2 main databases were used in the analysis in the Project: the Stanford HIV Resistance Database, and the Palma Database. Both are public databases that have a whole litany of information from previous studies that are allowed to be used for meta-analysis and data analysis for public health and socioeconomic reasons.

The Stanford Data database columns:

Country - The country where the study took place.

Period - The year the study occurred. Studies with a duration of longer than three years were excluded from the analysis.

NRTI - Levels of Nucleoside reverse transcriptase inhibitor Resistance per study (%)

NNRTI - Levels of non-nucleoside reverse transcriptase inhibitor resistance per study (%)

Num.People - The number of people in the study. Studies with less than 40 people were removed from the database and not analyzed to help reduce the variance.

Additionally, 2 columns were made during the cleaning of the data. These columns were made for ease of analysis:

Num.NRTI - This was calculated by multiplying the percent of NRTI resistance by the number of people in the study. The decimals were truncated.

Num.NNRTI - This was calculated by multiplying the percent of NNRTI resistance by the number of people in the study. The decimals were truncated.

The Palma ratio data set featured 3 important columns:

Entity - The sovereign nation at the time whose Palma ratio belongs to.

Year - The year corresponding to the Palma ratio value.

Palma Ratio (Pre-Tax) - The top 10% in income compared to the bottom 40% in income per country in that year.

The HIV resistance database was separated by year, and studies that had a multiyear length were averaged if the duration was under three years with the decimal being truncated. Or, the study was removed if the iteration was over three years. Additionally, the two made columns were self-calculated for ease of analysis with a Poisson distribution regression as listed above. All studies that had participants from multiple countries were removed from the database, so they cannot be analyzed due to being unable to isolate each study member from their respective country for

privacy reasons. Additionally, all studies with less than 40 participants were removed from analysis to help reduce variance.

The Palma ratio database was then joined to the corresponding year and country once the Entity and Year columns had been transformed into Countries and Period. These ratio values, once added to their corresponding country and year, were summarized into a separate dataset called Country_Resistance by country in year so that only 5 columns Remained (Countries, Year, mean_people, mean_num.NRTI, mean_num.NNRTI, palma ratio).

Note: These columns refer to the averaging done of the number of people in each study per country per year, as well as the number of people who had NRTI &NNRTI resistance per country and year. The Palmer ratio was not average.