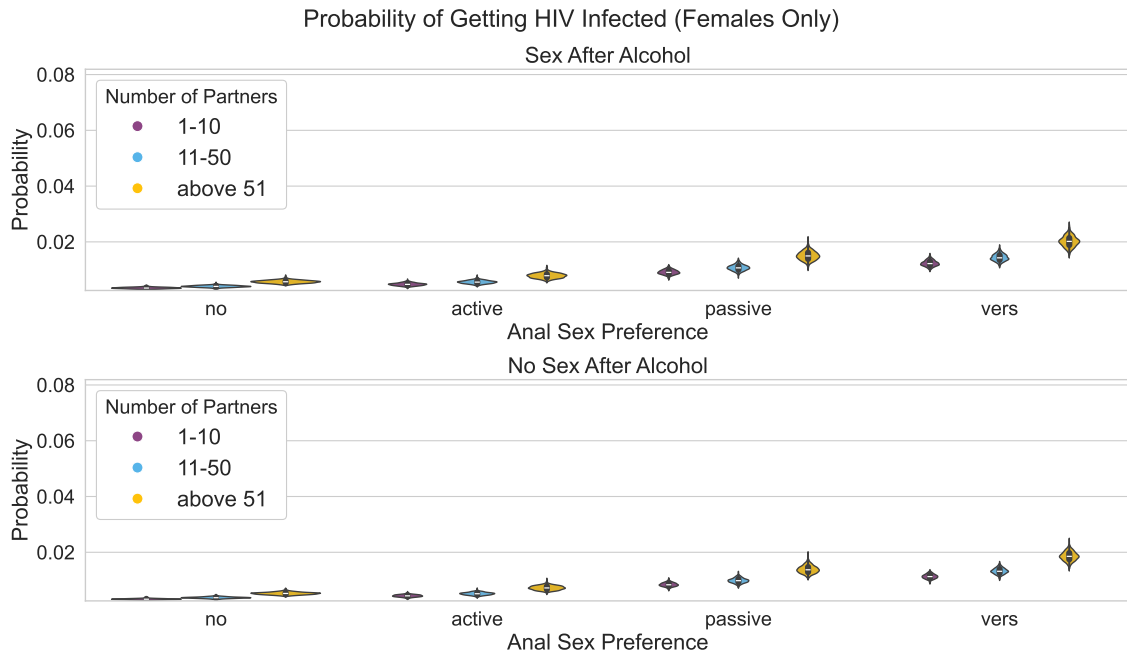
Figure 5: This is a twin visualization covering risky categories. This plot focuses exclusively on females, and the difference in the probability of getting infected is striking.

All these factors contribute to the definition of the most high-risk profile: a male with more than 51 sexual partners, a preference for versatile anal intercourse, and engaging in alcohol-induced sex. It's evident that each of these categories individually increases the probability of infection.

Defining a High-Risk Profile for HIV Infection

The next step in my research involves defining a Risk Profile (RP) that maximizes the chances of HIV infection. This profile will **exclude sexual identity and gender** in order to empasize the influence of other variables. I will introduce the profile as a new binary variable in the dataset. This approach is simpler than constructing a model that accommodates all these variables and creates different categories for each of them.

The variables I will use to create the RP are:

- Alcochol-induced sex,
- Having more than 10 sexual partners per year,
- Anal sex preference: passive or versatile,
- Anal sex protection use: sometimes or never.

In Figure 6, which represents the probability of infection for the RP, we can observe that the probability for males in the RP group is higher than the probability in our first model (represented in Figure 2), which showed a high probability of infection for non-
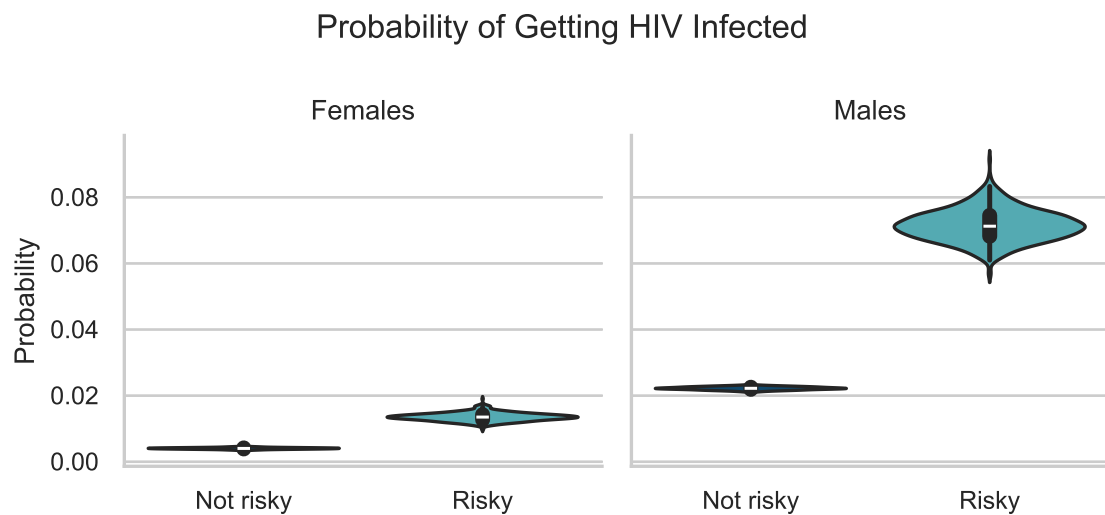
Figure 6: The visualization represents the probability of getting infected for RP holders, divided into females and males.

hetero males. This suggests that RP profile holders are at a greater risk of infection than non-hetero males. But what is the relationship between these two variables? We will test this with our next model.

```
C:\Users\nikod\AppData\Local\Temp\ipykernel_15480\1486812129.py:14: FutureWarning:

Passing 'palette' without assigning 'hue' is deprecated and will be removed in v0.1

  sns.violinplot(data=DFPost_Post_RISK_NormGendernoHetero, x="Gender", y="Probabili
C:\Users\nikod\AppData\Local\Temp\ipykernel_15480\1486812129.py:32: FutureWarning:

Passing 'palette' without assigning 'hue' is deprecated and will be removed in v0.1

  sns.violinplot(DFPost_POST_RISK_Hetero, x="HeteroNorm", y="Probability", ax=axs[1
```

In the pair of visualizations in Figure 7, we can see the models that were given gender and sexual identity as predictors to predict being in the most risky group. As it can be seen, non-hetero individuals and males have the highest probability of being in that group.

Returning to representing the causes of HIV infection, the variables that increase the probability of infection are better at getting you infected. Therefore, considering them as causes is more reasonable. The Directed Acyclic Graph (DAG) that will represent the relationship between the considered variables, after our analysis, will look something like the one shown in Figure 8.
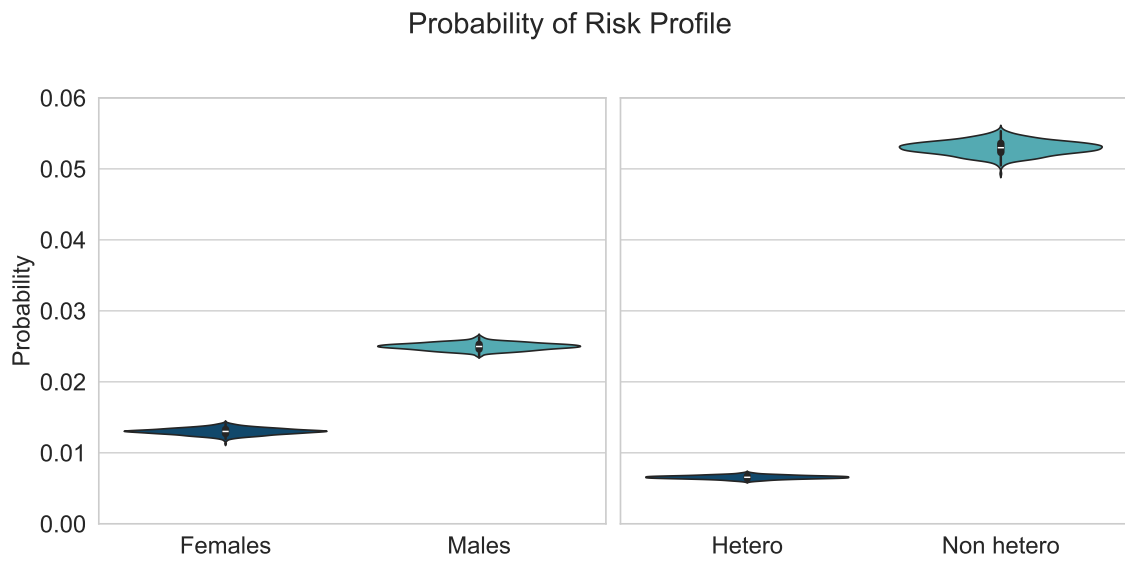
Probability of Risk Profile

Figure 7: The visualization represents the probability of being in RP, with a division by gender on the left and by sexual identity on the right.
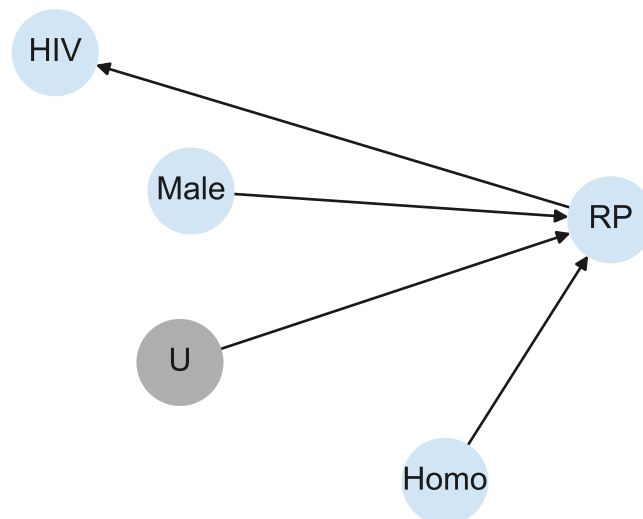

Figure 8: This DAG represents the relationship between demographic variables, the RP (risk profile), and HIV infection.

Conclusions

In the Figure 8, the DAG represents the relationship between the predictors we analyzed. It seems to represent cause and effect more accurately than just relying on the notion that non-hetero males are 'special' in some way, leading to HIV infection. In fact, non-hetero males have a high probability of being in an RP group, and risky behaviors associated with RP status are the causes of HIV infection. This includes having anal intercourse and a high number of sexual partners, which is associated with non-hetero males. The reason why non-hetero males choose to be in that group is not known, represented by the Unobserved node.

Is the stereotype true then? Well, non-hetero males are at the highest risk of getting HIV infected, but their sexual identity is not a direct cause of that. Hetero sexual males who enjoy similar activities, and thus are in an RP, also have higher chances of getting infected. The more direct cause is risky behavior, and non-hetero males tend to engage in more risky behaviors. Therefore, the story is more complex. One thing is certain: the sole fact of being a non-hetero person does not make you more susceptible to getting HIV infected.

An interesting observation in the field of interest to specialists from different scientific branches is the fact that women get infected far less often than males.

One thing worth noting is the concerning 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, assuming this infection is mainly associated with non-hetero males. This analysis has shown that there is a risky behavior pattern that all people should be aware of, as 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.

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.