Technical updates for UNAIDS HIV estimation tools

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

16-19th May 2022

Hotel Victoria, Glion, Switzerland

REPORT & RECOMMENDATIONS

Index

Abbreviations

AIM	AIDS Impact Model
ANC-RT	Antenatal Clinic Routine Testing
ART	Antiretroviral Therapy
ART-CC	ART Cohort Collaboration
ASM	Age Structured Model
CDC	US Centres for Disease Control and Prevention
CSAVR	Case Surveillance and Vital Registration
DHIS	Division of Health Informatics and Surveillance
DQA	Data Quality Assessment
DTG	Dolutegravir
EPP	Estimation and Projection Package
FSW	Female Sex Worker
GAM	Global AIDS Monitoring
GBD	Global Burden of Disease
HIC	High Income Country
IBBS	Integrated Biological and Behavioural Surveillance Survey
IeDEA	International Epidemiology Databases to Evaluate AIDS
IRR	Incidence Rate Ratio
KOS	Knowledge of Status
KP	Key Population
LMIC	Low- and Middle-Income Countries
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
PMTCT	Prevention of Mother to Child Transmission
PWID	People Who Inject Drugs
SSA	Sub-Saharan Africa
TGW	Transgender Women
UNAIDS	Joint United Nations Programme on HIV/AIDS
VLS	VL Suppression
WHO	World Health Organization
WPP	World Population Prospects

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 B**).
Cari van Schalkwyk, June 2022

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, Imperial College London and the University of Cape Town. The May 2022 meeting was the first meeting where SACEMA joined the Secretariat.

Meeting Overview

The UNAIDS Reference Group held its first in-person meeting in two years in Glion, Switzerland from 16-19 May 2022. While primarily an in-person event, some participants joined and presented remotely via Microsoft Teams. The meeting featured presentations and group discussions to generate consensus recommendations, divided into the following 8 sessions:

This report presents a summary of the meeting presentations and discussions that underpin recommendations by the Reference Group. The presentations are available to meeting participants (**Appendix B**) at www.epidem.org (others, please contact the Secretariat via epidem@sun.ac.za). The final recommendations can be found at the end of this report. The recommendations (**Appendix A**) drafted at these meetings provide UNAIDS with guidance on generating HIV estimates, review current approaches, and identify required data to further improve HIV estimates. The meeting agenda and objectives are in Appendix C. Previous meeting reports are available at www.epidem.org. This transparent process aims to allow the statistics and reports published by UNAIDS and partners to be informed by impartial, scientific peer review.www.epidem.org. This transparent process aims to allow the statistics and reports published by UNAIDS and partners to be informed by impartial, scientific peer review.

Introduction

Mary Mahy opened the meeting and welcomed everyone on behalf of Peter Ghys. She thanked Jeff Eaton and Oliver Stevens for their leadership of the Reference Group Secretariat for the last four years and introduced the new leads at the South African Centre for Epidemiological Modelling and Analysis (SACEMA) at the University of Stellenbosch—Cari van Schalkwyk and Faikah Bruce. Jeff Eaton and Leigh Johnson will continue as co-chairs of the Reference Group for continuity.

Mahy described that we have the final kilometre to go in our HIV control efforts, but in this last kilometre, we need to be measuring at the meter—we need to be measuring at a much more granular level and as a result our models need to be more and more precise. The Reference Group is critical for ensuring we can reach that precision, while recognising that models can only be as good as the incoming data, so we need to focus on how to improve without getting lost in the weeds of that process.

After a brief overview of meeting objectives, Mahy presented an **overview of the preliminary 2022 UNAIDS estimates.** Estimates were produced for 172 countries in this round, representing 99% of the world population (22 small population countries do not have estimates). She compared historical trends between the final 2021 estimates and preliminary 2022 estimates and highlighted that some countries in regions with the largest changes are still discussing to finalise results. Mahy summarized the main issues that arose in the 2022 estimation round, as background information for sessions dedicated to each:

- ANC Data:
    - Continued challenges with antenatal clinic (ANC) routine data including comparability of ANC1 visits to births (session 1)
    - Challenges with registries capturing known positives and use of those data in the calculation of prevalence (session 8)
- Case surveillance and vital registration (CSAVR) (session 2)
    - Each CSAVR model gives considerably different result – suggesting back calculation from case reports and deaths does not uniquely specify an epidemic trajectory.
    - Adding CD4 or standard adjustments for misclassification of deaths has not resolved the differences
    - Comparisons between EPP and CSAVR suggest discrepant interpretations of the data
- Increased need for more precision around ART coverage.
    - High burden countries are reaching over 100% coverage, require more precision on treatment gaps. Ideas for improving the measurement and visualization of treatment “churn” (session 3)
- Continued demand for improved key population data
    - Move of concentrated epidemics to CSAVR means that we lose ability to estimate mode of transmission. Is there reliable data on modes of transmission from case surveillance? (session 4)
    - Key population workbook in sub-Saharan Africa: gaps in data for SSA (Sessions 5 and 8)

Session 1 – Review of ANC testing data

Mary Mahy started this session with an overview of **routine ANC testing and PMTCT data in Spectrum/EPP/Naomi**. Routine ANC data consist of aggregate reporting of number of women attending ANC services, HIV status and testing outcomes of pregnant women, and women receiving and initiated on ART for prevention of mother-to-child transmission (PMTCT). ANC testing data are, ideally, a complete recording (census) of all pregnant women attending ANC in the country. These data are used in three places in the estimation:

1) The HIV prevalence trend among women attending ANC is used in EPP fitting to infer population HIV incidence trends,
2) The HIV prevalence level among women attending ANC is used in the Spectrum AIM module to calibrate fertility among women with HIV and resulting pregnant women prevalence, and thereby women who need PMTCT, and
3) Informing the spatial pattern in HIV prevalence by district in Naomi.

Underestimating ANC HIV prevalence leads to overestimating PMTCT coverage, and underestimate child infections. Inaccurate data on HIV prevalence trends at ANC results in biased estimates for HIV incidence; for example, spuriously declining ANC HIV prevalence results in underestimating incidence.

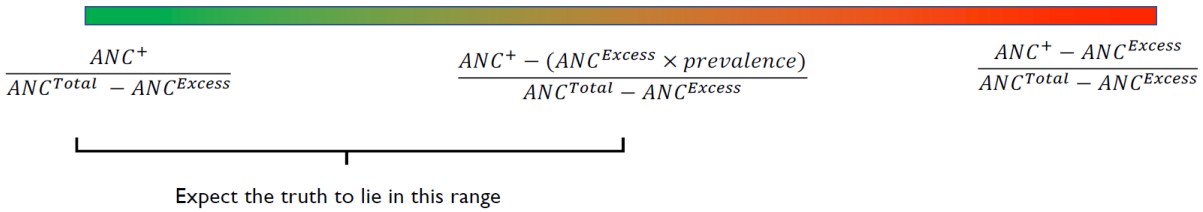
There have been obvious data problems in recent routine ANC data reported by countries in Spectrum estimates, notably, more ANC1 visits than births, and epidemiologically implausible prevalence trends that either vary more year-on-year or decline more rapidly than could be explained by true changes in prevalence among pregnant women. In any of the places that ANC data are entered and fitted (AIM, EPP, Naomi), prevalence may be biased down if re-tests are counted in the total women tested (as if being a new test), since HIV negative women are more likely to be retested within the same pregnancy.

John Stover followed with a presentation on considering possible **adjustments to ANC/PMTCT inputs and coverage for settings where ANC clients exceed births**. He explained why PMTCT coverage can be over 100%: program data on women receiving ART exceed Spectrum's estimate of pregnant women with HIV, which is a function of prevalence in women of reproductive age, overall fertility, and the reduction in fertility due to HIV infection. In a third of Spectrum files, the ratio of ANC1 visits to births is above 1. In 5% of countries, more women are tested for HIV than there are ANC1 visits recorded, and in 10%, all ANC1 (100%) got a test. **Stover therefore concluded that routine ANC data from recent years are not high enough quality to use in estimation.**

Hypotheses for why the number of ANC1 visits are greater than the number of births: visits may not be properly recorded [Salomon]; women hear they're pregnant at clinic close to work, go for ANC near home and thus get recorded ANC1 twice [Stover]; not every woman who goes to ANC will have a (live) birth [S Patel].

Next, **Oliver Stevens** presented a case study of the **impact of ANC routine testing surveillance data on HIV prevalence and ART coverage estimates in Mozambique**. In all 11 province-level Spectrum files for Mozambique (2022 estimates round), ART and PMTCT coverage was above 100%. An apparent rapid decline in ANC HIV prevalence is most likely explained by double counting (probably, some among positive tests, but more among those negative at initial test).

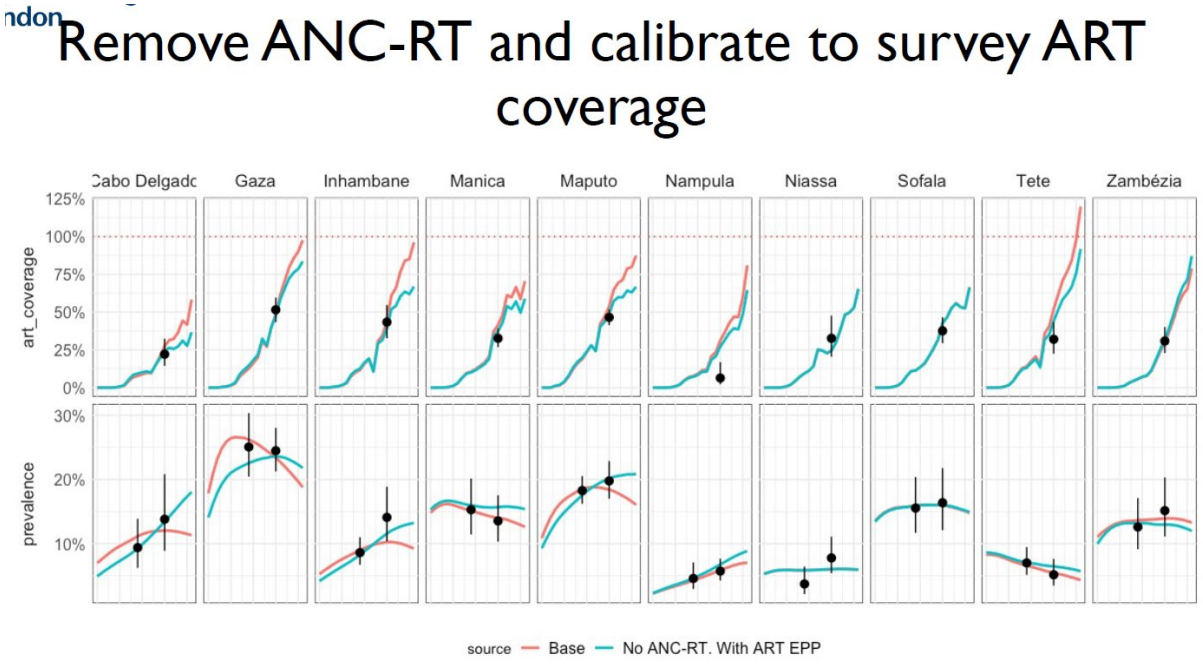
$$ANC\ prevalence = \frac{ANC^{+}}{ANC^{Total}}$$



Stevens discussed three indications that the Mozambican ANC-RT data may be skewed:

1) One concern is that the number of births, forming the denominator for ANC coverage, may be underestimated. However, Stevens described that an implausibly large increase in fertility rates, compared to recent survey and census data, would be required to increase the number of births sufficiently that ANC coverage would reduce to plausible levels. Recent census and survey data in Mozambique provide confidence about fertility estimates. Therefore, this suggests the issue is primarily overstating the number of ANC1 visits (the numerator) rather than related to number of births denominator.
2) A local fertility adjustor in Spectrum-AIM to Mozambique's ANC prevalence according to the routine data, is fitted to a value much below the default for Eastern Africa region.
3) Very sharp increases in Spectrum ART coverage estimates between 2020 and 2021.

To illustrate the effect of these data on HIV estimates, Stevens fitted Spectrum with and without the ANC-RT data and included survey ART coverage in the fitting. Excluding ANC-RT drastically changed prevalence trends in Cabo Delgado, Inhambane and Maputo provinces, with much higher current and recent prevalence in all three.



In the final presentation of the session, **Victor Kabwe** and **Patrick Amanzi** presented results from an **ANC data quality assessment exercise** performed in 34 sites across 10 provinces **in Zambia**. Teams visited facilities and collected data from registers and compared this to data from DHIS2 and PEPFAR's DATIM system (DATIM: Data for Accountability, Transparency and

Impact). For PMTCT indicators (numbers testing; known HIV; new HIV), entries in DATIM were within ±5% to the DQA in only 12-29% of facilities over time. Data from DHIS2 'matched' in 6-15% of sites. Older data (2018) were more likely to be incomparable (couldn't find the old registers), but as incomparability decreased over time, highly discrepant results (>10% difference) increased.

The recommendations following discussions after these presentations are captured in Appendix A.

Session 2 – CSAVR

The Case Surveillance and Vital Registration (CSAVR) model is a tool within Spectrum for estimating HIV incidence trends from reported numbers of annual HIV diagnoses and numbers of AIDS deaths recorded through vital registration. The model is primarily used in middle- and high-income countries with relatively concentrated epidemics and relatively complete case surveillance and vital registration systems. Due the complex nature of these data, there have been persistent challenges with the model reconciling the various data inputs, including the distribution of CD4 at diagnosis and providing an optimised user interface.

This session was introduced by a presentation from **Eline Korenromp** providing an **overview of and challenges faced in countries that used CSAVR for estimates in 2022**. In this estimates round, several countries used CSAVR that did not in the previous round – increasing the total to 59.

2022 → 2021	EPP Gen	EPP Conc	AEM	CSAVR	ECDC	Direct Incid.	Total 2021
EPP General.	37	PNG					**38**
EPP Concentr.		37		ARM			**38**
AEM			13				**13**
CSAVR		TUN		52	AUS	HUN	**55**
ECDC				BGR,HRV, NOR	9		**12**
Direct Incidence				MEX, JOR, KWT		13	**16**
Total 2022	**37**	**39**	**13**	**59**	**10**	**14**	**172**
PLHIV 2020 (L-2021)	**17.6 M**	**4.8M**	**2.1 M**	**2.6M**	**0.12 M**	**10.5M** (ZAF, BRA)	**37.7M**

Korenromp noted that country teams appreciated the increased speed at which CSAVR runs, which enabled running many iterations using different data/assumptions. Most countries (30/59) used incompleteness-adjusted estimates of AIDS deaths from GBD 2020 results, and 14/59 used CD4 at diagnosis in calibration (although it does not have a noticeable effect on results). She followed with some examples of issues with CSAVR estimates: estimated KOS lower than program-reported number receiving ART; temporary drops in testing and diagnoses rates during COVID potentially biasing results; representation of immigrant diagnoses. She proposed recommendations to overcome these – these are captured in **Appendix A**.

Next, **Guy Mahiane** presented details on why the issues that Korenromp highlighted arise, and proposed solutions. Instances of KOS estimated below AIM-recorded numbers of people on ART can occur because CSAVR sends the proportion of people who know their status (comprised of PLHIV on ART and PLHIV knowing their status but not on ART) to AIM, whereas CSAVR and AIM do not always estimate the same numbers of PLHIV, leading to a discrepant

number who know their status between CSAVR and AIM. The solution proposed is that CSAVR send its KOS estimate as a proportion of PLHIV not on ART, which AIM can then add to its recorded ART to calculate overall KOS (on and off ART).

Reasons for CSAVR overestimating CD4 at diagnosis, Mahiane suggested, could include, firstly, that CD4 data are not aligned with the other data types: A non-representative, biased subset of ART patients may receive CD4 at diagnosis—although coverage of CD4 testing is typically not below 80% in CSAVR countries [Korenromp].

Second, CD4 progression parameters may be problematic. Stover commented that rates of ART initiation in Spectrum were derived from SSA PHIA data, which may cause problems in CSAVR country fits. This conversation was continued in Session 8, with recommendations captured there.

Regarding the reduced HIV testing during COVID, the discussion concluded that although results will smooth out once adding post-2020-21 data, it would be worthwhile to capture HIV testing data in CSAVR and explore fitting to it. Recommendations following Mahiane's presentation are shown in **Appendix A.**

Then, **Rob Glaubius** presented modelling challenges accounting for user-inputted **immigrants who live with HIV**. In a handful of countries, mostly small European countries, HIV-infected immigrants constitute a substantial fraction of the overall HIV-positive population.

Imported cases potentially confound Spectrum's incidence estimates. Unless Spectrum models HIV-positive immigration, its only mechanism to reproduce observed HIV prevalence and mortality is to increase incidence, i.e., immigrants who got infected elsewhere, would be mis-attributed to incidence in the country.

For this reason, Spectrum has several mechanisms to represent migration:

1) DemProj/Spectrum assumes HIV prevalence in immigrants and emigrants equals prevalence of all residents, and accordingly subtracts or adds PLHIV in proportion to net (all-age both-sexes) migration (a crude representation, evidently not adequate given age/sex patterns in migration and in HIV; and when net migration is high to low prevalence settings, or vice versa)
2) Countries can input migrants into Spectrum's HIV-positive population, by age, sex and calendar year. (Limitations: countries record in-migrants but not out-migrants, and data may be otherwise incomplete or lack the age/sex disaggregation required in the current Spectrum).

Currently, using (2) may cause double counting with (1). The proposed solution is to suppress mechanism (1) for years that mechanism (2) is used.

A further challenge is to represent the effect of HIV-positive immigration on KOS estimates by CSAVR, and correctly distinguish HIV-positive immigrants who 1) were diagnosed before migrating, versus 2) were diagnosed at migration, or even 3) who immigrate before getting diagnosed.

As also in the ECDC case-based model, the recommended practice is to enter and fit only new diagnoses, i.e., exclude people who received prior diagnosis in another country. CSAVR, to this end, has the option to enter previous diagnoses among migrants by year, age and sex, which CSAVR then (although labelled as new, considered previous) deducts from the 'New

diagnoses' (even though the immigrant entry tab is mislabelled 'New immigrant diagnoses instead of 'Previous immigrant diagnosis). (A limitation, illustrated by Korenromp, is that incomplete time series of immigrant data causes jumpy results, for both the estimated diagnoses, incidence, and CD4 counts)

Glaubius proposed to limit user entry of HIV-positive immigrant to one place in AIM (instead of in AIM and in CSAVR) and to output user-entered HIV-positive immigrants against the Spectrum-calculated net HIV migration so users can cross-check the representation (and guard against artefacts related to incomplete migration data). To accommodate countries lacking age/sex stratification in their migration data, and possibly also refine the age/sex pattern in net migration calculated within Spectrum, Glaubius proposed to explore age-sex distributions in countries with such data.

Finally, it was noted that *out* migration of PLHIV may be less relevant for CSAVR countries, given these (higher income) countries have much lower HIV rates than (poorer) countries with net emigration.

Key point from the discussion:

- Countries may define 'migrants' differently (e.g., Dominicans of Haitian descent have lived in the DR for generations but are still named migrants in surveys; Rohingya in Southern Bangladesh are staying but are still considered migrants). We need to provide guidance on how we want countries to define migrants [Keith Sabin, Irum Zaidi, Stover]

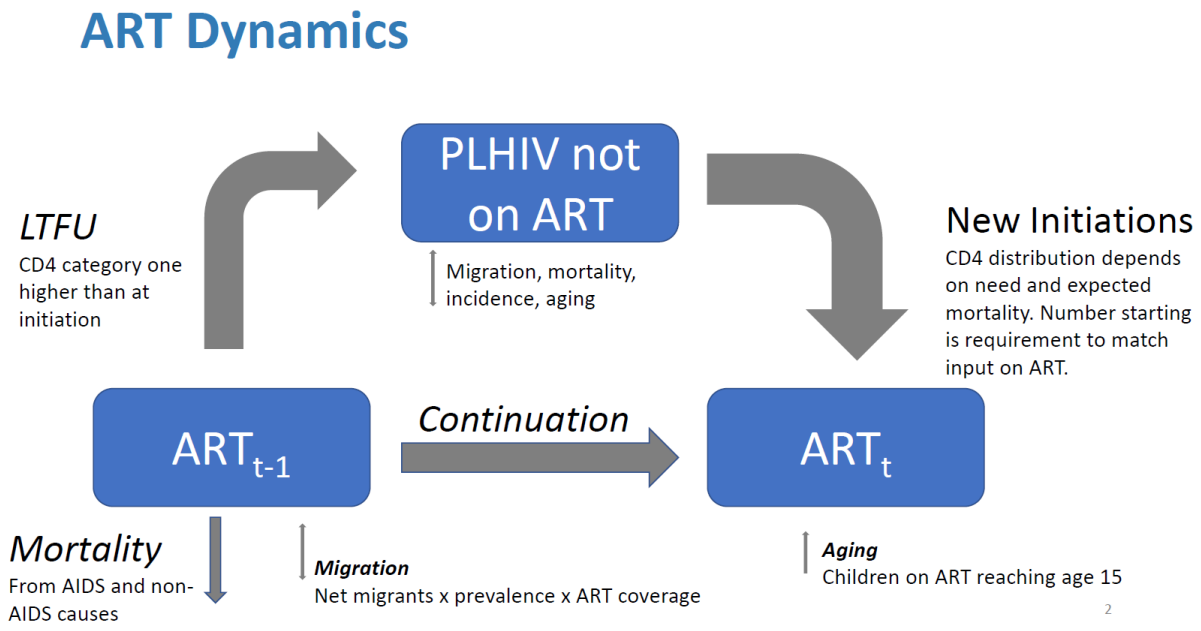
It was decided to launch a working group on issues regarding migration of PLHIV. Recommendations from the session is in Appendix A.

Session 3 – Testing and treatment churn

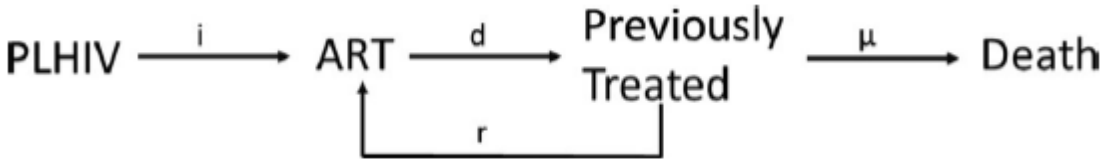
Jeff Eaton introduced this session highlighting the three main issues with how testing and treatment is handled in the current tools:

1) There are increases cases in which the input number of people on ART (from programme data) are higher than the model-estimated number of PLHIV (estimated by Spectrum).
2) Discrepancies between ART coverage by age and sex in Spectrum results and ART program data.
3) Non-representation of a previously treated or treatment-interrupted population and the non-reconciliation of people retesting and being reinitiated onto treatment.

In the first presentation of the session, **John Stover** reviewed **Spectrum's ART model**.



Number of people on ART each year is input data, and Spectrum calculates the number of new initiations to match the ART change from year *t* to *t-1*, also considering losses from death, emigration and treatment interruption. People can start ART at any CD4 counts, and they stay in the CD4 category they started for the duration of treatment. However, at time of interruption, we assume they are in one CD4 category higher than when initiated.



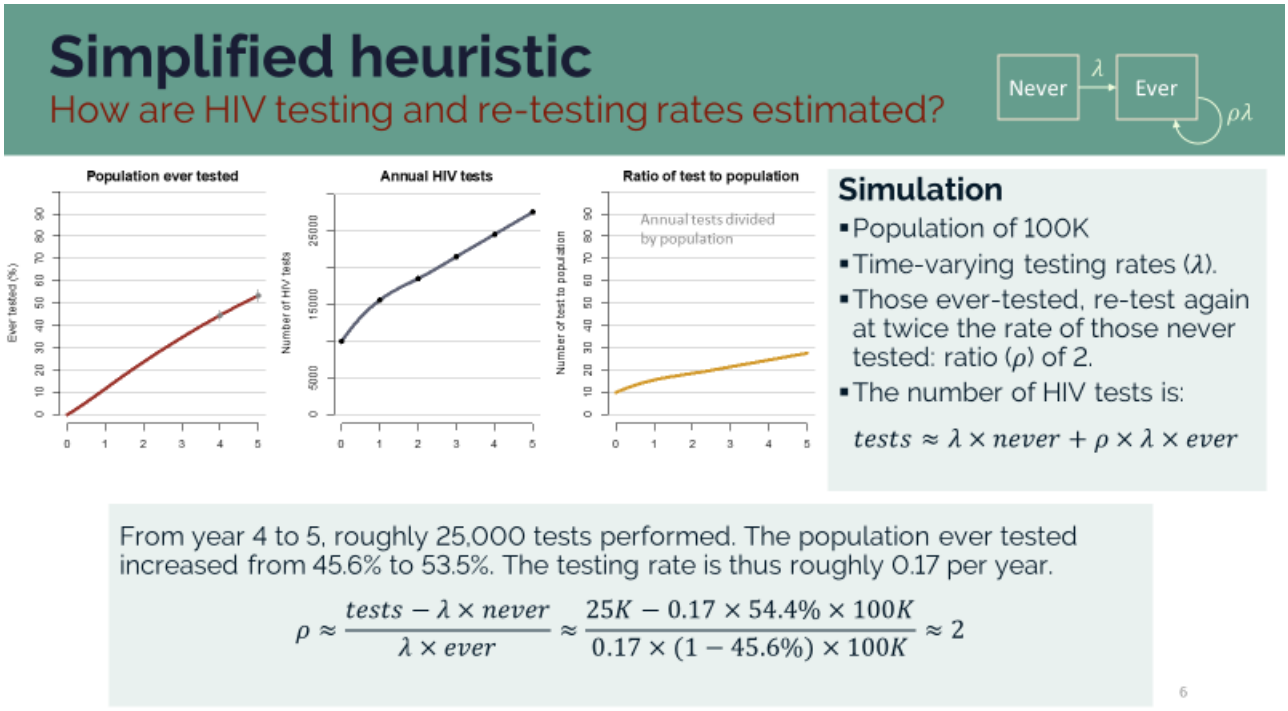
Spectrum calculates the previously treated population, which is shown as a Validation output. The calculation considers numbers who disengaged from care (not all countries report this), the rate of re-engagement, and mortality of PLHIV who were previously on ART. Currently, users can change these parameters to see how it affects their results, but to make this part of formal Spectrum outputs, we need data to inform the parameters.

Key point discussed: Zaidi asked why this information is important. Eaton responded, for all the dynamics, since people who reinitiate will on average be older and so the distribution between re-initiations vs new initiations will affect the age distribution in ART, and thereby the age distribution of new infections.

Ian Wanyeki briefly presented **knowledge of status in West and Central Africa**, estimated using the Shiny90 model, linked to Spectrum. In Sao Tome and Principe and Cape Verde, KOS was estimated to be over 100%, with female KOS above 120% in both. Four other countries also had KOS estimated above 100%.

Eline Korenromp followed with an overview of **loss to follow up from ART in 2022 Spectrum files**. 64 countries inputted rates for adult LTFU for at least one year since 2010. The mean inputted LTFU rates were about 3-4% annually. Some countries reported this data but not consistently; only 22 filled in all years. 38 countries had inputted child LTFU for at least one year, 9 for all 12 years, at an average of 2% LTFU per year. Only 19 of 239 Spectrum files entered data on ART **re-initiations**.

Next, **Mathieu Maheu-Giroux** gave an overview of Shiny 90 estimations, conducted in the 2022 round for 41 generalized epidemics with a household serosurvey available (all in sub-Saharan Africa). Modelling **re-testing** proved critical to reconcile test positivity rates reported with incidence, numbers of PLHIV, and knowledge of status. They estimated that in 2020, as much as 60% of PLHIV tested were not tested for the first time, i.e., they were not new diagnoses. To validate this Shiny90 result, Maheu-Giroux presented an external (non-Shiny 90) simple model simulation illustrating how retesting rates are inferred by the model:



This calculation is more complex in Shiny 90, which considers births, deaths, HIV incidence, ageing. Maheu-Giroux furthermore cited results of KOS and testing from PHIA surveys, that also confirm high rates of retesting.

Next, Maheu-Giroux addressed the problem of **perceived KOS estimates that are inaccurately high,** examples for which were given in the previous presentation by Wanyeki. There were four potential reasons:

1) Within-country heterogeneity in HIV testing, beyond age, sex and HIV status.
2) Programs may overstate numbers of people on ART, which elevates Shiny's KOS estimate for any age/sex/year group where the initial estimate was below ART – thus inflating the overall KOS estimate
3) Issues with population denominators (WPP): If population is underestimated, PLHIV will be underestimated.
4) Biases in either the household survey data or the routine HIV testing program data.

Two additional reasons highlighted by Eaton:

5) Inaccurate incidence estimates in earlier years are not generating enough people with HIV.
6) Inaccurate incidence rate ratios by age and sex imposed by Spectrum.

Maheu-Giroux then discussed treatment churn, showing two examples: 1) 32% of people (re)initiating ART in KZN had VLS at time of (re-)initiation (Sithole 2021) and, 2) 37% of people tested for HIV in Zambia had VLS (Tessema 2022), both suggesting prior and recent ART exposure. Explicitly modelling treatment churn and HIV testing (and re-testing) could avoid some of the issues (e.g., KOS=>ART). However, there is a lack of programmatic data that capture these dynamics.

Next, **John Stover** presented the **Goals testing model**, designed to determine cost-effective testing strategies in the context of high KOS. This model stratifies the population into up to 13 categories (pregnant women, partners of pregnant women, STI-, TB- patients, people with HIV symptoms, FSW, MSM, PWID, partners of PLHIV, exposed infants, children, and other adults). People can be tested through different modalities, and for each combination of population and testing approach the model considers (user-specified) cost per person testing, linkage-to-care rates and proportion re-testing. The model is calibrated to testing data and provides estimates of testing yield per testing modality, impact of alternative strategies on progress towards increasing awareness of status, and other programmatic outcomes.

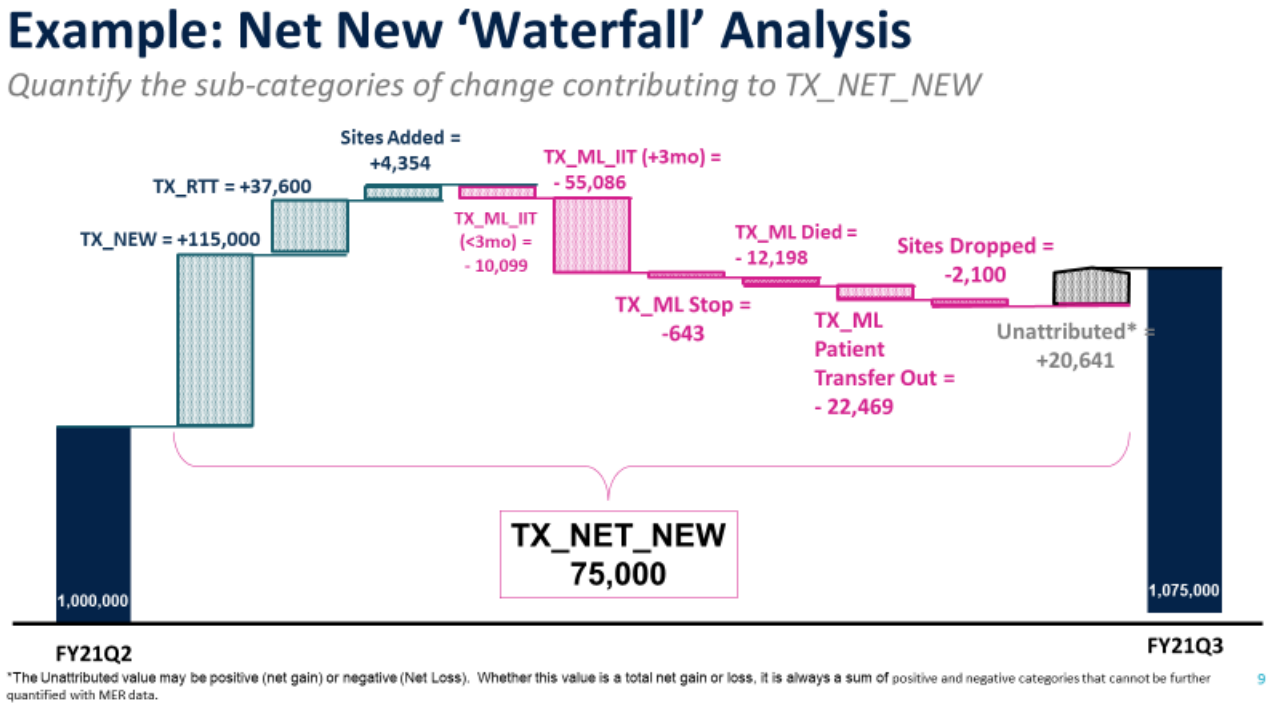
Leigh Johnson presented an overview of the **Thembisa** (re)testing and ART (re)initiation model. The testing model is similar to Shiny 90, which was guided by the Thembisa testing model. Unlike Shiny90, Thembisa explicitly models linkage to care and ART initiation. Rates of linkage to ART after (re)diagnosis depend on time (highest within first month), age, sex and CD4 at diagnosis. Some parameters are estimated from South African studies, while others are calibrated to fit reported numbers of PLHIV on ART. People interrupt ART at rates with priors determined from South African studies, and a female:male ratio from IeDEA-SA data. Rates of ART resumption (a function of rates of initial initiation) are consistent with results from three South African studies.

Johnson also briefly showed results of changes in testing strategies (index partner testing and self-testing) on HIV testing yields. These are testing modalities that are not explicitly represented in the Shiny90 model but may be important to capture in future due to their scale-up in many countries and potential impact on interpretation of observed HIV testing positivity. Historical implementation of passive partner notification (index partner testing) was estimated to increase HIV testing yield modestly by around 1%-point during the late 1990s and early 2000s, but with negligible effect during the 2010s because passive partner notification results in earlier diagnosis and less late diagnosis. Impacts of active partner notification or assisted

partner notification might be larger. HIV self-testing slightly increased positivity, as expected, but the magnitude was negligible because the number of HIV self-tests distributed thus far is small.

Next, **Tim Wolock** showed a modelling process for Malawi which simultaneously estimates HIV incidence and rates of ART initiation at district level. This model calibrates to number of ART patients (for years entered; but users can omit incomplete or unreliable data) and imposes internally consistent estimates of PLHIV and number on treatment for each year, adjusting ART by probabilistic projection and thus avoiding issues of inconsistent numerators and denominators as highlighted above for Spectrum. This method worked well for Malawi; it remains to be evaluated in for (the more typical) settings with less consistent ART programme data.

Sadhna Patel presented **PEPFAR's Waterfall analysis**—a monitoring exercise to measure continuity of treatment, interruption of and return to treatment, by visualising PEPFAR's monitoring, evaluation, and reporting (MER) indicators. She showed examples from Eswatini and Nigeria. This analysis provides a visual description of changes to numbers currently on treatment between two reporting periods, to highlight drivers of change and guide real-time discussions.



The following morning, **Harriet Nuwagaba-Biribonwoha** presented evidence of retesting from PEPFAR HIV recent infection surveillance programs. Eswatini, DRC, Rwanda and Zambia conducted such surveillance, among adult PLHIV identified as being newly diagnosed. People identified by the recency assay as recently infected but who are virally suppressed are redefined as re-testers; this concerned 44%, 39%, 41%, and 62% of PLHIV in DRC, Rwanda, Eswatini and Zambia, respectively. In Eswatini, this fraction ranged between 32 and 52% across different testing modalities (highest at mobile outreach/community-based testing).

Following the presentations, five working groups were convened to consider the evidence and recommendations for the session. **Jeff Eaton** introduced the working groups with an overview of this session and introducing three questions for each group to discuss:

- What strategies should we use or explore to address discrepancies between reported number on ART and (fewer) PLHIV, or inconsistent ART coverage by sex and age?
- How granularly should HIV testing, linkage, and ART initiation be represented in Spectrum? Should testing modalities be more explicitly represented? Should we explicitly representing testing and retesting linkage and ART initiation or keep them separate?
- How granularly should components of change in the ART population be represented in Spectrum? Currently Spectrum just models annual net changes needed to match the trend in§ reported ART, without considering user-entered data on new ART initiations. Typically, annual new ART initiations calculated and applied by Spectrum are far below reported new initiations, probably due to high rates of retesting and re-initiation (misreported as new initiations) as discussed before.

Working groups then discussed questions, including possible model structures and approaches; information and data gaps to fill; and assumptions required.

The recommendations that followed these working group discussions are captured in Appendix A.

Session 4 – Key population stratified estimates in concentrated epidemics

Mary Mahy opened this session stressing that the Global AIDS Strategy focuses on identifying the gaps. In most countries, key populations are being left behind. If estimation models do not allow us to understand epidemics by key populations, countries may not be able to close these gaps, especially given the challenging and stigmatizing legal environment that key populations live in. Sessions 4 and 5 focused on developing strategies for improving HIV estimates among key populations in the UNAIDS estimates.

First, **Keith Sabin** presented estimates of proportion of new infections in each population group (general population, FSW, PWID, MSM, etc.), produced by UNAIDS in the last 5-6 years (commonly referred to as the 'donut' plots). Lack of data is one of the greatest inequalities that we face with these groups; lacking commitment by national AIDS programmes is a major problem.

Next **Tobi Saidel** showed **results of a review on quality and biases in HIV case and death reports** by mode of transmission, drawing examples from settings selected for having good such data, as well as key population specific surveys. Specifically, countries were selected that had annual data on new diagnoses covering all different testing services, which were deduplicated, and with a valid breakdown by mode of transmission. Saidel and Yoko Shimada had contacted 45 countries and engaged with 36 (with language as a barrier in some of the remaining 9), from Eastern Europe and Central Asia, Latin America, and Middle East and North Africa region, as well as some from Asia and sub-Saharan Africa. With examples from China, Armenia, Ukraine, Cambodia and Philippines, Saidel illustrated that countries have different ways of categorising modes of transmission, and different ways to ascertain information from testing clients to categorize each case. She listed the following characteristics of countries with stronger mode of transmission information in case report data.

Findings

Characteristics of countries with stronger case report data with MoT

- Broad range of testing facility types (including several likely to capture high risk groups such as STI clinics, narcological clinics, NGOs with KP dedicated services, and prisons)
- Availability of VCT/HCT counsellors (outreach workers and social workers) to assess risk and assign preliminary mode of transmission, which can be revised later
- Obligatory reporting of HIV infections from all testing sites using standardized format
- Outreach workers who accompany clients who screen positive for confirmatory testing and follow-up
- Centralized confirmatory testing and care services
- Unique identifying codes linked to personal data (e.g. National ID, Passport, etc) so that a person only receives one code for life – (strongest deduplication)
- Cross-checking of personal data with information in electronic database to verify that the person does not already have a code before registering them as a new case
- Index testing of partners of HIV positive persons
- Focus on getting clients to "self-assess" substantial risk (more in LAC)
- Do not allow data to be uploaded into system unless it's complete (China)

Accordingly, Saidel proposed the following countries ed as candidates for a planned pilot of a CSAVR with MoT/KP breakdown (CSAVR-KP) are Morocco, Oman (MENA); Bolivia, Cuba, Peru (LAC); Armenia, Kazakhstan, Ukraine (EECA); and Cambodia, China, Philippines, Sri Lanka (APAC).

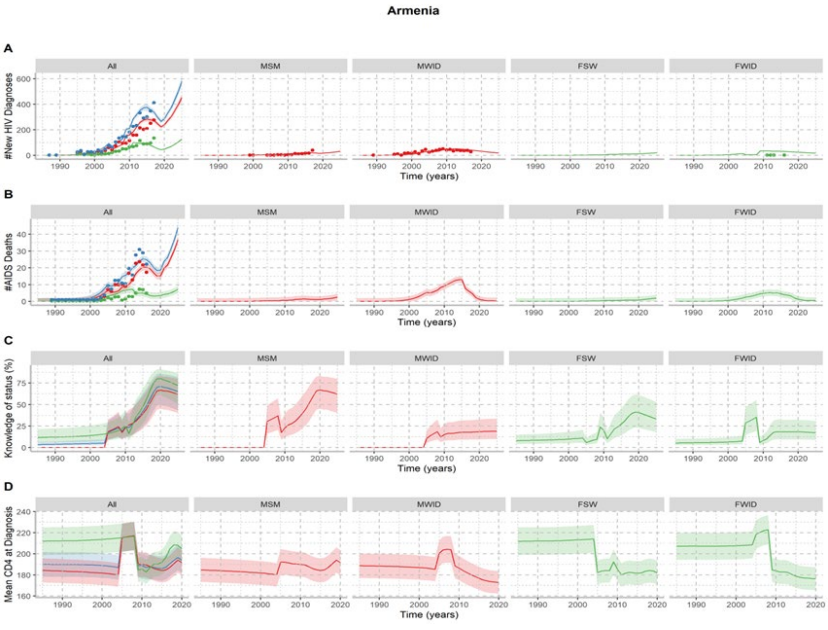
Guy Mahiane gave **an overview of CSAVR-KP**, an extension of the CSAVR model to separately model HIV incidence among multiple population groups. While data exist on HIV diagnoses by KP or probable mode of transmission, AIDS death data is not disaggregated this way. Therefore, survey data of prevalence among KPs need to be incorporated in fitting.

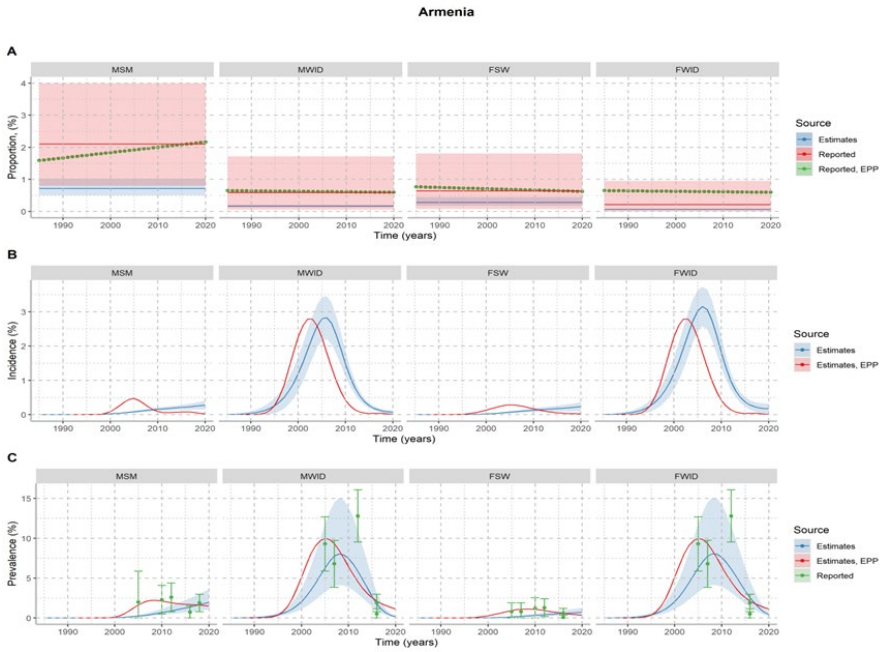
Key assumptions of CSAVR-KP are:

1) The KP proportion of the total population remains constant over time,
2) No turnover for MSM,
3) Age distribution of KPs time-constant, as that of the general population of same sex in 1970,
4) HIV incidence in KP is proportional to overall incidence of the same sex each year, and
5) the annual diagnoses in each KP is proportional to annual overall diagnoses of the same sex.

Mahiane showed an example application for Armenia, fitted to survey-based prevalence and new diagnoses by KP. The plots below show that CSAVR's overall case notifications and AIDS deaths match well to the data, and so do case notifications among MSM and MWID. However, the calibrated estimate for the proportion of MSM in the population was below national size estimate (i.e., mode of transmission for MSM may be undercounted). Also, CSAVR-estimated prevalence did not match historical survey estimates very well, demonstrating the need to include survey prevalence in the likelihood. Finally, modelled CD4 at diagnosis was below that in the available data (as for the overall CSAVR) and declined over time for unknown reason.

Mahiane also showed estimates for Portugal and Austria, but these countries had sparse prevalence data so were less informative pilots so far.





Key points from discussion:

- Armenia reported 2-3-fold more diagnoses amongst men than women. In a heterosexual epidemic, such a sex ratio is atypical, so this ratio suggests a predominantly MSM and injecting drug epidemic. But in model output, diagnoses among MSM and MWID are only a small part of overall male diagnoses, which suggests that reporting by mode of transmission undercounted numbers of MSM and MWID infections. The current model would tell us that this is a heterosexual epidemic, while there are probably just biases in this case report data. (Adam Trickey, Oliver Stevens, Eaton)
- Tim Brown asked about the contribution of labour migrants (mostly men, who become 'general male population'). Sabin confirms that these men get infected (probably from IDU) in Russia and get deported upon diagnosis there.
- Use additional Armenian data from Saidel's review to see if we can improve fits.
- Incidence which is occurring in other populations is getting diagnosed in the general population. Maybe we need a way to visualise turnover in the model (Brown).

In the next presentation, **Tim Brown** gave **a brief overview of EPP for concentrated epidemics**. The structure of EPP is determined by the users, driven by the data they have on prevalence and population size in sub-populations (can be regional, or by key populations). EPP also models remaining populations. EPP concentrated has no internal age structure. Sex is only specified at the highest level, not in terms of separate compartments, but by specifying the percent that are males in a particular subpopulation. This is used to calculate the sex ratio at the end.

EPP concentrated was used in 36 countries in 2022. Often the amount of surveillance data for fitting the model was low. For example, eight had fewer than 6 data points for MSM. Also, Integrated Biological and Behavioural Surveillance Surveys (IBBSs) are often done in different geographic locations, definitions of the populations being sampled vary over time, and sometimes they use a different sampling methodology. This reduces the comparability of data across surveys, and raised into question whether it is appropriate to estimate time trends on this data. Few of these data points can result in wildly varying fits like Uruguay MSM and Ecuador FSW). Changes in key population sizes not being smoothed can also result in

unexpected sharp changes to estimates. Brown showed examples of the impact that estimates of turnover can have on incidence and prevalence estimates.
He then discussed that age/sex IRRs used in Spectrum are the same for all populations, while we might expect that these differ by key populations, and by country. However, adding age/sex structure for key populations would be a challenge due to limited age stratified data.

Deepa Jahagirdar presented work in progress on the **introduction of age structure into EPP for concentrated epidemics**. Jahagirdar gave an overview of the differences between CSAVR and EPP for concentrated epidemics and showed how EPP-ASM (age structured model) will be different. She noted that despite limited availability of age-stratified data in concentrated epidemic settings, reasons for adding age structure are more accurately modelling shifts in age distribution of HIV prevalence, ART initiation and ART coverage over time. In EPP general, people enter the population at age 15, but people enter key populations at some age distribution.
Some assumptions she made in moving EPP-ASM towards KP estimation were 1) no migration into KPs, 2) survival rates are the same as the general population, 3) women enter FSW at age-specific rates from Thembisa, 4) number on ART in the key population is in proportion to their share in PLHIV/prevalence, and 5) those in turnover populations initiate ART at same rate as the key population.
She showed example of this model's fit for Senegal, with different age distributions at entry, different turnovers, tighter ANC bias parameter prior, etc. In most cases, EPP-ASM for KPs match data (and EPP) well.

Key points about assumptions from discussion:

- IDU may have lower, not equal survival rates as the general population (Trickey)
- Comparing remaining population prevalence to household surveys is not very accurate, since FSW and MSM also live in households (Mahy)
- If you include age structure, it may be more appropriate to also include age structure in turnover rates (Tim Brown)
- Maheu-Giroux and Johnson suggested calibrating to age-specific prevalence.
- Stover said that Spectrum needs an adjustment factor to match Senegal's prevalence, and it might be useful to see if feeding Deepa's EPP-ASM results (instead of EPP-Classic) into Spectrum might solve this problem.
- Assumption that FSW, clients, MSM have same age distribution at entry may be problem. Johnson mentioned that clients are older. Would be possible to get age distribution of clients of sex workers from household surveys? (Mahy)

The group split into four working groups following these presentations to discuss whether we should pursue continued development of CSAVR and EPP-ASM for key populations, under which conditions to pursue this, and to suggest other approaches if not. A summary of the discussions is captured in the recommendations.

Session 5 – Key population stratified estimates in sub-Saharan Africa

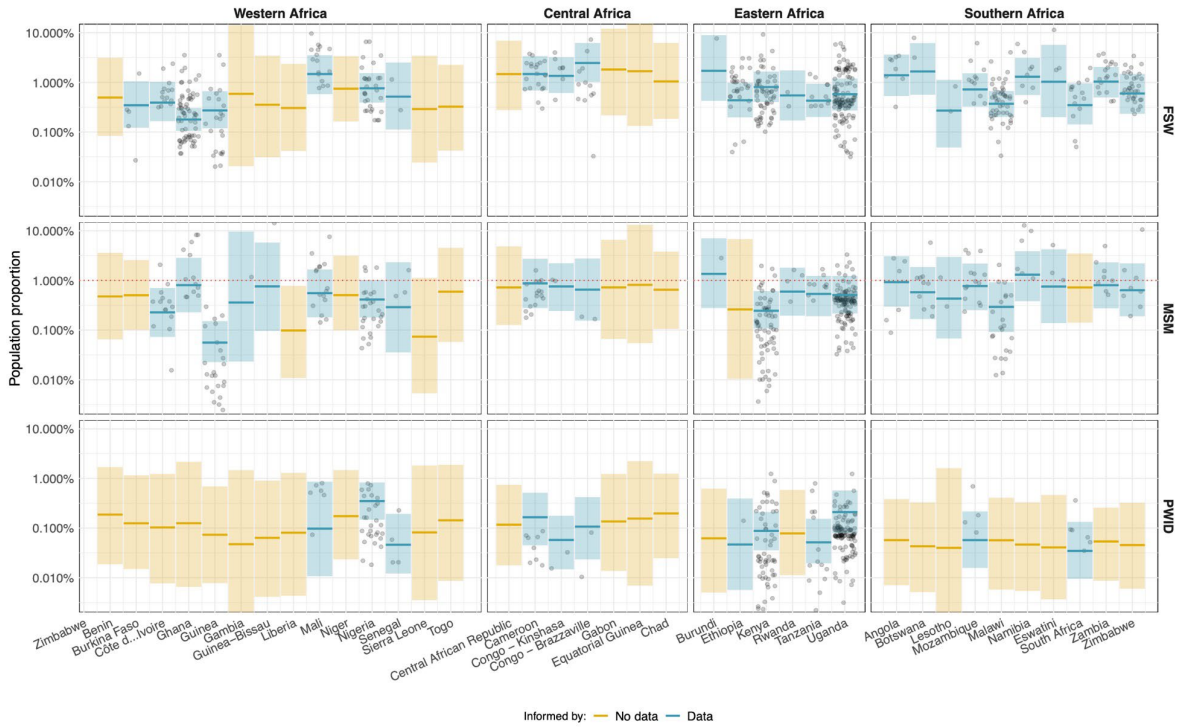
The objective of Session 5 was to plan modelling strategies and future development to represent key populations in HIV estimates in sub-Saharan Africa.

Oliver Stevens opened this session with a presentation titled '**Key population size, HIV prevalence, and ART coverage in sub-Saharan Africa: systematic collation and synthesis of surveillance data**'. He described the process followed to collect data from 2010 to present from all SSA countries about FSW, MSM PWID and transgender on three indicators: population size, HIV prevalence and ART coverage data. He acknowledged the large number of people and organisations who responded to requests for data. Data from the KPs in different countries at different time points were synthesised with regression models to estimate the 3 indicators for each KP and country. The population size estimate (PSE) model controls for sampling method and spatial correlation between national neighbours; the HIV prevalence model assumes that age-location-year matched total HIV prevalence can predict KP prevalence and a spatial correlation; and the ART coverage model makes these same assumptions (total coverage predicts KP coverage).

When discussing results, Stevens quantified data availability on each of the KPs for each of the 3 indicators. Only Kenya and Mozambique had data for all 3 indicators times 4 KPs, and several countries (South Sudan, Chad, CAR, Gabon, Liberia, Niger) had data on none.

		PSE		HIV prevalence		ART coverage	
Region	KP	Data points	Countries with data (%; n/N)	Data points	Countries with data (%; n/N)	Data points	Countries with data (%; n/N)
SSA	FSW	583	68 (26/38)	543	84 (32/38)	126	50 (19/38)
SSA	MSM	402	66 (25/38)	338	79 (30/38)	78	45 (17/38)
SSA	PWID	253	32 (12/38)	147	37 (14/38)	19	16 (6/38)
SSA	TG	59	16 (6/38)	95	53 (20/38)	16	26 (10/38)
ESA	FSW	382	94 (16/17)	225	94 (16/17)	86	76 (13/17)
ESA	MSM	283	82 (14/17)	113	82 (14/17)	49	65 (11/17)
ESA	PWID	203	35 (6/17)	43	35 (6/17)	13	24 (4/17)
ESA	TG	43	24 (4/17)	36	59 (10/17)	11	41 (7/17)
WCA	FSW	201	48 (10/21)	318	76 (16/21)	40	29 (6/21)
WCA	MSM	119	52 (11/21)	225	76 (16/21)	29	29 (6/21)
WCA	PWID	50	29 (6/21)	104	38 (8/21)	6	10 (2/21)
WCA	TG	16	10 (2/21)	59	48 (10/21)	5	14 (3/21)

The thick lines in the next plot shows the median PSEs for each KP in each country, with orange indicating that there was no PSE for these countries.



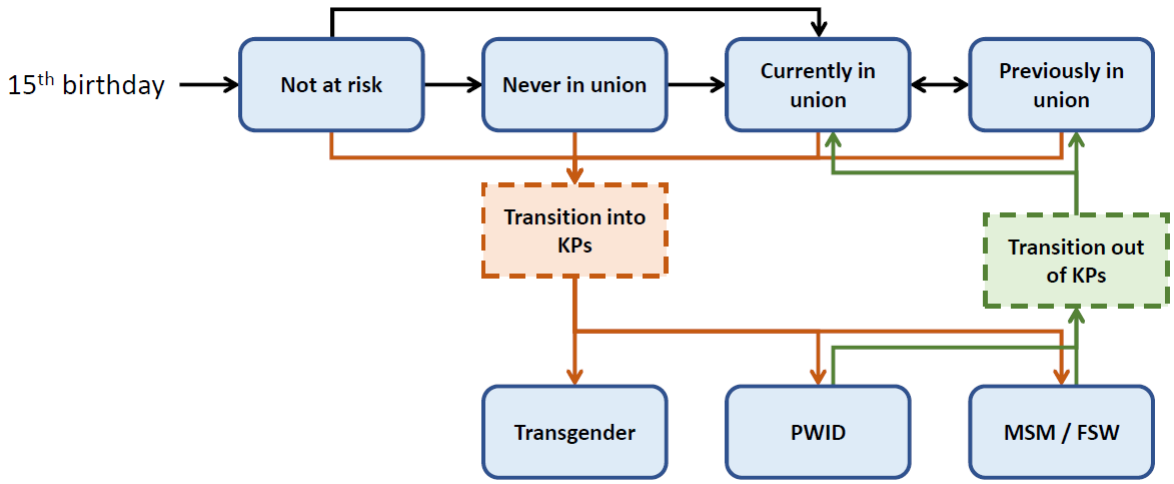
An example of results for HIV prevalence is that in an Eastern/Southern African country with 15% overall HIV prevalence, prevalence in FSW is estimated to be 40%, in MSM 19%, in PWID 45% and in TGW 24%, with large uncertainty bounds on all. In a country with 80% overall ART coverage, FSW ART coverage is estimated to be 68%, MSM 61% and PWID 59%, also with wide uncertainty. In SSA combined, KPs were estimated to be 1.1% of the adult (15-49 years) population, but 5.1% of adult PLHIV, with 500,000 FSW living with HIV (300,000 to 1,000,000), 300,000 MSM living with HIV (100,000 to 900,000) and 80000 PWID living with HIV (30,000 to 300,000).

Stevens summarised that there were large data gaps and large heterogeneity in observations within the same country, which resulted in substantial uncertainty in modelled regional estimates. These estimates should serve as a foundation to guide future surveillance activities and stimulate in-country conversations about data quality, data review and data processes. However, estimates are likely insufficiently precise for most key population planning, target setting and monitoring purposes.

Key points from discussion

- Compare PSE with estimates used in the Global AIDS strategy (Stover).
- Wolfgang Hladik raised the point that young MSM are overrepresented in IBBS surveys. Stevens said that HIV prevalence could be adjusted to account for this (using the age distribution in high income countries), but that it would be difficult to adjust the PSE.
- Mahiane asked why no data before 2010 were included. Most data are from post 2010, but since we are modelling estimates as cross-sectional (even though there were substantial changes in e.g., ART coverage over the last decade) it is perhaps more robust to not include earlier estimates.
- Eric Remera queried whether uncertainty in data of PSE can be used in the analysis (giving more weight to estimates with less uncertainty), but Stevens said most estimates do not have any uncertainty quantification, or if yes, it was not always clear how this was estimated.

In the next presentation **Rob Glaubius** described development of a **Goals age-risk model**. Currently, Avenir Health has two versions of the Goals models—one where heterogeneity in HIV transmission is driven by behavioural risk (KPs, low/medium/high), and the other by age. The new model Glaubius presented combines the two risk mechanisms. The model is still in prototype development stage. Age, sex, and HIV stage compartments are similar to AIM/Goals, with exceptions that it also stratifies PLHIV according to awareness of HIV status and includes compartments for PLHIV previously on antiretroviral treatment. Risk behaviour groups are: Not sexually active, MSM, FSW, PWID, transgender, and heterosexuals. The heterosexual population stratified by marital/cohabiting relationship status: never, currently, and previously.



Entrants into KP populations are determined by 1) proportion of population in KP, 2) the time in the group (1/turnover), and 3) the age distribution of the group. There is no explicit recruitment rate into any of the key populations; instead, a dynamic calculation on a year-by-year basis calculates the number of entrants required to ensure that size and age are stable over time. KP size are determined from data from IBBS surveys.

The force of infection has two components: transmission within marital/cohabiting partnerships, and within other partnerships. Data from DHS surveys are used to inform priors for numbers of partnerships and partner age differences. Mixing between the 12 groups (3 KPs and 3 risk groups for each sex) occur on a 'never mix', 'can mix' and 'prefer to mix' basis. Choosing values for this mixing matrix can be controversial, and regardless of how it is parameterised, the data may not be able to inform assortativeness parameters well.

Glaubius stressed that the model itself will not overcome the KP data sparsity in SSA, that compartmentalization of sexual identity and preferences is fraught and that moving towards a transmission model for the UNAIDS HIV estimates will require a consensus mixing matrix.

Key points from discussion:

- Excluding clients of sex workers as a KP means assuming they mix homogeneously with the general population, which may lead to quite different dynamics versus having explicit risk heterogeneity (Johnson, Eaton)

In the next presentation **John Stover reviewed simpler models** that have been used for estimating the distribution of infections and HIV transmissions from key populations: the Modes of Transmission (MoT) model and the Incidence Patterns Model (IPM).

The MoT model disaggregates Spectrum estimates of total prevalence and new infections into estimates for key populations using user input information on KP size, prevalence, and behaviours. All these inputs are not typically available, but users can put in numbers and adjust them to make the results sum to Spectrum totals.

The Incidence Patterns model (IPM) also fits to Spectrum estimates of incidence but uses DHS or PHIA survey data to disaggregate new infections into groups measured by the surveys (e.g., by circumcision status, by marital status), in the year of the survey. This is done in a formal Bayesian fitting process. KPs are an add-on, since data on these populations are typically not available from DHS/PHIA. Both MoT and IPM provide static estimates (for 1 year) and therefore do not estimate the role of KPs in transmission over time. These estimates could be improved by using information from studies, such as a Johns Hopkins meta-analysis result of a 13-fold risk of prevalent infection in FSW compared to all women in LMIC.

Stover suggested that instead of pursuing a dynamic version of IPM, we should consider using already developed dynamic models such as Goals or Optima, which have been applied to ~37 and ~18 countries in SSA, respectively. He showed an example case in which IPM, Goals and Optima results in the same country produced quite different estimates, but highlighted that inputs to these three were different. As an exercise, we should compare results when giving these models comparable inputs.

To address the limitation of IPM not being time-dynamic, **Oliver Stevens** proposed a synthesis between IPM and EPP, leveraging the demographic structure, age structure and dynamic compartmental modelling from EPP-ASM and the population-stratified IPM to produce KP estimates. This would better use available country KP surveillance data, compared to the current systematic, regional-level reviews extrapolated back to country estimates. The steps to do this would be:

- Calibrate EPP-ASM to household survey and ANC data to estimate a total population incidence and force of infection
- At each timestep, decompose force of infection by population group using mixing matrix between all population groups
- Estimate HIV incidence by population group with key population HIV prevalence and ART coverage as available (and by age)

This proposed dynamic approach integrates a time series of HIV prevalence in KP and changing population size over time. The model can quantify the network impact of KPs and KP transmissions over time and enforce consistency between HIV prevalence and incidence over time between all groups.

Stevens also gave an overview of the country **KP workbook** process developed last year for country HIV estimates teams to review and synthesise national KP survey data. Data collated and used in the analysis of his presentation were prefilled in a structured Excel spreadsheet and sent to countries to guide them to reach a consensus estimate of HIV prevalence, population size and ART coverage. Countries could enter additional data missing from the workbook use validation plots to compare their data with other countries in the region. The final sheet consolidates consensus country data-based estimates, and another comparing those consensus estimates to those of Goals and Optima country representations. If those estimates compare well, countries can choose to adopt the modelled output of new infections per KP from Goals or Optima. If not, Avenir could rerun Goals using the new country consensus estimates.

Unfortunately, only 7 countries completed the workbook this process in 2022 estimates, perhaps because the estimates process was completed virtually rather than guided through regional workshops. Users struggled with identifying the urban-rural ratios in KP population sizes.

After these presentations, the group split into four working groups to discuss the following questions:

- What strategies should we pursue for supporting countries to report key population estimates (size, PLHIV, infections, [ART coverage]) as part of the UNAIDS estimates process?
- Key population workbook: data review, report KP infections results from an external model (Goals, Optima, other locally used dynamic transmission model)
- Enhanced IPM model [static disaggregation of Spectrum infections]
- 'Dynamic' / time-changing IPM/MoT variant [disaggregation of Spectrum total population infections]
- Other ideas?

The main points arising from these discussions are captured in the recommendations in Appendix A.

Session 6 – On-ART mortality

Session 6 focused on considerations and potential changes to how mortality on ART is modelled in Spectrum estimates. The current approach models mortality on ART as a function of age, sex, CD4 category at ART initiation, and duration on ART (<6 months, 6-12 months, and >1 year). Mortality on ART is allowed to decline over time. Parameter estimates are derived from analysis of ART clinical cohort data from the IeDEA Network for global regions and the ART-CC collaboration for European regions. ART mortality estimates are adjusted for under ascertainment of deaths among patients lost-to-follow-up using data from tracing studies, with different assumptions across regions (see UNAIDS Reference Group meeting from October 2018 for the current approach and parameters).

Viral load suppression among persons on ART has been steadily increasing in recent years and transition to Dolutegravir (DTG)-based first line regimens are anticipated to improve viral suppression and treatment outcomes. These factors are not currently explicitly represented in the Spectrum mortality model, leaving some concern that contemporary AIDS deaths could be over-estimated, and country-specific progress are not reflected in the estimates of AIDS deaths outcomes.

There are also several regions, particularly Western Europe (WCENA), Eastern Europe (EECA), and southeast Asia (AP) where there are persistent challenges reconciling estimates of AIDS deaths from the Spectrum mortality model with observed data from HIV programmes or national vital registration. These discrepancies may be related to assumptions about derivation of the mortality parameters or region-specific adjustment for under-ascertained mortality among those lost to follow-up.

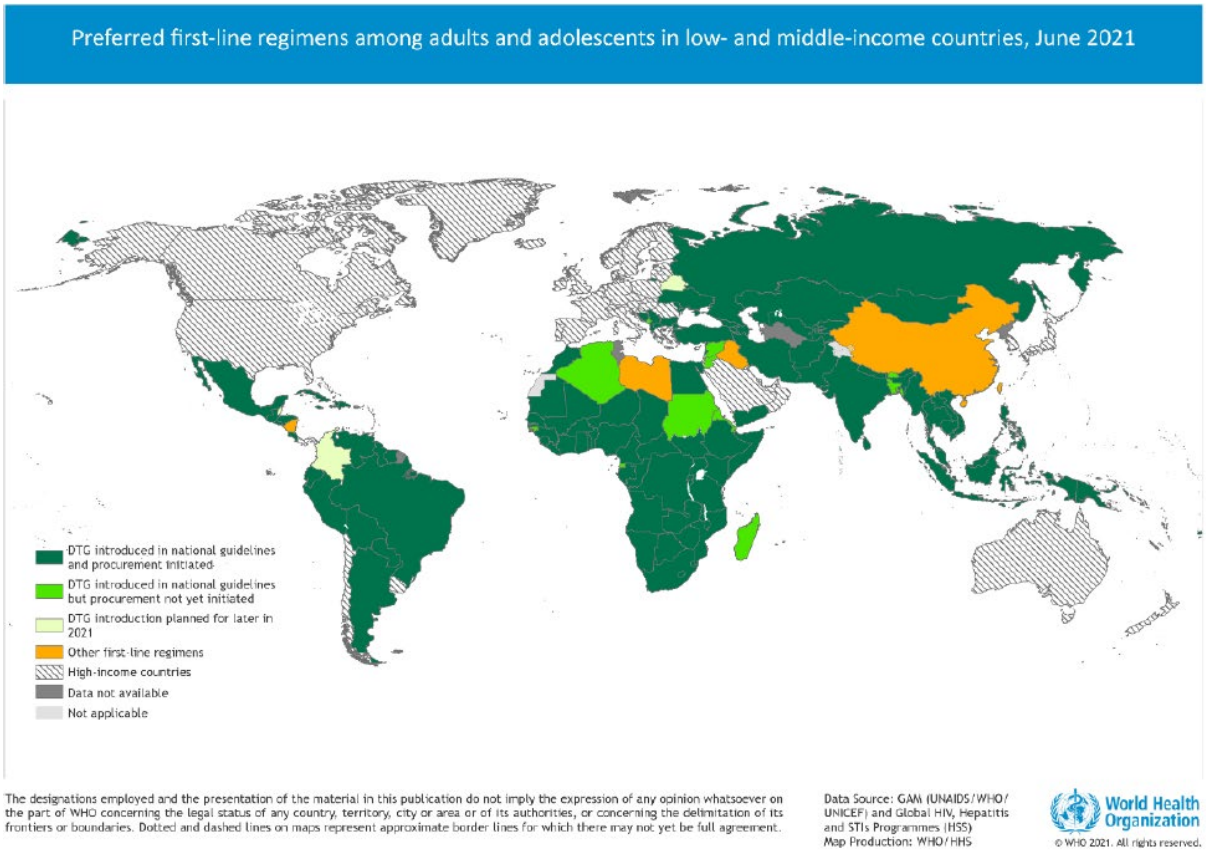
Session 6 addressed the following objectives:

- Are there sufficient data to move towards relating on ART mortality to VLS over time?
- Review evidence on the impact of transition to DTG on mortality
- Triangulate estimates of Spectrum deaths in WCENA, EECA, and AP with VR and programme data

Mary Mahy opened this session with an **overview of viral load (VL) monitoring globally**. In the Spectrum files, countries should enter numbers of PLHIV who received a VL test each year, and how many of those were virally suppressed. VL suppression from routine monitoring results are not used as national or in regional estimates if fewer than 50% of PLHIV on ART received a VL test. Only routine VL tests (excluding clinically indicated VL measurements), at least 6 months after treatment initiation, should be included in these numbers, but it is unclear to what extent data systems are able to distinguish routine versus clinically indicated viral load tests. Of 143 countries, 121 reported VL data between 2015 and 2020, but only 87 of those had VL tested in more than 50% of those on ART. Most countries reported increasing numbers of VL tests performed over time, with a decline in 2020 due to COVID. Percentages of those on ART who are virally suppressed (third 90) are consistently increasing in all regions, up to 96% in Asia and the Pacific. In countries that have not rolled out DTG, VL suppression (VLS) have also increased, except Libya.

In the next presentation **Marco Vitoria** from WHO gave an **overview of the global roll-out of dolutegravir (DTG)**. This drug leads to rapid VLS, increased tolerability, few drug interactions, only requires one pill a day, which all leads to better adherence and increased survival. This

drug has been adopted by 80% of PLHIV in 110 LMICs by mid-2021. The number is expected to increase to 94% by 2025.



Leigh Johnson presented results of a rapid **literature review of the relationship between VLS and on-ART mortality**. Currently, the UNAIDS estimation models assume that adult mortality rates after ART initiation depend on age, sex, baseline CD4, time since ART initiation and calendar year (informed by data from e.g., IeDEA and ART-CC cohorts). However, there is extreme variation in on-ART mortality across regions even when controlling for these factors. We are not currently adjusting for country variation in VLS.

The review aimed to assess whether there is evidence that mortality rates on ART vary systematically with viral load suppression, and therefore should be considered as a determinant of national ART mortality in the Spectrum ART mortality model. The updated IeDEA cohort analysis will investigate this (at aggregate level); the completed review assessed studies that measured the relationship between VLS and on-ART mortality at the individual level. Johnson found four studies that all controlled for at least age, sex, and CD4 count. All four found a significant, positive relationship between viral load and mortality. He a did a simulation for each study to estimate what increase in mortality per log increase in viral load would be needed to explain the observed relationship between viral load and mortality. The hazard ratio per unit change in log VL varies between 1.1 and 1.4, that is a 10 to 40% increase in mortality for each log increase in viral load. This relationship may be confounded by variables such as hazardous drinking, smoking, depression and hepatitis B and C, factors associated with both VLS and all-cause mortality. A main point from the discussion is that there is heterogeneity among studies in how they control for CD4 count (e.g., only at baseline and/or most recent, the baseline losing relevance after long treatment duration) which probably explains some of the variation across studies in observed associations between viral load and mortality.

Study	Effect of VL (aHR, 95% CI)	aHR per unit change in log viral load
May et al (2007)	VL <500: 0.31 (0.24-0.40) VL 500-9999: 0.49 (0.37-0.66) VL 10000-99999: 0.61 (0.45-0.82) VL ≥100000: 1	1.4
Brennan et al (2013)	VL <400: 1 VL ≥400: 1.8 (1.5-2.1)	1.2
May et al (2014)	For men aged 20-44, from 1 yr duration: aRR relative to VL <400 is 3.4 for CD4 <200, 1.9 for CD4 200-349, 1.9 for CD4 ≥350	1.2
Lee et al (2017)	VL <20: 1 VL 20-399: 1.12 (0.87-1.45) VL 400-999: 1.74 (1.10-2.74) VL ≥1000: 1.96 (1.56-2.46)	1.1

In the next presentation, **Reshma Kassanjee** and **Renee de Waal** provided an update on progress of their analysis of **on-ART mortality from recent IeDEA cohort data.** The previous analysis covered data from 2004-2017, and the new analysis will extend to 2021. Kassanjee briefly recapped the methodology of the unadjusted analysis (multivariable mixed effects Poisson regression controlling for region, sex, current age, CD4 at ART initiation, current ART duration and calendar time, with a random effect for heterogeneity between treatment programmes) and the adjusted analysis (same as unadjusted but simulating outcomes for those lost-to-follow-up based on data from tracing studies). The (adjusted) analysis including 2018-2021 data will also include data from a new tracing study and consider controlling for other variables to explain remaining variation in estimates within regions.

Renee gave a brief update on progress in receiving the data necessary for analyses. They are still awaiting IRB approval for data from North America but received and started cleaning data from 2 other regions (out of 7). Since they have not yet begun analyses, and know that the Reference Group needs results as soon as possible, they used their presentation time to ask questions on requirements from the analysis:

- Should they continue to exclude people who started ART younger than 15, like the previous analysis? [Not answered.]
- In previous analyses, people without a CD4 count at ART initiation were excluded, because these people were started for clinical reasons which might skew results. In recent years, however, up to 80-90% of people do not have a CD4 count at initiation. Should they continue to be excluded? [Do a descriptive analysis comparing mortality and other factors between those who have and don't have a CD4 count at baseline. The fact that this data is not collected anymore suggests we should change the structure of how we model On-ART mortality.]
- Should they include only the new tracing study data, or a combination of all? [Eaton asked for an explanation for the large difference in mortality for those in the new tracing data (~11%) vs ~20% in the older studies: are patients being initiated and traced healthier now? Nina Anderreg, who was involved in the studies, said that it may partly

be that the new study was a truly random sample (despite possible bias, as only 2/3 were traced) while the previous studies had convenience samples.]

- Should people who were initiated above the eligibility threshold still be excluded (applicable to the pre-treat-all era) since these thresholds are not meaningful anymore? Including them means that relationships between CD4 and mortality may change over time [Not answered.]
- Which additional indicators (e.g., VL) are useful to include, and how will the models use these? [Johnson and Kassanjee both thought to include a VLS suppression indicator at the programme level, Eaton thought at the individual level.]
- Since they're still waiting for data and just started cleaning the data they already have, would sending the unadjusted analysis be useful as it becomes available? [Previous analyses for Asia Pacific, North America and South Africa were not adjusted, so yes for those settings (Johnson). Eaton asked if we could learn something about trends in on-ART mortality from the unadjusted analyses; Johnson and Kassanjee commented there may also be a trend in tracing data, so these may offset each other.]

Adam Trickey described **patterns in causes of death among PLHIV on ART in the ART-CC cohort** (Europe and North America). The data comes from PLHIV who started triple therapy at age 16 or above from 16 cohorts. 70% of causes of deaths were ascertained by a clinician and an algorithm using ICD9/10 codes and confirmed by another clinician if discordant. All-causes and AIDS mortality significantly decreased from 1996-99 to 2016-19: from 16.8/1000py to 6.9/1000py and from 8.4/1000py to 1.0/1000py, respectively, both with consistent decrease throughout time.

		Adjusted mortality rate ratios [MRRs] (95% confidence intervals) vs 1996-99					Continuous
Cause of death	N Deaths	2000-03	2004-07	2008-11	2012-15	2016-19	MRR per 4 years
Overall	12,468	0.90 (0.81-1.00)	0.67 (0.61-0.75)	0.51 (0.46-0.56)	0.42 (0.38-0.46)	0.34 (0.31-0.38)	0.79 (0.78-0.80)
AIDS	3,016	0.92 (0.79-1.07)	0.62 (0.53-0.72)	0.41 (0.35-0.48)	0.29 (0.25-0.34)	0.21 (0.18-0.25)	0.71 (0.69-0.72)

Trickey also presented **associations between modern first-line regimens with all-cause mortality in the ART-CC cohorts and the UK-CHIC cohort** (16 cohorts included in analysis). Integrase strand inhibitor (INSTI)-based vs non-INSTI-based ART regimens were compared for prognosis regarding 1) virological failure and 2) all-cause mortality. Adjusted hazard ratios for each 3rd drug comparison showed significantly higher mortality only for raltegravir compared to other drugs; whereas only dolutegravir had significantly better VLS than the other drugs studied. Another analysis considered whether unsuppressed viral load or low CD4 count at 1.5 years after initiation predicted subsequent mortality, and the hazard ratio for this relationship was 2.5 (2.1-3.2).

In the next presentation **Eline Korenromp** and **Kelsey Case** highlighted that **countries in Western Europe had to use Spectrum's multiplier that scales down mortality on ART,** to match VR deaths reported or their numbers of PLHIV. The multiplier fitted ranged from 0.85 in Sweden to 0.3 in Greece. Alternatively, for some countries the needed lower on-ART mortality was obtained by replacing the high-income country (HIC) mortality pattern by the Asia pattern, which has lower rates, notably for the highest CD4 count category. It seems counter-intuitive that Asia has lower on-ART mortality rates than HIC.

Tim Brown followed with a presentation highlighting on-ART mortality issues in the Asia-Pacific region (EPP: Fiji; Asian Epidemic Model (AEM): Malaysia, Mongolia, Philippines, Sri Lanka and Thailand). He compared Spectrum estimated AIDS deaths to WHO raw numbers and GBD 2020 estimates. In Malaysia and Thailand, Spectrum estimates are in similar range as GBD,

but trends do not match. In Philippines and Fiji, GBD estimates are 5-10 times higher, whereas in Sri Lanka and Mongolia, Spectrum estimates are above GBD.

Of note, 5 countries (Cambodia, Lao PDR, Philippines, Thailand and Viet Nam) used **on-ART mortality multipliers ranging from 1.5-6.0** to match in-country on-ART mortality data. I.e., Spectrum's Asia default rates were too low for Asia. Also, in the Philippines, off-ART deaths exceeded on-ART deaths, and this may require further investigation. Brown suggested to add an output to Spectrum that shows all-cause mortality on-ART (instead of only total AIDS deaths), since this is data that countries are collecting.

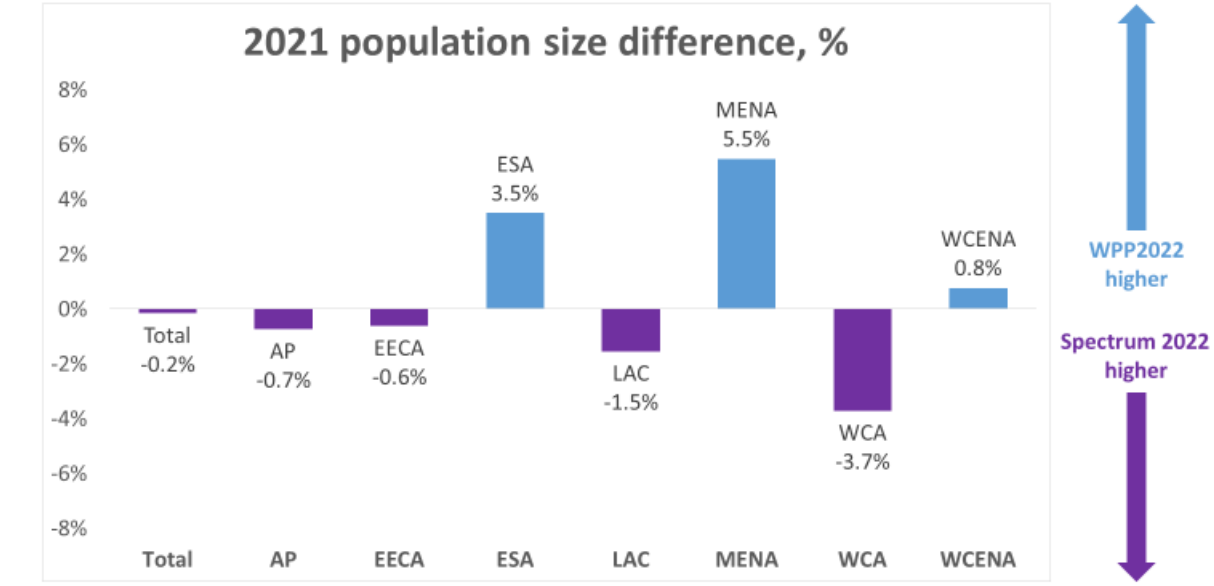
The main points arising from the discussions following these presentations are captured in the recommendations in **Appendix A.**

Session 7 – Population estimates

In this session, **Patrick Gerland presented on the methodological updates for the UN World Population Prospects (WPP) 2022**. The main changes that may directly influence the Spectrum model or operations are listed here:

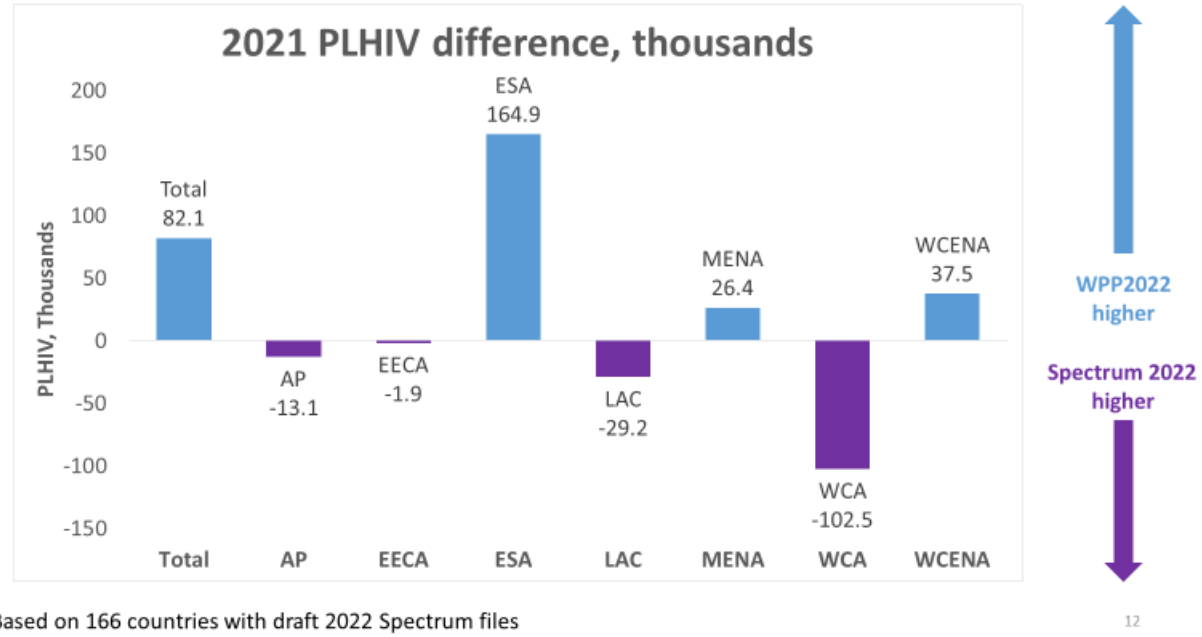
- Transition from the historical practice of estimating for 5×5 age groups and periods to a framework by single year of age and 1-year periods of time – obviating the earlier need to interpolate WPP's 5×5-year age groups to 1×1-year inputs for Spectrum.
- Enhanced data portal may make it possible for Spectrum Web to directly draw population data in (since no interpolation/post-processing is required any more).
- This version of WPP standardised the demographic impact of 'mortality crises', such as mass killings (genocide, war, etc.), natural disasters (floods, tsunamis, earthquakes), famine and COVID-19.
- UNDP switched from providing mid-year estimates, to estimates at the beginning of the year.
- WPP still lacks estimates at sub-national level.
- WPP 2022 extended fertility to 10–14- and 50–54-year-olds.

Rob Glaubius followed with a brief **comparison of Spectrum 2022 and preliminary WPP 2022**, based on provisional WPP estimates of total population sizes (not by age and sex) shared by Gerland. In some regions, WPP 2022 estimated total population sizes are up to 5.5% higher than Spectrum, but in other regions up to 3.7% lower.



Glaubius briefly examined possible reasons for large discrepancies in the 5 countries with the highest positive or negative *absolute* changes. In China (-2.6%), Mexico (-7.8%) and Pakistan (+2.8%), more recent census data contributed to the changes in WPP 2022. In DRC (-18.2%), Mexico (-7.8%), Bangladesh (-3.6%), the US (+3.7%) and the UAE (+1721.8%), the Spectrum files used pre-2019 WPP revisions. For UAE, net migration was manually set to zero in Spectrum (overwriting WPP's net immigration, as a way of estimating HIV for only the citizen population). Other countries with large differences were Ethiopia (+14.7%) and Nigeria (-2.6%), whose Spectrum total is the sum of sub-national files, which were not closely related to

WPP. For some countries with large *proportional* differences—Serbia (-16.6%), Comoros (-8.3%), Malta (+16.3%) and Central African Republic (+11.8%)—Spectrum matched WPP 2019 well, but not WPP 2022 and we will have to examine these files when updated.
Glaubius also explored effects on national PLHIV estimates, as the product of all-age HIV prevalence and Spectrum versus WPP 2022 all-age population estimates (i.e., assuming that age distribution is similar in WPP 2019 and 2022). WPP 2022 population size estimates would raise 2021 global PLHIV estimates by 0.2% (as a crude estimate), but with varying effects (up or down) by region.



While some large differences are due to WPP revisions, others reflect the use of subnational Spectrum files, Spectrum files using pre-2019 WPP inputs, or countries having overwritten WPP inputs into Spectrum to change the scope of the estimation from de jure to de facto population or to match official population estimates and projections produced by the relevant National Statistical Office.
The main points arising from the discussions following these presentations are captured in the recommendations in **Appendix A.**

Session 8 – UNAIDS estimates processes

Eline Korenromp presented ideas about **improved guidance for CSAVR users**. The presentation focused on:

1) Which mortality data to use (e.g., original, adjusted GBD),
2) When to fit to CD4-at-diagnosis data,
3) When to fit sex and/or age IRRs on sex and age patterns in reported diagnoses and deaths,
4) When to adopt CSAVR-fitted IRR in AIM, and
5) When to override case data-based KOS with CSAVR's KOS estimate.

When to fit CD4-at-diagnosis data: 14/59 countries had data on the distribution of CD4 at diagnosis and included them in CSAVR fitting. In none of these countries (which all also inputted death data) did CD4 data made a noticeable difference to incidence estimates.

Key points in discussion:

- The lack of influence of CD4-at-diagnosis on CSAVR fits is troubling and may indicate an undesirable rigidity in mortality assumptions. In countries with very high treatment coverage and low AIDS mortality, it is undesirable that AIDS mortality strongly influences the incidence estimate. Rather, incidence should be driven by new diagnoses and CD4 at diagnosis. Refining CSAVR to reconcile and balance between mortality and CD4 data is a priority for model development. [Eaton]
- CD4 at diagnosis data may be biased if only a selected (non-random) part of ART patients had CD4 measured and reported at diagnosis – explore applying CD4 completeness thresholds. [Glaubius]

The recommendations in **Appendix A** captures the decisions made on additional analyses required to answer this question.

W*hich mortality data to use:* the GBD 2020 data consists of country-reported vital registration AIDS deaths plus the IHME's estimate of AIDS-attributable deaths misclassified to other causes (including 'garbage codes'). Thirty countries used the GBD adjusted data, which UNAIDS recommended to all; whereas 14 opted to use their own (unadjusted) vital registration data. A problem with the GBD dataset is that, despite annual updating, the data inputted (after countries report these to WHO) lag behind the vital registration data that the countries have access to. Some countries therefore used GBD for years with such an estimate, supplemented with original VR data for more recent years ('hybrid' dataset).

Also discussed was the UNAIDS rule, adopted in 2022 estimation round, to publish incidence trends (2010-2021 this year) from CSAVR estimates, only if this had including death data from 2019 or later in fitting. Model developers commented that the recent incidence trend is not specifically influenced by deaths in recent years, so lacking recent death data should not prevent publishing incidence estimates. UNAIDS accepted to relax this rule, still just for the 2022 round of estimates.
These and related discussions are captured in the recommendations (**Appendix A**).

Korenromp then talked about **fertility adjustments among HIV positive women in concentrated epidemic settings**. The strong default fertility reduction, based on data from Sub-Saharan Africa, with relatively higher overall fertility, in many concentrated epidemics results in too few pregnant women with HIV and PMTCT and/or paediatric ART coverage estimates of over 100%. Countries that had and used ANC prevalence data remedied this by fitting and applying

a >1.0 ‘local adjustment factor’; and some countries lacking ANC prevalence data simply set a ‘local adjuster’ to >>1.0, sometimes at the maximum of 2.0, to prevent >100% PMTCT. Actions to refine HIV-related fertility effects for these settings are captured in the recommendations (**Appendix A).**

Next, **Jeff Eaton** discussed potential **new outputs for Naomi-based, district-level ART target setting**. Currently, Naomi outputs PLHIV by district (where people reside), as well as people receiving ART in the district irrespective of whether they reside there (for example, people may not reside in Gaborone, yet still receive ART there). Eaton suggested to output PLHIV by the district attended, so that treatment gaps can be calculated by the district where people seek treatment. He also suggested to output results not only 9-months ahead (which countries need for PEPFAR COP planning), but also for 21 months ahead (the end of the COP target period).

Eaton then discussed **the interpretation of Naomi district-level ANC outputs**. Spectrum focuses on modelling population-level birth outcomes (live births, births to women with HIV, ART to pregnant women with HIV) which is required for estimating MTCT rates. Modelling (P)MTCT rates and numbers at district level will require reconciling Naomi’s ANC data with national-level Spectrum inputs. Naomi currently, for its short-term projections for programmatic planning, simply calibrates to total ANC clients, numbers HIV positive and numbers on ART – without reconciling these with births, paediatric infections, or MTCT at district level. This causes confusion for users that the Naomi ANC outcomes don't match birth outcomes in Spectrum, and secondly, it means that we don't have any district outputs for PMTCT coverage.

Eaton posed two questions to the group:

1) Should Naomi ANC outputs be calibrated to match Spectrum’s birth and PMTCT outcomes? (i.e., the sum of all districts’ ANC clients matches total births, ANC HIV-positive match births to women with HIV) Given the problems with ANC-RT data discussed in previous sessions, we will revisit this question after data quality assessments and implied adjustments to Spectrum input data are completed.
2) Should Naomi ANC outputs be extended to include PMTCT? There is no current modelling strategy to do this. Of note, most (sub-Sahara African) districts with a Naomi estimate nowadays have very small PMTCT ‘gaps’ (fraction of HIV-positive ANC clients not accessing PMTCT), relative to the large uncertainties in fertility and ANC attendance.

Next, **Ian Wanyeki summarized current uses and successes of the AIDS data repository** (ADR) system. He suggested some enhancements that are captured in the recommendations **Appendix A.**

Taavi Erkkola gave an overview of **new developments in capacity building led by UNAIDS**. A new team based in Nairobi of 5 people with different roles will take on capacity building globally. Regional workshops will resume, but adding focus on: 1) inequalities, 2) communicating (possibly changed) results, and 3) triangulating across data and information sources (including policies). The new team will intensify efforts to enable country teams to produce estimates themselves, with reduced input from facilitators. Erkkola showed an example schedule of a series of 1-week workshops and a draft agenda for a SSA workshop that includes the three focus areas above.

Sonia Arias Garcia presented on **key populations data availability in SSA, reported through the Global AIDS Monitoring (GAM) tool**. For three key indicators (population size estimates, HIV

prevalence and ART coverage), most countries had weak or no data for all four key populations (FSW, MSM, PWID and transgender people).

Keith Sabin next presented on '**Developing and piloting a simplified biobehavioural survey methodology for key populations**' or BBS-Lite and its applications in Uganda and Georgia.

Appendix A

Recommendations

Recommendation	Lead person(s)	Timeline
Session 1: Review of ANC testing and PMTCT inputs to Spectrum in UNAIDS 2022 Estimates and proposed approach to adjusting ANC testing inputs		
• Add outcome 'total live births in health facilities' and 'known HIV negative' to ANC testing inputs in Spectrum/Naomi	Avenir Health, Imperial College, Fjelltop	October 2022
• Review Zambia Spectrum estimates before and after data quality assessment (DQA) adjusted ANC testing data inputs, to understand impacts of data improvements	UNAIDS	October 2022

• Support countries to conduct DQAs using supervisory visits of their routine ANC and PMTCT data before 2023 HIV estimates round and on an annual basis as a routine surveillance activity.
  ○ From a representative sample of health facilities, to avoid the limitations of the non-representativeness and resulting biased estimates from the ANC surveillance that we used to have.
  ○ To provide more reliable estimates of HIV prevalence and ART coverage in pregnant women
  ○ Quantify reporting completeness in the routine ANC data
  ○ Focus on understanding the causes of the data quality problems and how to solve them, rather than just diagnosing the problems.
	UNAIDS, Working group	Working group first meeting: June 2022
		Present progress: October 2022
		Use results: 2023 estimates

• Develop a simple "screening tool" to identify health facilities in which there are likely to be data quality problems, based on metrics such as completeness of data, variability of indicators over time, consistency of indicators (ratios <100%). ○ Validate screening tool using Rwanda classification of facility reporting from recent data quality review exercises.	TBD	May 2023
• Consider whether facility births or ANC1 attendance might be useful in disaggregating national-level WPP fertility estimates down to sub-national level to improve estimates and target setting for ANC and PMTCT outcomes.	TBD	May 2023

Recommendation	Lead person(s)	Timeline
• Develop an algorithm on weighting of ANC routine testing (ANC-RT) data in EPP estimation, and a (possible) within-model adjustment: ○ Depending on results of DQAs, determine an optimal statistical approach: in which situations should we not fit ANC-RT data at all, down-weigh and/or adjust the data before fitting, or model the bias due to over-counting HIV-negative and/or HIV-positive pregnant women as part of the calibration process.	UNAIDS, Working group	October 2022
• Continue to co-ordinate with WHO around plans to introduce individual-level data systems, as long-term resolution to improved routine programme data quality.	UNAIDS/WHO	
• Produce Viewpoint on priorities for HIV surveillance in sub-Saharan Africa during the transition to sustained epidemic control monitored through routine and individual-level data for HIV surveillance.	UNAIDS Reference Group Secretariat	2023 journal supplement
• UNAIDS to publish (and share with countries) guidance on where to focus their surveillance efforts with a special focus on what is needed for models.	UNAIDS	July 2022
• Remove the HIV-related fertility adjuster data editor, to instead read-in ANC prevalence and denominator from the (national-level) ANC testing editor	Avenir Health	October 2022
Session 2: CSAVR		

• Input and visualise HIV testing data (annual numbers tested and positivity) in CSAVR.
  ○ Consider whether a consolidated HIV testing data editor in AIM could be used for both Shiny90 and CSAVR HIV testing data inputs.
  ○ Collate testing data (2022 Spectrum files, GAM, European Centre for Disease Control (ECDC), by sex) to prepopulate 2023 CSAVR files
	Avenir Health
	
	UNAIDS
		October 2022

• Explore using data on HIV testing volume and positivity in CSAVR fitting, to distinguish effects on diagnoses and knowledge of status from possible changes in new infections versus changes in testing effort.	Avenir Health	May 2023
• Enable CSAVR data editors to retain data on diagnoses, deaths and CD4 for all years, but	Avenir Health	October 2022

Recommendation	Lead person(s)	Timeline
selectively exclude some data points from model fitting (e.g., data of poor quality or completeness). • Allow for entry and saving of mortality data from multiple sources (original VR, misclassification adjusted GBD 2020, etc.) in parallel, and choose one – or a user-defined hybrid – for fitting	Avenir Health	October 2022
• To avoid people on ART outnumbering PLHIV with knowledge of status, CSAVR should send to AIM proportions of PLHIV *not on ART* knowing their status, rather than overall proportions knowing their status.	Avenir Health	October 2022
• Migrants: o For years that users input HIV-positive migrant data into AIM, AIM should not apply its background calculation of net HIV-positive migration – but visualize and smooth transitions between years with and without user-inputted migrant data. o Expand user guidance on the best ways to handle HIV-positive migrants (new and known positives) in AIM and CSAVR and standardize the scope of epidemic estimates across low-HIV, high immigration countries to be the de facto (not de jure) population, for programmatically relevant national estimates, valid regional sums and alignment with WPP 2022. One common data editor used by AIM and (optionally) CSAVR; with option restored to enter immigrants without age/sex disaggregation and develop default patterns of the distribution of HIV-positive migrants by age and sex using data from countries that have it (Netherlands, Chile, Norway, Ireland). Failing that, use the age/sex distribution in the resident HIV population.	Working Group	Working Group first meeting: August 2022
Session 3: Testing and treatment churn		
• ART data quality assessments: o Strongly encourage countries without recent DQA to conduct these of their ART data and propose standardized methodology for DQAs o Stratifying DQA results by sex/age/location is important.	UNAIDS Avenir Health	2023 estimates

Recommendation	Lead person(s)	Timeline

• Explicitly record and represent DQA results in Spectrum
  o Input year and result of DQA in Spectrum
  o Show and store both reported ART data and adjusted ART data time series after DQA
	Imperial	October 2022

• Develop example of using DQA results to adjust spatial ART by district inputs in Naomi		October 2022
• For countries where number on ART is greater than PLHIV: o Test incorporating survey ART coverage data (by age/sex) into Spectrum incidence rate ratio fitting tool and allow users to fit to multiple data sources in turn, and select.	Avenir Health	October 2022
o Explore rejection of EPP fits where ART coverage is over 100%, i.e., require PLHIV estimates to be above reported ART numbers. This will lead to higher national incidence estimates and PLHIV, which will force users to 'reckon' with program data.	Imperial College	October 2022
• Visualise 'waterfall' results estimated by Spectrum versus reported data on ART initiation and interruption	Avenir Health	October 2022

• Further disaggregate HIV testing data requested by Shiny90:
  o self-testing
  o index testing
  o HCT by sex / age (GAM age groups)
	Avenir Health, McGill, Imperial, Fjelltop	2023 estimates

• Explore country examples of electronic medical records (EMR)-based data input to models: compare outputs when using aggregated vs individual level data as inputs. Prepare guidance document. o Proposed countries: Kenya, Tanzania, Namibia, Botswana, and eSwatini	UNAIDS	October 2022
Session 4: Key population stratified estimates in concentrated epidemics		

• Continue developing CSAVR-KP and:
  o Review model fits to countries proposed by Tobi with better-quality case data by modes of transmission
  o Triangulate data sources and/or consider imposing model assumptions to mitigate
	Guy Mahiane, Tobi Saidel, Working group	Working group to meet July 2022. Preliminary results

Recommendation	Lead person(s)	Timeline
under-reporting of modes of transmission (e.g., sex ratio in diagnoses among heterosexuals, MWID and/or FWID) o Explore the sensitivity of model outputs to each data source included o Allow for time-changing key population sizes (as an exogenous user input rather than fitted – as in EPP) o Consider informing modelled sub-population distributions in incidence and diagnosis by sub-population-specific testing numbers or rates (Bulgaria, Ukraine, Hong Kong). o Visualize the impact of group turnover on results.		October 2022.
• Develop guidance for countries on documenting test volumes and case diagnoses by mode of transmission, for Spectrum estimates and more broadly for meeting strategic priorities.	UNAIDS	2023 estimates
• Modes of transmission data quality may vary within countries between different partners and organizations. CSVAR currently handles this via priors on model parameters. These assumptions may suffice to address the issue, but assumptions should be reviewed for appropriateness and validity (use the data-rich countries that Tobi identified).	Guy Mahiane	October 2022
• EPP-ASM for KP: o Complete and document demonstration; explore impact on EPP/AIM incidence-prevalence adjustment. o Incorporate in future integrated model, not EPP 2022-23.	Deepa Jahagirdar	May 2023
Session 5: Key population estimates in sub-Saharan Africa		
• Workbook-based data collation and review should continue as part of the annual estimates process		
• Consider adding clients of sex workers as a KP (at least in outputs for distribution of modes of transmission); review other population groups and definitions	UNAIDS, Imperial College	October 2022
• Review outputs of workbook data submitted in 2022 round against Goals and Optima estimates; Goals and Optima to refit to workbook data	Avenir Health, Imperial College, Burnett Institute	October 2022
• Consensus not to recommend further development of static / cross-sectional models for routine use in		

Recommendation	Lead person(s)	Timeline
HIV estimates process (e.g., Modes of Transmission Model, Incidence Patterns Model) as these approaches do not capture progress in preventing transmission among key populations.		
• Develop TOR for proposals for further development of the dynamic IPM concept as progress towards the Symphony model that consolidates all data sources, transmission dynamics, burden estimates, and prevention and clinical impact, across multiple countries and epidemic types.	Reference Group Secretariat	TOR: January 2023 Proposals: May 2023
Session 6: On-ART mortality		
• Create an additional Spectrum output for all-cause deaths among persons on ART (complementing the existing, HIV/AIDS-related deaths).	Avenir Health	2023 estimates
• With corresponding validation plot, comparing this Spectrum estimate with country-inputted all-cause deaths on-ART deaths.	Avenir Health	2023 estimates
• Mini literature review suggests some evidence for relationship between log VLS and mortality on ART. It is uncertain if this translates to meaningful change in mortality for VLS change from e.g., 85% to 92%. o Do simulation exercise to explore this. o Further follow-up on ART-CC regression for more information on relationship between log VLS and mortality	Leigh Johnson	October 2022
• ART-CC collaboration data suggest that in high-income countries DTG improves virologic outcomes compared to efavirenz, but not mortality. However, generalisability of this to other regions including sub-Saharan Africa is uncertain. o Insufficient data to draw conclusions about mortality changes in mortality due to VLS and DTG. o No change in model structure now, but continue to monitor evidence, as a priority.		
• IeDEA: Review preliminary analyses for updated global mortality rates, with priority for: o Mortality trends during the past 5 years (since previous IeDEA update of Oct 2018). Review preliminary analyses to reach interim recommendation on default parameters about mortality rates on ART over time, for Spectrum 2023. o Mortality among cohorts with and without baseline CD4 measurement, to assess	Reshma Kassanjee	October 2022

Recommendation	Lead person(s)	Timeline
representativeness of available data on CD4 at ART enrolment.		
• Recent AIDS cause of death data from high-income countries implies systematically lower ART mortality than Spectrum defaults o Updated Trickey et al. analysis suggests continued decline in ART mortality; review Spectrum assumptions about mortality time trend o Review CD4 >500 mortality rate in Spectrum o Review whether Spectrum mortality rates reflect 'excess mortality' or 'HIV/AIDS mortality' from ART-CC o Consider separately modelling excess non-AIDS mortality and AIDS mortality	Leigh Johnson, Adam Trickey, Eline Korenromp, Avenir	July/August 2022 Presentation at October meeting
• Asia-Pacific: Several countries *substantially* increase ART mortality rates to match programme data o Review IeDEA mortality rates stratifying cohorts linked or not to deaths from vital registries, to identify any potential under-ascertainment bias affecting IeDEA-based mortality rate assumptions for Asia-Pacific region o Consider revising default ART mortality assumptions for Asia region accordingly.	Reshma Kassanjee, East-West Centre	October 2022
• Working group to prepare suggestions for required data and model structure	Working group	Working group first meeting: August 2022
Session 7: Population estimates		
• Countries should update Spectrum demographic inputs to WPP 2022 as default option unless there is specific reason not to. Countries that do not adopt WPP 2022 should review demographic inputs to ensure they are current.	UNAIDS	2023 estimates
• Countries using subnational Spectrum files should compare the sum of populations across their files with WPP 2022 and consider to re-align subnational demographic inputs to match WPP 2022 totals.	UNAIDS	2023 estimates
• Countries with large changes to population in WPP 2022 may require additional support to investigate and explain the drivers of those changes	UNAIDS	2023 estimates

Recommendation	Lead person(s)	Timeline
• DemProj: Consider change from mid-year to beginning of year, similar to WPP 2022.	Avenir Health, Imperial, East-West	First meeting: August 2022 Preliminary results: October 2022
• Explicit statement on move to de facto population for UNAIDS estimates, with asterisks noting countries where a different population concept is adopted	UNAIDS	2023 estimates
Session 8: Processes		
When to include CD4 count data in CSAVR fit • Preliminary recommendation: always fit CD4 count data for years these meet a completeness threshold (proposed: 80-95%) • Sensitivity analysis on threshold, and of biases related to less or more selective CD4 testing to explore whether these can explain why CSAVR typically estimates higher-than reported CD4. • Explore why CD4 at diagnosis often does not affect model fit when mortality is included: because the model mortality is rigid, limiting inferences about time or regionally varying relationships between diagnosis and mortality. o Proposed simulation study to understand model fit in cases where the assumed ART mortality and ART coverage data were severely mis-specified. • Review Croatia as an example where CD4 at diagnosis did affect model fit.	Avenir Health	October 2022
CSAVR sex/age IRR fitting • The default expectation is that countries will use CSAVR's sex and age IRRs, when fitted • If fitted IRRs are not used (for example, if resulting in KOS inconsistent with PLHIV or ART), **CSAVR should be refit with custom-set IRRs,** such that the final incidence fit is consistent with the IRRs used by AIM. • Similar to guidance for EPP, which needs to be refit if IRRs have changed a lot. This is likely more important for CSAVR, because CSAVR fits sex/age stratified data, which EPP does not.	UNAIDS	2023 estimates

Recommendation	Lead person(s)	Timeline
HIV/AIDS-related deaths data source(s) for CSAVR • Choice of mortality data source should be based on review of data sources **before CSAVR model fitting** and not selected based on comparing CSAVR model fits to alternative sources. • GBD misclassification-adjusted vital registration outputs remain the recommended primary source for AIDS deaths inputs. However, adjustments should be reviewed for plausibility prior to input to model and triangulated with other HIV epidemiology data sources and local expertise, especially in countries with large, stipulated proportions of missing cause of death / garbage codes. • In cases where GBD adjusted death data are not plausible or consistent with other HIV surveillance sources, this should be fed back to the IHME HIV and CoD team, for consideration in future revisions.	UNAIDS	2023 estimates

Mortality data availability criteria for publishing CSAVR-based incidence
• Availability of HIV/AIDS deaths data in period 2019-2021 should not be used as a threshold criterion for determining whether to publishing current HIV incidence estimates.
  ○ This is because there is not expected to be a strong signal about recent infection trends specifically from more recent AIDS deaths data, because (1) AIDS deaths primarily reflect infections that occurred at least 8-15 years ago, and (2) in the ART era, the relationship between AIDS deaths and incidence depends on many factors.
  ○ We recognize (1) the importance of implementable criteria for ensuring minimum data to inform trend estimates, and (2) the importance of avoiding suppressing estimates except when absolutely necessary.
	UNAIDS	2023 estimates

Choice of CSAVR or EPP • Choice of model should be based on data availability and quality (over time and by sex or KP) and plausibility of results (e.g., coherent cascades over time and across groups), not precedents of past estimates or country preference based on comparative results.	UNAIDS	October 2022

Recommendation	Lead person(s)	Timeline
Triangulation of both models is recommended, if capacity allows, to assess directions and magnitude of uncertainties. We also recognize the challenges explaining large year-on-year changes in country estimates if data updates do not by themselves necessarily justify this. • ACTION: Refine guidance on model choice and triangulation, considering data availability, using country case studies from 2021 and 2022 estimation rounds.		
Fertility rate ratio local adjustments • Fertility rate ratio local adjustment substantially above 1.0 in many LAC, CAR, and EECA countries either by fitting routine ANC testing prevalence or user-set to avoid PMTCT and/or paediatric ART coverage of more than 100%. • However, there is a concern that ANC testing is risk-based in many concentrated epidemic settings and therefore fitting to routine ANC prevalence may over-state true prevalence among pregnant women. Conversely, higher-risk women may be under-sampled if they avoid first ANC visits due to stigma. • Also, many countries have PMTCT and paediatric ART program numbers declining faster than Spectrum-estimated PMTCT and paediatric ART need, leading to recent or ongoing drops in estimated coverage often felt to not reflect the reality. • The fertility rate ratios in Spectrum are derived in sub-Saharan Africa settings with high fertility. These are almost certainly not generalizable to low fertility settings. John Stover proposed changing FRR parameters to 1.0 instead of current for all age groups in concentrated epidemic settings, but unsure if this reasonable. • Recommendations: o Test impacts of changing FRR to 1.0 for low fertility settings. o Triangulate data on paediatric HIV diagnoses and deaths, with Spectrum-modelled MTCT and child PLHIV in concentrated epidemics, to see the magnitude and direction of divergence. [Note: previous analysis by Amy Zhang found large discrepancies between paediatric AIDS deaths from vital	Avenir Health, UNAIDS	October 2022

Recommendation	Lead person(s)	Timeline
registration and predicted AIDS deaths from Spectrum; worthwhile updating with more recent data].		
Naomi outputs • Change to using 'plhiv_attend' indicator representing the estimated number of PLHIV in catchment of ART facilities for a district as denominator for target setting in PEPFAR Data Pack. • Project PLHIV forward by 21 months (end of COP planning year) for Data Pack target setting in addition to 9 months (start of COP planning year) and 12 months (calendar year). • No change to ANC testing outputs from Naomi. Provide users clearer guidance about the purpose of Naomi ANC testing outputs, and how and why these differ from Spectrum national ANC results. Discuss at paediatric Reference Group meeting in October 2022. • Review population data sources and consider any relevant updates. • Review availability and quality of subnational VLS data.	Imperial College	October 2022
AIDS Data Repository/Navigator • Move workbook KP data to ADR • Ensure shiny90 is interoperable with Spectrum online • Consider future updates to manage the correspondence with countries • Also consider adding all data to ADR in a csv file and pulling it into the respective models from the same resource sheet (CSV file) • Explore pros and cons of moving CSAVR estimates to ADR.	Avenir Health, Fjelltop	October 2022
KP data • Request KP data before the workshop along with data review cycle	UNAIDS	2023 estimates

Appendix B

Participants

Name	Organisation
Adam Trickey	ART-CC, Bristol University
Adrien Allorant	McGill University
Cari van Schalkwyk	SACEMA
Chibwe Lwamba	UNICEF
Debra ten Brink	Burnet Institute
Deepa Jahagirdar	Stanford University
Eline Korenromp	UNAIDS
Eric Remera	RBC Rwanda
Faikah Bruce	SACEMA
Guy Mahiane	Avenir Health
Harriet Nuwagaba-Biribonwoha	ICAP Swaziland
Hmwe Kyu	IHME
Ian Wanyeki	UNAIDS
Irum Zaidi	PEPFAR
Isaac Taramusi	MoH Zimbabwe
Jeff Eaton	Imperial College London
Jinkou Zhao	GFATM
John Stover	Avenir Health
Josh Salomon	Stanford University
Juliana Daher	UNAIDS
Katharine Kripke	Avenir Health
Keith Sabin	UNAIDS
Kelsey Case	consultant
Laura Porter	CDC
Leigh Johnson	University of Cape Town
Leigh Tally	CDC Zambia
Luisa Frescura	UNAIDS
Marco Vitoria	WHO
Mary Mahy	UNAIDS
Mathieu Maheu-Giroux	McGill University
Mehran Hosseini	GFATM
Melissa Arons	CDC
Monita Patel	CDC
Nina Anderegg	University of Bern / University of Cape Town
Noah Bartlett	(CENSUS/POP FED)
Oli Stevens	Imperial College London
Parviez Hosseini	PEPFAR
Patrick Amanzi	USAID Zambia

Name	Organisation
Patrick Gerland	UN Pop Division
Peter Ghys	UNAIDS
Ray Shiraishi	CDC
Renee De Waal	CIDER, University of Cape Town
Reshma Bhattacharjee	USAID
Reshma Kassanjee	CIDER, University of Cape Town
Rob Glaubius	Avenir Health
Rowan Martin-Hughes	Burnet Institute
Sadhna Patel	CDC
Samuel Dupre	(CENSUS/POP FED)
Shona Dalal	WHO
Shufang Zhang	GFATM
Sonia Arias Garcia	UNAIDS
Suzue Saito	ICAP/Columbia
Tim Brown	East West Center
Tim Wolock	Imperial College London
Tobi Saidel	Consultant
Victor Kabwe	PEPFAR Zambia
Will Probert	WHO
Wolfgang Hladik	CDC
Yoko Shimada	Consultant

Appendix C

Agenda

All times are GMT+2 (Glion, Switzerland)
Day 1:

Time	Duration (mins)	Topic	Presenter(s)/ Lead Discussant
9.00	10	Welcome and introductions	Mary Mahy
9.10	10	Meeting objectives	Jeff Eaton
9.20	40	2022 Estimates review and challenges arising	Mary Mahy
Session 1: Review of ANC testing and PMTCT inputs to Spectrum in UNAIDS 2022 Estimates and proposed approach to adjusting PMTCT inputs (chaired by Leigh Johnson) **Objective:** • **Formulate guidance for countries with first ANC visits exceeding number of births and numbers of women receiving PMTCT exceeding estimated number of pregnant women with HIV**			
10.00	15	Routine ANC testing and PMTCT data in Spectrum/EPP/Naomi	Mary Mahy
10.15	15	Proposed adjustments to ANC/PMTCT inputs and coverage for settings where ANC clients exceed births	John Stover
10.30	15	Impact of ANC routine testing surveillance data on HIV prevalence and ART coverage estimates in Mozambique	Oli Stevens
10.45	10	Break	
10.55	15	Interpretation of ANC testing data and indicators in Naomi and alignment to Spectrum	Jeff Eaton
11.10	60	Discussion	

Session 2: CSAVR (chaired by John Stover) **Objectives:** • **Review trends in tests, new diagnoses and CD4 at diagnosis during COVID** • **Reach recommendation for including or adjusting case report data during COVID** • **Review robustness of Knowledge of Status, and fitted CD4 at diagnosis – including during COVID** • **Discuss approach to adjust for HIV+ immigrant diagnoses and deaths**			
12.10	25	Overview and challenges with CSAVR estimates in 2022 • Trends in tests, new diagnoses and CD4 at diagnosis during COVID-19 • Availability of CD4 data at diagnosis and use as calibration data • Age/sex disaggregation in cases & deaths; age/sex IRR fitting • High knowledge of status and CD4 at diagnosis, relative to data • HIV+ Immigration versus incident infections	Eline Korenromp/ Keith Sabin
12.35	25	CSAVR model development • Unexpected high knowledge of status and CD4 at diagnosis • Use of HIV testing volume data • HIV+ Immigration versus incident infections	Guy Mahiane
13.00	60	Lunch	
14.00	60	Discussion	

Time	Duration (mins)	Topic	Presenter(s)/ Lead Discussant
Session 3: Testing and treatment churn (chaired by Jeff Eaton) **Objectives:** • **Determine a model structure and approach for integrating HIV testing and treatment in Spectrum** • **Review existing model approaches to estimate testing and re-testing rates** • **Reach consensus on using HIV testing data as calibration for ART initiation rates** • **Feasibility of integrating Shiny 90 into Spectrum**			
15.00	5	Session background and objectives	Jeff Eaton
15.05	10	Spectrum ART model	John Stover
15.15	5	Overview of • Estimates of knowledge of status in West and Central Africa	Ian Wanyeki
15.20	10	• LTFU in 2022 estimates • Initiation and re-initiation data in Spectrum	Eline Korenromp
15.30	10	Break	
15.40	15	Shiny 90 model: • Approach to estimating HIV testing and re-testing rates • Review of West and Central Africa knowledge of status estimate queries • Linking testing and re-testing to treatment initiation and re-initiation in Shiny 90	Mathieu Maheu-Giroux
15.55	15	Goals testing model	John Stover
16.10	15	Thembisa testing model	Leigh Johnson
16.25	10	Estimating rates of treatment initiation in Malawi	Tim Wolock
16.35	20	PEPFAR ART waterfalls framework	Sadna Patel
16.55	45	• Q&A for presentations • Summary of the working group objectives and process for Day 2	Jeff Eaton (chair)
17.40		CLOSE	

Day 2:

Time	Duration (mins)	Topic	Presenter(s)/ Lead Discussant
Session 3: Testing and treatment churn (continued)			
9.00	30	Evidence on HIV retesting from TRACE study	Harriet Nuwagaba-Biribonwoha
9.30	60	Working groups	
10.30	10	Break	
10.40	60	Discussion	
Session 4: Key population stratified estimates in concentrated epidemics (chaired by Rob Glaubius) **Objectives:** • **Review availability and reliability of routinely reported mode of transmission data** • **Review KP implementation of CSAVR** • **Brainstorm approaches for estimating KP stratified estimates for concentrated epidemic settings**			
11.40	10	UNAIDS strategic priorities for 2023	Mary Mahy
11.50	15	Describe the existing data	Keith Sabin
12.05	40	Case surveillance data by mode of transmission	Tobi Saidel
12.45	60	Lunch	
13.45	40	CSAVR model development • Overview of KP model structure • Examples of the model • Case study comparison with EPP	Guy Mahiane
14.25	10	EPP and EPP-ASM	Tim Brown
14.35	15	Introducing age structure to EPP for concentrated epidemics	Deepa Jahagirdar
14.50	40	Discussion about technical aspects of CSAVR and EPP-KP	
15.30	10	Break	
15.40	50	Working groups	
16.30	60	Discussion	
17.30		Close	

Day 3:

Time	Duration (mins)	Topic	Presenter(s)/ Lead Discussant

Session 5: Key Population estimates in SSA (chaired by Mary Mahy)
 	
Objectives:
 	
• **Review 2022 estimates and consolidated KP data**
  ○ **Use of KP workbook**
  ○ **Estimates of new infections by KP**
  ○ **Comparison with donut estimates**
• **Develop plan for future model development for key populations stratified estimates for sub-Saharan Africa**			

9.00	30	• 2022 KP estimates • Estimates of population size, HIV prevalence and ART coverage	Oli Stevens
9.30	30	Development of a Goals model stratified by age and behavioural risk	Rob Glaubius
10.00	30	Proposed modelling approaches for KP stratified estimates • 'Dynamic IPM' disaggregation of force of infection • KP enhanced static IPM • KP workbook approach (user selected model)	Oli Stevens/John Stover
10.30	10	Break	
10.40	60	Working groups	
11.40	60	Discussion	
12.40	60	Lunch	
Session 6: On ART mortality (chaired by Josh Salomon) **Objectives:** • **Are there sufficient data to move towards relating on ART mortality to VLS over time?** • **Ascertain the impact of transition to DTG on mortality** • **Triangulate estimates of Spectrum deaths in WCENA, EECA, and AP with VR and programme data**			
13.40	10	Changes in VLS over time	Mary Mahy

Time	Duration (mins)	Topic	Presenter(s)/ Lead Discussant
13.50	10	Scale-up of DTG	Marco Vitoria
14.00	10	Literature review of impact of VLS on mortality	Leigh Johnson
14.10	15	IeDEA analyses updates and plans • Revised parameters for all regions • Impact of DTG transition • Linking VLS to mortality estimates • Exploring additional covariates for mortality model	Reshma Kassanjee
14.25	15	Association of viral suppression and ART regimens on-ART mortality in WCENA	Adam Trickey
14.40	10	• Review the source of on-ART mortality estimates in EECA • Validation of Spectrum results with VR data in EECA • Distribution of on-ART mortality multipliers in 2021 & 2022 WCENA files	Rob Glaubius/ Keith Sabin/ Eline Korenromp
14.50	10	Estimates of on-ART mortality in AP • Validation of Spectrum results with VR data • Review the use of the mortality parameter multiplier	Tim Brown
15.00	10	Break	
15.10	60	Discussion	

Session 7: Population estimates (chaired by Mathieu Maheu-Giroux)

Objective:

- **Identify large differences in populations used in 2022 estimates with new WPP estimates**

16.10	40	• Overview of WPP 2022 estimates • DemProj and WPP 2022 • Discussion	Patrick Gerland Rob Glaubius
16.50	10	Recommendations	
17.00		CLOSE	

Day 4:

Time	Duration (mins)	Topic	Presenter(s)/ Lead Discussant
Session 8: Processes (chaired by Jeff Eaton and Mary Mahy) **Objectives:** • **Standardise decision tree guidance on data usage and model options for CSAVR** • **Consider revised default fertility rate ratio parameters for concentrated epidemic settings** • **Improve usage of district estimates of unmet need for ART programme target setting** • **ADR/Navigator – guidance to improve data compilation process** • **Identify data gaps for key population data in SSA**			
9.00	60	CSAVR data usage and model option decision trees	Eline Korenromp
10.00	30	Fertility rate ratio (FRR) local adjustment factor in concentrated epidemics	Eline Korenromp
10.30	10	Break	
10.40	60	Interpretation of district estimates for target setting in Data Pack	Jeff Eaton
11.40	30	Expansion of ADR use and Navigator	Ian Wanyeki
12.10	30	2023 Workshops	Luisa Frescura
12.40	60	Lunch	
13.40	60	Key populations data in SSA • Data availability • What KP data were collected in SSA and gaps • Closing the Gaps	Keith Sabin/Sonia Arias Garcia
14.40	20	How can KP workbook data be used in Data Pack?	Oli Stevens
15.00	10	Concluding remarks	Peter Ghys
15.10	60	Software developers planning session	
16.10		CLOSE	