Technical updates for UNAIDS HIV estimation tools

Report and recommendations from a meeting of the UNAIDS Reference Group on Estimates, Modelling, and Projections

29 July – 1 August 2024
Addis Ababa, Ethiopia

Table of Contents

Abbreviations

AIM	AIDS Impact Model
ANC (-RT)	Antenatal Clinic (Routine Testing)
ART	Antiretroviral Therapy
ART-CC	ART Cohort Collaboration
ARM	Age and Risk Structured Model
ASM	Age Structured Model
Africa CDC	Africa Centre for Disease Control and Prevention
CSAVR	Case Surveillance and Vital Registration
DQA	Data Quality Assessment
EPP	Estimation and Projection Package
FSW	Female Sex Worker
GAM	Global AIDS Monitoring
IBBS	Integrated Biological and Behavioural Surveillance Survey
IeDEA	International Epidemiology Databases to Evaluate AIDS
KP	Key Population
(K)PSE	(Key) Population Size Estimate
LTFU	Loss to Follow-up
MENA	Middle East/North Africa
MSM	Men who have Sex with Men
PEPFAR	President's Emergency Plan for AIDS Relief
PHIA	Population-based HIV Impact Assessment
PLHIV	People Living with HIV
PrEP	Pre-Exposure Prophylaxis
PWID	People Who Inject Drugs
SSA	sub-Saharan Africa
TGW	Transgender Women
UNAIDS	Joint United Nations Programme on HIV/AIDS
VLS	Viral Load Suppression
WHO	World Health Organization

The meeting of the UNAIDS Reference Group on Estimates, Modelling, and Projections was organised for UNAIDS by the Secretariat of the Reference Group (www.epidem.org), managed at SACEMA, Imperial College London, and the University of Cape Town. Participants of the meeting are listed at the end of this document (**Appendix C**).

Background

UNAIDS Reference Group on Estimates, Modelling, and Projections

The Joint United Nations Programme on HIV/AIDS (UNAIDS) relies on impartial scientific advice from international experts in relevant subject areas to provide guidance on how to best calculate estimates and projections of the prevalence, incidence, and impact of HIV/AIDS globally. The UNAIDS Reference Group on Estimates, Modelling, and Projections acts as an 'open cohort' of epidemiologists, demographers, statisticians, and public health experts to provide scientific guidance to UNAIDS and partner organisations on the development and use of the tools used by countries to generate annual HIV estimates, which are the source for UNAIDS Global HIV epidemic estimates. The Group is coordinated by a Secretariat hosted at SACEMA, Harvard School of Public Health, and the University of Cape Town.

Meeting Overview

The UNAIDS Reference Group held a meeting in Addis Ababa, Ethiopia, from 29 July – 1 August 2024. The meeting featured presentations and plenary discussion to generate consensus recommendations, organised into the following 12 sessions:

1. Advanced HIV Disease
2. Causes of death among PLHIV
3. Mortality among PLHIV on ART
4. Excess mortality among PLHIV
5. Treatment interruption
6. World Population Prospects 2024 update
7. Dynamical modelling of HIV trends in Key Populations
8. New Infections by Key Population and partners
9. Sub-national HIV estimates
10. Sub-national EPP stratification
11. ART coverage data discrepancies
12. Age distribution of New Infections

This report contains a comprehensive summary of presentations and discussions from the meeting, forming the foundation for the Reference Group's recommendations. Meeting participants (Appendix C) can access these presentations on www.epidem.org. For those not in attendance, please reach out to the Secretariat at epidem@sun.ac.za for access requests. **Appendix A** presents the final recommendations. These recommendations guide UNAIDS in reviewing HIV estimates, assessing current methodologies, and identifying data for refining HIV estimates. The meeting agenda and objectives are in **Appendix D**. For records of previous meetings, visit www.epidem.org.

Meeting Introduction

Mary Mahy, **UNAIDS**, opened the meeting, sharing three key points to set the stage for the upcoming discussions:

1. Location choice: Addis Ababa

Chosen to foster engagement with the Africa Centres for Disease Control and Prevention (Africa CDC), raising awareness of the Reference Group's process of making HIV estimates as reliable as possible.

2. Surveillance systems: current and future state

With shrinking budgets and limited personnel, the focus is on improving models by addressing under- and overcounting issues, discussed in multiple sessions.

3. Simplification through integrated tools

Integrated tools aim to reduce data discrepancies, avoiding overly complex models for smaller teams and budgets.

Overview of Africa CDC and strategic engagement

James Guwani, Africa CDC, provided an overview of Africa CDC's structure and strategic initiatives. He emphasized integrating data science to inform policy and programmatic decision-making and capacity building within national public health institutes, aligned with UNAIDS' goals of improving HIV estimates and data reliability.

Guwani outlined efforts to develop an HIV program within Africa CDC, traditionally focused on pandemic and epidemic response, and addressed data quality challenges. He highlighted initiatives to leverage UNAIDS data to enhance the data consistency across African Union member states and discussed potential UNAIDS-Africa CDC collaborations on data collection, surveillance and systematic reporting.

Overview of 2024 estimates

Eline Korenromp, UNAIDS, presented key challenges, trends, and implications of the 2024 estimate.

Spectrum incidence models: 172 countries.

- **EPP Generalized Epidemic model:** 38 countries (~50% global HIV burden).
- **EPP Concentrated Epidemic model**: 38 countries with key population epidemics.
- **AEM model**: 13 Asian countries.
- **CSAVR:** 71 countries.
- **ECDC model:** 2 countries.
- **Bespoke model:** 10 countries.

Key 2024 Spectrum model updates:

1. **Treatment Interruption Rates**: Defaults set at 1.6% (high-income) and 5% (other countries). Countries provided data or used defaults.
2. **Post-2018 Mortality:** Flatlined using IeDEA cohort data, leading to increases in AIDS deaths, infections, and PLHIV.
3. **Post-2018 Mortality:** Flatlined using IeDEA cohort data, leading to increases in AIDS deaths, infections, and PLHIV.
4. **Guidance for countries**

    - **EPP/AIM incidence/prevalence adjustment**: Reinforced default maximum 10, to better align AIM with EPP
    - EPP-Concentrated: Recommend **R-Hybrid** as default/option, including for sub-populations with limited prevalence data.

Factors influencing 2024 estimates:

1. **Data Quality:** Persistent double counting of patients and unadjusted deaths. Non-universal ANC testing affected concentrated epidemic data quality.
2. **Model Adjustments:** ART number reallocations (e.g., Kenya) impacted estimates.
3. **Prevalence Data:** Adjustments in countries like Uganda improved plausibility.
4. **Epidemic Curve Changes:** Hybrid models in Azerbaijan and Colombia altered epidemic timelines.

Global and regional trends

- 24 countries lacked plausible estimates or engagement with UNAIDS.
- AIDS deaths rose 7% in 2022, while other global indicators showed minor changes.

Country-level

- Key population estimates remain sparse, critical for 2030 planning.
- Lower burden countries, like Comoros and Colombia, had significant estimate changes. Larger countries, such as Nigeria, showed noticeable absolute shifts.
- Uncertainty persists in low-burden countries, with stable trends in most African nations.
- Countries with significant discrepancies between mortality, incidence, and prevalence require further scrutiny and validation, like Tajikistan and Kazakhstan, where infections decreased but AIDS deaths increased.

Discussion

Following the presentation, several key inquiries and concerns were raised:

- **Decline in ART reporting countries**: Some countries couldn't validate final numbers for 2023, leading to non-publication. In other cases, the ART numbers exceeded the number of PLHIV, resulting in over 100% coverage, which couldn't be resolved and thus, wasn't published.
- **Breakdown by region**: No regional breakdown was shown. Western European countries lacked Dublin Declaration data for 2023, affecting completeness. The Dublin Declaration involves annual reporting of HIV estimates to UNAIDS by the European Centre for Disease Control (ECDC).
- **Country data diversity**: Estimates are published without detailing the data's reliability, suggestion to develop a framework to score the reliability of data from different countries to identify gaps and improve standardization.

Meeting overview

Cari van Schalkwyk, UNAIDS Reference Group Secretariat, emphasized themes aimed at improving HIV estimates and facilitating programmatic decisions:

- Enhancing Advanced HIV Disease (AHD) estimates and causes of death.
- Updating mortality trends and ART data quality.
- Differentiating between AIDS and non-AIDS deaths.
- Evaluating treatment interruption impacts and new World Population Prospects estimates.
- Updating models for key populations and sub-national estimates.

Session 1: Advanced HIV Disease (AHD)

The objectives of the first session were to:

- **Validate Spectrum estimates of AHD (defined as CD4 < 200 cells/mm³)**
- **Identify needs for future AHD estimates from modelled results and surveillance data sources**

Chair, Jeff Imai-Eaton, Imperial College, opened the first session by outlining five key points where recommendations from the Reference Group will be sought:

1. **Spectrum CD4 < 200 cells/mm³ as estimates of AHD:** Spectrum provides estimates for the number of people living with HIV with CD4 < 200 cells/mm³ and the number of people initiating treatment at this stage, but these outputs have not been a primary focus. The Reference Group's comfort level in using Spectrum CD4 < 200 cells/mm³ results as reliable estimates of AHD is sought.
2. **Using Spectrum AHD estimates as inputs for other estimates:** Considering using Spectrum AHD estimates as inputs for other morbidity or mortality estimates, like cryptococcal meningitis.
3. **Additional Spectrum outputs as part of AHD estimates:** Discussing potential outputs, such as numbers of hospitalizations and individuals with AHD entering care.
4. **Incorporating new data for AHD:** Identifying new data on AHD to be incorporated to improve future estimates.
5. **Capturing data inputs in Spectrum:** Determining which AHD-related data inputs and indicators would be valuable for the estimation process.

Nathan Ford, **WHO, Department of HIV, Hepatitis, and STI**, presented three key studies, focusing on the global and regional prevalence of AHD, causes of hospital admission among people living with HIV (PLHIV), and the current state and future directions of CD4 monitoring, summarised in Table 1.1.

Table 1.1: Summary of studies on AHD, causes of hospital admission among PLHIV, and current state of CD4 monitoring.

Study	Methodology	Key findings	Limitations
Study 1: Global and regional estimates of the prevalence of Advanced HIV Disease	Systematic review and meta-analysis (under review) Jan 2015 – Mar 2024, 117 studies, over 1.8 million individuals from 2 databases and IeDEA	**Meta-analysis** **AHD prevalence:** High heterogeneity. - **Inpatients:** 44% (range: 39 – 50%), heterogeneity, 25 – 65%. - **Outpatients:** 34% (range: 32 – 35%), heterogeneity, 8 – 65%. **Subgroup-analysis** -Higher prevalence observed in inpatients compared to outpatients. - Highest prevalence in WCA, SEARO, PAHO, and EMRO. - Similar prevalence across age, ART status, time since diagnosis, study date (pre- and post-2020), and income level.	- Search strategy may have missed relevant studies. - Non-random CD4 testing could introduce bias. - Subgroup analyses used aggregate data.
Study 2: Causes of hospital admission among PLHIV	Systematic review updating a 2016 review (110 studies, AHD in title, treat-all era)	**Causes of admission:** - 42% AIDS-related, including 19% TB. - 26% bacterial infections. - NCDs relatively uncommon. **Trends over time:** - ART at admission: rose from 43% (2007-2014) to 56%. - AIDS-related and bacterial infections: Slight decrease in admissions from 46% (2007-2014) to 42% (current review). - Mortality: decreased from 20% (2007-2014) to 17%.	Not discussed
Study 3: Current state of CD4 monitoring and future directions	Overview of CD4 monitoring challenges and issues with PIMA device distribution and functionality	- 60% sensitivity and 73% specificity. Emphasis on the importance of CD4 testing over staging alone. - Across IeDEA regions, CD4 testing availability: 2005: 58% - 86%; 2015: 48% - 87%. - Regional variation in 2019: • 61% - 86% in North America, Latin America, Asia-Pacific, and South Africa.	- Limited and uneven distribution of functional PIMA devices. Manufacturer estimates that 60% of non-functional devices can be serviced.

Study	Methodology	Key findings	Limitations
		• Rapid decline in CD4 testing availability where the burden of disease is highest: 13% - 53% in other sub-Saharan African countries. - Shift towards lateral flow assays (LFA) and increased funding requests for CD4 scale-up. - Enhanced procurement by PEPFAR and GFATM. - PEPFAR modifying its MER indicators for AHD, with forthcoming data. - CHAI actively supporting network optimization. - WHO to update guidelines on CD4 testing vs. staging to identify AHD by November 2024.	

Future research directions

- Continue monitoring trends in causes of hospital admission among PLHIV.
- Evaluate the long-term impact of increased ART use on hospital admissions and mortality rates.
- Support further studies on the effectiveness of POC CD4 testing and its integration into routine care.

Discussion

The summary below captures the key points discussed following the presentation.

1. **ART experience stratification in AHD:** Subgroup analyses showed a mix of treatment-naïve at diagnosis, newly diagnosed and treatment-experienced individuals at return to care. No significant differences were found in those with AHD and main outcomes in hospital. Specific differences in individual comorbidities could be investigated further.
2. **Decline in sub-Saharan Africa CD4 testing:** The cause is unclear, with differing views in the IeDEA executive committee. Some believe sicker patients are tested more, but it also depends on CD4 testing availability.
3. **Regional breakdown in hospitalisation review:** Ongoing systematic review of causes of death among hospitalised PLHIV show heterogeneity within and between regions. AIDS-related causes of hospital admission remain dominant in most regions (slides added to presentations listing Nathan Ford as speaker).
4. **WHO staging in studies:** Some studies reported WHO staging, but the focus was on CD4 cell counts, and WHO stage 3-4 disease was not reported.

Advanced HIV disease estimates from PHIA surveys

Shona Dalal, WHO, presented AHD prevalence estimates among PLHIV using Population-based HIV Impact Assessments (PHIA) survey data.

Method

The analysis included data from 13 most recent PHIA surveys, focusing on individuals with CD4 < 200 cells/mm³, defined as AHD due to the absence of WHO clinical staging in the surveys. The aim was to estimate the prevalence of AHD by demographic factors and the testing and treatment cascade, disaggregated by age, gender, socioeconomic status, rural vs. urban residence, and wealth quintiles.

Key findings

- 28,000 individuals with CD4 results were analysed.
- Pooled AHD prevalence: 9.8% of PLHIV have CD4 < 200 cells/mm³.
- Significant prevalence and number of people with AHD in sub-Saharan Africa, prevalence varied across countries but no significant differences by age, urban/rural residence, or wealth.
    - Total PLHIV: 24.2 million.
    - Total with AHD: 1.9 million (range: 1.59–2.22 million).
    - Viral suppression among AHD: 44% (range: 37–52%).
    - Men had higher AHD prevalence (13.2%) vs. women (8%).
        - Women: 0.93 million with AHD (range: 0.78–1.08 million), 39% virally suppressed (range: 33–46%).
        - Men: 0.98 million with AHD (range: 0.82–1.15 million), 49% virally suppressed (range: 41–58%).
- Household surveys may underestimate AHD prevalence, missing hospitalized patients.
- Two-thirds of people with AHD are on ART, and nearly half are virally suppressed; initiation and sustained treatment alone are insufficient to manage AHD.
- Timely CD4 testing and linkage to care are crucial for AHD diagnosis and treatment.
- Implementation of WHO AHD package is necessary for better patient outcomes and reduced mortality.

Discussion

Following Shona Dalal's presentation on estimates of AHD prevalence from PHIA surveys, several key inquiries and concerns were raised by the discussants:

- **CD4 distribution in Spectrum:** During discussion about comparing the prevalence of AHD among *all PLHIV* from PHIA surveys versus Spectrum, participants were reminded that the CD4 distribution of people on ART reflects CD4 count at treatment initiation rather than current CD4 counts.
- **Representativeness of PHIA survey data:** Concerns were expressed about the representativeness of PHIA survey data, which might not fully capture the burden of AHD, particularly among those not included in household surveys (e.g., hospitalized individuals).
- **High proportion of viral suppression among those with AHD:** This was a surprising finding but could be explained by rapid viral suppression following ART initiation, while CD4 recovery takes longer.
- **Gender differences:** A higher prevalence of AHD among men compared to women, was attributed to late diagnosis and lower ART coverage.

Comparing Spectrum-estimated AHD prevalence among PLHIV to empirical data

CD4 modelling in Spectrum

Spectrum captures CD4 counts at ART initiation but does not model subsequent CD4 recovery or declines in those who remain on treatment, whereas empirical data account for CD4 changes during treatment.

Oliver Stevens, **Imperial College**, compared Spectrum-estimated AHD prevalence against empirical data, including:

1. PHIA surveys (Population-based HIV Impact Assessments)
   - Nationally representative, cross-sectional household surveys providing HIV prevalence and AHD among PLHIV.
2. GAM reports (Global AIDS Monitoring)
   - Country program data on AHD at diagnoses or ART (re-)initiation including demographic and clinical information on PLHIV.
3. WHO systematic review
   - Comprehensive review of AHD prevalence studies using data from different settings and populations.
4. IeDEA Data (International Epidemiology Databases to Evaluate AIDS)
   - Clinical data on AHD prevalence at ART initiation across multiple regions.

Summary of methods

Regression models compared Spectrum model outputs for CD4 < 200 cells/mm³ to estimates from PHIA, GAM, WHO, and IeDEA data, analysing trends in AHD prevalence over time.

Results

AHD at ART initiation from IeDEA and Spectrum

Spectrum estimates in 2019 are **5-15%** higher than IeDEA data, with differences ranging from 4 to 16 percentage points (pp), depending on the African region (Figure 1.1).

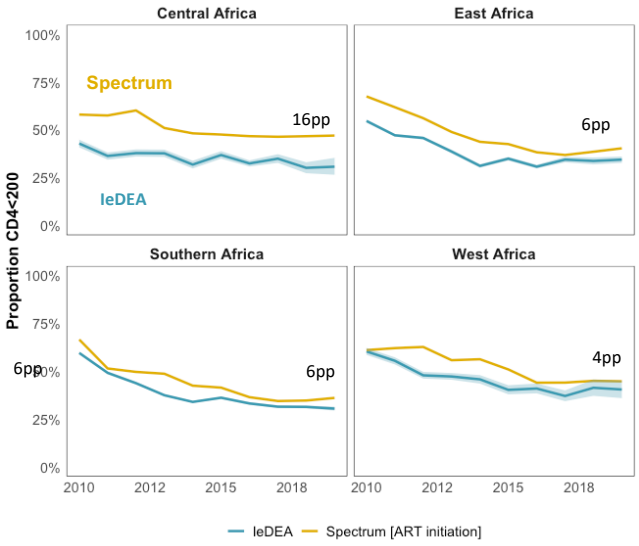


Figure 1.1: Comparison of AHD prevalence at ART initiation between Spectrum and IeDEA across four African regions in 2019.

1. Reference Group Technical Meeting on methods to estimate TB deaths among PLHIV: May 2024. Available at: https://epidem.org/wp-content/uploads/2024/10/Estimating-TB-deaths-among-PLHIV_Minutes.pdf

Clinic/surveillance studies vs Spectrum

A declining trend in the proportion of PLHIV with AHD (CD4 < 200 cells/mm³) across all regions, with varying degrees of alignment between different data sources and Spectrum (Figure 1.2).

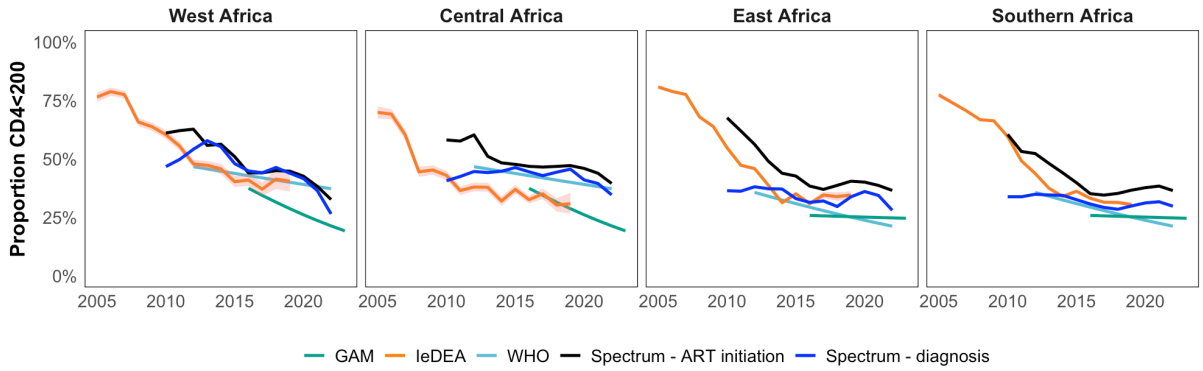


Figure 1.2: Proportion of PLHIV with AHD, CD4 < 200 cells/mm³ in four African regions (West Africa, Central Africa, East Africa, and Southern Africa) from 2005 to 2020.

AHD among all untreated PLHIV: PHIA vs Spectrum

AHD prevalence among untreated PLHIV ranged from 10-25% in PHIA, remaining stable between survey rounds, while Spectrum showed a wider range (10-40%). Where two rounds of PHIA surveys were available, some countries showed an increase between rounds (Figure 1.3).

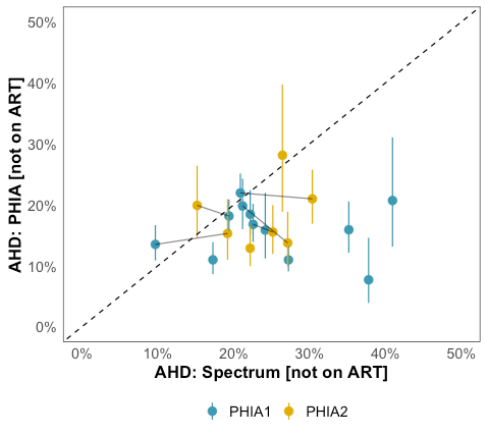


Figure 1.3: Comparison of AHD prevalence among untreated PLHIV between Spectrum estimates and PHIA survey data across multiple countries.

Summary

When comparing Spectrum estimates of AHD at diagnosis, the results aligned well in both level and trend with the GAM and WHO review data. However, Spectrum estimates at ART initiation were consistently higher than those observed in the IeDEA data. Furthermore, Spectrum estimates indicated higher levels of AHD among all PLHIV not on treatment compared to the levels reported by PHIA surveys.

Discussion

Following the presentation, several key points were discussed:

- **Data quality**: There were concerns about fluctuations in GAM data and changes in indicators. It was noted that even before the indicator change, the only countries with credible trends were Mozambique, South Africa, Nigeria, and Mali, with high variability year-to-year in other countries.
- **Hospital vs clinic**: All studies from the WHO review were pooled, but the comparative analysis could be influenced by separating hospital vs clinic-based studies.

- **ART initiation reporting**: It was highlighted that the IeDEA data field intended to capture the first ART start does not always accurately reflect this (due to people who have interrupted or transferred), leading to potential inaccuracies in the analysis. Exploring viral load suppression at initiation might impact AHD estimates.
- **Consistency of model outputs**: Outputs from different models (e.g., Shiny 90, AIM, and CSAVR) should be checked for consistency.
- **Treatment interruptions**: It was noted that Spectrum currently models the same rate of CD4 decline among treatment naïve people and people who have interrupted treatment, but data suggest a more rapid drop after treatment interruption.
- **Disaggregation of TX_NEW indicator**: CD4 disaggregation was added to the MER indicator TX_NEW in 2024. Of those newly initiating, 57% did not have CD4 measured, 12% had CD4 < 200 and 31% had CD4 > 200 cells/mm³.
- **Expand AHD definition:** There was support for potentially expanding the AHD category to include individuals with CD4 < 250 cells/mm³, to avoid missing individuals who may fall into WHO stages 3 or 4 of HIV disease progression.
- **CD4 after ART initiation**: Figure 1.4 shows a steep increase in AHD prevalence among those diagnosed but untreated in Spectrum (left) compared to the much more gradual increase in PHIA (right), raising concerns.

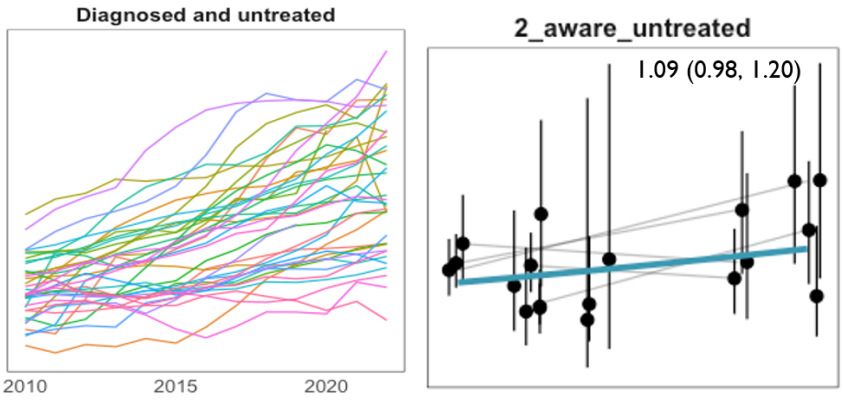


Figure 1.4: AHD prevalence among diagnosed but untreated individuals. Spectrum (left) showed a steep increase while PHIA survey data (right) showed a less pronounced increase.

This is likely because CD4 recovery is not modelled in Spectrum, and after interruption, individuals move to only one CD4 category higher than at treatment initiation, regardless of treatment duration. However, it should be investigated whether the increase in Spectrum is driven only by those who interrupt, or also by the treatment naïve. In addition, PHIA surveys are likely to underestimate the previously treated group, since this is a self-reported category, and individuals may not be willing to disclose stopped treatment.

Detailed recommendations are available in **Appendix A – recommendations**.

Session 2: Causes of death among PLHIV

Objective: Review implementation of CM and TB death estimates in Spectrum, within the envelope of Spectrum-estimated HIV-related deaths

Tuberculosis

Mathieu Bastard and **Nim Arinamanpathy** from the **WHO Global Tuberculosis Programme**, presented an overview of WHO's methods to estimate TB-HIV deaths, and showed estimates of WHO TB-HIV deaths as a proportion of UNAIDS' AIDS deaths by country and region.

In most countries, WHO estimates of TB deaths among PLHIV largely depend on two key factors: the estimates of TB incidence among PLHIV and the associated case fatality rates (CFRs) for TB/HIV coinfection. In countries that experienced severe disruptions due to COVID-19, estimates of TB/HIV deaths are derived using a transmission dynamic model. Refer to the minutes of the meeting held in May 2024[1] for a recap on methods for TB mortality estimation.

WHO TB-HIV deaths as a proportion of UNAIDS' AIDS deaths

Countries within the sub-Saharan Africa (SSA) region

As seen in Figure 2.1, Gabon, Central African Republic, Guinea Bissau, and Liberia report that over 75% of HIV-related deaths are due to TB. Countries like South Africa and Lesotho also show significant TB mortality.

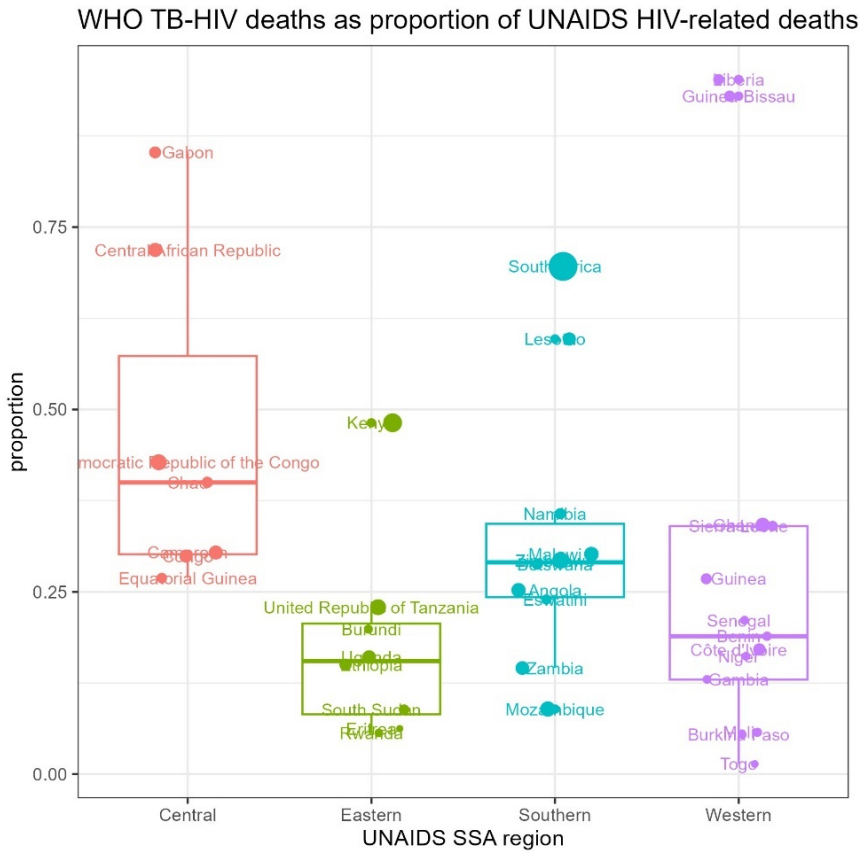


Figure 2.1: WHO TB-HIV deaths as proportion of UNAIDS HIV-related deaths in 2022, sub-Saharan Africa region.

Global perspective

1. Reference Group Technical Meeting on methods to estimate TB deaths among PLHIV: May 2024. Available at: https://epidem.org/wp-content/uploads/2024/10/Estimating-TB-deaths-among-PLHIV_Minutes.pdf

As shown in Figure 2.2, there is variation in TB-related mortality among PLHIV across different WHO regions with high TB-related HIV mortality in countries like Peru, Ecuador and Colombia in the Americas, and Cambodia in the Western Pacific Region.

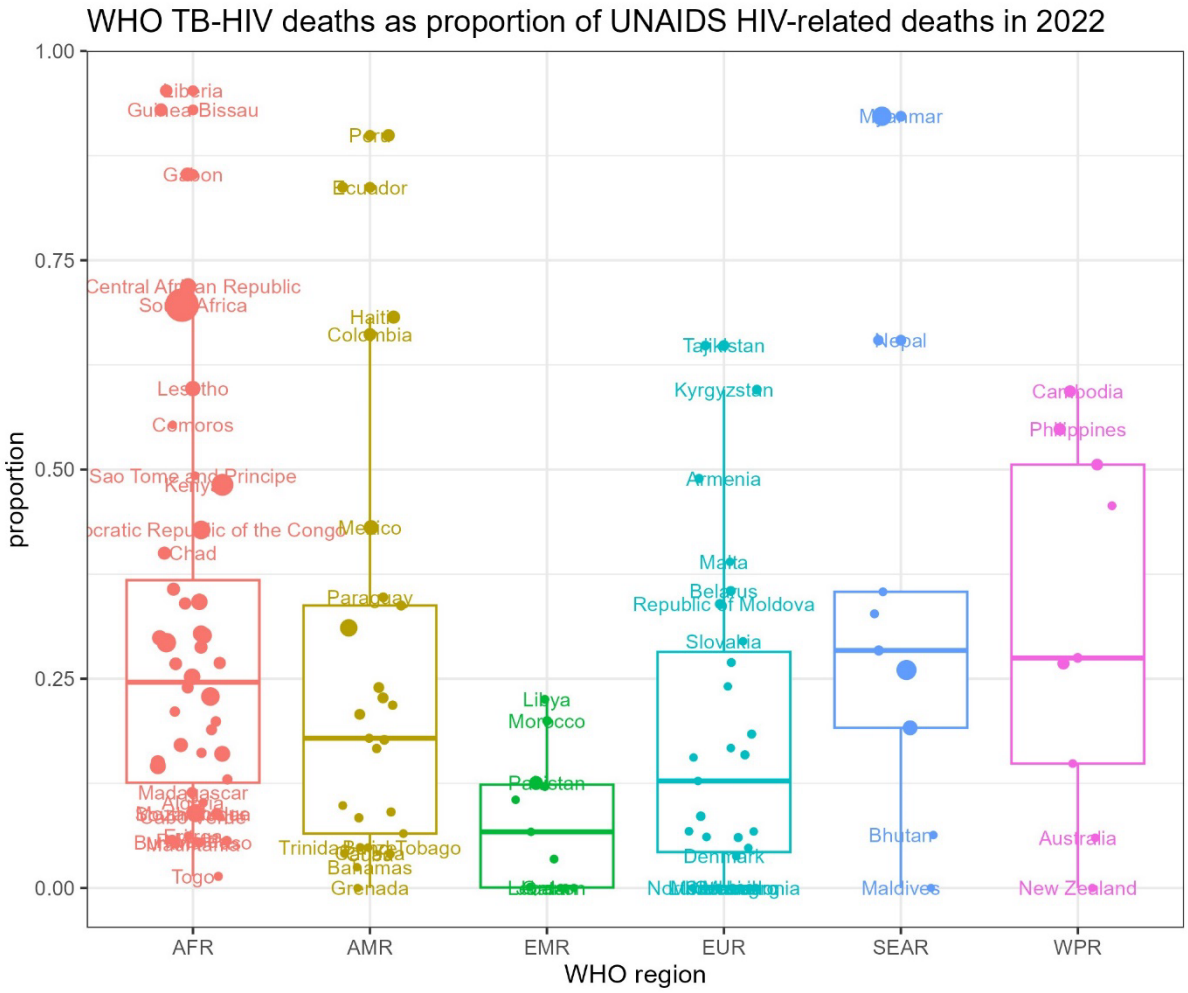


Figure 2.2: WHO TB-HIV deaths as proportion of UNAIDS HIV-related deaths in 2022, WHO regions.

Countries with prevalence surveys have more reliable incidence estimates, however case fatality ratios may require updating. The ongoing review of methods and the expected discussions at the September Task Force meeting are anticipated to lead to important updates in the TB estimation process.

A summary of the discussion key points and inquiries is listed below:

- **Stratification of ART coverage when applying case fatality rates (CFRs) to TB incidence in PLHIV:** Data collected from countries include HIV testing and HIV prevalence among TB cases, and ART coverage of those with TB-HIV. Case fatality rates depend on ART duration, stratified by less than one year or more than one year. However, obtaining accurate data on the duration of ART remains challenging.
- **Decentralising TB burden estimation to the national level to ensure local ownership of the estimates:** WHO does release initial estimates for country feedback, and there is a process for countries to provide input. Efforts are being made to enhance this process – like Spectrum's approach.
- **Case Fatality Rates (CFRs) are time-constant in increasing drug resistance context:** The CFRs currently used are constant, as derived from earlier literature. WHO is exploring the possibility of updating CFRs, potentially during the upcoming Task Force Meeting.
- **Use of Spectrum estimates for ART duration as an input to the TB model**: Spectrum estimates for ART duration as an input to TB model will be explored.

- **High proportion of TB-HIV deaths among HIV deaths in Peru and Colombia:** There are ways to potentially reduce the number of TB-HIV deaths: either reduce TB incidence or change the CFRs. The TB incidence estimates for Peru and Colombia were based on TB case notifications data, which leaves little room for adjusting TB incidence. Reducing the high proportion of HIV deaths due to TB would require a drastic reduction in TB case fatality rates, indicating a need for closer investigation into HIV-related deaths.
- **High proportion of TB-HIV deaths among HIV deaths in Cambodia and Myanmar:** In these countries, HIV is not a major driver of TB, accounting for less than 10% of TB incidence. The high proportions therefore require closer investigation.
- **Potentially unrealistically low proportions of TB-HIV deaths among HIV deaths** also require closer investigation.
- **High or low proportions of HIV deaths due to TB, might not have a universal explanation and could be country specific.**
- **TB-HIV incidence influencing HIV mortality:** It was noted that TB-HIV incidence could be factored into calculating country-level HIV mortality rates. WHO supported this idea, acknowledging the significant role TB plays as a cause of death among PLHIV.
- **Uncertainty in TB-HIV estimates:** Uncertainties in TB incidence among PLHIV and from CFRs are used to produce confidence intervals.

Recommendations

Including TB-related AIDS death proportions in Spectrum outputs

- **Option 1:** Include estimates for all countries in Spectrum for interpretation by users.
- **Option 2:** Include estimates for a selected set of countries in Spectrum, while withholding it from others that require more investigation.
- **Option 3:** Delay implementation in Spectrum until more thorough investigations are completed, especially considering upcoming methodological changes from the Task Force meeting.

There was agreement that it would be premature to include the proportions of TB-HIV deaths among HIV deaths across all countries in Spectrum without further investigation. WHO suggested waiting for input from the Task Force, given potential upcoming changes in both WHO methodologies and UNAIDS mortality estimates.

More detailed recommendations are captured in **Appendix A.**

Cryptococcal meningitis

Dr Radha Rajasingham, from the University of Minnesota (UMN), presented updates on methods used in the 2020 global burden of HIV-associated cryptococcal meningitis (CM) study using internal Spectrum estimates. Details of the 2020 study are in the February 2024 meeting minutes.[2]

UMN's 2020 methodology calculated the number of PLHIV in three categories, using estimates from aidsinfo.org:

- **Adults with HIV who do not know their status**: Adults with HIV * (1-%PWH who know their status)
- **Adults with known HIV, who are not on ART:** Adults with known HIV – Adults on ART
- **Adults on ART with virologic non-suppression**: Adults on ART * (1-% with suppressed viral loads)

These numbers would be similar with the 2024 methodology update, with numbers in each category directly drawn from internal Spectrum estimates.

The major change to the methods lies in how the number of adult PLHIV with advanced HIV disease (AHD) in each of the three categories are calculated:

Category	UMN 2020	2024 Update
Adults with AHD who do not know their status	% with CD4<200 not previous diagnosed (PHIA studies n=9) • Median = 16.5%	% with CD4<200 from Shiny 90
Adults with AHD who know their status but are not on ART	% "Late HIV diagnosis" from aidsinfo.org	% with CD4<200 from Shiny 90
Adults with AHD who are on ART, but are not virologically suppressed	% with CD4<200 on ART (PHIA studies n=5) • Median = 8.9%	Use PHIA estimates (including studies from 2020 and 2021) to approximate AHD prevalence of 30%

Key findings

Spectrum estimates of AHD prevalence are significantly lower in Botswana, Malawi, and Zimbabwe, resulting in lower estimates of the percentage of AIDS deaths attributable to CM (Table 2.2), when compared to the UMN model and updated AIDS Info data.

Lower CM death estimates are most prominent among individuals who:

- Known HIV but are not on ART.
- On ART with virologic non-suppression.

Table 2.2: Summary of CM estimates (% of AIDS deaths).

Country	UMN model 2020	AIDSInfo (2024 estimates)	Spectrum estimates 2023
Botswana	925 (21%)	660 (15%)	289 (7%)
Malawi	2976 (23%)	2719 (21%)	1312 (10%)
Zimbabwe	2994 (13%)	3163 (14%)	943 (4%)
Uganda	2669 (13%)	7062 (34%)	2264 (11%)

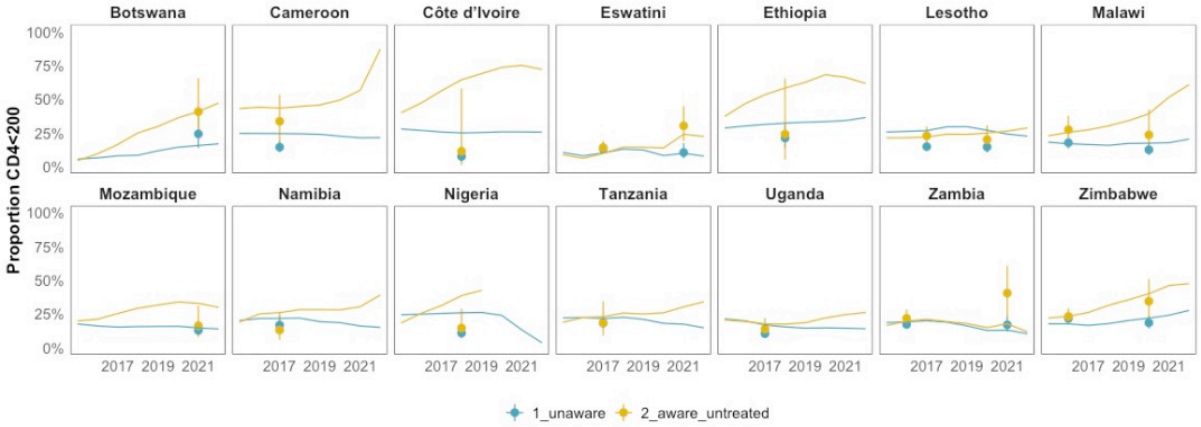
Discussion

Following the presentation, several key points were discussed:

- **Estimate discrepancies:** It was clarified that estimates in aidsinfo.org also come from Spectrum and should align if the same time point was used for most countries. However, in table 2.2, three different time points were at play: (2020 aidsinfo estimates, 2024 aidsinfo estimates, and 2023 Spectrum estimates).

2. Technical Meeting Estimating Cryptococcal Meningitis Deaths Among PLHIV – February 2024. Available at: https://epidem.org/wp-content/uploads/2024/10/UNAIDS-Technical-Meeting_Estimating-cryptococcal-meningitis-deaths-among-PLHIV-Minutes.pdf

- **Consistency of Spectrum and PHIA estimates of AHD prevalence**: Although not shown in Stevens' main results, Spectrum and PHIA estimates of AHD prevalence was consistent, across these countries:



- **Inclusion of virally suppressed individuals and risk for opportunistic infections:** The consensus was that while virally suppressed individuals with CD4 < 200 cells/mm³ should be included in estimates of AHD, they should be recognized as low risk for opportunistic infections, including CM. In the previous AHD session, it was shown that approximately 40% of people with CD4 < 200 cells/mm³ in the PHIA surveys are on ART and virally suppressed and are excluded from the CM model due to their low clinical relevance.
- **Complexity of cause of death attribution.** The challenge of double-counting deaths when multiple causes are involved was discussed. It was suggested that initially, CM be treated as a single category in Spectrum, with potential adjustments made in the future.

Recommendations

Cryptococcal meningitis will be incorporated in Spectrum as the first of several AIDS-related causes of death in Spectrum, with TB potentially being added in the future.

- **Default assumptions and outputs**
  - UMN's compiled global database will serve as the default value for prevalence, with countries having the option to replace these with updated local data. Uncertainty around progression and mortality rates will be used to generate uncertainty estimates within Spectrum.
  - The outputs will include key indicators such as the number of CrAg-positive cases and CM-related deaths.

- **Country-specific program data that users will be able to input**
  - % PLHIV screened for CrAg.
  - % PLHIV on fluconazole preventive therapy.
  - % PLHIV with CrAg who present to medical care.
  - For countries lacking program data, regional defaults or other interim measures were suggested as a solution.

- The consensus was to proceed with implementation in the next iteration of Spectrum, and validation can be done in selected countries such as Botswana and South Africa, where good CM data exists.

The detailed recommendations that followed this session are captured in **Appendix A.**

Session 3: Mortality among PLHIV on ART

Jeff Imai-Eaton prefaced the session by highlighting the significant limitations of available data to appropriately model mortality among PLHIV on ART. The most recent reliable data from the IeDEA consortium, which provides the mortality estimates used in Spectrum, is from 2017, prior to the COVID-19 pandemic and ART regimen changes.

He encouraged the Reference Group to consider:

- What data and contextual and anecdotal information can help evaluate HIV mortality estimates.
- How to ensure mortality estimates reflect local program outcomes, rather than relying solely on high-level modelling assumptions.

The session focused on the following objectives:

- **Validate Spectrum estimates of on-ART mortality against various sources**
- **Present new approach to Spectrum mortality structure given incomplete CD4 data in most IeDEA cohorts**

Trends in mortality among PLHIV in the Africa Health Research Institute Population Cohort, KwaZulu Natal

Mark Siedner and Avi Kenny presented on a study conducted within the AHRI population cohort in KwaZulu-Natal, South Africa. This long-term demographic health surveillance site (DHSS) collects data on births, deaths, and other health-related outcomes from annual home visits.

The key goals of the study were to:

- Estimate HIV-related outcomes: morbidity, hospitalization, and death.
- Assess how policy/regimen changes impact outcomes.
- Determine the effect of increased ART coverage on cause-specific mortality.
- Identify highest burden/priority conditions to improve care delivery.

Method

DHSSs often capture HIV status and deaths at regular intervals, leading to interval-censored data on HIV acquisition and ART initiation. In traditional HIV incidence models, if an individual is censored after their last negative test, they have zero future risk of acquiring HIV, often leading to biased or inaccurate mortality risk estimates. There are also no established methods for modelling time-varying exposure with interval censored data. To address this, a new model was developed to estimate the joint probability of HIV serostatus, mortality, and testing, that discretizes time into intervals (e.g., months, years). Serostatus, death, and testing data were integrated into the model, with parameters estimated via maximum likelihood.

Cohort characteristics	
Age range	13-60 years
Time period	2010-2022
Total participants	144,371
Women	76,941
Men	67,430
Total observation time	1,074,152 person-years
Median observation time/person	7 years
Total HIV seroconversions	12,827
Total deaths	6,630

Preliminary results

Mortality rates (per 1,000 person-years) trends

	2010	2022
Total Population with HIV	15.17	3.11
Total Population without HIV	1.95	1.66
Hazard ratio	7.85	1.88

Limitations

- **Model assumptions:** Assumes correct specification of hazard models for HIV seroconversion, mortality, and testing probability.
- **Covariate set:** Limited to sex, age, and calendar time.
- **HIV status:** Treated as a binary predictor rather than continuous (e.g., time since seroconversion).
- **ART status:** Not currently included in the model, reflecting a mixed population of people on and not on ART.

Next steps

Future work will incorporate ART status into the model, assess the differences in HIV- and ART-related mortality between pre- to post-dolutegravir roll-out and adapt the model to account for competing risks. Cause of death data (verbal autopsy and hospital-assessed) will be incorporated to assess impact of HIV on NCD and HIV-related causes of death.

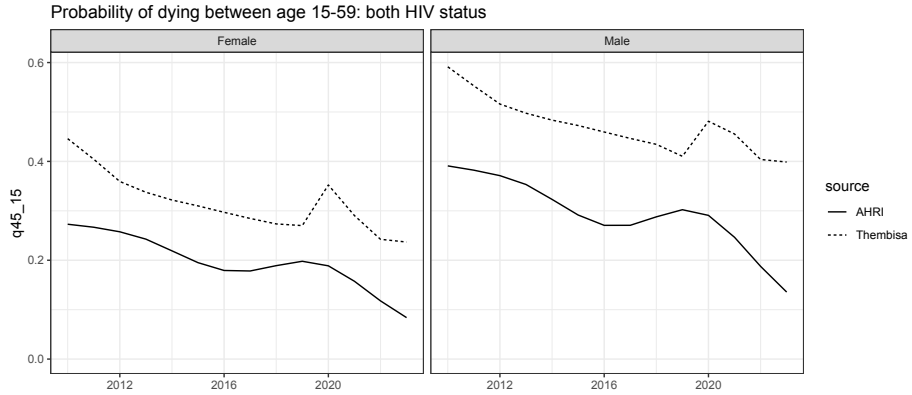
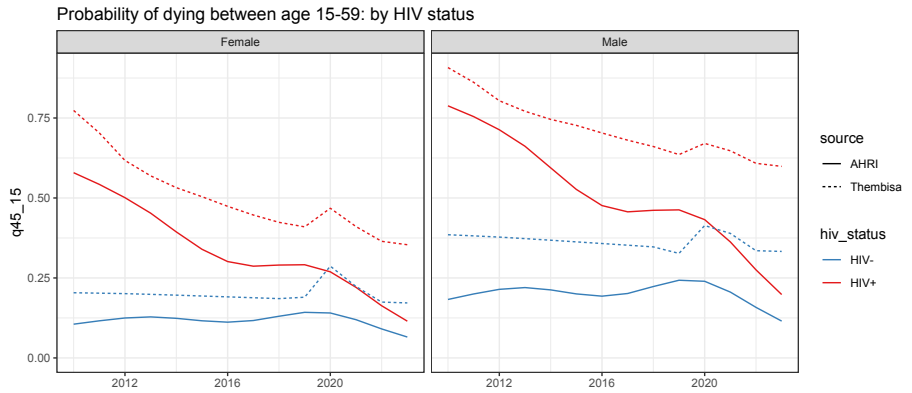
Discussion

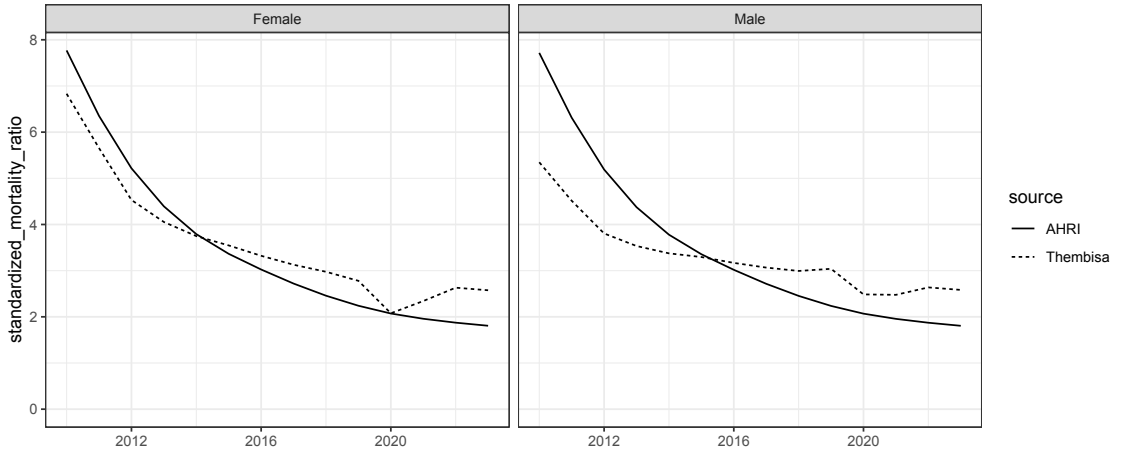
A summary of the key inquiries following the presentation, is listed below:

- **Linkage to the vital registration for ascertaining mortality:** The AHRI DHSS does not link to vital registration but collects data on deaths among residents of the catchment area through home-based surveys conducted three times a year and hospital records, thus providing unbiased death detection.
- **Flatline in mortality trends between 2016 and 2019:** The model's use of splines between time periods caused an apparent flatline, though mortality increased for both HIV-positive and HIV-negative individuals due to COVID-19.

AHRI mortality rates were benchmarked against Thembisa 4.7 for Kwa-Zulu Natal (Table 3.1).

Table 3.1. AHRI mortality rates, disaggregated by HIV status, were applied to the Thembisa 4.7 provincial population to calculate aggregate mortality statistics.

Metric	Observation
45q15 (Overall mortality) 	The overall probability of dying between ages 15 and 59 declined similarly in both models. However, AHRI estimates are somewhat lower compared to Thembisa.
Overall mortality by HIV status 	**Women with HIV:** Both models showed a relatively similar decline in mortality rates. **Men with HIV:** The AHRI estimates suggested a larger mortality decline compared to Thembisa. **Post-2020, for both**: A steeper decline in AHRI estimates than Thembisa estimates, flat mortality amongst adults on ART since 2020.

Metric	Observation
Marginal mortality ratio: ratio of mortality applied to standardized population structure [age/sex] • Applied 5-year age group mortality rates to Thembisa total population Standardized mortality ratio 	**Women:** AHRI ~2 vs. Thembisa ~2.5; AHRI shows a slightly steeper decline. **Men:** Similar range; slightly steeper decline in AHRI.

Mortality among people on ART from routine monitoring data through PEPFAR quarterly MER reporting

Sara Herbst, PEPFAR, presented a comparative analysis of mortality rates among people living with HIV (PLHIV) on antiretroviral therapy (ART) using Spectrum estimates and PEPFAR program data.

Primary objectives:

1. Compare Spectrum mortality estimates among PLHIV on ART with crude mortality from PEPFAR data.
2. Assess systematic variations in mortality by year, age, sex, and country.
3. Investigate potential reasons for any differences observed between Spectrum and PEPFAR mortality rates.

Data sources

PEPFAR program data

- Quarterly MER results from FY2020 through FY2024, analysing cumulative totals by fiscal year and disaggregating by age/sex groups (excluding paediatric data for this analysis).
- Numerator: TX_ML_died; Denominator: TX_CURR.

Spectrum

- Mortality estimates for calendar years 2020 to 2024, disaggregated by age/sex groups.
- Numerator: AIDS-deaths on ART + non-AIDS-deaths on ART; Denominator: Population on ART.

$$\text{Mortality rate (\%)} = \left(\frac{\text{\# deaths among treatment cohort}}{\text{\# PLHIV on treatment}} \right) \times 100$$

Key findings

PEPFAR reported deaths accounted for a median of one-third of deaths Spectrum estimates, ranging from one-sixth to 1.5 times the estimates. There is significant under-reporting of deaths in PEPFAR data, with inconsistent death ascertainment across countries. Discrepancies between PEPFAR and Spectrum is most pronounced among young adults, requiring further investigation (Figure 3.1). Additionally, PEPFAR data indicates a larger male-to-female mortality ratio in older age groups compared to Spectrum, seen consistently across most countries analysed.

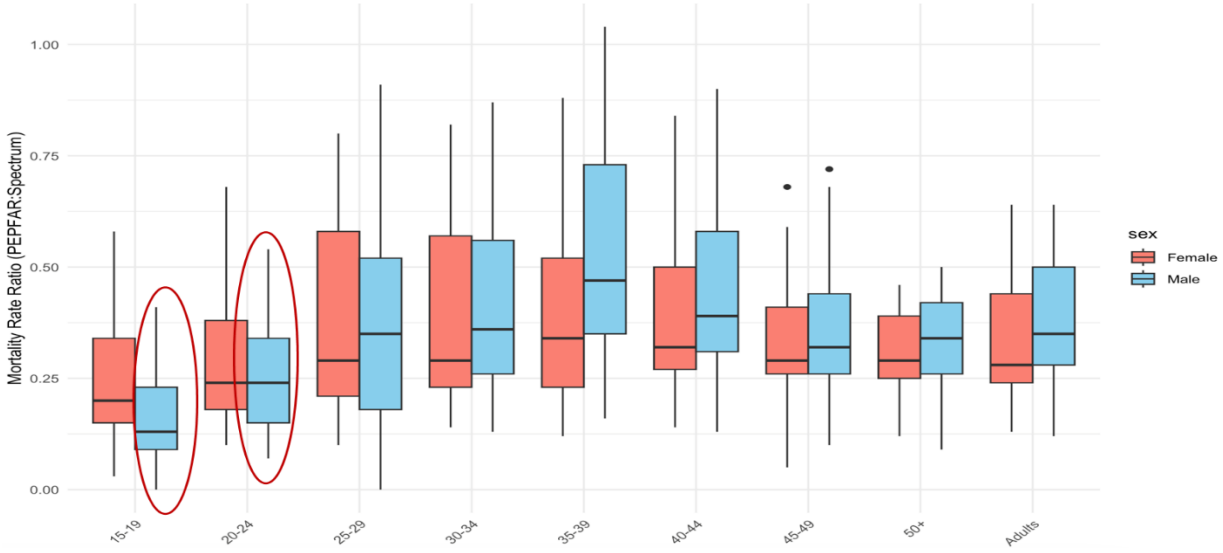


Figure 3.1: Mortality rate ratios among adult PLHIV on ART in 2023 between PEPFAR reporting and Spectrum estimates across multiple countries, by age and sex.

Proposed further analyses

- Correlating mortality rates with ART coverage across countries.
- Examining trends in quarterly program data.
- Further disaggregate mortality data by geography and healthcare facility.
- Reversing the mortality rate ratio comparison to compare Spectrum to PEPFAR, which may offer more interpretable insights.

Discussion

- **Enhancing PEPFAR mortality estimates:** The current mortality indicators used by PEPFAR were not originally designed to capture mortality accurately, since it cannot ascertain outcomes (mortality, transfers etc) among those lost to follow-up. Line list data from countries could be useful for identifying deaths among this group.
- **IeDEA data:** The mortality estimates in Spectrum are heavily informed by IeDEA data, which may not always represent the broader national population, as IeDEA sites are better resourced, potentially leading to biases.
- **Tracing studies** in IeDEA data, to adjust mortality among lost to follow-up, are seen as both a strength and a limitation. While they help correct mortality estimates, these studies are based on small samples and may introduce biases.
- **Underreporting of deaths:** In South Africa, only 35% of deaths in vital registration data were recorded in cohorts, a trend also observed in Ethiopia – indicating that many deaths are hidden within losses to follow-up, highlighting gaps in reporting.
- **Higher mortality among young people in Spectrum (via IeDEA data) than PEPFAR:** Spectrum shows higher mortality among young people. Spectrum's estimates are based on 2010-2017 data, while PEPFAR uses more recent 2023 data, likely reflecting improved survival due to better treatment. The IeDEA analysis also need to be revisited to determine whether the age and sex patterns result from the data or from the adjustments made using the tracing studies.
- **Duplications in TX_CURR:** Concerns were raised about duplications in TX_CURR (current on treatment), which likely lead to an overestimation of the denominator and reduced mortality rates, resulting in potential inaccuracies in the estimate.

New approach to Spectrum mortality structure with reduced CD4 data availability

Cari van Schalkwyk introduced the background to the ongoing discussion on mortality estimates, particularly focusing on the current limitations of mortality estimates from the IeDEA consortium (which informs Spectrum mortality), with the following key considerations:

- **Reduced CD4 data availability:** There is a decline in the proportion of people who initiate ART who receive a CD4 count, which could introduce potential biases if there are biases in who receives a CD4 test. This affects the accuracy of mortality estimates which are stratified by baseline CD4 counts.
- **ART duration uncertainties**: The uncertainties related to ART duration, including silent transfers and patients returning to care after interruptions, complicate mortality tracking.
- **Representativeness**: IeDEA sites may not be representative of all facilities in the broader regions to which mortality estimates are extrapolated. These sites are typically urban sites, and some are linked to tertiary hospitals, so they may be slightly better resourced than the average.
- **Recency of estimates**: Since the last reliable data are from 2019, the impact of COVID-19 and the introduction of DTG on ART outcomes and mortality rates are not integrated in the mortality estimates.
- **Tracing studies**: IeDEA mortality estimates are adjusted for mortality among those lost to follow-up using data from tracing studies – roughly half of deaths are imputed in this way. However, this imputation is based on few tracing studies, from few countries, and no recent tracing studies have been incorporated.

Participants were divided into Working Groups to discuss and report back on the following two questions related to the use of IeDEA data and the need for a new approach to mortality estimates in Spectrum:

- *Can we use IeDEA data differently? (such as not disaggregate mortality by CD4)*

- *How can we supplement/replace with other data?*

Several suggestions were made, as summarised below:
Can we use IeDEA data differently? (such as not disaggregate mortality by CD4)

- **Calibration using IeDEA data:** Use IeDEA as a calibration target by aggregating baseline CD4 counts to validate Spectrum outputs instead of using it as a direct input.
- **Comparative analysis:** Compare unadjusted IeDEA data with PEPFAR MER data (TX_CURR, deaths, LTFU) from the same sites to identify discrepancies and validate findings.

How can we supplement/replace with other data?

- **Find alternatives to CD4:** Explore using viral load suppression (VLS) or other markers instead of CD4 counts in regions where CD4 testing coverage is low. Consider a hybrid approach that uses CD4 data from the past and other markers moving forward.
- **Country tracing studies:** Use tracing studies data from countries like Ethiopia to supplement IeDEA analysis and enhance mortality estimates.
- **Review tracing studies:** Conduct a comprehensive review of existing tracing studies, including new studies that feature additional indicators such as age and sex strata, to ensure that all relevant data is included.
- **EMR person-level outcomes:** Leverage EMR data on person-level ART outcomes from countries such as Kenya, Zimbabwe, Rwanda, and Botswana to provide additional data sources for mortality analysis.
- **Mortality from DHSS sites:** Incorporate mortality data from demographic and health surveillance sites to provide a broader base for comparison.
- **Compare IeDEA with countries that have good Vital Registration (VR):** Analyse and compare IeDEA mortality data with mortality data from countries that have strong VR systems, like Botswana, to validate and potentially improve estimates.
- **Community-led monitoring:** Investigate how community-led monitoring can inform retention rates and treatment interruptions, and how these insights can be used to refine mortality estimates.
- **Cross-validate mortality patterns:** Cross-validate patterns of mortality by age with viral load suppression data from surveys, such as PHIA and household surveys, to improve accuracy and consistency in mortality estimates.

The full feedback from the Working Groups is captured in **Appendix B**

Session 4: Excess mortality among PLHIV

John Stover chaired the session, with the objective of **reviewing the method to disaggregate AIDS vs non-AIDS deaths**.

Excess mortality: AIDS vs non-AIDS

Rob Glaubius introduced a proposed methodology for disaggregating AIDS-related and non-AIDS-related mortality in Spectrum.

Background

In countries with high quality case-reporting and vital registration data, CSAVR is calibrated so that modelled deaths among PLHIV minus age-sex matched background mortality, or 'excess deaths', match AIDS-attributed deaths reported through vital registration. However, if this definition of model outputs representing 'excess mortality' among PLHIV are different from the definition of AIDS-related deaths recorded in vital registration, models may underestimate total HIV-related deaths and as a result underestimate total epidemic size.

Adam Trickey performed a systematic review[3], presented to the Reference Group at the October 2023 meeting[4], to estimate the proportion of excess mortality among PLHIV that are due to AIDS. This systematic review found 8 studies, all performed in high-income countries. The review included a meta-regression analysis, which estimates that at 50% ART coverage, AIDS accounts for 60% of excess deaths, and at 72% ART coverage, AIDS accounts for 55% of excess deaths.

Method

The proposed method involves deriving separate rates for AIDS and non-AIDS excess mortality that adds up to the current total excess mortality rates used in Spectrum and is consistent with the percentages estimated in the Trickey meta-regression. In the models, total excess mortality varies by CD4, age, sex, and duration on ART. AIDS mortality likely also varies by all the variables, but we do not have sufficient data to confirm this. The proposed method focuses on non-AIDS excess mortality, employing a linear model with parameters for age, CD4 count, and treatment duration. Spectrum outputs from 31 out of 36 files for the Western, Central Europe, and North America (WCENA) region from the 2023 estimates round were used to calculate the percentage of excess deaths due to AIDS for candidate sets of non-AIDS excess mortality rate inputs, using Bayesian methods to sample candidate sets.

Using France as an example, Glaubius showed good alignment with the 55% AIDS excess target, and higher non-AIDS rates at older ages and at lower CD4 counts, and that the coefficients for sex and ART duration were near zero with confidence spanning zero. The model was then refitted without sex and ART duration in each of the 31 countries, and the proposed default non-AIDS excess mortality rates are the median of best-fitting parameter estimates across countries.

Participants raised several important points during the presentation:

- **CD4 count and AIDS mortality**: A participant questioned the counterintuitive finding that the proportion of AIDS-related mortality decreases with increasing CD4 counts during the first 12 months on ART. Glaubius acknowledged this as an artifact of the model's simplifications and suggested that refining the model could address these issues if deemed necessary.
- **Regional applicability**: Concerns were expressed about applying data from high-income countries to model epidemics in high-burden, low-income regions. Glaubius acknowledged the lack of data from low-income regions, leaving the group with the options of using the same rates everywhere or assuming no non-AIDS excess mortality outside of high-income countries.
- **Non-AIDS mortality in Key Populations**: The importance of considering non-AIDS mortality specific to key populations, such as men who have sex with men and transgender women, was highlighted. Adam Trickey added that non-AIDS excess mortality includes factors beyond lifestyle, such as higher rates of cancer and cardiovascular disease among PLHIV, even with viral suppression.

3. **Trickey A, Ambia J, Glaubius R, van Schalkwyk C, Imai-Eaton JW, Korenromp EL, Johnson LF.** Excess mortality attributable to AIDS among people living with HIV in high-income countries: a systematic review and meta-analysis. *J Int AIDS Soc.* 2024 Nov;27(11). doi: 10.1002/jia2.26384.
4. Technical Updates Reference Group Meeting Report: October 2023. Available at: https://epidem.org/wp-content/uploads/2024/07/October-2023-Reference-Group-meeting-Report.pdf

Effects on Spectrum estimates

Three scenarios were run to assess the impact of implementing the disaggregated mortality rates across 173 countries: (1) the status quo where all excess deaths are classified as AIDS deaths, (2) a scenario where only excess deaths on treatment are disaggregated, and (3) a scenario where excess deaths both on and off treatment are disaggregated.

The results indicated that while disaggregation significantly reduces AIDS death estimates in high-income countries, its impact in other regions is less pronounced due to higher overall mortality rates.

France

Disaggregation had a more significant impact post-1997, with non-AIDS excess mortality gradually increasing from that point onward. Overall estimates of excess mortality were not significantly altered, except in the early 2000s due to differences in how mortality time trends were applied.

High-income vs. low-income countries

The model aligned well with expectations when applied to high-income countries. However, when applied to low-income countries, where excess mortality rates are much higher, the disaggregation had minimal impact, with over 90% of excess deaths still attributed to AIDS.

Glaubius concluded by outlining the decisions that need to be made:

1. The group needs to decide whether the proposed disaggregation methodology should be implemented using the current rates or an elaboration of those rates.
2. The decision needs to be made on whether to apply the disaggregation only to people on treatment or to all PLHIV, and whether this should be implemented only in high-income countries or globally.

Conclusion

While disaggregation could reduce estimates of AIDS deaths by 3% to 29%, depending on the region, the most significant impact would be seen in regions like WCENA, where CSAVR is heavily used.

Impact of proposed excess mortality disaggregation on CSAVR estimates

Guy Mahiane, Avenir Health Inc, presented an assessment of the proposed disaggregation of excess mortality people living with HIV (PLHIV) into AIDS-related and non-AIDS-related deaths on CSAVR estimates.

Method

A version of CSAVR that allows excess non-AIDS deaths for PLHIV was implemented, using data from 73 countries from UNAIDS regions, including Asia and Pacific (AP), Eastern Europe and Central Asia (EECA), Latin America and Caribbean (LAC), Middle East and North Africa (MENA), and Western and Central Europe and North America (WCENA).

The same non-AIDS excess mortality rates were applied for PLHIV both on and off ART and the model was fitted to data using the incidence option selected by the country:

- **Status quo**: No disaggregation of excess deaths.
- **2.A**: Disaggregation of excess deaths among PLHIV on ART into AIDS-related and non-AIDS-related deaths.
- **2.B**: Disaggregation of excess deaths among all PLHIV (regardless of ART status) into AIDS-related and non-AIDS-related deaths.

Key findings

Trajectories of key indicators for individual countries were compared and aggregated by UNAIDS regions, as summarised in Table 4.1.

Limitations

Model selection was not performed, meaning results could differ if a different model, like one using the AIC, had been chosen. Additionally, parameters derived from high-income countries were applied broadly to other regions, including Latin America, without region-specific adjustments, potentially affecting accuracy.

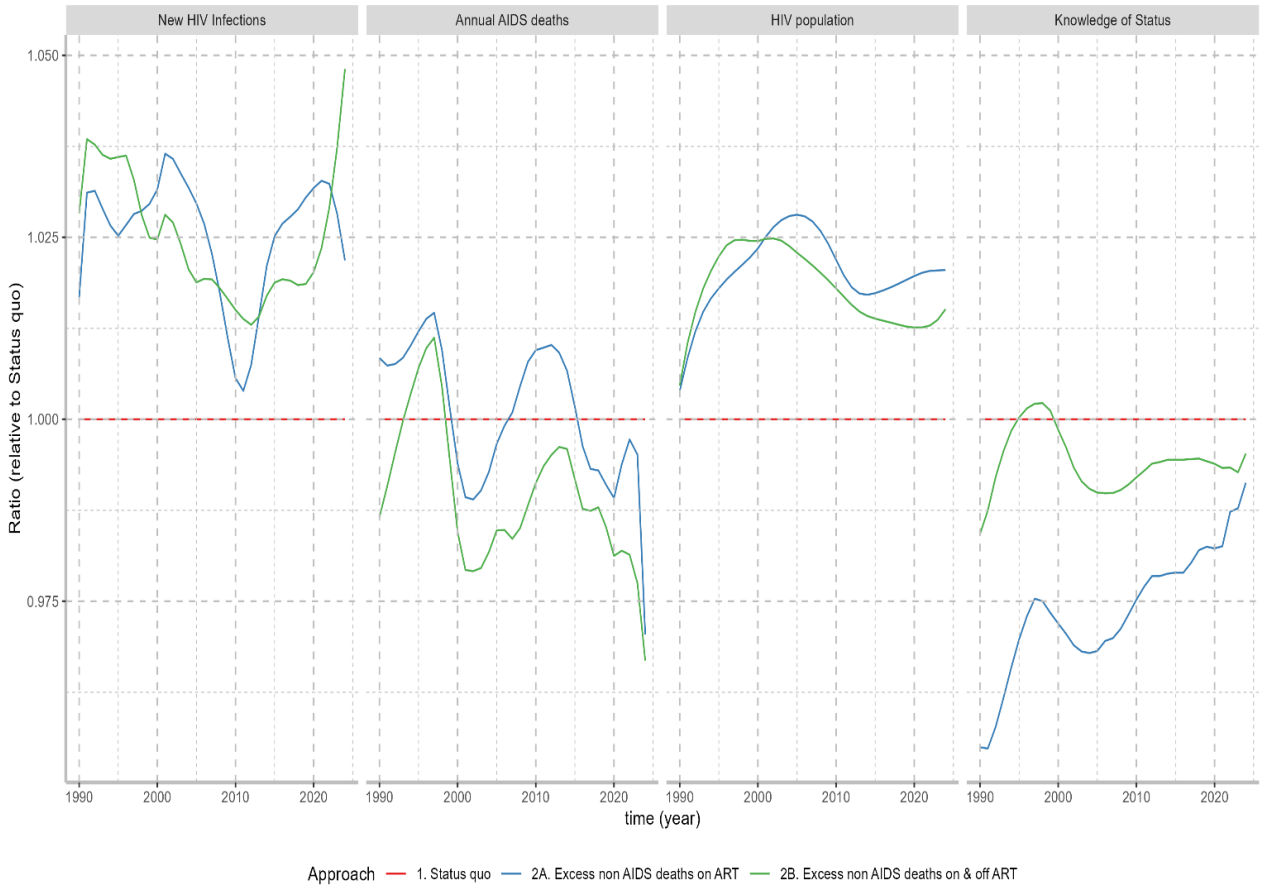
Conclusion

Disaggregating excess deaths among PLHIV into excess AIDS-deaths and excess non-AIDS deaths using the proposed approaches would result in

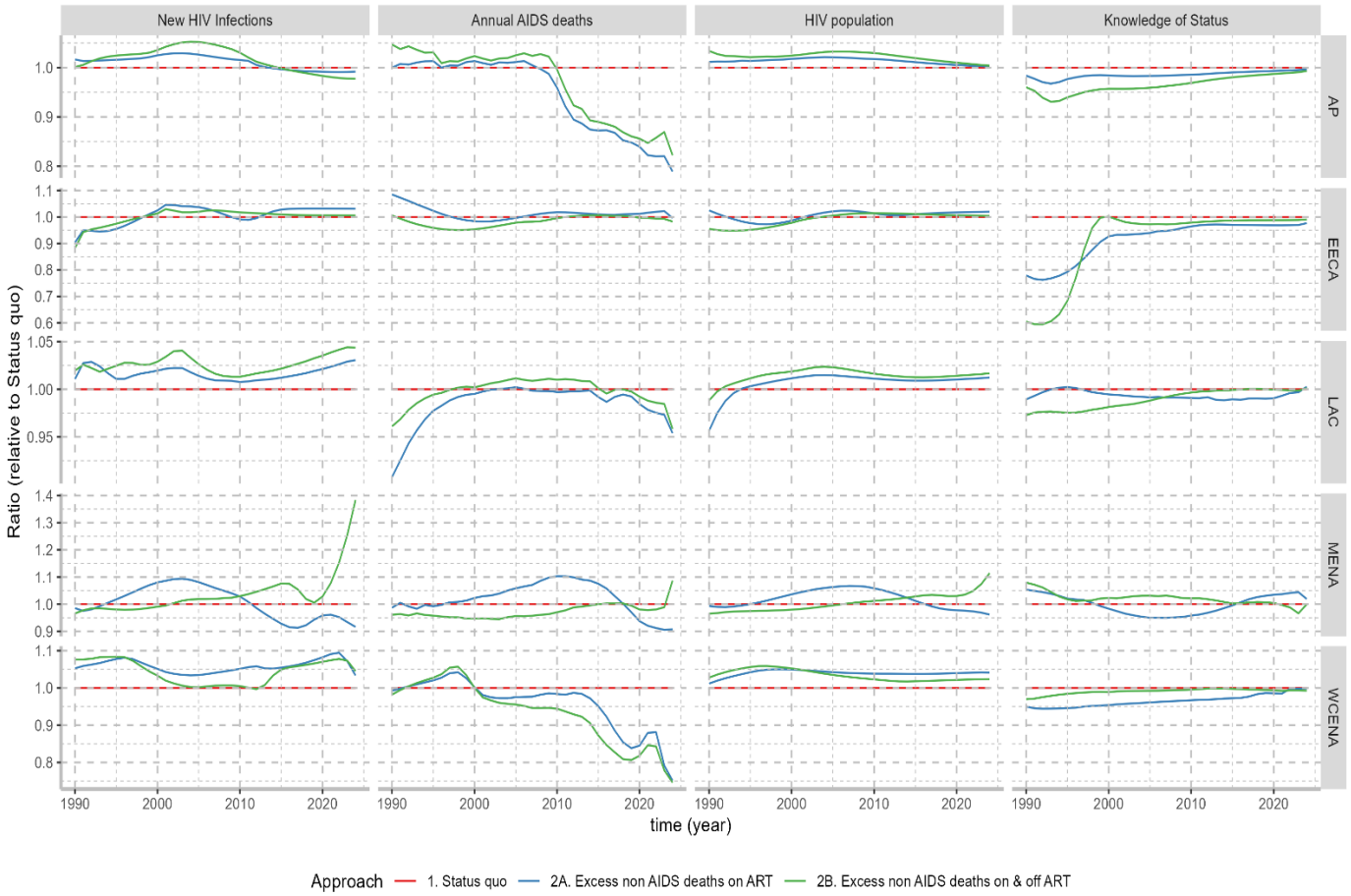
- small increase in new HIV infections and HIV populations (resp. <3% and <5%) in countries using CSAVR
- fewer AIDS deaths in recent years (after 2020, >10% in some countries)
- slightly lower estimates of KOS

The impact would vary across countries: relative difference will be influenced by the size of the epidemic, ART coverage and span of the data (AIDS-related deaths and new diagnoses time series).

Table 4.1. Regional and country-specific analyses for CSAVR countries.

Category	**Main findings per key indicator** (Sums: New HIV infections, Annual AIDS deaths and HIV populations; KOS is the ratio of #PLHIV knowing their status to the #PLHIV).
Aggregated results (All CSAVR countries) 	New HIV infections: Small increase (<5%) observed under 2.A and 2.B compared to the status quo. AIDS-related deaths: Fewer AIDS-related deaths under 2.B, with the maximum difference in 2023 (less than 5%). HIV population: Slight increase (1%-4%) in the number of PLHIV observed under 2.A and 2. B. Knowledge of Status (KOS): Lowest KOS estimates under 2.A, with differences shrinking over time.

UNAIDS regional variations



New HIV infections: Difference <5% from 2000-2015 in most regions; >10% for some years in MENA and WCENA after 2020.
AIDS-related deaths: Differences exceeded 15% in AP and WCENA regions.
HIV population: Largest differences (>5%) observed in MENA region.
KOS: Lowest estimates under 2.A or 2.B in most years; convergence in 2023 across regions.
France: Largest differences for AIDS-related deaths (>15%).
Mexico: Largest differences for 2.A observed before 2000, affecting deaths and PLHIV numbers.
Slovakia: Largest differences for AIDS-related deaths before 1995 and after 2015.

Discussion

1. **General support for high-income countries**

The proposed changes generally lead to small effects in the right direction, addressing some of the calibration issues observed in CSAVR estimates, particularly in high-income and some middle-income countries.

2. **Concerns about applicability in low-income regions**

Participants expressed concerns about the applicability of the methodology in sub-Saharan Africa, where the context differs significantly from high-income countries. They questioned whether the same assumptions could be applied given the differences in healthcare infrastructure, epidemiology, and social determinants of HIV.

3. **Concerns about explaining reduction to local stakeholders**

Concern: How the methodology might reduce reported AIDS deaths by 8% in countries like Kenya and Ethiopia, and the implications of explaining this reduction to local stakeholders. Any change would require clear communication, particularly in training workshops where these adjustments would need to be explained.

4. **Potential reversal of the proposed disaggregation methodology**

It was discussed that while it is possible to reverse the decision, doing so could complicate the explanation to stakeholders. The group stressed the importance of making a well-considered decision based on the best available data.

5. **Regional implementation considerations**

There was a suggestion to apply the disaggregation by region rather than income group, possibly excluding sub-Saharan Africa from this implementation. This approach might avoid the complexities and potential inaccuracies when applying the methodology to regions with less data or different epidemiological contexts.

6. **Impact of adjustment in multipliers**

It was noted that in regions like Asia, where AEM (Asian Epidemic Model) is used, any changes might be counteracted by adjustments in multipliers within Spectrum, making the overall impact less significant.

7. **Lack of data from low-income countries:**

Acknowledged the challenge of lacking data to disaggregate excess mortality in low-income countries. The group discussed the need for more studies or data to inform this methodology before its broader application.

Recommendations

1. Implement the disaggregation of AIDS-related and non-AIDS-related mortality for the next round of estimates, in countries outside of SSA.

Detailed recommendations arising from this session are available in **Appendix A**.

Session 5: Treatment interruption

Session 5, **chaired by Cari van Schalkwyk**, focused on two primary objectives:

- **Review outcomes from 2024 estimates**
- **Provide recommendations on guidance and definitions for calculating Spectrum interruption inputs**

Impact of new default rates, how many countries adopted defaults vs. entered national program data

Eline Korenromp / Rob Glaubius

Background

Treatment interruption has significant effects on disease progression and mortality for individuals on ART. In Spectrum, treatment interruptions influence the age distribution of patients on ART, the distribution of all people living with HIV (PLHIV), and the estimation of new infections. These interruptions also impact the estimation of HIV-positive pregnant women, mother-to-child transmission (MTCT), and child ART coverage. Previous rounds of Spectrum estimates had indicated older age and higher CD4 counts among PLHIV than observed in PHIA surveys.

UNAIDS/Spectrum 2024 guidance on treatment interruption

- **Mandatory entry of treatment interruption rates:** For the first time, all countries were required to enter non-zero treatment interruption rates for all years with ART entries in Spectrum.
- **Replacement of 'Lost to Follow-Up' terminology:** The term 'treatment interruption' replaced the previous term 'Lost to Follow-Up'.
- **Default rates:** Countries were advised to use default interruption rates where reliable national program data was not available:
  - **1.6% per year for high-income settings** (based on ART-CC data).
  - **5% per year for other settings** (based on expert opinion).
- **No fixed definition for treatment interruption:** UNAIDS did not provide a fixed definition for treatment interruption, recognizing that suitable metrics and existing data may vary depending on each country's clinical, program, and data systems.
- **Importance of long-term interruptions:** Long-term treatment interruptions, which have clinical and mortality impacts, are more critical to model in Spectrum than short-term interruptions, which are often recorded for the purpose of tracing and re-engaging patients.
- **Submission of country-specific definitions:** Countries were asked to submit the definitions of treatment interruption that matched the data they entered into Spectrum. However, only a few countries provided these definitions, and there was considerable variability in the definitions used.

Key findings

In the 2023 round, 52% of countries entered 0% interruption rates, while in the 2024 round, this dropped significantly to only 4% (6 countries).

For the 2024 round:

- 63% applied the new default rates.
- 34% entered non-zero, non-default values.
- 30% of files showed no change from the previous round.
- High-income countries tended to report higher-than-default interruption rates.
- Other income groups averaged slightly above the 5% default, with a noticeable peak during 2020-2021 likely related to COVID-19.

Figure 5.1 shows the comparison of interruption rates from the 2023 and 2024 Spectrum files, indicating that most countries that reported 0% interruption in 2023 moved to non-zero rates in 2024.

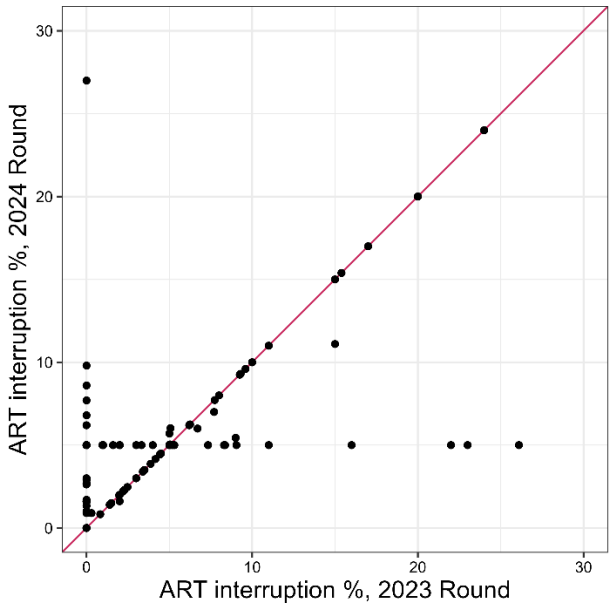


Figure 5.1: Comparison of ART interruption rates between the 2023 and 2024 Spectrum rounds. The diagonal red line indicates no change between rounds.

Impact of varying ART interruption rates on key HIV indicators

Interruption rates were varied from 0% to 25% per year to observe potential changes in Spectrum estimates.

- **ART coverage:** Expected to shift towards younger adults as interruption rates increase.
- **HIV-related mortality:** Expected to shift towards older adults with higher interruption rates.
- **PLHIV estimates:** Expected to remain stable, regardless of the variation in interruption rates.

Conclusion

Korenromp raised the question of whether the default ART interruption rates for adults, particularly the 1.6% rate for high-income countries, should be updated considering the few non-default entries provided in the 2024 round.

GAM-reported cohort retention on ART, data by country and over time

Eline Korenromp/ John Stover / Anna Yakusik

GAM indicator

The GAM indicator for retention on ART measures the percentage of patients who remain on ART over time, considering those who have stopped treatment, died, or are lost to follow-up.

Strengths

- **Comprehensive tracking:** The GAM indicator captures a wide range of outcomes (stopped treatment, died, or lost to follow-up), providing a holistic view of patient retention on ART.
- **Consistency:** It allows for consistent reporting and comparison across different countries and cohorts, offering valuable insights into the effectiveness of ART programs over time.

Weaknesses

- **Data quality issues:** The accuracy of the indicator is dependent on the quality of the underlying data, which can vary significantly across countries, leading to potential inconsistencies.
- **Inconsistent definitions:** Variations in how countries define and report “lost to follow-up” and other categories can affect the comparability of the retention data, reducing its reliability in cross-country comparisons

Retention rates were generally higher in high-income countries, among adults, and among women. However, retention decreases over time within any given cohort, which may be indicative of trends in treatment interruption.

WHO definitions of Loss to Follow-Up and treatment interruption

Hiwot Haile-Selassie, WHO Strategic Information and Analytics, presented the WHO 2022 HIV Strategic Information Guidelines.

Key recommendations on Treatment interruption

The guidelines recommend continuous assessment of treatment interruption using person-centred patient data. The goal is to enhance re-engagement and retention in care. Monitoring systems should be robust enough to identify treatment interruptions promptly, allowing for timely interventions, including patient tracing.

Definition of Loss to Follow-Up

WHO defines loss to follow-up as patients who have not been seen at a facility or community service delivery site for 28 days or more since their last scheduled appointment, including missed ARV drug refills. This definition was revised from 90 days to 28 days to support earlier identification of treatment interruptions and enable quicker re-engagement in care.

Differentiating treatment interruption from LTFU

There is a need for harmonization and standardization of treatment interruption definitions across countries and partners.

- LTFU refers to patients who are no longer in contact with the health system (with outcomes unknown)
- Treatment interruption involves the cessation of ART, usually unplanned, for a specific period.

Clinical implications of Treatment interruption

- **Viral load and transmission:** Interruption of ART can lead to rapid viral rebound, increasing the risk of HIV transmission and disease progression.
- **CD4 count and opportunistic infections:** Stopping ART decreases CD4 counts, heightening the risk of opportunistic infections and advanced HIV disease.
- **Drug resistance:** Treatment interruptions can lead to HIV drug resistance, potentially compromising the effectiveness of current ARV regimens.
- **Adherence management:** Prolonged interruptions may obscure adherence issues, complicating the management of HIV care.

Programmatic and monitoring considerations

- Harmonized definitions of treatment interruption are crucial for consistent monitoring across different service delivery models.
- Definitions must be simple and actionable to support program implementation and timely intervention.
- Overestimating the impact of treatment interruption is safer than underestimating it, especially for modelling the incidence, morbidity, and mortality of HIV.

Next steps

WHO plans to convene a technical consultation with the HIV patient monitoring group in Q3-Q4 2024 to agree on a standardized definition of treatment interruption. This process will be integrated into the development of the WHO HIV treatment guidelines, with ongoing inputs and updates from global partners.

Discussion

This discussion emphasized the need for clear, standardized definitions of LTFU and treatment interruption to ensure that Spectrum can provide accurate and meaningful insights into HIV epidemics and the effectiveness of interventions.

In the context of the PEPFAR IIR indicator, concerns were raised about the overlap between LTFU and treatment interruption, particularly regarding the implications for understanding patient mobility, mortality, and viral suppression.

The response highlighted that in Spectrum, the number of individuals on treatment is treated as a fixed input. The interruption rate mostly shifts around the age distribution of the people on treatment, which can have impact in some countries where incidence is estimated by age based on the numbers on treatment by age. Consequently, if the age distribution is inaccurately represented, it can lead to misestimations in age-specific HIV incidence. This was the main motivation in closely examining interruption rates to ensure that what is entered into Spectrum accurately depicts real world dynamics.

Proposal: Default relationship between WHO treatment interruption and treatment discontinuation

Jeff Imai-Eaton/ Rob Glaubius, focused on clarifying and refining the concepts of loss to follow-up (LTFU) and treatment interruption within Spectrum, with an emphasis on how these definitions impact HIV estimates and model outputs.

Key concepts

Loss to Follow-Up (LTFU)
What programs can measure / calculate

- Defined as patients not seen at a service delivery site more than 28 days after a missed appointment. It excludes individuals who are recorded as deceased or have transferred out of the program, but includes those with unknown outcomes, such as non-ascertained deaths or silent transfers.

Treatment interruption
What is important for models + currently asking users to divine for Spectrum input

- Refers to a clinically meaningful gap in treatment, which is critical for models like Spectrum to accurately predict outcomes such as mortality and disease progression.

Current Spectrum input

Within the current Spectrum model, users are asked to input the number of patients receiving ART each year and the percentage of those who interrupt treatment. However, the model does not account for patients who have died or switched service sites, which could lead to inaccuracies if LTFU rates are used instead of actual interruption rates.

Considerations for Spectrum model

Why do we need to model 'interruption rate' instead of simply inputting observed LTFU rates?
Interruption rate does not affect the total number of patients on treatment or the overall treatment coverage because Spectrum matches the input number for total ART. If the interruption rate is too high, the model will cycle through more ART initiations, potentially overstating mortality rates among people who may have simply transferred or are incorrectly categorized. One of the significant impacts of overstating interruption rate is on the age distribution of patients on ART. If the age distribution is misrepresented due to incorrect interruption rates, it can lead to errors in estimating HIV incidence by age. This is particularly important in countries where incidence is estimated based on the age distribution of people on treatment.

Proposed revisions to Spectrum input

Explicitly represent **"Lost to follow-up"** and **"Treatment interruption"** as separate concepts in Spectrum

To address these issues, several revisions were proposed on how LTFU and treatment interruption are represented in Spectrum:

1. Users would input separate percentages for LTFU and treatment interruption, aligning with WHO-recommended definitions and **what programs can realistically measure**.
2. Specify default relationship between between 'Percent LTFU' and 'Percent interrupted', which users could modify 'interruption rate' if they have local data or specific insights into actual interruption rates.
3. Model uses **percent interrupted** input for flow from 'on ART' to 'untreated'. A 'waterfall' visualization would help users distinguish between the LTFU observed in their programs and the modelled 'interrupters'. This would make the data and its implications clearer.

The strengths of the proposed revisions are that what users are asked to report in Spectrum are more consistent with the guideline (WHO, GAM, etc) and what programs can calculate, making the distinction between lost to follow up that tis being observed and interruption that the model is using, more explicit to users.

Further discussion

- The definition of what constitutes a 'clinically meaningful interruption' is still under discussion, with WHO planning to finalize this definition by the end of the year.
- The need to differentiate between short-term and long-term interruptions within the model. This distinction would require new compartments within the model, which could complicate its structure.
- Possible options for setting default interruption rates in the model:
    - Continue using the 5% default interruption rate (or 1.6% for high-income settings) regardless of the LTFU percentage entered.
    - Establish a scalar relationship between observed LTFU rates and interruption rates. This could involve analysing available LTFU data and determining an average relationship to specify a default scalar.
- The current model inputs vary only by year, but there is increasing capability among programs to calculate LTFU rates stratified by sex, age, and ART duration. Should entry of LTFU or interruption rates **vary by sex / age / ART duration**, to reflect the variability in interruption by duration on treatment as seen in GAM data.
- Review and potentially adjust the assumptions about CD4 recovery following treatment interruption, ensuring they align with observed CD4 distributions. The 2024 estimates round highlighted that few countries can accurately calculate treatment interruptions, making it challenging to validate these assumptions.

Discussion

Key points from discussion:

1. Relationship between Loss to Follow-Up and treatment interruption

The discussion highlighted whether loss to follow-up (LTFU) rates are systematically related to treatment interruption rates. It was noted that if LTFU is used as a proxy for interruption, it could lead to overstating mortality and ART re-initiation rates due to the high mortality associated with early treatment interruptions.

2. Feasibility of implementing a differentiated approach

Participants questioned whether the model should differentiate between short-term and long-term treatment interruptions and whether it should account for variation by age, sex, and ART duration. It was agreed that the current proposal aims to simplify this by separating LTFU and interruption rates within Spectrum, with the potential to build a relationship between them based on available data.

3. **Understanding the relationship between loss to follow-up and treatment interruption**

During the discussion, several data sources were mentioned, to help define the default relationship between these two metrics within the Spectrum model.

- **Tracing studies**: One of the outcomes recorded in tracing studies is whether a patient is on treatment elsewhere. Data from Malawi was highlighted as an example where patients flagged as lost to follow-up were later identified as being on treatment elsewhere. Tracing studies could help in distinguishing between true treatment interruptions and administrative loss to follow-up.

- **Viral suppression**: In Zambia, 60 to 70% of people initiating treatment are virally suppressed. This type of population-level data could be used to help trace individuals and inform the relationship between loss to follow-up and treatment interruption.
- **Tier.net data:** Includes approximately 260,000 patients, mostly women, with records from 2017 to 2022. The suggestion was made to re-analyse this data to examine the correlation between loss to follow-up and treatment interruption. However, it was noted that tier.net has limitations – specifically, that outcomes are not well captured, and loss to follow-up simply indicates that a person disappears from the dataset without being linked across IDs or facilities. Despite these limitations, re-analysing the data could still offer insights, particularly if it proves useful in understanding the relationship between loss to follow-up and treatment interruption.

4. Input on retention rates over time in Spectrum

The possibility of entering retention rates over time into Spectrum was raised, to provide a more comprehensive way of distinguishing between short-term and long-term effects on patients, particularly in relation to loss to follow-up and treatment interruptions. It was suggested that resurrecting the concept of tracking retention rates over time, including loss versus follow-up, could offer valuable insights and better validation of data within the Spectrum user interface. As a significant amount of data entry is already required, there was hesitation about whether the added benefit of detailed retention tracking would justify the increased burden on users.

5. Proposed changes

It was reiterated that the proposed changes involve asking countries to enter measurable LTFU rates based on the 28-day WHO definition, with Spectrum then applying a default relationship to estimate treatment interruptions. This aims to align user input with what programs can realistically measure while maintaining the model's accuracy.

Recommendations

In summary, no change to Spectrum default treatment interruption rates were recommended on the basis that (1) national ART programs are not able to accurately measure rates of treatment interruption (different from programme observed loss to follow-up) and (2) there is not a clear relationship between treatment interruption rates and LTFU rates.

The full recommendations arising from this session are shown in **Appendix A**.

Session 6: World Population Prospects 2024 update

The session's primary objective was outlined by chair **Oliver Stevens**: **Review impact of new population estimates on HIV estimates**.

Impact of WPP 2024 vs 2022 estimates on HIV estimates

Rob Glaubius

Background on Spectrum demographic inputs

- Spectrum's default demographic inputs are derived from the World Population Prospects (WPP) revisions, which include parameters such as 1970 population size, fertility rates, sex ratio at birth, life tables, and net migration.
- The WPP 2024 was released on July 11, 2024, and new demographic inputs based on these revisions are in progress.
- Countries often customize Spectrum's default inputs to align with their own national estimates.

Population comparison (2023)

A comparison of population size ratios against WPP 2024 estimates (Figure 6.1) shows:

- 64% of countries had AIM population estimates smaller than WPP 2024.
- 9 countries (5%) had AIM populations more than 5% larger than WPP 2024 estimates.
- 23 countries (13%) had AIM populations more than 5% smaller than WPP 2024 estimates.
- Differences were observed between AIM and WPP 2024 estimates, with significant deviations noted in countries such as Sao Tome and Principe, South Sudan, and Central African Republic.

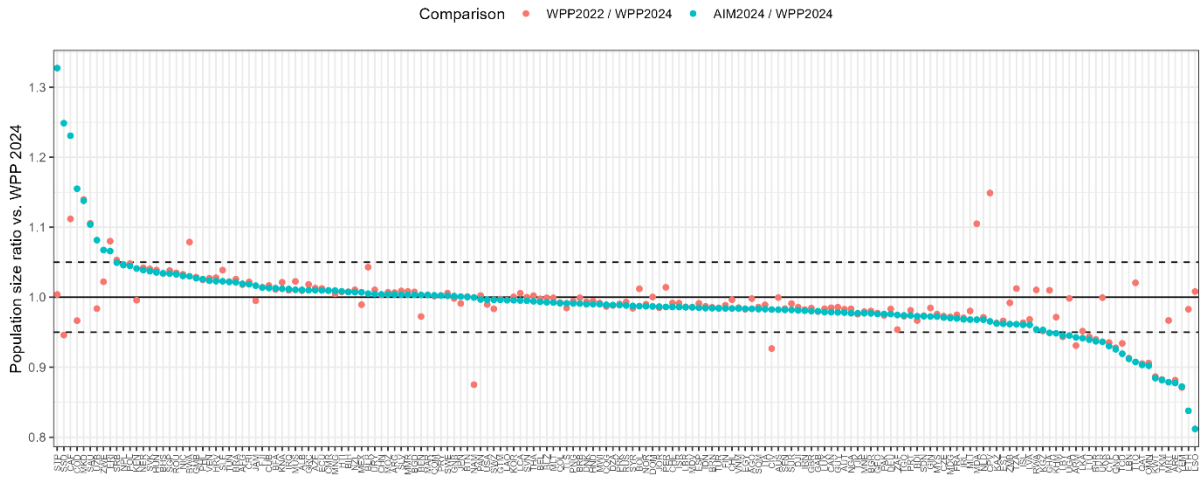


Figure 6.1: Comparison of population size ratios against WPP 2024 estimates. The plot shows the population size ratio for each country, comparing WPP 2022 to WPP 2024 (in red) and AIM 2024 to WPP 2024 (in blue). The horizontal dashed lines represent a 5% deviation threshold. Countries with ratios above or below these lines indicate significant differences between their population estimates in WPP 2022 and AIM 2024 compared to WPP 2024.

AIM population differences compared to WPP 2024

- **Larger AIM population:** There are nine countries where the AIM population is at least 5% larger than in the WPP 2024 estimates. These countries might see substantial changes in their HIV estimates if they adopt the new population data. Some of these countries, such as those without a recent census, face a lot of uncertainty regarding their actual population sizes.
- **Closer to WPP 2022:** Certain countries, such as **Yemen**, **United Arab Emirates**, **Turkmenistan**, and **Kuwait**, have AIM populations that align more closely with WPP 2022 than with WPP 2024. This suggests

that these countries may be slower to adopt the latest WPP figures or might continue using their existing population data.

- **Zimbabwe:** Zimbabwe is an interesting case because it uses subnational AIM files for estimates, aligning more closely with their national Statistical Office projections rather than the latest UN Population Division estimates. This might result in fewer adjustments when transitioning to WPP 2024 data.
- **Middle Eastern countries:** Many Middle Eastern countries have AIM estimates well-aligned with previous WPP figures, but recent updates in WPP rounds have led to significant changes. If UNAIDS or these countries adopt WPP 2024, they may experience substantial shifts in their population estimates.

Implications for HIV estimates

- **Population sizes:** Most (64%) countries would see larger population size estimates with the adoption of WPP 2024.

1. HIV incidence calibrated to HIV prevalence data

- When HIV incidence is closely linked to HIV prevalence data, an increase in population size will result in a proportional increase in the number of PLHIV.
- However, if the number of people on ART does not increase at the same rate, this could lead to a lower percentage of ART coverage.

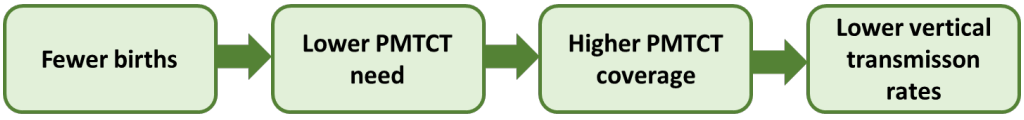
2. HIV incidence calibrated to new diagnoses and/or deaths

- In contrast, when HIV incidence is more closely linked to new diagnoses and/or deaths, the number of PLHIV may not increase significantly despite a larger population.
- This could lead to a lower HIV prevalence rate (because the population size is larger), but ART coverage as a percentage might remain similar if both the number of people on ART and PLHIV increase proportionately.

AIM birth differences compared to WPP 2024

- **Higher AIM births**: 57% of countries have more births in AIM than what is estimated in WPP 2024 overall. 64 countries (37%) have AIM births that are more than 5% higher than the births estimated in WPP 2024. These countries might see significant adjustments if they switch to WPP 2024 data.
    - **AIM births within 5% of WPP 2022:** Philippines, Poland, Argentina, Hungary, China, Republic of Korea
- **Lower AIM births**: 37 countries (21%) have AIM birth estimates that are more than 5% lower than WPP 2024. These countries may experience significant discrepancies when compared to the latest WPP estimates.
    - **AIM births within 5% of WPP 2022:** Yemen, Qatar, Turkmenistan, Kuwait, Lebanon, Bahrain

Implications for HIV estimates



- **Fewer estimated births**: With the adoption of WPP 2024, most (57%) countries are expected to have fewer estimated births.
    - Fewer births will reduce the overall need for PMTCT services, leading to lower demand for resources and possibly more effective allocation of existing resources.
    - As the demand for PMTCT services decreases, the percentage of coverage within the PMTCT program could increase, ensuring that a higher proportion of pregnant women receive necessary interventions.
    - With improved PMTCT coverage, the rates of vertical (mother-to-child) HIV transmission are expected to decrease, contributing to better health outcomes and a reduction in new paediatric HIV cases.

Discussion

- The adoption of WPP 2024 by over half of the countries could lead to significant changes in estimated births, potentially inflating PMTCT coverage rates above 100%. These countries may need help to maintain accurate paediatric HIV estimates.
- Several Middle Eastern countries may see larger population sizes and lower ART coverage if WPP 2024 is adopted.
- Countries using subnational AIM files may diverge from WPP estimates, which is acceptable if demographic inputs are regularly reviewed and updated.

Next steps

- Prepare demographic inputs based on WPP 2024 for use in Spectrum.
- Share tools with UNAIDS to help identify countries with substantial shifts in population and/or birth estimates for enhanced demographic support.

Discussion

Key inquiries from discussion:

1. **Impact of birth estimates on PMTCT coverage**:

It was noted that reduced birth estimates could push PMTCT coverage over 100% in regions outside of Africa. This issue is more pronounced in Latin America and the Caribbean, where treatment coverage among children has plummeted. Some countries might need to realign their estimates to match national statistics better.

2. **Discrepancies in population projections**

A significant issue was raised regarding the large differences in population estimates between WPP and national projections. In some countries, this discrepancy is as large as 20 million, which significantly impacts HIV estimates. The UN Population Division was acknowledged for being open to reviewing these discrepancies and refining their methods if needed.

3. **Variation in birth modelling**

An observation was made that while discrepancies between AIM and WPP population estimates were consistent, the variation in birth estimates between versions was not, leading to questions about the birth modelling process. There was no immediate explanation, indicating a need for further investigation.

4. **Fertility rate adjustments**

The use of fertility rate adjustment factors in optimizing PMTCT cascades was discussed. However, the potential impact of these adjustments considering the new WPP data remains unclear. It was noted that deviations from fitted values could lead to unpredictable outcomes.

5. **Challenges with ANC data and birth estimates**

Efforts to triangulate birth estimates using ANC data have been challenging. There appears to be a disconnect between program-related birth data and estimates from spectrum or other sources, with concerns about data double counting and inconsistencies.

The main recommendations following these presentations are captured in **Appendix A.**

Session 7: Dynamical modelling of HIV trends in Key Populations in sub-Saharan Africa

In previous years, UNAIDS primarily relied on simple static models for producing key population HIV estimates, which were increasingly recognized as inadequate for assessing trends. In response, a decision was made in 2022 to support the development of dynamic models. A call for proposals led to the selection of teams to develop and refine dynamic models specifically designed for key population estimates.

Key Population Modelling Technical Working Group: aims, processes, and deliverables

Sharmistha Mishra, **University of Toronto**, gave an update on processes and deliverables of the Working Group as they worked towards integrating Key Population HIV estimates into national processes for sub-Saharan Africa (SSA).

For a detailed recap of the recommendation on the proposed approach and development process to integrate Key Population modelling in HIV estimates process, refer to the May 2023 UNAIDS Reference Group meeting report.[5]

Aims

The primary aims of the Key Population Modelling Technical Working Group include:

1. **Developing Goals-ARM:** To estimate Key Population HIV indicator trends and transmission dynamics in SSA.
2. **Convene a Working Group:** To provide guidance on the development and review of Goals-ARM for estimating Key Population indicators.
3. **Pilot processes:** Engage teams from 2-5 countries to develop and pilot processes for producing Key Population HIV estimates as part of the nationally led HIV estimates process.

Processes

1. Proposed approach and process

- Integrate Key Population HIV estimates into national HIV estimates.
- Engage a broad range of relevant experts and stakeholders in Key Population data and programs.
- Review and collate relevant national data.
- Determine relevant counterfactual-based indicators and ensure appropriate and non-stigmatizing interpretation and communication of indicators.

2. Progress to date

- **Goals-ARM development:** Address dynamics of early epidemic take-off.
- **Technical Working Group:** Convene a group of modelers and epidemiologists to review and guide the model's development.
- **Country engagement:** Engage teams from selected countries to develop and pilot the process.

Key Population Working Group deliverables

1. **Reports**
    - Detailing reviews and analyses to inform model structure and assumptions by Key Population.
    - Describing methods that apply across Key Populations (e.g., mixing matrix model fitting).
    - Outlining remaining limitations and proposed future work.
2. **Model development and modifications**
    - Prepare data and develop structures and equations to avoid underestimating new HIV infections among Key Populations and their partners.
    - Capture uncertainties in calibration targets and behavioural inputs through consistency checks and triangulation.
3. **Internal and external validity checks**
    - Conduct internal validity checks of model outputs.
    - Compare outputs with reference Key Population models to ensure accuracy.

5. Technical Updates Reference Group Meeting Report: May 2023. Available at: https://epidem.org/technical-updates-meeting-may-2023/.

Composition of Key Population Modelling Technical Working Group

- **Epi-modelers:** 15-20 persons, virtual meetings, inviting teams/persons working on Key Population-modelling across epidemic contexts. The group includes diversity, including new/junior trainees, modelers/epidemiologists in Ministries of Health (MOH) receiving Goals training, and academics supporting MOH with HIV models (bespoke, or may use Goals, IPM, etc.), with potential overlap with the 2-6 countries where Goals-ARM can be first pilot-tested / test-cases.
- **Program experts and community leaders:** 5-8 members, separate expert advisory meetings (virtual). A consultative, rather than participatory research approach to community engagement will be used, with introductory sessions to address knowledge asymmetry and co-develop expectations, followed by itemized input into decisions impacting Key Population indicators and addressing uncertainty or missing data.

Plans for pilot testing

- Engage teams from 2-5 countries to develop and pilot a process for producing Key Population HIV estimates. Given the potential timelines for pilot in early 2025, 2 countries are feasible.
- Use Key Population Workbooks and Imperial/UNAIDS data to input and review calibration targets and parameter inputs.
- Conduct and interpret internal validity checks.
- Apply and run Goals-ARM model for Key Population indicators.

Model development and modifications guiding principles

1. **Structure and equations**
    - Reduce potential to underestimate new HIV infections in the context of Key Population and their partners.
    - Prepare for new Key Population indicators, such as long-term epidemic consequences of population-specific prevention/treatment gaps.
2. **Use of data**
    - Capture uncertainty in calibration targets and behavioural inputs by formalizing consistency checks, triangulation, and adjustments using Imperial/UNAIDS Key Population datasets and syntheses.

Preparation for country pilot/test case

Model development and modifications

1. **Model structure (Q4 2024):** Prioritization based on first Technical Working Group recommendations.
    - Application of STI for early epidemic take-off.
    - Key Population strata, client population size & overlap.
    - Force of infection: multiple exposures, partnership overlap, approximate partnership durations.
    - Turnover and periodic risks.
2. **Model checks**
    - Internal validity checks of outputs.
    - External validity: compare outputs with reference Key Population models (e.g., ratio of 1-year new HIV infections FSW: clients).
3. **Estimation of indicators with uncertainty**

Data preparation

1. **Calibration targets:** From Key Population Workbooks and Imperial/UNAIDS Key Population datasets.
    - Develop formal checks and adjustments for use in the model (e.g., FSW: client ratio).
2. **Input parameter synthesis:** From Key Population Workbooks and Imperial/UNAIDS Key Population datasets.
    - Per-country and regional (for missing country data).

- Develop formal triangulation checks and adjustments (e.g., condom-use over time if from multiple sources).

Goals-ARM development update

Rob Glaubius, **Avenir Health Inc,** gave a technical development update on the Goals-ARM model, particularly in the context of sexually transmitted infection (STI) symptom prevalence time trends, including how these updates were applied in various country case studies.

Technical update to integrate STI symptom prevalence time trends

The Goals-ARM was updated and now incorporates STI symptom time trends – borrowed from calibrated Goals risk-structured models. Populations were mapped between models to ensure consistency. STI symptom prevalence trends are categorized by age, gender, and risk groups. These trends are calibrated using Demographic and Health Surveys (DHS) and Key Population study data and align with HIV prevalence patterns from the AIM model.

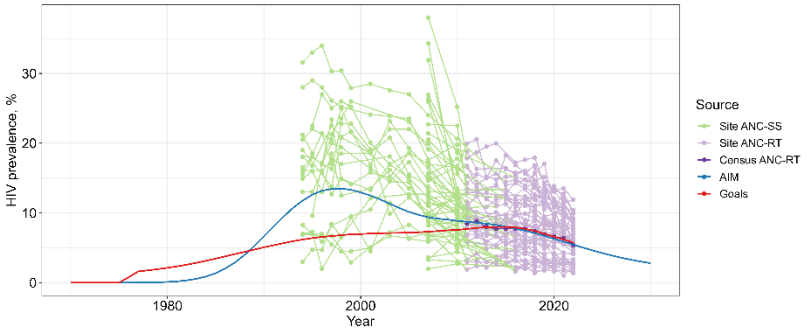
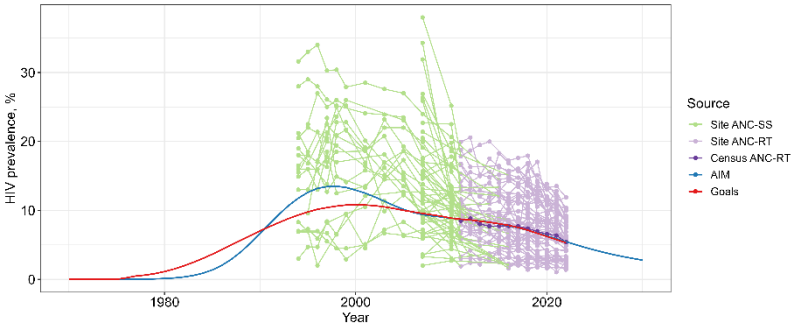
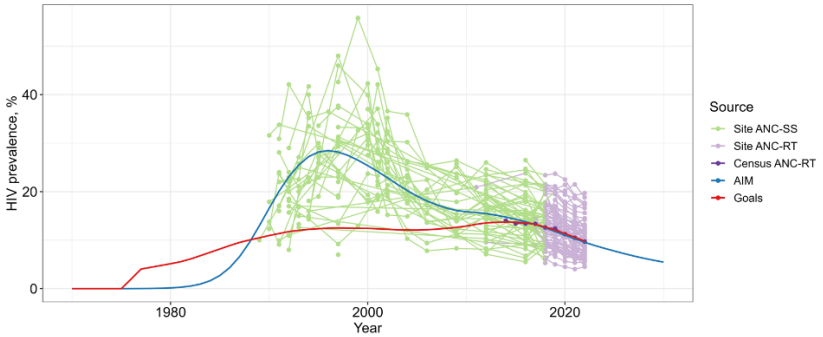
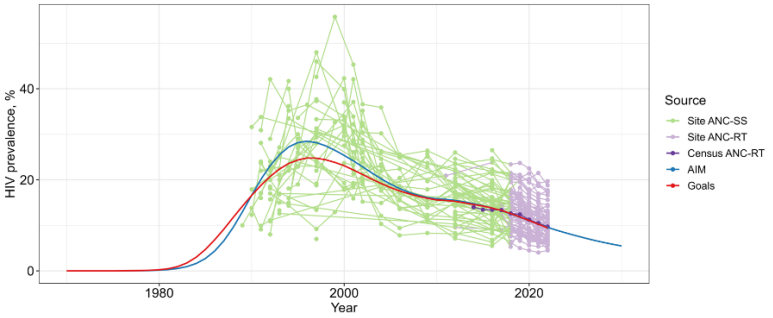
Early prevalence levels are assumed to be high to reflect initial epidemic conditions. Self-reported STI symptoms in household surveys (HHS) were used to estimate age patterns, which are assumed to be constant over time. A simplifying assumption was made, applying patterns for previously married/cohabiting populations to Key Populations. There is a preference to use Key Populations study data, as Key Population studies often test for specific STIs, while HHS data usually report self-reported symptoms. Reconciling these different data sources is a complex, ongoing task.

Case studies: Performance of updated model

The updated model's performance was tested, using data from Malawi and Zimbabwe, captured in Table 7.1 below.

While the inclusion of STI effects has improved the alignment of the model with historical HIV prevalence data, there are still areas where the model's fit can be refined, particularly concerning age-specific prevalence and key population data.

Table 7.1: Goals-ARM was updated to include modelling of sexually transmitted infection (STI) symptom prevalence over time. The updated model was tested with two countries, Malawi and Zimbabwe, modelling HIV prevalence in pregnant women.

Country	Without STI effects	With STI effects
Malawi		 Including STI effects results in a higher initial prevalence and a sharper decline, indicating a more significant impact of STIs on HIV transmission dynamics.
Zimbabwe		 The updated Goals-ARM model provides a better fit to the observed data, especially in capturing the early peak and subsequent decline in HIV prevalence.

5. Technical Updates Reference Group Meeting Report: May 2023. Available at: https://epidem.org/technical-updates-meeting-may-2023/.

The desk calibration in South Africa included:

1) HIV prevalence data from household surveys (2005, 2008, 2012, 2016, 2017, and 2022)
2) HIV prevalence data from 38 Key Population studies covering:
    - 18 studies on female sex workers (FSWs) and men who have sex with men (MSM), respectively.
    - 3 studies on people who inject drugs (PWID) and transgender women (TGW), respectively.

3) Mortality data

- Data on all-cause deaths by age and sex from 1997 to 2018.

Adjustments for missing data and reporting completeness were made, the adjustments for reporting completeness were shared by Leigh Johnson and the likelihood guided by the Thembisa manual.

In South Africa, the model fits reasonably well with data but struggles with heterogeneity in key populations which are often not nationally representative. Studies on men who have sex with men (MSM) and people who inject drugs (PWID) show considerable heterogeneity, as seen in Figure 7.1 and 7.2. The approach used in concentrated epidemics for dealing with data heterogeneity may not generalize well to South Africa, where data comes from diverse studies with varying methodologies.

Next steps

- Address spatial/methodological variability.
- Consider excluding lower-quality data or using bias-absorbing terms in the model.

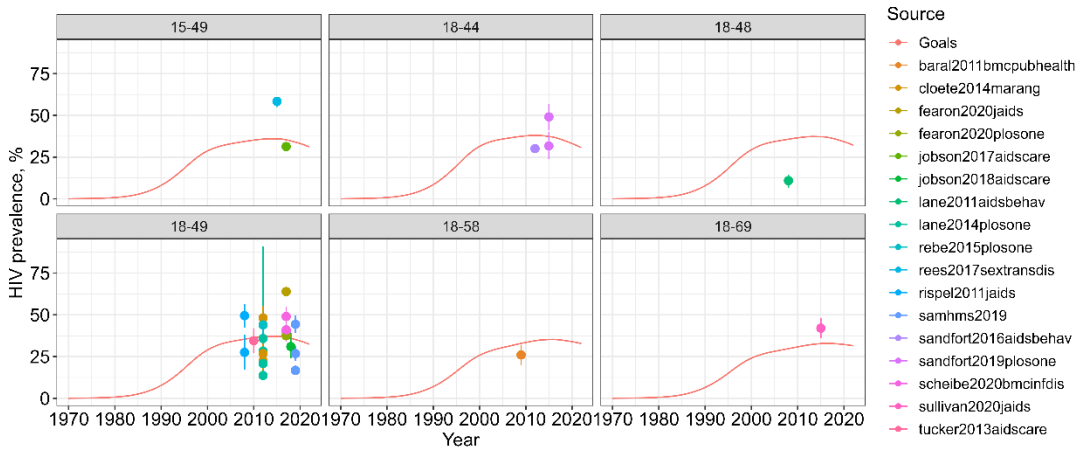
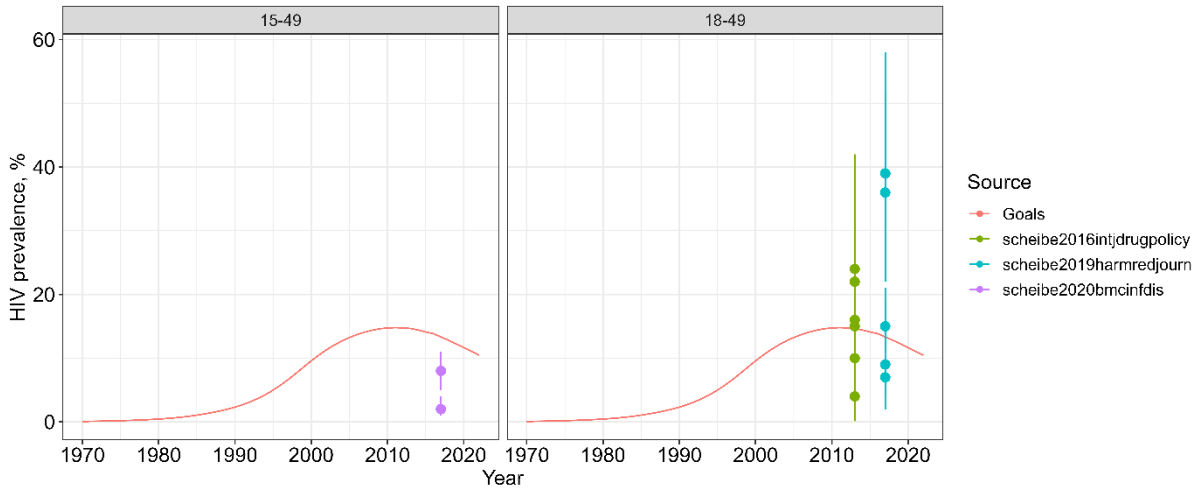


Figure 7.1. The models fit to HIV prevalence data among men who have sex with men, highlights data heterogeneity across various studies.



5. Technical Updates Reference Group Meeting Report: May 2023. Available at: https://epidem.org/technical-updates-meeting-may-2023/.

Figure 7.2. The models fit to HIV prevalence data among people who inject drugs. Like the MSM data, the red line indicates the model's estimated HIV prevalence, with the coloured points representing observed data from different studies.

Discussion

Key inquiries & clarifications:

1. **Flexibility in defining key populations, such as prisoners**
   The preference is to keep key population definitions based on behavioural risks rather than structural factors like being in prison. Prisoners are not considered a risk group per se; rather, the conditions leading to imprisonment (such as criminalization of certain behaviours) might be associated with higher risk. For example, if a country criminalizes men who have sex with men (MSM), the risk isn't necessarily because they are in prison but because of the reason for their imprisonment. The model should focus on behavioural risks, and while prisoners might be a monitored population, the model would not include them as a distinct key population unless the behavioural risk factors warrant it.
2. **AIM vs. Goals benchmarking**

- The intention is not to match AIM results precisely, as AIM might be more sensitive to specific data points, especially outliers. The Goals-ARM model, which includes additional behavioural data, should reflect a broader and potentially more accurate range of dynamics, particularly when introducing co-factors like STI prevalence. However, alignment is essential to ensure consistency, especially given the built-in acceptability of AIM among countries.
- South Africa: benchmarking against Thembisa. Additionally, specific details regarding mortality calibration targets in the South African context need to be clarified and refined.

3. **FSW calibration targets**
   Female sex worker age and region-specific prevalence ratios could be used as calibration targets, with overall prevalence used for validation.
4. **Primary HIV infection co-factor in the early epidemic phases**
   The model should incorporate primary HIV infection as a significant co-factor, especially in the early phases of the epidemic. This may explain the high STI cofactor effects observed, as these early infections combined with low condom use and high STI rates contribute significantly to transmission dynamics.
5. **STI co-factor effects**
   While it might be acceptable to allow the STI co-factor effect to absorb other behavioural risks, there is a strong recommendation to develop another behavioural co-factor.
6. **Consistency in surveillance data**
   Consistency in surveillance data varies widely by region. The working group needs to develop methods for dealing with inconsistent data, as consistent site surveillance over time is rare, especially in sub-Saharan Africa. Sampling methods and geographic variability should be carefully considered to ensure robust model outputs.

The following process recommendations were made to facilitate the pilot's reach for the Goals-ARM model:

1. **Early country engagement**
   Engaging countries early is important to build familiarity and gather input, so the model meets the needs of the countries. The focus is on engaging stakeholders in the technical aspects of the model and ensuring that relevant experts are involved. The broader engagement with stakeholders, including communities, will be carefully planned, potentially through UNAIDS.
2. **Focus on countries with established relationships**
   Prioritize countries where there are already established relationships and where the necessary data is more readily available. This approach will ensure more reliable feedback and smoother implementation during the initial pilot phase.
3. **Incorporate program and community feedback**
   Engage program experts, community leaders, and behavioural scientists and those that are not necessarily modellers, but use the data, in the process to ensure that the model addresses the needs and concerns of these stakeholders. This inclusion will help ensure that the model is not only technically sound but also practically useful and sensitive to the realities on the ground.

4. **Maintain flexibility in model application**
   Allow for flexibility in the model to accommodate country-specific data and KP definitions. This flexibility will enable countries to adapt the model to their local contexts, making it more relevant and accurate for their specific needs.
5. **Phased implementation approach**
   Implement the pilot in multiple stages. Start with a desk-based review and data collection, followed by in-person workshops where the model can be stress-tested.
6. **Incorporate technical support and training**
   Provide technical support and training as part of the pilot process, particularly through existing platforms such as the upcoming Goals workshops. Integrating Goals-ARM into these workshops could help build capacity and ensure that the country teams are well-prepared to use the model effectively.
7. **Prepare for long-term integration**
   While the pilot phase focuses on testing and refinement, there is a long-term vision for Goals-ARM to eventually replace Goals-ASM, and Symphony will replace AIM.

The main recommendations arising from the discussions following these presentations are captured in **Appendix A.**

Session 8: New infections by Key Populations and their partners

Chaired by **Sharmistha Mishra**, the objective of this session was to: **Review methods of estimating uncertainty in estimates**.

Keith Sabin opened the session with a discussion about the ongoing efforts to improve data collection from key populations and how these data are used. He highlighted the importance of engaging with communities and network populations to ensure that models reflect both the known realities and acknowledge the unknowns. The shift from using regional proportions for size estimates to national size estimates was a significant change in this round, aiming for more accurate and reflective data. Sabin underscored that this decision was practical and pragmatic, though it remains open for future discussions.

John Stover provided an update on the calibration and use of the Goals RSM model, which is currently implemented for 111 countries, sources for population size estimates and infections and provided insights into the results.

Key steps in updating the Goals models include:

- Updating program data using the latest data on ART and PMTCT services from the 2024 AIM files.
- Country-specific assumptions were incorporated into the model, including adjustments for ART, AIDS mortality, and migration.
- The latest key population size estimates from UNAIDS were integrated into the model.
- A significant update involved the inclusion of surveillance data from various studies, including those from the Estimation and Projection Package (EPP), the Asian Epidemic Model (AEM), and the database that Oli Stevens collated, covering 38 countries and 1,169 records for Sub-Saharan Africa (SSA).
- The model was re-fitted to ensure its accuracy and consistency with the updated data inputs

Table 8.1 shows the updated population size estimates for female sex workers (FSW), men who have sex with men (MSM), and people who inject drugs (PWID) across different regions. These updates reflect the variability within and across regions, a change from the previous approach where uniform regional estimates were used.

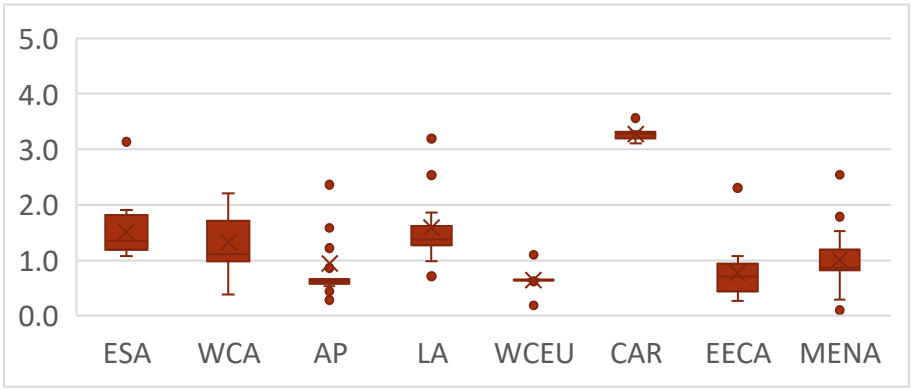
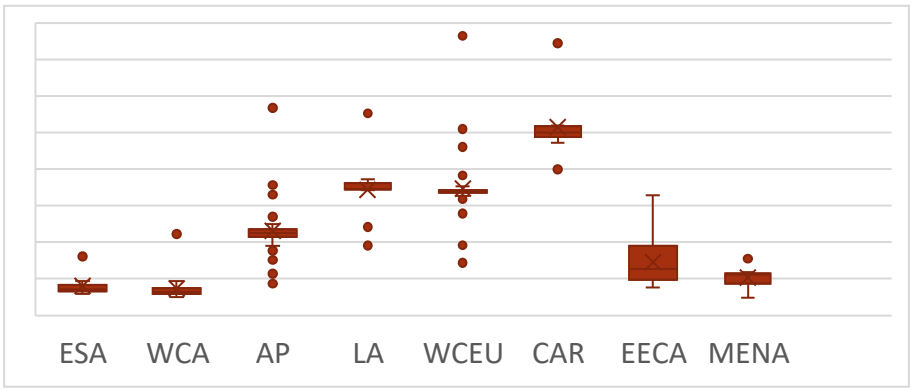
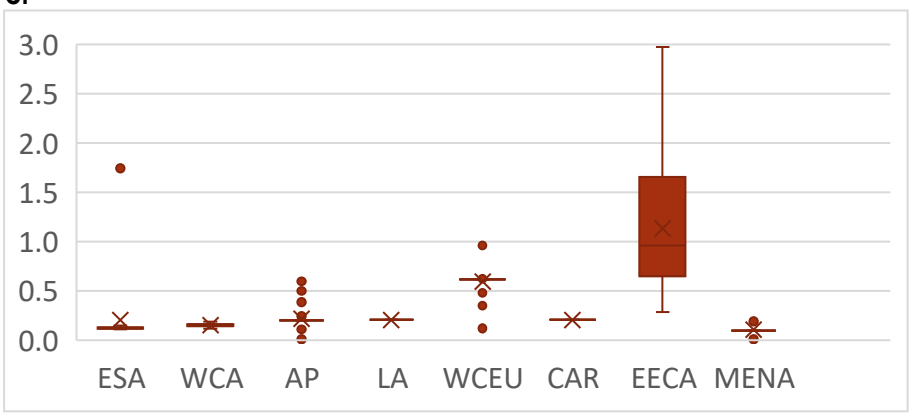
Plots	Observation
A. 	Percentage of female sex workers (FSW) relative to the female population aged 15-49 across regions. It shows the variation in population size estimates both within and across regions, highlighting the diversity in the prevalence of FSW in different settings.
B. 	Percentage of men who have sex with men (MSM) as a proportion of the male population aged 15-49 across regions. Like the FSW plot, it demonstrates significant variability in estimates, with some regions showing notably higher percentages.
C. 	Percentage of people who inject drugs (PWID) as a portion of the total population aged 15-49. Reveals wide-ranging estimates across regions, with some regions exhibiting considerably higher prevalence rates.

Table 8.1: A. Percentage of female sex workers (FSW) relative to the female population aged 15-49 across regions. **B.** Percentage of men who have sex with men (MSM) as a proportion of the male population aged 15-49 across regions. **C.** Percentage of people who inject drugs (PWID) as a portion of the total population aged 15-49.

Stover emphasized the integration of updated surveillance data into the Goals model, particularly for key populations. The inclusion of nearly 1,200 records from various studies, particularly from sub-Saharan Africa, allowed for more precise calibration of the model.

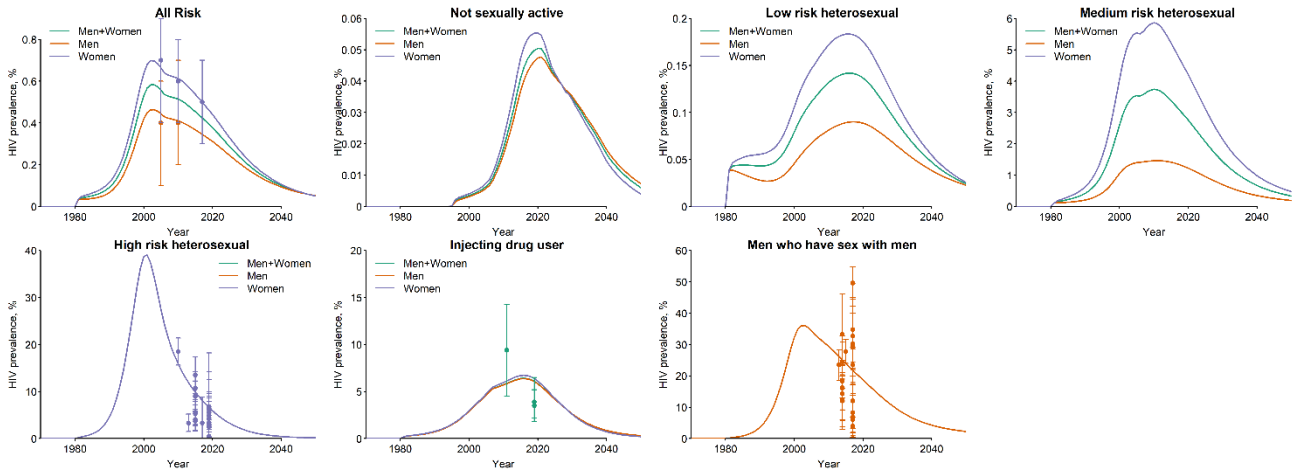


Figure 8.1. HIV prevalence trends in Senegal by risk group, 1980-2050, among different risk groups, including All Risk, Not Sexually Active, Low Risk Heterosexual, Medium Risk Heterosexual, High Risk Heterosexual, Injecting Drug Users, and Men Who Have Sex with Men (MSM).

For Senegal (Figure 8.1), the model successfully captured the overall HIV prevalence trend among the general adult population. However, when it came to key populations like MSM, the model's dynamics were more reflective of the general epidemic trends in the country rather than specific key population data. This is because the available data for key populations were clustered around one or two years, which limited the ability to model long-term trends accurately. Many key population studies are not nationally representative and often come from specific sites with varying methodologies. Stover acknowledged the inherent limitations, emphasizing the need for continued data collection and the importance of future rounds of estimates to improve the accuracy.

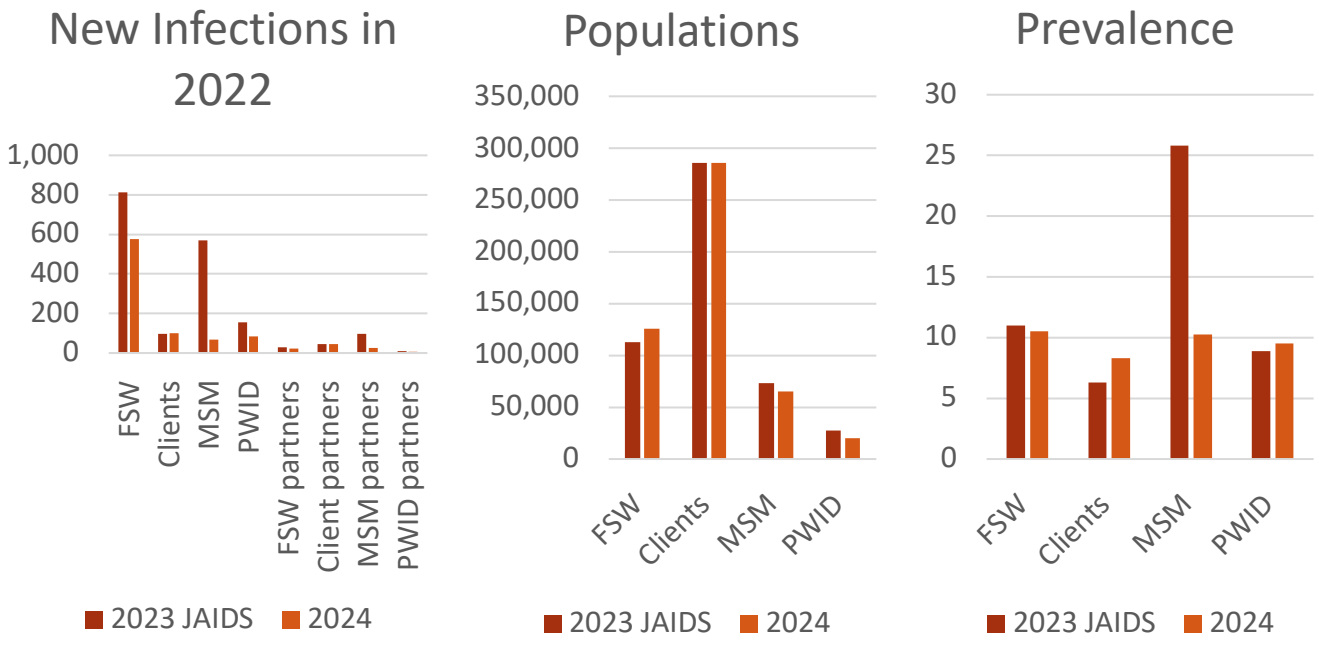


Figure 8.2. HIV prevalence and incidence estimates for key populations in Cameroon. The left chart compares the new infections estimated in 2022 between the previous model and the updated model. The middle chart displays the population size estimates for different key populations. The right chart shows the prevalence estimates for MSM, where new surveillance data caused reduction in MSM prevalence.

The results from Cameroon (Figure 8.2) illustrate how the updated model estimates compare to the previous year. The incorporation of more accurate and recent surveillance data led to a reduction in the estimated prevalence among MSM. Similar changes were observed in other countries due to various factors:

- **Tanzania**: The new estimate for HIV infections was higher due to a significant increase in the AIM file estimates for new infections among the 15-49 age group compared to last year.
- **Russia**: The estimates remained largely unchanged.
- **India**: A significant change was observed, particularly in the estimates for PWID and FSW. This was due to differences in the shape of the incidence curves despite using the same prevalence data as last year.

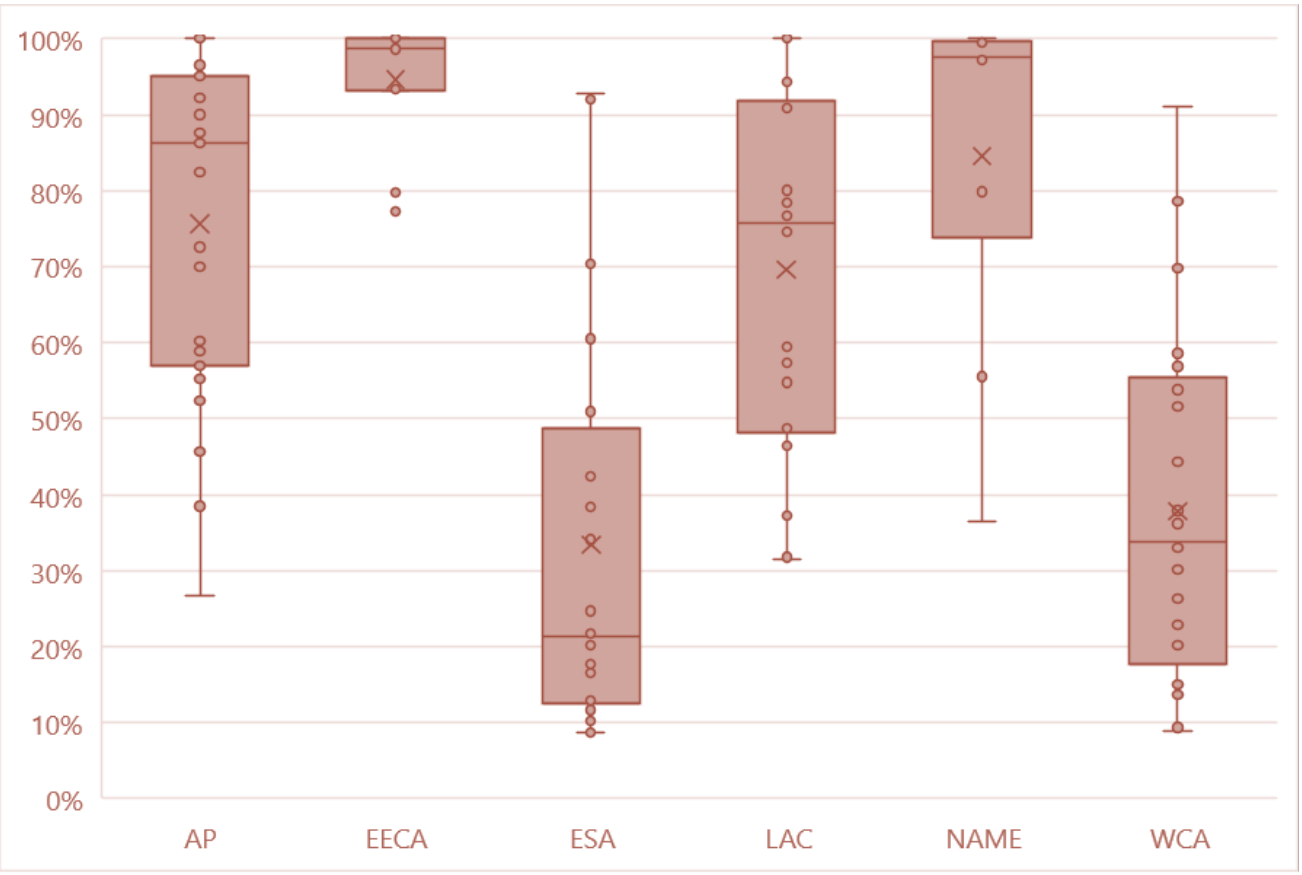


Figure 8.3. Percentage of new HIV infections occurring among Key Populations and their partners across various regions in 2023.

Figure 8.3 shows significant regional and within-region variation, which has increased due to the updated population size estimates and surveillance data. Some regions like EECA and NAME exhibit higher concentrations of new infections within key populations and their partners, while others like ESA display more variability.

Eline Korenromp presented updates on changes since the last JAIDS report (January 2024), including:

- **Temporal coverage expansion:** Shifted from 2 years (2010 & 2022) to 14 years (2010-2023).
- **Data summation:** Introduced population-weighted summation for all metrics instead of regional aggregates.
- **Input source flexibility:** All available sources stored, allowing users to switch preferred sources if multiple is available.
- **Data use:** Use of all data years from 2010 to 2023, with time trend imputation applied within each country and key population.
- **Uncertainty bounds:** Introduced uncertainty ranges for key results, summarizing new infections across key populations and by single key populations.

In the 2024 round, several countries transitioned to using the Goals model for key population size estimates, including Paraguay, Mexico, Ukraine, Jamaica, Croatia, Mongolia and Philippines amongst others. The primary reason for switching, is that Goals provided country-specific feedback and allowed for more accurate and comprehensive estimates across all key populations, which was a significant improvement over the previous models used.

Last year, a regional median was used for all countries, relying on data reported through the Global AIDS Monitoring (GAM) system, which was directly incorporated into Goals and used for extrapolations. However, this year, the preferred source for sub-Saharan Africa was a systematic review.

For countries not included in the review, other sources were considered. For example, if an Optima model was available and calibrated with country input, it was used for sex workers and MSM populations, except for transgender populations. The extrapolation for transgender groups, in cases where there is no direct data available, is based on the MSM estimates rather than using regional medians. Since the data for transgender populations is sparse, the method involves assuming about what proportion of the MSM population might be transgender. This approach avoids relying on the regional median for transgender estimates and instead derives the figures from the more robust MSM data. In cases where these sources were unavailable, the country's report to GAM was used, provided it was nationally representative and evaluated as adequate by UNAIDS. If none of these options were applicable, a regional median PSE was used as a last resort.

The resulting PSEs were expressed as percentages and assumed to remain constant over time due to insufficient evidence to suggest significant changes. Compared to the previous round, some key changes included a decrease in the proportion of MSM in Sub-Saharan Africa with the new source, while the proportion for MSM increased in Asia Pacific, Latin America, and the Middle East. These changes also affected the transgender population estimates, which were derived as a proportion of MSM estimates. Additionally, some high-income and Central Asia & Latin America/Caribbean countries used case-based TESSy data for MSM and PWID, while transgender estimates were derived from the ECDC/WHO-Euro 2023 surveillance report. Missing data within each dataset was interpolated using methods such as linear, moving average, or time-constant mean approaches.
Regarding diagnosis by mode of transmission, it is now possible to interpolate missing data, improving the accuracy of mode-specific estimates.

Next steps: The team plans to finalize the updates by August, with further discussions on how to communicate and present these findings in upcoming UNAIDS reports. Additionally, the same estimates of PSEs and new infections will be integrated into the SHIP tool, with training for the 2025 round of estimates anticipated.

Discussion

Key inquiries & recommendations:

1. Age bias in MSM surveys: It was acknowledged that MSM surveys often oversample younger individuals, which can lead to biased estimates when generalized to the entire MSM population. The model currently does not adjust for this age bias, but it was noted as a valid concern that could be addressed in future model iterations. The new Goals-ARM model aims to better match the age range of studies with the modelled age bands during the fitting process.

2. Age distribution in FSW estimates: The model currently uses the demographic shapes of populations to apply estimates, which can sometimes lead to overestimation of older female sex workers (FSWs). This was acknowledged as a potential issue, and it was suggested that further refinement is needed to accurately reflect the age distribution within the FSW population.

3. Geographic and demographic within-country variations – estimates: It was clarified that the model does apply percentage estimates uniformly across countries, but there is ongoing work to refine these estimates to better reflect geographic and demographic variations within countries, particularly in cases where urban-rural divides may lead to different HIV prevalence and key population sizes.

4. Uncertainty

- **Approach to uncertainty:** The discussion focused on the need to incorporate uncertainty bounds in the estimates of new HIV infections among key populations. The current method involves calculating the variability across countries within a region to infer uncertainty. This approach provides an interquartile range (IQR) or a 95% confidence interval (CI) to capture the variation in new infection estimates across different countries and populations.

- **Challenges with adding uncertainty bounds for individual populations**: While it is technically feasible to estimate uncertainty for specific key populations like MSM, there was concern that presenting these bounds separately might lead to misinterpretation or selective use of data by various stakeholders, such as NGOs or advocacy groups. The key issue is that summing the upper bounds across all key populations would result in an unrealistically high total estimate, while summing the lower bounds would result in an unrealistically low estimate.
- **Consideration of cross-country variation**: The current approach to estimating uncertainty by considering cross-country variation was seen as a pragmatic solution for the current round of estimates. However, it was recognized that this method primarily reflects variability rather than true statistical uncertainty, which would be more complex to calculate and interpret.
- **Long-term considerations**: There was a suggestion to explore more sophisticated statistical approaches, such as multiple runs of models that sample from the entire range of possible outcomes. However, this approach would be much more labour-intensive and is not feasible for the current round of estimates. Another option is to aggregate error as weighted averages of individual country standard errors.
- **Need for clarity in communication**: It was emphasized that any uncertainty estimates provided should be communicated clearly, ensuring that they do not mislead stakeholders about the precision of the estimates. The importance of framing the uncertainty as reflecting variability, rather than precise statistical confidence, was highlighted.

Session 9: Sub-national HIV estimates

The session's primary objective was outlined by **chair Cari van Schalkwyk**: **Review the SHIPP tool methods and make recommendations about its suitability and/or future development**.

Sub-national HIV estimates In Priority Populations (SHIPP) workbook tool

Katie Risher, **Penn State University**, presented on the SHIPP workbook tool.

Background

The process of generating HIV estimates at the national level is conducted through Spectrum, which aggregates data to provide an overview of the HIV epidemic within a country. These national estimates are disaggregated to the district level using the Naomi model, generating district-level estimates that are stratified by age and providing key HIV indicators such as prevalence, incidence, and the number of people living with HIV. While the Naomi model provides data at the district level, it does not account for variations in risky behaviour that can significantly impact HIV transmission dynamics. To address this gap, the SHIPP tool was developed as an extension of the Naomi model, to estimate populations at higher risk of HIV who require intensified HIV prevention programs. This estimation is achieved by disaggregating data by sex, age, and sexual behaviour at a sub-national level in 34 countries.

Methods

The SHIPP tool integrates data from multiple sources, including:

- **Population-based surveys:** Data on sexual behaviour is drawn from national population-based surveys, such as the Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS).
- **Naomi model outputs:** The SHIPP tool uses outputs from the Naomi model, including estimates of population size, HIV incidence, prevalence, and the number of new infections at the district level.

Key Population data

Data for key populations, including female sex workers (FSWs), men who have sex with men (MSM), and people who inject drugs (PWID), is incorporated from national workbooks.

- Sub-national disaggregation: KP data is disaggregated at the sub-national level using KP admin-1 level estimates.[6]
- HIV prevalence: Calculated using these sub-national estimates, providing a more accurate reflection of HIV prevalence among key populations.
- HIV incidence: Informed by national-level data from the Goals Model.[7]

6. **Stevens O, et al.**Population size, HIV prevalence, and antiretroviral therapy coverage among key populations in sub-Saharan Africa: collation and synthesis of survey data 2010–2023. *medRxiv.* 2022 Jul; doi: https://doi.org/10.1101/2022.07.27.22278071.

7: **Stover J, Glaubius R, Teng Y, Kelly S, Brown T, Hallett TB, et al.** Modeling the epidemiological impact of the UNAIDS 2025 targets to end AIDS as a public health threat by 2030. *PLoS Med.* 2021 Oct;18(10). doi: https://doi.org/10.1371/journal.pmed.1003831.

- Age disaggregation: key population sizes are disaggregated by 5-year age groups to better capture age-specific HIV risk within these populations.
  - For MSM and FSW: The SHIPP tool uses a combination of age at first sex data and Thembisa age distribution (Gamma distribution) to estimate age distribution.
  - For PWID: literature-based estimates.

Behavioural proportions and Population Size Estimation

- **Behavioural categories:** The SHIPP tool categorizes individuals based on their sexual behaviour, including those with no sexual activity in the past year, those with one regular partner, and those with non-regular partners. Key populations are categorized separately.
- **Geospatial modelling:** A space-time multinomial model is used to estimate the proportions of individuals in each behavioural category at the district level. This model accounts for spatial, age, and temporal variations in sexual behaviour.
- **Population Size Estimation:** The estimated proportions are multiplied by the Naomi model's population estimates to derive the population sizes for each behavioural category.

Estimation of HIV prevalence and incidence

- **HIV prevalence:** HIV prevalence is estimated by behavioural category using national-level log odds ratios from household survey data for general populations and KP-specific data for key populations.
- **HIV incidence:** The SHIPP tool maintains Naomi's incidence rate estimates at the district level and uses literature-based incidence rate ratios (IRRs) to estimate incidence rates by behaviour. This approach allows for the estimation of new HIV infections within each behavioural category.

The SHIPP tool provides estimates of:

- Behavioural proportions by sex, age, and district.
- HIV-negative population sizes by behaviour.
- Estimated new HIV infections and incidence rates by behaviour and district.

Priority development areas

The SHIPP tool has already been used by country teams, particularly focusing on women aged 15 to 29, excluding female sex workers indicated by size estimation surveys. Feedback on how countries used the tool, especially regarding sex work, is essential for future development. There is an ongoing review of the methodology and its application in different countries for programming, resourcing, and preparation purposes. The current version of the SHIPP tool has several updates compared to the one disseminated two years ago. The updates are documented in a paper under preparation.

The tool's estimates, particularly those related to KP sizes, behavioural data, and incidence rate ratios, have a high degree of uncertainty. This uncertainty is not fully captured in the current version of the tool, highlighting the need to:

- Incorporate empirical data to disaggregate 15-49 to 5-year age groups.
- Incorporate space-time-behaviour interactions in the small area estimation model for sexual behaviour.
- Varying IRRs for those with non-regular partner(s) relative to single regular partner.
- Improved visualization tools and streamlined outputs, particularly for mapping results at the district level.

Discussion

The key clarifications and discussion points were:

1. Application of the SHIPP tool across countries

The SHIPP tool has been developed and applied in 34 countries. The process for in-country validation is still being refined. The tool has been used in a few countries, particularly for Global Fund applications, where it helps estimate population sizes for proposals.

2. Integration with other models

The SHIPP tool draws from the Naomi and Goals models and is built on their foundations, so it is not entirely standalone, but users can interact with the tool independently.

3. Issues with Key Populations exhausting total new infections

The discussion also touched on the challenges of accurately estimating key population sizes and their integration into the SHIPP tool. One of the key issues raised was the variability in KP estimates across different sources and the difficulty in

ensuring these estimates are accurate and representative at the sub-national level. The tool's reliance on various data sources, such as the KP workbooks and Goals model, can lead to discrepancies, particularly in settings where data is sparse or inconsistent. Risher acknowledged that issues had arisen in previous applications of the SHIPP tool, particularly when KP estimates were too high relative to the total number of new infections in a district, leading to implausible results. Risher noted that the team is working to troubleshoot these problems with the new estimates being developed for the current year.

4. Handling challenges with sparse data and risks

There was also concern about the potential homogenization of data in countries where KP estimates are sparse. In such cases, the SHIPP tool may apply a uniform estimate across large geographic areas, potentially obscuring important regional variations. Participants discussed the need for more granular data and better modelling techniques that can account for these regional differences, even in data-poor settings.

5. Role of KP Workbooks

The role of KP workbooks in the SHIPP tool's estimation process was a significant topic of discussion. Some participants questioned the value of these workbooks, noting that in some cases, they did not add new or more accurate data compared to other sources like the Goals model. However, it was emphasized that KP workbooks serve an important function by providing nationally owned and adopted KP estimates, which are crucial for countries when submitting data for applications such as the Global Fund. This national ownership and validation process is essential for ensuring that the estimates used in tools like SHIPP are aligned with the figure's countries consider most accurate and representative.

6. Sensitivity of assumptions

Another critical point raised was the sensitivity of the assumptions used in the SHIPP tool, particularly regarding the prevalence and incidence ratios among key populations. For example, the assumption that a certain percentage of the male population engages in same-sex behaviour can significantly impact the estimates produced by the tool. The 1% threshold is based on national estimates that meet certain criteria. However, there are uncertainties, and the threshold may not perfectly represent the risk. Participants suggested that the SHIPP tool might benefit from sensitivity analyses that could explore the impact of these assumptions on the tool's outputs. This would help to identify potential biases or inaccuracies in the estimates and allow for more informed decision-making.

7. Ownership and validation of estimates

The final part of the discussion focused on the importance of ownership and validation of estimates by the countries themselves. It was stressed that for the SHIPP tool to be effective and widely adopted, countries need to feel confident in the estimates it produces. This requires a robust validation process, where countries are actively involved in reviewing and approving the estimates that will be used in national HIV prevention and treatment programs.

Naomi issues

Rachel Esra, Avenir Health Inc, provided an overview of the Naomi model, the new features planned for the upcoming round of estimates, and the challenges encountered in accurately modelling HIV-related data at the sub-national level.

Naomi model approach

- **Cross-sectional estimations:** uses a cross-sectional Bayesian small area estimation statistical model, smoothing sparse data at the district level by borrowing information from neighbouring areas.
- **Integration with Spectrum:** disaggregates national HIV projections from Spectrum down to the district level, incorporating program data like ANC surveillance and ART coverage
- **Short-term projections:** provides short-term projections used for planning and targeting, with a focus on the most recent survey data and future projections informed by Spectrum assumptions.

Challenges

1. Sub-national population inputs

- Naomi requires district-level population data by 5-year age and sex groups. Ideally, this comes from recent census data, but in many cases, it relies on publicly available gridded population datasets, from sources like GPW (Gridded Population of the World) or WorldPop.
- **Challenges with gridded population data:** These datasets often lead to inaccurate and inconsistent subnational estimates over time. For example, unusual population structures (e.g., a university district with a high student population) can distort PLHIV cohort migration in the model, leading to issues in projecting HIV prevalence and

treatment needs. Case studies from Ghana and Cameroon are presented in figure 9.1, illustrate extreme age/sex distributions, where PLHIV cohorts may exceed the total population denominator in the migrated age bands.

- **Solution:** The team is developing a beta version of the model to handle unusual population structures more effectively.

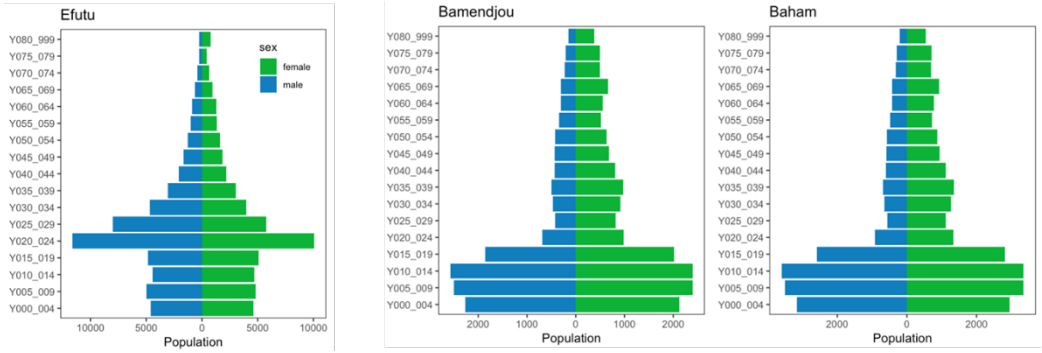


Figure 9.1: Extreme district-level age/sex distributions in Ghana (2021 census), left, and Cameroon (2020 WorldPop data), middle, right.

2. Patient mobility

- Naomi estimates the number of people receiving ART in a district based on where they reside, not where they receive treatment. This distinction is crucial for accurate ART coverage estimates.
- **Challenges in mobility modelling:** Patient mobility, particularly in areas affected by conflict (e.g., Tigray in Ethiopia, Cabo Delgado in Mozambique), complicates these estimates. Patients may seek treatment in neighbouring or distant regions, leading to discrepancies between reported treatment numbers and the resident PLHIV population in those areas. Figure 9.2. illustrates the challenge in Soroti, Uganda, where the model's estimate of HIV prevalence was significantly overstated due to unaccounted-for patient mobility.
- **Solution:** The team is working on solutions that better account for movement between regions, particularly in conflict-affected areas. For example, in Addis Ababa, the model adjusts for the influx of patients from neighbouring regions, which is critical for accurate ART coverage estimates.

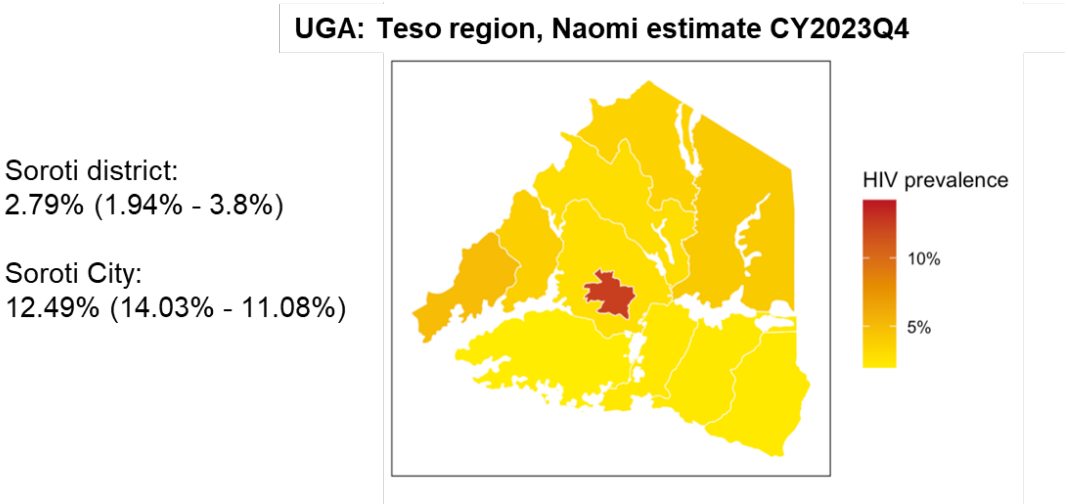


Figure 9.2: Patient mobility and HIV prevalence in Soroti City, Uganda, with disparity in HIV prevalence estimates between Soroti district (2.79%) and Soroti City (12.49%), highlighting the potential impact of patient mobility on the accuracy of sub-national estimates.

New features

- The Naomi model will now include the ability to download SHIPP tool outputs directly from the Naomi interface, ensuring alignment with recent Naomi estimates.

- **Improved Spectrum integration:** The team plans to enhance the Naomi-Spectrum integration by allowing users to compare subnational data with aggregated Spectrum data earlier in the process, reducing the need for re-calibration and improving the overall workflow.

Esra concluded by inviting feedback on the proposed improvements and encouraging ongoing collaboration with country teams to refine the Naomi model further.

Discussion

1. **Maintaining consistency and updating outputs**

- It was emphasized that using outdated estimates can lead to significant issues. The SHIPP tool must be updated to reflect the most current Naomi results to maintain consistency and reliability in the data used for planning and funding applications.
- Frequent requests for updated SHIPP tool outputs from country teams highlight the need for a clear consensus and communication plan regarding the availability of updates, the timing, and the round of estimates to which they will be linked.

2. **Revisions and enhancements**

- There is a need for further review on implementing the full SHIPP tool, with potential for peer review from other modelling experts. The tool requires refinement, particularly in handling data aggregation by age and spatial distribution.
- Utility for programs: Ideas were sought to make SHIPP tool estimates more actionable for HIV programs and related exercises, such as the Kenya epidemic appraisal. The goal is to ensure the tool's outputs are directly useful for strategic HIV prevention and treatment planning.

3. **Better integration**

- The SHIPP tool should integrate more effectively with existing models like Goals and Naomi to avoid discrepancies, particularly in Global Fund applications where differing estimates can cause issues.
- Routine updating: A recommendation was made to integrate the routine updating of SHIPP tool outputs into the Spectrum-Naomi estimates process, ensuring regular updates align with the latest data and methodologies.

4. **National vs. subnational estimates**

- The group noted that while national estimates are essential, subnational estimates often provide better information. The goal is to use the most specific data available, recognizing that the choice between national and subnational data may depend on the context and data availability.
- Where subnational data is not available, national estimates should be used, with an emphasis on using local behavioural and biological surveillance surveys (BBSS) to inform district-level needs.
- **Mitigating discrepancies:** There was a suggestion to highlight the importance of local BBSS data in informing these estimates, as national key population size estimates might not be as useful without subnational granularity.

5. **Population size estimates (PSE)**

The SHIPP tool currently uses a combination of Workbooks and the Goals model for population size estimates (PSEs), prioritizing Workbooks as the primary source and Goals as the secondary option.

- It was recommended that UNAIDS validate these Workbooks to ensure their reliability as the primary source of PSE data. The process of reviewing Workbooks and Goals data before updating SHIPP tool outputs should be streamlined and begin earlier in the timeline to allow sufficient time for review and integration into the estimates process.
- Addressing the urban-rural divide in PSEs remains a challenge. The discussion acknowledged the difficulty in achieving accurate data without clear urban-rural distinctions, and there was a suggestion to improve methodologies to better capture these dynamics.

Session 10: Sub-national EPP stratification

The session's primary focus, chaired by **Sharmistha Mishra**, was to address two key objectives: **reviewing the likelihood structure for replacing sub-national EPP (Estimation and Projection Package) stratification with a national-level approach, and discussing potential advancements in the EPP toolkit**.

Replacing sub-national EPP stratification: Proposed likelihood structure for weighting ANC data

Jeff Imai-Eaton led the discussion by presenting the challenges and considerations involved in moving from sub-national to national stratification within the EPP framework.

EPP – a back-calculation engine

EPP is a tool designed to back-calculate what HIV incidence has been based on trends in HIV prevalence and mortality, doing so in an epidemiologically principled way. To achieve this, two fundamental components are required: a trend in HIV prevalence within the population and assumptions about survival following seroconversion, along with the impacts of ART on survival. The process within EPP involves modelling the HIV transmission rate to represent a consistent epidemic prior to the availability of prevalence data. If a complete time series of prevalence data were available from the beginning of the epidemic, it would be possible to simply calculate backwards to determine the number of infections. However, since prevalence data typically begins partway through the epidemic, the transmission rate is used to infer the earlier stages of the epidemic.

Furthermore, EPP accounts for the impact of ART coverage on transmission to inform the recent trend in HIV incidence. This adjustment is necessary because recent prevalence data does not provide sufficient insight into recent incidence trends. Since it takes time to accumulate enough data, additional structure is imposed on the recent prevalence data to accurately model the epidemic's progression.

Why model subpopulations in EPP?

Historically, EPP has employed stratification by subpopulations, such as urban versus rural or by province, to achieve more accurate prevalence trends within different regions of a country. The rationale for this approach stems from the need to account for the varied dynamics of the epidemic across different geographic and demographic groups. Early antenatal clinic (ANC) sentinel surveillance data, which were often non-representative, made it necessary to fit separate epidemic models to each region. These regional fits were then weighted and combined to produce a national estimate.

In recent years, however, HIV prevalence data available through household surveys and routine ANC testing provides nationally representative trends, reducing the need for complex stratification that aims to correct for biases in earlier, less representative data.

Uganda's ANC prevalence data from 1990 to 2000 were used to illustrate the process of fitting separate epidemic models to urban and rural regions, then combining these models to create a national prevalence estimate.

The prevalence trends shown in figure 10.1 highlight the disparities between urban and rural areas. The urban areas showed higher prevalence rates, which is consistent with the greater concentration of the epidemic in these regions during the period studied. In the third panel below, the urban and rural prevalence data was combined to produce a national estimate. Figure 10.2 demonstrates the impact of sub-national stratification on the national incidence trends.

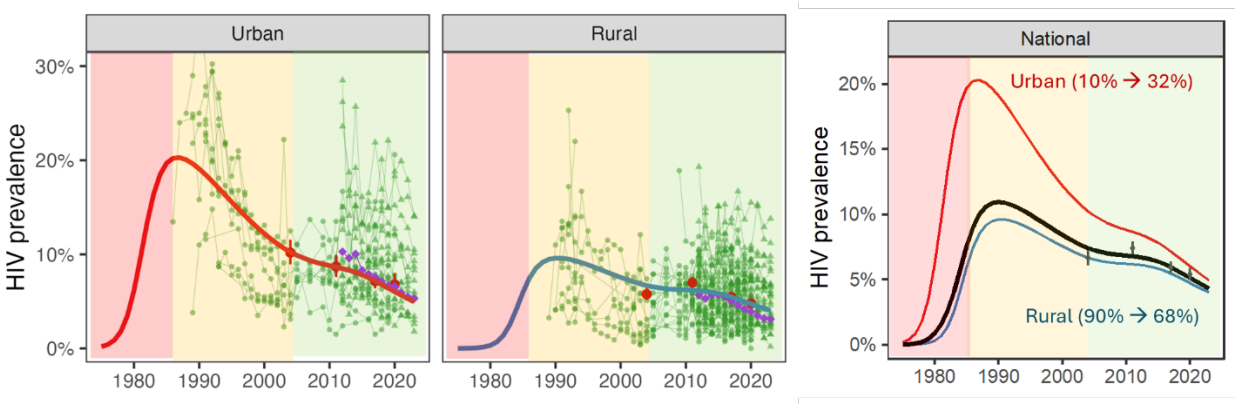


Figure 10.1: HIV prevalence curve fitted to Ugandan ANC data from 1990 to 2000, comparing 1. urban (red line) and **2.** rural (blue line) regions**. 3.** National prevalence is shown as a black line, which is a weighted average of the urban (red line) and rural (blue line) prevalence curves.

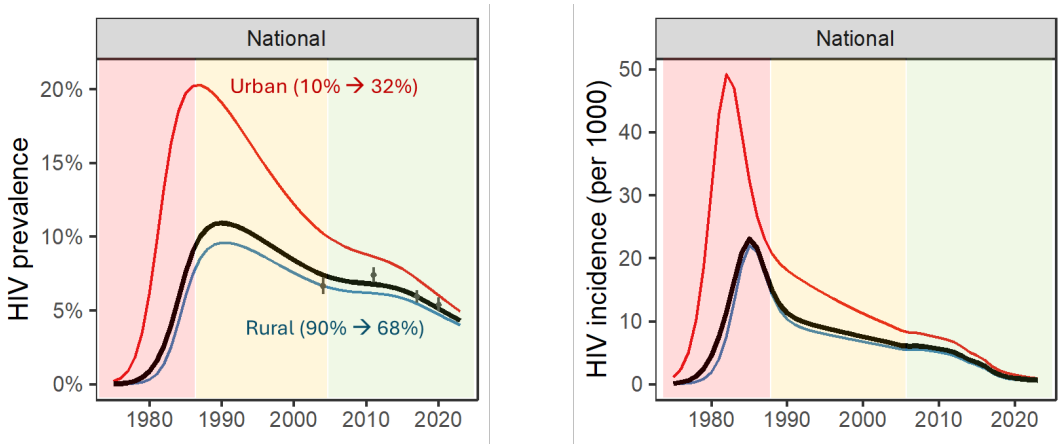


Figure 10.2: Weighted HIV incidence plot, derived from the combined Ugandan sub-national prevalence data. The black line represents the national weighted incidence, which is a combination of the contributions from urban (red line) and rural (blue line) regions.

The session highlighted several challenges associated with sub-national EPP fitting:

1. **Complexity in disaggregating data:** Disaggregating program data into urban and rural regions presents significant challenges. These stratifications often do not align well with existing health information systems, making it difficult to accurately allocate ART and other resources based on these divisions. The need to maintain a hidden sub-national population structure, originally created in 2017 using World Urbanization Prospects and sub-national PJNZ files, was emphasized. This structure requires updating to reflect current demographic trends and to reduce uncertainty in regional prevalence estimates.
2. **ART coverage and Attend vs. Reside:** One of the major challenges discussed was the allocation of ART coverage by urban/rural or other sub-regional divisions. ART coverage varies widely between regions, complicating the modelling process. Unlike the Naomi model, which effectively addresses the "attend vs. reside" issue - where individuals might receive ART in a different location from where they reside - EPP does not have a built-in mechanism to manage this discrepancy. This limitation of EPP further complicates the accuracy of sub-national stratifications and the allocation of resources based on these models.

Table 10.1 summarizes the different stratification methods employed by countries in Western, Central, Eastern, and Southern Africa in their EPP models, including urban/rural, subnational region, and single national stratifications.

Table 10.1: 2024 Generalised Epidemic EPP stratification.

	Urban / Rural	Subnational region	Single national
Western and central Africa (Total n = 19)	Burundi, Benin, Burkina Faso, Cen. Afr. Repub., Chad, Côte d'Ivoire, Cameroon, Dem. Rep. Congo, Congo, Gabon, Ghana, Guinea, Gambia, Guinea-Bissau, Liberia, Mali, Togo (n= 17)	Nigeria (37 states) (n = 1)	Equatorial Guinea (n = 1)
Eastern and southern Africa (Total n = 16)	Angola, Botswana, Eritrea, Rwanda, South Sudan, Uganda (n = 6)	Eswatini, Ethiopia†, Kenya*, Mozambique, Namibia, Tanzania, Zimbabwe* (n = 7)	Lesotho Malawi Zambia (n = 3)

* Kenya and Zimbabwe use province-level Spectrum PJNZ files
† Ethiopia uses province PJNZ files, further stratified by urban/rural within provinces

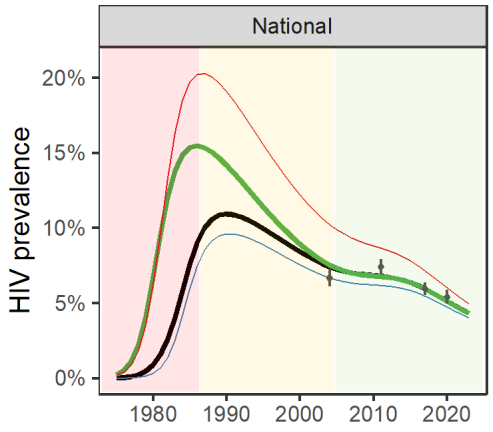
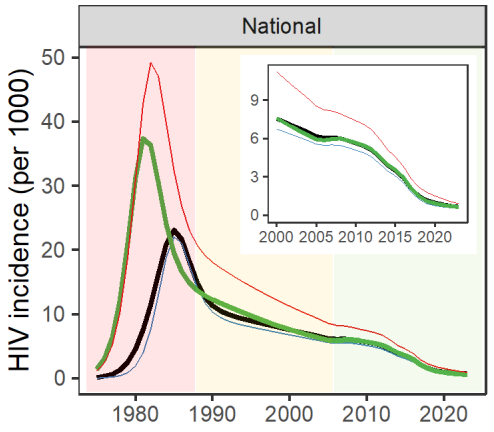
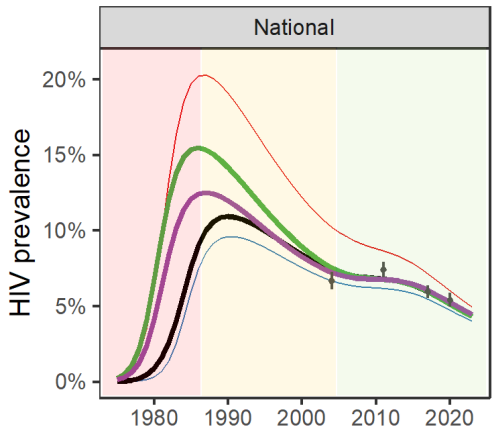
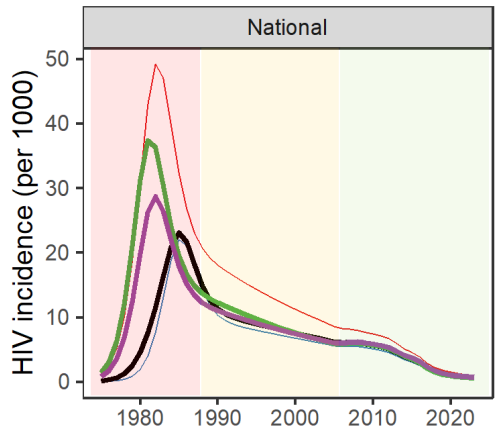
Statistical accounting for EPP subregion prevalence difference:

1. **Simple option:** This option involves fitting separate epidemic models to EPP subregions by combining all ANC site data into a national fit. Initially, urban and rural data are equally weighted before representative data is available, leading to a combined national fit. As representative data becomes available, the national fit and subpopulation trends converge, simplifying the modelling process.

2. **Proposed statistical adjustment:** This option adjusts the likelihood structure to account for systematic differences between urban and rural prevalence trends. It simulates a single national epidemic model and uses a linear regression model to disaggregate national prevalence trends while maintaining urban/rural weighting.

Table 10.2 summarizes the two approaches to modelling HIV prevalence and incidence in subregions, illustrating their key features and how they handle urban and rural ANC site data.

Table 10.2: Statistical accounting for EPP subregion prevalence difference.

Option	Plots
1. Simple option: Red curve: Urban ANC site data Blue curve: Rural ANC site data Black curve: national fit obtained by combining and weighting regional fits Green curve: All prevalence estimates are equally weighted, to obtain a combined national fit. During the period where representative data is available, the trends (green and black curves) converge, demonstrating the alignment between the combined national estimate and individual subpopulation trends as data becomes more comprehensive.	 
2. Proposed statistical adjustment: This plot demonstrates the statistical adjustment applied to account for differences between urban and rural prevalence trends during the period informed by only ANC data (1990-2000). The purple curve shows the adjusted national fit, which incorporates a linear regression model to capture systematic urban/rural differences.	 

Conclusions

The primary motivation for maintaining separate EPP subpopulations has been to account for non-representative HIV prevalence trend data, especially the disproportionate representation of ANC sentinel sites in urban areas where the epidemic was initially more concentrated. However, recent data, which is more representative of national populations, shows less variation between urban and rural areas compared to the early epidemic data. Since 2010, trends in HIV prevalence have been similar regardless of whether subpopulations or aggregated national data are used. While it is possible to partially adjust for different prevalence trends in early data, handling variations in trends before data collection began remains challenging.

Proposed recommendations

1. **Guidance on subpopulation Choices:** It is recommended to provide more detailed guidance on the rationale for subpopulation stratifications, particularly to adjust for historical differences in data availability.
2. **Encouraging national fit:** Countries should not be forced to adopt a single national fit but should be encouraged to consider it, especially if they are struggling with EPP inputs or aligning coverage. However, the use of EPP outputs for subnational epidemic information is discouraged, as Naomi is more suited for this purpose.
3. **Aggregation over stratification:** Further subnational stratification in EPP should be discouraged, and countries should be encouraged to consider more aggregation. For example, Nigeria might benefit from moving to seven zones instead of using 37 state files.
4. **Review of historical representativeness:** Consideration should be given to how much priority should be placed on refining statistical adjustments to account for historical representativeness in likelihood calculations. It is important to assess the value of accurately capturing the early epidemic in the aggregated national fit.
5. **Review of population inputs:** For countries continuing to use subnational stratifications, a review of sex/age population inputs is recommended to ensure consistency with Naomi district populations and to address issues with subnational PJNZ inputs.

Discussion

The key discussion points were:

- **Sub-national trends and Prevention of Mother-To-Child Transmission (PMTCT) indicators**: Although the Naomi model could replace the need for sub-national models for most indicators, Naomi does not produce trends over time or PMTCT indicators, which are important for advocacy and planning.
- **Challenges with sub-national PMTCT data**: Although these indicators are important for sub-national planning, there is large uncertainty in these estimates, driven by uncertainty in sub-national estimates of the population size, the total fertility rate, the prevalence of pregnant women, and whether women are receiving services in their own region or elsewhere. This uncertainty may be larger than the gap in care that the models are aiming to measure.
- **Data needs by region for second estimation option**: Imai-Eaton clarified that, for the second estimation option, which proposes a statistical adjustment in the likelihood approach, the requirement for ART data by region is eliminated. Instead, only population size by region is necessary for this method.
- **Changes in urban/rural classification over time**: Another strong reason to move away from the urban/rural stratification in EPP is that classification has changed in some countries since the start of HIV epidemics, and EPP has no way to handle such changes.
- **Changes in indicators other than prevalence and incidence**: Several participants suggested also considering changes to indicators such as deaths and number of orphans when considering the different methods of moving from sub-national to national models. Since the options only produce vastly different curves in the early epidemics (before surveillance data became available), it is likely that differences in these other indicators will be negligible for recent years.
- **Fitting one curve through a large 'cloud' of data**: Concern was expressed over the statistical appropriateness of fitting one national curve through many data points associated with very different contexts, potentially obscuring regional differences. Visualisations of the posterior trends at the site level could address this concern.

Session 11: ART coverage data discrepancies

The session's objectives were outlined by **chair Leigh Johnson**:

- **Plan further work related to ART coverage data discrepancies**
- **Develop guidance and examples for countries to apply**
- **Review current uncertainty on ART numbers and proposed adjustments**

ART adjustments in the 2024 HIV estimates

Rachel Esra /Ian Wanyeki

- Summary of country adjustments
- Rationale/ evidence supporting adjustment
- Summary of ongoing DQA activities

Background

Discrepancies between estimated ART coverage and reported numbers on treatment were identified, especially in countries with generalized epidemics, thus requiring adjusting of ART numbers.

ART coverage estimates are modelled in Spectrum:

$$\text{ART Coverage} = \frac{\text{ART programme data}}{\text{Modelled PLHIV [Prevalence} \times \text{Population size]}}$$

ART coverage estimates rely on several data inputs, including:

- Demographic projection models: Using population estimates from the World Population Prospects (WPP).
- Nationally representative household surveys
    - Demographic household survey (DHS): HIV prevalence
    - Population-based Impact Surveys (PHIA): HIV prevalence, incidence, and ART coverage.
- Routine antenatal data and historical sentinel surveillance: HIV prevalence among pregnant women.

The 2021 PHIA survey indicated high levels of ART coverage:

- 95% of individuals were aware of their HIV status.
- 98% were receiving ART.
- 98% of those on ART were virally suppressed.

Expected: Saturation effect at high levels of programme coverage

Countries nearing high ART coverage levels were expected to show a plateau in the rate of scale-up, indicating the exhaustion of eligible individuals. However, in several countries, this plateau was not observed, raising concerns about the accuracy of reported ART numbers. Countries such as Kenya, Zambia, Uganda, and Malawi showed steady ART scale-up without the expected saturation, indicating potential overcounting.

Factors contributing to ART overcounting

- Data infrastructure: Issues like lack of unique identifiers and siloed electronic medical records (EMRs), hinder the ability to accurately track patients across facilities.
- Inaccurate outcome reporting: Including failure to account for loss to follow-up (LTFU), facility transfers, or deaths.
- Artificial recordings: Instances of treatment collection visits being artificially recorded to maintain TX_CURR (treatment current) statistics.
- Re-initiation of active patients: Approximately 30% of patients in some contexts were found to be virally suppressed at initiation, driven by stigma, mobility, and targets set by implementing partners.

Methods to adjust for ART overcounting

- Cameroon: Adjustments were based on comparing HIV prevalence to ART coverage across different data sources, including DHS and PHIA.

- Zimbabwe: Used historical survey data to align ART coverage estimates with actual population data, addressing discrepancies in reported numbers.
- Namibia: The national DQA indicated differences between reported numbers and actual patient records, causing a 3% overreporting on ART. Preliminary 2022 census (3 million total pop) vs WPP projection (2.6 million), necessitated adjustment to population denominators.
- Tanzania: Applied survey data to adjust ART coverage and calculated an adjustment factor (~20% reduction in number on ART).

2024 Country-specific adjustments

Data Quality Assessments (DQA):

- Namibia: Adjusted ART coverage by 3% based on national DQA findings.
- Morocco: Adjusted ART coverage by 22%, considering discrepancies between reported numbers and patient files.
- Nigeria: Adjusted downwards by 6% based on DQA findings from PEPFAR and national assessments.
- Ghana: Adjusted ART coverage by 50% due to an algorithm bug in their DHS.

Data triangulation with historical household surveys and proxy data sources:

- Tanzania: Triangulated ART coverage data from the 2016 and 2022 PHIA surveys with historical data, aligned with proxy data from women entering antenatal care (ANC), 22% reduction.
- Uganda: Used Spectrum scenario modelling with various discounting factors to align with the 2020 PHIA survey data, 12% reduction.
- Malawi: Triangulated program data with ANC clients and STI clinic attendees, reduction of 5% for men and 10% for women.
- Haiti: The PHIA data implied lower ART coverage than Spectrum/survey/surveillance program data and demography. Used patient-level extracts to revise down the aggregated program ART numbers.

No proxy data available to inform exact adjustments needed:

- India: Adjustments were made for child ART coverage, which was above the estimated number of children living with HIV (CLHIV) due to double-counting and children aging out or not being removed from the ART database.
- Saudi Arabia: Heuristically fitted DQA factor (2009-2012) ensuring a plausible scale-up. MOH data cleaning efforts in 2014, giving an abrupt drop in unadjusted ART numbers, notably for men.
- Sierra Leone, Burundi, Mali, Niger, Burkina Faso: Adjustments were made based on the best available data.

ART adjustments observations

- Proxy data can provide a more accurate picture of ART coverage, particularly in settings where direct data may be unreliable or incomplete.
- The experience in Malawi highlighted the usefulness of commodities data in estimating ART coverage. This approach requires close collaboration with country teams familiar with commodities data extraction and interpretation to ensure accurate results. However, this method was not as successful in other countries such as Uganda, Zambia, and Sierra Leone.

Overall, the need for political will to transparently adjust reported numbers and improve data infrastructure was stressed.

Implications for sub-national estimates

ART numbers are a fixed model input in the Naomi framework, which affects subnational estimates of ART coverage and unmet need.

Proposals for calculating sub-national targets for unmet need using discounted ART data

1. If ART overreporting impacts both unmet need and number on ART:

- Use discounted subnational ART data for Naomi fit.
- Calculate % PLHIV untreated = untreated_plhiv_attend / plhiv_attend
- Calculate "programmatic" untreated = reported number on ART * % PLHIV untreated

2. If ART overreporting impacts only the number on ART and not the estimate of PLHIV eligible for treatment:

- Use discounted subnational ART data for Naomi fit.
- Add untreated_plhiv_attend to the reported number on ART.

ANC ART coverage analysis and impact

Jeff Imai-Eaton presented an update to determining ART coverage among pregnant women as a proxy for population ART coverage, comparing recent program data with survey estimates from 2020-2022, following earlier comparisons from 2015-2019.

Using ANC data as a proxy: HIV status is routinely recorded for all pregnant women during their first ANC visit, providing near real-time data on ART coverage. This data is crucial, especially in regions lacking regular household surveys, like much of Western and Central Africa.

Comparison Between Program Data and Survey Data

- The first round of PHIA surveys compared the total number of adults (15+) on ART, estimated from surveys, with ART program data reported in Spectrum files.
- In the earlier PHIA surveys (2015-2019), ART coverage estimates from program data generally aligned well with survey estimates, showing no systematic bias. However, in the more recent PHIA2 surveys (2020-2022), program data often indicated higher ART coverage than survey data, particularly in countries like Lesotho and Malawi, suggesting possible discrepancies.
- Differences in ART coverage between men and women, particularly in the 25–39 age group, remain consistent over time and space.
- Significant discrepancies were identified in West and Central Africa, where ANC data suggested substantially lower ART coverage than what was reported in program data. This discrepancy raised concerns about potential inaccuracies in ANC reporting or higher stigma leading to non-disclosure of ART use.
- In Ghana, after accounting for a 50% reduction in reported ART patients, an 8% predicted under-enumeration strongly suggests discrepancies in program data.
- Subnational data quality (DQ) assessments are a challenge, particularly where ANC data quality varies.
- Country-specific contexts, such as fertility levels and health system variations, require tailored adjustments. For example, data from countries like Eritrea and Mali highlight significant deviations due to local factors.

Secondary Data Analysis Concerns

- Using secondary data from ANC facilities raises concerns about duplication and data accuracy. Women often visit multiple facilities during pregnancy, complicating reliable ART coverage estimation.
- Unique ART identifiers in some countries mitigate duplication but remain inconsistent.

Strength of ANC Data as a Proxy

- ANC data provides an advantage by measuring both the numerator (ART patients) and denominator (HIV-positive individuals) within the same population, avoiding reliance on external estimates.
- Systematic biases, such as the younger age profile of pregnant women and gaps in ART coverage among men and young adults, may affect the representativeness of ANC-derived ART coverage as a proxy.

Recommendations

- Baseline viral load testing among pregnant women with HIV at select sentinel facilities. This approach would help determine the true ART coverage and identify gaps due to non-disclosure or reporting errors.
- Shift focus from trying to perfectly align population estimates with program data to accurately quantifying gaps in ART coverage. More reliance on proxies like ANC data could be key to monitoring and guiding program strategies effectively.

Systematic differences between the ANC population and the general population, such as age and HIV acquisition timing, could influence ART coverage estimates. Thus, while ANC data is a useful proxy, it should be part of a broader set of tools used for monitoring ART coverage.

Discussion

1. Deciding between using World Population Prospects (WPP) and national population data when there are discrepancies between the two.
The UN's World Population Prospects (WPP) generally applies rigorous demographic techniques, including adjustments for census undercounts, which might be more accurate than national estimates. However, WPP data can be outdated. If a country's statistical bureau has more recent data, it might be preferable. Countries should discuss with their statistical bureaus to decide which data best reflects their situation. A follow up comment noted the importance for the Group to work with the UN Population Division to understand any adjustments they make and ensure that these are considered at the country level.
2. Are there any countries where ART coverage estimates from survey-based populations and reported numbers align well? What can be learned from countries that are managing this well?
The issue is more about the EMR infrastructure rather than its availability. Even in countries with individual-level EMR data, such as Tanzania or Nigeria, there are problems with data linkage across facilities, leading to potential duplication in national aggregates. While there are problem cases, there are also countries with better results, although these are not the focus of the current discussion.
3. Were the ART adjustments applied at the Naomi level adopted into the official DHIS data or specific to Spectrum only?
The reductions are not reflected back into the data system. The data pulled needs to be adjusted, and the record of that adjustment exists in Spectrum and AIM. ART overreporting may not be consistent over space, especially at the district level, and currently, a national reduction factor is applied. There is insufficient data to inform subnational adjustments.
4. Clarification requested on whether the ART numbers reported are aggregated from individual-level data and if comparisons have been made with recalculated numbers from country teams.
Some countries, such as Zambia and Uganda, have attempted probabilistic matching exercises, yielding varied outcomes. In Uganda, the government chose not to use any presented options due to the differing results.
5. Application of a 5% reduction across all countries for ART adjustments.
The 5% reduction was not uniformly applied; it was a country-specific decision based on their best estimates, with some countries applying reductions of 10% or 15%.
6. Clarification sought on the method of setting targets in Naomi, specifically if using the discounted ART number to calculate unmet need might underestimate the target.
Two indicators were discussed: one for true unmet need (people with HIV not on treatment) and another for programmatic coverage. Using discounted ART numbers might make more sense, considering the overreporting in program data.
7. Comparison between PHIA surveys and program data regarding ART coverage.
The comparison shows no systematic bias in program data relative to PHIA surveys from 2015-2019. However, in the 2020-2022 PHIA surveys, most program data indicated more people on treatment than the surveys suggest.
8. Use of ANC data alongside DQAs to address uncertainties in subnational ART coverage estimates.
ANC data can inform subnational DQAs, but the quality of the ANC data must be considered. ANC clients are prone to double counting due to multiple visits, which must be accounted for in any analysis.
9. Use of ART coverage among pregnant women as a proxy for population ART coverage.
ART coverage among pregnant women before their first ANC visit is a strong proxy for monitoring population ART coverage, showing a high correlation with overall population ART coverage trends.
10. Status of viral load testing among pregnant women at delivery within PEPFAR-supported programs.
Viral load testing is not systematically done at delivery, but it is performed based on risk assessments. There is a need for consistent and comprehensive testing at the first ANC entry point.

The discussion highlighted that the focus should shift to supporting countries in improving the data going into the system. Continuing to apply methods without addressing underlying data quality issues could delegitimize the estimation process.

Recommendations
The Reference Group identified the following key areas for further action:

- UNAIDS should provide clear guidelines for countries lacking recent census, survey, or DQA plans to ensure transparency in adjustments made during the estimation process. The guidance should include methods for evaluating and improving data quality and examples of successful adjustments and the impact on ART coverage estimates.

- Flagging high levels of coverage: Number on ART > PLHIV warning mostly ignored in Spectrum and very hard to diagnose in Naomi. Stricter validation rules should be imposed to flag implausibly high ART coverage relative to

the estimated number of PLHIV. There was a consensus on the need for a fundamental change in how ART counts are handled in Spectrum. ART data should be treated as another calibration data source rather than an exact total to match.

- Ownership/transparency in adjustments that happen during file review. There was consensus on the need for a sustainable approach where countries can make these decisions themselves in future.
    - It was suggested to include an option in the ART editor where countries can enter an adjustment to account for double counting. The ART editor could feature a button that pops up a new screen, guiding users through potential adjustments, such as those based on ANC data or common adjustments found through PQAs.
    - Allow countries to see multiple country data in Spectrum. This transparency would enable countries to compare their data trends with others and identify common patterns and discrepancies. Overlaying indicators from multiple countries would help in synthesizing data and providing a clearer picture of ART coverage trends.

- Sub-national targets: Consensus on calculation of subnational targets for unmet need using discounted ART data.

Methods to interrogate and adjust ART programme data in Malawi

Andreas Jahn /Stone Mbiriyawanda

In his presentation, **Andreas Jahn** provided an analysis of Malawi's ongoing efforts to interrogate and adjust ART programme data. The interrogation of Malawi's ART programme data was prompted by the observation of implausibly high ART coverage in the 2024 HIV estimates. This trend was especially noticeable in the Spectrum and Naomi models, where ART coverage growth did not show the expected deceleration as the TX_CURR (current on treatment) numbers approached saturation. Instead, there was a significant divergence between ART coverage trends at the population level and those observed among antenatal women and STI patients.

Key findings

1. **Divergent ART coverage trends**

- **Population vs. antenatal and STI data:** While population-level data continued to show a steep increase in TX_CURR, the ART coverage among antenatal women and STI patients exhibited a distinct levelling off, suggesting potential overreporting in the programme data.

2. **Investigating possible explanations**

- **Population projections:** The demographic projections, including fertility rates and census adjustments, were reviewed by experts, concluding that these were not the source of the discrepancies.
- **Subnational population distribution:** The possibility of errors in subnational population distribution was also considered but ultimately dismissed as a primary factor.
- **HIV prevalence estimates:** The age and sex-specific HIV prevalence estimates were consistent with two PHIA surveys conducted in Malawi, indicating that the model estimates were accurate.
- **ART programme data:** The primary concern was the potential overreporting of TX_CURR, with several contributing factors identified:
    - **Foreign nationals:** While there is some evidence that foreign nationals may be counted in Malawi's ART data, this would need to be a significant number to impact national estimates.
    - **Malawians in South Africa:** There are reports of Malawians living in South Africa receiving ART from Malawian clinics, which could contribute to inflated TX_CURR numbers.
    - **Overreporting of TX_CURR:** Documented incidents of overreporting have been noted, motivated by the pressure to meet targets and possibly due to drug theft.

3. **Examples of overreporting**

- **Facility-level trends:** Several facilities were identified where sudden drops in TX_CURR were observed, often following data cleaning exercises or investigations. This suggests previous overreporting that went unnoticed for years.
- **District-level analysis:** Some districts showed unexpectedly high and continuous growth in ART coverage, which was not consistent with expected trends based on population data.

4. **Triangulation methods**

- **ARV consumption gap:** By comparing the ARVs issued to facilities with the estimated consumption based on reported TX_CURR, significant gaps were identified, suggesting overreporting.
- **Cohort growth and Loss to Follow-Up:** A slowing in ART cohort growth was expected due to high coverage, but some sites showed a net decline in loss to follow-up, which may indicate overreporting.
- **Viral load monitoring coverage:** Low viral load monitoring coverage at certain sites was another indicator of potential overreporting.

Challenges and solutions

1. **EMR data analysis**

- **Data quality issues:** The analysis of visit-level EMR data revealed several challenges, including back-entered visits and systematic errors, making it difficult to identify genuine visits. However, some facilities were flagged for further investigation based on these indicators.

2. **Triangulation findings**

- **ARV supply gap:** This was one of the most informative indicators, showing an increasing gap between ARVs supplied and those needed to maintain reported TX_CURR, with a 5.7% gap identified in 2023.
- **Selected sites for DQA:** A total of 129 sites were selected for targeted DQA visits, based on triangulated data from various sources.

Conclusions and recommendations

1. **Adjustment of ART data**
   A national adjustment was made based on the ARV supply gap, subtracting 5.7% from the 2023 TX_CURR, equivalent to 59,000 patients. District-level adjustments were also made using a modified Naomi model.
2. **Root causes of overreporting:**
   The main drivers of overreporting were identified as ambitious and unattainable output targets, difficulty in auditing programme data, and a general lack of incentives to question positive results.
3. **Proposed actions**
   - **Redefining target-based implementation:** Targets should be used as a neutral planning tool rather than for punitive measures.
   - **Incentivizing accurate reporting:** It is crucial to move away from payment systems that encourage overreporting and to emphasize the importance of accurate data across the programme management cascade.
   - **Data validation tools:** There is a need to design more efficient tools for data validation, ensuring that extreme results are questioned and triangulated with other data sources.

While the margin of adjustment may seem small, Jahn emphasized the importance of incentivizing honesty in data reporting and rewarding genuine data cleaning efforts to maintain the integrity of ART programme data.

Discussion

A summary of the key discussion points:

1. **Were revised ART numbers back corrected in EMR, and how often would correction be done?**
   Data cleaning was carried out retrospectively back to 2020, and the two district hospitals have been able to maintain more plausible reporting. It is hoped that a one-off large-scale cleaning exercise would instil a sense of data integrity and that a promising strategy would be to communicate that there are no specific ART targets that need to be met.
2. **Inclusion of ANC data in upcoming DQA**
   ANC data will not be included, as it has shown very consistent trends. There are no specific ART targets, which is thought to be the main driver of distortions in ART program reporting.
3. **Systematic misclassification of known positives on ART being reported as new positives and new treatment initiations**
   There's potential for systematic misreporting and biases in data tools, although ANC data in Malawi has not been significantly affected, there is an ongoing viral load baseline survey among new positives to investigate potential misclassification.

4. **For countries that may not have the capacity to replicate Malawi's work, what changes do you suggest for EMR systems to make them more practical for data validation once aggregated?**
   Jahn advocated for improved data collection in supply chain management as patient data can be triangulated with ARV supply data. He also pointed out the issues with back-entered visits in EMR systems and the need for better control over dispensing data through a pharmacy module, integrated into the point of care EMR. Biometric identification is unfeasible at a national level in Malawi, and a disproportionate investment is required for implementation.

ART dashboard and validation process in Kenya

Morris Ogero, National Syndemic Diseases Control Council (NSDCC), presented on the ART dashboard and validation process in Kenya.

The ART dashboard integrates data from various sources (Kenya Health Information System, MOH 731 forms, National Data Warehouse) and includes tools for validating data at various levels. It gives an overview of key metrics related to ART services, disaggregated at the national, regional, county, sub-county, ward, and facility levels, includes an alert system that notifies users and can generate reports. User access is controlled to ensure data security and confidentiality.

The ART verification process in Kenya is focused on ensuring accurate data through cross-checking patient records, deduplication, and validating National Unique Patient Identifier (NUPI) that are created for PLHIV.

While no ART adjustments are made, there are several key issues:

- Transcription errors were found in the KHIS data, but ART numbers remained consistent across different platforms.
- Significant service quality gaps were identified in ASAL counties.
- The 2023 data quality audit revealed discrepancies in viral load data, unreliable nutrition data, and incongruities among key population (KP) registers, indicating a need for improved data management.
- Despite 94% of NUPIs being traced:
    - 11% of PLHIV lack a NUPI, especially in remote areas.
    - 3% missing age data, 0.7% in multiple facilities.
    - 6% of IDs not traced in the national system.
    - Cross-border patients receive alternative IDs like passport numbers.

Proposed refined method to estimate uncertainty on Spectrum country-level ART numbers

John Stover presented a refined method to estimate uncertainty in Spectrum country-level ART (antiretroviral therapy) numbers. The current assumption in the AIDS Impact Model (AIM) is that program data on the number of people on ART has an uncertainty range of -12% to +4%, based on early Data Quality Audits (DQAs).

The presentation discussed whether these bounds should be revised and proposed methods to determine a new range.

Three approaches to estimating uncertainty

1. **Review latest DQA studies:** Analyse the most recent DQA findings to understand the current state of data accuracy and identify potential sources of error.
2. **Compare Spectrum with PHIA:** Compare ART coverage data from the Spectrum model with Population-based HIV Impact Assessment (PHIA) surveys to identify discrepancies and calculate uncertainty.
3. **Examine historical program data changes:** Evaluate changes in ART program data over time by comparing historical data to current estimates to detect patterns and deviations.

ART coverage among adult men and women (2024): Comparison of AIM and PHIA data

- Significant differences were observed, with PHIA generally reporting higher ART coverage than AIM.
- Median differences: -8% for men and -12% for women; mean differences: -21% for men and -27% for women.
- The largest differences were noted in older surveys from countries such as Cameroon (2017), Malawi (2015-16), Uganda (2016-17), and Zambia (2016).

Historical program data changes

ART program data from 2020, comparing estimates from the 2021 and 2024 Spectrum files.

The percent differences in ART estimates highlighted substantial variability, suggesting that historical data changes can inform the uncertainty range.
Stover summarized the key points and posed two critical questions for the UNAIDS Reference Group:

- Should the uncertainty range for program ART data be changed?
- If so, what is the best approach to determine a new range?

Discussion

Addressing large uncertainty in ART coverage

- Identify a set of countries with significant uncertainty in ART coverage data. Develop strategies to quickly gather reliable data to provide accurate ART coverage estimates.
- Emphasize the importance of viral load testing and other indicators to validate ART coverage and address data discrepancies.
- Explore the application of ANC data adjustments in countries where ANC data has shown to be reliable. Validate the impact of these adjustments through detailed reviews and comparisons with program data, especially in countries with known discrepancies.
- **Validation for CSAVR countries**
- Conduct a thorough validation of the adjustments applied to CSAVR countries, assessing the consistency and reliability of the adjustments and ensuring they align with the local epidemiological context.
- Organize a focused meeting on Concentrated Epidemics, possibly in September, to review these adjustments and their implications.

Session 12: Age distribution of new infections

The session's objectives were outlined by **chair John Stover**:

- **Review update to generalized epidemic default age incidence rate ratio pattern**
- **Review evidence on incidence age pattern shifting to older ages**
- **Consider default pattern and model specification for time-varying age incidence rate ratios**

Oli Stevens summarized the recommendations from the recent Reference Group meeting on HIV incidence in sub-Saharan Africa. [8]

1. Large HIV incidence declines recorded in Eastern and Southern Africa
    - **Findings:** Since 2010, cohort studies have consistently recorded significant declines in HIV incidence in Eastern and Southern Africa. This trend contrasts with incidence rates found in the control arms of HIV prevention trials, which remain systematically higher.
    - **Considerations:** There remains substantial spatial heterogeneity and regions and sub-populations with higher risk still require focused attention and intervention.
    - **Implications:** Messaging around "low" HIV incidence should be contextualized within the broader landscape of risk and spatial heterogeneity. In areas with concentrated epidemics, messaging should emphasize the continuing risk and the importance of targeted prevention efforts.

2. Model alignment with prospective cohort data
    - **Findings:** Models used for estimating HIV incidence are generally well aligned with data from prospective cohort studies.

3. Empirical evidence on time-varying incidence rate ratios
    - **Aging of the epidemic:** Evidence from cohort studies consistently shows the HIV epidemic aging, with faster declines in incidence among younger age groups (15-24 years) compared to older groups (25+ years).
        - Cohort studies from various sites, including AHRI in South Africa, KEMRI in Kenya, RCCS in Uganda, and Manicaland in Zimbabwe, support this finding.
        - Incidence rates decline by 3% slower per year in the 25+ age group compared to the 15-24 age group.
    - **Spectrum vs. PHIA surveys:** While Spectrum estimates assume a constant proportion of new infections among the 15-24 age group, PHIA surveys show different trends, with the share of incidence rising among 15–24-year-olds.
        - **Possible explanation:** The role of false recency rates in older populations with high prevalence needs further investigation.

A summary of the key discussion points is listed below:

- **Prevalence trends in younger age groups:** Observation of prevalence trends indicating lower prevalence in the 15-19 age group but higher prevalence in the 20-24 age group raised concerns about potential spikes in prevalence as individuals leave school or enter relationships. Emphasis on the importance of focused programming for the 20-24 age group to address these spikes.
- **Impact of perinatal transmission programs on age-specific prevalence:** The perinatal transmission prevention programs have likely contributed to the reduction in prevalence among the 15-19 age group. The higher prevalence in the 20-24 age group could be due to the period before these individuals received the benefits of perinatal transmission prevention, necessitating further study.
- **Possibility of adjusting models to reflect changes in the age distribution of new infections:** The current stable assumptions do not capture the evolving dynamics of the epidemic. Future models should incorporate time-varying parameters.
- **Detecting trends in PHIA data despite limited cases:** Although individual survey data points have high uncertainty, pooling data across surveys can reduce uncertainty and the aggregated data provides a clearer picture of incidence patterns.
- **Enrolment practices can lead to underrepresentation of younger age groups,** which could skew the analysis. Adjustments for this issue were mentioned as a necessary step in data analysis.

8.https://epidem.org/wp-content/uploads/2024/06/UNAIDS-Reference-Group-meeting-on-HIV-incidence-in-sub-Saharan-Africa-Report.pdf

Rob Glaubius, Avenir Health Inc, gave an update to the generalized epidemic default age incidence rate ratio pattern. Default age incidence rate ratios: Update the generalized epidemic pattern considering recent incidence and serology surveys

- Simultaneous fitting to survey prevalence, survey ART coverage, and programme ART by age

Spectrum handles patterns of incidence by age and sex

- **Background:** Spectrum currently uses a default generalized epidemic pattern to disaggregate input HIV incidence trends by sex and age, based on HIV incidence by age from ALPHA Network sites, published in 2014.
- **Current use:** Few countries extensively use this default pattern; many prefer custom or survey-based patterns.
- **Proposed updates:** The need for a new default pattern is being considered, with proposals to fit incidence by age and sex to HIV prevalence in countries with adequate data and to evaluate potential proxy countries for those lacking sufficient data.

Fitting IRRs to multiple data sources

- To improve accuracy, Spectrum supports fitting incidence rate ratios (IRRs) to multiple data sources, including HIV prevalence in household surveys, numbers on ART from program data, and ART coverage by age and sex from household surveys.
- **Model fitting:** The IRR model uses a lognormal distribution to fit age and sex patterns. Bayesian maximum a posteriori estimation is used for model fitting, with priors derived from ALPHA Network incidence estimates.

Simultaneous fitting to survey and program data

- **Scenarios examined:** The fitting scenarios include combinations of survey HIV, survey ART, and program ART data to evaluate the best approach for accurate estimates.
- **Challenges and findings:** Fitting to multiple sources can sometimes pull the model away from observed HIV prevalence, and ART program data can dominate other sources, leading to significant deviations.
- **Implications:** Survey data are preferred for fitting incidence patterns due to better alignment with observed HIV prevalence. Using survey ART coverage data alone is not recommended, but it can enhance the model when combined with other sources.

Table 12.1. Country case studies simultaneously fitting to survey-based HIV prevalence, survey-based ART coverage, and program ART data.

Country	Data	Findings
Botswana	4 HIV surveys, 2 ART coverage surveys, 2 years of ART program data.	Adjusted ART program data showed minimal misalignment. Using combined data sources improved incidence estimates.
Cameroon	4 HIV surveys, 1 ART coverage survey, 6 years of ART program data.	Adjusted ART program data showed good alignment, demonstrating the value of combining survey and program data.
Eswatini	4 HIV surveys, 2 ART coverage surveys, 7 years of ART program data.	Fitting to combined data sources provided a comprehensive view of incidence patterns, with minor adjustments needed for program data.
Rwanda	4 HIV surveys, 1 ART coverage survey, 5 years of ART program data.	Consistent results from combined data sources, with program data requiring careful validation to avoid misalignment.
Tanzania	5 HIV surveys, 2 ART coverage surveys, 6 years of ART program data.	Adjustments to ART program data were necessary, but combining multiple sources led to robust incidence estimates.
Zambia	6 HIV surveys, 2 ART coverage surveys, 5 years of ART program data.	Combining data sources provided stable and accurate incidence estimates, with program data aligning well after adjustments.

Model for time-varying incidence rate ratios and default pattern

Jeff Imai-Eaton/Rob Glaubius, **Avenir Health Inc,**

Rob Glaubius assessed the feasibility of fitting IRRs simultaneously to multiple data sources:

- HIV prevalence from surveys,
- Survey-based ART coverage,
- ART programme data by age.

Key insights:

- **Programme ART data** tend to dominate the fitting process and distort estimates away from observed HIV prevalence.
- **Survey ART coverage** data alone are not reliable for IRR fitting and provide little added value when combined with other sources.
- Simultaneous fitting results in unstable outcomes and should be applied only with strong justification and careful review.

Validation issues are exacerbated when input data across tools (e.g., AIM vs EPP) are inconsistent due to survey timing or disaggregated entry issues.

Time-varying IRRs

Time-varying IRRs in AIM currently interpolate between survey years. This often results in abrupt changes in incidence patterns that are artefactual rather than epidemiologically plausible. Comparisons with Goals ASM demonstrated more coherent ageing of incidence over time.

A proposal was made to develop a simplified, mechanistic IRR model inspired by Goals ASM that includes:

- Age mixing,
- Age-specific acquisition risk,
- HIV prevalence and ART coverage trends,
- ART effect (e.g., viral suppression).

This structure avoids reliance on survey timing and may provide a more realistic and stable way to represent time-varying incidence.

Static IRRs	Dynamic IRRs
Age-specific IRRs remain constant over time.	Age-specific IRRs are specified for each survey.
Uses six parameters in total.	Uses six parameters per survey. IRRs vary linearly between surveys.

Discussion

- **Survey-based treatment coverage adds little to IRR fitting:** Such as in Ethiopia. In urban surveys, misrepresentation of rural dynamics may still be a concern and should be reviewed country by country.
- **Full implementation of age-mixing and dynamic IRR models in Spectrum is a longer-term goal:** In the interim, the group supported defining an average ageing pattern, similar to how default sex ratios were developed.

- **Time-varying IRRs in Tanzania produced unrealistic fluctuations:** These were linked to survey years, whereas Goals ASM provided smoother and more epidemiologically plausible results. However, it was clarified that while aggregate outputs (e.g., total PLHIV or new infections) may be similar, age-specific indicators differ, justifying the need to improve age pattern representation in Spectrum.
- **Auto-selecting survey data for IRR fitting is technically feasible:** However, some countries with good data still revert to the default pattern, possibly due to internal preferences or concerns over survey reliability.
- **Inconsistencies between modules in survey year and age groupings were flagged:** Especially where surveys span across years or use different age brackets (15–49 vs five-year groups). A need for harmonization of data inputs and synchronization between AIM, EPP, and Naomi was highlighted.
- **Full age mixing matrix in AIM/EPP not supported at this time:** It was considered too large a model change. Instead, participants supported revisiting prior work (e.g., 2020 incidence pattern comparisons) to derive a static ageing pattern based on multiple models, including Goals ASM and others.
- **ART programme data often distort estimates when combined with other sources:** The analysis showed that programme data can overwhelm survey inputs, pulling Spectrum's incidence and prevalence estimates away from observed values. Participants expressed discomfort, noting that over-reporting and duplication biases are common. Several pointed out that deduplication efforts in some countries could reveal age-specific inflation, helping assess the reliability of programme data.

Appendix A – Recommendations

Recommendation	Lead person(s)	Timeline
Session 1: Advanced HIV disease **Objectives:** • Validate Spectrum estimates of AHD prevalence (defined as CD4 < 200/mL) • Identify needs for future AHD estimates from modelled results and surveillance data sources		
Spectrum AHD outputs: Overall, Spectrum and Shiny90 model results for the percentage of CD4 <200 were consistent with available PHIA survey, GAM, IeDEA and systematic review data on CD4 <200 among untreated PLHIV (PHIA surveys; African countries) and among those diagnosed or initiating ART (GAM, IeDEA, WHO systematic review). Based on this, the Reference Group recommended to add new Spectrum outputs for estimates of Advanced HIV Disease (AHD): • Report outputs for AHD both among all people living with HIV and AHD at ART initiation (relatable to clinical programme data). • Stratify Spectrum outputs showing estimated numbers and proportions with AHD in the following sub-groups: unaware of HIV positive status, Know status but not on ART, On ART but not VLS, and on ART and VLS. • *Provisional recommendation:* Add PLHIV untreated with CD4 200-249 to 'AHD population' to approximately reflect those with AHD-defining clinical conditions, but CD4 >200, consistent with previous practice for ART eligibility CD4 200. ○ Provisional pending review of impact on Spectrum AHD estimates compared to available data. PHIA survey data indicate that around 4% of PLHIV who are virally suppressed have CD4 <200. This represents a large proportion (~half) of all PLHIV with CD4 <200, but clinically those VL suppressed are perceived to have low risk of opportunistic infections; this may vary for different conditions. Therefore, recommended to (1) include those with VLS in modelled AHD outputs, but (2) stratify reported AHD estimates among those on ART by VL suppression status.	Avenir Health	Oct 2024 Reference Group meeting

Spectrum AHD data input and validation:
• Add input editors for users to record routine programme data on AHD that are currently reported to GAM:
  ○ CD4 distribution among new diagnoses and ART initiators
  ○ Cryptococcal meningitis treatment cascade
• Add validation plots to compare routine programme data on AHD and cryptococcal meningitis with Spectrum outputs.
	Avenir Health	Oct 2024 Reference Group meeting

Spectrum model structure and assumptions:
• Among those on ART, Spectrum only represents CD4 at ART initiation, not current CD4 category. To produce modelled estimates for AHD among those on ART, review data to develop model approximation for:
  ○ Proportion with CD4 <200 among those on ART with unsuppressed VL, by duration on ART (<6 months, 6-12 months, 12+ months)
  ○ Proportion with CD4 <200 among PLHIV on ART with suppressed VL
• Spectrum model outputs indicated a rapid increase in recent years in percentage with CD4 <200 among the 'diagnosed but untreated population' (largely treatment interrupters), which was inconsistent with data from PHIA surveys. This may be an artefact of moving interrupters only one CD4 category higher, i.e. someone who initiated at CD4 of <100 will be in 100-200 when interrupting (i.e. still defined as AHD) regardless of duration on ART before interruption. *Recommendation:*
	TBD
	
	TBD
		Oct 2024 Reference Group meeting
		
		Oct 2024 Reference Group meeting

Recommendation	Lead person(s)	Timeline
○ Review evidence on CD4 recovery on ART and CD4 decline following treatment interruption ○ Assess whether alternative assumptions about CD4 recovery are more consistent with data on AHD among diagnosed but untreated		
• The Shiny90 sub-model is used to produce estimates for CD4 <200 stratified by undiagnosed and diagnosed but untreated for countries in sub-Saharan Africa. Confirm that Spectrum AIM simulation and Shiny90 model simulation produce internally consistent outputs for number with CD4 <200.	Avenir Health	Oct 2024 Reference Group meeting
• Extend evaluation to concentrated epidemic countries: confirm CSAVR and Spectrum-AIM results for CD4 <200 are internally consistent and review CSAVR outputs for CD4 <200 at diagnosis vs. available data.	Avenir Health	Oct 2024 Reference Group meeting
<u>Further data triangulation and validation:</u> • Review available data from new PEPFAR disaggregation of TX_NEW indicator by CD4 at initiation. ○ Compare with Spectrum outputs ○ Assess evidence of association between CD4 testing coverage and advanced HIV disease, indicating potential selection bias in those receiving CD4 testing in settings with incomplete AHD testing coverage	Oli Stevens/ Sara Herbst	May 2025
• Review data on CD4 and VL measurement at ART initiation to assess AHD among those virally suppressed at ART initiation ○ Baseline CD4 and VL measurement collected in Zambia surveillance ○ VL suppression at ART initiation likely largely reflect clients who are on ART and silent transferring	Jake Pry, CIDRZ, Zambia	May 2025
Session 2: Causes of death **Objectives:** • Review implementation of CM and TB death estimates in Spectrum, within the envelope of Spectrum-estimated HIV-related deaths		
<u>Tuberculosis:</u> Comparison of WHO-published estimates of TB-HIV deaths and UNAIDS-published estimates of AIDS-related deaths showed unexpectedly high or low proportions of all AIDS deaths due to TB in some countries. The Reference Group recommended to not yet implement output for TB-HIV deaths as proportion of AIDS-related deaths in the next round of Spectrum estimates, pending further review of estimates. Form a Working Group with WHO TB team to review unexpectedly (high and low) implausible proportions of WHO estimated TB-HIV deaths relative to UNAIDS estimated AIDS deaths) in some country-years. • Review consistency of AIDS deaths estimates with surveillance data in countries where TB estimation data perceived strong; examples: Colombia, Peru, Cambodia, Myanmar Review implications of any methodological changes to TB estimation methods result from September 2024 WHO Task Force meeting. Working group activities will commence after these potential changes have been determined. Compare case notification data on HIV prevalence and ART coverage among TB cases, used by WHO TB model, to UNAIDS/Spectrum modelled HIV estimates results across countries. WHO TB-HIV mortality model currently uses these case notification data to determine how TB case fatality ratios (not receiving ART or receiving ART) are applied to the HIV population. Considerations for potential development of TB-HIV deaths modelling strategy:	Working Group	Meetings to commence after the Task Force meeting in September Next feedback to the Reference Group in May 2025

Recommendation	Lead person(s)	Timeline
• Higher rates of TB testing among PLHIV than HIV-negative people may bias HIV prevalence among TB notifications • Request for gender stratification in TB-HIV death estimates • Update the review of TB case fatality ratios (CFR) among people living with HIV • Consider using Spectrum model outputs for duration on ART (by country and calendar year) as input to TB model for applying stratified CFRs • Update the systematic review of the proportion of HIV deaths that are due to TB. Previous two systematic reviews were from the pre-ART-for-all era.		
Review evidence and consider modelling strategies for whether assumed national mortality rates among people with HIV in Spectrum should vary according to national TB burden	TBD	May 2025
<u>Cryptococcal meningitis (CM):</u> *Spectrum model outputs:* Implement the University of Minnesota model in Spectrum, updated using Spectrum internal calculations as inputs for numbers with CD4 <200 by HIV care cascade stages • Report CM model estimate outputs for: o Cryptococcal meningitis cases among PLHIV o Number of CM-HIV deaths o CM-HIV deaths as a proportion of AIDS deaths o Disaggregate cases and deaths results by HIV care cascade stages (unknown HIV status, aware but untreated, on ART). *Spectrum model inputs:* • Create CM programme data input editor in Spectrum for users to input annual data on: o Coverage of screening of PLHIV for CrAg, prevalence of CrfAg and treatment with fluconazole preventive therapy if diagnosed (the cascade piloted in 2024 GAM – in numbers and %s) o • Where possible, prepopulate input editor with 2024 GAM-reported CM screening coverage, CrAg prevalence and treatment coverage. For CrAg prevalence, a second default for countries not reporting this to GAM 2024, can be the value used in UMN's analysis. • In validation plots, compare CM cascade programme data and modelled CM cases *Further model validation:* • Validate CM results for Botswana and South Africa from work resulting from the Ambition trial in Botswana, and the NICD in South Africa. • Review alignment of CM indicators specified in GAM, PEPFAR MER, and UNAIDS 2030 targets.	Avenir Health TBD	Oct 2024 Reference Group meeting May 2025
Session 3: Mortality among PLHIV on ART **Objectives:** • Validate Spectrum estimates of on-ART mortality against various sources • Present new approach to Spectrum mortality structure given incomplete CD4 data in most IeDEA cohorts		
Comparison of ART mortality rates recorded through PEPFAR TX_ML indicator vs Spectrum estimates for mortality on ART: • Across countries, there was wide variation in mortality rates reported through PEPFAR routine TX_ML indicators, ranging between 50% to 80% lower than Spectrum modelled estimates. This likely reflects variation in completeness of death ascertainment in clinical records, rather than providing evidence about overall mortality rates. • *Review further evidence on sex ratios in ART mortality*. Across countries, PEPFAR TX_ML reporting consistently implied higher male-to-female mortality ratios than Spectrum on-ART estimates.	Sara Herbst, Jeff Imai-Eaton, Reshma Kassanjee	May 2025

Recommendation	Lead person(s)	Timeline
• *Review ART mortality rates among young adults 15-24 years old from IeDEA and other sources*. In PEPFAR data, mortality increases by age, while in Spectrum mortality is higher among 15–24-year-olds than among 25-29-year-olds.		
Review new ART patient tracing studies, including routine programme tracing results that may not be published in academic journals. *Example:* Ethiopia cohort review.	UNAIDS	Feedback by May 2025
IeDEA analyses: • Compare all-cause mortality on ART (aggregating over age/sex/CD4) between IeDEA and Spectrum (by region) • Compare mortality among ART clients recorded through IeDEA surveillance to mortality recorded through routine PEPFAR reporting at the same health facilities.	Avenir Health / Reshma Kassanjee	May 2025
Analyse mortality among ART patients from clinical records in countries with strong electronic medical record systems. Example countries: Kenya, Zimbabwe, Rwanda, Botswana, Eswatini, Zambia	TBD	May 2025
Compare excess mortality among people living with HIV from population HIV cohort studies to Spectrum (AHRI, South Africa; Rakai, Uganda; Siaya, Kenya; Manicaland, Zimbabwe)	TBD	May 2025
Compare vital registration AIDS mortality recorded in Botswana to Spectrum.	TBD	May 2025
Review Mozambique and Sierra Leone sample vital registration data for AIDS deaths or deaths among PLHIV	TBD	May 2025
Session 4: Excess mortality among PLHIV **Objective:** • Review method to separate AIDS vs non-AIDS deaths within AIM's IeDEA-based all-cause mortality rates		
Implement in Spectrum a non-AIDS excess mortality rate among people living with HIV by sex / age / CD4 category. The non-AIDS excess mortality will apply to all PLHIV irrespective of ART status. For all regions outside sub-Saharan Africa, establish default parameters for non-AIDS excess mortality based on proposed method, fitting to the percentage of non-AIDS deaths from Trickey *et al.* review in high-income countries. Adjust AIDS-related mortality rates on ART, such that total all-cause mortality among PLHIV remains consistent with IeDEA-based all-cause mortality rates on ART between AIDS. Retain standard AIM outputs at all-cause deaths among PLHIV and AIDS-related deaths. Add an additional output reporting the modelled excess mortality among PLHIV stratified by AIDS-related and non-AIDS deaths. *Impact of this change is anticipated to be largest in high-income countries, with high ART coverage and low mortality rates among those on ART. Impacts are anticipated to be modest in other global regions where AIDS-related mortality rates are higher among those on ART.*	Avenir Health	2025 estimates
Review assumptions of derived excess non-AIDS mortality rates to address unexpected pattern that more excess deaths are due to non-AIDS causes at lower CD4 categories.	Avenir Health	Oct 2024 Reference Group meeting
Recommend not to apply excess non-AIDS mortality in Africa region, due to lack of data quantifying non-AIDS excess mortality among PLHIV in this region, and uncertainty about generalizing patterns derived from high-income countries to more generalised epidemic patterns in sub-Saharan Africa.		

Recommendation	Lead person(s)	Timeline
Data identified in Trickey *et al.* review were exclusively from high-income settings, there is low confidence in generalizing these patterns to other regions, and priority to identify additional mortality data for country and region-specific recommendations for all non-high-income regions.		
Session 5: Treatment Interruption **Objectives:** • Review outcomes from 2024 estimates • Provide recommendations on guidance and definitions for calculating Spectrum interruption inputs		
No change to Spectrum default treatment interruption rates, on the basis that (1) national ART programs are not able to accurately measure rates of treatment interruption (different from programme observed loss to follow-up) and (2) there is not a clear relationship between treatment interruption rates and LTFU rates.		

Reinforce to country teams in Spectrum user guidance, how treatment interruption differs from that for LTFU.
• **Lost to follow-up:** Not seen at service delivery site >28 days after missed appointment
  • Excludes individuals recorded as deceased or transferred
  • Includes those with unknown outcomes—including non-ascertained deaths, silent transfers
• **Treatment interruption:** 'Clinically meaningful' gap in treatment
 	
Users should not use LTFU measured in routine programme reporting as Spectrum's treatment interruption input (including the PEPFAR IIT [interruption in treatment] indicator, which actually measures LTFU and not interruption). It is anticipated that true treatment interruption rates are substantially lower than reported LTFU / PEPFAR IIT rates, given high rates of silent transfer.		

Session 6: World Population Prospects 2024 update **Objective:** • Review impact of new population estimates on HIV estimates		
Do a preliminary Spectrum analysis with updated demographic inputs to determine countries most likely to have substantial changes in HIV estimates due to changes in demographic input parameters. Then provide support to countries to anticipate and understand changes in HIV estimates resulting from large shifts in population and/or births. Review subnational demographic inputs (fertility rates, mortality rates, population structure, and net migration patterns) for all countries using sub-national Spectrum files.	UNAIDS	2025 estimates
Session 7: Dynamical modelling of HIV trends in key populations in sub-Saharan Africa **Objective:** • Review proposal for technical working group process and composition • Review updates to Goals-ARM		
<u>Key populations modelling Technical Working Group (TWG)</u> TWG to provide expertise and input from intersecting and overlapping (community, national programme experts, strategic information experts) perspectives on the representation of key populations in the Goals-ARM model being developed, prior to its piloting in 2025. 1) TWG: Engage experts beyond epidemiologic modellers to experts in key population behavioural science and epidemiology. 2) TWG: Establish terms of reference for the TWG defining expectations, roles, responsibilities for members of the TWG. 3) Community engagement composition: consultative approach recommended engaging via UNAIDS Community Leadership and leveraging partnerships with community partners experienced in working with modellers of HIV in key populations. 4) Expertise in HIV key population programmes and programmatic data collection: consultative approach recommended via leveraging partnerships with program experts engaged and/or interested in key population-modelling.		

Recommendation	Lead person(s)	Timeline
5) Country Pilot: two sub-Saharan African countries in the 2025 round, selected for comprehensive and high-quality key population data. Tag mini-workshop to end of December 2024 Estimates Workshop, following the training on Goals-RSM, on Day 5 of the main workshop with all countries participating. The scope of the TWG and above community and program engagement will focus on seeking inputs and validation related to the generation of key population indicators, for which Goals-RSM is currently being used in sub-Saharan Africa. The Key Population modelling TWG will not evaluate or make recommendations about application of Goals-ARM as an alternative to EPP/Spectrum-AIM for producing full national epidemiologic estimates.		
<u>Goals-ARM development</u> Reconciling epidemic growth rates during the early epidemic period: • Prevailing modelled estimates for early epidemic growth rates are heavily dependent on model assumptions and are informed by limited surveillance data. Therefore, while aiming for Goals-Arm calibrations to roughly reproduce AIM/EPP-estimated historic HIV trends, the AIM/EPP estimates for early epidemic growth rates should not be considered as a 'gold standard' for evaluating accuracy of Goals-ARM. Early epidemic growth in Goals-ARM results should be evaluated according to epidemiologic plausibility and triangulation with data, rather than based on discrepancy with existing results for the early epidemic period. • STI co-factors: o Although STI co-factor effects facilitated model alignment (with EPP/Spectrum) for the early epidemic growth rate, there was some concern that allowing the co-factor strengths to vary in calibration might lead to implausible values, "absorbing" other factors not accounted for. o Could add another calibrated factor to absorb these residual 'unexplained' country-specific drivers of transmission. o The scarcity of STI data for key populations and uncertainty on plausible cofactor values separately for symptomatic STI (syndromes) versus any STI infection, were also noted. • Transmission during primary HIV infection: Ensure that the primary infection transmission rate ratio is sufficiently high– and not absorbed in the STI cofactor. Initial peaks in primary infections, combined with typical low condom usage and high STI rates are important for reconciling rapid growth rates in early epidemics. Key population data likelihood and calibration: • Approach to calibrating key population epidemic: Given the lack of nationally representative key population prevalence levels and trends and heterogeneity in key population surveys, consider calibrating to ratios of prevalence between key population and the general population, adjusted by age and location. • Likelihood definition: Consider likelihood approaches (e.g. random effect models) that account for heterogeneity across key population samples rather than assuming key population surveillance data are nationally representative. (example in https://doi.org/10.1136%2Fsti.2009.037341) South Africa case-study recommendations: • Surveillance data suggest minimal heterosexual HIV transmission in South Africa before the late 1980s. Consider seeding the SA epidemic around 1985, to better align with early growth estimated by Thembisa • Consider calibrating to ANC data similar to Thembisa. • Investigate why Goals-ARM estimated all-cause mortality among PLHIV increases rapidly after the most recent year of deaths data used in model calibration.		
Session 8: New infections among key population and their partners **Objective:** • Review methods of estimating uncertainty in estimates		
It is important for dissemination of distribution of new infections to clearly communicate that there is substantial uncertainty in current estimates at country, regional, and global level. Formal statistical uncertainty ranges are difficult to quantify in the multi-source method used to derive estimates across countries. For 2024 UNAIDS update of distribution of infection estimates, use variation across countries within a region in the percent of new infections in each key population as a heuristic representation of the amount of variation in these estimates. Recommend against interpreting variation across countries as a measure of statistical uncertainty ranges. Also caution against adding up upper-bounds or lower-bounds across key populations, which would not reflect the correlated uncertainty across countries and populations.	UNAIDS	2024 update

Recommendation	Lead person(s)	Timeline
Session 9: Sub-national HIV estimates in sub-Saharan Africa **Objective:** • Review the SHIPP tool methods and make recommendations about its suitability and/or future development		
SHIPP tool: The group reviewed the methods and workflow of the Subnational HIV Estimate in Priority Populations (SHIPP) tool, a tool to disaggregate district/sex/age results for population size and new infections from Naomi to risk populations (not sexually active, single cohabiting/marital partner, non-regular sexual partner) to support country programme planning aligned to the Global AIDS Strategy 2021-2026. There were concerns about disaggregation of key population sizes and new infections by district, sex, and age group, considering the extremely sparse stratified data informing these stratifications. Therefore, the group recommended the SHIPP tool be revised to **only report national-level key population disaggregation**, and key population results should not be reported at the district level. Further development suggestions: • Review assumptions and additional available data for disaggregating key population estimates by age. • Investigate sensitivity of results when distributing men who have sex with men population size estimate only among sexually active men, vs. classifying some into the not sexually active group. Reiterate need for new data sources to guide subnational key population size and epidemiologic estimates, particularly use of key population programme data.	Katie Risher	Oct 2024 Reference Group meeting

Key population size estimates
• *Provisional recommendation:* If a consensus national key population size estimate was reported by country team in the key population workbook submitted with their Spectrum file, use this as the consensus PSE for other estimates and modelling purposes.
  ○ Recommendation is pending validating and finalising remaining pending queries on key population 2022 and 2023 workbooks submitted to UNAIDS.
• In the absence of an updated (or reconfirmed old) workbook, use PSEs determined by UNAIDS and used in 2024 Goals calibrations for 2024 regional key population estimates. These were determined according to a hierarchy:
  ○ For SSA countries, the Stevens et al. *Lancet Glob Health* 2024 synthesis, and
  ○ For 6 smaller concentrated-epidemic countries in SSA, used latest EPP estimate or an imputed (ESA or WCA) regional median.
• To enable logical workflow and alignment across tools (Spectrum, Goals, SHIPP) and results (national, subnational and UNAIDS regional KP estimates), *the key population workbook updating with country teams should be completed as a separate process before the Spectrum estimation starts (recommended timeline September to November)*
	UNAIDS	September 2024

Naomi • District population estimates: Display district-level population pyramids to support review and validation of inaccurate district population inputs • Implement validation comparison of Spectrum and Naomi ART and ANC programme data inputs to alert users to discrepancies in district-level and national programme data reports • Commission development of Microsoft Excel dashboard tool to enable more interactive review of district HIV estimates	Avenir Health	Oct 2024 Reference Group meeting May 2025
Session 10: Sub-national EPP stratification **Objective:** • Review likelihood structure and make recommendations about removing sub-national (including urban/rural) EPP stratification.		

Recommendation	Lead person(s)	Timeline

• Provide improved guidance to users about rationale for EPP subpopulation stratification:
  ○ Clarify primary rationale: adjusting for historical non-representativeness of surveillance data
  ○ This has lesser importance for recent estimates in current era, thanks to more nationally representative surveillance (household surveys, routine ANC testing)
• *Provisional recommendation:* Encourage users with urban/rural population structures to consider switching to a consolidated single national EPP region.
• Confirmation of this recommendation is provisional on further review impact on changes to estimates in pre-data period on outputs of interest, e.g., incidence, prevalence, AIDS deaths, number of orphans, and children living with HIV
	UNAIDS
	
	Avenir Health
		2025 estimates
		
		Oct 2024 Reference Group meeting

• For countries with sub-national stratification of EPP fits or subnational Spectrum files, share further information about the rationale for more aggregated regions in EPP to enable more precise trend estimates, but allow users to continue to choose regional or national stratification to EPP fitting ○ There is perceived high importance of region-specific results from EPP and Spectrum which are not available from Naomi, namely HIV incidence trends, PMTCT indicators, and AIDS-related deaths estimates. • Further study accuracy of applying EPP at regional versus national level according to statistical goodness of fit criteria, such as out-of-sample predictions to guide recommendations and user decisions about level of stratification and confidence in subnational trend estimates from EPP.	UNAIDS	2025 estimates
• For users who retain subnational Spectrum PJNZ files, provide improved guidance and tools for reallocating ART, ANC, and PMTCT programme data to align numerators and denominators and ensure accurate service coverage, service, need and programme impact estimates. • Communicate to users the intrinsic challenge that uncertainty in numerators (with some women attending ANC outside from the province they live in) and denominators (e.g., sub-national fertility rates, sub-national population estimates) is often greater than the margin of the service coverage gaps that Spectrum seeks to inform.	UNAIDS	2025 estimates
Session 11: ART coverage data discrepancies **Objectives:** • Plan further work to resolve or reduce ART coverage data discrepancies • Develop guidance and examples for countries to improve validation and data quality improvement of ART results • Review current uncertainty on ART numbers and propose adjustments		
<u>Improved guidance to users</u> • Develop guidance for countries on common reporting discrepancies and reasons for adjustments to programme data. • Collate and provide users with cases studies of ART programme data adjustments inputted to Spectrum estimates, including strategies and data sources for determining magnitude of adjustments. • Establish default adjustment options in Spectrum editor. As a default specify adjusted ART coverage that is halfway between ART coverage predicted from routine ANC data and reported ART programme data • Encourage detailed triangulation of facility ART reporting data with national supply chain and dispensing data as part of data review process preceding HIV estimates updating. This data review process will require more time and preparation before estimates process. • Discourage ART adjustments focused solely on resolving logically inconsistent estimates (such as >100% coverage for the latest estimated year.	UNAIDS	2025 estimates

Recommendation	Lead person(s)	Timeline
<u>Spectrum Program Statistics inputs editor and validation visualisations</u> • Place the ART adjustment multipliers in the main Spectrum (adult and child) ART editors, instead of pop-out editor, to give more prominence to the adjustment tool and set expectation this may be used • Add visualizations of how national ART scale-up compares with trends in other countries, to assist countries to identify and address implausible or otherwise abnormal entries • Communicate in user guidance that errors and duplication in routine data are expected and therefore a non-zero adjustment is often appropriate. • Move the comparison of ART coverage by ART program data versus ANC/PMTCT data (% of PMTCT initiated before current pregnancy) into the Program Statistics ART editor [currently displayed in Validation editor] • Add visualization(s) of the effect of alternative adjustments, e.g. aligning with ANC/PMTCT-based coverage, or halfway between ANC-based adjustment and unadjusted.	Avenir Health	Oct 2024 Reference Group meeting
<u>Recommendations for national ANC data</u> Accurate ascertainment ART usage prior to first ANC visit are critical data for triangulating population ART coverage. Accurate reporting of routine ANC data is weak in some countries, particularly in West and Central Africa where data also imply larger discrepancies with prevailing national ART coverage estimates from ART programme data. Reporting in these countries may also be susceptible to higher rates of non-disclosure of ART use related to higher stigma and services access challenges (e.g. stock outs). *Recommendation:* UNAIDS and partners should support national programmes to review and ensure accurate reporting of routine ANC indicators. *Recommendation:* To rapidly address uncertainties about population ART coverage and routine ANC reporting, particularly in Western and Central Africa countries, support countries to conduct viral load testing among pregnant women newly diagnosed or 'known positive' with HIV at first ANC presentation at a representative sample of ANC clinics. • May also consider tenofovir urine assay to assess ARV usage (easier, less expensive, directly measures whether on ART).	UNAIDS UNAIDS	2025 estimates September 2024
<u>Use of routine ANC HIV testing data in concentrated epidemic settings</u> Consider the current use and UNAIDS guidance on routine ANC HIV testing data to triangulate HIV prevalence and ART coverage in concentrated HIV epidemic settings. Potential uses of routine ANC testing and prevalence data: • Triangulating with CSAVR-estimated historic HIV prevalence and ART coverage trends, to select among alternative CSAVR incidence curves. • Fitting relative fertility among PLHIV via the local adjustment factor, to estimate PMTCT (and paediatric ART) coverage. • Validating incidence sex ratios (from CSAVR, concentrated-epidemic defaults, or fitted to ART by age data) by comparing ART coverage of pregnant versus all adult women. Recommended analyses: • Coverage of ANC HIV testing, out of total births and ANC clients • Consistency over recent years in annual ANC testing volume, and positivity • HIV prevalence among all adult women vs. routine ANC HIV prevalence, adjusted for testing coverage • ART coverage among all adult women vs. pregnant women attending routine ANC	Jeff Imai-Eaton, Maggie Walters, Eline Korenromp	May 2025
<u>Uncertainty in ART numbers</u>	Avenir Health	May 2025

Recommendation	Lead person(s)	Timeline
• Recommend not change to the current approach of proportional uncertainty ranges on ART numbers based on pre-2019 DQAs • For countries with direct estimates of ART coverage from household surveys (e.g. PHIAs), compare survey confidence intervals and current Spectrum uncertainty ranges to assess whether Spectrum ART coverage uncertainty ranges are systematically wider than survey uncertainty ranges ○ This may imply a revised approach, to obtain narrower ranges in countries with a recent PHIA survey whose ART coverage estimate was used in EPP fitting. • Investigate impact of approaches to expressing uncertainty in the percentage ART coverage rather than in the absolute number on ART, such that uncertainty in ART numbers and in PLHIV become related.		
<u>District level ART coverage, unmet need, and target setting</u> Develop guidance for how to propagate national ART adjustments to district ART coverage and unmet need in Naomi model estimates Develop guidance for users on how to specify district ART targets in settings where ART programme data are adjusted downwards for HIV estimates, resulting in difference in ART data recorded in DHIS reporting and ART totals implied in modelled ART coverage numerators.	Rachel Esra PEPFAR/GHSD, UNAIDS, Avenir Health	2025 estimates
Session 12: Age distribution of new infections in generalized epidemics **Objectives:** • Review the need to update generalized epidemic default age incidence rate ratios • Review evidence on incidence shifting to older ages • Consider Spectrum default patterns and specification for time-varying age incidence rate ratios		
Recent empirical HIV incidence data from Eastern and Southern African settings and mathematical modelling demonstrate moderately strong evidence that the age distribution of HIV infections is increasing as incidence and prevalence among young adults decline. In contrast, default parameters in Spectrum used by most countries assume static age pattern of HIV incidence over the course of the epidemic. It was recommended not to implement a full age mixing matrix in EPP-ASM or Spectrum-AIM. While this approach could capture underlying transmission dynamics underpinning aging incidence, it would be a large model change; the Reference Group instead recommends to rather focus efforts on the full transmission dynamic model Goals-ARM. As an interim approach, recommended to review incidence patterns by age from empirical data and mathematical models and calculate an average aging pattern for generalized epidemics to serve as updated default incidence rate ratios including ageing of the epidemic. Specifically, review: • Fitted HIV prevalence residuals over time to inform empirical evidence for an average change over time in data • Mathematical model comparison results on age pattern of incidence from October 2020 incidence patterns review (see October 2020 UNAIDS Reference Group meeting)	Avenir Health, Jeff Imai-Eaton, Oliver Stevens	Oct 2024 Reference Group meeting
Do not recommend update of generalised epidemic default age incidence rate ratio pattern. Most countries with a generalized epidemic calibrate country-specific incidence rate ratios use – or should use - household survey HIV prevalence; only very few should still need the default pattern. Provide examples, user guidance, and user support to identify and resolve instances where alternative data sources (survey HIV prevalence, survey ART coverage, ART programme data by age) imply inconsistent sex/age patterns in HIV prevalence or people living with HIV. • Add validation plots illustrating comparisons of survey-based versus program-reported adult ART numbers	Avenir Health	Oct 2024 Reference Group meeting

Recommendation	Lead person(s)	Timeline
• Consolidate validation plots for multiple data sources about prevalence and PLHIV by age, in the IRR fitting tool		

Appendix B – Feedback from Working Groups – Session 3

1. Can we use IeDEA data differently? (such as not disaggregate mortality by CD4)
2. How can we supplement/replace with other data?

Working Group	Response
1	**Calibration using IeDEA data:** Use IeDEA as a calibration target by aggregating baseline CD4 counts to validate Spectrum outputs rather than as input. **Sensitivity/variant analysis of IeDEA data:** Address assumptions on CD4 counts among patients missing CD4 tests and differentiate between LTFU vs. treatment interruptions, considering effects on VLS and CD4. **Triangulate data sources:** Use ART cohort studies with outcome tracing (e.g., Ethiopia) and high-quality (sub-) national EMR person-level ART outcomes. Consider tracing and cause of death from verbal autopsy and AI algorithms, and link patient records to civil registration.
2	**Comparison with PEPFAR:** Compare raw IeDEA data with PEPFAR MER data (TX_CURR, deaths, LTFU) from the same sites to identify discrepancies. **Additional data sources:** Search for more tracing studies, analyse treatment interruptions by age in MER, examine mortality data from DSS sites, compare IeDEA mortality with data from countries with high-quality mortality data (e.g., Brazil), and assess total mortality within IeDEA data.
3	**Update and review tracing studies:** Incorporate additional insights from updated tracing studies. **Triangulate data at aggregate level:** Compare long-term ART mortality data by age and sex using IeDEA, PEPFAR, and other country sources (e.g., South Africa, some Asian countries). **Alternative indicators:** Develop models using indicators other than CD4, such as VLS, and hybridize historical CD4 data with current VLS metrics.
4	**Tracing studies:** Consolidate all existing tracing studies, including non-IeDEA sources. Encourage new studies to include additional indicators (e.g., age/sex strata) capturing regimen shifts. **Other data sources:** Use EMRs (Rwanda), VR data (Botswana), sample/death registration data (Africa CDC), HDSS/cohorts, and insurance/banking data on all-cause mortality. **Spectrum/PEPFAR comparison:** Drop <90% coverage countries and perform finer disaggregation by AIDS/non-AIDS mortality. Some countries will have cause of death data.
5	**Environmental scan:** Conduct scans for countries with line lists, assessing coverage and tracing efforts. Evaluate variability in LTFU and treatment interruptions by age, gender, urban/rural areas, and Key Populations. Ask SIAs/UCOs about awareness of tracing efforts. **Community-led monitoring:** Explore community-led monitoring to understand retention and treatment interruptions, and cross-validate mortality patterns by age against VLS data from surveys. **Systematic review:** Perform a systematic review of tracing studies and individual meta-analysis.
6	**Leverage local systems:** Focus on sentinel sites that can report data centrally via EMR surveillance or national reporting systems. Kenya and Ethiopia have strong contact tracing protocols and data points that could be reported nationally. **Guidance and dissemination:** Provide clear guidance on data collection, reporting, and utilization to benefit national mortality surveillance and Spectrum estimates.

Appendix C – Participants

In-person participants

Name	Organization
Cari van Schalkwyk	SACEMA
Eleanor Gouws	UNAIDS
Eline Korenromp	UNAIDS
Elleni Seyoum	UNAIDS
Guy Mahiane	Avenir Health Inc
Ian Wanyeki	UNAIDS
James Guwani	UNAIDS
Jasmine Buttolph	PEPFAR
Jeffrey Imai-Eaton	Harvard University
John Stover	Avenir Health Inc
Keith Sabin	UNAIDS
Laura Porter	CDC
Leigh Johnson	University of Cape Town
Mary Mahy	UNAIDS
Melaku Dessie	USAID
Monita Patel	CDC
Morris Ogero	Kenya MoH
Oli Stevens	Imperial College London
Rachel Esra	Avenir Health Inc
Ray Shiraishi	CDC
Reshma Bhattacharjee	USAID
Rob Glaubius	Avenir Health Inc
Sara Herbst	PEPFAR
Sharmistha Mishra	University of Toronto
Shona Dalal	WHO
Wolfgang Hladik	CDC
Yuri Munsamy	SACEMA

Virtual participants

Name	Organization
Andreas Jahn	Malawi MoH
Avi Kenny	Duke University
Mathieu Bastard	WHO
Rebecca Bunnell	PEPFAR
Kelsey Case	Independent consultant
David Boulware	University of Minnesota
Hiwot Haile-Selassie	WHO
Jinkou Zhao	Global Fund
Joseph Jarvis	London School for Hygiene and Tropical Medicine
Mark Siedner	AHRI
Mary-Ann Davies	UCT, CIDER
Mathieu Maheu-Giroux	McGill University
Nim Pathy	WHO

Nathan Ford	WHO
Peter MacPherson	Glasgow University
Radha Rajasingham	University of Minnesota
Ajay Rangaraj	WHO
Francoise Renaud	WHO
Renee de Waal	UCT, CIDER
Reshma Kassanjee	UCT, CIDER
Kathryn Risher	Penn State University
Italia Rolle	UNAIDS
Stephen Olivier	AHRI
Tim Brown	East West Centre
Marco Vitoria	WHO
Wiwat Peerapatanapokin	East West Centre
Anna Yakusik	UNAIDS

Appendix D – Agenda

Monday, 29 July

Time	Duration (mins)	Topic	Presenter(s)/ Lead Discussant
9.00	20	Welcome and introductions	Mary Mahy
9.20	30	Overview of 2024 estimates	Eline Korenromp
9.50	10	Meeting objectives	Cari van Schalkwyk
Session 1: Advanced HIV disease (chair: Jeff Imai-Eaton) **Objective:** • Validate Spectrum estimates of AHD (defined as CD4 < 200) • Identify needs for future AHD estimates from modelled results and surveillance data sources			
10.00	15	• Overview of WHO AHD work area and country implementation plans of CD4 testing • Summary of WHO systematic review of AHD	Nathan Ford
10.15	10	Advanced HIV Disease estimates from PHIA surveys	Shona Dalal
10.25	20	Comparing Spectrum estimated AHD prevalence among PLHIV to: • PHIA-reported AHD prevalence among PLHIV • Country program GAM-reported AHD at diagnoses or ART (re-)initiation • WHO systematic review of AHD prevalence • AHD at ART initiation (IeDEA)	Oliver Stevens
10.45	15	TEA	
11.00	60	Discussion	
12.00	90	LUNCH	
Session 2: Causes of Death among PLHIV (chair: Shona Dalal) **Objectives:** • Review implementation of CM and TB death estimates in Spectrum, within the envelope of Spectrum-estimated HIV-related deaths			
13.30	90	Tuberculosis	WHO
15.00	10	TEA	
15.10	90	Cryptococcal Meningitis	Radha Rajasingham
Session 3: Mortality among PLHIV on ART (chair: Jeff Imai-Eaton) **Objectives:** • Validate Spectrum estimates of on-ART mortality against various sources • Present new approach to Spectrum mortality structure given incomplete CD4 data in most IeDEA cohorts			
16.40	50	Mortality trends among PLHIV in AHRI population cohort, KwaZulu Natal	Mark Siedner
17.30		CLOSE	

Tuesday, 30 July

Time	Duration (mins)	Topic	Presenter(s)/ Lead Discussant
Session 3: Mortality among PLHIV on ART – continued			
9.00	30	Mortality among people on ART from routine monitoring data through PEPFAR quarterly MER reporting	Sara Herbst
9.30	60	New approach to Spectrum mortality structure with reduced CD4 data availability	
10.30	10	TEA	
Session 4: Excess mortality among PLHIV (chair: John Stover) **Objective:** • Review method to separate AIDS vs non-AIDS deaths			
10.40	30	Excess mortality AIDS vs non-AIDS: • Proposed approach for additive mortality • Proposed method for regions other than WCENA • Impact on estimates by region	Rob Glaubius/Jeff Imai-Eaton
11.10	20	Impact of proposed excess mortality split-off on CSAVR estimates	Guy Mahiane
11.30	60	Discussion	
12.30	90	LUNCH	
Session 5: Treatment interruption (chair: Cari van Schalkwyk) **Objectives:** • Review outcomes from 2024 estimates • Provide recommendations on guidance and definitions for calculating Spectrum interruption inputs			
14.00	15	Impact of new default rates, how many countries adopted defaults vs. entered national program data, etc	Eline Korenromp / Rob Glaubius
14.15	15	WHO definitions of loss to follow-up and treatment interruption	Hiwot Haile-Selassie
14.30	15	GAM-reported cohort retention on ART, data by country and over time	Eline Korenromp/ John Stover / Anna Yakusik
14.45	15	Proposal: Default relationship between WHO treatment interruption and treatment discontinuation	Jeff Imai-Eaton/ Rob Glaubius
15.00	60	Discussion	
16.00	10	TEA	
Session 6: World Population Prospects 2024 update (chair: Oli Stevens) **Objective:** • **Review impact of new population estimates on HIV estimates**			
16.10	20	Impact of WPP 2024 vs 2022 estimates on HIV estimates	Rob Glaubius
16.30	30	Discussion	
17.00		CLOSE	

Wednesday, 31 July

Time	Duration (mins)	Topic	Presenter(s)/ Lead Discussant
Session 7: Dynamical modelling of HIV trends in key populations in sub-Saharan Africa (chair: Leigh Johnson) **Objective:** • Review updates to Goals-ARM			
9.00	20	Key population modelling technical working group aims, processes, and deliverables	Sharmistha Mishra
9.20	20	Goals-ARM development update	Rob Glaubius
9.40	65	Discussion	
10.45	15	TEA	
Session 8: New infections by Key Population and their partners (chair: Sharmistha Mishra) **Objective:** • Review methods of estimating uncertainty in estimates			
11.00	5	Introduction	Keith Sabin
11.05	10	Key population size estimates used in Goals	Keith Sabin
11.15	15	Methods, including uncertainty intervals, hierarchies of models & sources for population size estimates and infections	Eline Korenromp
11.30	15	Goals calibrations, including country inputs and feedback & data and model triangulations	John Stover
11.45	15	Results, interpretation and next steps	Rachel Esra / Eline Korenromp
12.00	45	Discussion	
13.00	60	LUNCH	
Session 9: Sub-national HIV estimates (chair: Cari van Schalkwyk **Objective:** • Review the SHIPP tool methods and make recommendations about its suitability and/or future development			
14.00	75	Sub-national HIV estimates In Priority Populations (SHIPP) workbook tool	Katie Risher
15.15	30	Naomi issues	Rachel Esra
15.45	15	TEA	
Session 10: Sub-national EPP stratification (chair: Sharmistha Mishra) **Objective:** • Review likelihood structure and make recommendations about removing sub-national EPP stratification			
16.00	60	Replacing sub-national EPP stratification: Proposed likelihood structure for weighting ANC data	Jeff Imai-Eaton
17.00		CLOSE	

Thursday, 1 August

Time	Duration (mins)	Topic	Presenter(s)/ Lead Discussant
Session 11: ART coverage data discrepancies (chair: Leigh Johnson) **Objectives:** • **Plan further work related to ART coverage data discrepancies** • **Develop guidance and examples for countries to apply** • **Review current uncertainty on ART numbers and proposed adjustments**			
9.00	30	ART adjustments in the 2024 HIV estimates: • Summary of country adjustments • Rationale/ evidence supporting adjustment • Summary of ongoing DQA activities	Rachel Esra / Ian Wanyeki
9.30	10	Estimating subnational unmet need for ART for using discounted ART data	Rachel Esra
9.40	20	ANC ART coverage analysis and impact	Jeff Imai-Eaton
10.00	20	Methods to interrogate and adjust ART programme data in Malawi	Andreas Jahn / Stone Mbiriyawanda
10.20	20	ART dashboard and validation process in Kenya	Morris Ogero
10.40	40	Discussion	
11.20	10	TEA	
11.30	20	Proposed refined method to estimate uncertainty on Spectrum country-level ART numbers	John Stover
11.50	40	Discussion	
12.30		LUNCH	
Session 12: Age distribution of new infections (chair: John Stover) **Objectives:** • **Review update to generalized epidemic default age incidence rate ratio pattern** • **Review evidence on incidence age pattern shifting to older ages** • **Consider default pattern and model specification for time-varying age incidence rate ratios**			
14.00	15	• Summary of recommendations from 'HIV incidence in Sub-Saharan Africa' Reference Group meeting • Empirical evidence on time-varying incidence rate ratios	Oli Stevens
14.15	25	• Default age incidence rate ratios: Update the generalized epidemic pattern considering recent incidence and serology surveys • Simultaneous fitting to survey prevalence, survey ART coverage, and programme ART by age	Rob Glaubius
14.40	20	Model for time-varying incidence rate ratios and default pattern	Jeff Imai-Eaton/ Rob Glaubius
15.00	60	Discussion	
16.00		CLOSING REMARKS	