Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Natalya Kostandova Date

Simulating the effect of evaluation unit size on eligibility to stop mass drug administration for lymphatic filariasis in Haiti

By

Natalya Kostandova
Degree to be awarded: Master of Public Health

Global Epidemiology

_____

Dr. Amy Kirby
Faculty Thesis Advisor

_____

Dr. Katherine Gass
Thesis Field Advisor

Simulating the effect of evaluation unit size on eligibility to stop mass drug administration for lymphatic filariasis in Haiti

By

Natalya Kostandova

B.S., California Institute of Technology, 2010

Faculty Thesis Advisor: Amy Kirby, Ph.D., M.P.H.
Thesis Field Advisor: Katherine Gass, Ph.D., M.P.H.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2016

Abstract

Simulating the effect of evaluation unit size on eligibility to stop mass drug administration for lymphatic filariasis in Haiti
By Natalya Kostandova

As more and more countries have implementation units that are entering their sixth year of annual mass drug administration (MDA) for lymphatic filariasis (LF), there is a need to assess whether transmission has been reduced below the threshold required for sustainable transmission and MDA can be stopped. Currently, the main tool used to assess this threshold is the Transmission Assessment Survey (TAS). The guidelines for TAS limit the population in the evaluation units (EUs) to <2 million. This study uses simulations to investigate the effect of population size on the classification of EUs as either passing or failing TAS.

The data come from TAS conducted in 14 EUs in Haiti during 2014-2015. To simulate the effects of population size, larger "combination-EUs" were created by forming eight combinations of adjacent EUs. Several approaches to simulate TAS were carried out, with the intent to see how the classification of EUs as passing or failing TAS would change when the larger units were considered.

The results of the simulations show that in some combinations, the vast majority of the time the classification of the combination-EU would agree with both the expected decision, obtained by passing or failing the combination-EU based on the overall expected prevalence, calculated from positive results, sample size and target population from the original EUs, and the desired decision, that is the programmatic decision to fail any combination-EU in which one or more of the original EUs failed. However, a non-negligible proportion of combination-EUs were misclassified.

Misclassifying an EU as failing would result in continued MDA in a region where prevalence is lower than the threshold required for sustainable transmission, translating into additional rounds of MDA and TAS. Misclassifying an EU as passing when transmission persists is even more troubling, given the programmatic implications of stopping MDA too early. The guidelines should be carefully reconsidered to ensure that TAS is a valid and reliable tool to assess the possibility of stopping LF MDA.

Simulating the effect of evaluation unit size on eligibility to stop mass drug administration for lymphatic filariasis in Haiti

By

Natalya Kostandova

B.S., California Institute of Technology, 2010

Faculty Thesis Advisor: Amy Kirby, Ph.D., M.P.H.
Thesis Field Advisor: Katherine Gass, Ph.D., M.P.H.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2016

Acknowledgments

I would like to thank my advisors, Dr. Katherine Gass and Dr. Amy Kirby, for their unrelenting support and guidance throughout this process.

Table of Contents

Chapter I: Background

Lymphatic filariasis

Lymphatic filariasis (LF) is a vector-borne disease caused by nematodes, or roundworms, that invade lymphatic tissue (1). With over 120 million people affected by the disease, and 1.3 billion at risk, lymphatic filariasis is the second most common cause of physical disability among neglected tropical diseases. While LF does not have a single causative agent, *Wuchereria bancrofti* is the most common cause, followed by *Brugia malayi* and, very rarely, *Brugia timori* (Ibid).

LF is spread through mosquitoes, specifically those belonging to the *Culex*, *Anopheles*, *Aedes*, or *Mansonia* genera. The life cycles of *W. bancrofti* and *B. malayi* have some variation, although both include a portion of a cycle in the vertebrate host and a portion that takes place inside the mosquito.

As is the case for *W. bancrofti*, a mosquito takes a blood meal, during which it deposits third stage filarial larvae onto the skin of the human it bites. The larvae enter into the bite wound and develop into adults. These adult nematodes tend to migrate to the lymphatics, with female worms reaching up to 10 cm in length, while the males are about half the size and much smaller in diameter (2). Within the human host, the adults then produce microfilarae (Mf), which can develop up to ~300μm in length. The Mf are protected by a sheath, and follow a circadian rhythm with nocturnal periodicity. Specifically, Mf are released into the peripheral blood at night, with the exception of Mf in the South Pacific (Appendix 1) (Ibid).

The Mf migrate to lymph and blood channels, and may be ingested by a mosquito that takes a blood meal from an infected human. Inside the mosquito, Mf lose their protective sheath, and some of them manage to move through the walls of the mosquito's mid-gut to the thoracic muscles (2). In these muscles, Mf develop into first stage larvae and eventually reach third stage, a point at which they are considered infective. The third stage larvae migrate to the mosquito's proboscis and are deposited into a human's skin when the mosquito takes a blood meal, thus completing the life cycle (Ibid).

While some 67% of the infections are asymptomatic, in some patients the infection can manifest in swelling, specifically in chronic lymphedema and elephantiasis (12.5% of infected patients), which affects mostly limbs and, in some cases, the genital region (1, 3). Chyluria and hydrocele can also take place (20.8% of infected patients). Overall, the Mf infection impedes the functioning of the lymph system, and the clinical manifestations, which affect some 44 million people, can lead to debilitating disability, as well as stigma, psychological problems, and lowered quality of life (1, 4).

While there exist strategies to manage symptoms, treatment and reversal are difficult. In particular, acute dermatolymphangioadentitis, which has clinical symptoms that resemble that of cellulitis, is usually managed with antibiotic treatment. Lymphoedema and elephantitis are usually managed by patients themselves. Management techniques include wound care, foot care, limb washing, use of suitable footwear, and physiotherapy and exercise. Note that therapy for acute episodes should be addressed by medical professionals. Hydrocele can be managed through lymphatic drainage, while more advanced hydrocele can be addressed through hydrocoelectomy, a type of surgery (5). Because symptom management is challenging and difficult to

reverse, the efforts have been concentrated on prevention of LF transmission and infection.

Mass Drug Administration (MDA)

Once the clinical manifestations of LF set in, they are difficult to reverse, so the main cornerstone of LF control is prevention through Mass Drug Administration (MDA). The primary objective of MDA is to lower the level of microfilaraemia in individuals that have been infected so that, even after MDA is stopped, transmission cannot continue (6). The World Health Organization recommends annual distribution of selected drugs to all those living in areas at risk. Specifically, in areas with endemic LF that do not have onchocerciasis co-endemicity, the drugs of choice are diethylcarbamazine (DEC) and albendazole. In these areas, all individuals are treated with exception of children under 2 years old, pregnant women, and individuals that are severely ill. Where onchocerciasis is co-endemic with LF, the drug combination of choice is ivermectin and albendazole, which are administered to all except for pregnant women and women lactating in the first week after birth, severely ill, and children that are shorter than 90 cm in height. While alternative prevention methods have been considered and implemented, such as fortification of table salt with DEC for 1-2 years, MDA is most common. The main course of action of the drugs administered is to lower the number of Mf that circulate in blood, thus preventing further transmission when mosquitoes take blood meal (Ibid).

Effectiveness of MDA has been shown in many endemic areas, such as Egypt. Following 4-6 annual rounds of MDA, with effective coverage of over 65%, LF transmission was greatly reduced and, in some cases, driven below the threshold of

sustainability (1, 6). In addition, this drug regimen led to the elimination of adult worm nests formed by worms inside lymph nodes and lymph vessels post-treatment, and has reversed sub-clinical lymphatic pathology following 3 years of bi-annual treatment in children in India (1). However, the effect of MDA on progression of LF is not entirely clear, as a study in Haiti showed that MDA had no effect on measures of foot, ankle, or leg circumference, as well as on volume displacement of leg and stage of lymphedema in LF patients. Frequency of acute dermatolymphangioadenitis episodes, a common manifestation of LF, was similar for patients that received MDA and those that did not. At the same time, patients receiving MDA reported improvement in quality of life, so there may be psychological benefit to DEC administration to those affected by LF (7). Ivermectin and albendazole are effective in reducing the number of circulating microfilarae, yet they do not target adult worms in infected patients. DEC does both, but its function only destroys about 50% of adult worms (1).

MDA, along with morbidity management, is the main pillar of the Global Programme to Eliminate LF (GPELF), which was established in 2000. In the first 8 years of program implementation, it is estimated that a total of 9.5 million people were protected from LF infection, with a corresponding 6.0 million averted cases of hydrocele, and 3.5 million averted cases of lymphedema. A total number of disability-adjusted life years (DALYs) averted is 26 million (4). In addition to the direct benefits through LF protection, there are indirect benefits due to MDA. Specifically, throughout the scope of this MDA, it is estimated that over 56.6 million children were protected against soil-transmitted helminthes (STH), since albendazole is used in STH MDA, while 45 million people were treated with ivermectin, which has benefits in control of onchocerciasis,

scabies, and lice. Overall, over 560 million people have been treated through GPELF activities (Ibid).

A key change in 2000 was a push for integrating LF MDA with MDA for other neglected tropical diseases, including onchocerciasis, schistosomiasis, STH, and trachoma. Because many programs have responded to this push, the way that MDA is carried out has to take into account not only dosage and continuation recommendations for LF, but for other diseases as well, especially STH (8).

Transmission Assessment Surveys and Guidelines for Stopping MDA

After GPELF started implementation of the first MDA campaigns in 2000, many countries initiated and expanded their LF control and elimination programs. By 2009, some 37 countries were in the process of completing at least five years of MDA in some of their implementation units (6). MDA is an undertaking that requires considerable commitment not only from the ministries of health, local health workers, and the community, but also comes with non-negligible costs. While the drugs have been donated by pharmaceutical companies, significantly reducing costs to implementing partners, the costs can still be quite high, especially considering demands for resources for other programs. A 2011 study of communes in Haiti that received MDA found that the cost of MDA distribution in the first year of the national strategic plan in just 9 out of 55 communes to be $264,970. Extending this cost to all of the communes in the program amounts to about $1,214,102 for just one year. The estimates do not include costs of albendazole (9). Thus, there is a very real cost to implementing MDA, and it is important to know when it is time to stop the program.

The World Health Organization developed guidelines for determining whether MDA can be stopped. The areas where a decision about MDA will be reached are called evaluation units (EUs), and are usually some combination of the areas where MDA was implemented. An EU should be somewhat homogeneous as far as epidemiologic characteristics go, and transmission rates as determined by sentinel and spot-check sites should be similar in different parts of an EU (6). The size of an EU is not specified, although the upper limit on population in the unit is placed to be at 2 million. As an initial consideration, and EU must have had at least 5 rounds of MDA, with coverage of at least 65%, and the rates of microfilariae in all sentinel and spot-check sites in the EU must be less than 1%, indicating low transmission. If all of these conditions are satisfied, a Transmission Assessment Survey (TAS) is carried out to determine whether MDA can be stopped. TAS should be carried out at least 6 months following the last MDA round and, pending results, a 6^{th} round of MDA is usually carried out a year after the fifth, even if there are reasons to believe that TAS will show that transmission is sufficiently low to allow suspension of MDA (Ibid).

TAS is a survey that relies on randomization of sites of testing, as well as of survey participants. It is an example of a modified Lot Quality Assurance Sampling method, which samples a pre-determined number of participants to see if the number of these participants that tests positive for antigen or antibodies is greater than an allowed threshold. If the number is greater, the EU "fails," and continues MDA; if the number is less than or equal to the threshold, the EU is considered to "pass," and can suspend MDA (6). Following this suspension, other LF control activities should be conducted and surveillance should be continued. In addition, other factors may affect the decision to

continue MDA even if the unit "passes," given integration of soil-transmitted helminthiasis and other NTD programs with LF MDA (4, 10).

The target population for TAS is children ages 6 to 7 years, because if MDA has been successful, they should be protected against LF. Older children and adults may have antigenaemia from previous infections, yet in young children, this antigenaemia serves as a proxy for recent infection. In areas where over 75% of children are enrolled in primary schools, school-based surveys are used for TAS, whereas community-based surveys are carried out in areas with lower school enrolment (6). The tests and critical thresholds differ based on the type of LF and its vector (Appendix 2). In areas where *Wuchereria bancrofti* is the endemic parasite, and the mosquito vector is *Culex* or *Anopheles*, the upper 1-sided 95% confidence limit of antigenaemia prevalence has to be less than 2% in order for EU to "pass." The sample sizes and cut-off threshold for passing are calculated accordingly. Where the mosquito vector is *Aedes*, passing EUs have to have an upper 1-sided 95% confidence limit of antigenaemia prevalence less than 1%, because *Aedes* is a more efficient vector. If the parasite endemic to the area is *Brugia spp*., the target threshold for antigenaemia prevalence is an upper 1-sided 95% confidence limit of less than 2% (Ibid).

Stopping MDA early could lead to significant setback in LF control, so validity and reliability of TAS as a transmission assessment tool is of huge importance. A study by Chu et al, carried out in 2013, aimed to evaluate the robustness of the TAS classification over time. Specifically, the research team carried out two TAS, approximately 2 years apart, in 11 countries. All of the EUs chosen satisfied the WHO guidelines, as described above. Either school-based or community-based surveys were

conducted, based on primary school enrolment, except in case of Dominican Republic, and the parasitic species endemic to the areas varied from *W. bancrofti* to *Brugia spp*. In areas where *W. bancrofti* was endemic, an inmmunochromatographic (ICT) test was used to measure prevalence of filarial antigens. In areas where *Brugia spp*. is endemic, a BmR1 antibody test was used as the diagnostic test. When either ICT or antibody test was positive, a three-line blood smear test or real-time PCR was conducted to test for microfilariae. The study found that the results for both TAS were consistent; that is, in areas where MDA was stopped after it "passed" according to results of TAS-1, the EUs still "passed" when TAS-2 was conducted (11). However, while the study supports reliability of TAS, it does not discuss validity of TAS as a tool to assess stopping of MDA.

Need for Reevaluation of Guidelines

As the guidelines currently stand, TAS is the main tool used to make a decision with regards to stopping or continuing MDA in evaluation units, granted that the evaluation units have to satisfy initial conditions in order to qualify for TAS. Thus, it is of paramount importance to ensure that TAS is a sufficient tool to identify areas with persistent transmission where MDA should not be stopped. There is some evidence that TAS, in the way that it is currently defined, may not be a sufficient tool.

One of the first countries to implement MDA as a part of its LF elimination program, which follows GPELF guidelines, was Sri Lanka. Sri Lanka, all 8 of whose districts were endemic for LF, carried out 3 rounds of MDA with DEC alone, followed by 5 rounds of MDA with albendazole and DEC. A 2014 study by Rao et al. compares

the results of TAS conducted 6 years after the 8 years of MDA with surveillance activities carried out in smaller areas that were purposively selected based on high prevalence prior to MDA or based on post-MDA surveys. The TAS was implemented according to WHO standards, with ICT used as an initial test, and follow-up Mf blood smear testing for those that had positive ICT results. The surveillance activities included school and community surveys for antigenemia and antibodies using a card test for circulating filarial antigenemia, IgG_4 antibody test, and three-line blood smears for microfilaria, and DNA detection in mosquitoes using qPCR. The results from the TAS in all EUs were reassuring, as all EUs "passed," with the number of positive results below the critical cut-off, as determined by the WHO guidelines. Microfilarea rates were also below the accepted 1% in all of the areas tested. The study suggests a revision of upper confidence limits for circulating filarial antigenaemia in community surveys to a threshold of 2% prevalence. However, the antibody rates passed this threshold in 10 of the 19 EUs. Thus, in this case, TAS was not sufficiently sensitive to detect continual transmission of LF, which has worrisome implications not just for Sri Lanka, which has suspended its MDA due to previous TAS results, but for other programs that are nearing decision-making periods. The authors suggest that some areas of improvement for TAS include reducing the size of EUs, given that LF is a focal disease; carrying out antibody testing in school-aged children rather than antigenaemia rates; or testing adults for antigenaemia rather than children. The authors conclude that TAS, in the form that it is implemented right now, is simply not sufficient to show that LF transmission has been interrupted (12).

The characteristic of LF being a focal disease in particular places the size limit on EUs into question. Two million people in a densely populated country with homogeneous vector distribution would likely present a very different LF profile than two million people in a sparsely populated country with varying altitudes, humidity, climate zones, and vector distribution. In a heterogeneous environment, sampling from a large population may result in missing pockets of higher LF prevalence.

A paper by Drexler et al. supports the idea that LF is really a focal disease. In the study, the group tested children in schools in 5 different communes in Haiti, and then carried out systematic sampling of neighboring households of those students that tested positive. Over 40% of people that tested positive lived within 20 m of the child that was identified through the school survey. This clustering indicates that LF transmission may, in fact, be occurring at a microfocal level (13).

While reducing the size of EU may improve the chances of including pockets that may have higher prevalence of LF, if they exist, reducing the size of EU and thus increasing the number of EUs would come with increased costs. The average cost of a community-based TAS, based on a 2013 study in 11 countries, is $26,800, whereas the average cost of a school-based TAS is $24,900 (14). Given the limitation of resources available to LF elimination programs, the guidelines for EU size should thus be carefully evaluated.

Diagnostic tests

Given the need to correctly identify cases of LF infection as well as individuals who are not infected, diagnostic tools are of great importance. There does not currently

exist a gold standard diagnostic tool, but, as in some other neglected tropical diseases, LF saw development of faster and easier tests.

A method that has been in use the longest is the use of blood films to detect the presence of microfilarea. Prior to GPELF, LF control programs used this technique to mass screen for infection. One of the difficulties associated with this technique is periodicity of Mf. Microfilarea have nocturnal periodicity, except in the South Pacific (2), so samples collected for use in blood films had to be collected at night (6, 15). In addition, while this test is quite specific, it is not very sensitive, and may miss active infections and those with low Mf counts (15).

A more sensitive test than the blood film approach to Mf detection is antigen testing. This test is available for *Wuchereria bancrofti*, and can be done either as a lab-based ELISA test, Og4C3, or as a rapid immunochromatographic test, ICT. The test is convenient, because it can be done at any time, and ICT in particular is applicable in settings where access to lab is difficult to obtain. However, as the test measures presence of adult nematode antigen that is circulating in blood, antigen may stay in blood for several months or even years after adult worms and microfilariae die and disintegrate, such as after MDA. In addition, antigen testing is costly, and is not available for *Brugia spp* (6, 15).

Filarial antibody tests are available for all species, including *Brugia spp*. Specifically, the Brugia Rapid™ test detects antibodies to both *B. malayi* and *B. timori*. Antibody testing is highly specific, and detects IgG_4 antibodies to filarial antigens like Bm14 and B*m*R1, depending on the species. Antibody testing tends to be more sensitive than antigen and Mf rates in young children in areas endemic for LF, but these tests

cannot distinguish between infections that are current or that took place in the past (6, 15).

Microscopy or PCR to assess mosquito infection levels have also been used in the past. Microscopy has been used after a mosquito has been dissected to verify whether there are filarial larvae within mosquitoes (assessing mosquito infection) and whether the larvae present are infective (assessing infectivity). However, in areas where mosquito infection rates are low, such as after several rounds of MDA, this technique is not practical (15). An alternative technique is molecular xenomonitoring followed by PCR for presence of the parasite in mosquitoes. While this allows assessment of mosquito infection, it does not allow one to assess infectivity or rates of transmission (6, 15).

Physical observation, such as tracking of hydrocele, lymphedema, and elephantiasis presentation, is not sensitive, given that majority of LF infection does not manifest in symptoms (1, 4, 15).

Finally, urinary tests are in development, as there are some concerns with cross-reaction to loiasis infection and onchocerciasis (6).

As seen from the description above, no single test can be used as a gold standard, and yet programmatic decisions are made with reliance on these tests. A 2012 study by Gass et al. analyzed effectiveness of different tests in detection of *Wuchereria bancrofti* in a region that has implemented several rounds of MDA. The expected prevalence of infection was between 0.5 and 2%, and samples were acquired through community and school surveys. The following tests were used: blood smear and Og4C3 to measure Mf; ICT and Og4C3 for antigen detection; and PanLF, Bm14 and Urine SXP antibody tests. Blood collection for microfilaremia had to be performed at night, given that *W. bancrofti*

has nocturnal periodicity. The study found a high range of specificity and sensitivity of each test in different countries and settings. For example, ICT had 61% sensitivity in Ghana, but 79% in Haiti and French Polynesia; its specificity ranged from 89% in Haiti to 94% in Ghana. Lab-based tests had greater sensitivity as compared to rapid tests, and yet this advantage was outweighed by convenience and standardization of rapid tests. The Bm14 antibody test produced the highest prevalence of positive results, followed by PanLf and urine SXP. The higher prevalence of positive antibody tests makes sense, since the tests react both to present and past infections. Among antigen tests, ICT produced highest prevalence of positive results, followed by PCR and blood smear. The paper argues for the use of ICT as a primary tool in the TAS, given its quick results, ease of use, relative affordability, and practicability. ICT detects the presence of adult worms, which can serve as a proxy of continual transmission. However, some additional training may be necessary to ensure that those who administer the test know how to interpret a result with a weak signal (16).

Haiti: Overview

Haiti, a country with population of 10.6 million people in 2014, has a population density of 378.5 people per square kilometer of land area. As of 2013, the average life expectancy in Haiti is 62.4 years, with mortality under 5 years of age standing at 73.1 deaths per 1,000. In 2014, 57.5% of population in Haiti had access to a water source that could be classified as improved, a definition that includes public water taps, rainwater collection, piped water in the household or nearby, tube wells or boreholes, protected dug wells, or protected springs. Access to sanitation facilities classified as improved was available to 27.4% of Haitians in 2014. Since 1960s, there has been a decrease in mean

annual rainfall by 5 mm per month per decade, whereas vulnerability to floods has increased, due to increase in deforestation and lack of infrastructure that allows for drainage (17).

Following the 2010 earthquake, many of the public health programs suffered a setback. The earthquake, which killed 230,000 people and injured another 300,000, destroyed large amounts of infrastructure, and resulted in over 1 million internally displaced people (IDPs) (18). HIV and TB services, like screening and therapy enrolment, initially dropped dramatically, but have recently recovered. For instance, the percentage of pregnant women identified as HIV positive and receiving antiretroviral treatment through the United States President's Emergency Plan for AIDS Relief (PEPFAR) dropped from 44% in 2009, just before the earthquake, to 32% in 2010, but increased to 87% by 2014. A similar trend was seen in TB therapy enrolment, with increase in detection and notification of active TB cases boosted by improvements in national lab capacity. The national sentinel surveillance system has similarly grown to 153 sites nationwide, and provides surveillance for immediately notifiable diseases. In addition to HIV and TB, there are other control and elimination programs in Haiti, including neglected tropical diseases, malaria, and rabies, among others. In particular, Haiti has the highest incidence of human rabies in the Western Hemisphere, with number of canine rabies cases rising sharply from 2012 to 2013 (19).

The post-earthquake period has seen a sharp increase in funding to disease control programs in Haiti. The Water, Sanitation, and Hygiene (13) sector in particular has received a lot of attention. In 2013, Haiti launched the National Plan as a part of a larger regional "Call to Action" to eliminate cholera within 10 years. This plan includes

interventions in 4 areas – WASH, health care services and management, health and hygiene promotion, and epidemiology and surveillance, and has a budget of over $1.6 billion for WASH activities alone. Regardless of this investment, progress in WASH as well as other health sectors has had its challenges, as much of post-earthquake response was focused on recovery rather than longer, more sustainable programs. In its sixth year after the earthquake, Haiti is seeing many of the shorter-term programs wrapping up. In addition, this period has been marked by low level of coordination between different organizations providing interventions. In the WASH sector alone, over 100 NGOs responded to the earthquake, and while many of their interventions overlapped, there was little synergy. Aside from the post-earthquake challenges, the WASH program in Haiti is characterized by disparity between rural and urban populations, with 85% of the urban population having access to an improved water source as compared to 51% of the rural population in 2010. Similarly, 24% of the urban population had access to improved sanitation, as compared to only 10% of the rural population in the same year. For comparison, the regional average improved sanitation coverage in 2010 was 80% (18).

Lymphatic Filariasis in Haiti

Haiti is one of four countries in the Americas endemic for LF, bearing some 90% of LF disease burden in the region. The species endemic to Haiti is *Wuchereria bancrofti*, the primary vector of which is the *Culex quinquefasciatus* mosquito. In 2001, the antigen prevalence among children aged 6 to 11 was between 0 and 45%, with over 88% of all communes showing prevalence greater than 1% and thus qualifying for MDA according to WHO standards (13). In 2000, with support from the Ministry of Public Health and the

Population (MSPP), National Program to Eliminate LF (NPELF) was established. LF mapping was carried out, using ICT as a diagnostic tool, and communes were designated as either high prevalence (≥10% positive ICT results), moderate (5 – 9.9%), low (0.1 – 4.9%) or non-endemic (0%). Twenty communes were classified as high prevalence, and 13 had moderate prevalence of LF. The prevalence of LF was not equally distributed across all communes, with high prevalence communes located in the northern plains and to the coastal plains north, west, and east of the capital city. In addition, there was evidence for focal infection (20).

In the first years of NPELF, treatment was limited to high prevalence communities, with the scope of the program increasing as allowed by budget. The first commune to receive MDA was Léogâne, with the program implemented in partnership with University of Notre Dame, Hôpital Ste. Croix, Centers of Disease Control and Prevention, and Interchurch Medical Assistance. MDA was limited to Léogâne from 2000 to 2003. However, due to political crises and violence in 2003-2005, followed by withdrawal of funding, the LF program experienced a setback and saw an increase in prevalence. After funding was renewed and more partners supported the NPELF, including the Bill & Melinda Gates Foundation, USAID, CBM, the Inter-American Development Bank (IDB), the Abbott Foundation, PepsiCo, and the Frank Eck Family Foundation, MDA was scaled up and the cost of treatment per person dropped from $2.23 in 2000 to $0.50 in 2009. Even with hurricanes, a devastating earthquake, and a cholera outbreak, by 2012, NPELF was able to implement MDA nationwide, reaching more than 8 million people, with estimated coverage around 71% (10).

While remarkable progress has been made in Haiti, there are still many challenges to face. By 2011, all endemic areas in Haiti have received at least one round of MDA, except for the capital, Port-au-Prince, and five surrounding communes. These communes are considered to be a "challenging area" because of the many internally displaced people, high levels of migration, disruption of infrastructure, and lack of access to natural resources following the 2010 earthquake. The lack of healthcare resources has been exacerbated, and the national health system currently reaches 47% of population. While MDA was carried out in Port-au-Prince and surrounding areas in November 2011 – February 2012, coverage was low. Some 9.7 million Haitians are considered to be living at risk for or infected with LF (21).

An additional challenge in Haiti is that in some regions, LF transmission continues even after many rounds of MDA. A 2010 study done in Léogâne after 7 rounds of MDA found that LF prevalence was still greater than 10%, which would place the commune in the high prevalence classification. However, in 2005, Mf prevalence was shown to be less than 1%, which would make the commune eligible for stopping MDA. This continual transmission even after stopping criteria has been met is alarming, to say the least. The authors of the study identified MDA compliance as a factor contributing to continued transmission. It appears that some individuals have never participated in the MDA, and serve as a reservoir for infection. Additional factors that may be considered are heterogeneity of transmission, as authors found significant clustering of antigen positive households, as well as vector density and a missed round of MDA (22). This clustering of households is consistent with findings by Drexler et al. and Beau de Rochars et al., which suggests that transmission may occur in microfoci (13, 20), and transmission

could be partially explained by cane cultivation and processing that are key features of Léogâne commune and a good habitat for *Culex (20)*.

It is of note that NPELF is not simply limited to MDA. The program is considering a wide distribution of salt fortified with DEC, which has been locally manufactured since 2005, and has implemented programs targeting exposure to mosquitoes, such as distribution of long-lasting insecticide-treated nets. However, the *Culex* vector is not as sensitive to the insecticides on the nets. Finally, there has been push for morbidity management programs, patient support groups, and training programs for hydrocele surgery. While lymphedema management support has been established at two referral centers, resulting in treatment of more than 1,500 patients, the program has been plagued by funding issues and has struggled with continuity (10).

Vector distribution

The vector of *W. bancrofti* in Haiti is *Culex quinquefasciatus*, which is quite effective. An experiment carried out in 1998 showed that 21 days after non-infected mosquitoes fed on people with microfilaria, 216 out of 476 mosquitoes that were dissected yielded 860 infective larvae, which amounts to quite a high infectivity rate. In addition, mosquitoes that took their blood meal on people with microfilaria had similar survival rates to those mosquitoes that fed on amicrofilarial people, showing that the mosquito was not harmed by the parasite (23). Another study addressed a question of whether patients that had low levels of microfilaria were an important part of LF transmission. The study was conducted in Léogâne, Haiti, which is known to be a high prevalence community. Participants of the study were treated with diethylcarbamazine

citrate (DEC-C), and lab-bred mosquitoes were released to feed on the patients under mosquito nets while the patients slept. While in general, the microfilarial uptake was proportional to the patient's Mf level, the mean density of uptake was much higher than expected. The infectivity ratio was 0.3 third-stage (infective) larvae per mosquito in the ultralow density group, and was 0.8 third-stage (infective) larvae per mosquito in the low density group, which suggests that people with low levels of microfilaria are still important carriers and players in transmission. It is also a testimony to effectiveness of *Culex quinquefasciatus* as a vector (24).

As for the habitat itself, *C. quinquefasciatus* requires nutrient-rich standing water for ovipository. If the water evaporates before eggs hatch or the larvae cycle is complete, the mosquito progeny cannot survive. Thus, access to standing water is key for continued transmission. Bird baths, tires, any containers that hold water are common depositories for the mosquito eggs (25). Rum distilleries, as common in Léogâne, have also been identified as good *Culex* habitats (24). In Hawai'i, many anthropogenic and naturally-occurring habitats of *Cx. quinquefasciatus* have been identified. Examples of anthropogenic habitats include open-topped cisterns and short sections of exposed pipes; naturally-occurring habitats include stream drainages, tree holes and cavities, and rock holes. Proportion of larvae in anthropogenic habitats was found to be higher than in naturally occurring ones in many areas (26).

In Haiti, human-made changes, as well as the earthquake in 2010, have led to changes in the environment that affect mosquito habitat and ecology. Displacement of some 2.3 million people following the earthquake increased the number of informal settlements and camps. Breakdown in infrastructure, overcrowding, and poverty can lead

to change in vector population density, minimum infection rates, biodiversity, and vector development sites. In addition, change in land cover due to urbanization and displacement can increase availability of mosquito breeding sites, particularly through water storage practices, ponding, and waste disposal. Tires, bottles, buckets, ditches, and temporary pools of water comprise some of the common human-made habitats in Haiti. *Cx. quinquefasciatus* in particular has high probability of residing around urban areas, and displacement increases risk of LF transmission, as supported by the species distribution model. Finally, it is not uncommon to see flooding in some areas of Haiti during the rainy season, which, combined with accumulation of trash and drain cloggage, results in standing stagnant water (27).

All of these factors contribute to persistence of lymphatic filariasis in Haiti, and must be taken into account in any LF elimination program.

References

1. Shenoy RK, Bockarie MJ. Lymphatic filariasis in children: clinical features, infection burdens and future prospects for elimination. *Parasitology* 2011;138(12):1559-68.
2. Centers for Disease Control and Prevention. Parasites - Lymphatic Filariasis. CDC; 2013. (http://www.cdc.gov/parasites/lymphaticfilariasis/index.html). (Accessed January 20, 2016 2016).
3. Brady MA, Hooper PJ, Ottesen EA. Projected benefits from integrating NTD programs in sub-Saharan Africa. *Trends in parasitology* 2006;22(7):285-91.

4. Ottesen EA, Hooper PJ, Bradley M, et al. The Global Programme to Eliminate Lymphatic Filariasis: Health Impact after 8 Years. *PLoS neglected tropical diseases* 2008;2(10).

5. World Health Organization. Lymphatic Filariasis: Managing Morbidity and Preventing Disability, an Aide-Mémoire for National Programme Managers. 2013.

6. World Health Organization. Lymphatic Filariasis: TAS Manual. 2011.

7. Eddy BA, Blackstock AJ, Williamson JM, et al. A longitudinal analysis of the effect of mass drug administration on acute inflammatory episodes and disease progression in lymphedema patients in Leogane, Haiti. *The American journal of tropical medicine and hygiene* 2014;90(1):80-8.

8. World Health Organization. Preventive chemotherapy in human helmianthiasis: Coordinated use of anthelminthic drugs in control interventions: a manual for health professionals and programme managers. *WHO technical report* 2006.

9. Goldman AS, Brady MA, Direny A, et al. Costs of integrated mass drug administration for neglected tropical diseases in Haiti. *The American journal of tropical medicine and hygiene* 2011;85(5):826-33.

10. Oscar R, Lemoine JF, Direny AN, et al. Haiti National Program for the elimination of lymphatic filariasis--a model of success in the face of adversity. *PLoS neglected tropical diseases* 2014;8(7):e2915.

11. Chu BK, Hooper PJ, Bradley MH, et al. The economic benefits resulting from the first 8 years of the Global Programme to Eliminate Lymphatic Filariasis (2000-2007). *PLoS neglected tropical diseases* 2010;4(6):e708.

12. Rao RU, Nagodavithana KC, Samarasekera SD, et al. A comprehensive assessment of lymphatic filariasis inSri Lanka six years after cessation of mass drug administration. *PLoS neglected tropical diseases* 2014;8(11):e3281.

13. Drexler N, Washington CH, Lovegrove M, et al. Secondary mapping of lymphatic filariasis in Haiti-definition of transmission foci in low-prevalence settings. *PLoS neglected tropical diseases* 2012;6(10):e1807.

14. Chu BK, Deming M, Biritwum NK, et al. Transmission assessment surveys (TAS) to define endpoints for lymphatic filariasis mass drug administration: a multicenter evaluation. *PLoS neglected tropical diseases* 2013;7(12):e2584.

15. Weil GJ, Ramzy RM. Diagnostic tools for filariasis elimination programs. *Trends in parasitology* 2007;23(2):78-82.

16. Gass K, Beau de Rochars MV, Boakye D, et al. A multicenter evaluation of diagnostic tools to define endpoints for programs to eliminate bancroftian filariasis. *PLoS neglected tropical diseases* 2012;6(1):e1479.

17. The World Bank. Data: Haiti. 2016.

18. Gelting R, Bliss K, Patrick M, et al. Water, sanitation and hygiene in Haiti: past, present, and future. *The American journal of tropical medicine and hygiene* 2013;89(4):665-70.

19. Domercant JW, Guillaume FD, Marston BJ, et al. Update on progress in selected public health programs after the 2010 earthquake and cholera epidemic--Haiti, 2014. *MMWR Morbidity and mortality weekly report* 2015;64(6):137-40.

20. Beau de Rochars MV, Milord MD, St Jean Y, et al. Geographic distribution of lymphatic filariasis in Haiti. *The American journal of tropical medicine and hygiene* 2004;71(5):598-601.

21. Byrne SK, Collins SD. Lymphatic Filariasis in Children in Haiti. *MCN The American journal of maternal child nursing* 2015;40(4):227-33.

22. Boyd A, Won KY, McClintock SK, et al. A community-based study of factors associated with continuing transmission of lymphatic filariasis in Leogane, Haiti. *PLoS neglected tropical diseases* 2010;4(3):e640.

23. Calheiros CM, Fontes G, Williams P, et al. Experimental Infection of Culex (Culex) quinquefasciatus and Aedes (Stegomyia) aegypti with Wuchereria bancrofti. *Memórias do Instituto Oswaldo Cruz* 1998;93:855-60.

24. Lowrie RC, Jr., Eberhard ML, Lammie PJ, et al. Uptake and development of Wuchereria bancrofti in Culex quinquefasciatus that fed on Haitian carriers with different microfilaria densities. *The American journal of tropical medicine and hygiene* 1989;41(4):429-35.

25. Hill SC, R. Featured Creatures: Southern house mosquito. University of Florida; 2013. (http://entnemdept.ufl.edu/creatures/aquatic/southern_house_mosquito.htm). (Accessed 17 January 2016 2016).

26. Reiter ME, Lapointe DA. Larval habitat for the avian malaria vector Culex quinquefasciatus (Diptera: Culicidae) in altered mid-elevation mesic-dry forests in Hawai'i. *Journal of vector ecology : journal of the Society for Vector Ecology* 2009;34(2):208-16.

27. Samson DM, Archer RS, Alimi TO, et al. New baseline environmental assessment of mosquito ecology in northern Haiti during increased urbanization. *Journal of vector ecology : journal of the Society for Vector Ecology* 2015;40(1):46-58.

Chapter II: Manuscript

Simulating the effect of evaluation unit size on eligibility to stop mass drug administration for lymphatic filariasis in Haiti

Natalya Kostandova[1], Amy Kirby[1], Katherine Gass[7]

[1]Rollins School of Public Health, Emory University, USA [2]NTD Support Center, Task Force for Global Health, USA

Abstract

Background: As more and more countries have implementation units that are entering their sixth year of annual mass drug administration (MDA) for lymphatic filariasis (LF), there is a need to assess whether transmission has been reduced below the threshold required for sustainable transmission and MDA can be stopped. Currently, the main tool used to assess this threshold is the Transmission Assessment Survey (TAS). The guidelines for TAS limit the population in the evaluation units (EUs) to <2 million. This study uses simulations to investigate the effect of population size on the classification of EUs as either passing or failing TAS.

Methodology: The data come from TAS conducted in 14 EUs in Haiti during 2014-2015. To simulate the effects of population size, larger "combination-EUs" were created by forming eight combinations of adjacent EUs. Several approaches to simulate TAS were carried out, with the intent to see how the classification of EUs as passing or failing TAS would change when the larger units were considered.

Principal findings: The results of the simulations show that in some combinations, the vast majority of the time the classification of the combination-EU would agree with both the expected decision, obtained by passing or failing the combination-EU based on the overall number of positive results, sample size and target population from the original

EUs, and the desired decision, that is the programmatic decision to fail any combination-EU in which one or more of the original EUs failed. However, a non-negligible proportion of combination-EUs were misclassified.

Conclusions: Misclassifying an EU as failing would result in continued MDA in a region where prevalence is lower than the threshold required for sustainable transmission, translating into additional rounds of MDA and TAS. Misclassifying an EU as passing when transmission persists is even more troubling, given the programmatic implications of stopping MDA too early. The guidelines should be carefully reconsidered to ensure that TAS is a valid and reliable tool to assess the possibility of stopping LF MDA.

Introduction

Lymphatic filariasis (LF) is a vector-borne disease caused by nematodes, or roundworms, that invade lymphatic tissue (1). With over 120 million people affected by the disease, and 1.3 billion at risk, lymphatic filariasis is the second most common cause of physical disability among neglected tropical diseases. While LF does not have a single causative agent, *Wuchereria bancrofti* is the most common cause, followed by *Brugia malayi* and, very rarely, *Brugia timori* (Ibid).

While some 67% of the infections are asymptomatic, in some patients the infection can manifest in swelling. Specifically, 12.5% of infected people develop chronic lymphedema and elephantiasis, which affect mostly limbs and, in some cases, the genital region (1, 2). Chyluria and hydrocele can also take place (20.8% of infected). Overall, the microfilarea (Mf) infection impedes the functioning of the lymph system,

and the clinical manifestations, which affect some 44 million people, can lead to debilitating disability, as well as stigma, psychological problems, and lowered quality of life (1, 3).

Once the clinical manifestations of LF set in, they are difficult to reverse, so the main cornerstone of LF control is prevention through Mass Drug Administration (MDA). The primary objective of MDA is to lower the level of microfilaraemia in individuals that have been infected so that, even after MDA is stopped, transmission cannot continue (4). The World Health Organization recommends annual distribution of selected drugs to all those living in areas at risk. The drugs distributed are diethylcarbamazine (DEC) and albendazole in areas without co-endemicity with onchocerciasis, and ivermectin and albendazole where onchocerciasis is co-endemic. The main course of action of the drugs administered is to lower the number of Mf that circulate in blood, thus preventing further transmission when mosquitoes take blood meal (Ibid).

MDA, along with morbidity management, is the main pillar of the Global Programme to Eliminate LF (GPELF), which was established in 2000. In the first 8 years of implementation of the program, over 560 million people have been treated, with an estimated total of 9.5 million people protected from LF infection, averting 26 million disability-adjusted life years (3).

By 2009, some 37 countries were in the process of completing at least five years of MDA in some of their implementation units (4). A 2011 study of communes in Haiti that received MDA found that the cost of MDA distribution in the first year of the national strategic plan in just 9 out of 55 communes to be $264,970. Extending this cost to all of the communes in program amounts to about $1,214,102 for just one year, not

including the cost of albendazole (5). Thus, there is a very real cost to implementing MDA, and it is important to know when it is time to stop the program.

The World Health Organization developed guidelines for determining whether MDA can be stopped. The areas where a decision about MDA will be reached are called evaluation units (EUs), and are usually some combination of the areas where MDA was implemented. An EU should be somewhat homogeneous as far as epidemiologic characteristics go, and transmission rates as determined by sentinel and spot-check sites should be similar in different parts of EU (4). The size of an EU is not specified, although the upper limit on population in the unit is placed to be at 2 million. As an initial consideration, an EU must have had at least 5 rounds of MDA, with coverage of at least 65%, and the rates of microfilariae in all sentinel and spot-check sites in EU must be less than 1%, indicating low transmission. If all of these conditions are satisfied, a Transmission Assessment Survey (TAS) is carried out to determine whether MDA should be stopped (Ibid).

The target population for TAS is children of ages 6 to 7 years. In areas where over 75% of children are enrolled in primary schools, school-based surveys can be used for TAS, whereas community-based surveys are required in areas with lower school enrolment (4). The tests and critical thresholds differ based on the type of LF and its vector. In areas where *Wuchereria bancrofti* is the endemic parasite, and the mosquito vector is *Culex* or *Anopheles*, the upper 1-sided confidence limit for the antigenaemia prevalence has to be less than 2% in order for EU to "pass." The sample sizes and cut-off threshold for passing are calculated accordingly. Where the mosquito vector is *Aedes*, passing EUs have to have an upper 1-sided confidence limit for the antigenaemia

prevalence less than 1%, because *Aedes* is a more efficient vector. If the parasite endemic to the area is *Brugia spp*., the target threshold is based on LF antibodies and is set at an upper 1-sided confidence limit of 2% (Ibid).

TAS is an example of a modified Lot Quality Assurance Sampling method, which samples a pre-determined number of participants to see if the number of participants that test positive for antigen or antibodies is greater than an allowed threshold. If the number is greater, the EU "fails" and continues MDA; if the number is less than or equals the threshold, the EU is considered to "pass," and can stop MDA (4).

Haiti is one of four countries in the Americas endemic for LF, bearing 90% of LF disease burden in the region. The species endemic to Haiti is *Wuchereria bancrofti*, the primary vector of which is the *Culex quinquefasciatus* mosquito. In 2001, the antigen prevalence among children aged 6 to 11 was between 0 and 45%, with over 88% of all communes showing prevalence greater than 1% and thus qualifying for MDA according to WHO standards (6). In 2000, with support from the Ministry of Public Health and the Population (MSPP), National Program to Eliminate LF (NPELF) was started. Even with hurricanes, a devastating earthquake, and a cholera outbreak, by 2012, NPELF was able to implement MDA nationwide, reaching more than 8 million people, with estimated coverage around 71% (7). Many of the implementation units have reached or are now reaching 5th year of implementation, qualifying for TAS and potential suspension of MDA.

There is some evidence that TAS, as it is currently defined, may not be an effective tool for determining MDA stoppage (8). The focality of LF infection places the liberal size allowance for EUs into question. Two million people in a densely populated

country with homogeneous vector distribution would likely present a very different LF profile than two million people in a sparsely populated country with varying altitudes, humidity, climate zones, and vector distribution. In a heterogeneous environment, sampling from a large population may result in missing pockets of higher LF prevalence.

While reducing the size of an EU may improve the chances of including pockets with persistent transmission of LF, if they exist, reducing the size of an EU and thus increasing the number of EUs would come with increased costs. The average cost of a community-based TAS, based on a 2013 study in 11 countries, is $26,800, whereas the average cost of a school-based TAS is $24,900 (9). Given the limitation of resources available to LF elimination programs, the guidelines for EU size should thus be carefully evaluated, in light of feasibility of implementation. At the same time, the additional costs of TAS in smaller EUs should be weighed against the costs of additional rounds of MDA, as well as the costs of misclassifying EUs. In this study, the effect of using larger EUs for classifying the area as either passing or failing will be explored through formation of combinations of adjacent EUs.

Methods

Dataset

The dataset utilized in this study is a subset of data from a TAS-STH-Malaria survey conducted by the Centers of Disease Control and Prevention (CDC) in 2015 in Haiti. The transmission assessment survey was conducted in 14 evaluation units (EUs), with each unit comprised of one or more communes, with the exception of one evaluation

unit that was smaller than a commune. All evaluation units had completed TAS requirements as established by WHO: at least 5 rounds of MDA, with coverage over 65%; microfilarea prevalence at sentinel and spot-check sites of <1%; and a total population under 2 million people. TAS was conducted as a randomized cluster or systematic survey targeting children 6-7 years old, with schools as the primary sampling unit. Immunochromatographic card test (ICT) was the diagnostic tool used to test for the presence of antifilarial antigens. The data includes the names of each evaluation unit, the names and locations for each school, the ages and sex of the children tested, and the ICT results. The Survey Sample Builder data provided by the CDC was used to obtain information about the target population, total number of schools, and expected absentee rates for each EU.

Forming conglomerates

Eight combinations of evaluation units were obtained by combining adjacent evaluation units. Each of these EU combinations – hereby referred to as 'conglomerates' – represents an alternative EU that the NPELF could have designated for basing its stopping MDA decision, since they satisfy all of the guidelines specified by WHO. Target populations for each conglomerate were obtained by combining target populations for each EU comprising the conglomerate. The total number of schools in the conglomerate was taken to be the sum of schools in each participating evaluation unit. The expected absentee rate for each individual evaluation unit varied from 10% to 15%; since each of the conglomerates contained at least one evaluation unit with expected absentee rate of 15%, all of the conglomerates were assigned the expected absentee rate

of 15%. Because the target population of each of the combinations exceeded 1000 and the number of schools in each combination exceeds 40, cluster sampling was assumed, as per WHO TAS protocol. Following the same protocol, combinations with target population below 2400 people were assigned a design effect of 1.5; those achieving and exceeding 2400 people were assigned a design effect of 2. The WHO TAS table was used to obtain the necessary sample size for a transmission assessment survey in the conglomerates (4). This sample size was scaled according to the expected absentee rate to obtain the target sample size for that combination. An average number of students per school was obtained by dividing the total target population by the total number of schools. The sample size was divided by the average number of students to obtain the number of schools that needed to be sampled for each conglomerate.

Passing or failing decision

The desired decision for a conglomerate here is defined as passing if all individual EUs pass according to guidelines presented in the WHO TAS manual[1], which classifies a unit as either passing or failing by comparing the number of positive ICT results in the unit to a threshold, based on an adapted LQAS method. For sample sizes larger than those listed in the LQAS table for *Culex* vector, passing or failing was determined based on hypergeometric probability. If any of the individual EUs fail, the desired decision for the conglomerate is to fail.

---

[1] Target threshold of antigenaemia prevalence is 2%, since Haiti is endemic for *W. bancrofti*; the critical cut-off for this threshold is identified so that each conglomerate has no more than 5% of being misclassified as passing, and at least 75% chance of correctly passing if the prevalence of antigenaemia is 1.0%

The expected decision for the conglomerate is obtained by weighing individual evaluation units and calculating the expected point prevalence of positive ICT results in the conglomerate. If this prevalence exceeds or equals 2%, then the expected decision is to fail the conglomerate; if the prevalence is less than 2%, then the expected decision is to pass.

Bootstrapping

For all subsequent analysis, SAS v9.3 (Cary, NC) was used. Bootstrapping here refers to sampling with replacement from the TAS data.

In **Phase 1** of the analysis, for each EU conglomerate, the required number of school clusters for a TAS was sampled with replacement at the school level from the available data. For each selected school, all of the observations for that school were included. This sampling was replicated 1000 times, resulting in 1000 TAS results. The total number of positive ICT results in each of the bootstrap replicates was obtained, and a decision to pass or fail the TAS as a conglomerate was made, based on the pre-determined threshold of cases for that conglomerate, as defined above. These thresholds were used even if the target sample size was not achieved. The proportion of passing and failing replicates for each conglomerate was calculated.

Because in Phase 1 the thresholds were calculated for a sample size that was often not achieved, the next step was to adjust the thresholds appropriately. In **Phase 2**, the same bootstrap replicates were used, but the thresholds were adjusted according to the actual sample size obtained in each replicate. That is the threshold applied to each bootstrap replicate was taken from the WHO TAS manual table according to the actual

sample size achieved, and not the target sample size. Most often this resulted in a reduction in the threshold of positive cases; however, if the sample size was larger than the values listed in the WHO TAS manual, the replicate passed if the hypergeometric probability was smaller than 0.05. The total number of passing and failing replicates for each conglomerate was obtained.

While the thresholds in Phase 2 were more appropriate, given that they were calculated based on actual sample sizes achieved, the sample sizes were still often less than desired. In a real TAS setting, if target sample size was not achieved, additional schools would be visited until the requisite sample size was reached. Phase 3 provides this simulation. In **Phase 3**, the same bootstrapping process was carried out as in the previous phases. The sum of positive and negative ICTs was obtained for each bootstrap replicate. If this sample size did not reach the sample size required, as determined by WHO TAS standards, additional schools were bootstrapped and incorporated into that replicate, until every one of the 1000 bootstrap replicates for each conglomerate had at least achieved the target sample size. A passing or failing decision was made for each replicate, as in previous phases, and a total number of passing and failing replicates for each conglomerate was obtained.

The overwhelming undershooting of target sample size had occurred because of two principal factors – low rate of signed parental consent and higher absentee rate than expected. Because expected absentee rate affects the number of clusters selected for initial sampling, it may be of interest to explore the effect of expected absentee rate on passing or failing of EUs. In **Phase 4**, sensitivity analysis for different expected absentee rates was performed for absentee rates of 30%, 40%, and 60%. The new required sample

sizes were obtained for each of the conglomerates based on these absentee rates, and the number of schools to be sampled was obtained by dividing these numbers by the average number of children in schools. Using this new number of schools, analyses for Phase 1, Phase 2, and Phase 3 were repeated.

In **Phase 5**, the upper 1-sided 95% confidence interval for the prevalence of positive ICTs was calculated for direct comparison against the TAS threshold (2%). The conglomerates were bootstrapped 1000 times as in previous phases. A "proc surveyfreq" procedure was performed to obtain the Clopper-Pearson exact confidence intervals, as well as estimated design effects in each of the bootstrapping replicates for each of the conglomerates. If the upper confidence interval exceeded 2% prevalence, then the replicate was identified as failing; if the upper confidence interval was less than 2%, then the replicate was considered to pass. A total number of passing and failing replicates for each of the conglomerates was calculated.

A flow chart of Phases 1 through 5 is included as **Figure 1**.

The next phase of analysis addressed the issue that schools from EUs that employed systematic sampling in the original TAS (e.g., EUs with few schools) were overrepresented in bootstrapping when compared to schools from EUs with cluster sampling because schools in the component EUs were bootstrapped with equal probability. In **Phase 6**, the total number of schools selected from each conglomerate was the same as in Phases 1-3, and Phase 5; however, school selection was stratified by EU and schools were bootstrapped independently from each EU, with the number of selected schools proportional to the total number of schools in the EU. The schools obtained from

each of the EUs for each replicate were combined together. A passing or failing decision was reached, as before, and further analysis remained the same.

Because only a fraction of students from each school was sampled in systematic EUs, the number of schoolchildren from such EUs is consistently misrepresented. To address this issue, the children sampled in systematic EUs are assumed to be representative of all schoolchildren in the EU. In **Phase 7**, bootstrapping was completed for conglomerates as in Phase 6, with the number of schools selected in each individual EU proportional to the total number of schools in that EU. However, for EUs that were sampled systematically rather than through cluster sampling, additional bootstrapping was completed at the schoolchild level in order to obtain necessary sample size. The number of schoolchildren selected in these schools was equal to the average number of children per school, calculated for the individual conglomerates. For EUs with cluster sampling, bootstrapping was only done at the school level. The number of passing and failing replicates was calculated as before.

In order to address the replicates that did not achieve desired sample size in Phases 6 and 7, additional schools were sampled until all replicates reached the target sample size. Sampling until target sample size was reached for Phase 6 was denoted as **Phase 8,** whereas amending Phase 7 to reach target sample size became **Phase 9**. Because systematic EUs in Phase 7 already reached sample size, in Phase 9, only schools in cluster EUs were sampled.

A flowchart of Phases 6 – 9 is included as **Figure 2**.

Results

The TAS Dataset

Information pertaining to characteristics of the evaluation units and TAS results is presented in Table 1. Fourteen total EUs were sampled in TAS, with target population ranging from 707 children in EU #5 to 35,357 children in EU #2. The number of schools in the EUs ranged from 17 in EU #5 to 721 in EU #2. The range of average students in target grades is from 29 in EU #9 to 64 in EU #7. EUs #1 – 6 had expected absentee rate of 10%, whereas EUs #7 – 14 had expected absentee rate of 15%.

The number of schools tested per EU spans from 16 in EU #7 to 53 schools in EU #2. Four of the EUs were small enough to be sampled systematically. That is, all schools that were accessible were sampled. The remaining ten EUs were sampled through cluster survey, with number of schools selected from those EUs ranging from 31 to 53. In cluster surveys, all children in target grades were tested for antigenaemia using the ICT test, whereas in systematic EUs, only a fraction of students in target grades at each school were tested. The total number of children tested per EU ranged from 364 in EU #5 to 1986 in EU #13.

Two of the EUs, EU #11 and EU #12, failed. That is, the number of positive ICT results obtained in these EUs exceeded the cut-off threshold for that EU. EU #13 came close to reaching the threshold, with 19 positive ICTs, as compared to the threshold of 20. All other EUs passed, with the number of positive ICT results seen lower than the acceptable cut-off threshold.

The EUs and the locations of schools where the survey was conducted are displayed in **Figure 3**.

Forming conglomerates

Conglomerates were formed by combining adjacent EUs. Combinations of EUs with no positive ICT results or an extremely small number of results, such as EU #2 and EU #1, or EU #7 and EU #8, were not considered. In total, eight conglomerates were formed, as presented in Table 2.

As seen in the table, the conclusion reached using expected prevalence of positive ICT results differs from the desired conclusion for 7 out of 8 combinations. That is, while the desired programmatic conclusion is for the new combination to fail if at least one of its component EUs fails, and pass only if all of its component EUs pass, in all combinations, the expected point prevalence is less than 2%. From here on, this decision will be referred to as expected decision.

All of the combinations were large enough to merit a cluster survey. The target sample size was calculated using the WHO guidelines. The number of schools to be sampled from each conglomerate was calculated as described in the Methods section, by dividing the target sample size, adjusted by expected absentee rate, by the average number of students in the combined EU.

Phase 1 – Sampling a pre-determined number of schools per EU conglomerate with a fixed cut-off threshold

As described in the methods, in Phase 1, a pre-determined number of clusters, in this case schools, were randomly selected with replacement from EU conglomerates 1000 times. The number of positive ICT results in each replicate was compared to a pre-determined cut-off threshold. This threshold did not vary based on sample size reached in

each of the replicates. As seen in **Table 3**, with the exception of the conglomerate composed of EU #5 and EU #11, an overwhelming majority of the replicates passed. Specifically, five of the combinations did not have any replicates that would have failed, given the fixed cut-off thresholds. However, a large proportion of replicates in each of the conglomerates did not achieve the target sample size, so using a fixed threshold does not seem appropriate, since the threshold was calculated given the target sample size. In fact, in four of the eight conglomerates, over 96% of replicates were undersampled; and only one combination had over 75% of the replicates reaching the target sample size. The likely reason for this failure to achieve the desired sample size is the low rate of parental consent for testing, as well as lower than expected school attendance rate. Not reaching the target sample size increases the probability of committing a type I error, resulting in falsely passing an EU that should fail.

Phases 2 and 3 – Sampling a pre-determined number of schools per EU conglomerate with variable cut-off threshold; Sampling up to desired target sample size

In order to address the undersampling issue of Phase 1, two approaches were explored, with results presented in **Table 4**.

As seen in the table, in Phase 2, the percentage of replicates that did not reach the target sample sized for each of the combinations remained the same as in Phase 1. However, the passing / failing decisions were quite different, since the cut-off threshold was recalculated for each replicate based on the sample sized reached. As a result, many more replicates failed. The failure rate ranged from 0.6% of replicates in case of the

combination comprised of EU #10 and EU #14 to 67% of replicates bootstrapped from combination of EUs #5 and #11. This high number of failures in the case of the combination of EUs #5 and #11 is somewhat reassuring, since the desired conclusion is to fail the EU; however, while the majority of the time the bootstrap results would agree with this conclusion, 33% of the time the results would not.

When the Phase 3 approach was used, which consisted of sampling additional schools until each replicate reached target sample size, there was quite a difference in the results, as compared to Phase 2. It is good to note that for some EU combinations, such as combination of EUs #11, 5, 4, and 6, more than twice the expected number of schools had to be sampled in order to achieve desired sample size, which is a testament to the gross underestimation of the expected absentee rate. A reassuring result is seen in the case of a combination of EUs that would have passed based on both desired conclusion and the expected conclusion – 99.9% of the bootstrapped replicates would have passed as well. Thus, there is little misclassification in this instance.

Similarly, 76.3% of replicates resulting from bootstrapping combination of EUs #11 and #5 per Phase 3 protocol would have resulted in a failing decision, which is consistent with the desired conclusion. Less than 24% would have falsely passed in this instance.

Combination of EUs #11, #5, #4, and #6 would have passed 98.4% of the time and failed 1.6% of the time; whereas combination of EUs #11, #4, and #6 would have passed 98.4% of the time and failed 11.6% of the time. It should be noted that for both of these combinations, the desired conclusion is a failure, and the expected conclusion is a pass.

For the remaining 4 combinations, the failure rate ranged from 18.9% to 36.5%. In all of these cases, the desired conclusion was failure, and the expected conclusion was a pass. The high rate of misclassification of the decision in the conglomerates, especially if the desired decision is failing, is alarming.

Phase 4 – Expected absentee rate sensitivity analysis

The results from Phase 4, expected absentee rate sensitivity analysis, are presented in **Table 5**, **Table 6**, and **Table 7**.

For Phase 1, the results of which are presented in **Table 5**, higher expected absentee rate corresponds to a higher number of initial schools that are sampled. In the first phase, because more schools are being sampled, as expected, a higher number of replicates are failing. For the last four combinations of EU, however, all of the replicates are still passing, because the sample sizes achieved are still far below the target sample size.

For the second phase, the results presented in **Table 6** show that with increasing the sample size, the passing / failing decisions generally tend to resemble the results derived from the expected decision, as expected, since the more schools selected, the more representative the sample becomes. The only exception is the combination of EUs #12 and #9, where an increase in expected absentee rate corresponds to an increase in number of failing replicates. This can be explained by the characteristic of EUs that make up the combination. EU #12 is almost twice as large as EU #9, with 37 schools as compared to 24 schools from EU #9. As more schools are selected, they are more likely to be selected from EU #12, which is an EU with 15 positive ICT results, as compared to

0 positive results from EU #9. Thus, it's not all that unexpected that the number of failing replicates fluctuates and rises by 4% from expected absentee rate of 15% to expected absentee rate of 40%.

The results for Phase 3, corresponding to varying ranges of expected absentee rates, are presented in **Table 7**. The general change in number of replicates passing or failing with increase in expected absentee rate is not as large as in the other phases, with the largest change corresponding to a 23% increase in number of passing replicates for combination of EUs #12 and #13. All remaining changes in number of passing replicates were less than 10%. This can be explained by the nature of the Phase – the only change that occurs is the initial number of clusters that are selected. With an increase in the initial schools selected, some of the schools that would have reached target sample size with a smaller number of schools are now being oversampled, with the additional schools selected more likely to come from evaluation units that are have larger number of schools. In four of the combinations, the number of replicates whose passing decisions corresponded to the expected conclusion increased. The combination of EUs #12 and #9 saw a slight increase in number of failing replicates, as 61% of schools with data in this combination come from the failing EU #12; thus it's not surprising that the number of positive ICT results grows as more schools are selected. The results for the combination of EUs #11, #4 and #5 barely change, with number of failing replicates increasing from 95 to 100 and then decreasing to 98; it is unlikely that this change is significant. Similarly, the change in results for combination of EUs #11, #5, #4, and #6 is minimal, going up from 16 failing replicates to 19 as expected absentee rate increases from 15% to 60%, just as the number of negative replicates in combination of EUs #11, #4, and #5

changes from 16 to 21. This seems to indicate that the Phase 3 approach is more robust than the other approaches up to this point.

Phase 5 – Upper 1-sided 95% confidence interval for prevalence of positive ICTs in bootstrapped replicates

The upper 1-sided 95% confidence interval (10) was calculated for the number of positive results in the bootstrapped replicates. The replicates were sampled as in Phases 1 and 2; that is, the number of schools selected for each of the conglomerates was equal to the number of schools calculated based on the 15% expected absentee rate.

The results are presented in **Table 8**. As expected based on results from previous phases, the combination of two passing EUs #10 and #14 had the highest number of passing replicates based on the values of upper 1-sided 95% CI. Fifteen out of the 1000 replicates for this combination failed, which is higher than in the previous phases. The highest proportion of failing replicates was once again observed in combination of EUs #11 and #5, with 85.1% of replicates failing and 14.9% passing.

As for the remaining 6 EU combinations, the proportion of replicates for which the upper 1-sided 95% CI contains 2% prevalence of positive ICT results ranges from 20.3% for combination of EUs #11, #4, and #6 to 46.3% for combination of EUs #12 and #9. For all six of these conglomerates, the desired decision is to fail, whereas the decision based on expected prevalence of positive ICTs is to pass. For the four combinations that have percentage of failing replicates ranging from 41.1% to 46.3%, the chances of passing or failing the combination are close to a coin toss, which is rather concerning.

Phase 6 – proportional school sampling

This phase addresses the issue of proportionality. The results comparing the approach in Phase 2 with the current approach are presented in **Table 9**. To recall, Phase 2 randomly sampled with replacement the expected number of schools needed to achieve target sample sized, assuming 15% absentee rate. The cut-off threshold was adjusted based on the actual sample size achieved by each replicate. The Phase 6 results presented in the table use the same decision making process, but weight the number of schools selected from each EU comprising the conglomerate, in order to make the sample more representative.

For 5 of the 8 combinations, the proportional sampling brings a higher number of replicates closer to the expected decision. For the other 4 combinations, the number of passing replicates per combination increases, consistent with the expected decision. Notably, the number of failing replicates for both combination of #11, #5, #4, #6 and combination of #11, #4, and #6 decreases from 57 and 58, respectively, to 0. That is, all of the replicates obtained through bootstrapping after schools were sampled proportionately would have passed.

There are three combinations that actually saw an increase in failing replicates when the expected decision is to pass. The number of failing replicates for combination of EUs #11 and #5 increases when schools were sampled proportionally, which is consistent with the desired conclusion but not the expected conclusion. In combination of EUs #12 and #9, the failing EU #12 has a larger number of schools than the passing EU #9. Because proportional sampling increases the chances that schools selected would come from EU #12, it makes sense that the number of replicates that fail increases.

The combination of EUs #10 and #14 tends toward failing more than in Phase 2 results, which at first appears surprising, given that both of the individual EUs pass. However, EU #10 has 0 positive ICTs, and #14 has 10. Because there are 2.1 times more schools in EU #14 than in EU #10, proportional selection of schools would pick up more of the positive ICTs, which would explain the higher number of failing replicates. It is necessary to note that the number of negative replicates is still very low – only 16 out of 1000 replicates would fail using proportional sampling, as compared to 6 out of 1000 replicates obtained using procedure for Phase 2.

Phase 7 – Proportional sampling with bootstrapping at schoolchildren level in systematic EUs

While the Phase 6 approach seems to be a closer approximation of real conditions, many of the replicates sampled in Phase 6 do not reach the desired sample size. A characteristic of systematic sampling is that only a portion of schoolchildren in target grades attending a selected school is sampled. This, in addition to the fact that systematic EUs tend to have smaller schools, accounts for a consistent undershooting of sample size and contributes to under representation of schoolchildren coming from systematic EUs as compared to schoolchildren from cluster EUs. The results of bootstrapping at schoolchildren level at schools selected in systematic EUs until the average number of schoolchildren in conglomerate is reached in those schools are presented in **Table 9**.

The first thing to note is that results for combinations of EUs #12 and #13, as well as for combination of EUs #11, #14, and #6, did not change at all when bootstrapping was done at schoolchildren level in systematic EUs. This is because the two combinations

do not contain systematic EUs, as all individual EUs are sampled through cluster sampling.

More of the replicates obtained through bootstrapping in systematic EUs are achieving target sample size, since we are essentially forcing the systematic EU samples to meet the target size. However, because the cluster EUs are still being sampled only at school level, there remain replicates that are under target sample size.

For three of the combinations, bootstrapping at schoolchildren level in systematic EUs brings the higher number of replicates to the expected decision.. In the first combination, that of EUs #4 and #5, EU#5 is a systematic cluster with 1 positive ICT result. This result is likely to be magnified if the school with the positive result is among those selected, since each school in this EU will be bootstrapped to obtain the average sample size per school in the conglomerate, 54. Thus, the number of failing replicates for this combination grows from 42 to 97.

In the second and third combinations, those of EUs #11, #5, and #4, and of EUs #11, #5, #4, and #6, respectively, a similar situation takes place. Because the proportion of the schools selected from the systematic EU #5 is smaller, the change in the number of failing replicates is relatively small, as it grows from 42 to 97 in the first combination, and from 0 to 9 out of 1000 replicates in the second.

Phases 8 and 9 – Proportional sampling of schools, reaching target sample size for each replicate

The final step in analysis is to address the issue of small sample size in Phases 6 and 7. Phases 8 and 9 correspond respectively to Phases 6 and 7 by randomly sampling

additional schools when target sample size is not reached. In Phase 9, additional schools are only sampled from cluster EUs, because systematic EUs are forced to reach target sample size through bootstrapping of schoolchildren at school level. The results of Phases 8 and 9 are presented in **Table 10**.

The change from Phase 6 to 8 is generally not very dramatic, with the number of passing replicates changing at most by 18%. The two combinations that always passed remained the same, which is consistent with the expected decision. Three remaining combinations approached the expected decision, whereas one combination remained the same. In case of combination of EUs #12 and #3, a higher number of replicates failed in Phase 8 as compared to Phase 6. Both EUs have a rather high number of positive ICT results, so if the additional schools selected are the ones with positive results, the number of failing replicates would potentially increase. It should be noted that additional number of failing replicates is not very high; the percentage of failing replicates increases from 14.9% to 16.3%, while 83.7% of the replicates would pass. A similar parallel can be drawn for combination of EUs #11 and #5, which saw a small increase in number of failing replicates from 75.1% to 80.9%.

In the results that show change from Phase 7 to Phase 9, the results for the first combination of EUs, EU#12 + EU#3 are identical to results from Phases 6 and 8 because there are no systematic EUs in the combination.

The two overwhelmingly passing combinations, EUs #10 and #14, as well as combination of EUs #11, #4, and #6, remained exactly the same, as the total number of positive ICTs in the two combinations is well under the cut-off threshold. The remaining combinations become proportionately weighted more towards cluster EUs, because we

are now sampling schools from cluster EUs until the target sample size is reached. Because in each of the remaining combinations, at least one cluster EU had failed, we are more likely to pick up positive ICTs, which would explain the general trend of increasing number of failing replicates in Phase 9 as compared to Phase 7. The exception is the combination of EUs #12 and #9, in which the number of failing replicates actually decreased from 295 out of 1000 to 290; however, this change is quite small.

Overall, the results from Phases 8 and 9 are generally consistent with the expected decision. However, for the one combination that has the highest expected prevalence of positive ICT results (1.54%), the bootstrapped results would disagree with the expected decision about 80% of the time.

For the remaining combinations that would be classified as passing using the expected decision, the agreement from bootstrapped results is rather high. For Phase 8, 90-100% of replicates in five of the combinations show agreement, whereas the remaining two combinations agree 65-85% of the time. For Phase 9, three of the combinations show 90-100% agreement, and the remaining four combinations agree 77-86% of the time.

Discussion

Forming conglomerates and making passing / failing decisions

Upon formation of the conglomerates, two separate decisions were formed based on the available data. The desired conclusion is the one that would recognize the conglomerate as failing if at least one of its composite evaluation units fails. The second

type of conclusion, referred to as expected decision, is based on the expected point prevalence of positive ICT results in the combined data. Expected decision is to pass if the point prevalence is smaller than 2%, and to fail if the expected true prevalence is greater than or equals 2%. It should be noted that the desired and the expected decisions are different for 7 combinations and are concordant for 1. In the combination for which both decisions are to pass, both of the evaluation units comprising the combination pass individually. The disagreement in other combinations suggests that even by directly combining available data, misclassification of EUs, defined here as making a decision different from the desired conclusion, would have happened 7 out of 8 times. Seven of the combinations that we would like to classify as failing would have been identified as passing. In effect, the positive results are diluted when the larger combinations are considered.

Expected absentee rate

The issue of underestimating absentee rate, which in this case encompasses both absence of schoolchildren in school and lack of parental consent, is explored in Phase 4, and shows that picking a correct initial expected absentee rate has a non-negligible effect on classification of combinations as either passing or failing. When a number of schools is selected based on this absentee rate, and the cut-off threshold is adjusted based on the achieved sample size in each of the replicates, as is done in Phase 2, the difference can be quite striking. For example, a combination of two failing EUs, EU #11 and #5, would have failed 67.0% of the time if the number of schools that were sampled was determined based on the 15% expected absentee rate. However, when the expected absentee rate is

increased to 60%, 79.9% of the replicates would have failed. Similarly, a combination of two EUs with mixed individual decisions would have failed 22.5% of the time when the number of schools selected was based on expected absentee rate of 15%; the percentage of failing replicates would have decreased to 10.1% when the expected absentee rate increases to 60%, converging towards the expected decision.

The effects of changing the expected absentee rate are less dramatic when, after going to the initial number of schools, additional schools are sampled until the target sample size is reached, per protocol in Phase 3. The biggest deviation from a majority decision among the replicates is demonstrated by a combination of EUs #12 and #13, with increase from 63.5% of replicates passing to 78.1%. In remaining combinations, the change is much smaller. This phase is, by definition, more robust, because increasing the initial number of schools sampled does not change the number of replicates that are undersampled, since the method assures that every single replicate will at least achieve its target sample size. However, higher expected absentee rate, and hence a higher number of schools initially selected, produces oversampling in some replicates, which would explain change in results. A large change seen in the combination of EUs #12 and #13 is, however, alarming, as one would expect that, after reaching the desired sample size, the sample is somewhat representative of the population.

Thus, even if schools will be sampled until the target sample size is achieved, estimating correct absentee rate plays a role on quality of the results.

Bootstrapping

The Phases generally progress towards a scenario that more realistically represents real-life. The first approach, in which a number of schools to be sampled is determined based on expected average number of children per school and an expected absentee rate, with a fixed cut-off threshold based on target sample size, is almost entirely uninformative, since vast majority of the time, the target sample size is not reached. The second and the third phases address the sample size issue by adjusting the cut-off threshold based on achieved sample size in each replicate and by randomly sampling (with replacement) additional schools until the desired sample size is reached. The two methods give rather similar results, as in both approaches, the majority of replicates for each of the EU combinations agree with the expected decision for the corresponding combinations. The change in the percentage of replicates with the same decision as the expected never changes by more than 9%, with all combinations except for two becoming more concordant with the expected decision as the methodology transitions from Phase 2 to Phase 3. The one exception, combination of EUs #12 and #9, changes from passing 72.2% of the time to 71.6% of the time, which is unlikely significant. This result is consistent, because higher sample size in a random sample should produce a more representative sample, which should be consistent with the expected decision. The other combination, EUs #11 and #5, sees an increase in failing replicates, converging towards the desired decision and away from the expected decision.

Proportional sampling of schools based on the total number of schools in each of the EUs making up a combination attempts to closer simulate real-life conditions, as does bootstrapping at schoolchildren level in schools selected from systematic EUs, given the comparable small sample size of schoolchildren sampled in schools in such EUs. The

results in the two general approaches are even closer to the expected decision. When sampled following protocol for Phase 7, in which the combinations were sampled until target samples size is reached in all replicates, proportional sampling resulted in five combinations with less than 10% failing replicates when the expected decision was to pass. One combination would have failed 14.9% of the time when the expected decision was to pass, and another combination would have failed 29.5% of the time when the expected decision was to pass. However, the combination of EUs #11 and #5 would have passed 20.1% of the time while failing the remaining 79.9% of the time based on bootstrapping, even though it would have passed based on the expected decision and failed based on desired conclusion.

The results obtained by Phase 9, which introduces bootstrapping until the average number of children per school is achieved in systematic schools, are overall similar, with three combinations disagreeing with expected decision less than 5% of the time, and four combinations disagreeing with the decision less than 29% of the time. However, the same combination that passed 24.9% of the time in Phase 7 when the desired conclusion was to fail, combination of EUs #11 and #5, passed 17.0% of the time.

Notably, in all of these approaches, the vast majority of the replicates conform with the expected decision. In only one of the combinations does this decision coincide with the desired decision – to fail the combination if at least one of its EU components fails. If the most realistic simulated scenario, Phase 9, is considered, in seven of the combinations, over 77% of the time, the combinations will pass when the desired decision is to fail.

Design effect and 1-sided 95% CI

The mean design effect calculated for non-proportionately sampled combinations ranged from 2.06 to 2.34. It should be noted that, following the WHO TAS standards, two of the combinations were expected to have design effect of 1.5, and the remaining – the design effect of 2. Because design effect serves as a proxy for inner-EU heterogeneity, an elevated design effect would suggest a higher heterogeneity than expected, which is of importance here because lymphatic filariasis is a focal disease. If the real design effect is higher than expected, then using the standard design effects of 1.5 or 2 is not appropriate. In particular, the maximum design effect seen in the replicates of the eight conglomerates ranged from 2.04 in combination of EUs #10 and #4 to design effect of 6.27 in combination of EUs #12, #9, and #13.

The pass/ fail decisions made using the upper 1-sided 95% CI decision that's made for each of the replicates resulting from bootstrapping the combinations are consistent in their majority decision with the overall expected decision, but the proportion of passing replicates in each conglomerate is much closer to chance than in previous phases, with three of the conglomerates having over 40% that disagree with the majority decision, three conglomerates disagreeing between 20 and 33 percent of the time, and one conglomerate disagreeing only 1.5% of the time. The high percentage of disagreement is concerning. If this approach were used to make a decision to pass or fail an EU combination, the decision in at least three of these conglomerates would be uncomfortably close to a coin toss.

Programmatic implications

The high rate of disagreement of results with both the crude decision and the desired decision is concerning. TAS is used as a primary way to assess whether the annual mass drug administration for lymphatic filariasis can be stopped or should be continued. Falsely failing a passing evaluation unit means that mass drug administration will continue when it should not be. This could have high additional costs, as estimates for one year of MDA in Haiti in 2011 were $1,214,102 (5). In addition, it should explored whether continuing MDA over many years may result in volunteer fatigue and decreased compliance, which may have an effect not only on success of LF programs, but of other programs as well. If continued MDA without end in sight has effect on the trust of population towards the implementers, the government, and partners, compliance may be affected (11).

The alternative misclassification, which would result by falsely passing a combination that would have failed, could have drastic consequences. MDA would be stopped, and two years would pass before the second TAS is carried out. If transmission continues unabated, it will not be caught for at least two years, when another TAS should be carried, which will set back control and elimination.

Haiti is unusual in that its evaluation units are small compared to the size required by TAS. Based on the results seen in this study, it would seem that this strategy had averted potential misclassification of at least some of the EUs.

Limitations

The way that EU combinations were formed has some limitations. The conglomerates satisfy the broad WHO standards, as they are well below the 2 million

population threshold, are contiguous, have carried out at least 5 rounds of MDA with >65% coverage, and have similar low levels of suspected prevalence. However, directly combining the data from two or more combinations may not be an appropriate way to estimate a real-life scenario and to draw conclusions about what decision would have been made if the conglomerates were formed in real life. The primary concern with formation of combinations, as well as in carrying out bootstrapping, is in combining systematic with cluster EUs. The ways schoolchildren were tested are not identical. In cluster sampling, a random sample of schools is selected from a list of all schools in the EU. Then, all children for whom parental consent could be obtained were tested. In systematic sampling, a fraction of children were selected from each of the schools in the EU for testing. Thus, the tests results in the two types of EUs may not be representative to the same extent. Simply combining them and calculating an expected decision based on expected true point prevalence may be an oversimplification and may not be appropriate.

References

1. Shenoy RK, Bockarie MJ. Lymphatic filariasis in children: clinical features, infection burdens and future prospects for elimination. *Parasitology* 2011;138(12):1559-68.
2. Brady MA, Hooper PJ, Ottesen EA. Projected benefits from integrating NTD programs in sub-Saharan Africa. *Trends in parasitology* 2006;22(7):285-91.

3. Ottesen EA, Hooper PJ, Bradley M, et al. The Global Programme to Eliminate Lymphatic Filariasis: Health Impact after 8 Years. *PLoS neglected tropical diseases* 2008;2(10).

4. World Health Organization. Lymphatic Filariasis: TAS Manual. 2011.

5. Goldman AS, Brady MA, Direny A, et al. Costs of integrated mass drug administration for neglected tropical diseases in Haiti. *The American journal of tropical medicine and hygiene* 2011;85(5):826-33.

6. Drexler N, Washington CH, Lovegrove M, et al. Secondary mapping of lymphatic filariasis in Haiti-definition of transmission foci in low-prevalence settings. *PLoS neglected tropical diseases* 2012;6(10):e1807.

7. Oscar R, Lemoine JF, Direny AN, et al. Haiti National Program for the elimination of lymphatic filariasis--a model of success in the face of adversity. *PLoS neglected tropical diseases* 2014;8(7):e2915.

8. Rao RU, Nagodavithana KC, Samarasekera SD, et al. A comprehensive assessment of lymphatic filariasis inSri Lanka six years after cessation of mass drug administration. *PLoS neglected tropical diseases* 2014;8(11):e3281.

9. Chu BK, Deming M, Biritwum NK, et al. Transmission assessment surveys (TAS) to define endpoints for lymphatic filariasis mass drug administration: a multicenter evaluation. *PLoS neglected tropical diseases* 2013;7(12):e2584.

10. Hamlin KL, Moss DM, Priest JW, et al. Longitudinal monitoring of the development of antifilarial antibodies and acquisition of Wuchereria bancrofti in a highly endemic area of Haiti. *PLoS neglected tropical diseases* 2012;6(12):e1941.

11. Krentel A, Fischer PU, Weil GJ. A Review of Factors That Influence Individual Compliance with Mass Drug Administration for Elimination of Lymphatic Filariasis. *PLoS neglected tropical diseases* 2013;7(11).

Tables

EU#	Target population	# of schools in EU	Average # of students in target grades	Expected absentee rate	# Schools tested	Type of survey	# Positive ICTs	Total Sample Size in data	Cut-off threshold	Pass/ fail decision
1	14,813	367	40	10%	36	Cluster	0	1494	16	Pass
2	35,357	721	49	10%	46	Cluster	3	1659	18	Pass
3	2,442	67	36	10%	53	Cluster	2	1231	14	Pass
4	6,821	120	57	10%	45	Cluster	0	1528	18	Pass
5	707	17	42	10%	16	Systematic	1	364	3	Pass
6	18,977	333	57	10%	42	Cluster	2	1617	18	Pass
7	1,597	25	64	15%	25	Systematic	0	551	6	Pass
8	20,833	441	47	15%	47	Cluster	2	1587	18	Pass
9	754	26	29	15%	24	Systematic	0	587	6	Pass
10	1,875	36	52	15%	30	Systematic	0	672	7	Pass
11	1,336	42	32	15%	31	Cluster	19	858	9	Fail
12	1,634	48	34	15%	37	Cluster	15	1037	11	Fail
13	9,299	199	47	15%	32	Cluster	19	1984	20	Pass
14	4,038	74	55	15%	33	Cluster	10	1414	16	Pass

Table 1. Characteristics of individual evaluation units (EUs) and TAS results. Target population is the expected number of school children enrolled in 1st and 2nd grades of primary schools. Number of schools in EU denotes the number of schools that exist in the evaluation unit. Number of schools tested is the number of schools that were selected in TAS, and for whom there is at least one ICT results present in the data. Total sample size in data is the number of positive and negative ICT results that were recorded in the EU. If the number of positive ICT results in the EU is greater than the cut-off threshold, the EU is said to fail; else, the EU passes.

EUs	Positive ICTs	Cut-off threshold	Decision	# Schools tested	Survey type	Target sample size	# Schools to be sampled	Desired programmatic conclusion	Expected true prevalence	Expected conclusion
12	15	11	Fail	37	Cluster	1540	41	Fail	1.03%	Pass
13	19	18	Pass	32						
12	15	11	Fail	37	Cluster	909	33	Fail	0.99%	Pass
9	0	4	Pass	24						
12	15	11	Fail	37	Cluster	1540	43	Fail	0.96%	Pass
13	19	18	Pass	32						
9	0	4	Pass	24						
11	19	9	Fail	31	Cluster	909	31	Fail	1.54%	Pass
5	1	4	Pass	16						
11	19	9	Fail	31	Cluster	1532	36	Fail	0.36%	Pass
4	0	18	Pass	45						
5	1	4	Pass	16						
11	19	9	Fail	31	Cluster	1556	34	Fail	0.20%	Pass
4	0	18	Pass	45						
5	1	4	Pass	16						
6	2	18	Pass	42						
10	0	7	Pass	30	Cluster	1392	31	Pass	0.48%	Pass
14	10	16	Pass	33						
11	19	9	Fail	31	Cluster	1556	34	Fail	0.20%	Pass
4	0	18	Pass	45						
6	2	18	Pass	42						

Table 2. Characteristics of conglomerates formed from adjoining evaluation units. Positive ICTs, Cut-off threshold, Decision, and # schools tested all refer to individual characteristics of EUs that form the conglomerates. Target sample size is the number of ICT results that should be obtained in bootstrapping to achieve desired power and alpha levels. Number of schools tested is the expected number of clusters to be selected in the conglomerate in order to achieve desired sample size. Desired programmatic conclusion is to fail if at least one of the individual EUs is said to fail; if all individual EUs comprising the conglomerate pass, the desired conclusion is to pass. The expected true prevalence is the expected percent of school children aged 6-7 years in the conglomerate that would have tested positive if all of them were tested. The expected decision is to fail the conglomerate if the expected true prevalence is greater than or equals 2%, and to pass otherwise.

EU combination	Target sample size	Cut-off threshold	Total Positive ICT results	Desired programmatic conclusion	Expected conclusion	# Replicates failing / passing in bootstrapped data	% of Replicates in bootstrapped data not reaching target sample size
12+13	1540	18	34	Fail	Pass	7 / 993	24.4%
12+9	909	11	15	Fail	Pass	57 / 943	61.2%
12+13+9	1540	18	34	Fail	Pass	0 / 1000	40.3%
11+5	909	11	20	Fail	Pass	585 / 415	73.3%
11+4+5	1532	18	20	Fail	Pass	0 / 1000	97.4%
11+5+4+6	1556	18	22	Fail	Pass	0 / 1000	97.9%
10+14	1392	16	10	Pass	Pass	0 / 1000	96.2%
11+4+6	1556	18	21	Fail	Pass	0 / 1000	96.5%

Table 3. Bootstrapping results for Phase 1 - sampling a pre-determined number of schools per EU conglomerate, with fixed cut-off thresholds. EU combination lists the individual EUs that were combined to form a conglomerate. Target sample size is the number of ICT results that should be obtained in bootstrapping to achieve desired power and alpha levels. Cut-off threshold is the number of positive ICT results that can be obtained in a sample of target sample size before the combination is said to fail. Total positive ICT results refer to the number of positive ICT results obtained by combining the positive ICT results in all individual EUs making up the conglomerate. Desired programmatic conclusion is to fail if at least one of the individual EUs is said to fail; if all individual EUs comprising the conglomerate pass, the desired conclusion is to pass. The expected conclusion is to fail the conglomerate if the expected true prevalence is greater than or equals to 2%, and to pass otherwise. Number of replicates failing / passing in bootstrapped data is the number of replicates in which the number of positive ICT results exceeds the cut-off threshold, out of 1000 replicates obtained through bootstrapping.

EU combination	Target sample size	Cut-off threshold	Total Positive ICT results	Desired programmatic conclusion	Expected conclusion	Phase 2		Phase 3	
						# Replicates Failing / Passing in bootstrapped data with variable threshold	% of bootstrapped replicates not reaching target sample size	# Replicates Failing / Passing in bootstrapped data with variable threshold	Min / Max # of schools sampled
12+13	1540	18	34	Fail	Pass	372 / 628	24.4%	365 / 635	41 / 61
12+9	909	11	15	Fail	Pass	278 / 722	61.2%	284 / 716	33 / 54
12+13+9	1540	18	34	Fail	Pass	225 / 775	40.3%	189 / 811	43 / 68
11+5	909	11	20	Fail	Pass	670 / 330	73.3%	763 / 237	31 / 63
11+4+5	1532	18	20	Fail	Pass	182 / 818	97.4%	95 / 905	36 / 78
11+5+4+6	1556	18	22	Fail	Pass	57 / 943	97.9%	16 / 984	34 / 73
10+14	1392	16	10	Pass	Pass	6 / 994	96.2%	1 / 999	31 / 64
11+4+6	1556	18	21	Fail	Pass	58 / 942	96.5%	16 / 984	34 / 67

Table 4. Bootstrapping results for Phases 2 and 3 - sampling a pre-determined number of schools per EU conglomerate with variable cut-off threshold, and sampling up to desired target sample size. Target sample size is the number of ICT results that should be obtained in bootstrapping to achieve desired power and alpha levels. Cut-off threshold is the number of positive ICT results that can be obtained in a sample of target sample size before the combination is said to fail. Total positive ICT results refer to the number of positive ICT results obtained by combining the positive ICT results in all individual EUs making up the conglomerate. Desired programmatic conclusion is to fail if at least one of the individual EUs is said to fail; if all individual EUs comprising the conglomerate pass, the desired conclusion is to pass. The expected conclusion is to fail the conglomerate if the expected true prevalence is greater than or equals 2%, and to pass otherwise. Phase 2 refers to sampling a set number of schools in each EU, and making a pass / fail decision by comparing the number of positive ICT results in each replicate with a cut-off threshold that is calculated for each replicate based on the sample size achieved. In phase 3, additional schools are sampled until the target sample size is achieved; the pass / fail decision is calculated as for Phase 2. The min / max # of schools sampled corresponds to the minimum and the maximum number of schools, respectively, that had to be sampled in Phase 3 in order to achieve the target sample size.

EU combination	Desired programmatic conclusion	Expected true prevalence	Expected conclusion	Phase 1 : # of Bootstrap Replicates Failing / Passing			
				Original	Abs=30%	Abs=40%	Abs=60%
12+13	Fail	1.03%	Pass	7 / 993	17 / 983	36 / 964	66 / 934
12+9	Fail	0.99%	Pass	57 / 943	69 / 931	92 / 908	134 / 866
12+13+9	Fail	0.96%	Pass	0 / 1000	0 / 1000	0 / 1000	0 / 1000
11+5	Fail	1.54%	Pass	585 / 415	672 / 328	756 / 244	858 / 142
11+4+5	Fail	0.36%	Pass	0 / 1000	0 / 1000	0 / 1000	0 / 1000
11+5+4+6	Fail	0.20%	Pass	0 / 1000	0 / 1000	0 / 1000	0 / 1000
10+14	Pass	0.48%	Pass	0 / 1000	0 / 1000	0 / 1000	0 / 1000
11+4+6	Fail	0.20%	Pass	0 / 1000	0 / 1000	0 / 1000	0 / 1000

Table 5. Results for sensitivity analysis following Phase 1 protocol based on expected absentee rates (Abs=15%, 30%, 40%, 60%). Desired programmatic conclusion is to fail if at least one of the individual EUs is said to fail; if all individual EUs comprising the conglomerate pass, the desired conclusion is to pass. The expected true prevalence is the expected percent of school children aged 6-7 years in the conglomerate that would have tested positive if all of them were tested. The expected conclusion is to fail the conglomerate if the expected true prevalence is greater than or equals 2%, and to pass otherwise. Original, Abs=30%, Abs=40%, and Abs=60% refer to the different levels of expected absentee rate that were used in calculating the number of schools necessary to be sampled in order to achieve the target sample size.

EU combination	Desired programmatic conclusion	Expected true prevalence	Expected conclusion	Phase 2 : # of Bootstrap Replicates Failing / Passing			
				Original	Abs=30%	Abs=40%	Abs=60%
12+13	Fail	1.03%	Pass	372 / 628	333 / 667	288 / 712	232 / 768
12+9	Fail	0.99%	Pass	278 / 722	301 / 699	290 / 710	316 / 684
12+13+9	Fail	0.96%	Pass	225 / 775	174 / 826	150 / 850	101 / 899
11+5	Fail	1.54%	Pass	670 / 330	736 / 264	771 / 229	799 / 201
11+4+5	Fail	0.36%	Pass	182 / 818	184 / 816	167 / 833	148 / 852
11+5+4+6	Fail	0.20%	Pass	57 / 943	62 / 938	55 / 945	36 / 964
10+14	Pass	0.48%	Pass	6 / 994	5 / 995	8 / 992	3 / 997
11+4+6	Fail	0.20%	Pass	58 / 942	39 / 961	43 / 957	31 / 969

Table 6. Results for sensitivity analysis following Phase 2 protocol based on expected absentee rates (Abs=15%, 30%, 40%, 60%). Desired programmatic conclusion is to fail if at least one of the individual EUs is said to fail; if all individual EUs comprising the conglomerate pass, the desired conclusion is to pass. The expected true prevalence is the expected percent of school children aged 6-7 years in the conglomerate that would have tested positive if all of them were tested. The expected conclusion is to fail the conglomerate if the expected true prevalence is greater than or equals 2%, and to pass otherwise. Original, Abs=30%, Abs=40%, and Abs=60% refer to the different levels of expected absentee rate that were used in calculating the number of schools necessary to be sampled in order to achieve the target sample size.

EU combination	Desired conclusion	Expected true prevalence	Expected conclusion	Phase 3 : # of Bootstrap Replicates Failing / Passing			
				Original	Abs=30%	Abs=40%	Abs=60%
12+13	Fail	1.03%	Pass	365 / 635	319 / 681	287 / 713	219 / 781
12+9	Fail	0.99%	Pass	284 / 716	284 / 716	319 / 681	312 / 688
12+13+9	Fail	0.96%	Pass	189 / 811	191 / 809	157 / 843	113 / 887
11+5	Fail	1.54%	Pass	763 / 237	775 / 225	801 / 199	794 / 206
11+4+5	Fail	0.36%	Pass	95 / 905	100 / 900	97 / 903	98 / 902
11+5+4+6	Fail	0.20%	Pass	16 / 984	21 / 979	14 / 986	19 / 981
10+14	Pass	0.48%	Pass	1 / 999	0 / 1000	4 / 996	0 / 1000
11+4+6	Fail	0.20%	Pass	16 / 984	25 / 975	15 / 985	21 / 979

Table 7. Results for sensitivity analysis following Phase 3 protocol based on expected absentee rates (Abs=15%, 30%, 40%, 60%). Phase 3 samples additional schools at random until target sample size is achieved in each replicate. Desired programmatic conclusion is to fail if at least one of the individual EUs is said to fail; if all individual EUs comprising the conglomerate pass, the desired conclusion is to pass. The expected true prevalence is the expected percent of school children aged 6-7 years in the conglomerate that would have tested positive if all of them were tested. The expected conclusion is to fail the conglomerate if the expected true prevalence is greater than or equals 2%, and to pass otherwise. Original, Abs=30%, Abs=40%, and Abs=60% refer to the different levels of expected absentee rate that were used in calculating the number of schools necessary to be sampled in order to achieve the target sample size.

EU/ combo	% Prevalence in data	Mean % Prevalence	Upper 1-sided CI*			Design effect			Fail / Pass	# Replicates with no ICT+
			mean	min	max	mean	min	max		
12+13	1.12	1.10	1.91	0.39	3.86	2.06	0.00	5.03	425 / 575	1
12+9	0.92	0.87	2.20	0.51	4.81	2.31	0.00	6.14	463 / 537	19
12+13+9	0.94	0.91	1.75	0.29	3.83	2.19	0.56	6.27	321 / 679	0
11+5	1.64	1.70	3.41	0.84	7.36	2.19	0.00	4.04	851 / 149	2
11+4+5	0.73	0.73	1.83	0.39	4.58	2.16	0.00	4.17	411 / 589	22
11+5+4+6	0.50	0.52	1.46	0.32	4.53	2.03	0.00	4.30	214 / 786	41
10+14	0.48	0.48	1.13	0.36	2.59	1.10	0.00	2.04	15 / 985	13
11+4+6	0.52	0.51	1.43	0.32	4.06	2.07	0.00	4.22	203 / 797	32

Table 8. Upper 1-sided 95% confidence interval for prevalence of positive ICT results in bootstrapped replicates. Percentage (%) prevalence in data is the percent of total positive ICT results in all the ICT results obtained in the conglomerate. Mean % prevalence is the average prevalence of ICT results in 1000 replicates obtained through bootstrapping in each EU. Mean, min, and max of upper 1-sided CI refer to mean, minimum and maximum 95% 1-sided confidence limits for the 1000 replicates in each conglomerate, respectively. Note that in calculation of mean, minimum, and maximum upper 1-sided CI, replicates that had no positive ICT results were excluded. These values are denoted with asterisk (*). The number of replicates excluded in calculations for each of the combinations is recorded as “Number of replicates with no ICT+.”

EU combination	Desired programmatic conclusion	Expected true prevalence	Expected conclusion	EU	# schools selected proportionately	Phase 2 - # of replicates failing / passing	Phase 6 - # of replicates failing / passing	Phase 6 - # of replicates smaller than target sample size	Phase 7 - # of replicates failing / passing	Phase 7 - # of replicates smaller than target sample size
12+13	Fail	1.03%	Pass	12 13	8 33	372 / 628	149 / 851	44	149 / 851	44
12+9	Fail	0.99%	Pass	12 9	22 12	278 / 722	345 / 655	511	295 / 705	233
12+13+9	Fail	0.96%	Pass	12 13 9	8 31 5	225 / 775	109 / 891	26	71 / 929	5
11+5	Fail	1.54%	Pass	11 5	22 9	670 / 330	751 / 249	708	799 / 201	516
11+4+5	Fail	0.36%	Pass	11 4 5	9 25 4	182 / 818	42 / 958	944	97 / 903	884
11+5+4+6	Fail	0.20%	Pass	11 5 4 6	3 2 8 22	57 / 943	0 / 1000	941	9 / 991	912
10+14	Pass	0.48%	Pass	10 14	10 21	6 / 994	16 / 984	911	1 / 999	435
11+4+6	Pass	0.20%	Pass	11 4 6	3 9 23	58 / 942	0 / 1000	956	0 / 1000	956

Table 9. Bootstrapping results for Phases 2, 6, and 7. Phase 2 samples schools randomly from the list of all schools for which data is available in the conglomerate. Phase 6 selects schools from individual EUs maintaining the proportionality of schools in individual EUs to the number of schools in the conglomerates. Phase 7 introduces bootstrapping at schoolchild level from selected schools in systematic EUs while maintaining proportionate selection. Desired programmatic conclusion is to fail if at least one of the individual EUs is said to fail; if all individual EUs comprising the conglomerate pass, the desired conclusion is to pass. The expected true prevalence is the expected percent of school children aged 6-7 years in the conglomerate that would have tested positive if all of them were tested. The expected conclusion is to fail the conglomerate if the expected true prevalence is greater than or equals 2%, and to pass otherwise. Number of schools selected proportionately refers to the number of schools selected from each individual EU to be representative of proportion of total number of schools in the EU to the number of schools in the conglomerate.

EU combination	Desired programmatic conclusion	Expected true prevalence	Expected conclusion	EU	# of schools selected from EU to achieve proportionality	Phase 6 - # of replicates passing / failing	Phase 8 - # of replicates passing / failing	Phase 7 - # of replicates passing / failing	Phase 9 - # # of replicates passing / failing
12+13	Fail	1.03%	Pass	12 13	8 33	149 / 851	163 / 837	149 / 851	163 / 837
12+9	Fail	0.99%	Pass	12 9	22 12	345 / 655	345 / 655	295 / 705	290 / 710
12+13+9	Fail	0.96%	Pass	12 13 9	8 31 5	109 / 891	89 / 911	71 / 929	222 / 778
11+5	Fail	1.54%	Pass	11 5	22 9	751 / 249	809 / 191	799 / 201	830 / 170
11+4+5	Fail	0.36%	Pass	11 4 5	9 25 4	42 / 958	12 / 988	97 / 903	149 / 851
11+5+4+6	Fail	0.20%	Pass	11 5 4 6	3 2 8 22	0 / 1000	0 / 1000	9 / 991	42 / 958
10+14	Pass	0.48%	Pass	10 14	10 21	16 / 984	0 / 1000	1 / 999	1 / 999
11+4+6	Pass	0.20%	Pass	11 4 6	3 9 23	0 / 1000	0 / 1000	0 / 1000	0 / 1000

Table 10. Bootstrapping results for Phases 6, 7, 8 and 9. Phase 6 replicates are obtained by proportional sampling. Phase 8 is obtained by same method as Phase 6, but additional schools are randomly sampled until the target sample size is reached. Phase 7 is proportional sampling, with bootstrapping of schoolchildren at school level in systematic EUs. Phase 9 is identical to Phase 7, but additional schools are selected in cluster EUs until desired sample size is reached. Desired programmatic conclusion is to fail if at least one of the individual EUs is said to fail; if all individual EUs comprising the conglomerate pass, the desired conclusion is to pass. The expected true prevalence is the expected percent of school children aged 6-7 years in the conglomerate that would have tested positive if all of them were tested. The expected conclusion is to fail the conglomerate if the expected true prevalence is greater than or equals 2%, and to pass otherwise. Number of schools selected proportionately refers to the number of schools selected from each individual EU to be representative of proportion of total number of schools in the EU to the number of schools in the conglomerate.

Figures

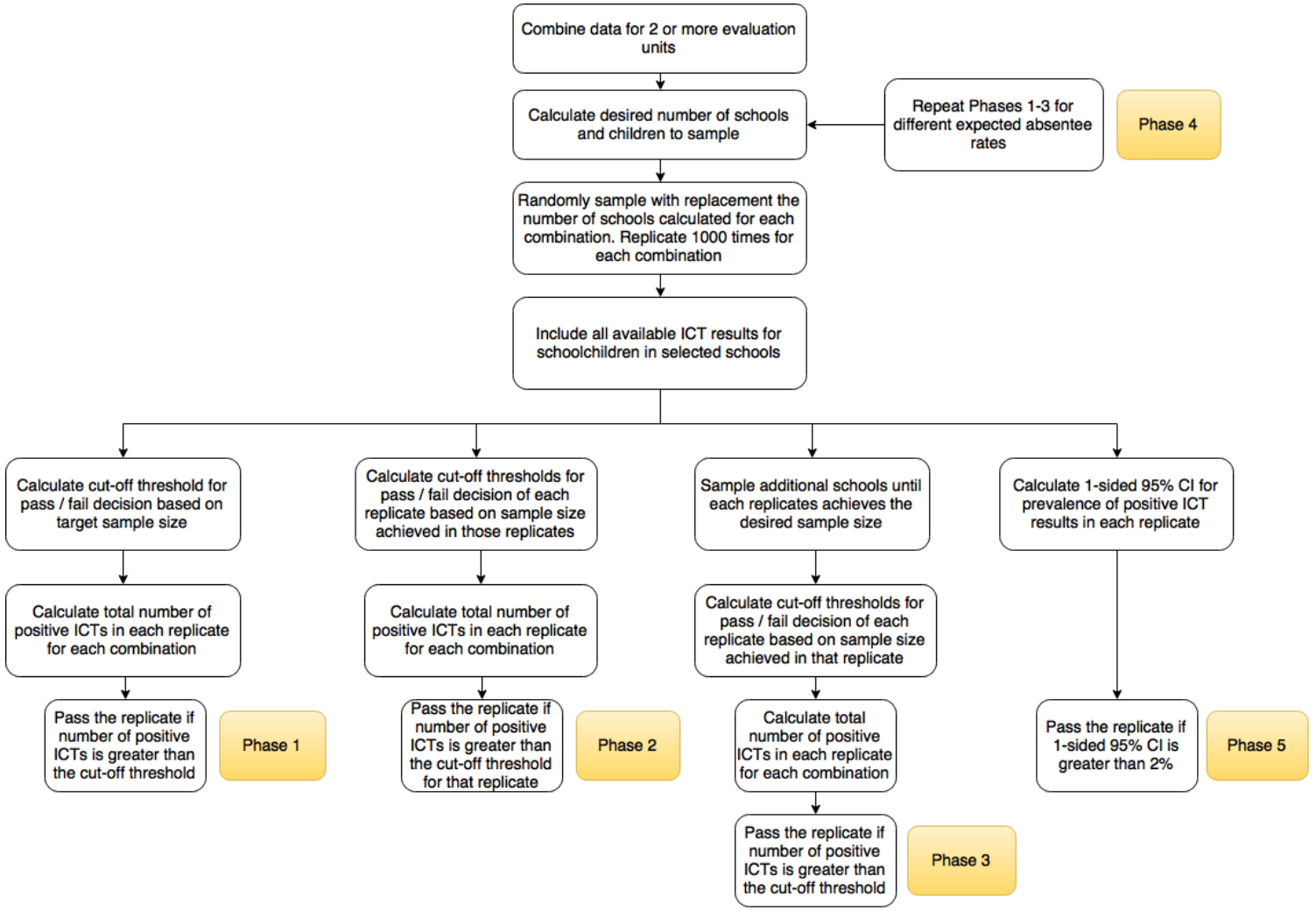


Figure 1. Flowchart of Phases 1 through 5 of data simulation

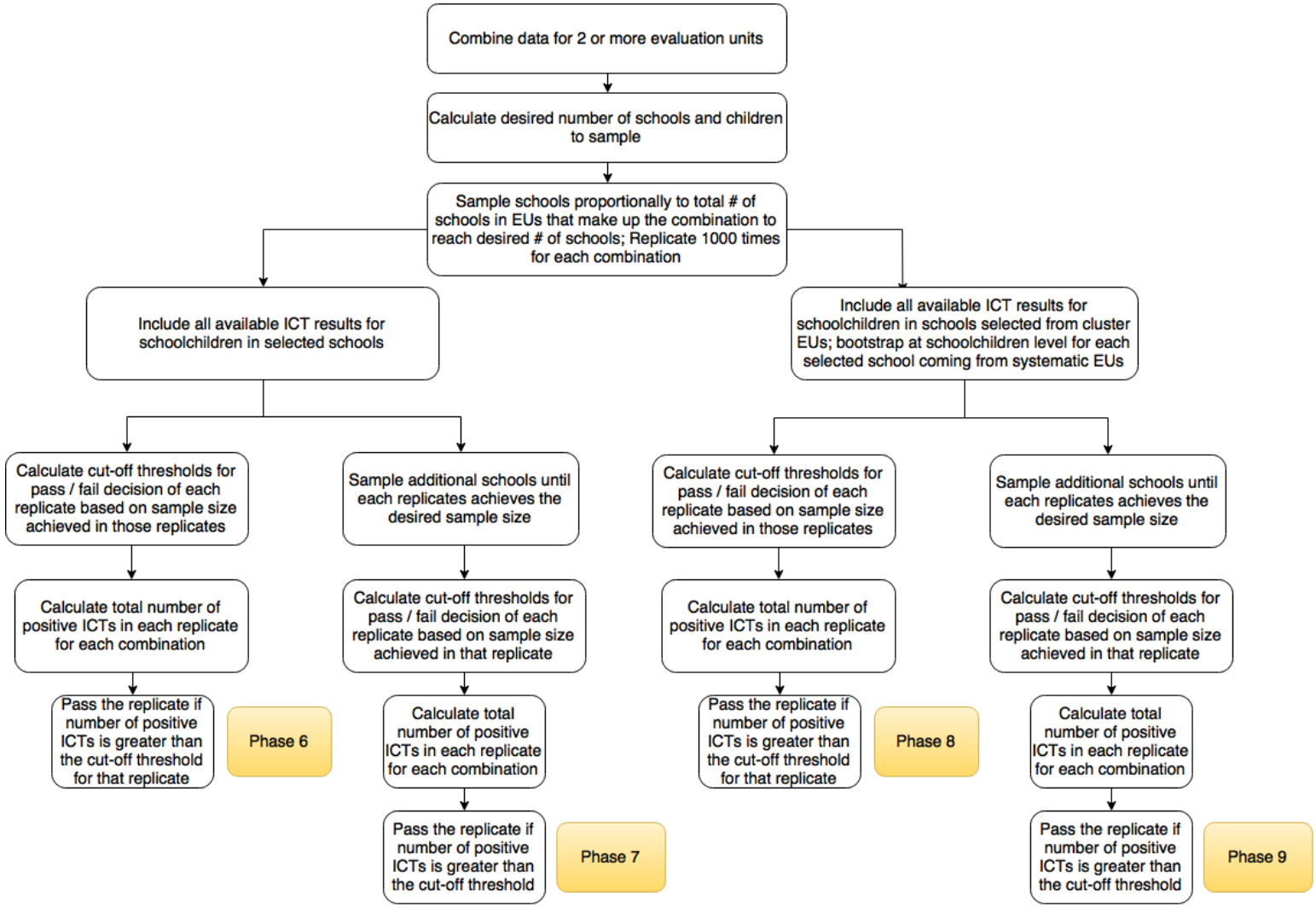


Figure 2. Flowchart of Phases 6 - 9 of data simulation

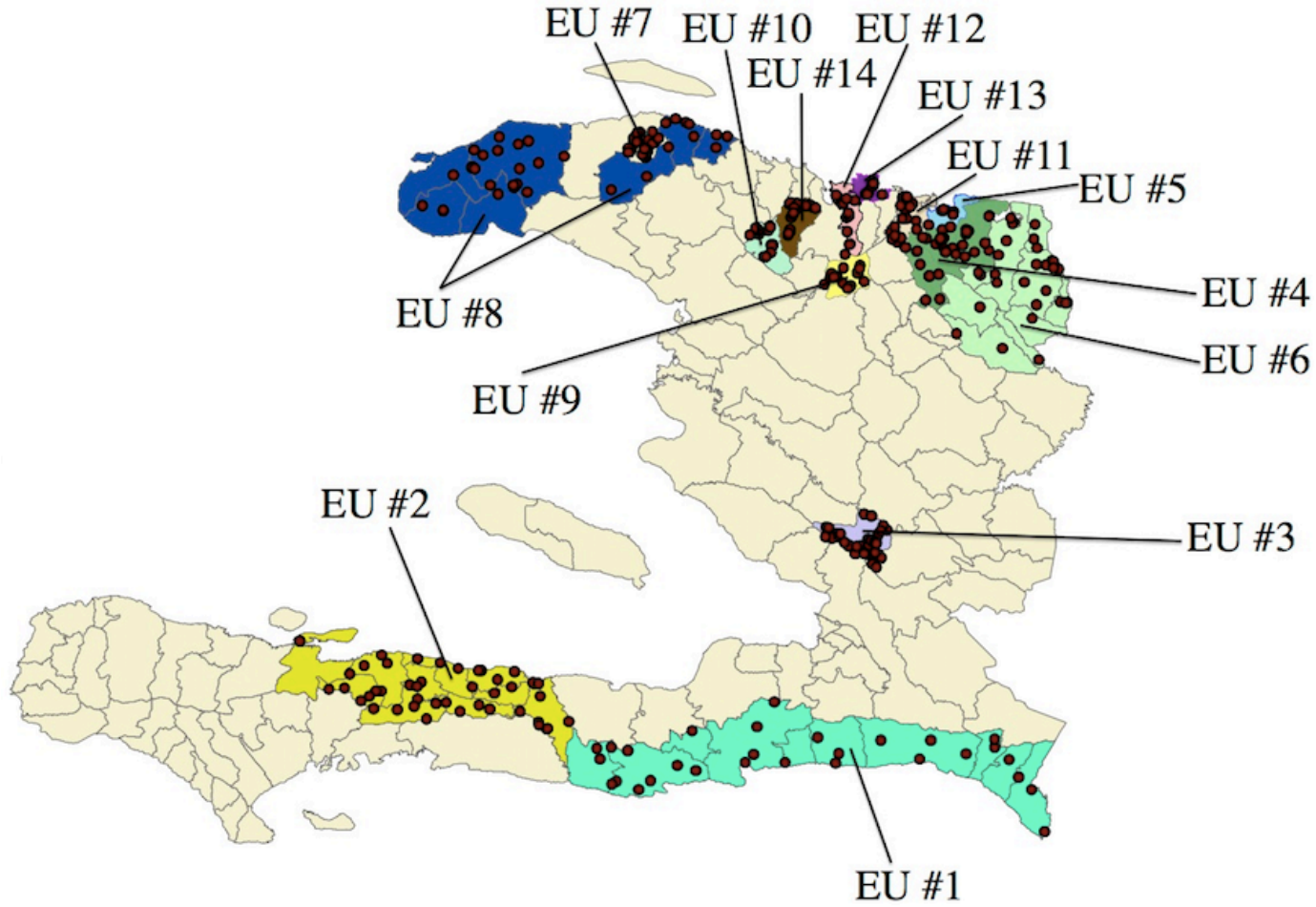


Figure 3. Sites of transmission assessment surveys and evaluation units. Red circles represent schools where schoolchildren in grades 1 and 2 were tested.

Chapter III: Summary, Public Health Implications, and Future Directions

Summary

As more and more countries are entering at least sixth year of annual mass drug administration (MDA) for lymphatic filariasis, there is a need to revisit guidelines to determining whether transmission has been reduced to below sustainable level, allowing for mass drug administration to be stopped. One of the guidelines that calls for attention is the population size in the units that are to be evaluated. The current guidelines place the upper limit on population size in the unit to be evaluated at 2 million (1). This study investigates the effect of target population size in the evaluation units on the classification of units as either ones where mass drug administration can be stopped or ones where MDA should continue.

The dataset used is the results of a transmission assessment survey (TAS) conducted in Haiti in 2014-2015 in 14 evaluation units. The Haiti evaluation units are of comparably small sizes; larger units were simulated by forming eight combinations of adjacent units. Several approaches to simulate real life TAS were carried out, with an intent to see whether classification of units as either eligible for stopping MDA or not would change when the larger units were considered, as compared to the decisions for the individual units that comprise them.

The results of simulations show that in some combinations, such as when both individual evaluation units making up the combination were eligible for stopping MDA, the vast majority of the time the larger unit would also successfully “pass” the TAS criterion, qualifying for stopping MDA. However, there were some combinations that

raised a cause for alarm. Specifically, a non-negligible proportion of the time, combinations with at least one failing evaluation unit would be classified as passing the criterion.

Misclassifying a passing evaluation unit as failing would mean continued MDA in a region where transmission is lower than the threshold suspected for sustainable transmission. This adds unnecessary costs of additional MDA and testing. A possibility of misclassifying an evaluation unit as passing when transmission persists in at least one part of it is even more troubling, given the programmatic implications of stopping MDA too early. The guidelines should be carefully reconsidered to ensure that TAS is a valid and reliable tool.

Public Health Implications

The implications of misclassifying an EU are extensive. The possibility of falsely failing an EU when it should be passed will result in programmatic costs due to continuation of MDA and further testing. MDA costs vary widely, so it is difficult to estimate the exact cost. However, a study by Goldman et al. in Haiti in 2011, the estimated total MDA program cost was $1,850,153 for 55 communes, which excludes cost of albendazole, donated from GlaxoSmithKline. Program cost per person treated was $0.42, while the total economic cost per person treated, which includes cost of donated albendazole, were $0.64. The average commune population size in the ten communes considered in the study was 87,187 (2). Assuming the EU that was misclassified was similar to a combination of EUs considered in our study, we can assume the population in the misclassified combination would be about twice the mean communal population, or 174,374 people. Assuming the programmatic cost of MDA in this unit is the same as the

programmatic cost of MDA in the Goldman study, and assuming 90% coverage, the cost of one round of MDA for this combination would be around $65,913. Based on WHO TAS guidelines, after TAS fails, MDA will continue while sentinel and spot-check site data is collected every two years until the criteria necessary for stopping MDA is met, at which time another TAS should be conducted. In best-case scenario, the first sentinel and spot-check site data will show that criteria needed to consider the EU for TAS is met. Because TAS can be conducted no sooner than 6 months after MDA took place, at least 2 rounds of MDA will have taken place before another TAS can be carried out. Costs of these 2 rounds of MDA can be expected to be around $131,827. A study by Chu et al. estimated that the mean cost of TAS conducted between 2009 and 2011 in an EU was $25,500 (3). The average EU size in the study was 248,121 people, which is 1.42 times the size of the combination EU that we have in consideration. Scaling the TAS cost proportionately results in a cost of $17,921 for TAS conducted in the combination EU. If at this point, TAS correctly identifies the combination as passing, MDA will be stopped. The cost of the two rounds of MDA and an additional TAS that it took to correctly classify the combination is $149,748, which does not include the additional cost of sentinel and spot-check surveillance. While this figure is estimated using gross assumptions, both about population size and scaling of costs per person treated, it should give an idea of a plausible cost of misclassification of an evaluation unit as failing when it does, indeed, meet the passing criteria. It should be noted that the size of combination considered here is well below the size of evaluation units in the Chu et al. study, so this figure is not implausible.

The costs of misclassifying a failing unit as passing are more difficult to quantify because the main effect of premature stopping of MDA is on disease transmission. If MDA is stopped before reaching the hypothesized threshold below which transmission cannot be sustained is reached, transmission will continue in the area for at least two years before another TAS can be implemented. The increase of microfilarea levels in the affected population will depend on the initial prevalence, but it can be expected that additional rounds of MDA will be required in order to lower the prevalence below the sustainable threshold. In addition to the costs of additional MDA, which have been discussed above, people whose infection could have been averted if MDA were continued as expected must be taken into account. While the burden of disease cannot be adequately captured through economic indicators, there is an economic cost, such as the disability-adjusted life years, to the disease. Thus, the cost of stopping MDA early is non-negligible.

While reducing the size of the population in the evaluation unit can reduce the likelihood of misclassification, it does come with a cost. Decreasing the size of an EU corresponds to an increase in number of EUs, which is likely to increase programmatic costs, including transportation and per diems, two categories that comprise majority of all programmatic costs (4). However, given the costs incurred in the misclassification scenarios, it is likely that choosing smaller EUs will be cost-effective in long term.

It is difficult to know which implementation units (IUs), or regions where MDA was carried out, can be combined for evaluation. In its TAS manual, WHO advises that IUs can be combined if they have had at least five rounds of MDA and share “similar epidemiological features” (1). The manual suggests that the epidemiological features of

interest may include rates of MDA coverage, as well as similar prevalence of microfilaraemia as estimated through sentinel and spot-check sites. Currently, the manual recommends that there be at least one sentinel site per 1 million population, with at least one corresponding spot-check site (1). As seen in Appendix 3, which is an ArcGIS-generated map of distribution of positive ICT results in some of the evaluation units of Haiti TAS, positive ICT results appear to have hotspots. Because of the focality of LF and possible hotspots of cases, it may be of value to increase the number of sentinel and spot-check sites prior to selection of EUs. It may be of value to limit the level of heterogeneity in EUs, so if a potential hotspot has been identified through spot-checks or sentinel testing, the implementation unit containing the hotspot should not be combined with other IUs.

Another possibility is to use mini-TAS to identify potential hotspots and similarly restrict IUs that are combined to form larger evaluation units. However, more research needs to be done in order to develop more specific guidelines.

Possible Future Directions

The antibody data for the TAS dataset used in this study has recently become available, and could provide an additional layer of analysis. The number of positive antibody responses is expected to be higher than the number of positive ICT results, so this data could be quite valuable.

Another possible direction to consider is analysis of mini-TAS, or a Transmission Assessment Survey that would be conducted in smaller sized EUs, but would test fewer schoolchildren. The decreased sample size would effective reduce power of the test. If the mini-TAS shows promise, there could be potential cost-saving implications to the

TAS procedure. Alternatively, the mini-TAS could be considered as an additional tool that would allow for identification of hotspots of positive results or would inform subsequent TAS.

Furthermore, spatial analysis could be carried out. One of the questions that this analysis could answer is whether there is clustering of positive ICT results at school or village levels. In addition, the role of environmental factors that may be involved in the lifecycle of vectors and their habitats could be evaluated. Some of the environmental factors to consider are population density, soil moisture level, rainfall, temperature, altitude, and presence of improved water and sanitation infrastructure.

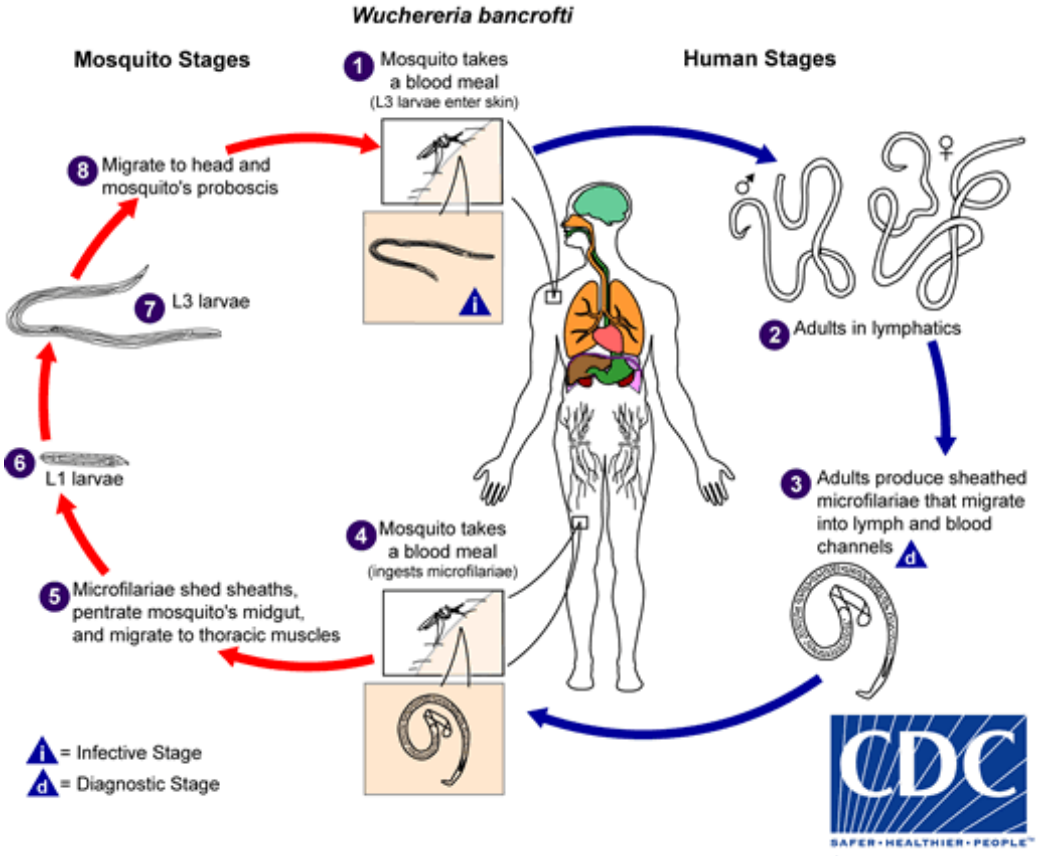
Finally, data from other countries could be used to repeat and expand on the analysis carried out in this study. Patterns of lymphatic filariasis transmission are expected to vary widely from country to country because of its focal nature and dependence on mosquito vectors. Thus, the generalizability of the study's conclusions could be explored through analysis of other datasets.

References

1. World Health Organization. Lymphatic Filariasis: TAS Manual. 2011.
2. Goldman AS, Brady MA, Direny A, et al. Costs of integrated mass drug administration for neglected tropical diseases in Haiti. *The American journal of tropical medicine and hygiene* 2011;85(5):826-33.
3. Chu BK, Hooper PJ, Bradley MH, et al. The economic benefits resulting from the first 8 years of the Global Programme to Eliminate Lymphatic Filariasis (2000-2007). *PLoS neglected tropical diseases* 2010;4(6):e708.
4. Chu BK, Deming M, Biritwum NK, et al. Transmission assessment surveys (TAS) to define endpoints for lymphatic filariasis mass drug administration: a multicenter evaluation. *PLoS neglected tropical diseases* 2013;7(12):e2584.

Appendices

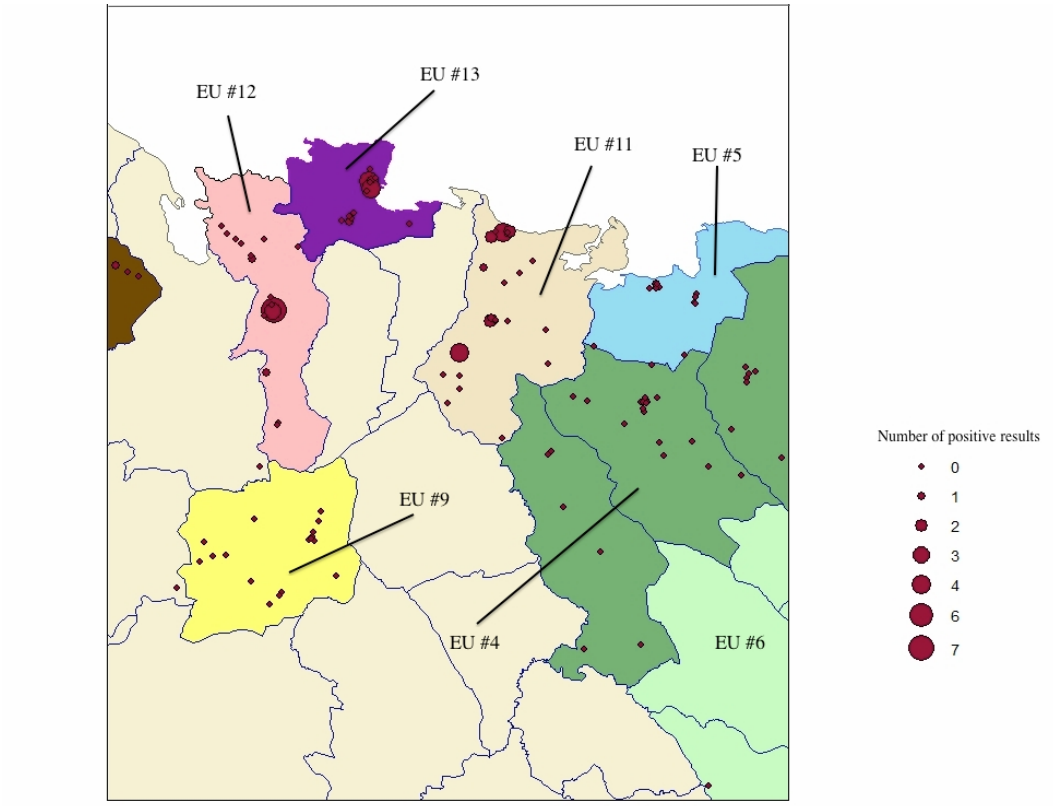
Appendix 1: Life cycle of W. bancrofti lymphatic filariasis. Source: Centers for Disease Control and Prevention. Biology - Life Cycle of Wuchereria bancrofti. 2013. (1)



Appendix 2: Transmission Assessment Survey cut-off criteria for given parasites. Based on World Health Organization. Lymphatic Filariasis: TAS Manual. 2011. (2)

Parasite	Mosquito vector	Target threshold prevalence	Prevalence measured
W. bancrofti	*Anopheles*; *Culex*	<2%	Antigenaemia
W. bancrofti	*Aedes*	<1%	Antigenaemia
Brugia spp.	*Mansonia*, *Anopheles*, *Aedes*	<2%	Antibody

Appendix 3: Spatial distribution of positive ICT results in Haiti TAS data.



References

1. Centers for Disease Control and Prevention. Biology - Life Cycle of Wuchereria bancrofti. 2013.

2. World Health Organization. Lymphatic Filariasis: TAS Manual. 2011.