**Background**

Human immunodeficiency virus (HIV), the virus causing acquired immunodeficiency syndrome (AIDS), is a preventable infection that can lead to severe, life-long, and often fatal consequences in humans. The HIV virus exists in two main infectious sub-types: HIV-1 and HIV-2. The most common ways that HIV can spread is through open-wounds or cuts where blood from an infected individual transfers to into circulation of another individual usually through injection-drug users who share needles with persons with HIV, anal sex, and, in some cases, vaginal sex. To accurately allocate resources to reduce the spread of HIV, updated monitoring of the main risk groups involved in spreading HIV infections, including persons who inject drugs (PWIDs), men who have sex with men (MSM), and heterosexual individuals are necessitated. Using retrospective data on HIV infections collected by the Epidemiological Surveillance Network of Castilla y León between 2009 and 2021, the etiology of HIV cases were analyzed and the relative risk of AIDS diagnoses among each risk group were compared.

**Objectives**

1. To determine the distributions of age and sex of the reported HIV cases between 2009 and 2021 and identify any shifts in the probability of infection by age and sex.

2. To determine the probaiblity of early, asymptomatic, and AIDS infections among HIV positive cases between 2009 and 2021.

**Data source**

Data were from the Publications Office of the European Union data repository (link: <https://data.europa.eu/en>). The data set was collected by the "New HIV Infections Register" of the Epidemiological Surveillance Network of Castilla y León (link: <https://data.europa.eu/data/datasets/https-datosabiertos-jcyl-es-web-jcyl-set-es-salud-registro-vih-1284868676866?locale=en>; retrieved 2023-08-21).

All downstream uses of the data follow Creative Commons Attribution 4.0 International (link: <https://creativecommons.org/licenses/by/4.0/>).

**Methods/Material**

*Data wrangling*

Data were cleaned using the `dplyr::` package in R (v4.3.1), specifically character values of `Risk Group` and `State of Clinical Infection` were translated to English using Google Translate.

*Inclusion criteria*

Adults (\>= 18), diagnosed with HIV-1 sub-type infections, and known risk group from Spain. After applying inclusion criteria 981 cases eligible for downstream analyses.

*Exclusion criteria*

`PWID + MSM/BISEX` and `Unknown` risk groups, ages over 58, and year 2021 were omitted from downstream analyses due to low sampling or ambiguity.

*Statistical analysis*

Descriptive statistics were calculated for age (median age at diagnosis, quantiles) and sex ( of cases by sex) among HIV-1 infections. Density plots were generated to visualize the distribution of age at diagnosis stratified by sex and diagnosis year, accompanied by bar graphs of age at diagnosis over years stratified by sex.

Splined multinomial logistic regression (MLM) was used to determine the probability of HIV diagnoses by age group [ref: 18-28] by years and adjusted for sex [ref: Male]. Due to the non-linear trend over years, diagnosis year was splined with 2 degrees of freedom to better fit the distribution without over-fitting.

MLM was used to determine the probabilities of each disease status (early infection [ref], asymptomatic, AIDS) by year adjusted for sex [ref: Male] and age group [ref: 18-28] among HIV-1 infections. Binomial logistic regression (BLM) was used with AIDS [1] and Early-infection [ref], adjusting for age (continuous) and sex.

BLM was used to determine the odds and probability of an AIDS diagnosis [1] given the risk group (HETERO [ref], PWID, MSM/BISEX) adjusted for age group and sex.

**Results**

*Distribution of age and sex among HIV-1 infected patients*

* More men were diagnosed with HIV than women between 2009-2021 (836 vs 145). This remained true regardless of the year of study.
* The overall median age of diagnosis was 38.5 (quantiles: 36-40). Males tended to be diagnosed earlier than women (overall median age: 37 vs 41). However, this was highly dependent on the year of study, especially towards the later years of the study, and was limited to few female cases.

**