MEETING REPORT
Schistosomiasis Oversampling Study: Survey Strategy Selection
Nairobi, Kenya - May 18 &19, 2023

Key Takeaways:

- There is a significant data gap across sub-Saharan Africa for supporting schistosomiasis (SCH) intervention strategies and progress towards global goals.
- Currently there is no evidence-based and standarised impact assessment approach which considers the varying level of heterogeneity in the distribution of SCH and is feasible and efficient for determining appropriate sub-district treatment classifications after at least 5 rounds of preventive chemotherapy.
- Schistosomiasis Oversampling Study (SOS) countries and partner country data were used to:
    - develop true prevalence surfaces using Bayesian geostatistical models, and then
    - use these truth surfaces to evaluate the performance of different impact assessment survey sampling methods in terms of correctly classifying treatment decisions in subdistricts.
- The SOS study highlighted that SCH is less focal than we might have thought in many areas; there is variation in prevalence, but often it is between 20% - 80% or 1% - 8%, these areas would still get the same treatment decision. The threshold that matters for treatment decisions is 10%.
- In response to the SOS results, a two-stage impact assessment algorithm was developed to harmonize the need to correctly classify sub-IUs as accurately as possible to minimize over- and under-treatment, while conserving resources in IUs where SCH prevalence is homogenous relative to the 10% threshold.
    - Practical assessments are district-level (IU) impact assessments designed to determine if the prevalence of SCH in the district is sufficiently homogenous, such that the same treatment decision would be appropriate for all sub-districts, to save SCH programs time and money.
    - Precision assessments are sub-district-level (sub-IU) impact assessments appropriate in areas where the prevalence is heterogeneous around the 10% threshold and are designed to classify the prevalence in the sub-district as above or below 10%.
- Program Managers and Experts from the African region reviewed the results of survey sampling simulations and ultimately agreed on the sampling strategy to pilot:
    - **Practical Assessments** should use a systematic sampling approach to select 15 sites, with 30 school-age children per site, per district.
    - **Precision Assessments** should use purposive sampling, based on sites expected to have the greatest risk for SCH, to select 4 sites and 20 school-age children per site, per sub-district.
- The next steps are to pilot this impact assessment approach in the context of ongoing SCH programmes to better understand the feasibility and utility for making treatment decisions.

ESPEN Impact Assessment Design and Gaps (Pauline Mwinzi and Jorge Cano)

- ESPEN portal endemicity maps highlight the progress that has been made in mapping and treating SCH; however, the disease is still largely endemic across the continent.
- By 2021 over 15% (n=672) of endemic IUs reported having received ≥5 rounds of effective preventive chemotherapy (PC), defined as ≥75% reported coverage.In 2023 and 2024 it is estimated that >1,000 IUs will be eligible for impact assessments, based on baseline endemicity and ≥5 rounds of effective PC.
- Countries are currently using a variety of different sampling designs to conduct impact assessments with and across countries.

Schistosomiasis Oversampling Study (SOS) design and country results (Stella Kepha and Joseph Opare)

- The objective of the SOS is to identify optimal survey sampling method(s) for conducting impact assessments that are feasible for country programs, cost-effective and result in appropriate treatment classifications.
- To evaluate the performance of different survey sampling methods, SOS took a multi-phased approach with health ministries and regional technical partners in the lead to: (I) develop country-specific protocols to (II) collect comprehensive parasitological data through oversampling surveys (the "SOS") in multiple settings after several treatment rounds of PC, then (III) these data were used to generate 'truth' surfaces, which were used as a basis to evaluate alternative impact assessment survey designs using simulations.
- SOS was conducted in 4 countries: Cote d'Ivoire, Ghana, Mali and Togo; these sites were chosen to represent the different archetypes (aka 'epidemiological settings) where schistosomiasis is commonly found.
- In each of the country sites, three contiguous districts were selected and 40% to 50% of communities were randomly selected.
- The primary outcomes were: prevalence (and intensity) of infection based on duplicate Kato-Katz thick-smear or single urine filtration and urine dipstick readings on single day stool / urine samples; the secondary outcomes of SOS were: socio-demographic, school attendance, behavioural exposures to surface water, water and sanitation access.
- In addition to the standard diagnostic tests, in a subset of sites (three per country) dried blood spots were created for each individual and these were frozen, along with extra urine, for future SCH diagnostic testing.
- The study represented a massive effort with over 200 communities visited and 6000 – 8000 SAC sampled per country site.
- Three of the four sites had *S. haematobium* as the predominant species, while only Cote d'Ivoire was predominant with *S. mansoni* in the selected districts.
- In addition to the four SOS country sites, three partner studies that followed a similarly dense sampling strategy in Burundi, Kenya and Ethiopia (all with predominant *S mansoni*) agreed to share their data to contribute to the survey design simulations.

Prevalence Data & Truth Surfaces (Penelope Vounatsou)

- Bayesian geostatistical models (BGM) were applied to the SOS survey data from each country to predict SCH prevalence at all villages within the selected districts (truth surface).
- The models included climatic, socio-economic (SES) and behavioral predictors. Climatic data were extracted from satellite sources (i.e. daily land surface temperature from Terra, Moderate Resolution Imaging Spectoradiometer) and other gridded data (i.e. rainfall estimates from the Climate Hazards Group InfraRed Precipitation with Station database) during 2017-2021 were processed to obtain 19 bioclimatic variables. Distance to permanent water and area covered by water within a buffer of 5km were calculated from Copernicus satellite data and the humanitarian OpenStreetMap data. The Normalized Vegetation Index (NDVI) was used as a proxy of humidity. SES (i.e. water and sanitation sources, occupation) and behavior data (i.e playing in water, fishing) were collected as part of the SOS surveys.
- BGM without covariates were fitted to SES and behavioral data to predict this information at the unsampled villages.
- Extensive BGM selection was carried out by fitting the models arising from all possible combinations of predictors. The model with the best predictive ability in terms of root mean squared error was used to obtain the truth surfaces.
- The truth surfaces were used to (i) estimate the geographical distribution of prevalence at village level; (ii) classify sub-districts and districts in the "true" treatment category (at 10% prevalence threshold) and (iii) compare the survey designs.
- Each survey design (Appendix 2), in each setting, was replicated 100 times and compared using 100 simulations of the true surface to take into account sampling variation (of the design) and prediction uncertainty (of the true surfaces). Therefore, for each design and setting, a total of 10,000 comparisons were performed between sampled and “true” survey data.
- The performance of sampling designs (Appendix 2) was assessed in terms of i) the proportion of IU (subdistricts) and villages classified in the correct treatment category and ii) the accuracy of prevalence estimation.

Summary of past SOS meetings leading up to the present (AFRO-SOS SCH Programme Managers questionnaire & FGD; COR-NTD Breakout session; SOS Technical Results Meeting) (Fiona Fleming)

- Greater granularity in treatment decisions are needed. Relying on mean prevalence of an IU for decision-making will mask heterogeneity; incorporating village-level prevalence into decision making is a priority.
- Survey design should be flexible to a choice of IU (although non-admin IU would be challenging for implementation). Could have a 2-stage design with a preference for school-surveys not community-based.
- In all three meetings, it was apparent that whilst there's future potential, at present geostatistical survey design would be challenging.

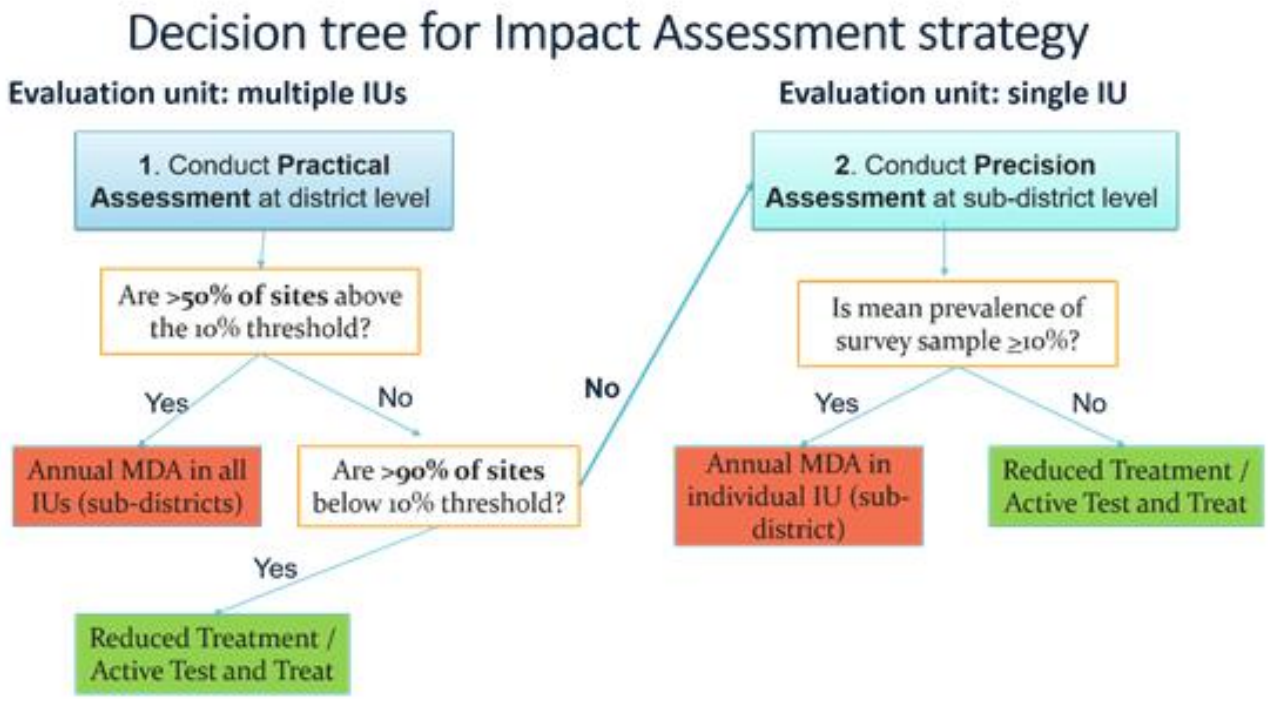
Impact Assessments (Fiona Fleming)

- What is an impact assessment (IA)? A survey in areas that have conducted multiple rounds of preventive chemotherapy and used to make future treatment decisions for an implementation unit = sub-district equivalent.
    - *IA are not* traditional SCH/STH 'impact surveys' which follow a limited number of sites longitudinally in an area/s and demonstrate changes in infection but don't support treatment decisions.
    - *IA are* more similar to reassessment /micro-mapping / precision mapping / granular mapping surveys which support sub-district decision making but not always cost efficient.
- How will IA be used by national NTD programmes? Following 5 or more rounds of PC with effective coverage (≥75%), the IA will be conducted to support treatment (and other control/elimination interventions for SCH) decisions at the sub-district level and will support:
    - a more efficient use of praziquantel in a country by reducing the number of treatments where they are not required/ directing treatment where most needed;
    - providing evidence to determine if WHO targets of Elimination as a Public Health Problem by 2030 have been reached within a country.
- The data from the IA will be incorporated into the ESPEN SCH Community Data Analysis Tool and the Joint Application Package for more accurate PZQ requests.

Introducing Practical and Precision Assessments (Katie Gass)

- Countries are encouraged to shift from district to a sub-district treatment strategy to ensure that treatment is making it to the populations that need it most, making the sub-district the preferred *implementation unit* for SCH.
- The outstanding question is what is the preferred *evaluation unit* (i.e., the geographical area across which a survey is implemented) for conducting impact assessments.
- The SOS study highlighted that SCH is less focal than we might have thought in many areas; there is variation in prevalence, but often it is between 20% - 80% or 1% - 8%, which isn't what matters, these areas would still get the same treatment decision.
- The threshold that matters for treatment decisions is 10%.
- In districts where SCH prevalence is **consistently low** or **consistently high**, conducting an impact assessment in each sub-district separately would be overkill because all would get the same treatment decision; these areas can be considered **homogenous.**
- In homogenous settings, it would be a more efficient use of resources if the **district** were the evaluation unit for the impact assessment survey.
- In other settings where SCH prevalence is heterogeneous around the 10% threshold within a district, it is necessary to conduct the impact assessment at the sub-district level (i.e. the evaluation unit), to avoid over- or under-treating large areas.
- The question then becomes, how does an NTD program determine whether a district is homogenous and thus a district-wide impact assessment would be appropriate vs. heterogeneous and thus a sub-district level impact assessments are necessary?

- A two-step impact assessment strategy can allow programs to identify whether a district or sub-district impact assessment is appropriate and to determine the treatment decision as follows:
    - Step 1: Conduct a **Practical Assessment** designed to test if the district has homogeneously high (≥10%) or homogeneously low (<10%) SCH;
    - Step 2: Where the Practical Assessment indicates the district has heterogeneous SCH, proceed to a **Precision Assessment** at the sub-district level to determine if the mean prevalence is ≥10%.
- Programs should decide whether it makes sense to start with the Practical Assessment first or go straight to the Precisions Assessment; the decision should be based on local knowledge:
    - **When to start with a Practical Assessment:** in areas with sparse data or little program knowledge; where existing data suggest that the prevalence of SCH is likely to be homogenously high or homogenously low; or anywhere else where, starting with the Practical Assessment is likely to be a better use of resources.
    - **When to start with a Precision Assessment:** In districts suspected as having heterogeneous SCH relative to the 10% threshold, based on historic data, treatment history, or the presence of focal sources of transmission (e.g., waterbody with snails that is present in only part of the district).



Survey sampling simulation results (Rachel Pullan)

- The table in Appendix 2 shows the different survey deigns that were tested for each approach.

- This session first presented results supporting the design for Practical Assessments, before presenting the results supporting the design for Precision Assessments
- **Practical Assessment results**:
    - **Practical Assessments** are conducted at the evaluation unit level (ie a district), to make decisions for multiple subIUs (ie sub-districts). This might be (i) a treatment decision, or (ii) a decision to do precision assessments. They are intended to identify areas that are **sufficiently similar** where assigning the **same treatment decision** to all subIUs in the implementation unit will not lead to unacceptable levels of under- or over-treatment.
    - As described in the figure above, there are **three possible outcomes** based on the proportion of sites exceeding the 10% prevalence threshold. These proportions were chosen because looking across all the available truth surfaces, the proportion of communities correctly classified, when applying these criteria, were equivalent to basing a decision on the population-weighted mean prevalence.
    - Survey sampling simulations considered **alternative sampling strategies** (simple random sampling and systematic sampling using a two-stage sampling procedure sampling subIUs and then sites within subIUs) and **sampling effort** (varying the number of children and number of sites sampled). To evaluate sampling strategies for practical assessments, the decisions resulting from simulated surveys were compared to the correct decision based on directly classifying the truth surface.
    - Overall, **systematic sampling performed better than random sampling**, although there were noticeable differences between countries. **Increasing the number of sites sampled from 10 to 15 improved classification accuracy** for most countries. In some countries increasing the number of children improved accuracy. **Most incorrect subIU decisions were in classifying subIUs as requiring further precision assessments when, in fact, a treatment decision would have been appropriate.** That is, this approach is conservative and inclined towards suggesting further precision mapping in areas of uncertainty, and only rarely are subIUs assigned the wrong treatment decision. In most countries, upwards of 90% of subIUs assigned a treatment decision based on the practical assessment were given the right decision.
- **Precision assessment results**:
    - **Precision assessments** are conducted at the subIU level (ie sub-district), to make decisions for that subIU only. They are intended to make an accurate treatment decision in areas that are heterogeneous around the 10% prevalence threshold. The two possible outcomes are classifying the subIU as above or below 10%, based on the sample mean.
    - Survey sampling simulations considered **alternative sampling strategies** (simple random sampling, purposive sampling with increased probability of sampling sites close to permanent water bodies, and a cluster LQAS design) and **sampling effort** (varying the number of children and number of sites sampled). Simulations also considered the implications of sampling across larger and smaller subIUs.
    - To evaluate sampling strategies for precision assessments, the decisions resulting from simulated surveys were compared to the correct decision based on directly classifying the truth surface.
    - In most settings, we saw similar patterns in terms of the proportion of subIUs correctly classified when comparing an approach using simple random sampling to select sites:

accuracy was seen to increase up to ~3 sites per subIU, and there were only marginal gains seen when increasing the number of children sampled.
    - Results were variable when considering alternative sampling approaches: purposive sampling resulted in considerable overtreatment in Ghana, although undertreatment was minimised, whereas in other settings there were minimal differences between purposive and random sampling. LQAS approaches tended to increase overtreatment but reduce undertreatment.
- After reviewing results for each sampling strategy, **indicative cost estimates** were provided. Survey cost estimations were performed using an ingredients-based costing model incorporating cost data from across multiple implementation settings including those covered by SOS, and included a series of assumptions around diagnostics, logistics and team compositions.
- Cost efficiency analysis was conducted to estimate the cost-per-subIU-correctly-classified by combining information on the truth surface, the performance statistics for each survey strategy, and the cost of implementing each survey strategy.
- Indicative costs were shared that compared the practical & precision approach to blanket precision assessments, assuming 10 - 20 sites and 20 and 30 children for practical assessments, and varying precision assessments between 2 sites and 20 children, and 5 sites and 50 children.

[Breakout Group Discussions]

According to the agenda, three breakout groups were formed on Day 2 (May 19):

1. Breakout 1a: Experts and Program Managers from the African region – Anglophone
2. Breakout 1b: Experts and Program Managers from the African region – Francophone
3. Breakout 2: External experts, WHO, donors

Selection of the preferred sampling strategy by regional experts (Breakout Group 1)

- Program managers and regional experts engaged in a lively discussion about how best to prioritize the sampling designs
- Interest was expressed in using the impact assessment data to make treatment classifications *and* to measure elimination as a public health problem (<1% heavy intensity infections), which would require a larger sample size given the low threshold.
    - Others clarified that the purpose of the impact assessment strategy is to make treatment decisions.
- Hotspot concerns were raised, particularly what to do if there is a community that exceeds 10% prevalence in a district or sub-district that falls below the threshold for annual treatment, as the WHO guidelines call for all communities above 10% to be treated.
- When discussing a 5-site x 50 SAC design, a bequest was made to factor in resource scarcity at the national program level, the ideal design should be the minimally viable design.

- **Practical Assessment Discussion**
    - Both the Anglophone and Francophone program manager breakout groups agreed that the Practical Assessment strategy that included 15 sites x 30 SAC was preferred, given that the 15-site design showed a marked increase in classification accuracy compared to

either the 10 or 20 site design. This improvement in 15 sites is due to the cutoff used to define areas with majority high prevalence (>50%). Since 15 is an odd number, the threshold for being >50% is 8 villages, which corresponds to a 53% prevalence.

- **Precision Assessment Discussion**
    - Some participants felt strongly that more data was better because it can lead to more precise estimates, and is, therefore, the best approach; however, when modeling data on the impact of # sites and # SAC on precision (as measured by mean absolute error) were subsequently presented, it became apparent that any gains in precision from adding sites leveled off pretty quickly after 3-4 sites per sub-district. Similarly, no gains in precision were seen when increasing the number of SAC per site.
    - Concern was expressed that a single design may not perform equally well in all settings, as the number of sub-districts per district, and the population size of districts and sub-districts, can vary widely between countries.
        - A suggestion was made that 3 sites may be sufficient for Precision Assessments in small sub-districts but 4 may be better in larger sub-districts
        - Because it is hard to determine what constitutes a 'big' vs. 'small' sub-district, it was agreed that defaulting to a 4 sites-per-subIU design will meet the needs of small and large districts while keeping the guidance for countries simple.
- There was a debate on the value of purposive vs. simple random sampling (SRS) for the Precision Assessment. The SRS design performed slightly better at making the correct treatment classification, though the median classification accuracy for both designs was not significantly different. The purposive sampling design was more likely to reduce under-treatment, while the SRS design was more likely to reduce over-treatment.
    - The group stated a preference for reducing under-treatment, as opposed to minimizing over-treatment
    - It was discussed whether the simulations could accurately capture purposive sampling, which was simulated by picking sites based on proximity to water. Concern was raised that this selection strategy may not represent the way purposive sampling would have been performed on the ground by programs – as it doesn't account for local knowledge. It was particularly challenging to accurately simulate purposive sampling in the SOS data, given that there was little association between SCH and proximity to water, with the exception of Ghana.
    - In Ghana, proximity to Lake Volta was strongly associated with SCH prevalence and in this setting the purposive design showed a marked improvement in reducing under-treatment, compared to SRS. It was believed that this example is more reflective of what one would expect from purposive sampling.

The Practical & Precision Assessment strategy that was agreed upon by the participants is included in Appendix 1.

Secondary analyses and operational research priorities from SOS (Breakout Group 2)

- External partners and donors led by Drs Anouk Gouvras and Evan Secor discussed questions around critical actions and gaps for schistosomiasis around the delivery and development of an M&E Framework. Including questions directly on which gaps the SOS data could address; at what frequency will impact assessments need to be conducted; and whether practical and precision assessments could be conducted in the same year or whether an interim treatment decision was required? The main themes coming from these discussions were:

- **Additional analysis / operational research needs**

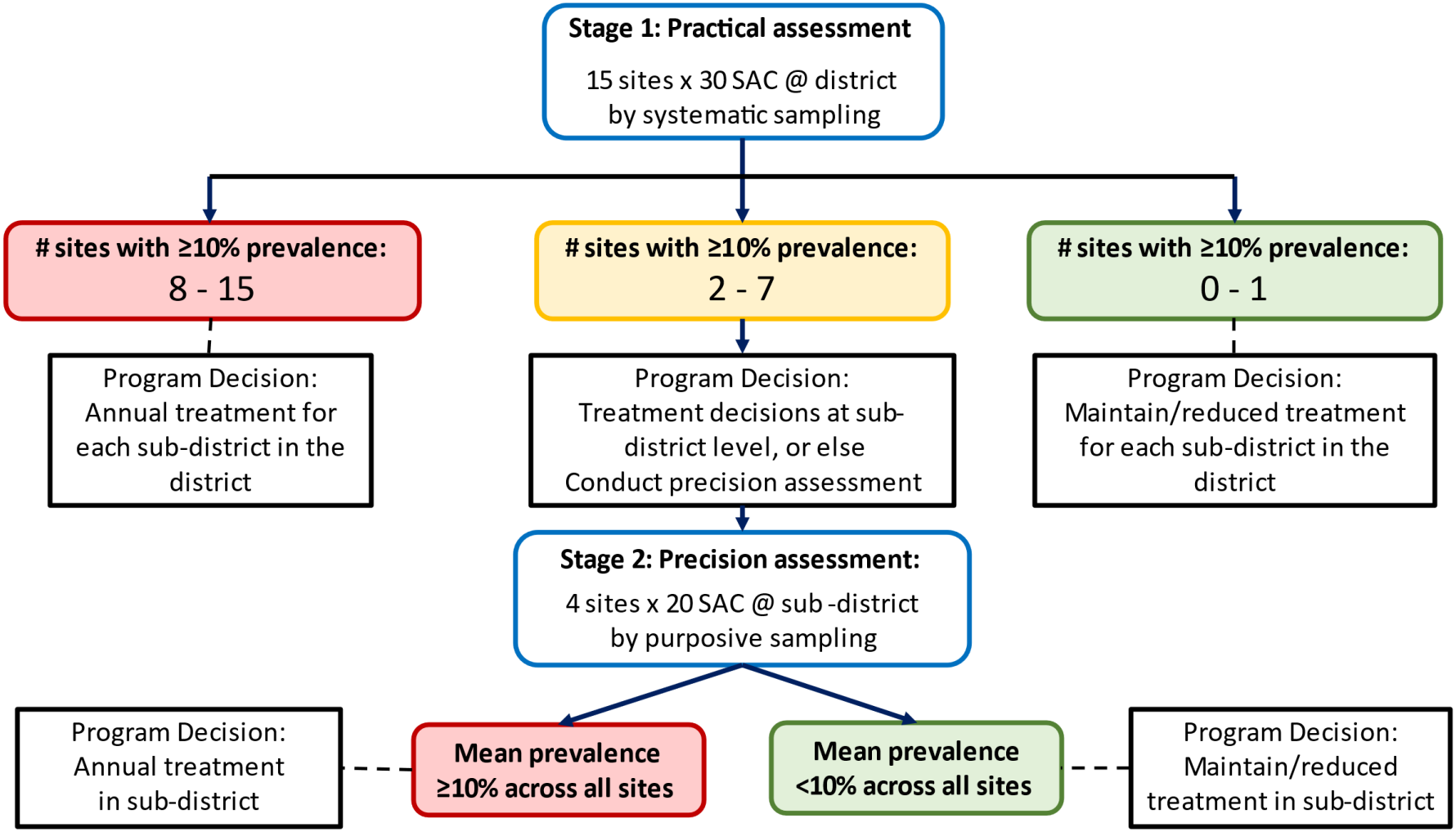
    - Intensity data is important for transmission models and for the current elimination as a public health problem target (<1% prevalence of heavy intensity) but programmatically it does not inform decision-making. Growing evidence shows that prevalence of micro-haematuria (non-visible blood-in-urine and a form of SCH-related morbidity) as diagnosed by hemastix / urine dipstick could be an alternative tool for determining EPHP for *S. haematobium* and it has a strong association with prevalence of infection and prevalence of heavy intensity of infection by urine filtration.
      ACTION: Further analysis of SOS data should be performed to explore the associations between prevalence in haematuria, egg-detected prevalence and intensity in the different archetypes.
    - Additional data collection in SOS sites may provide an opportunity to determine SCH-related morbidity in adults and the relationship with prevalence in SAC to determine the impact of >5 years of regular preventive chemotherapy (PC) and expected levels of adult morbidity.
    - How to tackle areas of persistent infection, hotspots
        - Is the WHO SCH guidelines definition of a hotspot programmatically feasible to determine? Are there alternatives? What is a minimum prevalence in a baseline survey to determine a hotspot e.g., 30%?
        - Are untreated adults perpetuating high prevalence areas? Data from Ethiopia, Zimbabwe, Malawi and Kenya suggest not but need to be synthesised with SOS data and analysed to demonstrate *Schistosoma* infection in adults in settings with >5 years of regular PC targeted at school-age children (SAC).
        - Are there programmatic coverage issues contributing to areas of persistent infection?
        - Is there PZQ resistance / tolerance?
        - What role does the force of infection play in addition to the above factors?
    - SOS plus data to demonstrate age-infection profiles and that school-attending school-age children are an appropriate group for sampling in practical and precision assessments.

- **Frequency and timing of impact assessments**

    - 5 years or 3 years were both thought to frequent enough for impact assessments for SCH where annual PC (prevalence >10%) with the discussion coming down to resources and the ability to change treatment strategies.
    - Three years for impact assessments would be better for locating and monitoring hotspots.

- Would impact assessments be used where prevalence is under 10% and the treatment strategy is maintained or reduced? What would a surveillance system look like for where EPHP has been achieved? What role could urine dipsticks play in surveillance?
- If programmatically feasible to roughly plan and budget, don't delay the MDA in between the 2-stage practical and precision assessments.

Appendix 1: Impact Assessment Strategy for Pilot Testing



Appendix 2: Survey designs tested in SOS analysis

Survey design aim: to determine a sub-district (implementation unit) decision for preventive chemotherapy.

survey designs	evaluation unit (EU)	definition / metric	# sites	# children	relative cost* per EU
Practical Assessment					
Systematic Sampling	District	A form of random sampling where ordering is used to select the sample. For practical assessments, this design is used to ensure that selected sites are evenly dispersed across the sub-districts within the EU (district) **Decision metric**: number of sites with mean prevalence ≥10%	Range 10 to 20	Range 20 – 50	Low to moderate sampling effort e.g. $4,842-$20,459 per district
Precision Assessment					
Purposive	Sub-district	As above **Decision metric**: classifying mean prevalence of the survey as above/below 10%	Range 1 to 5	Range 20 – 50	Low to moderate sampling effort e.g. $470 - $4,921 per sub-district
Simple random sampling (SRS)	Sub-district	A probability sampling approach where every site (e.g., school) has an equal chance of being chosen. **Decision metric**: classifying mean prevalence of the survey as above/below 10%	Range 1 to 5	Range 20 – 50	Low to moderate sampling effort e.g. $470 - $4,921 per sub-district
Cluster Lot Quality Assurance Sampling (CLQAS)	Sub-district	A sampling methodology that combines cluster sampling with lot quality assurance sampling techniques. **Decision metric**: classify the evaluation unit as above/below a target threshold (e.g., 10%) with a specified level of precision.	5	50	High sampling effort 5 sites x 50 SAC e.g. $470 - $4,921 per sub-district
Precision mapping *Based on WHO-ESPEN / frequently the status quo*	Homogenous zones in a district / sub-district	A non-probability sampling technique. In the context of this meeting, purposive sampling means selecting sites (e.g., schools) that are believed to have the greatest risk for SCH, such as proximity to a water body. **Decision metric**: highest prevalence site used to determine prevalence for sub-district	5	50	High sampling effort 5 sites x 50 SAC e.g. $470 - $4,921 per sub-district

* based on average costs for a moderate sized country, includes training costs, perdiem (average team of 5 inc. driver), fuel, vehicle hire, survey equipment and consumables, one day sampling = 1 slide urine filtration, 2 slides Kato-Katz, 1 urine dipstick and average of 60 slides / technician / day

Appendix 3: Definition of terms used

Cluster Lot Quality Assurance Sampling (CLQAS)	a sampling methodology that combines cluster sampling with lot quality assurance sampling techniques. The survey is designed to classify the evaluation unit as above/below a target threshold (e.g., 10%) with a specified level of precision.
Evaluation unit (EU)	the geographic area across which an impact assessment is conducted. For Precision Mapping, the EU is the sub-district; for Practical Mapping, the EU is the district. In both instances, the implementation unit would remain the sub-district.
Heterogeneous	when the prevalence of SCH within a given area differs substantially between sites, such that some sites have a prevalence <10% while others have ≥10% prevalence.
Homogeneous	when the prevalence of SCH within a given area is similar, relative to the target threshold, such that the majority of sites have <10% prevalence, or the majority have ≥10% prevalence.
High prevalence	above 10% prevalence as per WHO 2022 guidelines for SCH.
Impact Assessment (IA)	a survey in areas that have conducted multiple rounds of preventive chemotherapy and is used to make future treatment decisions. In this context, an IA is more similar to micro-mapping / precision mapping / granular mapping and not similar to traditional SCH/STH 'impact surveys', which follow a limited number of sites longitudinally.
Implementation unit (IU)	the geographic area for which a single treatment decision is made; for schistosomiasis programs and the purposes of this meeting, the IU is the sub-district.
Low prevalence	below 10% prevalence, as per WHO 2022 guidelines for SCH.
Over-treatment	when treatment is provided to people or geographic areas that do not merit it. In this meeting, over-treatment refers to classifying an IU where the true prevalence is <10% (and thus reduced treatment is called for) as incorrectly requiring annual treatment.
Practical Assessment	an impact assessment strategy conducted at the district level and designed to identify if all the IUs in a district can be classified as I) having a homogeneously low prevalence (<10%) and thus merit a *reduced treatment strategy,* II) majority high and thus merit an *annual treatment strategy,* or III) requiring Precision Assessment.
Precision Assessment	an impact assessment strategy conducted at the sub-district level and designed to classify sub-districts as ≥10% or <10%.
Purposive sampling	a non-probability sampling technique. In this meeting, purposive sampling means selecting sites (e.g., schools) that are believed to have the greatest risk for SCH, such as proximity to a water body.
Simple random sampling (SRS)	a probability sampling approach where every site (e.g., school) has an equal chance of being chosen.

Sites	these refer to the locations where the survey is conducted (i.e., primary sampling unit) and could be schools or communities/villages. The SOS data suggest sampling school-age children in schools would be more efficient and would not provide a biased sample.
Systematic sampling	a form of random sampling where ordering is relied on to select the sample. In this meeting, systematic sampling is used in Practical Assessments to ensure that the selected sites are evenly dispersed across the sub-districts within the EU (district).
Under-treatment	when treatment is not provided people or geographic areas that merit it. In this meeting, under-treatment refers to classifying an IU with a true prevalence $\geq$10% (thus meriting annual treatment) as incorrectly requiring reduced treatment.

AGENDA

Schistosomiasis Oversampling Study: Survey Strategy Selection
Nairobi, Kenya - May 18 &19, 2023

Objective: The purpose of the meeting is to review the Schistosomiasis Oversampling Study (SOS) simulation results comparing different impact assessment sampling strategies. Participants representing schistosomiasis endemic countries will be asked to lend their expertise and experience to the discussions regarding the relative merits of each approach and, ultimately, to reach an agreement on a single strategy that can be recommended to WHO as an impact assessment approach.

Location: Movenpick Hotel

May 18th, 2023 **AGENDA**

Room: Almasi 2, 4th Floor		
Time	**Welcome, Overview, Stage setting**	Presenter
9:00 – 9:15am	Welcome	Wycliff Omondi
9:15 – 9:30am	Introductions & Agenda Presentation	Day 1 AM Chair
9:30 – 9:45am	Opening Remarks	Amadou Garba
9:45 – 10:00am	ESPEN strategy - What is currently being done & gaps	Pauline Mwinzi & Jorge Cano
Time	**Presentation of what we've learned so far**	Presenter
10:00 – 10:10am	SOS Rationale	Stella Kepha
10:10 – 10:35am	Country sites, design and quick data overview (one single summary of what was done - not individual countries)	Joseph Opare
10:35 – 10:45am	Truth surface creation	Penelope Vounatsou
10:45 – 11:00am	BREAK	
Time	**Presentation of what we've learned so far**	**Presenter**
11:00 – 11:15am	Tying it all together: Summary of past meetings	Fiona Fleming
11:15 – 11:50am	Discussion	All (Day 1 AM chair)
11:50 – 12:00pm	Big picture of view of where the discussions are headed	Stella Kepha
12:00 – 1:00pm	LUNCH	
Time	**Presentation of Simulation Results**	**Presenter**
1:00 – 1:30pm	Practical & Precision assessment conceptual presentation	Katie Gass
1:40 – 2:00pm	Practical assessment simulation results + Discussion	Rachel Pullan
2:00 – 2:30pm	Precision assessment simulation results + Discussion	
2:30 – 3:00pm	Cost & feasibility summary tables + Discussion	Fiona Fleming
3:00 – 3:15pm	BREAK	
Rooms: Almasi 2, La Mesa, Turkana 4 and Turkana 6 for group work, Almasi 2 for summary and prep		
Time	**Group Work – digging into the interim strategy**	**Presenter**
3:15 – 4:45pm	Small group exercises with sample data	Day 1 Chair
4:45 – 5:00pm	Summary and prep for day 2	

May 19th, 2022 **AGENDA**

Room: Almasi 2, 4th Floor		
Time 8.00 – 9.00am	**Meet the experts** Sascha Gummin & Penelope Vounatsou (Swiss Tropical Public Health Institute) Joseph Timothy (London School of Hygiene & Tropical Medicine)	
Time	**Discussion of proposed sampling strategies**	**Presenter**
9:00am 9:00 – 9:45am 9:45 – 10:30am	Welcome Day 1 Breakout Recap Review of highlights from Day 1	Day 2 Chair All / Facilitators All
10:30 – 10:45am	Tea Break	
10:45 – 12:00pm	Breakout groups: 1) Experts/PMs from the region - Anglophone **Room: Turkana 6 (3rd Floor)** 2) Experts/PMs from the region - Francophone **Room: Turkana 4 (3rd Floor)** 3) External experts/donors/WHO - **Room: La Mesa (15th Floor)**	
12:00 – 1:00pm	**LUNCH**	
Time 1:00pm – 2:30pm	**Discussion of proposed sampling strategies** Breakout groups: 1) Experts/PMs from the region – **Room: Almasi 2 (4th Floor)** 2) External experts/donors/WHO – **Room: La Mesa (15th Floor)**	
2:30 – 3:00pm	Break	
Time 3:00 – 4:00pm 4:00 – 4:30pm	**Regroup and Discuss** Present group work; review & finalize recommendations Piloting Impact Assessment Strategy – where, what needs to be answered with pilots, what is success?	**Presenter** Facilitators SOS Organizers
Time 4:30pm – 5:00pm	**Closing** What have we accomplished & what is forthcoming Closing remarks from Donors Closing remarks from ESPEN Closing remarks from WHO	**Presenter** Day 2 Chair, SOS Organizers