The Statistical Illusion of a Global HIV Endgame

Shaoming Chen[1]*
[1]Rheast LLC, 1331 Lamar St, Houston, TX 77010, USA

Abstract

This study analyzes prevalence growth rates across regions using UNAIDS data and statistical modeling to forecast trends through 2050. Globally, the targets are projected for 2033, 2030, and 2025, driven by Eastern and Southern Africa's success from robust interventions. However, this masks rapid growth in Asia and the Pacific, the Caribbean, and the Middle East and North Africa, where rising infections and limited treatment coverage reveal stark disparities. These regional crises challenge the framework's global applicability, suggesting Africa's progress cannot be extrapolated universally. Relying on 95-95-95 risks an illusory "endgame" without tailored strategies for diverse epidemiological contexts. This study used a random forest model analyzing feature correlations with HIV growth. The first 95 targets exhibit strong negative correlations, indicating that higher testing substantially reduces growth. Conversely, the LGBT score and urban population show positive correlations, suggesting elevated growth in urbanized LGBT communities. The third 95 targets display weaker correlations, showing that growth is driven by those who have not been tested. Prevention should prioritize encouraging untested individuals to undergo testing in urban LGBT populations.

Keywords: HIV, AIDS, epidemiology, epidemic modelling

Background

UNAIDS, a key agency of the United Nations, has established the 95-95-95 Targets to end the AIDS epidemic by 2030 [1], aligning with global health objectives. These targets aim to reduce new HIV infections and AIDS-related deaths while ensuring that the majority of people living with HIV have access to life-saving treatment. Although significant progress has been achieved in some areas, the global response remains uneven, and multiple factors continue to hinder these efforts.

One notable achievement has been a 60% reduction in new HIV infections since 1995 [2], but many regions are falling behind on key targets. In Asia and the Pacific, for example, only 78% of people living with HIV know their status, which is far from the 95% target. Globally, 9 million people still lack access to treatment, increasing their risk of poor health outcomes. [3] Beyond economic factors, challenges such as inadequate healthcare infrastructure, political instability, social stigma, and limited access to education and prevention programs significantly influence the effectiveness of HIV interventions [4].

This research will analyze current data on new infections and related factors to project future trends. By examining these factors, the study aims to identify gaps in the response and offer actionable insights to improve intervention strategies.

Methods

This study employs a comprehensive approach to analyze trends in HIV/AIDS and forecast future trajectories through 2050, utilizing global health and demographic data. The methodology integrates quantitative data analysis and statistical modeling to evaluate progress toward the UNAIDS 2030 goals and predict long-term outcomes.

The primary data source for this research is the UNAIDS HIV estimates with uncertainty bounds, covering the period from 1990 to the present [4]. These documents provide detailed statistics on HIV prevalence, incidence, testing, treatment, and mortality, along with uncertainty ranges to account for variability in reporting. The analysis begins with an examination of global trends to establish an overall picture of the epidemic's evolution. Subsequently, the data are disaggregated to explore regional patterns, followed by a detailed investigation of individual countries where sufficient data are available. Countries with no data are excluded from the analysis.

Python is used as the primary programming language for data analysis, offering robust tools for handling large datasets and performing statistical modeling. Specifically, the Pandas [5] library is employed to process Excel tables from UNAIDS and the United Nations, enabling data cleaning, aggregation, and transformation into a suitable format for analysis. The mathematical models are applied to fit the historical data and project future trends. These models are selected to capture different potential trajectories of the epidemic, representing linear growth, slowing progression, or rapid escalation, respectively. The forecasting formulas for each model are outlined in **Table 1** below, providing a structured basis for extrapolation beyond the current data.

Model	Formula			
Linear	$y_i = \alpha + \beta x_i$	$\beta = \frac{n\sum(x_i y_i) - \sum x_i \sum y_i}{n\sum {x_i}^2 - (\sum x_i)^2}$	$\alpha = \frac{\sum y_i - \beta \sum x_i}{n}$	
Logarithmic	$y_i = \alpha + \beta \ln(x_i)$	$\beta = \frac{n\sum(y_i z_i) - \sum y_i \sum z_i}{n\sum {z_i}^2 - (\sum z_i)^2}$	$\alpha = \frac{\sum y_i - \beta \sum z_i}{n}$	$z_i = \ln(x_i)$
Sigmoid	$y_i = \frac{\alpha}{1 + e^{-\beta x_i}}$	$\beta = \frac{n\sum(x_i z_i) - \sum x_i \sum z_i}{n\sum {x_i}^2 - (\sum x_i)^2}$	$\alpha = \max_{0 \le y_i \le 1} y_i$	$z_i = \ln\left(\frac{y_i}{\alpha - y_i}\right)$

Table 1. Mathematical models

For the curve-fitting process, the NumPy [6] and SciPy [7] libraries are used, with the $curve\_fit$ function from SciPy enabling precise parameter estimation for each model. The fitted models are then used to forecast HIV-related metrics, such as new infections, treatment coverage, and viral suppression, up to the year 2050.

To assess the accuracy of the fitted models, the coefficient of determination R^2 is calculated for each model, providing a measure of how well it explains the variance in the historical data. R^2 values are computed using the $r2\_score$ function from the Scikit-learn [8] library, ensuring a standardized and reliable evaluation metric. The model with the highest R^2 is selected as the best fit for that specific dataset, indicating the most accurate representation of the observed trends and

the most reliable basis for forecasting.

This study used a random forest regression method to predict HIV growth rate, with features including LGBT score [9], sex education [10], urban population [11], funding [12], and three indicators of the 95-95-95 target [3]. Due to limitations in data sources, quantitative indicators of key variables such as social stigma and legal barriers were not included. Random forest was chosen because of its robustness to heterogeneous datasets and its ability to handle high-dimensional features. Random forest is an ensemble learning method that performs well in dealing with complex, nonlinear, multidimensional data by building multiple decision trees and aggregating their predictions. [13] This study set estimators to 1000 and random state to 0 to ensure the stability and repeatability of the model. Feature importance is calculated by averaging the feature contribution of each tree, providing the relative impact of each variable on the prediction, while the Pearson correlation coefficient is used to reveal the positive or negative relationship between the feature and the target variable.

Results

This analysis utilizes data from UNAIDS spanning 2013 to 2023 to fit the growth trend of the global HIV epidemic. Based on this projection, the first, second, and third targets of the 95-95-95 framework—95% of people living with HIV knowing their status, 95% of diagnosed individuals receiving treatment, and 95% of those treated achieving viral suppression are estimated to be achieved around 2033, 2030, and 2025, respectively. These trends are illustrated in **Figure 1** below.

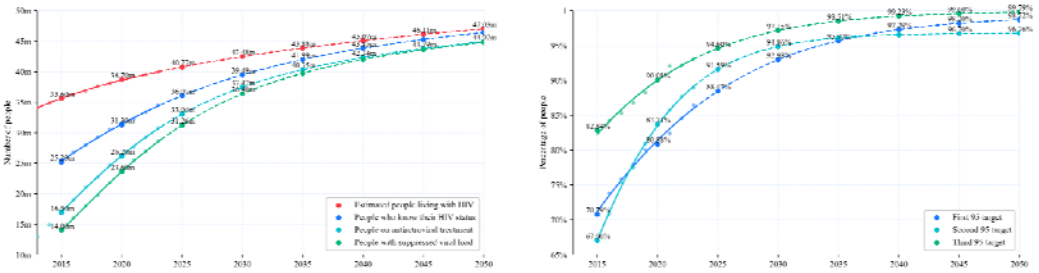


Figure 1. Projected timeline for achieving the UNAIDS 95-95-95 targets based on global HIV data from 2013 to 2023. The figure illustrates the estimated years of attainment for the first (95% awareness of HIV status), second (95% treatment coverage), and third (95% viral suppression) targets, projected to occur around 2033, 2030, and 2025, respectively.

However, when zooming in to specific regions, the conclusions diverge significantly. Eastern and southern Africa, the most populous region, accounted for 20.6 million people living with HIV in 2020, representing 53.2% of the global total of 38.7 million. The slowdown in this region has artificially elevated the global average, masking underlying disparities.

While Eastern and southern Africa, Western and central Africa, and Western and central Europe and North America exhibit a decelerating growth trend, other regions present a contrasting picture, as shown in the upper-left panel of **Figure 2**. Asia and the Pacific, depicted in the upper-right

panel, continue to experience rapid growth, rising from 5.9 million in 2015 to a projected 9.16 million by 2050. Similarly, the Caribbean and the Middle East and North Africa, shown in the lower-right panel, show steep increases, with the Caribbean projected to reach 497,450 cases by 2050 and the Middle East and North Africa nearing 400,550 cases. Eastern Europe and central Asia, as well as Latin America, also raise concerns, with their growth trajectories in the lower-left panel suggesting persistent challenges, reaching 3.38 million and 3.53 million respectively by 2050.

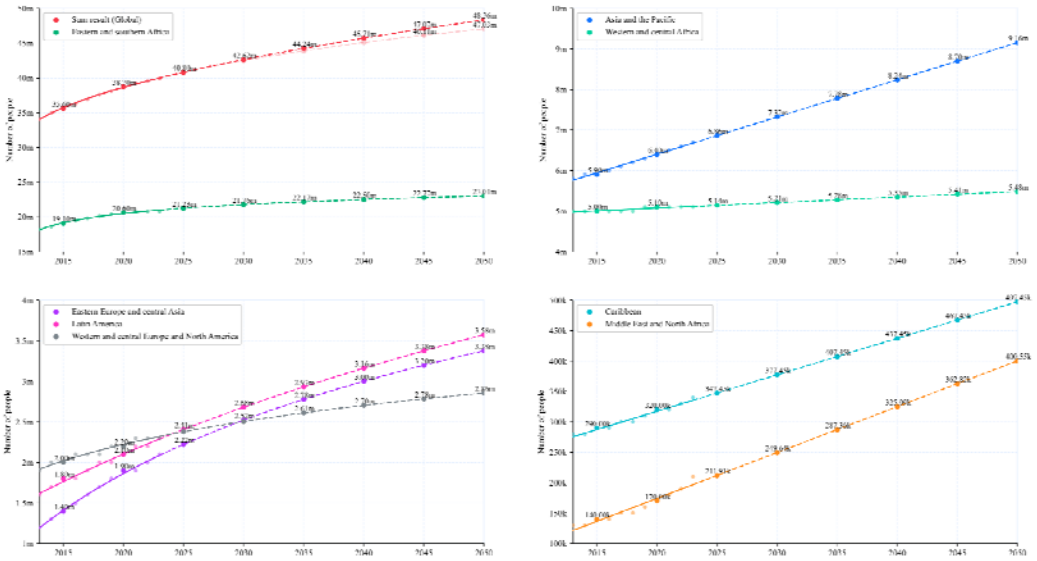


Figure 2. The figure illustrates the projected growth of HIV cases across eight distinct regions, including Eastern and southern Africa, Western and central Africa, Western and central Europe and North America, Eastern Europe and central Asia, Latin America, Asia and the Pacific, the Caribbean, and the Middle East and North Africa. The red curves represent the global data, with the dark red curve indicating the sum of individually fitted regional projections, while the light red curve shows the overall fitted global trend. The noticeable divergence between these two red curves highlights the disparities masked by averaging the regional data.

The red curve represents the global data, but unlike the previous curve fitting approach, this analysis calculates the actual global growth by summing the projections derived from fitting the data of eight distinct regions individually. A noticeable discrepancy emerges between the summed result (dark red) and the fitted global trend (light red), with the summed projection reaching 48.31 million by 2050 compared to the fitted global estimate of 47.03 million in 2050. This gap underscores how averaging obscures the true growth trajectory of the HIV epidemic.

This discrepancy arises because averaging smooths out regional variations, particularly the rapid growth in regions like Asia and the Pacific, the Caribbean, and the Middle East and North Africa, while underrepresenting their contribution to the global total. By aggregating data across regions with divergent epidemiological dynamics, such as the decelerating trends in Eastern and southern Africa versus the accelerating trends in Asia and the Pacific, averaging introduces a statistical bias that dilutes the impact of high-growth areas. This methodological flaw highlights the limitations of relying on global averages to capture the heterogeneous nature of the HIV pandemic, potentially

leading to over-optimistic conclusions about the trajectory of the epidemic.

For instance, according to UNAIDS statistics, the number of adults and children newly infected with HIV in Asia and the Pacific was 320,000 in 2013, yet by 2023, this figure remained alarmingly high at 300,000. [3] This marginal decline of only 20,000 over a decade highlights the persistent challenge of controlling new infections in the region, despite global efforts to curb the HIV epidemic. It underscores the limitations of current prevention strategies in addressing the unique social, cultural, and structural barriers in Asia and the Pacific, such as stigma, limited access to testing, and inadequate targeting of high-risk populations, indicating that the region remains a critical hotspot for the ongoing HIV crisis.

Due to incomplete data from countries around the world, only a dataset containing 81 countries was generated. The dataset contains a variety of socioeconomic and public health indicators, and random forest is able to capture the nonlinear interactions between these variables without assuming a specific functional form. In addition, the method is insensitive to noise and outliers, which is particularly important in cross-national HIV data because the data may vary due to statistical methods or reporting quality. However, as a "black box" model, random forest lacks transparency in its internal decision-making process. Although it can efficiently predict and quantify feature importance, it cannot directly provide clear coefficients or explanatory formulas similar to linear regression. These correlations are illustrated in **Figure 3** below.

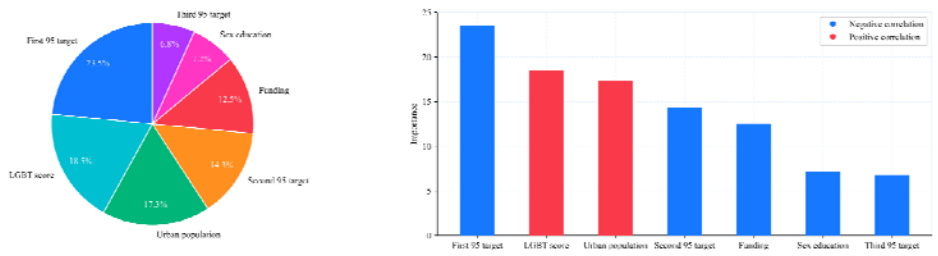


Figure 3. The chart shows a visual analysis of the importance of each feature in the prediction of HIV growth rate. The pie chart on the left uses different colors to represent the relative contribution of each feature, showing its proportion in the model. The bar chart on the right compares the importance of features by bar height, with red and blue marking the positive and negative relationship with the growth rate respectively. The legend distinguishes these two directional influences.

In predicting the HIV growth rate, the correlation between each feature and the growth rate reveals the direction and strength of its impact. The first 95 target (23.5%, negative) is highly correlated, while the second 95 target (14.3%, negative) is moderately correlated, indicating that high testing and treatment rates significantly reduce the growth rate. This aligns with the expectation that "informed and treated individuals reduce transmission."

The LGBT score (18.5%, positive) and urban population (17.3%, positive) are also highly correlated, suggesting that HIV growth is primarily driven by the LGBT community in highly urbanized areas. Funding (12.5%, negative) shows only a general correlation, indicating that its impact is not significant.

The third 95 target (6.8%, negative) has the lowest correlation, raising questions about its practical significance. This suggests that the population with viral suppression is not the main factor in virus transmission; rather, growth is driven by those who have not been tested. As a result, the third 95 target has limited predictive power and contributes minimally to the model.

Discussion

The question of whether the 95-95-95 targets truly represent a global endgame for the HIV epidemic remains contentious, particularly in light of the rapid growth observed in regions beyond Africa. While Eastern and southern Africa have demonstrated significant success, nearing these targets due to robust intervention programs, this progress does not equate to global triumph. The accelerating HIV incidence in regions such as Asia and the Pacific, the Caribbean, and the Middle East and North Africa underscores a stark disparity. Relying on the 95-95-95 framework as a marker of global success risks overlooking these regional crises, suggesting that the "endgame" may be an illusion unless tailored strategies address the diverse epidemiological landscapes beyond Africa's borders.

Additionally, HIV prevention and control strategies should focus on the LGBT community in highly urbanized areas, particularly those who have not been tested. The strong positive correlation between the LGBT score and urban population indicates that these areas have high population density and complex social networks, which may increase the risk of exposure and transmission within the LGBT community. Conversely, individuals who have received treatment, especially those whose virus is suppressed, require less attention. Given limited funding, resources should prioritize encouraging untested individuals to undergo testing.

Author contributions

Shaoming Chen wrote the manuscript.

Acknowledgments

None.

Conflict of interest statement

None.

Data availability statement

The datasets analyzed during the current study are available in the GitHub repository at https://github.com/rheast/epidemiology .The data that support the findings of this study are openly available in national HIV estimates files published through UNAIDS at https://www.unaids.org/en/resources/documents/2024/HIV_estimates_with_uncertainty_bounds_1990-present .

Ethics statement

No animals or humans were involved in this study.

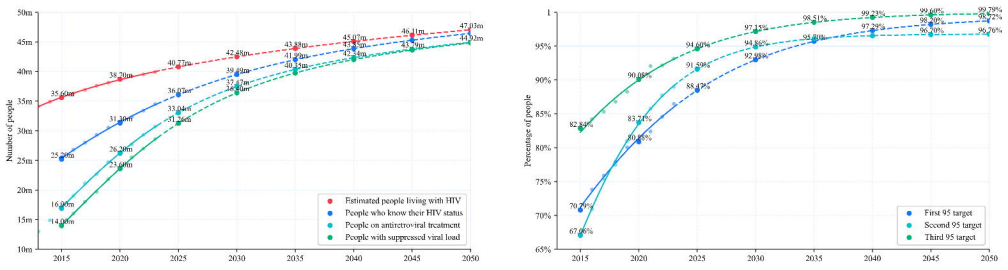
Authors and affiliations

Shaoming Chen. Rheast LLC, 1331 Lamar St, Houston, TX 77010, USA. research@rheast.com

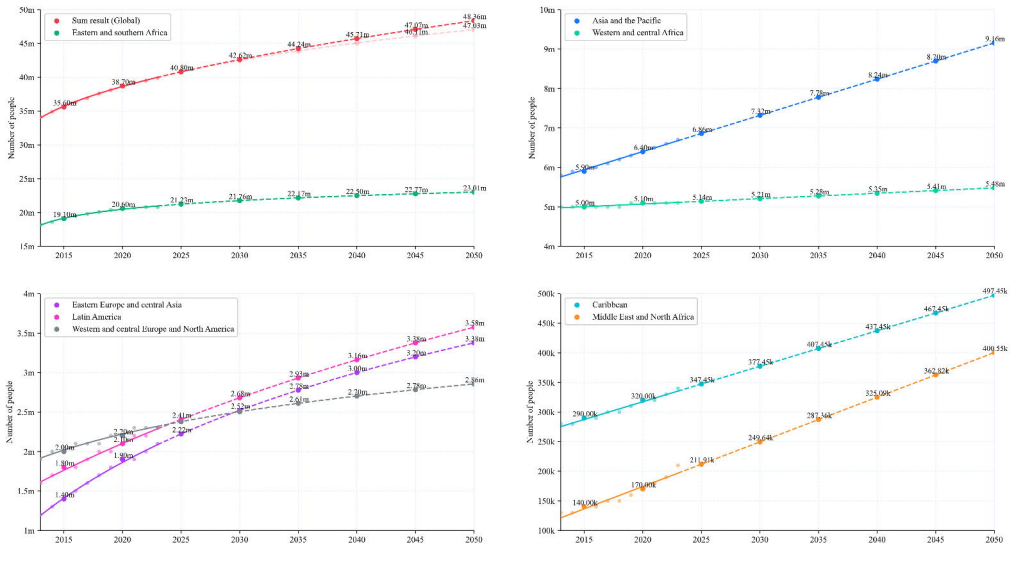
References

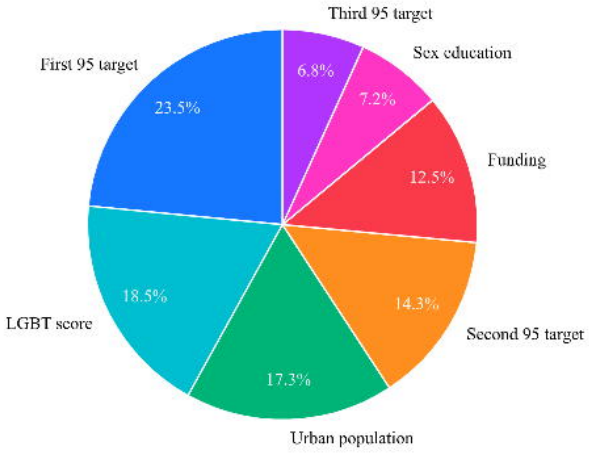
1. Khalifa A, Stover J, Mahy M, Idele P, Porth T, Lwamba C. Demographic change and HIV epidemic projections to 2050 for adolescents and young people aged 15-24. Glob Health Action. 2019;12(1):1662685. doi: 10.1080/16549716.2019.1662685.
2. Girum T, Wasie A, Worku A. Trend of HIV/AIDS for the last 26 years and predicting achievement of the 90-90-90 HIV prevention targets by 2020 in Ethiopia: a time series analysis. BMC Infect Dis. 2018 Jul 11;18(1):320. doi: 10.1186/s12879-018-3214-6.
3. UNAIDS. HIV estimates with uncertainty bounds, 1990–present. UNAIDS; 2024. https://www.unaids.org/en/resources/documents/2024/HIV_estimates_with_uncertainty_bounds_1990-present
4. Stover J, Glaubius R, Teng Y, Kelly S, Brown T, Hallett TB, Revill P, Bärnighausen T, Phillips AN, Fontaine C, Frescura L, Izazola-Licea JA, Semini I, Godfrey-Faussett P, De Lay PR, Benzaken AS, Ghys PD. Modeling the epidemiological impact of the UNAIDS 2025 targets to end AIDS as a public health threat by 2030. PLoS Med. 2021 Oct 18;18(10):e1003831. doi: 10.1371/journal.pmed.1003831.
5. McKinney, Wes. "pandas: a foundational Python library for data analysis and statistics." Python for high performance and scientific computing 14.9 (2011): 1-9. https://doi.org/10.25080/Majora-92bf1922-00a
6. Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. Nature 585, 357–362 (2020). https://doi.org/10.1038/s41586-020-2649-2
7. Virtanen, P., Gommers, R., Oliphant, T.E. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17, 261–272 (2020). https://doi.org/10.1038/s41592-019-0686-2
8. Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." the Journal of machine Learning research 12 (2011): 2825-2830. https://scikit-learn.org/
9. Equaldex. (n.d.). Equaldex: Collaborative LGBTQ+ rights knowledge base. Retrieved March, 2025, from https://www.equaldex.com
10. World Health Organization. (n.d.). Sexual and reproductive health data report [PDF]. Retrieved from https://cdn.who.int/media/docs/default-source/hrp/379607.pdf
11. World Bank. (n.d.). Urban population (% of total population) [Data set]. Retrieved from https://api.worldbank.org/v2/en/indicator/SP.URB.TOTL.IN.ZS
12. UNAIDS. (n.d.). GARPR16-GAM2024 programme expenditures [Excel file]. Retrieved from https://hivfinancial.unaids.org/GARPR16-GAM2024ProgrammeExpenditures.xlsx
13. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3),

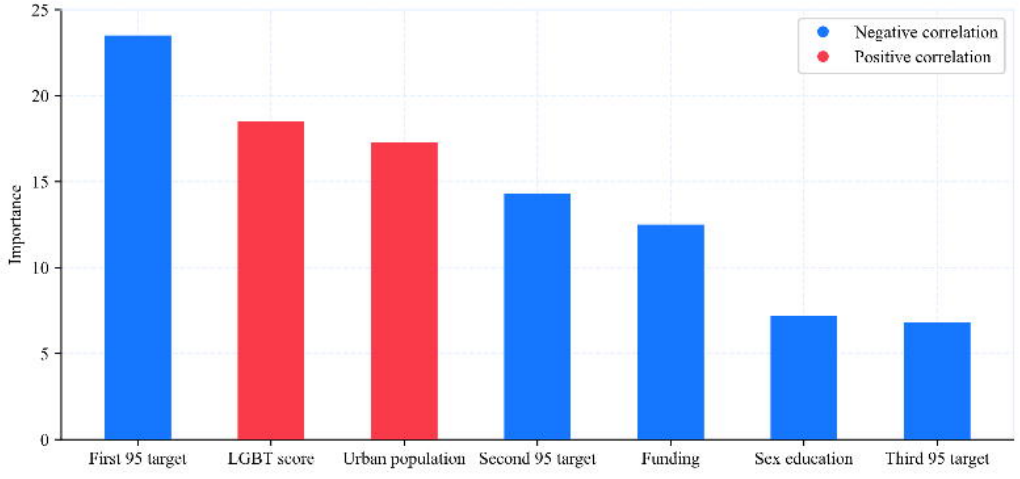
18-22. https://journal.r-project.org/articles/RN-2002-022/



Number of people
Estimated people living with HIV
People who know their HIV status
People on antiretroviral treatment
People with suppressed viral load
Percentage of people
First 95 target
Second 95 target
Third 95 target


Sum result (Global)
Eastern and southern Africa
Number of people
35.60m
38.70m
40.80m
42.62m
44.24m
45.71m
46.71m
47.07m
48.36m
47.03m
19.30m
20.60m
21.23m
21.76m
22.17m
22.50m
22.77m
23.01m
Asia and the Pacific
Western and central Africa
5.90m
6.40m
6.86m
7.32m
7.78m
8.24m
8.70m
9.16m
5.00m
5.10m
5.14m
5.21m
5.28m
5.35m
5.41m
5.48m
Eastern Europe and central Asia
Latin America
Western and central Europe and North America
2.00m
2.10m
2.22m
2.41m
2.52m
2.68m
2.78m
2.93m
3.00m
3.16m
3.20m
3.38m
3.58m
3.38m
1.40m
1.80m
2.61m
2.70m
2.76m
2.86m
Caribbean
Middle East and North Africa
290.00k
320.00k
347.45k
377.45k
407.45k
437.45k
467.45k
497.45k
140.00k
170.00k
211.91k
249.64k
287.36k
325.09k
362.82k
400.55k


Third 95 target
Sex education
First 95 target
6.8%
7.2%
23.5%
Funding
12.5%
18.5%
14.3%
LGBT score
Second 95 target
17.3%
Urban population


Negative correlation
Positive correlation
Importance
0
5
10
15
20
25
First 95 target
LGBT score
Urban population
Second 95 target
Funding
Sex education
Third 95 target