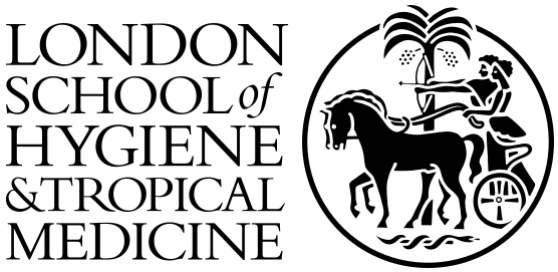


Exploring the spatial heterogeneity of trachomatous trichiasis

REBECCA MANN FLUECKIGER

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy

University of London

APRIL 2018

Department of Disease Control

Faculty of Infectious and Tropical Diseases

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by Helen Keller International, Sightsavers, and The World Health Organization

Declaration by candidate

I, Rebecca Mann Flueckiger, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed  Date 4 April 2018

Rebecca Mann Flueckiger

Abstract

Prolonged conjunctival infection with *Chlamydia trachomatis* leads to an inflammatory response, trachomatous inflammation follicular (TF). Over time, repeat infection can progress to scarring of the conjunctiva causing the eyelid to turn inward, resulting in lashes rubbing against the cornea. This painful stage of the disease is called trachomatous trichiasis (TT). TT can damage the cornea, leading to vision impairment or blindness.

Trachoma is targeted for elimination as a public health problem by the year 2020, which for TT is defined as less than 1 TT-positive person, who is not already known to the health system, per 1,000 population. For trachoma to meet the elimination targets, massive resources are required for both mapping and intervention. A particularly large knowledge gap exists around identifying areas where TT is likely to be found. To better align resources and plan for elimination, the trachoma community needs to understand how much TT currently exists and requires management, how to accurately measure TT prevalence, and where TT cases are mostly likely to be located. Understanding these elements will help position trachoma control programs to meet the TT elimination targets.

In my thesis I first calculate an updated global estimate of TT cases and describe the methods involved. Second, I provide a survey design for measuring TT with adequate precision for control activities, along with validation exercise results and a brief time-cost analysis. I then examine the spatial structure of TF and TT and identify areas of spatial autocorrelation. Finally, I identify environmental factors associated with higher than expected TT prevalence to identify TT hot spots.

The outcomes of these activities provide an updated global estimate of existing TT cases, a validated tool for measuring TT prevalence at the implementation unit (district) level, and insight on where to begin case finding activities in the context of the “end game”. These outputs are critical to the continued effort of trachoma elimination as a public health problem, specifically providing targeted direction for TT resources.

Acknowledgements

Most importantly, I offer my heartfelt thanks to my supervisors, Dr. Rachel Pullan and Professor David Mabey, for generously sharing their time and invaluable expertise in support of this thesis. I would also like to thank Professor Simon Brooker for his guidance and encouragement. I am ever grateful to the members of my advisory panel; Dr. Anthony Solomon, Dr. Katie Gass, and Dr. Charles Opondo. In particular, my sincerest thanks go to Dr. Solomon, for his patient mentorship and thoughtful comments and suggestions throughout this process.

In addition, I would like to thank Helen Keller International, Kilimanjaro Centre for Community Ophthalmology, Sightsavers, and the World Health Organization for funding many aspects of this work. I am especially grateful to Professor Paul Courtright for his technical guidance in developing the trachomatous trichiasis survey design.

I am indebted to the trachoma community, made up of colleagues around the world, who constantly inspire me as they fight towards ending this terrible disease. These brave and resilient men and women have dedicated their lives to improving the lives of others – and for this I will be forever grateful.

Finally, I could never adequately articulate my overwhelming gratitude to my husband, Peter, who has patiently supported me throughout this journey, helped me think through methods and approaches, and whose steady gentle presence provides unmeasurable comfort every single day.

Table of Contents

List of figures

List of tables

List of acronyms

AUC	*Area under the curve*
ADL	*Activities of daily living*
BLTR	*Bilamellar tarsal rotation*
CI	*Confidence interval (95%)*
CGIAR-CSI	*Consultative Group on International Agricultural Research Consortium for spatial information*
CO	*Corneal opacity*
CRS	*Cluster random sample*
DALY	*Disability-adjusted life years*
DE	*Design effect*
DFID	*United Kingdom's department for international development*
DRC	*Democratic Republic of the Congo*
EU	*Evaluation unit*
FPC	*Finite population correction*
GET2020	*Global elimination of trachoma by 2020*
GIS	*Geographic information system*
GLW	*Gridded livestock of the world*
GTMP	*Global trachoma mapping project*
GTSM	*Global trachoma scientific meeting*
HALY	*Handicap adjusted life years*
ICTC	*International coalition for trachoma control*
IDT	*Initial doubling time*
ITI	*International trachoma initiative*
ITM	*Integrated threshold mapping*
JMP	*Joint monitoring programme for water supply and sanitation*
KNN	*K-nearest neighbour*
LF	*Lymphatic filariasis*
LGA	*Local government area*
LISA	*Local indicators of spatial association*
LSHTM	*London school of hygiene and tropical medicine*
MBG	*Model based geostatistics*
MDA	*Mass drug administration*
MDG	*Millennium development goal*
MoH	*Ministry of health*
NGO	*Non-government organization*
NL	*Night-light*
NPV	*Negative predictive value*
NTD	*Neglected tropical disease*
OLS	*Operational Linescan system*
OSCE	*Objective structured clinical exam*
OR	*Odds ratio*
PBPS	*Population based prevalence survey*
PCR	*Polymerase chain reaction*
PET	*Potential evapo-transpiration*
PLTR	*Posterior lamellar tarsal rotation*
PPS	*Probability proportional to size*

PPV	*Positive predictive value*
RAAB	*Rapid assessment of avoidable blindness*
RMSE	*Root mean squared error*
ROC curve	*Receiver operating characteristic curve*
SAFE	*Surgery, antibiotic, facial cleanliness, environmental improvement*
SD	*Standard deviation*
SDG	*Sustainable development goal*
STAG	*Scientific technical advisory group*
TB	*Tuberculosis*
STH	*Soil transmitted helminths*
TF	*Trachomatous inflammation follicular*
TI	*Trachomatous inflammation intense*
TRA	*Trachoma rapid assessment*
TS	*Trachomatous scarring*
TT	*Trachomatous trichiasis*
UNdata	*United Nations population division*
USAID	*United States agency for international development*
VIF	*Variance inflation factor*
WASH	*Water, sanitation and hygiene*
WHO	*World health organization*

Chapter 1: **Introduction**

1.1 Background

Trachoma, a neglected tropical disease (NTD), is endemic in more than 50 countries [1] and affects the most impoverished people of the world. Improved living standards are credited for the disappearance of trachoma from Europe and North America, but in many less developed countries, trachoma is still a public health problem, and contributes to the continued suffering and deepening of poverty of millions of people. The World Health Assembly Resolution 51.11 of 1998 targets the elimination of trachoma as a public health problem [2].

For trachoma to meet the elimination targets, considerable resources are required for both mapping and intervention. An important knowledge gap exists around identifying areas where trachomatous trichiasis (TT) is likely to be found. This hinders the ability of trachoma programmes to efficiently reach TT patients with services. Geospatial techniques can provide valuable insight on the nature and distribution of trachoma, which in-turn can provide targeted direction for resources.

The utility of geospatial tools and techniques to identify disease clusters, patterns, and trends have been incorporated into public health operational research for decades. More recently, the NTD community has adopted the use of these methods [3-17]. Gains in efficiency in identifying areas of endemicity through geospatial approaches can lead to more targeted programmatic work.

The aim of this thesis is to inform estimates of the number of cases of TT (TT backlog) at global and national scales, assess a TT specific survey design, describe the

geospatial relationship between infectious trachoma (trachomatous inflammation follicular as a proxy) and trachoma morbidity (trachomatous trichiasis) and inform the possibility of identifying TT hot spots. To do so, I exploit a unique epidemiological resource generated through the Global Trachoma Mapping Project (GTMP), the largest ever exercise of its kind, which commenced in 2012. The aim of this project was to complete the global baseline mapping of trachoma using systematic and standardised methods. Through GTMP, 2.6 million people were examined representing 1,542 districts in 29 countries [18]. This geolocated data has been made available for operational research and provides an evidence base to explore the geospatial components of trachoma in a detail previously not possible.

This chapter provides background information on clinical presentation of trachoma as well as detection, diagnosis, and epidemiology. Next, mapping and intervention strategies are reviewed and lastly, spatial modelling and geographic information systems (GIS) approaches and their applications to TT are outlined.

1.2 Introduction to trachoma

The bacterial infection *Chlamydia trachomatis* is the causative organism of trachoma. Repeated ocular chlamydial infection results in chronic inflammation, characterised by sub-epithelial follicles in the tarsal conjunctiva, which may be sufficiently large and numerous to meet the definition of trachomatous inflammation follicular (TF), a key sign for assessing trachoma prevalence [19]. Over time, with repeated reinfection, scarring may develop; this scarring can eventually cause the eyelid to turn inwards in some people, resulting in eyelashes touching the globe. This stage is TT and is very

painful [20]. As an individual with trichiasis blinks, the eyelashes abrade the cornea, which can lead to corneal opacity and blindness [21].

1.2.1 Detection and diagnosis

NTD control programmes currently diagnose trachoma based on clinical presentation. The World Health Organization (WHO) developed a simplified grading system in an effort to standardise trachoma diagnosis in a public health setting [19]. The eye is inspected using a binocular magnifying loupe (x2.5). The examiner first inspects the lashes and cornea, looking for at least one lash rubbing the eyeball (TT) and easily visible corneal opacity over the pupil (CO). The upper lid is then everted, exposing the tarsal conjunctiva. The examiner looks for five or more follicles at least 0.5 mm in diameter (TF), pronounced inflammatory thickening that obscures more than half of the normal deep tarsal vessels (TI) and presence of scarring (TS) [19] (Figure 1.1).

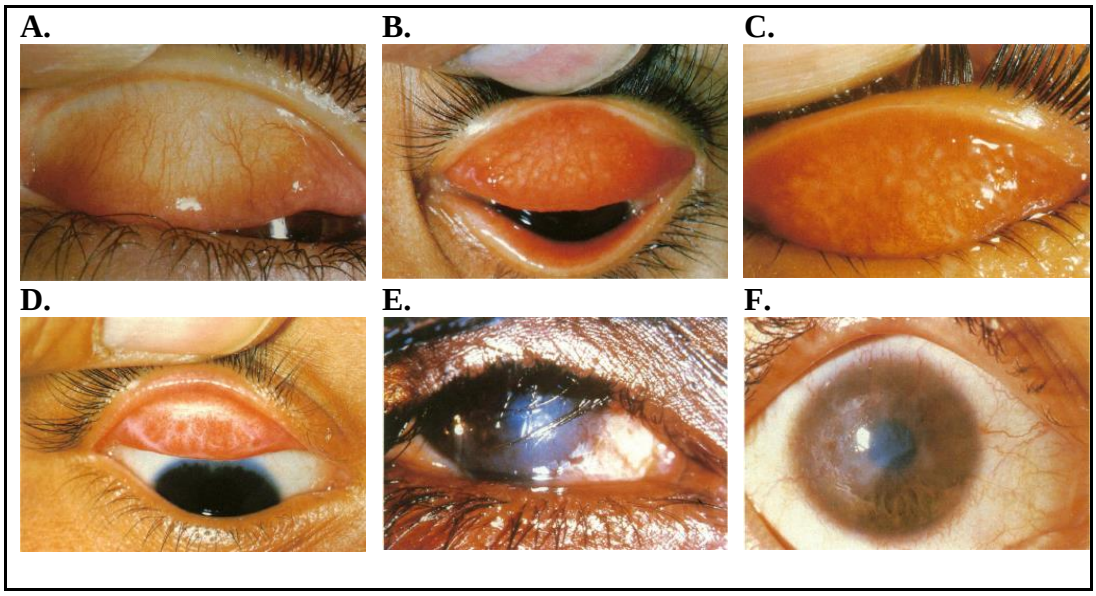


Figure 1.1 The WHO simplified grading system for assessment of trachoma, Adapted from Thylefors et al. (1987) [3]. A) Normal tarsal conjunctive, B) Trachomatous inflammation follicular (TF), C) Trachomatous inflammation follicular and intense (TF+TI), D) Trachomatous scarring (TS), E) Trachomatous trichiasis (TT), F) Corneal opacity (CO)

The simplified grading system identifies clinical signs that are not necessarily pathognomonic for trachoma. However, they are reasonably predictive of the community situation. In areas where trachoma is endemic, these signs are nearly always attributable to trachoma [22].

Despite the existence of a standardised grading system, diagnosis based entirely on clinical presentation may not always support elimination efforts. When examining eyes in a post-intervention setting, it is possible that residual clinical signs may bias survey results. Studies in The Gambia have shown, after multiple years of intervention, positive clinical signs remain high (TF 12.3%), whilst ocular *Chlamydia trachomatis* infection is low (0.3%) [23]. However, studies in Nepal and Tanzania have consistently shown good concordance between clinical and serological diagnosis in low endemicity settings [24-26].

Whilst the simplified grading system aims to provide a consistent diagnostic, it is still subjective to the examiners. Great effort has gone into training examiners to consistently identify each sign. Examiners participate in practical exercises and are graded on their consistent diagnosis. This method works well for identifying TF, which is relatively common in training settings. However, because TT is a much less common sign it is often difficult to find enough positive cases in a training setting to validate a consistent diagnosis.

It is not always possible to make presumptions as to the aetiology of trichiasis cases. Serological testing, similar to those used in the lymphatic filariasis [27] and onchocerciasis [28] programmes would be useful. Although current work is underway to develop an objective diagnostic tool for trachoma, we do not yet have a rapid test

available for routine programmatic use. Such a tool could be extremely useful in the "end game" of trachoma elimination. Generally serological surveys have been limited to research environments. For example, a 2014 study in Tanzania examined the use of serological tools for measuring trachoma in a post-intervention setting. In the community, 218 children were examined using the simplified grading system, additionally an ocular swab and blood sample were taken from participants. The TF prevalence was 6.5%, yet there were zero positive ocular swabs and a 3.5% prevalence from the serological assay (with specificity limits of 96-98%) [29]. This small study raises concern that residual clinical signs remain in the community after transmission is halted. However, several additional studies have collected serologic data along with clinical signs and show concordance between the methods. A 2014 cross-sectional survey of Achham district in Nepal found TF prevalence within 24 randomly selected communities to be 0.3% (CI 0.1%-0.8%) and prevalence of ocular chlamydia infection to be 0% (CI 0%-0.3%) [24]. A 2015-2016 cross-sectional survey of four districts in Nepal found TF prevalence <1% in all districts. These same districts showed seropositivity varying from 1.4% (CI 0.7%-2.6%) to 2.5% (CI 1.3%-4.5%) [26]. A 2017 cross-sectional study, of 38 communities of Kongwa district in Tanzania found TF prevalence to be 3.2% and ocular chlamydia infection to be 6.5% [25]. These studies show that there is often concordance, though it is important to acknowledge the occasional discordance between clinical identification of trachoma and serological tests.

1.2.2 Epidemiology

Chlamydia trachomatis is transmitted through close physical contact. It is reasonable to expect that young children infect each other through play and caretakers of these

children (often women) also become infected. Individuals with severe inflammatory disease are more likely to develop fibrotic sequelae and this is more common after repeat infection [30, 31]. Over time, chronic TF can lead to scarring which may results in the inverting of the eyelid and eyelashes scrapping the globe (TT) (Figure 1.2). I suspect that TF distribution dynamics will almost certainly influence TT distribution.

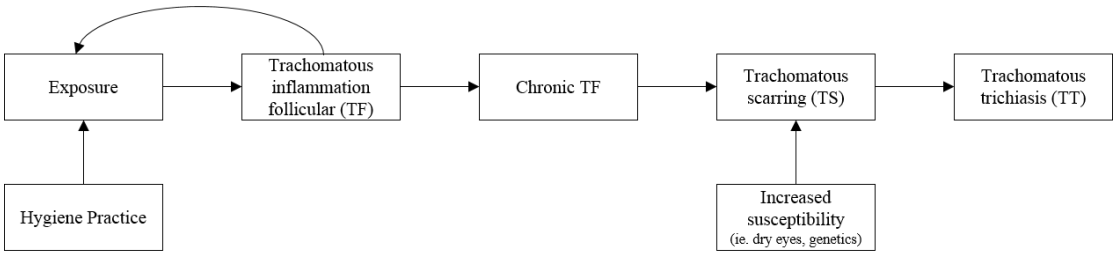


Figure 1.2 Description of pathway to trachoma infection and progress towards TT

Many studies have identified the importance of TF transmission between family members and close communities [32-42]. These studies consistently found a statistically significant relationship between positive TF and sharing a dwelling or living in close proximity. Bailey *et al.* even found that in one Gambian village there was clustering *within* a household [42]. Those who shared a bedroom were more genetically similar and/or share similar exposure risk. It is important to note that several studies have shown that immune response to *C. trachomatis* is connected to host genetic variation and genetic polymorphisms related to immune system function have been associated with TS [43, 44]. The results of these studies provide evidence that trachoma is related to close personal contact and that increased exposure increases risk of TF and/or a strong genetic predisposition to development of clinical signs exists.

There have also been several studies evaluating behaviour and environmental contribution to trachoma transmission. Numerous studies have demonstrated a

relationship between TF status and poor personal hygiene. A risk factor of note is facial cleanliness [32, 41, 45-48]. Nasal and ocular discharge [49] contain chlamydial bacteria. Washing the discharge away may reduce transmission.

Improved access to water is necessary to encourage personal hygiene practices. A relationship has also been found between distance to water and risk for TF [32, 50-52]. These studies conclude that distance to water is associated with the use of the water. The authors infer that increase risk of disease is not simply a direct function of access to water but of value placed on the water. It is reasonable to suggest that when water is more readily available it is more likely to be used for bathing (i.e. clean faces).

Furthermore, eye seeking flies have been shown as possible trachoma vectors. A series of investigations were conducted in The Gambia to identify potential fly vectors. Two species were found to be present on children's eyes. 90% of the flies found on eyes were *Musca sorbens*. Children with ocular and nasal discharge had twice as many flies than those with no discharge [53]. Additional studies have found a significant association between flies on faces and TF [38, 51, 53-56]. Facial cleanliness could reduce the transmission resulting from the presence of flies.

A study in Tanzania also identified altitude as significantly inversely associated with TF prevalence (OR 0.55). The authors note, households of higher socioeconomic status tend to reside higher up the slopes of Mt. Kilimanjaro. The association with altitude may also be a result of fly density decreasing as altitude increases. This theory has been demonstrated in previous studies conducted in central Ethiopia [57, 58]. However, increased altitude as a risk factor is inconsistent with finding in the Amhara region of Ethiopia where ordinal logistic regression showed a positive associated with

TF prevalence and altitude P-value <0.001 [59]. Additionally, there is evidence that perhaps fly density is not associated with increased TF in hyperendemic communities. A 2006 single-blind randomized trial in Tanzania where one dose of antibiotic was administered to residents in both intervention (community treated with insecticide and monitored for fly count) and control groups found no difference in trachoma rates [60]. Perhaps there are inherent differences in the socio-economic and geographic settings of these areas. There is likely a cause/effect relationship between the environmental factors and hygiene behaviour, ultimately resulting in increased exposure.

1.2.2.1 Understanding progression from TF to TT

Crucially, very few studies have been conducted to evaluate risk factors specifically for TT and our understanding of this outcome is limited. Overall, more research is needed to understand the relationship between community-level TF prevalence, geoclimatic factors and TT occurrence (Table 1.1). One study in the Amhara region of Ethiopia used multivariate logistic regression to explore the relationship between risk factors and TT [59]. The results show an association between TT and increased age (OR per 5-year increase 1.5), female sex (OR 4.5), increased proportion of children with TF (P-value 0.003) and increased altitude (P-value 0.015). These findings align with common belief that age and sex are associated with progression to TT. Altitude deserves further investigation as a risk factor. It is may also be reasonable to infer that increased occurrence of TF results in increased occurrence of TT. However, additional evidence is needed to generalize this relationship.

A 1993 study in Tanzania evaluated risk factors associated with TT in women [61]. This study matched women with TT to two women with no clinical signs for trachoma

of the same age and from the same village. Logistic regression models showed the following factors associated with TT: history of trichiasis in the women's mother (OR 3.6, CI 2.0-6.5), sleeping in a room with a cooking fire during childbearing years (OR 1.8, CI 1.2-2.8), a home of wood and earth during childbearing years (OR 2.1, CI 1.3-3.3), no adult education classes (OR 2.2, CI 1.4-3.4); and five or more deaths among her children (OR 2.6, CI 1.3-5.1). These factors all point to a strong socio-economic element at play in developing TT.

The frequency of *C. trachomatis* infection required to develop trichiasis is unknown. It is also unclear if other factors play a role in the speed and intensity of trichiasis development. Long-term longitudinal studies have found TS to be strongly associated with TF [62-65]. However, no association was found when *C. trachomatis* infection was compared with progression of previously established TS [63]. Perhaps, progression from scarring to TT may not be entirely dependent on re-infection, which would mean that there may be progression to TT even once infection is eliminated as a public health problem.

It is argued that conditions promoting eye to eye transmission exacerbate the progression from TF to TT [33]. In The Gambia a cohort of trichiasis patients who declined surgery were followed over four years [66]. It was observed that trichiasis continued to worsen over time, even with low TF prevalence in the community. The study found increased age, conjunctival inflammation (OR 2.04) and bacterial infection (OR 4.33) were associated with progressive trichiasis. The authors suggest chronic inflammation of the conjunctiva may be the central event in the pathogenesis of trachoma. This finding is consistent with results from a study in Tanzania, where a cohort of children were followed for seven years to determine incidence of scarring

[62]. The incidence in children with constant severe inflammatory trachoma (TI, or trachoma intense) was 29.2% (OR 4.85) [63].

Limited studies have demonstrated a potential geoclimatic element at play with the development of TT. A study in Mali found high rates of TF prevalence in the northern portion of the country (41.1%) [67]. However, prevalence rates of TT increased from north to south (1.0% north of the 15th parallel and 2.8% south of the 15th parallel: OR 2.91). The authors speculate this may be attributed to dry conditions in the north contributing to TF and the humid environment in the south contributing to blinding complications. It has also been suggested that areas in Sudan with frequent sandstorms result in eye trauma [68]. Irritaion of the eyes leads to rubbing with fingers. This perhaps hastens the progression of TT.

Table 1.1 Studies with focus on risk factors specific for TT

Year	Location	Study population	Key findings	Limitations	Citation
2008	Ethiopia	children aged 1–9 years and adults aged ≥15-years	Amhara region of Ethiopia used multivariate logistic regression to explore the relationship between risk factors and TT. The results show an association between TT and increased age (OR per 5-year increase 1.5), female sex (OR 4.5), increased proportion of children with TF (P-value 0.003) and increased altitude (P-value 0.015).	The sampling implemented for data collection used in this analysis was limited to 160 clusters and so may not appropriately represent the entire region of Amhara Ethiopia described in the findings.	Ngondi J, Gebre T, Shargie EB, Graves PM, Ejigsemahu Y, Teferi T, et al. Risk factors for active trachoma in children and trichiasis in adults: a household survey in Amhara Regional State, Ethiopia. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2008;102(5):432-8. doi: 10.1016/j.trstmh.2008.02.014. PubMed PMID: 18394663.
2006	The Gambia	Individuals with trachomatous trichiasis in at least one eye	In The Gambia a cohort of trichiasis patients who declined surgery were followed over 4 years. It was observed that trichiasis continued to worsen over time, even with low TF prevalence in the community. The study found increased age, conjunctival inflammation (OR 2.04) and bacterial infection (OR 4.33) were associated with progressive trichiasis.	In this study, 30% of target population was lost to follow-up, with 19% having died and so the findings may be an underestimate.	Burton MJ, Bowman RJ, Faal H, Aryee EA, Ikumapayi UN, Alexander ND, et al. The long-term natural history of trachomatous trichiasis in the Gambia. Investigative ophthalmology & visual science. 2006;47(3):847-52. doi: 10.1167/iovs.05-0714. PubMed PMID: 16505016.
2007	Mali	children aged 1–9 years and women aged ≥15-years	A study in Mali found high rates of TF prevalence in the northern portion of the country (41.1%). However, prevalence rates of TT increased from north to south (1.0% north of the 15th parallel and 2.8% south of the 15th parallel: OR 2.91).	This study included data from only 202 clusters across the entire country limiting the generalizability of the findings.	Schemann JF, Laffly D, Sacko D, Zephak G, Malvy D. Trichiasis and geoclimatic factors in Mali. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2007;101(10):996-1003. doi: 10.1016/j.trstmh.2007.05.015. PubMed PMID: 17658570.
1959-1969	Sudan	All reported trachoma cases	This study suggests that areas in Sudan with frequent sandstorms result in eye trauma. Irritation of the eyes leads to rubbing with fingers. This perhaps hastens the progression of TT.	The data used in this study was derived from ophthalmological records and so populations who do not access services are surely underrepresented in the findings.	Salim AR, Sheikh HA. Trachoma in the Sudan. An epidemiological study. The British journal of ophthalmology. 1975;59(10):600-4. PubMed PMID: 1191619; PubMed Central PMCID: PMC1017417.

In summary, most epidemiological research to date evaluates factors associated with TF. Factors shown to have an association are exposure, unclean faces, fly density (debatably), and access to water. These factors are interrelated. There may be a relationship between TT progression and individual bacterial load and geoclimatic factors. However, our understanding of the factors driving progression to TT in different settings remains too limited.

1.3 Mapping and control of trachoma

There has been substantial progress in trachoma mapping over the past five years. During this time, a standardised approach was used to complete the global baseline mapping of trachoma, with a strong focus on generating local prevalence estimates for TF. Less attention has been paid to mapping TT specifically. In this next section, I lay out the key strategies that have been used to map trachoma outcomes.

1.3.1 Mapping strategies

1.3.1.1 Rapid assessment method

The trachoma rapid assessment (TRA) is used to identify communities most in need of intervention, through an initial desk review and evaluation of high risk communities. In this assessment community members are asked a series of questions to help determine who in the community is most likely to suffer from TT. An eye exam is performed on these individuals to diagnose TT. To measure TF, the selection of households is “optimally biased” towards the least advantaged. A minimum of 50 children aged 1–9 years are examined from the selected households [69]. This method is a biased epidemiologic survey but is a useful cost-effective way of determining if a population based prevalence survey (PBPS) is warranted.

1.3.1.2 Current mapping gold standard

The WHO endorses a PBPS for trachoma, which is designed to calculate the prevalence of TF in children aged 1–9 years with an expected prevalence of 10%. This method assumes that children aged 1-9 years are the reservoirs of TF and the results can be generalized to the entire population. The design is a two stage cluster random sample survey, which uses probability proportional to size (PPS) to select 20-30 clusters and examines all children aged 1–9 years for TF and adults aged ≥15-years for TT in the selected clusters [22]. TF and TT are evaluated using the simplified grading system. This design is powered to provide TF prevalence estimates for the entire population. However, the design is not sufficient for estimating TT prevalence with high precision for the entire population. A survey powered to provide TT prevalence estimates is needed.

In 2013 the United Kingdom's department for international development (DFID) provided funding for GTMP which has served as a catalyst to systematically complete the global trachoma map using standardised techniques in both collecting and analysing survey data [70]. GTMP measured trachoma using the gold standard method described here. To prevent convenience sampling, the GTMP requires a fixed number of households to be enrolled per cluster. Adults living in the selected households are examined for trichiasis. These surveys accept the loss in precision in estimating TT [71]. The GTMP method was used in 29 countries where 2.6 million people were examined providing results for over 1,540 districts [72]. The methodology for analysing the datasets involves standardising the calculations by age and sex to account for the non-random difference in age and sex which systematically bias the results. Anecdotally it has been seen that during household surveys it is more likely that older

individuals and women are at home and therefore potentially over sampled. It is additionally assumed that persons with TT are more likely to be home as they may be unable to work. The findings of the massive mapping campaign show 36% of the districts mapped require antibiotic intervention (described in the following section) and nearly 80% of those districts are in Eastern Africa (Figure 1.3).

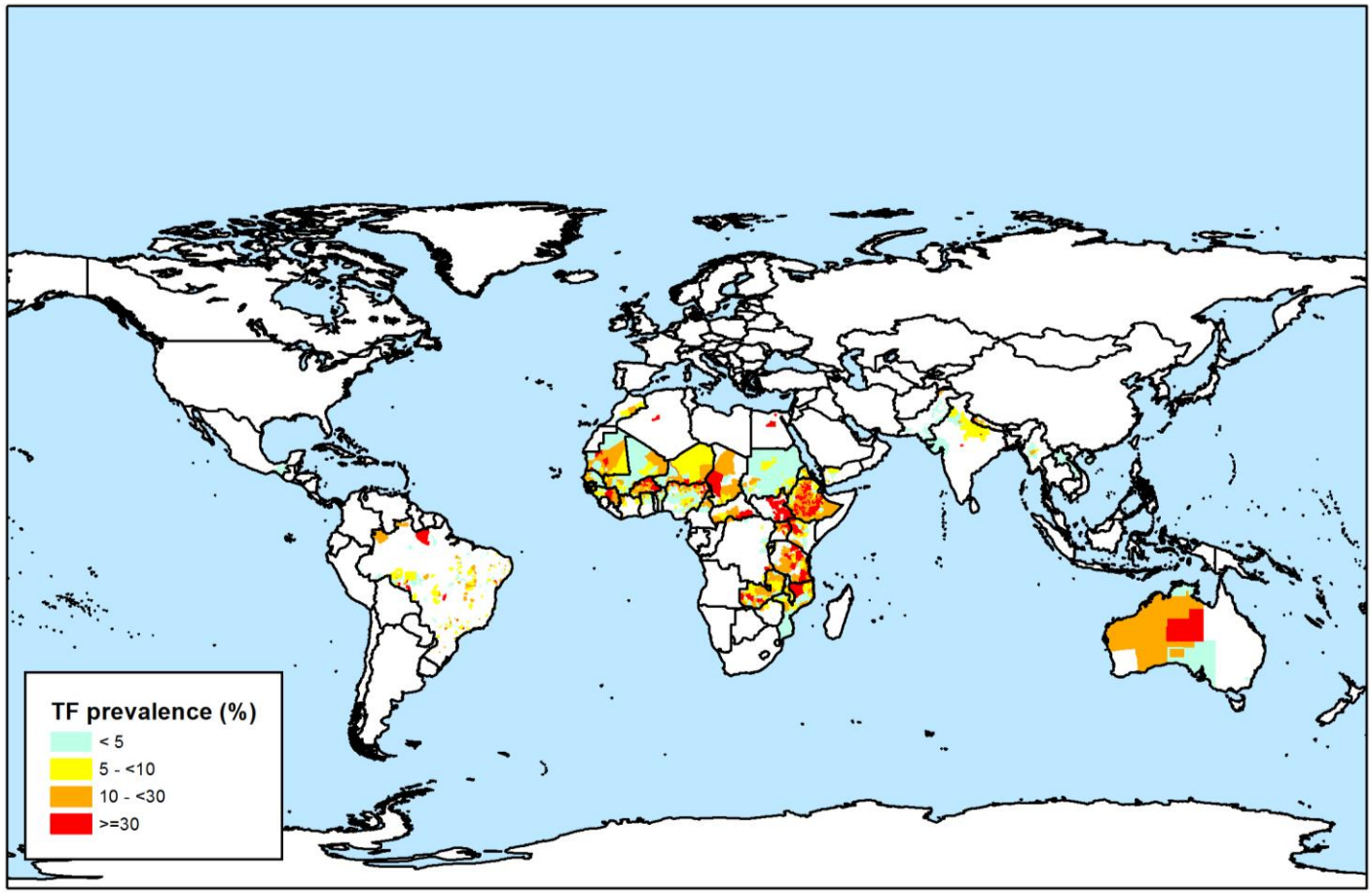


Figure 1.3 The global distribution of TF, where light green indicates district-level TF prevalence below the elimination threshold of 5%, yellow indicates district-level TF prevalence below the threshold for Mass Drug Administration (MDA) but above the elimination threshold, orange indicates district-level TF requiring a minimum of three years of MDA, and red indicates district-level TF prevalence requiring a minimum of five years of MDA.

1.3.2 Intervention strategies

The World Health Assembly Resolution of 1998 targets the elimination of trachoma as a public health problem by the year 2020 [2]. A series of global trachoma scientific meetings (GTSM) resulted in generating the elimination targets for TF as, less than 5% prevalence in children aged 1-9 years and TT as less than one person with positive TT in either eye, who is not already known to the health system per 1,000 population per health district (TT prevalence of less than 0.1%) [73].

The strategy to eliminate trachoma as a public health problem is community wide intervention, rather than individual treatment. The strategy is named **SAFE (**Figure 1.4**)** and involves (**S**) surgery to correct trichiasis, (**A**) mass drug administration (MDA) of azithromycin and improved sanitation and hygiene, focusing on (**F**) facial cleanliness and (**E**) environmental improvement [2]. This strategy is based on the identification of clinical signs, TF and TT.

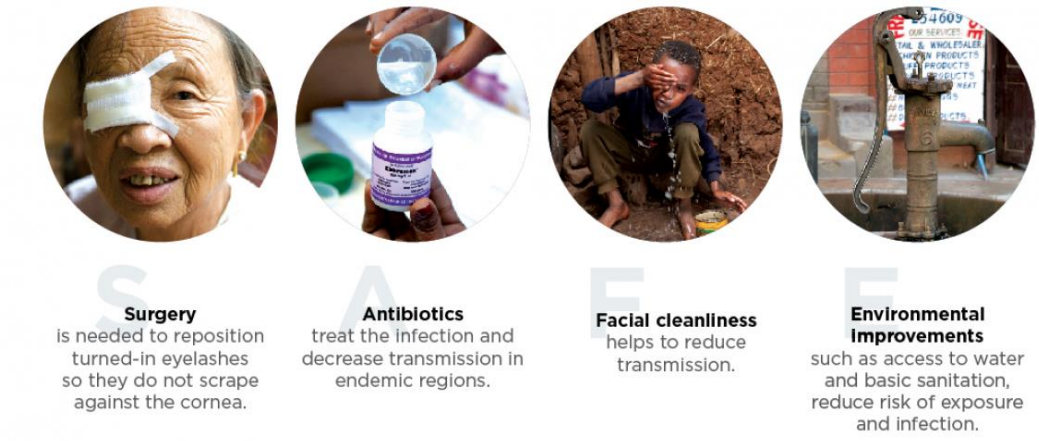


Figure 1.4 ***SAFE*** *diagram, provided by the International Trachoma Initiative*

The **A F E** intervention recommendations are based on district level TF prevalence. A TF prevalence of <5% warrants no intervention. A TF prevalence of 5-<10% warrants a sub-district level evaluation. A TF prevalence of 10-<30% warrants three years of MDA along with district wide facial cleanliness and environmental improvement campaign. A TF prevalence of >=30% warrants a minimum of five years of MDA along with district wide facial cleanliness and environmental improvement campaign. Once the appropriate number of treatment cycles are completed, impact assessments are conducted to re-evaluate the disease burden [73]. The result of the impact assessments informs the next steps, which are either continuing or stopping **A F E** interventions.

The **S** intervention is determined based on district level TT prevalence. A TT prevalence of <0.1% in the entire population (or 0.2% in adults aged ≥15-years) does not warrant surgical intervention. A TT prevalence of ≥0.1% (or 0.2% in adults aged ≥15-years) warrants surgical campaigns within the district. There is a lack of guidance on what to do if TF is below this threshold and TT is above this threshold as there is not currently a TT specific survey methodology (Figure 1.5).

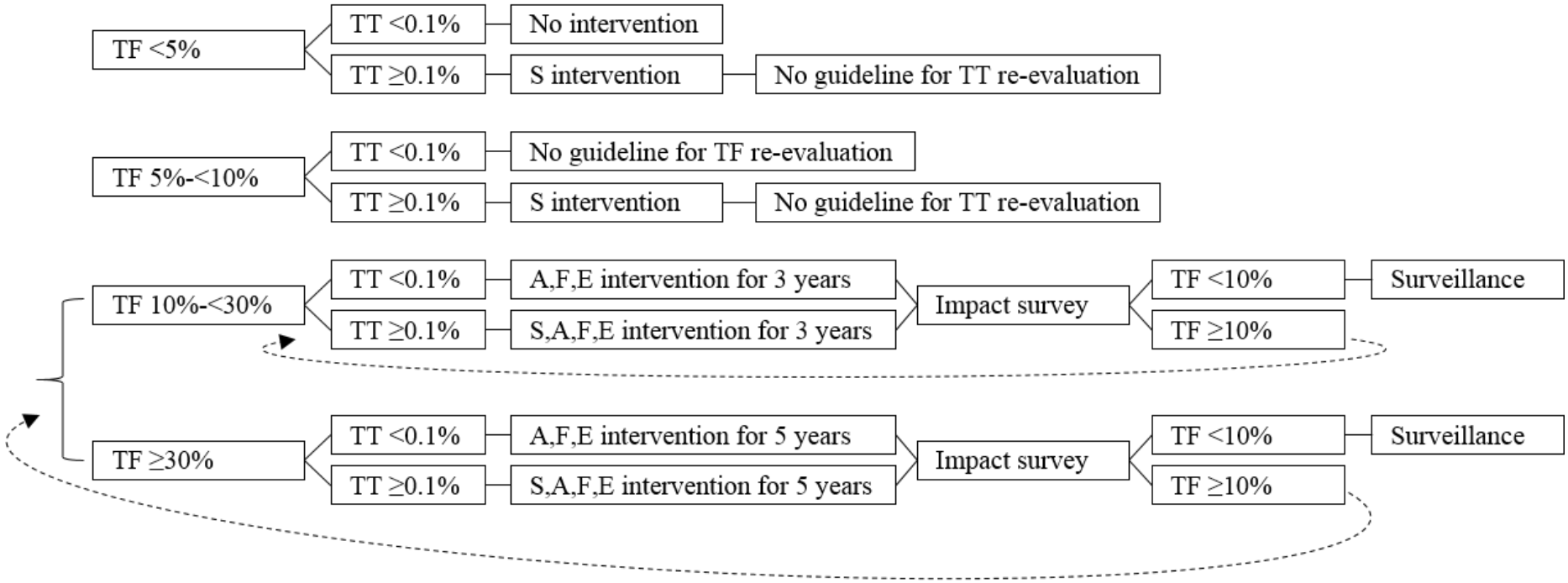


*Figure 1.5 Trachoma programme flowchart demonstrating the decision tree associated with district level prevalence outcomes. Where district level TF is below 10% in children aged 1–9 years no **AFE** interventions are warranted, where TF is between 10% and 30% **AFE** interventions are warranted for a minimum of three years, where TF is 30% or higher **AFE** interventions are warranted for a minimum of five years. Where district level TT prevalence in the entire population is less than 0.1% no **S** intervention is warranted and where prevalence is 0.1% or higher in the entire population **S** intervention is warranted. There are gaps in guidance around when to re-evaluate if the TF and TT prevalences are not aligned (i.e. there is transmission but no morbidity or there is morbidity and no transmission).*

These recommendations often leave country programmes confused. There is no mechanism for evaluating TT in absence of TF or clear surveillance strategy guidelines. In the following section I will review each component of the **SAFE** strategy, focusing on the practicality of the interventions.

1.3.2.1 Surgery

Trichiasis can be corrected through quality surgery. The decision on who should receive surgery is the first step. Burton *et al.* demonstrated that disease progression can move quickly, with 33% of minor TT cases progressing to major TT cases in the Gambia, over one year [66]. However, there are different thoughts on providing surgery to those individuals with minor TT. Some recommend providing surgery to all TT cases [74] whilst others suggest exploring epilation in cases where TT is minor [75]. Weighing the cost-benefit of surgery is certainly delicate and specific to individual circumstances. Whilst, trachoma examiners grapple with personal ethics on when and to whom surgery should be considered the WHO has provided clearer recommendations. The WHO recommends a posterior lamellar tarsal rotation (PLTR) as the surgical technique replacing the previously recommended bilamellar tarsal rotation (BLTR) [76], because a clinical trial showed that PLTR led to a lower rate of recurrence [77]. This randomised control trail was conducted in Ethiopia and included 1,000 participants with trichiasis. The study found 22% recurrence among the BLTR cohort and 13% among the PLTR cohort. Successful surgery can halt progression to corneal opacity and relieve pain (Figure 1.6).

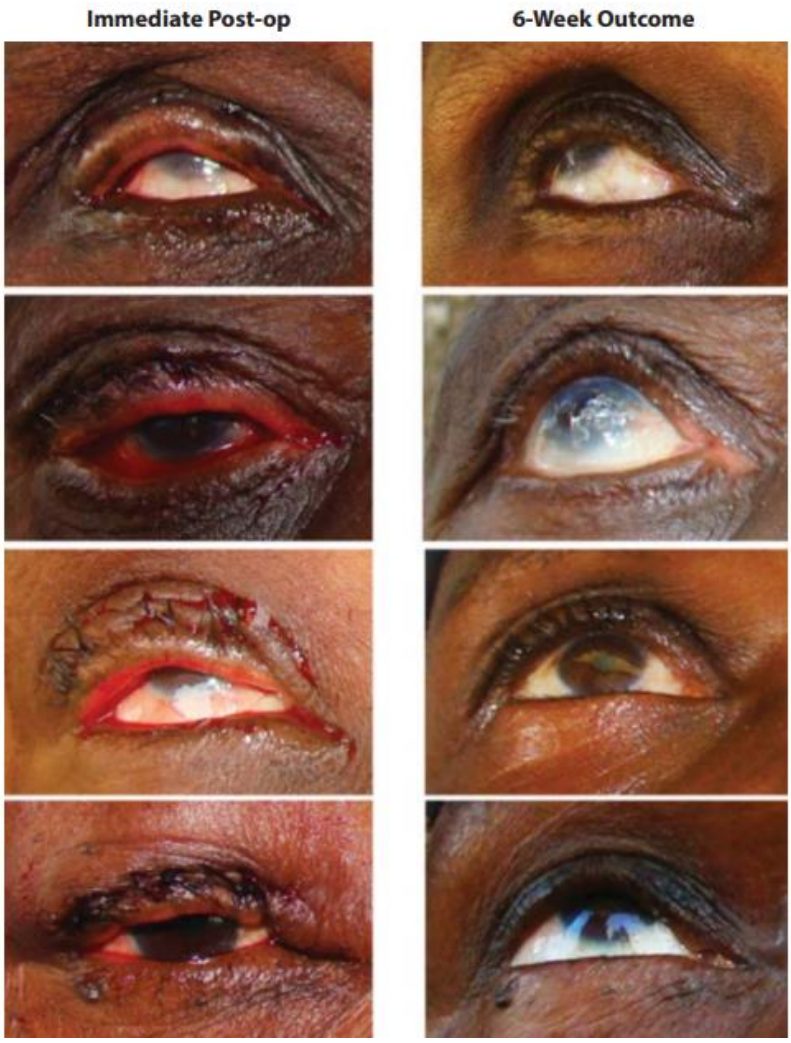


Figure 1.6 Images of trichiasis surgery outcomes adapted from Merbs et al. (2015) [78].

Post-surgery trichiasis reoccurrence is frequent and may occur for many different reasons, including; preoperative disease severity, quality of surgery, and surgical procedure performed [79]. A study in The Gambia was conducted to understand the long-term outcome of surgery. A cohort of patients who had undergone surgery three to four years earlier were re-assessed through a clinical exam, polymerase chain reaction (PCR) and bacterial culture. Examination were performed on 141 people and recurrent trichiasis was found in 41.6% of the operated eyes [80]. In 2010, an additional study was conducted reviewing the reoccurrence of TT post-surgery. The reoccurrence rate remained high ranging from 32%-41% [81].

To address surgical quality, the HEAD START programme was established. HEAD START involves an enhanced training system where all persons training in trichiasis surgery are taught by certified national master trainers and undergo hands on practical training using a mannequin [82]. The adoption of life-like simulations is recommended by WHO for both new trainings and refresher exercises. The broad adoption and success of this HEAD START for trichiasis surgeries has inspired the development of a similar life-like simulation mannequin for lymphatic filariasis (LF) hydrocele surgical trainings.

There has been a substantial recent uptake in trichiasis surgeries being performed, with 260,000 reported surgeries in 2016 [83] compared to 66,000 reported in 2011 [84]. This may partially be a result of an increase focus on TT surgery within trachoma elimination programmes along with the institution of HEAD START.

To plan for surgeries, it is fundamentally necessary to understand how many surgeries are needed and where these patients are most likely to live. Current estimates suggest 7.26 million people are living with TT [84]. This estimate was generated through a review of WHO 2011 country reports, where methodologies used in each country to generate estimates were not described and so this estimate is extremely limited. I will re-evaluate this estimate considering the availability of new data in Chapter 2.

Whilst my thesis focuses on the TT stage of trachoma, the distribution of TF likely influences the distribution of TT. Understanding the rationale around TF interventions is important in framing the context of TT distribution. Figure 1.7 demonstrates at what stage these intervention strategies effect trachoma progression. Delivery of these

interventions interrupt trachoma progression on the pathway to TT. Preventing re-infection should reduce the amount of TT.

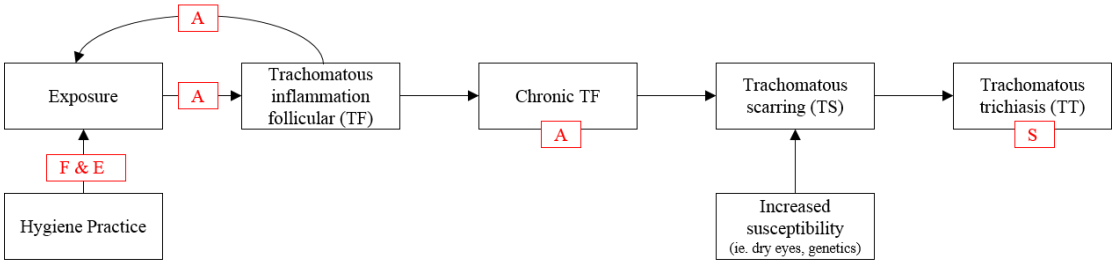


Figure 1.7 Trachoma progression and intervention

1.3.2.2 Antibiotic

Current guidelines recommend MDA with azithromycin once annually in trachoma endemic districts. A prolonged course of topical tetracycline or a single dose of Azithromycin are effective against chlamydial infections. A single dose Azithromycin is more effective due to improved compliance [85]. The pharmaceutical company Pfizer Inc. has pledged to donate the antibiotic through the International Trachoma Initiative.

Questions remain surrounding the most efficient treatment cycles of antibiotic. A cross-sectional survey in Tanzania with baseline trachoma and average MDA coverage data was conducted [86]. Coverage was defined as number of persons treated/number of persons in the community. A multivariate linear model of trachoma prevalence, adjusted for baseline trachoma rate and the average treatment coverage, suggested that each additional round of mass treatment beyond three years decreased the prevalence by an absolute value of 1.6%. The model suggests communities with baseline TF prevalence of 50% and an average annual MDA coverage of 75% would require seven years of MDA to bring the TF prevalence below 5%. The findings of this study are consistent with a global study conducted in 2014 which used linear and logistic regression models to show the probability of achieving the <5% TF target in different

scenarios [87]. The results showed the probability of achieving the target in areas with a baseline TF >=30% was <50%. Both studies are limited by the data available. MDA coverage is extremely difficult to measure. The numerator (number of persons treated) is routinely recorded as MDA is carried out. However, the denominator is challenging as it fluctuates and is difficult to accurately measure at the time of MDA. Because of the lack of reliability in coverage calculations, the second study simply used number of rounds of MDA. This is perhaps a more programmatically realistic approach but loses information around effective MDA.

The current guidelines recommend a single round of MDA annually. However, the value of biannual rounds of MDA has been explored [88]. Using an age structured mathematical model, estimates of the initial doubling time (IDT) for trachoma were developed. The model suggests MDA should be annually administered in areas where TF prevalence is less than 50% but in areas where TF is greater than 50% MDA should be administered biannually.

A 2011 cluster-randomised trial in Ethiopia, randomly assigned subdistricts to receive annual (mean baseline TF prevalence 41.9%) or biannual (mean baseline TF prevalence 38.3%) azithromycin treatment [89]. The prevalence of ocular chlamydial infection in children was not different between the two groups at 18, 30, and 42 months (P-value>0.99). However, a Cox proportional hazards model shows that the mean elimination time in the biannual subdistricts was 7.5 months earlier (2.3-17.3) than that of the annual group (P-value = 0.10).

Additionally, a 2016 cluster-randomized trial in Niger found no significant difference in TF prevalence among communities where children were treated annually and

biannually [90]. The mean baseline TF in these communities was 27.7% and 24.3% respectively.

All of these study results suggest that the azithromycin donation programmes should consider planning additional years of MDA and biannual treatments in areas with very high levels of TF prevalence.

1.3.2.3 Facial Cleanliness

Facial cleanliness is included as part of the trachoma intervention strategy because clean faces reduce the infectious ocular and nasal discharge which attract eye seeking flies [53]. This practice may also reduce the occurrence of auto reinfection by removing the pathogen from the eye [91]. An intervention trial was undertaken in Kongwa, Tanzania where villages were assessed on the impact of face washing intervention with MDA or MDA alone [92]. After 12 months of follow-up, the children who received face washing intervention along with MDA were 60% more likely to have clean faces than those who only received MDA, but there was no difference in the prevalence of TF between the intensive face washing and comparison arms. It was further determined that a clean face at two or more follow-up visits was protective for TF (OR 0.58). This was an intense, very expensive and time-consuming intervention. Sustained hygiene behaviour change is associate with reduced TF. However, face-washing is very difficult to implement on its own and so should be included in general hygiene behaviour practice programmes.

1.3.2.4 Environmental improvement

As discussed previously, the presence of *M. sorbens* is occasionally associated with trachoma risk. The larval medium for *M. sorbens* is human faeces [93]. However, larval

stages have not been found in latrines and adult *M. sorbens* have not been observed emerging from latrines [94]. A community-based cluster randomized controlled trail was conducted in The Gambia, where seven sets of three village clusters were randomly assigned to either (1) receive regular insecticide spray (space-spraying with water-soluble permethrin every two days for two weeks followed by maintenance spraying twice a week) (2) install pit latrines in each household or (3) receive no intervention (control).

In the insecticide villages, the number of flies found on children's eyes was 88% less than the control. In the pit latrine villages, the number of flies found on children's eyes was 30% less than the control. The cluster level trachoma prevalence in the insecticide villages demonstrated a statistically significant mean reduction of 56% and the pit latrine villages did not demonstrate a statistically significant difference in comparison with the mean rate change in the control [95].

This is a single study and it may not be appropriate to generalize to the larger trachoma endemic world. Whilst in The Gambia, *M. sorbens* seem to be an important vector for trachoma, this is not necessarily the case in other settings. A similar study was conducted in Tanzania and found that the trachoma rates were not significantly different between the intervention (community treated with insecticide and monitored for fly count) and control groups [60].

A systematic review of the effect of water, sanitation, and hygiene on the prevalence of trachoma was conducted in 2014 [96]. The review found evidence to support the **F** and **E** components of the **SAFE** strategy. However, the authors highlight the need for standardised approaches for measuring progress in this area. Programmes have

difficulty implementing a facial cleanliness and environmental improvement intervention as it is unclear what is meant by district wide facial cleanliness and environmental improvement campaigns. There are no concrete guidelines on what is necessary from the water and sanitation stand point to stop the transmission of trachoma.

The **S** component of the **SAFE** strategy is clearly the focus for addressing TT morbidity. However, the progression of TF to TT can be slowed or even halted through earlier interventions such as **AFE** and is some cases may eliminate the need for future surgical interventions.

1.4 Trachoma in global context

1.4.1 Sustainable development goals

In 2015, 193 world leaders committed to adopting the sustainable development goals (SDGs), which build upon the Millennium Development Goals (MDGs). The aims of the SDGs are to end poverty, protect the planet and ensure prosperity for all [97]. The SDGs specifically relevant to trachoma interventions are: (1) end poverty in all its forms everywhere, (3) ensure healthy lives and promote well-being for all at all ages, (4) ensure inclusive and quality education for all and promote lifelong learning, and (5) achieve gender equality and empower all women and girls. Callahan *et al.* evaluated trachoma control initiatives in the context of the MDGs [98] and this evaluation is still very applicable to the SDGs. This publication addresses the evidence suggesting **SAFE** interventions as part of trachoma control programmes play a valuable role in the reduction of extreme poverty.

It is well established that trachoma disproportionately effects women and children [99]. Because of the infectious nature of ocular chlamydia, children are the main reservoir for the disease. Children wipe nasal and ocular discharge on their clothes, share toys and personal space with other children and so continuously re-infect one another. Women are typically the caregivers of children in these communities. This means women are also more exposed to re-infection [100]. It is commonly believed that the re-infection of children and women leads to a pattern of disenfranchisement in women and girls. When a woman develops TT and is no longer able to perform household tasks someone else must take on her responsibilities. I suspect that this creates as a cycle, increasing risk of disease and lack of education in women and girls. For these reasons the elimination of trachoma as a public health problem is aligned with the SDG (1) end poverty in all its forms everywhere, (3) ensure healthy lives and promote well-being for all at all ages, (4) ensure inclusive and quality education for all and promote lifelong learning, (5) achieve gender equality and empower all women and girls, (6) ensure access to water and sanitation for all, and (17) revitalize the global partnership for sustainable development.

As outlined in the previous section, azithromycin is distributed to entire communities during MDA campaigns for trachoma. Azithromycin is a powerful antibiotic which has been proven to reduce malaria [101] and treat many sexually transmitted infections (STIs) [102, 103]. Furthermore, azithromycin has been shown to significantly reduce morbidity associated with diarrheal symptoms [104, 105]. The **A** component of the **SAFE** strategy directly links to the SDG (3) ensure healthy lives and promote well-being for all at all ages.

The **E** (environmental improvement) component of the **SAFE** strategy aligns nicely with the SDG (6) ensure access to water and sanitation for all. Trachoma programmes prioritize the construction of latrines and improved access to water sources.

Section 1.5 will discuss why the economic implications of trachoma are important, and how trachoma contributes to loss in productivity and increase in poverty [106] aligning with SDG (1) end poverty in all its forms everywhere.

Finally, successful trachoma intervention is only possible with the cooperation of many stakeholders, including ministries of health, pharmaceutical partners (Pfizer, donates azithromycin (Zithromax®) to endemic countries) and global partners such as WHO. Trachoma elimination programmes certainly adhere to SDG (17) revitalize the global partnership for sustainable development.

There is clearly a link between scaling up of trachoma intervention programmes (**SAFE**) and the SDGs. This link has been useful in advocating for trachoma intervention resources.

1.4.2 Trachoma elimination metrics and benchmarks

In 2011, *Accelerating work to overcome the global impact of neglected tropical disease: A roadmap for implementation*, was approved by the World Health Organization's Strategic and Technical Advisory Group for NTDs (STAG-NTD) [107]. This was a substantial moment in the NTD community, where WHO facilitated a clear set of targets and benchmarks from which progress to the 2020 elimination goals can be measured. Trachoma elimination targets set by country governments in 2011 are provided in the *roadmap*. According to the *roadmap*, to achieve the goal of global elimination of trachoma as a public health problem by the year 2020, 10% of endemic

countries would achieve elimination by the year 2013. This means five countries (10% of the 52 endemic countries) should demonstrate district level prevalence of TF of <5% and district level prevalence of TT of <0.1% in the entire population (or 0.2% in the ≥15-years population) in all previously endemic districts. Furthermore, 40% of endemic countries should achieve elimination by the year 2016. In addition, by 2015 50% of people at risk for trachoma should be receiving MDA and 100% of countries endemic for trachoma should have a plan of action for integrated MDA.

The *roadmap* elaborates on the need to plan for a post-endemic future. Many districts and countries are expected to reduce prevalence below the intervention thresholds soon and so plans for scaling down intervention and appropriate surveillance techniques need to be planned for. The *roadmap* especially notes the need for preparation and innovation now of epidemiological tools, diagnostics and models to help identify maintenance strategies.

Also in 2011, The end in sight: 2020 INSight [108], was published by the international coalition for trachoma control (ICTC). This document provided metrics and benchmarks for reaching 2020 and has played a critical role in securing resources for scaling up trachoma elimination efforts. The ability to measure and set clear expectations is crucial for advocacy and donor relationships.

Inspired by the *roadmap*, the London Declaration on Neglected Tropical Diseases (2012) Ending the neglect and reaching 2020 goals [109], renewed the call for control and/or elimination by 2020.

1.4.3 Scaling of resources

In part, as a response to the *roadmap* and 2020 INSight metrics the UK government (DFID) pledged to provide £10.6 million and the US government (USAID) pledged to provide an additional £6 million to complete the baseline mapping of trachoma. This initiative was named the GTMP and successfully completed trachoma baseline mapping in 29 countries (2.6 million people examined) between December 2012 and January 2016.

In 2013, the Queen Elizabeth Diamond Jubilee Trust pledged £42.8 million to fight trachoma in the commonwealth countries in Africa [110]. The specific plans for this funding are to (1) provide trichiasis surgery to an estimated 200,000 people, (2) facilitate MDA for an estimated 12.4 million people, (3) improve messaging around community hygiene and sanitation practices, and (4) work with stakeholders to improve access to safe water sources.

In 2014, DFID pledged £39 million to support the elimination of trachoma as a public health problem through the **SAFE** strategy [110].

1.5 Economics of trachoma

Along with the epidemiology of trachoma it is also important to understand the economic impact of the disease. Economics are a driving force in society and so the ability to frame a disease in the context of cost-benefit is crucial for programme growth. To understand the economic implications of trachoma, it is first necessary to determine the disability associated with TT. Because TT is the morbidity stage of trachoma it is the stage where economic implications can most clearly be identified. To truly appreciate the economic consequences of TT it is imperative to understand the burden.

The 2016 estimated disability-adjusted life years (DALYs) attributed to trachoma was 245 thousand (162-355 thousand CI). This estimate was made based on published data, with the most recent citation occurring in 2014. No results from the GTMP were included and much of the data was derived from publications of rapid blindness surveys. With an updated estimate of the global burden of TT available, the DALY estimate could be much improved.

1.5.1 Visual impairment

The most recently available estimate of visual impairment was published by WHO in 2012, representing the situation in 2010 [111]. The prevalence of visual impairment and blindness were calculated for each WHO region and can be found in Table 1.2.

Table 1.2 Adapted from Global Data on Visual Impairments 2010. Values represent estimated millions of people with visual impairment. (1) excludes India (2) excludes China

WHO Region	Total population (percentage)	Visual Impairment (percentage)
African	804.9 (11.9)	26.3 (9.2)
Americas	915.4 (13.6)	26.6 (9.3)
Eastern Mediterranean	580.2 (8.6)	23.5 (8.2)
European	889.2 (13.2)	28.2 (9.9)
South-East Asian[1]	579.1 (8.6)	27.9 (9.8)
western Pacific[2]	442.3 (6.6)	14.7 (5.2)
India	1,181.4 (17.5)	62.6 (21.9)
China	1,344.9 (20	75.5 (26.5)
World	6,737.5 (100)	285.4 (100)

A global estimate of the prevalence of trachomatous blindness and low vision was generated in 1990 [112, 113]. The study estimated a total of 2.9 million cases of trachomatous blindness and 3.8 million cases of visual impairment. Whilst, the estimate provided in Table 1.2 and the trachomatous blindness estimate generated in 1990 are taken from different time points and cannot be directly compared, these values do provide a context for the scale of visual impairment resulting from trachoma. Through

understanding the current burden and distribution of TT these values could be more discreetly defined.

The 1990 study incorporated the years of life lost resulting from premature mortality and the years lived in a handicapped state to create a single measure of disease burden, handicap adjusted life years (HALYs). The HALYs were multiplied by the prevalence of trachomatous visual impairment to generate an estimate of global burden of trachoma visual impairment of 80 million HALYs.

1.5.2 More than vision loss

Whilst the obvious economic measurement from trachoma is derived from visual impairment, it is also important to consider the pain and photophobia associated with trichiasis. This important element is pointed out by Frick *et al.* [114]. A study in Tanzania found that having trichiasis even without visual acuity loss limits the ability to perform ADL tasks (activities of daily living). The study also found those with visual acuity loss and trichiasis had a greater limitation than those with visual acuity loss from other causes [115, 116]. This is a relevant finding, highlighting the possibility that only using vision loss in calculating economic loss may be resulting in an underestimate. The fact that individuals with trichiasis report more difficulty in conducting ADL tasks than those with visual acuity loss from other causes means there is an additional factor influencing the ability to conduct these tasks, not only vision loss.

1.5.3 Cost effectiveness

The cost associated with lost productivity caused by trachoma was estimated in 1995 to be $2.9 billion (USD) annually [117]. More recent (2003) estimates suggest a total loss in productivity between $3.5 billion (USD) and 5.3 billion (USD) [117].

Trachoma exacerbates the cycle of poverty by increasing the burden of blindness and low vision [3]. Prevention of blindness along with reducing trachoma burden through the **S**, **F**, and **E** components of the **SAFE** strategy reduces disability and improves productivity in communities.

Trichiasis surgery is a very cost effective intervention [118]. A study in The Gambia estimated the average cost of untreated trichiasis to be $89 (USD), based on life-time lost economic productivity. In The Gambia, on average, the surgery costs $6.13 (USD). To fully appreciate the finances needed for TT elimination, understanding the burden and distribution of the burden is essential.

1.6 GIS and Geostatistical methods

The above review highlights the limited understanding of TT burden distribution. There is a great selection of GIS resources that can be used to improve this understanding, providing better characterisation and exploration. In this section I will describe existing GIS resource and spatial modelling for trachoma. I will then outline additional potential approaches to spatial modelling and their applications to trachoma.

1.6.1 Geostatistical methods

Tobler's first law of geography, "everything is related to everything else, but near things are more related than distant things" [119] is the guiding principle applied in geostatistical methods. Semi-variograms are used to test whether this law is applicable to a dataset. A semi-variogram illustrates the rate of decay of spatial autocorrelation between observations. Semi-variance is generally low when locations are near to each other and increase as distance grows between observations until a point is reached where the locations are independent of each other [120, 121].

1.6.1.1 Semi-variograms

A semi-variogram illustrates the rate of decay of spatial autocorrelation between observations through plotting half of the mean squared difference between a pair of observations as a function of the distance separating the observation location (lag) [121] (Figure 1.8). The semi-variogram formula is defined as

$$\hat{y}\,(h) = \frac{1}{2n(h)} \sum_{n(h)} (x(x_i) - z(x_i + h))^2$$

where x is the given location; z is the value of an attribute; h is a given separate distance between observations (lag); $n(h)$ is the number of pairs of sample points of observations of the value of z separated by distance h.

Semi-variograms invoke Tobler's first law of geography [119]. This means that semi-variance is generally low when locations are near to each other and increase as distance grows between observations until a point is reached where the locations are independent of each other. The nugget is the variability in data that cannot be explained by distance between observations. The sill is the maximum variability observed. The difference between the sill and the nugget (partial sill) is the amount of observed variation that can be explained by distance between observations. The range is the point where semi-variance stops increasing and observations are considered independent.

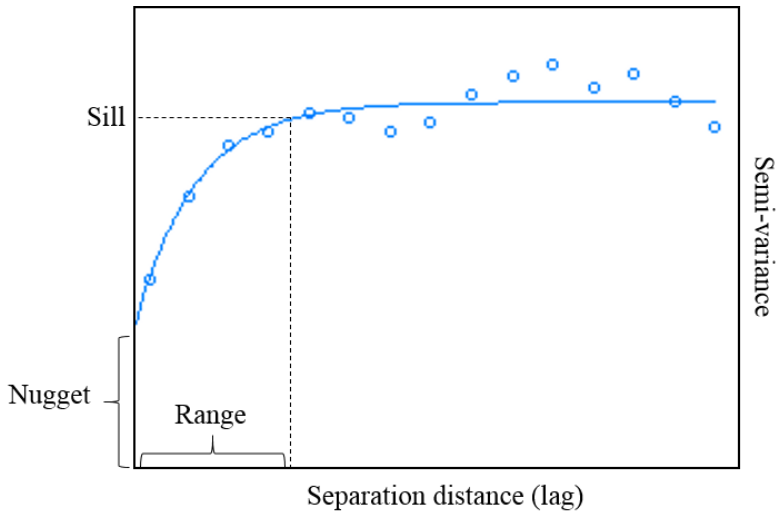


Figure 1.8 Example semi-variogram, where the sill is the maximum value of semi-variance, the range is the lag distance at which the sill is reached and the nugget is the variability at a distance smaller than the sample spacing

1.6.1.2 Cluster detection

It is often necessary to not only determine whether there is a spatial structure to the data, but if there is clustering within the data. Knox *et al.* defines a cluster as, a geographically and/or temporally bound group of occurrences of sufficient size and concentration to be unlikely to have occurred by chance [122]. A variety of methods can be used for cluster detection, outlined in Table 1.3.

Table 1.3 Overview of cluster detection methods for geostatistical analysis

Type	Method	Application	Disadvantages
Global	Cuzick and Edwards k-nearest neighbour	Tests for spatial clustering whilst taking into account the distribution of the population at risk. The test uses the locations of cases and randomly selected controls and includes a spatial scale parameter. The test counts how many of the k-nearest neighbours are also cases [123].	The user determines the value for the scale parameter, making this parameter subjective.
	Moran's I	Quantifies the similarity of a variable among areas that are defined as spatially related. A weights matrix is used to define the spatial relationship so that areas close in space are given greater weight then those that are distant [124].	Assumes population at risk is evenly distributed within the area and correlation is the same in all directions.
	Ripley's k function	Compares a given point distribution with a random distribution.	There are no first order effects in the spatial pattern. Variance increases with distance, not suitable for estimating clustering over large distances [125].
	Mantel's test	Compares inter-event distances in space and time against a null hypothesis that time and space distances are independent [126].	Assumes a linear correlation [127]. Low statistical power with small number of cases [128].
Local	Getis and Ord's local statistic	An indicator of local clustering that measures the concentration of a spatially distributed attribute variable. Through comparing local estimates of spatial autocorrelation with global averages, the statistic identifies hot spots [129].	Results are dependent on the size of the features being analysed. When large areas tend to have low values and smaller areas tend to have high values.
	Kulldorff's spatial scan statistic	SaTScan – for each specific location a series of circles of varying radii is constructed. Each circle absorbs the nearest neighbouring locations that fall inside it and the radius of each circle is set to increase continuously from zero until a fixed percentage of the total population is included [130].	Imposes a shape on the clusters, and irregularly shaped cluster may not be detected.
	Local Moran's I	Detects local spatial autocorrelation in aggregated data.	Small geocoding errors can significantly influence results [131].

Spatial clustering of disease can be measured through global or local methods [132]. Global methods are used to determine whether clustering is present throughout the study area by providing a single statistical measure of the degree of spatial clustering. These methods simply identify if spatial clustering exists, but they do not identify where. Importantly, these methods identify if observed patterns are significantly different from spatial randomness. Local methods define the location and extent of clustering. These methods scan an entire dataset, but only measure spatial dependence in pre-defined limited portions of the study area.

Understanding if clustering of TT exists - and then disentangling risk factors associated with identified clusters is important for designing intervention strategies. Identifying groups of community-level TT prevalence that are higher than expected from spatial randomness, could be used for targeting placement of TT surgical camps. As noted previously during the end-game of trachoma elimination TT prevalence is expected to be very low (approaching 0.2% in adults aged ≥15-years), making it difficult to find and offer services to those affected. Cluster detection and a stronger understanding of spatially distributed risk factors could guide strategies and improve efficiency in case finding.

1.6.2 GIS and trachoma

Trachoma is mainly found in Sub-Saharan Africa, the Middle East and Asia [133]. The Global Atlas of Trachoma (www.trachomaatlas.org) serves as the global repository for current prevalence information gained from PBPS and TRAs (Figure 1.9). The atlas uses maps to illustrate the distribution of both TF and TT. This resource has been invaluable in informing country level planning as well as informing global burden

estimates. Calculations based on information provided to the atlas estimate in Africa 129.4 million people live in trachoma endemic (TF >=5%) settings [134]. The maps provided in the atlas illustrate the cross-border prevalence of trachoma and highlight the great burden of disease in Eastern Africa.

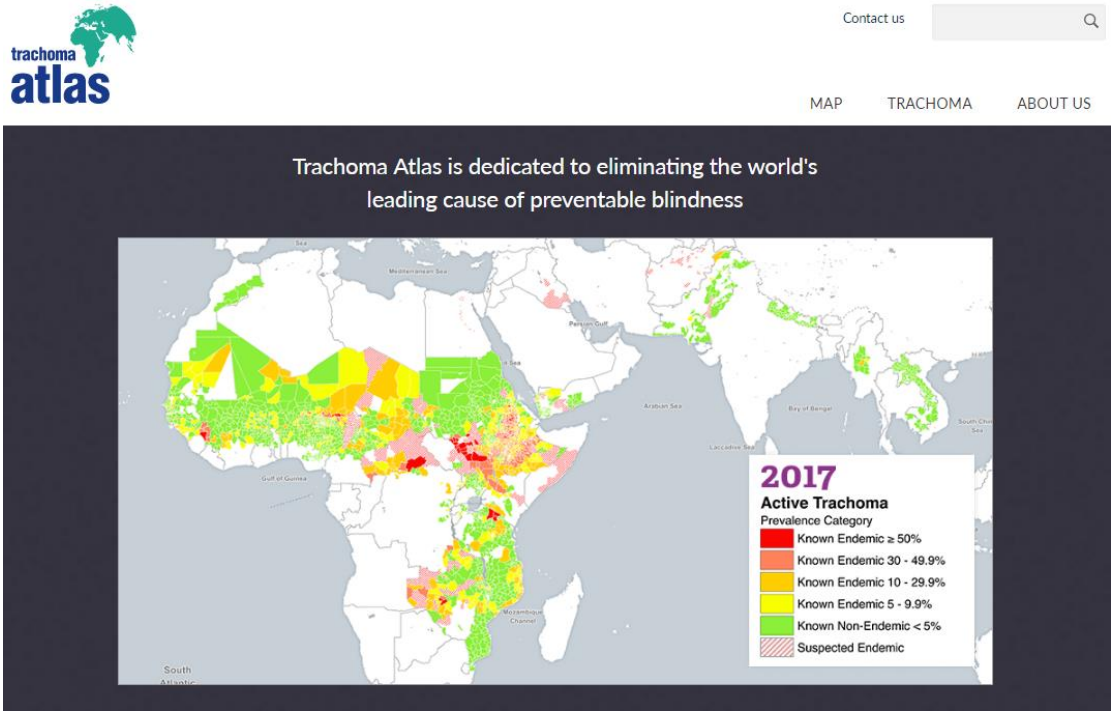


Figure 1.9 The global atlas of trachoma (www.trachomaatlas.org)

1.6.3 Characterising the spatial distribution of trachoma

Whilst the aggregate maps generated from these surveys are valuable planning tools, they are limited in that they do not provide information about how and why risk varies within areas. Understanding this variance would be particularly useful for TT intervention planning. Unlike TF, where an entire district receives treatment when prevalence is high, TT intervention (surgery or epilation) requires identification of actual afflicted individuals. As prevalence drops through both secular change and programme interventions, it will become more difficult to find the final TT cases. This

has been the experience of other disease elimination programmes such as Guinea worm [135] and leprosy [136-139]. Studies have identified spatial clustering of leprosy and have been useful in identifying high risk areas, informing where targeted intervention is needed [140-144]. Similarly, it may be possible to identify communities with a comparatively high number of TT cases, or hot spots.

A hot spot is often described as a small foci of intense disease within a generally low prevalence area. However, because available TT data was collected through a random sampling of clusters, it is expected that an occasional high prevalence community will be present in the data. I do not consider this single community a hot spot, rather an outlier or random artefact of the sampling technique that is not of public health significance. If the identified hot spot is larger than a single community, it is likely a true hot spot worthy of investigation. In my thesis I consider a TT hot spot, a community where the proportion of TT in surrounding communities is high and the value in the specified community is significantly higher than that in the surrounding communities.

Great efforts have been made in hot spot identification and analysis in malaria transmission. Specifically, predictive models have been developed to identify geographic areas of high levels of transmission intensity [145]. These models identify clusters of high intensity using a spatial scan method and then a regression model to determine significant covariates. Whilst the malaria community is at the forefront of hot spot identification and analysis, there are inherent differences in the epidemiology of malaria and trachoma. Malaria has an obligate vector and so clustering by indices for mosquitos makes biological sense. However, there is no obligate vector for trachoma and so the rational for suspected clustering is slightly different. As described in detail in section 1.2.2 trachoma is transmitted through close personal contact [30,

31], the transmissive nature of *chlamydia trachomatis* provides a strong hypothesis that local clustering of TF is expected. Additionally, the important role of sanitation and hygiene in TF transmission [32, 41, 45-48] suggests between community clustering. If it is also accurate to say that there is a spatial association between TF and TT (which is explored in Chapter 5:), there is reason to suspect that models could identify TT clustering (which is explored in Chapter 6:).

Very few spatially explicit analyses of trachoma have been published and of these, only two publications include TT in their analysis. These spatially explicit analyses of trachoma are outlined in (Table 1.4). With the availability of the GTMP geolocated data, it is now possible to build upon these earlier findings and expand analyses to multiple country settings.

Table 1.4 Summary of trachoma geostatistical published work

Year	Location	Summary	Limitations	Citation
2010	South Sudan	A trachoma risk map was created using Bayesian geostatistical models incorporating TF prevalence data in South Sudan TF in children along with long-term average rainfall, land cover, and geostatistical random effects describing spatial clustering of trachoma were included in the model. With a significant negative correlation between TF and rainfall (OR 0.21), Clements *et al.* concluded in South Sudan the spatial variation of trachoma is associated with aridity.	The model is limited in that it is static. Rainfall was used as a major predictor and seasonality may greatly affect the results. Hot and arid may simply be a proxy for poverty.	Clements AC, Kur LW, Gatpan G, Ngondi JM, Emerson PM, Lado M, et al. Targeting trachoma control through risk mapping: the example of Southern Sudan. PLoS neglected tropical diseases. 2010;4(8):e799. doi: 10.1371/journal.pntd.0000799. PubMed PMID: 20808910; PubMed Central PMCID: PMC2923154. [11]
2015	Nigeria	This analysis examined the relationship between climatic factors and TT and corneal opacity (CO) in Nigeria, whilst accounting for individual risk factors and spatial	The data used in the model was derived from a national blindness survey, including 304 clusters with large geographic areas not represented.	Smith JL, Sivasubramaniam S, Rabiu MM, Kyari F, Solomon AW, Gilbert C. Multilevel Analysis of Trachomatous Trichiasis and Corneal Opacity in Nigeria: The

		correlation. Data were fit to Bayesian hierarchical logistic models. The risk of TT/CO was associated with factors at both the individual and cluster levels. This study found strong evidence that environmental factors at the cluster-level were associated with a greater risk of TT/CO. The findings from this study are very promising and deserve further analysis in additional countries.		Role of Environmental and Climatic Risk Factors on the Distribution of Disease. PLoS neglected tropical diseases. 2015;9(7):e0003826. doi: 10.1371/journal.pntd.0003826. PubMed PMID: 26222549; PubMed Central PMCID: PMCPMC4519340. [146]
2010	Mali	The aims of this study were to disentangle the relative importance of clustering at different levels and to assess the respective role of individual, socio-demographic, and environmental factors TF prevalence in Mali. TF data were categorized into a hierarchy and village level environmental variables were identified. Bayesian hierarchical logistic models were fit to the data. Clustering was found to be significant at all four hierarchy levels. The model showed that individual level (age, dirty faces), caretaker-level (wiping after body washing), household-level (wealth), and village-level (women's association and climate) were associated with TF prevalence. This analysis highlights the need to not only focus interventions on positive cases, but also communities. This aligns with the existing **SAFE** strategy, where **AFE** are population based, rather than individual.	This study included data from only 202 clusters across the entire country limiting the generalizability of the findings.	Hagi M, Schemann JF, Mauny F, Momo G, Sacko D, Traore L, et al. Active trachoma among children in Mali: Clustering and environmental risk factors. PLoS neglected tropical diseases. 2010;4(1):e583. doi: 10.1371/journal.pntd.0000583. PubMed PMID: 20087414; PubMed Central PMCID: PMC2799671. [38]
2007	Mali	This study analyzed the role of geoclimatic factors in TF and TT distribution in Mali. Multiple linear regression models including latitude and	The limitations in this study are the same as those listed above.	Schemann JF, Laffly D, Sacko D, Zephak G, Malvy D. Trichiasis and geoclimatic factors in Mali. Transactions of the Royal Society of Tropical

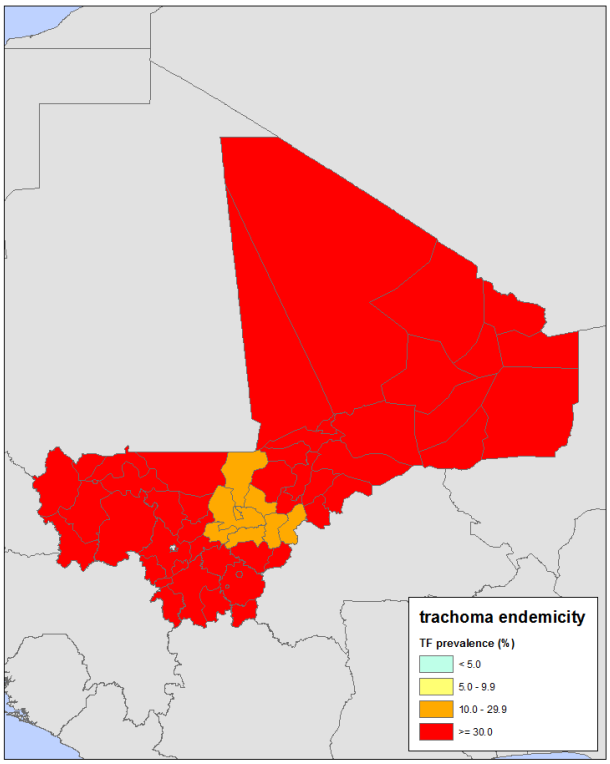
		longitude explained 6.1% of the variance for TF and 14.0% for TT. When environmental risk factors were added to the model, the coefficient of determination increased to 11.2% for TF and to 19.6% for TT. The outputs from these regression models were plotted in a predictive map and demonstrated an opposite spatial distribution of TF and TT, with a gradient from south-south-east to north-north-west for TF and a gradient from north-east to south-west for TT.		Medicine and Hygiene. 2007;101(10):996-1003. doi: 10.1016/j.trstmh.2007.05.015. PubMed PMID: 17658570. [67]

1.6.4 Using Geospatial approaches to guide post-treatment surveillance

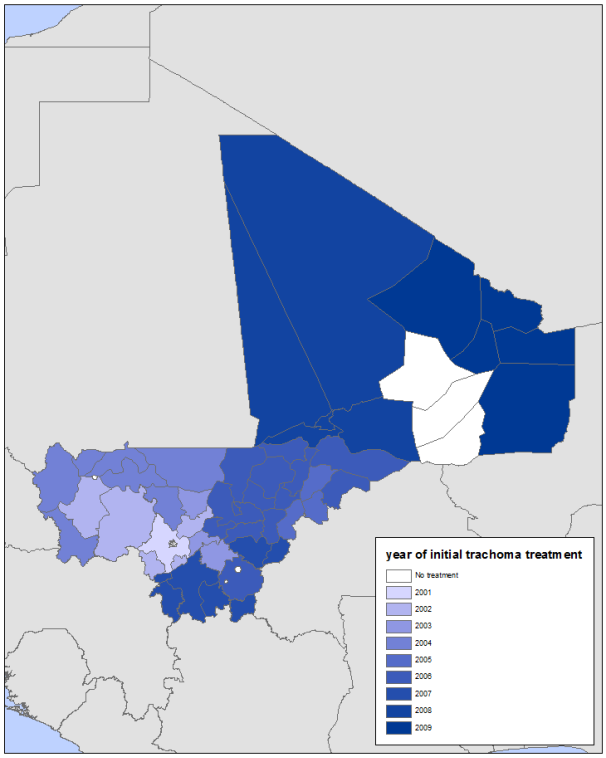
Many countries have implemented successful trachoma control programmes, and now need improved guidance on where and how to conduct post-MDA surveillance. Ghana is an example of a country that has implemented the **SAFE** strategy with great success. The post-intervention surveys show a reduction of TF below the elimination targets and so MDA has been discontinued in the country [147]. However, Ghana needs guidance on how to approach surveillance to ensure the disease does not return. The country also estimates nearly 5,000 people remain with TT; this backlog of surgical candidates clearly needs to be addressed.

Mali is also a country that has had great success with the **SAFE** strategy (Figure 1.10). The country has been implementing a community by community approach to surveillance. However, it is unrealistic for all programmes to survey every community within a district to ensure the disease is truly gone and so geostatistical techniques should be harnessed to improve efficiencies in surveillance activities.

Historic mapping



Treatment scale-up



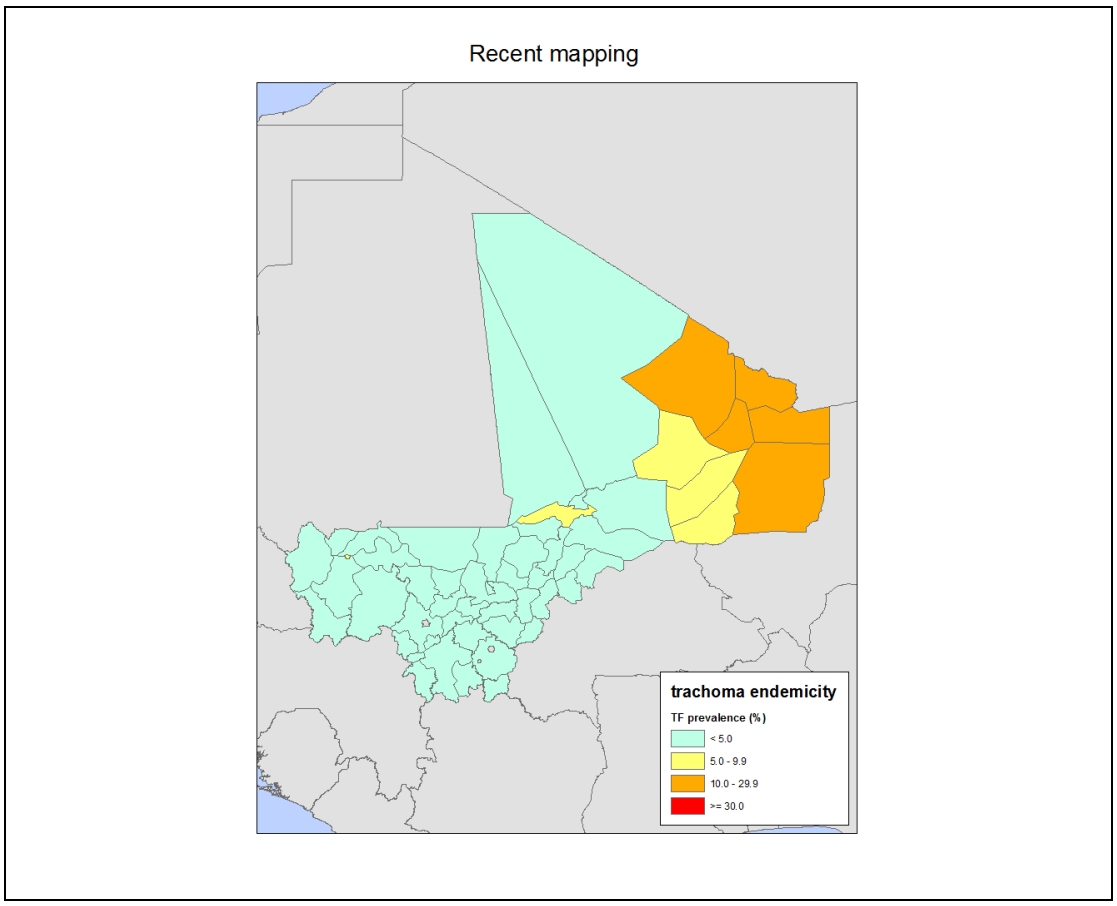


Figure 1.10 Impact of the SAFE strategy on TF prevalence in Mali. The historic map represents trachoma endemicity prior to MDA intervention, the treatment scale-up map identifies when MDA began in each district, and the recent mapping map represents trachoma endemicity after MDA.

Lietman *et al.* evaluated the use of mathematical models to predict community level infection in areas where trachoma is disappearing [148]. With longitudinal infection data from two sites in Ethiopia the fit of eight discrete distributions was assessed (geometric, binomial, Poisson, discrete weibull, negative binomial, beta binomial and zero-inflated geometric, and zero-inflated Poisson). In both survey sets the geometric distribution had the most parsimonious fit and the goodness-of-fit testing was also consistent with the community level data drawn from the geometric distribution. This model shows a long tail, which suggests that it is expected to have an occasional high prevalence community when the disease is on the way out and this community is not necessarily a sign of re-emergence. This work points to the ability to model at the end-stage. However, the longitudinal datasets used in this study come from Ethiopia.

Ethiopia has extremely high baseline prevalence values and information gleaned from this country may not be applicable to other settings.

Geospatial modelling may serve as a valuable tool in informing surveillance strategies for trachoma. With the large amount of data now available from GTMP, it may be possible to implement geospatial modelling techniques to explore the dynamics between trachoma and environmental factors in a global setting. This type of work has been done for other diseases, including tuberculosis [149, 150], podoconiosis [4], schistosomiasis [3, 5, 10, 13, 15, 151, 152], cholera [153] and leprosy [144]. These analyses have been used to identify spatial correlation with demographic and environmental factors which influence prevalence, as well as cluster detection.

The malaria community is at the forefront in applying geospatial methods in elimination programmes. Spatial prediction models as well as spatial decision support systems have been developed to enhance resource allocation [154]. Many studies have investigated the association between climatic and environmental drivers and malaria infection [155-161]. Associations have been found to exist between malaria prevalence and rainfall, temperature, normalised difference vegetation index, distance to rivers and streams, altitude, slope, land-cover class, and forest cover. These associations have been used to predict malaria outbreaks.

It may be possible to use additional elements as proxy measurements to predict where trachoma re-emergence is most likely to occur first. The ability to visualise the distribution of the disease at a micro level will lead to more focused and strategic intervention and surveillance strategies.

1.7 Justification

My thesis will focus on understanding the spatial heterogeneity of trachoma. I will particularly focus on the TT stage of the disease. As outlined in the previous sections there are knowledge gaps surrounding trichiasis surgical needs, with TT burden estimates outdated and generated from inconsistent methodologies. There is also a need for a purpose-designed, robust strategy for determining the TT burden at the implementation unit level. Extremely few studies have examined the spatial structure of TT nor have they disentangled the spatial association between TF and TT. A stronger understanding of TT distribution and spatial structure could provide valuable insight around where to focus TT intervention strategies. Understanding the spatial distribution may also inform whether the district (currently implementation unit) is the appropriate scale for addressing TT surgical needs. This work is timely, with the newly available GTMP data and specific requests from the global trachoma community for guidance on effectively reaching and providing surgery to those who need and want it.

1.8 Thesis aims and objectives

The dual aims of my thesis are to **quantify the current global distribution of TT**, and to **explore spatial heterogeneity in TT at fine spatial scales**. Addressing these aims will provide evidence to guide strategic use of resources for implementing TT management campaigns. These aims will be reached through the following objectives:

1. To quantify the global burden of TT in 2016, with associated uncertainty
2. To design and validate a robust strategy to assess TT burden in defined implementation units

3. To explore the relationship between community-level TF and TT, and identify factors that modify this relationship in a variety of country settings
4. To characterize the extent of spatial clustering of community-level TT, and identify factors influencing the level and degree of clustering

1.9 Thesis outline

Chapter 2 introduces the Global Trachoma Mapping Project (GTMP) dataset, which is referenced throughout this thesis. Chapter 3 describes the methodology and results of recalibrating and updating national and global TT burden estimates for 2016. Chapter 4 outlines the design and validation of a TT specific survey, which measures with precision appropriate for meeting the WHO elimination target. This chapter also assess the time costs associated with a TT specific survey. Chapter 5 evaluates the community level spatial distribution of TT and the association between TF prevalence and TT prevalence. This chapter further explores how additional risk factors effect this spatial relationship. Chapter 6 presents a methodology for identifying TT hot spots using geostatistical methods and identifies environmental factors associated with these hot spots. Lastly, Chapter 7 discuss the findings of this work and highlights implications on trachoma control programmes.

Chapter 2: **Understanding the GTMP dataset**

2.1 Overview

Much of the data used thought this thesis came from the GTMP. The GTMP was funded by (1) a grant from DFID (ARIES: 203145) to Sightsavers, and (2) USAID, through the ENVISION project implemented by RTI International under cooperative agreement number AID-OAA-A-11-00048, and the END in Asia project implemented by FHI360 under cooperative agreement number OAA-A-10-00051.

2.2 Background

The GTMP launched in December 2012 with the aim to complete the baseline map of trachoma through conducting standardised PBPS's. This project was funded by the UK government (£10.6 million) with additional funds provided by USAID (£6 million) and led by Sightsavers. The collaborative nature of the GTMP implementation ensured success. The GTMP partnership was made up 30 ministries of health (MoH), the London School of Hygiene & Tropical Medicine (LSHTM), the International Trachoma Initiative, the World Health Organization (WHO) and over 20 non-profits.

2.3 Methods

Standardised training materials were developed, and workshops conducted to ensure consistent quality grading of the trachoma clinical signs, particularly TF and TT. All GTMP graders participated in a "grader qualifying workshop", during which they received classroom- and field- based training. Only those who passed both a slide- and field-based test of diagnostic accuracy became members of the GTMP survey team.

Data collected during the GTMP was systematically entered into an android phone and synchronized to an amazon server. The recorders, who entered the data into the system, participated in the training workshop and only those recorders who passed a test of data capture accuracy became members of the GTMP survey team. Further detail of the training methodology is described elsewhere [71].

The PBPS were conducted at the evaluation unit- (EU) level, which typically equated to a district with a population of between 100,000-250,000. Within each EU, all residents one year or older from the randomly selected households in each of the randomly selected clusters were invited to be examined for trachoma clinical signs, specifically TF in children aged 1–9 years and TT in adults aged ≥15-years. Along with individual clinical trachoma data, the survey teams collected geolocated household-level water, sanitation and hygiene data.

2.4 Results

The GTMP completed in January 2016 and had collected data from 2.6 million people in 1,542 districts in 29 countries (Figure 2.1).

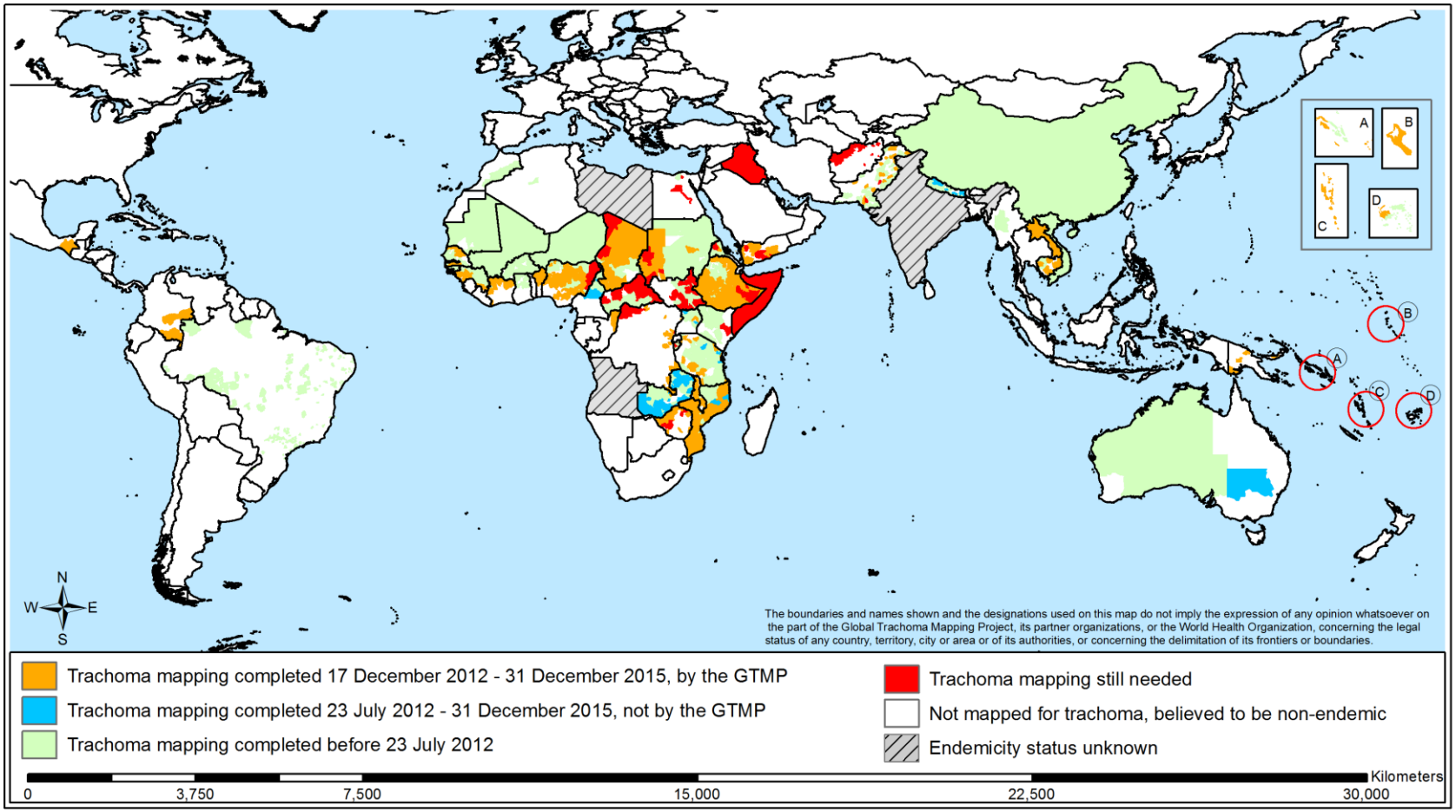


Figure 2.1 Map outlining where mapping was completed prior to GTMP, during GTMP (along with methods), and where mapping is still needed (all outstanding districts were unable to be reached through GTMP because of security risks to the survey teams). This map was generated for the "Completing the Global Trachoma Map" manuscript, which is pending publication.

Of the 1,542 districts, 47% were above the district-level TF elimination threshold of 5% in children aged 1–9 years and 53% were above the district-level TT elimination threshold of 0.2% in adults aged ≥15-years (Table 2.1).

Table 2.1 GTMP district count and distribution of TF and TT within each country by elimination threshold categories

Country	Districts	TF>= 5%	TT>= 0.2%
Benin	26	8	19
Cambodia	14	0	1
Chad	48	19	34
Colombia	3	3	1
Congo	6	0	0
Cote d' Ivoire	11	10	2
DRC	30	17	23
Egypt	4	4	4
Eritrea	8	8	4
Ethiopia	506	465	458
Fiji	1	0	0
Guinea	15	7	5
Kiribati	1	1	1
Laos	106	0	0
Malawi	25	14	8
Mexico	1	0	0
Mozambique	118	46	35
Nigeria	288	43	138
Pakistan	49	7	4
Papua New Guinea	7*	6	0
Senegal	17	1	8
Solomon Islands	4	4	0
Sudan	45	14	28
Tanzania	21	3	13
Uganda	4	0	1
Vanuatu	1	1	0
Yemen	164	31	19
Zambia	3	1	2
Zimbabwe	16	11	10
Total	1,542	724	818
*Data transmission challenges delayed the release of the final prevalences associated with one district in Papua New Guinea			

The largest number of districts surveyed were in Ethiopia (506) followed by Nigeria (288). Excluding, Ethiopia and Nigeria, the mean number of districts surveyed in each country was 28 (Figure 2.2).

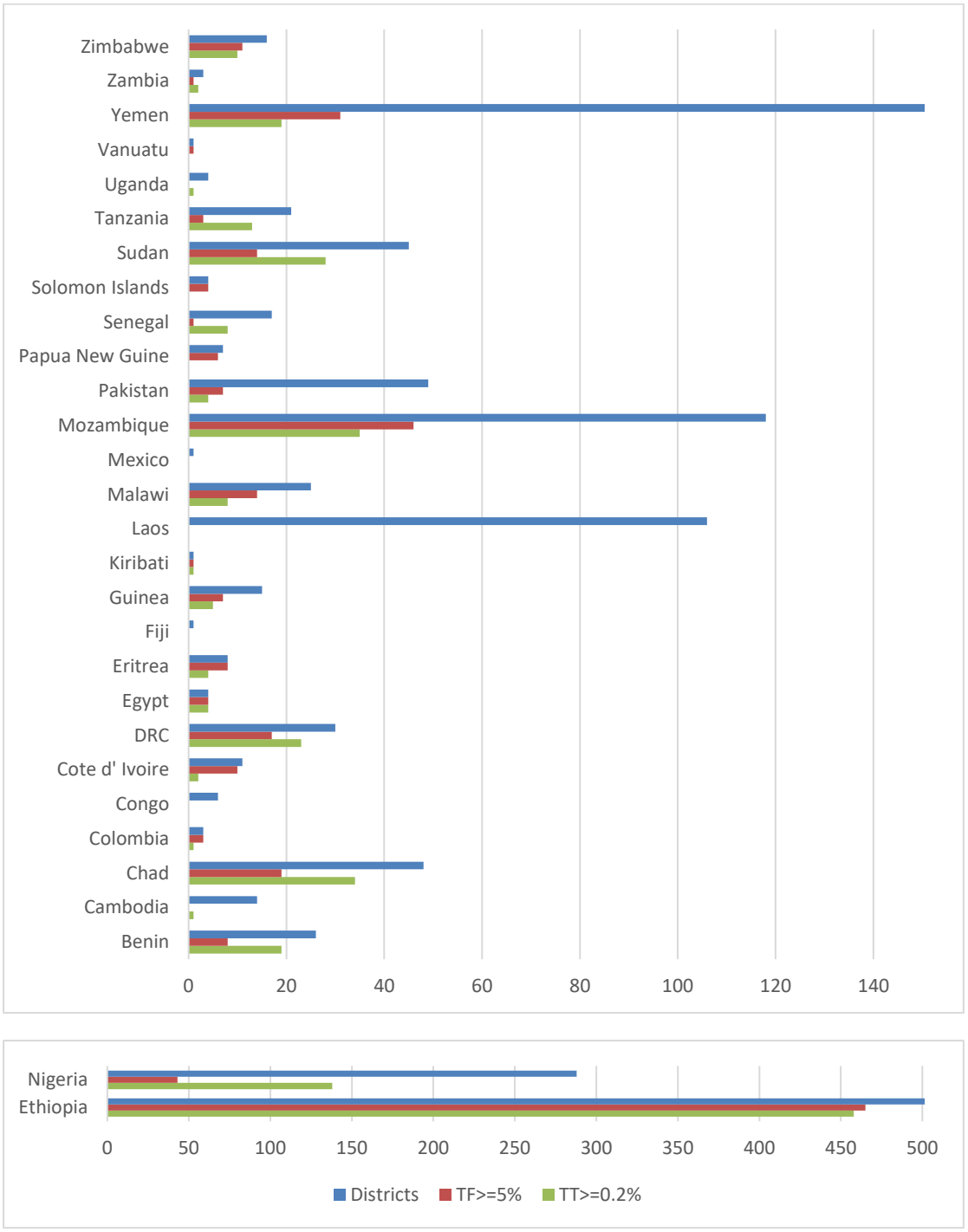


Figure 2.2 Distribution of GTMP surveys, by country and categorized by elimination targets. The blue bar represents the total number of districts surveyed, the red bar represents the number of districts surveyed showing TF above the elimination threshold of 5% and the green bar represents the number of districts surveyed showing TT above the elimination threshold of 0.2% in adults.

In Chapter 5 and Chapter 6, community level data is explored from ten GTMP collaborating countries, namely Benin, Cote d' Ivoire, Democratic Republic of the Congo (DRC), Ethiopia, Guinea, Malawi, Mozambique, Nigeria, Sudan and Uganda; encompassing 15,051 clusters (or communities) within 624 EUs.

To ensure country ownership, a systematic process for accessing the GTMP data was constructed. Study proposals desiring to conduct secondary analyses of GTMP datasets were submitted for review to the "GTMP Data Application Panel". After approved by the panel, data sharing agreements were distributed to relevant ministry of health officials on behalf of the researchers. Once the ministry of health authorized the data sharing agreements, the GTMP released the approved data to the research team. Through this system, countries generously provided the detailed data which I will describe here.

The dataset includes 15,051 communities, of which, 9,590 (63.7%) had zero TT cases, 9,657 (64.3%) had a prevalence of positive TT of <1% and 13,994 (93.0%) had a prevalence of positive TT of <5%. The TT and TF table presented here illustrate the cluster-level distribution of TT and TF within each country (Table 2.2). Table 2.2 clearly shows that Benin, DRC, Ethiopia and Nigeria have the greatest variation in community-level TT prevalence.

Table 2.2 Distribution of community-level TT and TF prevalence by country.

TF Prevalence						
Country	Minimum	First Quartile	Median	Mean	Third Quartile	Maximum
Benin	0.0%	1.5%	3.4%	7.7%	8.9%	65.9%
Cote d'Ivoire	0.0%	3.5%	8.5%	11.0%	15.0%	65.4%
DRC	0.0%	1.7%	6.4%	9.4%	13.8%	86.3%
Ethiopia	0.0%	3.0%	14.8%	21.9%	36.4%	100.0%
Guinea	0.0%	0.0%	2.5%	4.0%	5.8%	24.1%
Mozambique	0.0%	0.0%	0.0%	4.4%	5.7%	75.0%
Malawi	0.0%	0.0%	3.5%	6.2%	8.6%	50.9%
Nigeria	0.0%	0.0%	0.0%	3.0%	3.6%	59.7%
Sudan	0.0%	0.0%	0.0%	4.2%	5.0%	60.7%
Uganda	0.0%	0.0%	0.0%	3.1%	4.4%	19.2%
TT Prevalence						
Country	Minimum	First Quadrant	Median	Mean	Third Quadrant	Maximum
Benin	0.0%	0.0%	0.0%	1.6%	1.9%	52.6%
Cote d'Ivoire	0.0%	0.0%	0.0%	0.2%	0.0%	7.2%
DRC	0.0%	0.0%	0.0%	2.7%	3.2%	75.0%
Ethiopia	0.0%	0.0%	0.0%	1.9%	2.7%	34.5%
Guinea	0.0%	0.0%	0.0%	0.3%	0.0%	6.3%
Mozambique	0.0%	0.0%	0.0%	0.4%	0.0%	12.2%
Malawi	0.0%	0.0%	0.0%	0.5%	0.0%	9.5%
Nigeria	0.0%	0.0%	0.0%	1.4%	1.9%	21.8%
Sudan	0.0%	0.0%	0.0%	1.1%	1.9%	14.3%
Uganda	0.0%	0.0%	0.0%	0.3%	0.0%	3.1%

The GTMP dataset provides a robust resource for trachoma analysis and as a result it is now possible to generate detailed generalizable analyses around trachoma. In the next chapter I will use the GTMP dataset to update the global TT burden estimate, I will then inform the design of a TT specific survey through disentangling the design effect associated with TT and confirm the age distribution of TT cases using multiple country datasets. In Chapter 5, I will use the GTMP dataset to closely examine the community-level relationship between TF and TT whilst accounting for additional covariates. Finally, in Chapter 6, I will identify high TT prevalence clusters, hot spots, and their associated risk factors.

Chapter 3: **Describing the global trichiasis burden**

3.1 Overview

A series of global trachoma scientific meetings resulted in defining the elimination as a public health problem thresholds for trachoma as a prevalence of TF of less than 5% in children aged 1–9 years, and a prevalence of TT of less than one case unknown to the health system per 1,000 total population (0.1%), in each previously-endemic health district [73]. The TT threshold can also (perhaps more appropriately) be stated as a prevalence of TT unknown to the health system of <0.2% in adults aged ≥15-years [162]. Trichiasis is the blinding stage of trachoma, thus appropriate management of individuals with trichiasis is the priority of every trachoma elimination programme. An accurate estimate of the number of persons with TT and their geographical distribution can help to effectively align resources for surgery and other services. Previous work estimated global TT backlogs of 8.2 million (2009) [163] and 7.3 million (2012) [84].

This section describes the process of updating this estimate through (1) obtaining the most recent data for a district; (2) applying age- and sex-standardization to historic survey data, where raw data were available; (3) using historic prevalence estimates and adjusting for bias for districts for which raw data were unavailable; and (4) obtaining expert opinion for districts for which no data were available.

Data used for this study came from several sources. The GTMP outlined in Chapter 2, supported collection of the data described for 29 countries. The health ministries (or equivalent) in the following countries provided data used in this analysis: Benin, Cambodia, Cameroon, Central African Republic, Colombia, Congo, Côte d'Ivoire, Democratic Republic of the Congo, Eritrea, Ethiopia, Guatemala, Guinea, Guinea-

Bissau, Kenya, Lao People's Democratic Republic, Malawi, Mali, Mauritania, Morocco, Mozambique, Nepal, Niger, Nigeria, Pakistan, Papua New Guinea, Senegal, Solomon Islands, South Sudan, Sudan, Togo, Uganda, United Republic of Tanzania, Vanuatu, Yemen, Zambia, Zimbabwe. The study was funded by the World Health Organization (SPHQ14-APW-4886).

I assembled the datasets, conducted the analysis, and wrote the manuscript (pending publication) informed by the information presented in Chapter 3. These results were also presented at the second Global Scientific Meeting on Trachomatous Trichiasis.

3.2 Background

Through high quality surgery, which involves correcting the position of the in-turned eyelid [78] it is possible to reduce the number of people with TT. An accurate estimate of the number of persons with TT (the TT backlog) and their geographical distribution are needed in order to effectively align resources for surgery and other necessary services.

In 2009, Mariotti *et al.* estimated the global TT backlog to be 8.2 million people [163]. This estimate was derived by summarising a combination of published and unpublished information. First, a literature review was performed to identify published prevalence data. Where published data were not available, unpublished data were compiled from the Eleventh (2007) Meeting of the World Health Organization (WHO) Alliance for the Global Elimination of Trachoma by 2020 [164]. Where information was still missing, unpublished reports were provided by health ministries. Finally, if none of these resources were available, data were extrapolated from a proxy country believed to have common epidemiological conditions and demographic structure.

There are many uncertainties inherent in the 2009 estimate [163]. First, where PBPS data were available, the results were not systematically standardised by age and sex. This is problematic because I suspect that women and the elderly are more likely than men and younger adults, to be at home at the time that a house-to-house survey team calls, and to have TT. Second, in countries for which data were available, survey coverage was generally far from complete. Though not an invariable rule, surveys to estimate the prevalence of neglected diseases have a tendency to be done first in areas of high prevalence; for the 2009 estimate, if any data were available for a particular country, the TT prevalence figure from it was applied across the yet-to-be-mapped suspected-endemic population. Third, the use of proxy countries is extremely subjective.

In 2012, WHO collected and collated provisional 2011 country reports, and published an updated figure for the global TT backlog. The total given was 7.3 million people, but the methodologies used in each country to generate national backlog figures were not described [84]; they are likely to have been highly heterogeneous.

In July 2012, the DFID provided funding for the GTMP which sought to systematically complete the global trachoma map using standardised techniques for both collecting and analysing survey data [71]. GTMP measured trachoma prevalence using gold standard PBPSs conducted at district level. People of all ages living in selected households of 1,542 districts across 29 countries were examined for trachoma, resulting in the examination of 2.6 million people [18]. GTMP analyses included standardising trichiasis prevalence estimates against national population pyramids in an attempt to partially account for systematic differences in age and sex of those examined.

As a result of the GTMP, there are now high-quality PBPS data available for most suspected-endemic areas that were previously unsurveyed. The availability of these data has catalysed the current attempt to generate a new estimate of the global trichiasis backlog. The aim of this chapter is to describe the methodology used to update the backlog estimate and provide the national and global results, highlighting the limitations involved in this type of analysis.

3.3 Methodology

3.3.1 Assembling the datasets

I updated previous global estimates using the best available data, according to the following hierarchy. First, where GTMP data [71] were available, I used the age- and sex-standardised trichiasis prevalence estimates for adults in each evaluation unit. In August 2014, GTMP added examination for the presence or absence of TS [19] for all eyes determined to have trichiasis; prior to this, and for all non-GTMP data, trichiasis prevalence estimates used must be considered to be estimates of "all trichiasis" rather than "trachomatous trichiasis".

Second, where PBPSs had been done without the support of the GTMP, I requested raw survey data from national health ministries. Where those data were provided by 1 March 2016, I applied the same age- and sex- standardization as was used in the GTMP [71].

Third, where PBPSs had been done but raw data were not available, prevalence estimates (whether standardised for age and sex or not) were obtained from country programmes.

Fourth, if prevalence estimates were not available, I reviewed previous estimates through desk reviews and communications with country programmes. If there was adequate evidence to revise the estimates, new estimates were used; otherwise, the 2009 estimates [163] were retained.

The population data used in this analysis were derived from the UN population division (UNdata) [165] and www.worldpop.org [166]. As trachoma is a disease primarily affecting rural populations [21], rural population pyramids were obtained from UNdata. Microsoft Excel (2007) was used to organize the datasets into 5-year age bands stratified by sex for each country. The percentage of the population within each stratum was estimated from the bands. District level populations were derived from www.worldpop.org data [166] using the zonal statistics tool in ArcGIS 10.3 [167]. A sensitivity analysis was also performed, comparing the www.worldpop.org population estimates, and district population estimates provided by national trachoma elimination programmes. The mean ratio between the national programme estimates and www.worldpop.org was 0.97. Because of this close correlation, and for the purposes of standardising our methods, the www.worldpop.org data were used throughout this analysis.

3.3.2 Analysis

The statistical software package R [168] was used to perform age- and sex-standardisation. First, the crude prevalence was calculated for each cluster. Second, the prevalence was standardised for each cluster by weighting the proportion of each sex-specific five-year age band observed to have trichiasis by the proportion of the adults aged ≥15-years expected to have that age and sex in that district, if available, or (if not

available), nation-wide. Third, the un-weighted arithmetic mean of the standardised cluster-level trichiasis proportions was taken as the district-level prevalence. Once the initial standardization was completed, bootstrapping was undertaken on the dataset for each individual evaluation unit to derive 95% confidence intervals. For an evaluation unit surveyed by examining individuals in *n* clusters, this involved bootstrap resampling (with replacement) of *n* clusters, over 10,000 replications. The R code is provided in Appendix A.

The standardised prevalence estimates were then multiplied by the corresponding evaluation unit-level adults aged ≥15-years (projected to 2016) to provide an estimate of the number of persons with trichiasis.

A sensitivity analysis was performed using ratios from proxy countries in situations where communications with country programmes estimated a value of zero (Appendix B). Proxy countries were chosen based on similar geography, hygiene and sanitation situation. Hygiene and sanitation data were derived from the WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP) [169].

3.4 Results

Data from 1,355 districts in 31 countries were age- and sex- standardised and contribute an estimated 676 thousand cases to the global total. On average, adjusting prevalence estimates by age and sex reduced raw district-level estimates by a factor of 0.45 (range 0.03–2.28). Where it was possible to adjust the datasets by age and sex, the country level backlog reduced by a mean factor of 0.56 (range 0.22–0.94). Estimates from PBPSs in 398 districts could not be standardised by age and sex because the original datasets were unavailable; the unadjusted backlogs in these districts totalled 781

thousand cases. I adjusted these estimates by multiplying them across the board by 0.45, yielding 411 thousand cases. Based on new expert opinion, it was determined that eight countries no longer have a trichiasis backlog. For six countries, local expert opinion generated revised, non-zero estimates. For one country, other published estimates [108] were used. Finally, there were seven countries for which the 2009 estimates were retained. Overall, data included in these analyses had been collected from 2000 to 2016 (Appendix C).

Through the inclusion of GTMP data, adjusting older datasets by age and sex, and canvassing current local expert opinion, the global backlog estimate reduces from 8.2 million [163] (or 7.3 million [84]) to 2.8 million (Table 3.1). The updated estimate reduced to zero cases in twelve countries, namely; Angola, Botswana, Burundi, Congo, The Gambia, Namibia, Iran, Iraq, Morocco, Mexico, Fiji and Nauru. Updated survey data demonstrated a zero prevalence in Congo, The Gambia, Morocco and Fiji. Whilst, ministries of health in Angola, Botswana, Burundi, Namibia, Iran, Iraq, Mexico and Nauru reported no evidence of trichiasis in 2016. The sensitivity analysis performed in the countries where the ministry of health reported no evidence of trichiasis in 2016, suggests these zero estimates may contribute to a global underestimate of 13,285.

In the estimates, data collected in GTMP surveys set up prior to August 2014 include all trichiasis, irrespective of the presence or absence of trachomatous conjunctival scarring, whilst data collected in GTMP surveys set up after August 2014 provide estimates of trichiasis only in eyes that also demonstrated trachomatous conjunctival scarring (or had an eyelid that could not be everted, due to presumed dense scar), thereby including only those cases of trichiasis attributable to trachoma. In both scenarios the data represent both “managed” and “unmanaged” trichiasis, irrespective

of whether individuals have previously been offered corrective surgery or epilation [73].

Table 3.1 Estimated region-level trichiasis backlogs, 2016, with comparisons to the corresponding estimates for 2009

WHO Region	**2009 Estimate**	**2016 Estimate**				Proportion change
		Retained previous estimate or expert opinion	Unstandardized (reduced by a factor of 0.45)	Standardised (95% CI)	**Total estimate**	
African	**3,846,500**	*915,274*	*247,570*	*547,151 (283,052-913,841)*	**1,709,995**	0.45
Americas	**58,050**	*58,000*	*566*	*48 (0-123)*	**58,614**	1.01
Eastern Mediterranean	**660,000**	*126,875*	*66,171*	*106,003 (39,337-196,230)*	**299,049**	0.52
South-East Asia	**485,000**	*443,000*	*25,240*	*3,252 (462-7,323)*	**471,492**	0.97
Western Pacific	**2,630,000**	*166,900*	*71,288*	*5,947 (810-13,027)*	**244,135**	0.09
Total	**8,248,050**	*1,710,049*	*410,835*	*662,403 (323,661-1,130,5430)*	**2,783,285**	0.35

3.5 Discussion

Trichiasis remains a significant public health problem in many countries, with a global backlog estimated for 2016 at 2.8 million people, 61% of whom live in sub-Saharan Africa. Whilst there were methodological challenges in generating the 2009 estimate, it was the best estimate that could be made with the information available at the time. The considerable reduction from that estimate to the one generated here is likely to be the result of a combination of factors. First, there are now more, and better data derived from rigorous surveys. Second, in many countries, there has been an impressive programmatic scale up to manage TT, conducted by a complex network of governments and their partners. Third, there is likely to be an effect on the incidence of TT from the intensive efforts to reduce active trachoma prevalence in many contexts; such efforts have been ramping up in endemic countries since the World Health Assembly's 1998 commitment to global elimination of trachoma [2]. Teasing out the relative contribution of each of these factors is not possible at the present time, but regardless of cause, the reduction is welcome news for global health.

An understanding of the backlog and distribution of trichiasis cases is necessary to effectively plan for surgical services and other components of management of individuals with trichiasis. To reduce TT prevalence in each district of each endemic country to <0.2% in adults, which is the defined elimination threshold for TT [162], at least 2.0 million people will need to have their TT appropriately managed. From available data, it is estimated that 56% of people with trichiasis have bilateral trichiasis.

Whilst these calculations lessen the uncertainty around the global backlog estimate, important limitations remain. Expert opinion was used for 13 endemic countries (representing 37% of the estimate), in eight of which the estimate was zero cases. The sensitivity analysis performed on these eight countries suggests that the zero estimates may contribute to a global underestimate of 13,285. Ethiopia, Mali, and Niger have undergone intensive surgical scale up in the time since the most recent round of prevalence surveys. Because of this programmatic output, and a lack of consensus around how to counterbalance backlog reduction with new incident cases and post-surgical recurrence (both of which are inescapable, but presently impossible to quantify) these countries provided results based on expert opinion. Uncertainty is greatest among the nine countries in which previous estimates were retained; these countries accounted for 848 thousand cases of trichiasis (30% of the global estimate) and further investigation is needed. India, which accounts for 443,000 cases of trichiasis (52% of the total retained estimate) is a particular priority.

A second ongoing uncertainty stems from the unavailability of raw data for 398 districts, for which, as a result, age- and sex-standardization was not possible. As noted in the calculations, there is a mean prevalence reduction against the raw prevalence estimate of 0.45, which was applied to the unstandardized districts.

Ongoing lack of trichiasis (and TF) prevalence data for 235 suspected-endemic districts is due to local insecurity, and the global trachoma community stands ready to support national governments to undertake the needed mapping in those populations, when and if security conditions improve to allow safe conduct of fieldwork.

It is recognized [71] that the PBPS methodology described earlier is potentially imprecise in estimating TT against the WHO elimination target of ≤0.2% in adults. Thus, at district level, the uncertainty around the estimates of the TT backlog can be large. At national, regional and global level, I expect that the precision of overall estimates derived from multiple appropriately standardised PBPS datasets, in which the total number of clusters and total number of adults examined is very large, will be reasonably tight. In the next chapter, I will describe a more reliable methodology for estimating the district-level prevalence of TT.

At present, the diagnosis of TT is based on the presence of clinical trichiasis (one or more lashes touching the globe or evidence of recent epilation of in-turned eyelashes) plus residence in a (presumed) trachoma-endemic setting. This definition is by nature somewhat circular and may need review, as it inevitably leads to classification of trichiasis as TT regardless of whether trachomatous conjunctival scarring is present or not. Trichiasis without trachomatous conjunctival scarring may be age-related or due to trauma, distichiasis, epiblepharon or other inflammatory disease [170]; the sight-threatening potential and optimal management strategies for non-trachomatous trichiasis still require further investigation [171].

Assumptions were made when adjusting the data for age and sex. First, I assumed that five-year age bands were accurate. However, it is reasonable to hypothesise that during

a survey, individuals demonstrate a terminal age preference. Second, this analysis assumed that UNdata rural population pyramids used were representative of all districts for which survey data were available. Finally, because many studies have demonstrated a high correlation between trichiasis and increased age [33, 59, 66, 67], I assumed zero trichiasis cases in the population aged 14 years and younger.

These calculations of national and global TT burdens are point-prevalence estimates based on data of differing vintages; they will change as additional surveys, baseline or impact, are undertaken. Other than for Ethiopia, Mali and Niger, the estimates do not consider the number of surgeries done during the period between the most recent survey and this analysis. Nevertheless, estimates of current national TT backlog are essential for countries to appropriately allocate resources for surgical campaigns. Future district-level impact surveys or surveillance strategies in countries with trachoma elimination programmes will provide progressively improved evidence for countries to assess their residual TT burden and, ultimately, validate elimination of trichiasis as a public health problem. Baseline surveys of trachoma are still needed in a handful of countries (e.g., Egypt, Somalia and Central African Republic); these will lead to revisions in national estimates. In some settings, there will be a need to undertake trichiasis-specific surveys in order to assess progress to elimination. Unfortunately, insecurity may continue to limit surveys in a few, but not all, of the settings in which previous estimates were retained. Where possible, surveys should be undertaken. Such data will help drive towards, and demonstrate success in, the global elimination of trachoma as a public health problem.

With an updated estimate of the backlog it is now possible to organize resources to focus on intervention in the high burden areas. However, a survey suitable for

measuring TT prevalence with precision against the WHO elimination target is needed. In the next chapter, I describe the design of a TT survey as well as it's validation. I will then explore the spatial heterogeneity of TF and TT and ultimately evaluate the ability to identify TT hot spots.

Chapter 4: **Design and validation of a trachomatous trichiasis (TT) specific survey**

4.1 Overview

The previous chapter described the global burden of trichiasis estimated using the best data currently available. This information is valuable in aligning resources and planning for intervention campaigns. However, as country programmes approach the elimination targets, the necessity of a strategy to precisely assess TT burden within districts has been highlighted.

The district has typically been identified as the implementation unit for trachoma intervention and mapping for practical reasons. A district is often an administrative boundary comprising of between 100,000 to 250,000 individuals. There is often an organised administrative body associated with a district which can facilitate both MDA and mapping.

The WHO endorses a PBPS for trachoma, which is designed to estimate the prevalence of TF in children aged 1–9 years. The template design is a two stage cluster random sample survey, which uses PPS sampling to select 20-30 clusters; everyone aged over one year living in selected households is examined [22]. TF and TT are evaluated against the criteria set out in the WHO simplified trachoma grading system [19].

However, TT disproportionately affects older people [33, 66, 67], and is a much less common sign than TF, and so the number of adults examined is generally not sufficient for estimating TT prevalence with good precision. These surveys simply accept the loss in precision in estimating TT [71].

In this chapter I present the design of a TT specific survey and validation of the design. The health ministries (or equivalent) in Benin, Malawi and Nigeria provided data used to inform the survey design. The validation exercise was implemented in United Republic of Tanzania, Uganda, Cameroon, and Chad with funding from Hellen Keller International and Sightsavers. I was responsible for the design of the survey, coordinated the validation exercises and conducted the validation analysis. Recommendations generated by this work were presented at the WHO, Eighth NTD-STAG Global Working Group Meeting on Monitoring and Evaluation of NTDs held February 2017. A report of these findings and recommendations was published by WHO with the accompanying protocol made available for use in national programmes in 2017 [172].

4.2 Background

As trichiasis is the morbidity stage of trachoma, appropriate management of afflicted individuals is major priority of all trachoma elimination programmes. Precise data on TT prevalence is therefore essential to help programmes plan surgical services, monitor progress, and assess whether the trichiasis component of trachoma elimination has been successfully achieved.

A number of different approaches have been developed for estimating the burden of trachoma [173]. The low threshold for determining TT elimination triggers the question of whether a survey is the best strategy. In other disease control programmes with small elimination thresholds, such as Leprosy, active case detection is the typical strategy used. Leprosy control programmes follow up with newly diagnosed cases through home studies where family members and close neighbours are also examined [137, 138].

Another technique to consider is patient screening, where persons who seek health services are examined for TT as part of their clinical exam. Both of these strategies are problematic for the task at hand, TT is often found among the poorest and most remote individuals, these isolated individuals are expected to have greater barriers in seeking care [174, 175]. Appropriately powered household surveys would be more likely to find these patients.

Typically, trachoma surveys are conducted at the district-level and so for practicality this survey will be designed similarly. However, during the validation process, oversampling will be conducted in an effort to confirm a sufficient number of communities are included and that the design effect is adequate.

The WHO endorses a PBPS for trachoma [176], which is primarily designed to estimate the prevalence of TF in children aged 1–9 years [22]. This approach to mapping trachoma is cost effective and practical for a public health intervention.

Until now, the prevalence of TF (in children aged 1–9 years) and the prevalence of TT (in adults aged ≥15-years) have usually been measured at the same time. However, as programmes evolve, there are three broad scenarios where a TT specific survey may be necessary or desirable. In two of these scenarios the PBPS results may indicate either no need for intervention for TF and in-turn no need for re-evaluation of TF, or a long span of intervention for TF and subsequently a delay for re-evaluation of TT. In these scenarios, the programme may wish to re-estimate the TT prevalence without TF (Figure 4.1). The third scenario is if a survey at any stage of the programme estimated the prevalence of TT with a questionable methodological approach, the programme may wish to conduct a TT-specific survey.

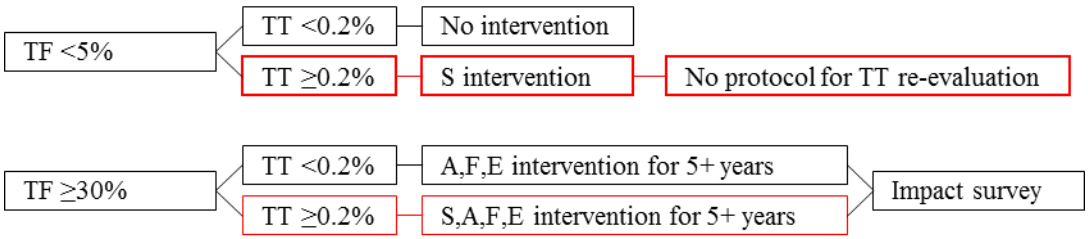


Figure 4.1 Flow diagram illustrating the time points when a TT specific survey may be needed

I set out to develop a methodology for undertaking a TT specific population based prevalence survey. This approach was maintained not only because programmes are familiar and comfortable with it, but also because it will provide robust estimates that enabling programmes to calculate their burden of disease. In developing the methodology, I felt it important to focus on an appropriate balance between feasibility and precision. In the context of a public health initiative it is necessary for a survey to be realistically costed, whilst maintaining sufficient precision. I explored various elements of the design through simulations with existing trachoma prevalence survey data, then field-tested a draft methodology in four trachoma-endemic districts at different stages of progress towards elimination in four different countries.

4.3 Informing a TT specific survey design

4.3.1 Data sources

The population data used in this analysis were derived from the United Nations Statistics Division [165] and www.worldpop.org [166]. Microsoft Excel (2007) was used to organize the datasets into 5-year age bands stratified by sex for each country. The percentage of the population within each stratum was estimated from the bands. The district level populations were derived from www.worldpop.org data [166] using

the zonal statistics tool in ArcGIS 10.3 [167]. A sensitivity analysis was performed comparing the www.worldpop.org population estimates, and district population estimates provided by national programmes. The mean ratio between the national programme estimates and www.worldpop.org was 0.97.

I used existing datasets to understand two key components influencing the optimal design of a TT-only survey: (1) the age distribution of TT and (2) the design effect (DE) of the TT estimate. Health ministries of Benin, Malawi and Nigeria provided datasets from 271 surveys undertaken from 2012–2016 with the support of the GTMP [71, 177-184]. These countries were selected based on the availability of high quality data and diversity of the known trichiasis situation (Table 4.1).

Table 4.1 Summary of survey data used in analysis

Country ([State], where applicable)		No. surveys	Range of trichiasis prevalences
Benin		27	0.0 – 1.9
Malawi		24	0.0 – 0.6
Nigeria	[Bauchi]	20	0.1 – 3.3
Nigeria	[Benue]	23	0.0 – 0.4
Nigeria	[FCT]	6	0.0 – 0.3
Nigeria	[Gombe]	11	0.5 – 3.9
Nigeria	[Jigawa]	4	1.9 – 3.1
Nigeria	[Kaduna]	23	0.0 – 0.8
Nigeria	[Kano]	44	0.0 – 2.9
Nigeria	[Katsina]	34	0.0 – 3.6
Nigeria	[Kebbi]	2	0.4 – 1.8
Nigeria	[Kogi]	4	0.0 – 0.0
Nigeria	[Kwara]	8	0.0 – 0.2
Nigeria	[Niger]	25	0.0 – 0.4
Nigeria	[Sokoto]	3	0.3 – 1.0
Nigeria	[Taraba]	13	0.0 – 0.8

Each one was a PBPS [185] using a methodology described in detail elsewhere [71]. These surveys did not collect information on presence or absence of TS of the conjunctiva [19] in eyes with trichiasis [186]. The surveys were conducted prior to the addition of this information within the GTMP's training and fieldwork systems [187, 188]: these datasets therefore include data on all trichiasis, irrespective of the presence

or absence of TS, and it is not possible to make presumptions as to the aetiology of the cases.

4.3.2 Analysis

4.3.2.1 Age Structure

National-level population pyramids were used. To ensure that national-level age distributions were representative of local age distributions, the national trachoma programme district level population datasets were compared to district level datasets extrapolated from www.worldpop.org (Figure 4.2). The national distributions track closely to the local distributions and so the former was used in the analyses.

Figure 4.2 Comparison of national and local population distributions.

4.3.2.2 Age prevalence curves

Using R statistical software, trichiasis prevalence by age and sex was calculated. The R script first groups raw data by cluster, then age and sex. The sum of residents examined along with the number of cases per group is determined, then each group is weighted by the expected proportion of residents with that age and sex. The unadjusted proportion is calculated per group by dividing cases over total examined. The proportion is then multiplied by the assigned weight. The sum of the weighted proportion for groups within each cluster is calculated as the cluster-level age- and sex-adjusted proportion. Finally, the mean of the cluster-level result is taken as the adjusted prevalence for the EU.

When exploring the requirements of a TT specific survey powered to estimate TT prevalence in different age ranges, I assumed that programme interventions [189] would reduce the prevalence of TT uniformly across all age groups; adults aged ≥40-years constitute 34% of the adults aged ≥15-years; and 85% of TT cases among adults aged ≥15-years occur in adults aged ≥40-years. Using these assumptions, a prevalence of 0.2% in those adults aged ≥15-years would correspond to a prevalence of 0.5% in those adults aged ≥40-years.

4.3.2.3 Design Effect

The DE estimates of variance arising from the cluster-sampled design and was calculated for each EU as $DE = 1 + m\alpha^2\mu$, where m is cluster size, α is the standard deviation over the mean and μ is the mean prevalence.

4.3.2.4 Sample size

Sample size was calculated as $n = DE \times \left(\frac{z_{\frac{a}{2}}^2 \times p(1-p)}{e^2} \right)$ where $z_{\frac{a}{2}}$ = the standard normal deviate corresponding to 95% confidence intervals, p = the expected prevalence, and e = the desired absolute precision, expressed as half the width of the desired confidence interval [190].

The number of clusters needed was calculated as $c = \frac{n}{(h \times a)}$ where h = the number of households that can be seen by one team in one day and a is the expected number of adult residents in each house.

4.3.3 Results

The EU-level prevalence curves and percent of cases by age were plotted to visualise the heterogeneity of trichiasis curves within a country. The mean of the EU curves was taken as the country curve (Figure 4.3). Variation in age prevalence is observed. However, there does appear to be a consistent increase in prevalence around age 30 - 40 years. Across the three countries, on average, trichiasis prevalence increases from 0.3% (adults aged 25-29 years) to 0.4% (adults aged 30-34 years), to 0.5% (adults aged 35-39 years) to 0.9% (adults aged 40-44 years). Trichiasis prevalence of 100% is reached among the 65-69, 70-74, and 75+ age groups in four EUs in Nigeria. However, in these scenarios only one individual was enrolled in the age group and so the denominator is one.

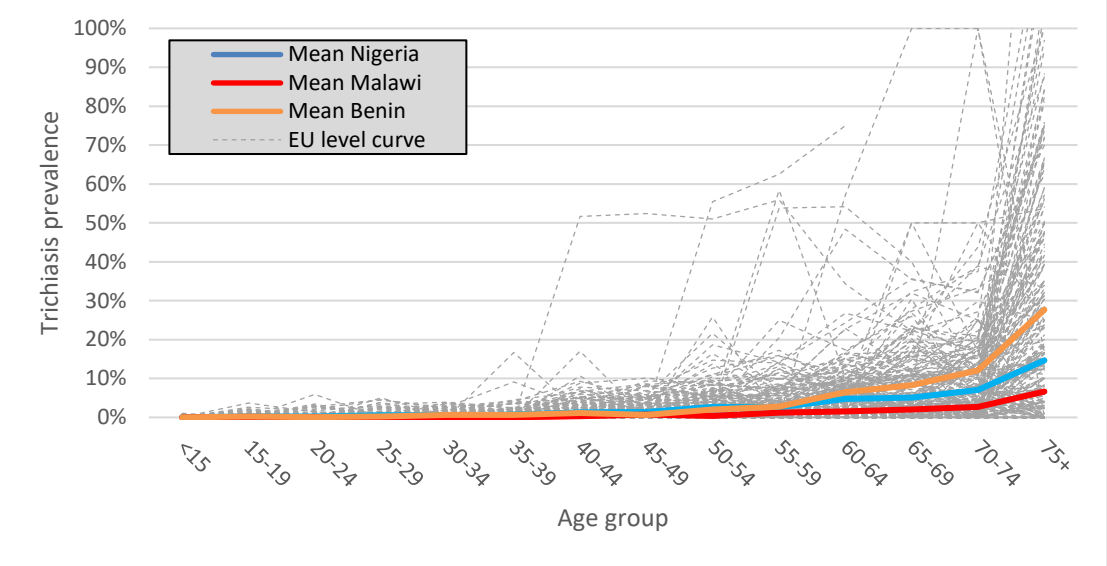


Figure 4.3 Trichiasis prevalence by age group. The solid lines represent the mean prevalence observed by age group within each country and the dotted lines represent the mean prevalence observed by age group within each EU. In total 271 EUs are represented. . A gap in the dotted line can be observed in one EU. This is a result of no individuals identifying themselves as being between the ages of 65 and 69 during data collection in this EU.

Further exploration of prevalence of trichiasis cases in relation to percentage of population within the age groups shows over 90% of cases are found in the adults aged ≥30-years and 86% in the adults aged ≥40-years (Table 4.2), with moderate variation between countries.

Table 4.2 Proportion of trichiasis cases within each age group

	≥15 years	≥30 years	≥40 years
Malawi	100%	92%	89%
Benin	99%	95%	85%
Nigeria	97%	92%	83%

The DE associated with the different surveys ranged from 1.1 to 5.1, and 92% of the surveys had a DE of 2 or lower (Table 4.3, Figure 4.4). Ordered from smallest to largest, the 75th centile of DEs in the 271 datasets was 1.47. The 1.47 value was carried forward in the sample size calculation for validation. Whilst it is reasonable to argue that a larger DE should be used in determining the sample size, feasibility and practicality must also

be considered. The validation exercises described in the next section demonstrate that a larger DE is not necessary.

Table 4.3 Distribution of design effects for trichiasis (Benin, Malawi and Nigeria), Global Trachoma Mapping Project, 2012–2016

DE interval	Cumulative Percent
1	23.85%
1.0-1.5	76.15%
1.0-2.0	91.74%
1.0-2.5	97.71%
1.0-3.0	98.62%
1.0-3.5	99.08%
1.0-4.0	99.54%
1.0-4.5	99.54%
1.0-5.1	100.00%

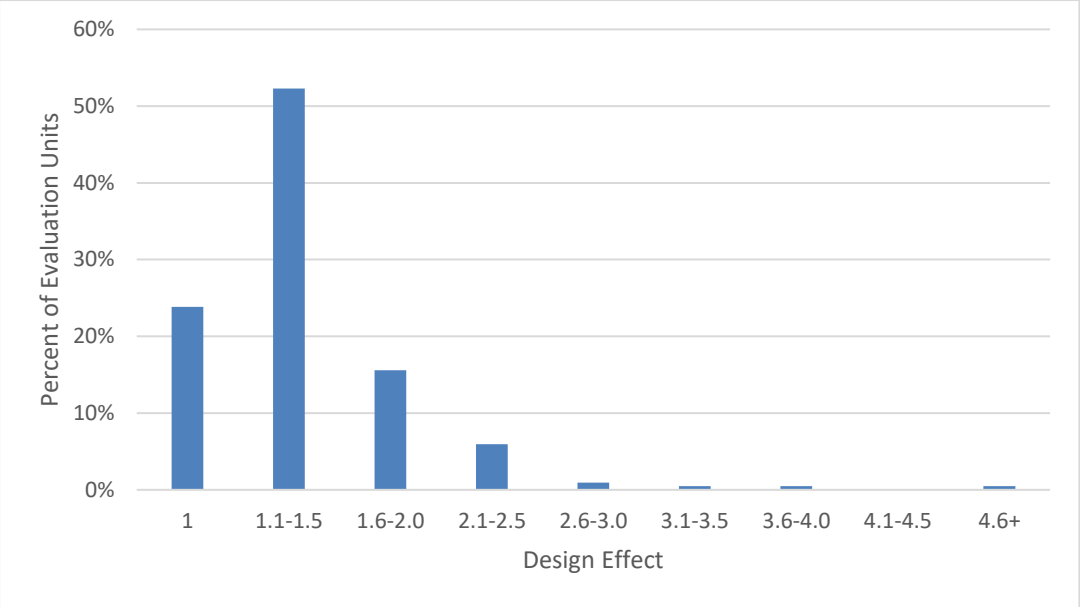


Figure 4.4 Frequency of design effect (DE) found among included datasets

To estimate an expected prevalence of 0.2% with an absolute precision of ±0.20%, the sample size would be: $n = DE \times \left(\frac{z^2 \times p(1-p)}{e^2}\right)$ = $n = 1.47 \times \left(\frac{1.96^2 \times 0.002(1-0.002)}{0.002^2}\right)$.

where z = the standard normal deviate corresponding to 95% confidence intervals, p = the expected prevalence, and e = the desired absolute precision, expressed as half the

width of the desired confidence interval. This gives a sample size of 2,818 adults aged 15 years and older.

On this basis, I calculated a range of sample sizes, displayed in Table 4.4. The sample size required decreases as the age group widens. This is because the variance increases as the expected proportion increases towards 50%, then declines again beyond 50%. If the required absolute precision is held constant, therefore, a larger sample size is needed the closer the expected prevalence is to 50%, to allow the signal to be discerned beyond the noise.

Table 4.4 Sample sizes required to estimate the elimination threshold prevalence of trachomatous trichiasis in different age groups for different precisions, design effect=1.47

Age group sampled	Expected prevalence (%)	Absolute precision			
		±0.15%	±0.20%	±0.25%	±0. 50%
adults aged ≥15-years	0.2	5010	2818	1803	451
adults aged ≥40-years	0.5	12487	7024	4496	1124

4.4 Validating a TT specific protocol

4.4.1 Protocol methods

In 2016, four field-based district-level surveys were executed to test the validity of this proposed design. Four districts (Am-Timan, Chad; Budaka, Uganda; Monduli, United Republic of Tanzania; and Touboro, Cameroon) were surveyed. Each district was in a different stage of progress towards trachoma elimination as a public health problem.

Protocols (Appendix D) were approved by the Cameroon Ministry of Public Health (18 July 2016); Chad Ministry of Health (002/PR/PM/MESRS/SG/CNBT/2014); Uganda Ministry of Health (HS 2012); National Institute for Medical Research, United Republic of Tanzania (NIMR/HQ/R.8a/Vol.IX/2085); and the Research Ethics Committee of the London School of Hygiene & Tropical Medicine (10360).

In the 491 the GTMP survey datasets from Benin, Malawi and Nigeria, there was a mean of 3.0 (survey-level range in means 1.4–6.1) people aged ≥ 15 years per selected household; a mean of 2.3 (1.2–4.6) people aged ≥ 30 years per selected household; and a mean of 1.5 (1.0–2.2) people aged ≥ 40 years per selected household. If 30 households are sampled per cluster (as was often done within the Global Trachoma Mapping Project), then 32 clusters would be needed to include 2,818 residents aged ≥ 15 years, ignoring non-response. The validation exercise involved increasing the number of clusters sampled to 60 to ensure enough data for computer simulations.

The validation exercises involved a two-stage sampling of one evaluation unit from each country. In Cameroon, Chad and Uganda all individuals living in the selected households in a random selection of half the selected clusters were examined for signs of TT based on the WHO Simplified Trachoma Grading Scheme and individuals aged 40 years and older were examined in the remaining clusters. In Tanzania, all individuals living in the selected households in all selected clusters were examined for signs of TT based on the WHO Simplified Trachoma Grading Scheme.

When TT was found in an eye, the eyelid of that eye was everted and inspected for TS based on the WHO Simplified Trachoma Grading Scheme. Data was collected and stored using the LINKS electronic data collection system.

4.4.1.1 Scarring and lower lid trichiasis

The second Global Scientific Meeting on Trachomatous Trichiasis was held in November 2015 [171]. A group of experts discussed the optimal diagnostic method for TT in a survey setting. Consensus was not reached for a change to the definition of TT to require trichiasis (or evidence of recent epilation of in-turned eyelashes) AND either

(1) the presence of TS in the same eye, or (2) inability on the part of the grader to evert the eyelid to examine the conjunctiva; rather, the meeting concluded that collection of data on TS should continue, with the question to be revisited at a later date. Trichiasis without TS could be age-related or due to trauma, distichiasis, epiblepharon or other inflammatory disease [170]. The meeting further recommended including both upper and lower lid trichiasis in the survey. This information was included in the TT specific survey validation protocol (Appendix D).

4.4.1.2 Training

To ensure high quality and consistent diagnostic practices, a standardised training system was developed by a consortium of experts, using as templates the GTMP training schedule [187]. The presentation slides and manual can be accessed at tropicaldata.knowledgeowl.com.

The manual highlights the need to use a variety of different teaching methods, including discussion with the assistance of PowerPoint slides, role play, practical exercise and evaluation of students. This training method focuses on involving the students in the discussion and demonstrations rather than lecturing.

Students are taught how to examine eyelids by practicing on one another, this is especially important as it reminds the student to be gentle when examining eyelids in the field. The training included an objective structured clinical exam (OSCE) to provide an objective measure for ensuring quality standardised grading.

4.4.1.3 Sample selection

Along with preparing the survey team for the examination of patients, the training period was also used as an opportunity to prepare the logistics of the survey work. The evaluation units to be surveyed were pre-determined and will be discussed in detail in the following sections.

In order to generate data for subsequent simulations, oversampling was undertaken in each survey: nearly two times the calculated number of clusters, $2 \times \frac{n}{(h \times a)} = 64$. For practicality in the field the teams sampled 60 clusters per district.

The first and second stage clusters within the evaluation unit were defined using the following method (1) clusters (villages or communities) were selected based on population proportion to size of 60 clusters using systematic random sampling procedures (2) 30 households within each selected cluster were randomly selected to be included in the survey.

4.4.1.4 Data Management

The exercise was carried out using LINKS [191], the electronic data collection system used for the GTMP. LINKS is a smart-phone- and cloud-based system for data collection and reporting which was used in 29 countries for the GTMP, and an additional six countries for surveying other NTDs. Best practices for data management were used, which included data managers independent of the national programme, regular calculation of descriptive statistics and generation of point maps showing cluster locations during the data collection process. Data were stored on a secure server which was backed up hourly. Age- and sex- adjusted prevalence calculations were

provided to the appropriate MOH for review and approval. Upon MOH approval, national programmes shared the results with WHO to facilitate global monitoring of progress and assistance with alignment of resources for intervention with country-specific needs.

4.4.2 Implementing the validation exercise

The four validation surveys commenced in 2016. The characteristics of these districts are summarized in Table 4.5.

.

Table 4.5 Characteristics of four districts involved in the trachomatous trichiasis (TT)-specific survey validation exercise, 2016.

District	Population	Proportion of adults within the ≥15-years population who are aged ≥40-years (%)	Baseline TT prevalence estimate in those adults aged ≥15-years (%) [year of survey]	Baseline TF prevalence estimate in children aged 1–9 years (%) [year of survey; year MDA commenced]	Next TF prevalence estimate due (year)	Rationale for conducting a TT survey
Am-Timan, Chad	233,447	30	6.2 [2002]	26.9 [2002, 2014]	2017	The survey estimated the prevalence of TT with a questionable methodological approach and at the region-level. The programme wished to conduct a TT-only survey.[1]
Budaka, Uganda	192,853	28	3.1 [2012]	2.2 [2012, not indicated]	Not indicated	At baseline the survey estimated prevalence of TF in children <5% and of TT in adults ≥0.2%, an impact survey to measure TF prevalence is not indicated; after interventions, a TT-only survey to re-estimate the TT prevalence is indicated.
Monduli, Tanzania	174,482	34	5.5 [2004]	57.6 [2004, 2015]	2018	At baseline the survey estimated prevalence of TF in children ≥30% and of TT in adults ≥0.2%, at least 5 years of A, F and E interventions are recommended before an impact survey to again measure the TF prevalence. During this time, the programme wished to undertake a TT-only survey to assess progress in addressing the TT backlog, facilitating adjustments in delivery of S interventions, if needed.
Touboro, Cameroon	287,087	35	0.5 [2011]	3.0 [2011, not indicated]	Not indicated	At baseline the survey estimated prevalence of TF in children <5% and of TT in adults ≥0.2%, an impact survey to measure TF prevalence is not indicated; after interventions, a TT-only survey to re-estimate the TT prevalence is indicated.

[1] Baseline survey conducted at region-level.

4.4.2.1 Am-Timan, Chad

Am-Timan District (population 233,447) is in Salamat Region of Chad. Adults aged ≥40-years make up 30% of adult aged ≥15-years [165]. A pre-intervention (baseline) region level survey conducted in 2002 demonstrated a TF prevalence of 26.9% and TT prevalence (in adults aged ≥15-years) of 6.2% (Figure 4.5). Since disaggregating the region level results dramatically reduces the precision of the district level estimates, a new survey was needed.

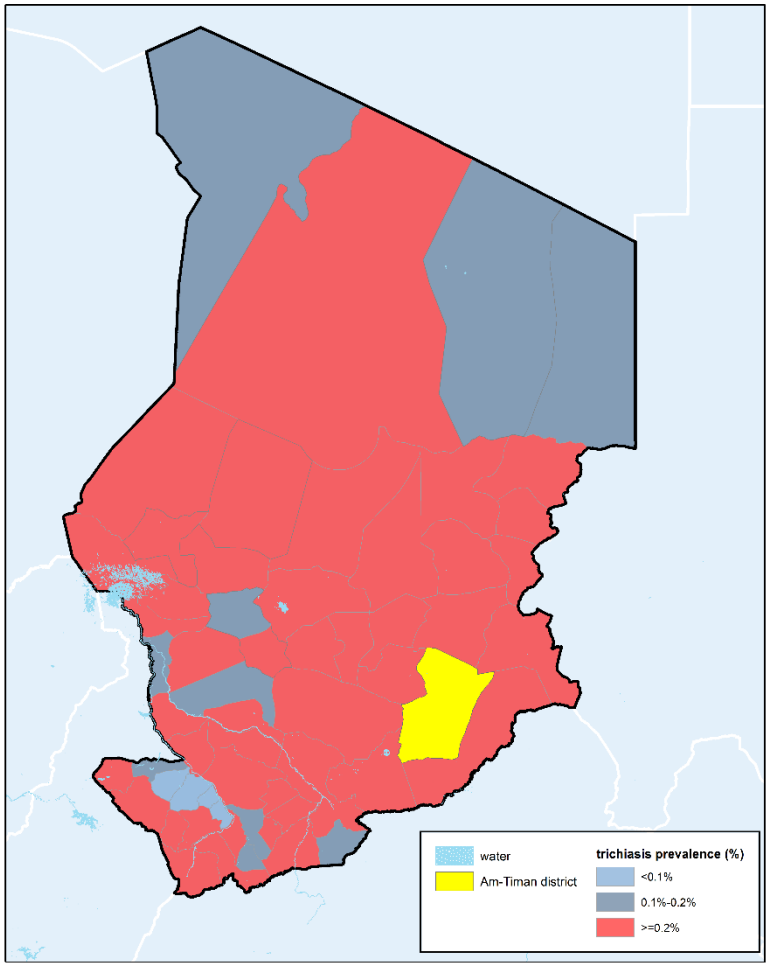


Figure 4.5 Location of Am-Timan district in Chad and the surrounding district level trichiasis prevalences from the most recently available surveys [17, 18]

4.4.2.2 Budaka, Uganda

Budaka District (population 192,853) is in Eastern Region of Uganda. Adults aged ≥40-years make up 28% of adults aged ≥15-years [165]. A pre-intervention (baseline) survey conducted in 2012 demonstrated a TF prevalence of 2.23% and TT prevalence (in adults aged ≥15-years) of 3.1% (Figure 4.6). Since the district was below the WHO recommended threshold for intervention with the **A**, **F** and **E** components of **SAFE**; no impact survey was planned. However, the TT prevalence was above the WHO threshold for declaring trachoma elimination as a public health problem and it was therefore important to re-evaluate the status of TT in this district.

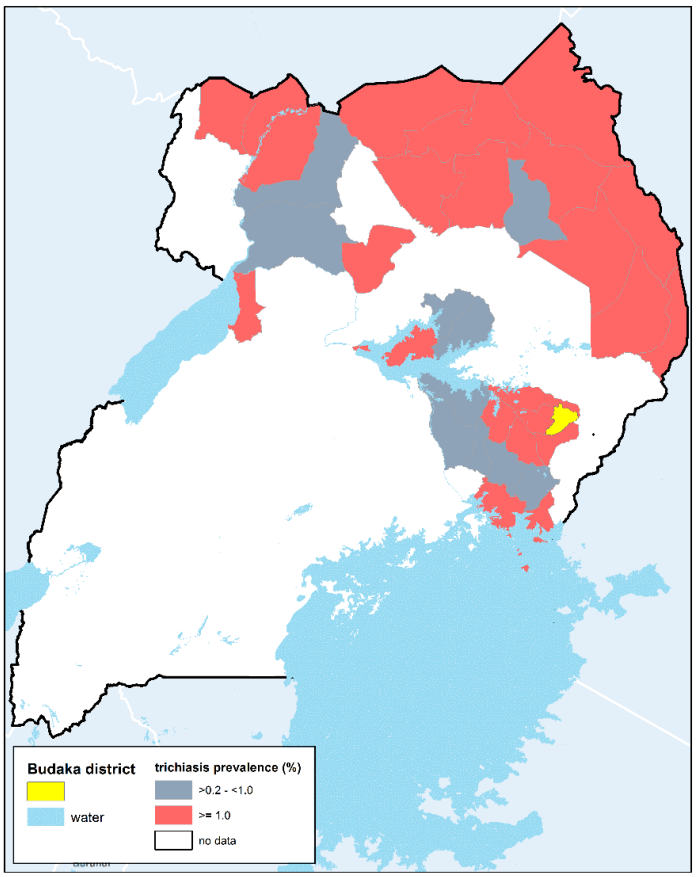


Figure 4.6 Location of Budaka district in Uganda and the surrounding district level trichiasis prevalences from the most recently available surveys [192, 193]

4.4.2.3 Monduli, Tanzania

Monduli District (population 174,482) is in Arusha Region of Tanzania. Adults aged ≥40-years make up 34% of adults aged ≥15-years here [165]. Pre-intervention surveys conducted in 2004 demonstrated a TF prevalence of 57.6% and TT prevalence (in adults aged ≥15-years) of 5.5% [63] (Figure 4.7). Systematic mass drug administration (MDA) was started in 2015 and there are plans to conduct an impact survey in 2018. Because 12 years have passed since the last TT evaluation, it was felt necessary to re-evaluate TT now rather than wait for the 2018 impact survey. An updated TT prevalence estimate is needed for resources to be appropriately aligned in Monduli.

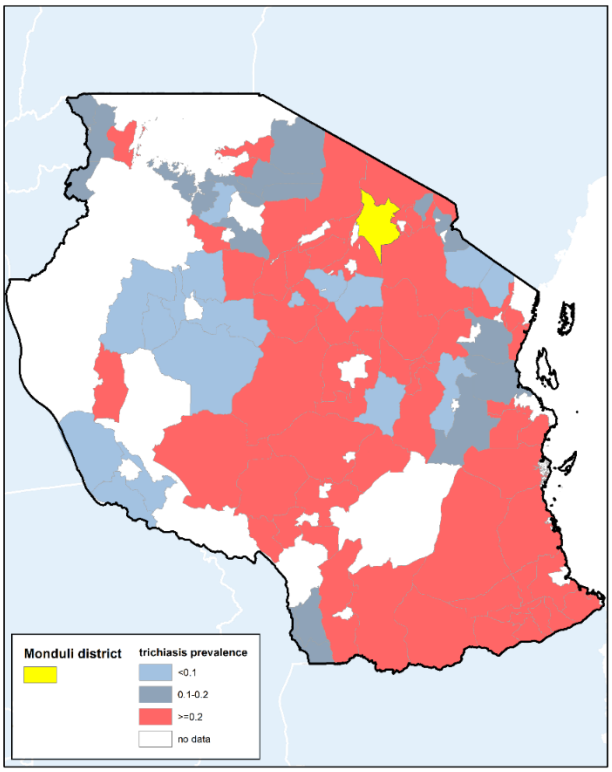


Figure 4.7 Location of Monduli district in Tanzania and the surrounding district level trichiasis prevalences from the most recently available surveys [19, 20]

4.4.2.4 Touboro, Cameroon

Touboro District (population 287,087) is in Nord Region of Cameroon. Adults aged ≥40-years make up 35% of adults aged ≥15-years [165]. A pre-intervention (baseline) survey conducted in 2011 demonstrated a TF prevalence of 3.0% and TT prevalence (in adults aged ≥15-years) of 0.5% (Figure 4.8). Since the district was below the WHO recommended threshold for MDA intervention no impact survey was planned. However, the TT prevalence was above the WHO threshold for declaring trachoma elimination as a public health problem and it was important to re-evaluate the status of TT in this district.

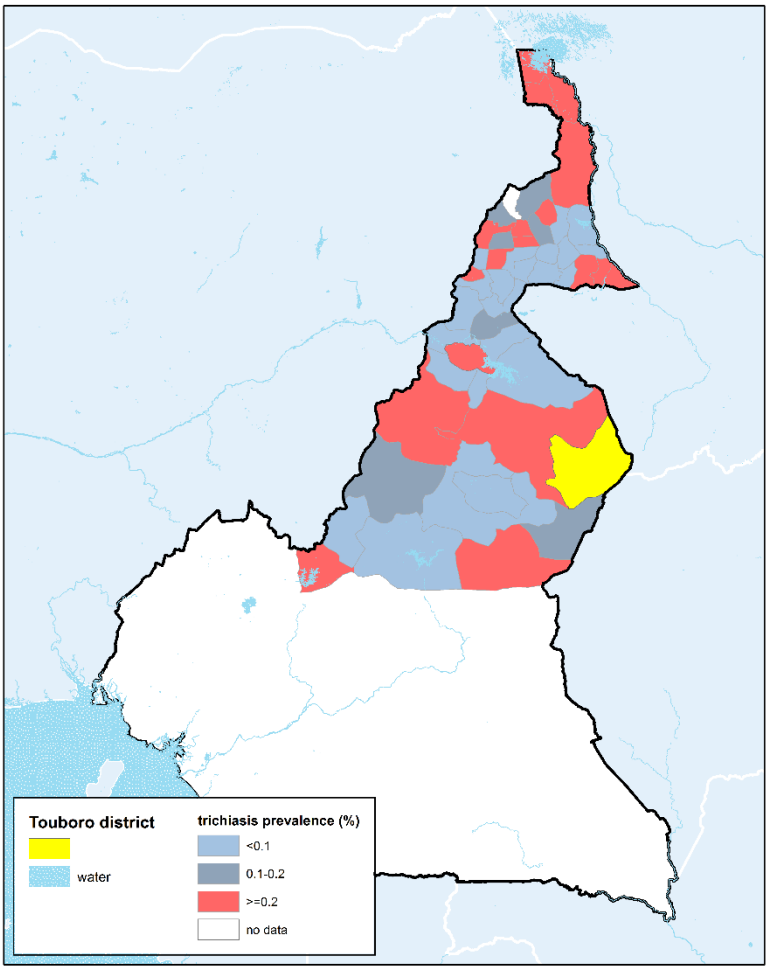


Figure 4.8 Location of Touboro district in Cameroon and the surrounding district level trichiasis prevalences from the most recently available surveys [17, 18]

4.4.2.5 Field work

In Monduli, all consenting individuals who were aged ≥1-year and living in the selected households were included in the survey. For Budaka, Touboro, and Am-Timan, in half of the clusters; eligibility was the same as in Monduli, and in the other half, all consenting individuals who were adults aged ≥40-years and living in the selected households were included in the survey. Field work began immediately following the three-day training. Monduli deployed 12 graders who also recorded the data. The survey teams in Budaka, Touboro and Am Timan were composed of a grader and a designated recorder (six graders and six recorders, five graders and five recorders, four graders and four recorders respectively).

4.4.3 Analysis methods

Analyses were conducted using R statistical computing packages [168, 194-198] (Appendix E). In each district, I calculated the proportion of all TT cases found within each age group (<15 years, 15–39 years, and ≥40 years), and the unadjusted and age- and sex-adjusted EU-level prevalence for adults aged ≥15-years and adults aged ≥40-years. Then, the cluster-level proportion of cases in the adults aged ≥40-years was calculated, along with the DE. Again using the adults aged ≥40-years, prevalence estimates were calculated with 95% confidence intervals determined by bootstrapping [199] over 10,000 replications with 1) resampling all clusters (to determine the 95% confidence intervals of the "true" prevalence), and 2) resampling half the clusters. In each bootstrapping set, the 2.5th and 97.5th centiles of the ordered means were used as the lower and upper bounds, respectively, for the confidence interval.

To understand the precision associated with reducing the number of clusters, data from the Monduli District dataset (in which everyone aged ≥1 year was invited to be examined) were

bootstrapped, with replacement, over 10,000 replications, three times: first selecting 30 clusters, then 40 clusters, then 50 clusters in each resample.

4.4.4 Validation results

In Monduli, 7,811 individuals from 1,894 households within 60 clusters were examined. In Budaka, 6,839 individuals from 1,729 households within 60 clusters were examined. In Touboro, 5,278 individuals from 1,816 households within 60 clusters were examined. In Am-Timan, 4,071 individuals from 1,798 households within 60 clusters were examined.

In Monduli, 96.5% (136) of the TT (trichiasis + scarring) cases were found in adults aged ≥40-years, 2.8% (4) of the cases were found in adults aged 15-39-years, and 0.7% (1) were found in children aged <15-years. For the other districts, the following results are derived from the clusters where individuals aged ≥1-year were examined. No TT cases were found in individuals under the age of 40 years in Budaka. In Touboro 86.7% (26) of the TT cases were found in adults aged ≥40-years, 13.3% (4) of the cases were found in adults aged 15-39-years and no cases were found in children aged <15-years. In Am-Timan, 89.2% (25) of the TT cases were found in the adults aged ≥40-years, 10.7% (3) of the cases were found in adults aged 15-39-years and again no cases were found in those aged <15-years (Table 4.6).

Table 4.6 Summary of TT results by age group

Location	Age group (years)	Persons examined	TT cases	Percent of cases	Unadjusted prevalence
Am-Timan	<15	1,713	0	0%	0%
	15-39	722	3	10.7%	0.4%
	40+	353	25	89.2%	7.1%
Budaka	<15	2,541	0	0%	0%
	15-39	1,542	0	0%	0%
	40+	1,340	50	100%	3.7%
Monduli	<15	2,877	1	0.7%	0.03%
	15-39	1,782	4	2.8%	0.2%
	40+	3,149	136	96.5%	4.3%
Touboro	<15	1,446	0	0%	0%
	15-39	1,501	4	13.3%	0.3%
	40+	1,160	17	86.7%	2.2%

In Am-Timan, the EU level age- and sex-adjusted prevalence of trichiasis in adults aged ≥15-years was 1.0% (CI 0.5%-1.5%) and that for the adults aged ≥40-years was 2.0% (CI 1.9%-3.9%). In Monduli, the EU level age- and sex-adjusted prevalence of trichiasis in the adults aged ≥15-years was 1.2% (CI 0.9%-1.7%) and that for the adults aged ≥40-years was 3.0% (CI 2.1%-4.1%). In Budaka, the EU level prevalence for the adults aged ≥15-years population was 0.6% (CI 0.3%-0.8%), being 2.5% (CI 1.7%-3.1%) in the adults aged ≥40-years. In Touboro, the prevalences for the adults aged ≥15-years and adults aged ≥40-years were 0.9% (CI 0.5%-1.2%) and 2.0% (1.4%-2.6%) respectively (Table 4.7, Figure 4.9).

Table 4.7 Summary of results

Location	Age group (years)	TT prevalence (unadjusted)	TT prevalence (age- and sex- adjusted)	Lower bound of CI (95%)	Upper bound of CI (95%)	Design effect
Am-Timan	15+	2.5%	1.0%	0.5%	1.5%	1.2
	40+	3.6%	2.0%	1.9%	3.9%	1.11
Budaka	15+	0.9%	0.6%	0.3%	0.8%	1.2
	40+	3.2%	2.5%	1.7%	3.1%	1.03
Monduli	15+	1.9%	1.2%	0.9%	1.7%	3.5
	40+	4.4%	3.0%	2.1%	4.1%	1.05
Touboro	15+	1.1%	0.9%	0.5%	1.2%	1.3
	40+	2.4%	2.0%	1.4%	2.6%	1.02

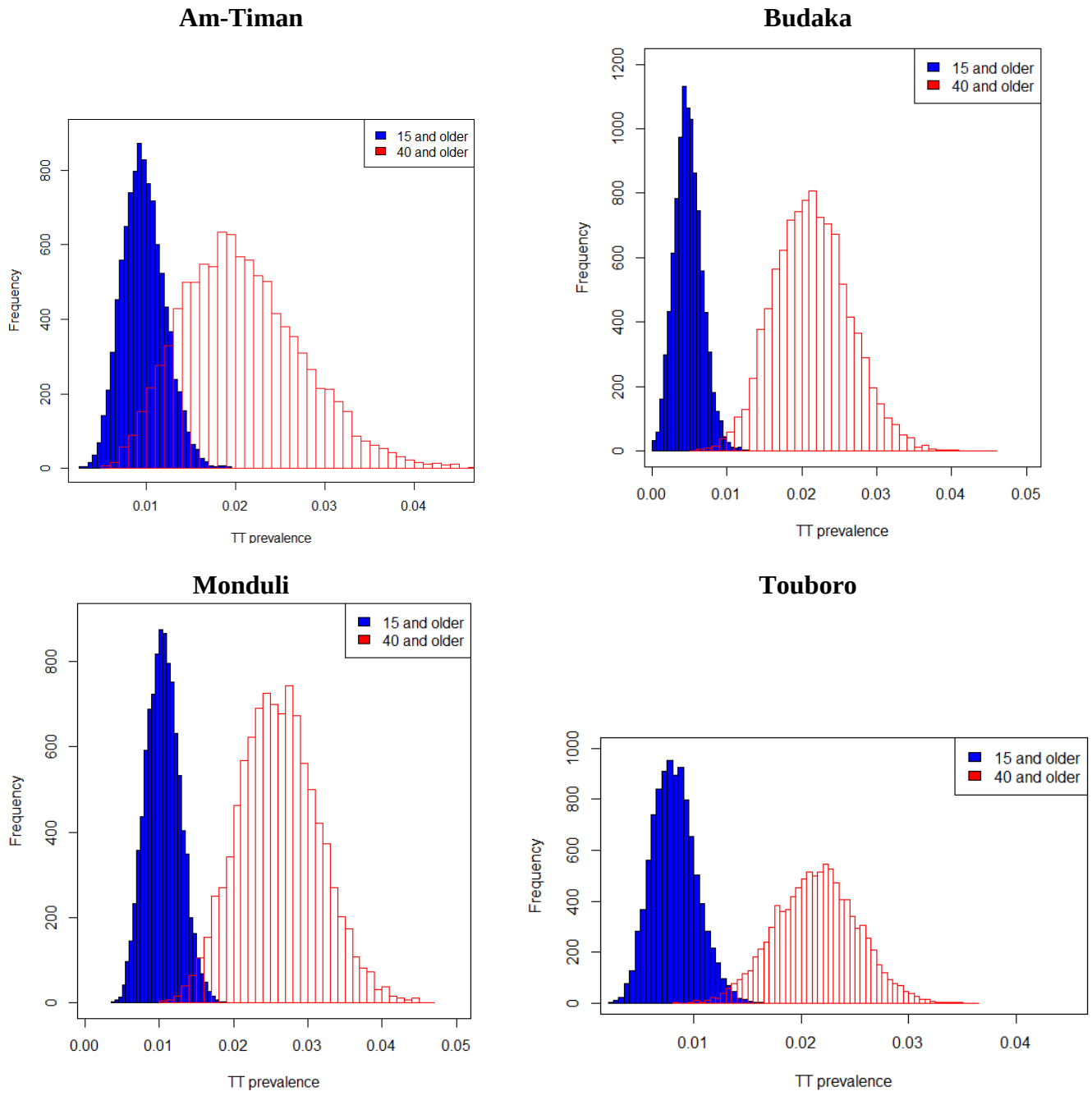


Figure 4.9 Comparative results of TT random selection with 10,000 replications from the adults aged ≥15-years and adults aged ≥40-years

The cluster-level age- and sex- adjusted proportions of TT in adults aged ≥40-years in Monduli ranged from 0% to 21.5% with a DE of 1.05. In Budaka, the proportions ranged from 0% to 11.3% with a DE of 1.03. In Touboro, the proportions ranged from 0% to 12.5% with a DE of 1.02. In Am-Timan the proportions ranged from 0% to 18.4% with a DE of 1.11 (Figure 4.10). These results tell us that observations within clusters are only very marginally more similar than observations from different clusters.

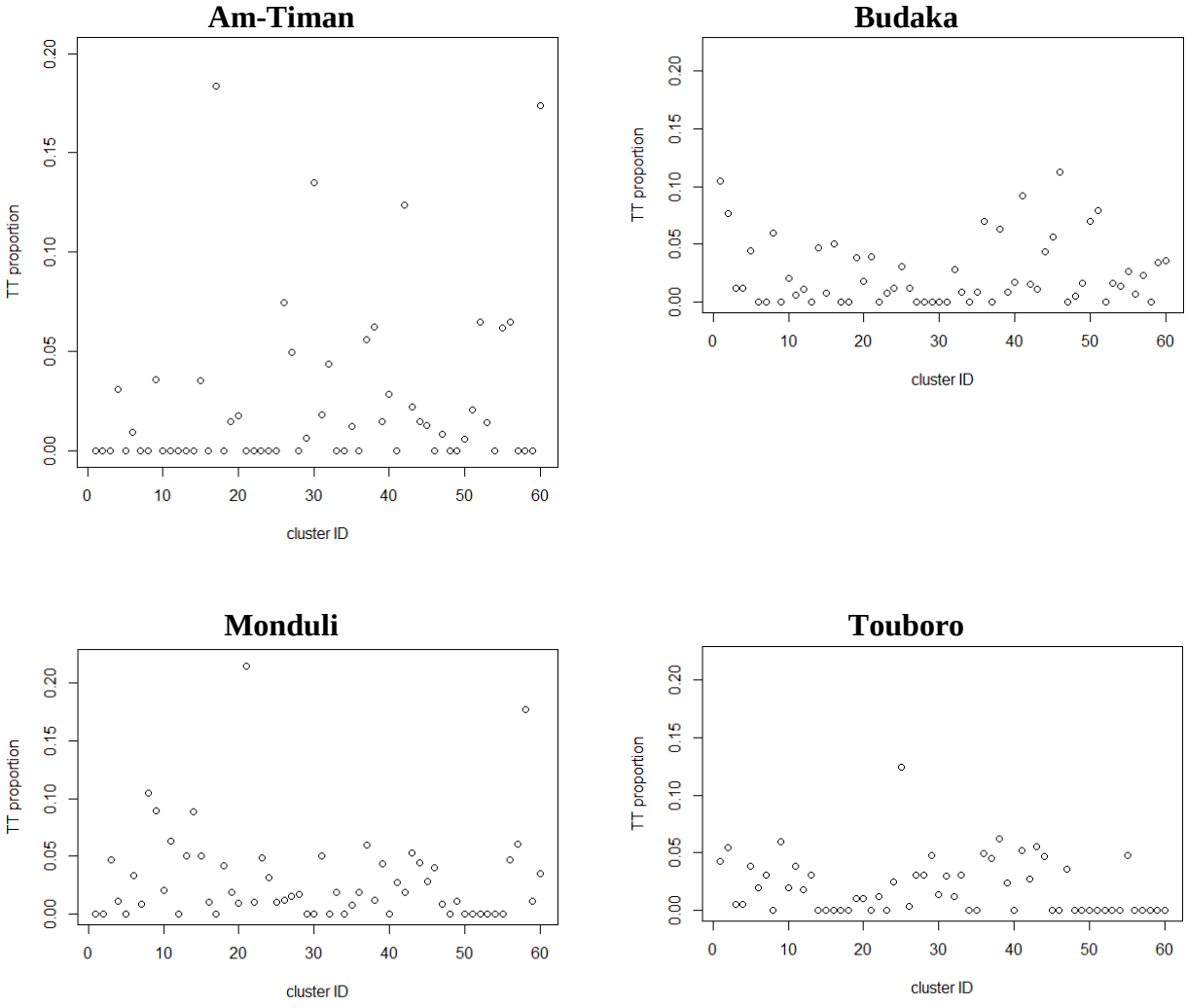
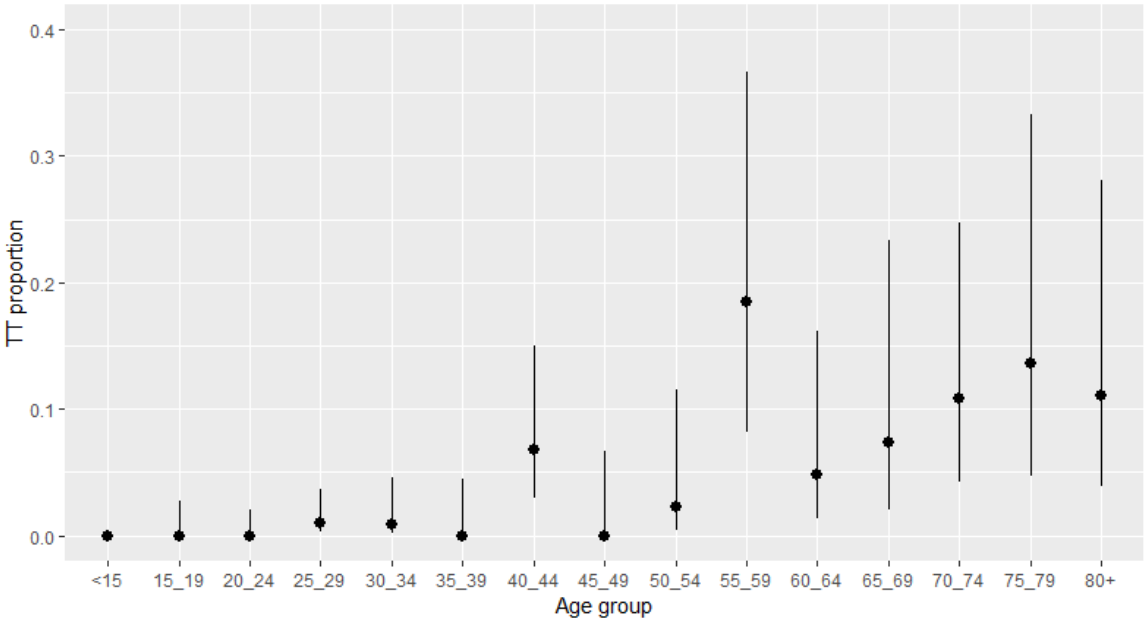


Figure 4.10 Cluster level proportions calculated for adults aged ≥40-years in Monduli, Budaka, Touboro, and Am-Timan

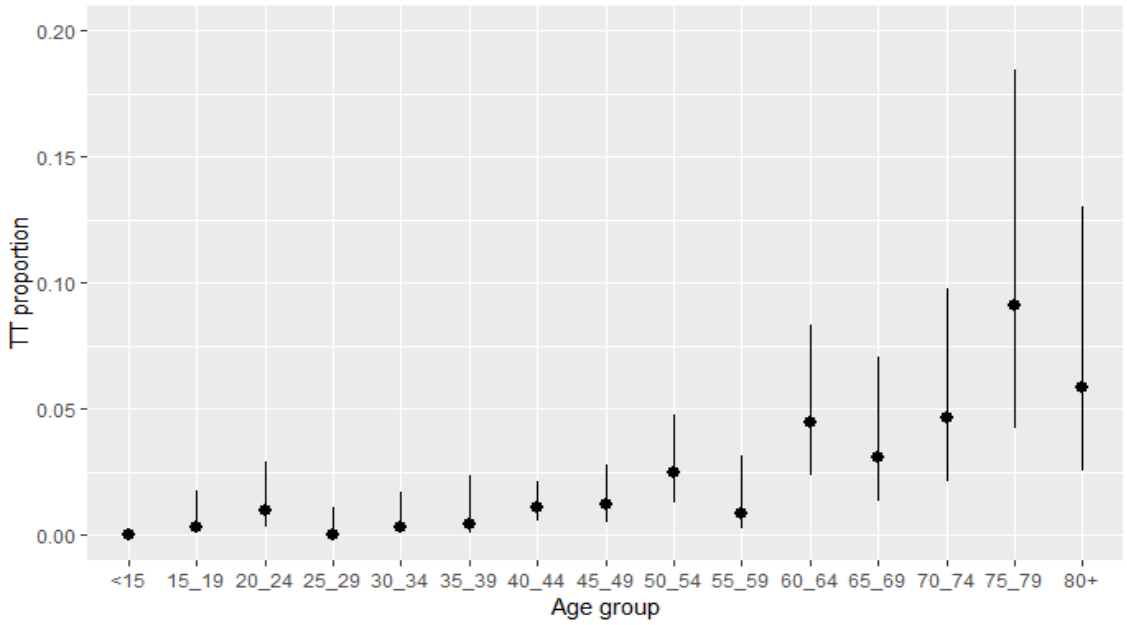
The non-response rates of the adults aged ≥15-years derived from our validation surveys in the clusters where individuals aged ≥1-years were examined ranged from 1.8% to 7.7%.

In all four scenarios the proportion of positive cases trends up as age increases and this uptick appears to occur between age 40 and 50 (Figure 4.11).

Am-Timan, Chad



Budaka, Uganda



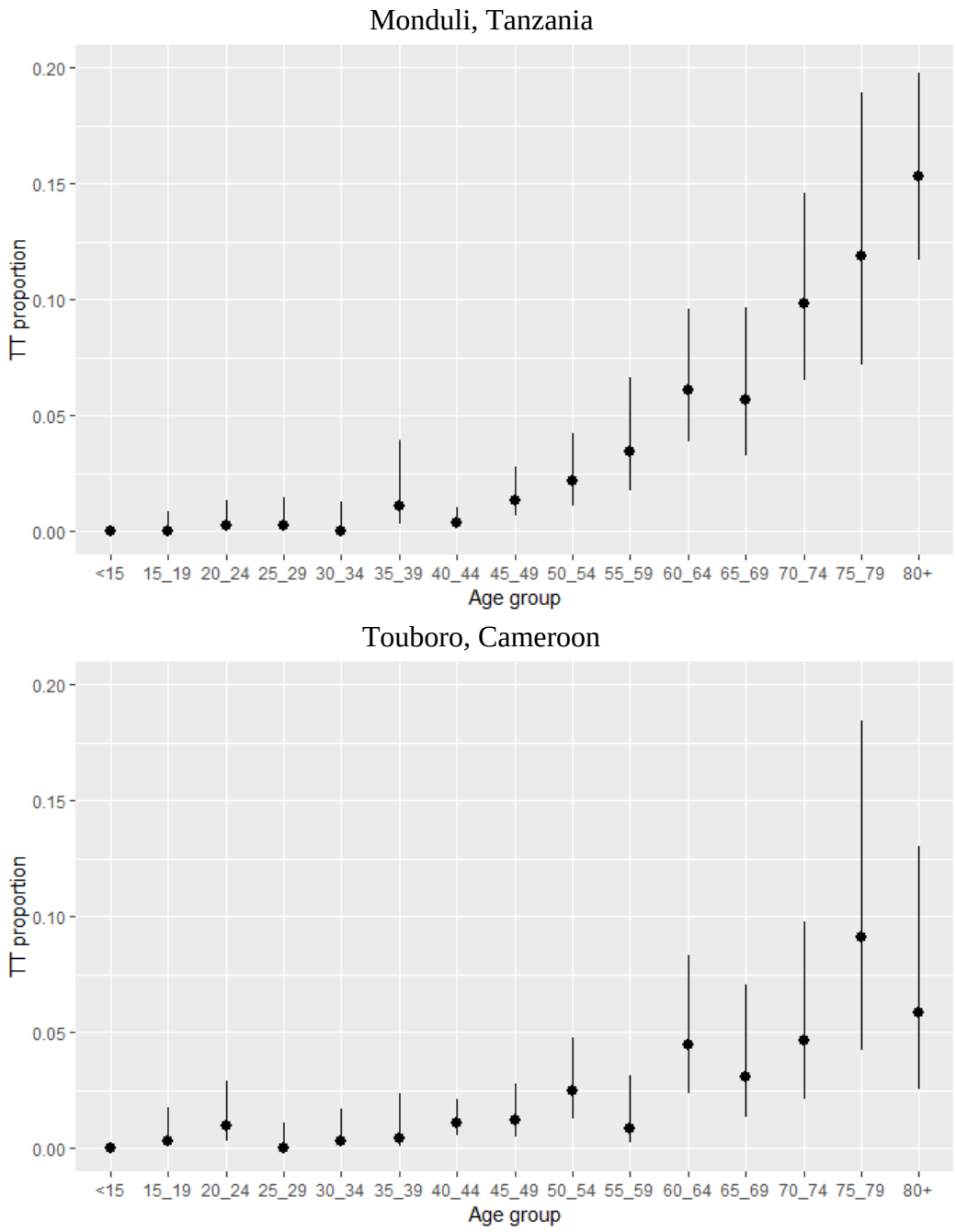


Figure 4.11 Demonstration of the proportion of TT cases found in each age group with 95% confidence intervals.

As expected, the number of individuals examined generally decreases with increased age. However, in Budaka, Monduli, and Touboro there is a marked spike in participation among the 40-44 age group (Figure 4.12).

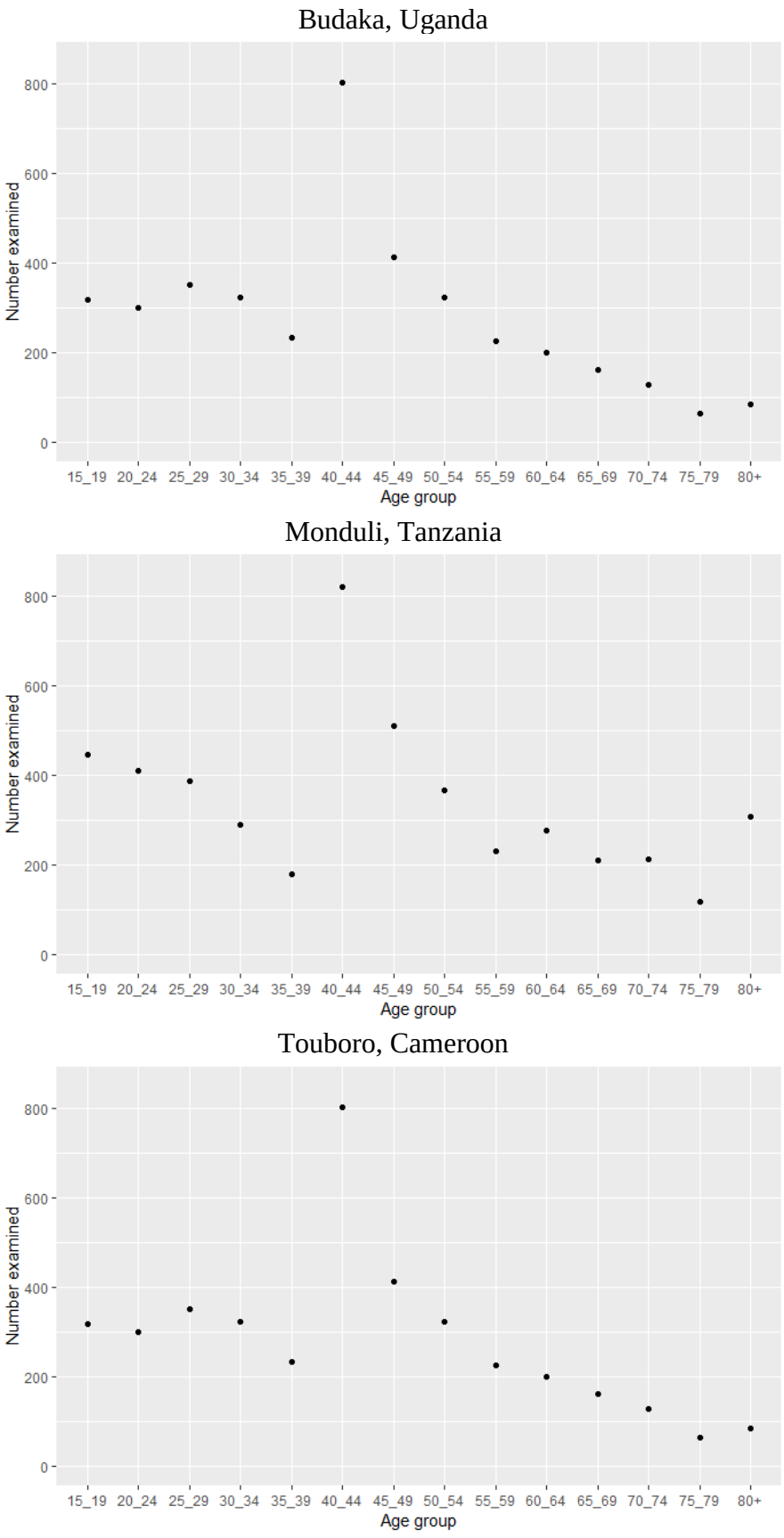


Figure 4.12 Age distribution of persons examined

In the Monduli dataset, a comparison of bootstrapping (with 10,000 replications) with selection of 60 clusters versus selection of 30 clusters, using only data from adults aged ≥40-years, produced confidence intervals of 2.1%-4.1% and 1.8%-4.6%, respectively. For Budaka, the same simulation exercise resulted in confidence intervals of 1.7%-3.1% and 1.5%-3.4%, respectively. In Touboro, the confidence intervals were 1.4%-2.6% and 1.3%-3.0%, respectively. In Am-Timan the confidence intervals were 1.9%-3.9% and 1.5%-4.4% (Figure 4.13).

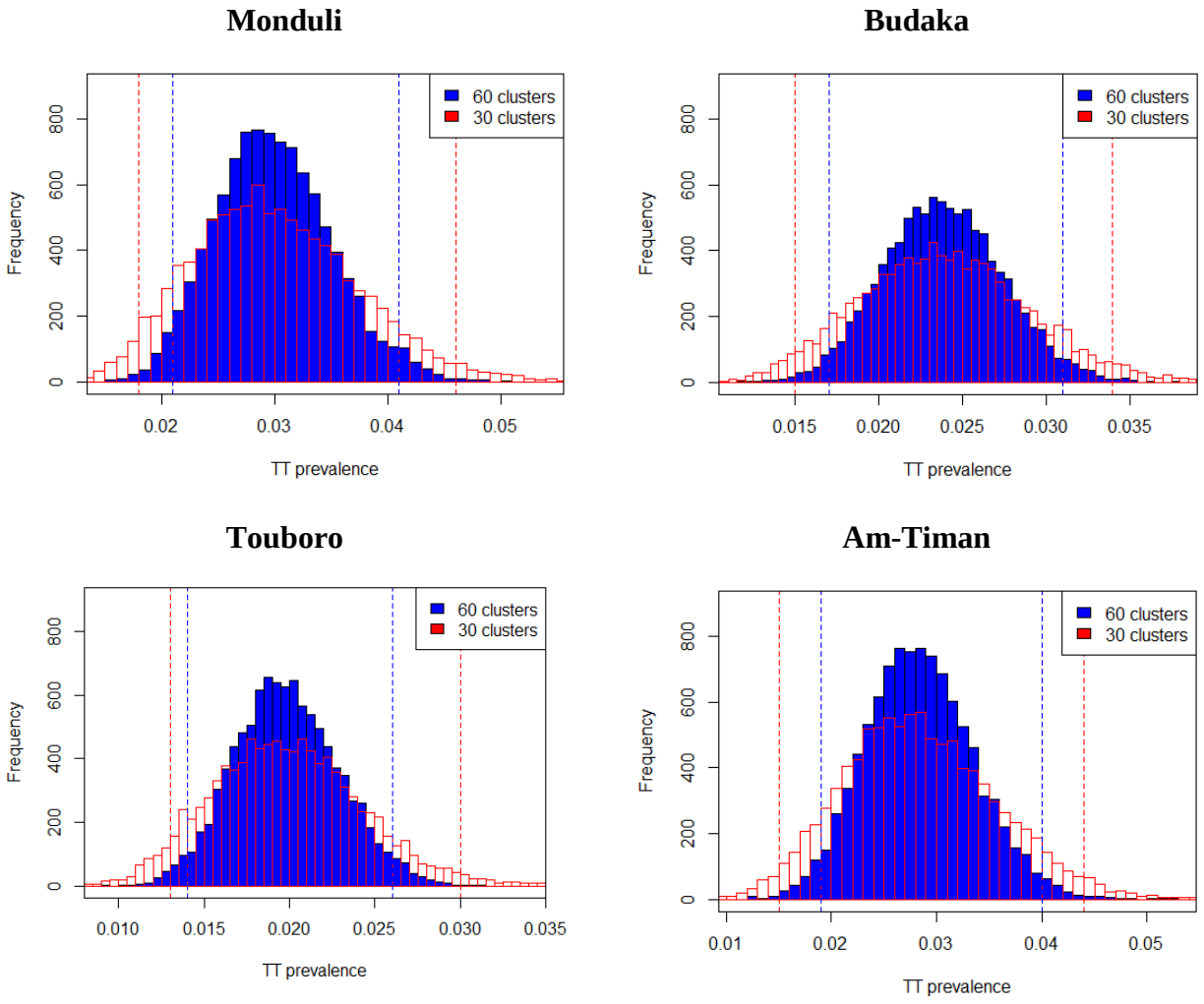


Figure 4.13 Comparing a selection of 30 clusters and 60 clusters in adults aged ≥40-years with 10,000 replications generated through bootstrapping cluster level results. Dashed vertical lines indicate 95% confidence intervals.

Results from bootstrapping with 10,000 replicates using the adults aged ≥15-years from the Monduli dataset are shown in Table 4.8.

Table 4.8 Results from randomly selecting 30, 40 and 50 clusters from the adults aged ≥15-years clusters from the Monduli dataset with 10,000 replications

Location	Number of clusters	Prevalence of TT			
		Mean	SD	Low CI (95%)	Upper CI (95%)
Monduli	30	1.23%	0.30%	0.69%	1.86%
	40	1.23%	0.26%	0.77%	1.77%
	50	1.22%	0.23%	0.80%	1.76%

4.4.5 Discussion

Precise data on the numbers affected with blinding trachoma are essential to enable national control programmes to effectively target surgeries, to evaluate ongoing interventions, and to determine if elimination thresholds have been reached. The design and validation of this TT specific survey provides a practical and reproducible tool to guide trachoma elimination programmes. Based on the evidence shown here, a TT specific survey targeting 30 clusters and examining at least 2,818 adults aged ≥15-years will enable programmes to estimate an expected TT prevalence of 0.2% with absolute precision of 0.2%. Results from the validation exercise suggest this design provides consistent prevalence estimates with reasonable precision for public health decision making.

PBPS's remain the gold standard methodology for obtaining accurate disease estimates when case detection and reporting through the health system is incomplete. Thirty clusters is often taken as a rule of thumb for PBPS, but there has been little empirical evidence to support this selection. Through bootstrapping, I evaluated the stability of a 30-cluster survey. As seen in Figure 4.13, the estimates for 30 clusters and 60 clusters were closely aligned with only slightly wider confidence intervals. This is especially evident when examining the Tanzania results in

Table 4.8, where all 60 clusters include data on adults aged ≥15-years, the precision gained in estimating the adults aged ≥15-years prevalence by including additional clusters is minimal and not an efficient use of resources.

In the context of a public health initiative, it is necessary that a survey does not only provide sufficient precision but is also affordable. In an effort to reduce the necessary sample size, and therefore lessen time in the field, I evaluated potential target age ranges, including the exclusion of adults aged ≥40-years. Whilst I consistently demonstrate across multiple settings that >85% of TT cases will be found in the adults aged ≥40-years, a comparison with the data obtained when including the adults aged ≥15-years highlighted some important limitations of age-specific recruitment. First, I observed a spike in the 40-44 years age group. This may represent, in part, a terminal digit preference for multiples of ten when self-reporting age, and/or demonstrate an age bias in individuals examined. Introducing a bias by limiting enrolment to those adults aged ≥40-years is concerning, since a desire to be examined on the part of individuals aged <40-years may systematically alter the prevalence estimate obtained. It is also notable that, within the four surveys conducted here, the ratio between the prevalence estimated in those adults aged ≥40-years and those adults aged ≥15-years, varied from 2.0 to 4.2. Continuing to use adults aged ≥15-years as the sampled population will help minimize bias and ensure practicality. Resulting prevalence calculations should of course incorporate adjustment for age and sex of those examined, using the methods published by the Global Trachoma Mapping Project [200].

It is important to note that, whilst this survey was designed for an expected prevalence of 0.2%, the TT prevalence estimates from all four validation districts are all higher. This does potentially limit generalizability to very low prevalence settings. Validation districts were selected under the assumption that they would have very low TT prevalence, based on previous

prevalence estimates and an understanding of the interventions that had taken place since. There are several potential explanations. First, prior TT prevalence estimates may have been inaccurate, especially when non-standardised approaches were taken, such as was seen for Am-Timan. Alternatively, surgery campaigns may have lost momentum or had poor uptake, as has been observed for the Maasai population living in Monduli [201]. This may have been a particular problem for Budaka and Touboro, where no other trachoma control activities (such as MDA) were indicated. Cross border dynamics may have also influenced survey results for rarer outcomes such as these.

The TT-specific survey methodology presented here provides an appropriate balance between feasibility and precision in a public health setting. This proposed methodology has been validated through simulations in four trachoma-endemic districts representing different elimination stages, and the results yielded are consistent and defendable. A report of these findings and the following recommendations was published by WHO [172] making the protocol described here available for use by national programs. To date this survey design has been implemented in over five countries with plans to expand to all areas that qualify for a TT specific survey.

4.5 Recommendations

When undertaken, a TT specific survey should be implemented as a PBPS designed to estimate the prevalence of TT in adults aged ≥15-years. The sample size is calculated to estimate, with 95% confidence, an expected TT prevalence of 0.2% with absolute precision of 0.2% and a design effect of 1.47, yielding 2,818 as the target number of adults aged ≥15-years to be examined. This should be appropriately inflated to account for the expected non-response rate. The number of clusters, *c*, that would ideally be included is given by $c = (2{,}818 \times [non -$

$response\ inflator$])/($h \times a$), where h is the number of households that can be seen by one team in one day, and a is the expected number of adults resident in each house, as determined by the most recent census or recent population-based trachoma survey experience. If c, determined by the above formula, is ≥30, 30 clusters should be used.

When trichiasis is observed, the eye should be evaluated for the presence or absence of TS, as defined within the WHO simplified trachoma grading scheme [19], and the subject should be asked scripted questions to determine whether interventions to manage the trichiasis in that eye have previously been recommended by health care workers [186, 187].

Prevalence calculations should incorporate adjustment for age and sex of those examined, using the methods published by the GTMP [200].

4.6 Time effectiveness of a trachomatous trichiasis (TT) specific survey

A public health survey must be time effective to be practical to implement. In the next section I will outline the time-costs associated with a TT-specific survey where adults aged ≥15-years are included in comparison with a survey where adults aged ≥40-years are included. This information is useful to country programmes when budgeting for mapping.

The purpose of this exercise is to compare the amount of time needed for a TT-specific survey where adults aged ≥15-years are included in comparison with a survey where adults aged ≥40-years are included to determine if the time saved in the ≥40-years survey is beneficial enough to further explore the terminal age preference identified in Chapter 4.4.4.

In the validation exercise, 60 clusters were selected per EU in both Monduli, Tanzania and Budaka, Uganda. In Monduli, all consenting individuals who were aged ≥1-year and living in the selected households were included in the survey. In Budaka, 30 clusters included

individuals aged ≥1-year and 30 clusters included only adults aged ≥40-years. For clarity in this chapter I will refer to the Budaka clusters where individuals aged ≥1-year were included as clusters(1) and where only adults aged ≥40-years were included as clusters(40). Time-stamps were included throughout the questionnaire, providing useful information for estimating time-cost at each stage of the survey.

The time associated with each component of the data collection process was calculated by determining the mean time across clusters. First, the difference between the time of arrival at the first household in a cluster (time point B) and the time of arrival at that cluster (time point A) was calculated. Second, the difference between the start of the first exam in a household (time point C) and arrival at that household (time point B) was determined. Then the difference between the start (time point C) and end of each exam (time point D) was calculated for each individual. The difference between the end of the final exam in a household (time point E) and the arrival at the household (time point B) was determined. Finally, the difference between completion of the final exam in a cluster (time point F) and the time of arrival in the cluster (time point A) was calculated (Table 4.9, Figure 4.14).

Table 4.9 Time point associated with each component of the data collection process

Time Point	Event
A	Team arrives in cluster
B	Team arrives in first household
C	Team starts first exam in first household
D	End of first exam in the first household
E	End of the final exam in the first household
F	End of last exam in the last household (departure from the cluster)

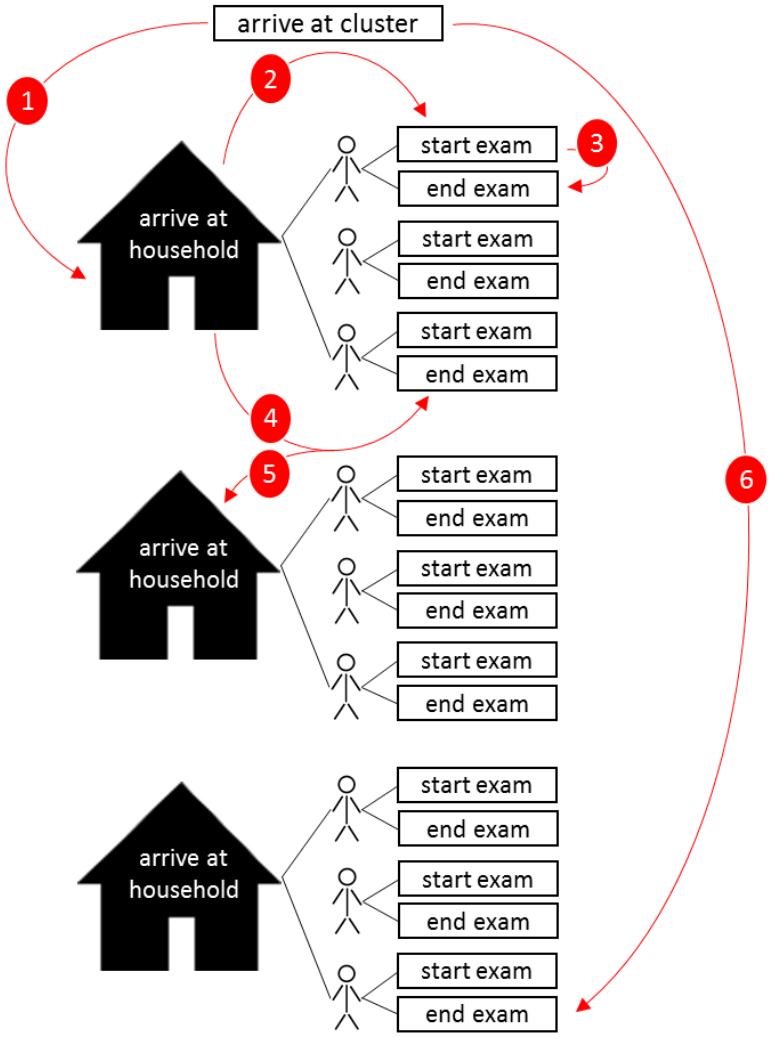


Figure 4.14 1) difference between arriving at household and arriving at cluster quantifies time spent preparing cluster (i.e. sensitizing village leader) 2) difference between start of first exam and arriving at household quantifies time spent organizing and explaining process to household 3) difference between start and end of each exam quantifies time spent examining each individual 4) difference between end of final exam in household and arrival at household quantifies time spent at each household 5) difference between end of final exam in final household and arrival at cluster quantifies time spent in each cluster

Next, time associated with examining individuals with positive TT was calculated and a comparison of time and positive cases was performed. Time spent in a cluster was determined using the following calculation: $t + (b \times h) + (h \times e \times a) + (c \times h)$, where t is the time preparing for data collection, b is average time spent organizing households, h is number of households, e is average time spent per examination, and a is average number of examinations per household and c is average travel time between households per cluster.

4.6.1 Results

The mean time spent in a household in Monduli was 15 minutes 10 seconds. In Budaka the mean time spent in a household was 4 minutes 58 seconds in cluster(1) and 5 minutes in cluster(40) (Table 4.10).

Table 4.10 Mean time spent at each stage of the survey in Monduli and Budaka with 95% confidence intervals

		Monduli	Budaka cluster(1)	Budaka cluster(40)
Time spent preparing to begin data collection		0:40:42 (0:33:15-0:48:10)	0:06:54 (0:04:59-0:08:49)	0:10:18 (0:08:07-0:12:29)
Time spent organizing household prior to first exam		0:03:38 (0:03:05-0:04:10)	0:01:08 (0:01:02-0:01:14)	0:02:18 (0:01:46-0:02:50)
Time spent examining individuals	≥1 year	0:01:27 (0:01:26-0:01:29)	0:00:30 (0:00:29-0:00:30)	
	≥15 years	0:01:36 (0:01:34-0:01:38)	0:00:33 (0:00:32-0:00:33)	
	≥40 years	0:01:43 (0:01:40-0:01:45)	0:00:39 (0:00:38-0:00:41)	0:00:59 (0:00:57-0:01:01)
Time spent at household		0:15:10 (0:13:31-0:16:49)	0:04:58 (0:04:40-0:05:17)	0:05:00 (0:04:15-0:05:46)

The mean time spent examining individuals with positive TT in Monduli was 3 minutes 35 seconds, in Budaka cluster(1) 1 minute 40 seconds, and in Budaka cluster(40) 2 minutes 12 seconds. Whilst the mean of negative cases was 1 minute 25 seconds, 29 seconds and 51 seconds in Monduli, Budaka cluster(1) and Budaka cluster(40) respectively (Table 4.11).

Table 4.11 Mean examination time for positive and negative TT cases in Monduli and Budaka with 95% confidence intervals

	Monduli	Budaka cluster(1)	Budaka cluster(40)
Time examining positive cases	0:03:35 (0:03:11-0:03:59)	0:01:40 (0:01:28-0:01:52)	0:02:12 (0:01:54-0:02:29)
Time examining negative cases	0:01:25 (0:01:23-0:01:28)	0:00:29 (0:00:29-0:00:30)	0:00:51 (0:00:50-0:00:52)

Overall the mean time per cluster to examine the ≥40 population was 4 hours 43 minutes, the adults aged ≥15-years was 5 hours and 25 minutes and the individuals aged ≥1-year was 6 hours and 47 minutes in Monduli. The mean time per cluster in the Budaka cluster(1) was 4

hours 40 minutes for the adults aged ≥40-years, 5 hours 16 minutes for the adults aged ≥15-years and 6 hours 30 minutes for the individuals aged ≥1-year. In Budaka cluster(40) the mean time was 3 hours and 51 minutes (Table 4.12).

Table 4.12 Mean time by age group

Project	Age	Household count	Persons examined per household (mean)	Time per cluster (mean)	Percent of positive TT cases found per age group
Monduli	≥1 year	32	4.1	6:47:32	0.71%
	≥15 years	32	2.6	5:25:19	2.84%
	≥40 years	32	1.8	4:43:08	96.45%
Budaka Cluster(1)	≥1 year	31	5.8	6:30:58	0.00%
	≥15 years	31	3.1	5:16:41	0.00%
	≥40 years	31	1.7	4:40:22	100%
Budaka Cluster(40)	≥40 years	27	1.7	3:51:32	100%

4.6.2 Discussion

Time spent preparing a cluster to begin data collection varies depending on the quality of sensitization provided prior to the start of the survey as well as the individual interest of the village leader and the experience of the examination team. Time spent preparing a household for examination also varies depending on the familiarity of the head of household with public health surveys and willingness to participate. In Budaka it took longer to prepare the households in the cluster(40) sample. It may be that more time was spent explaining why only some individuals would be examined.

The examinations in Monduli took longer than the examinations in Budaka. The Monduli survey teams did not include dedicated recorders, whilst the Budaka teams did. The two-person teams in Budaka were able to move through an exam quicker than the one-person in Monduli. However, the Monduli teams were able to enrol multiple households at a time so the time

essentially evened out between a two-person (recorder/examiner team) and two one-person teams.

A survey which only includes adults aged ≥40-years takes marginally less time than a survey which examines adults aged ≥15-years. The time difference in an adults aged ≥40-years and an adults aged ≥15-years survey in Monduli was 42 minutes. The additional 42 minutes taken to examine the individuals aged 15-39-years provided an extra 2.9% of TT cases. In Budaka the difference between examining adults aged ≥40-years in cluster(40) and adults aged ≥15-years in cluster(1) was 1 hour and 25 minutes. No TT cases were found in adults aged <40-years in Budaka and so the additional time did not add any value in that particular situation.

4.6.3 Conclusion

Limited additional information is gained by including individuals aged 15-39-years in the survey. However, the time saved by not including this age group does not reduce the direct costs. Trachoma survey teams are familiar with adults aged ≥15-years cut-off and confusion may arise if this cut-off is changed to ≥40-years. Additionally, as demonstrated in the previous chapter, a ≥40-years cut-off introduces a recruitment bias. It should be noted that whilst there are no direct cost implications, the enthusiasm of the survey team could be positively impacted by the saved time. The majority of time spent in the survey is sensitising village leaders and households, which is not strongly related to the age being examined. Continuing to include adults aged ≥15-years is desirable.

In the previous chapters I estimated the national and global burden of TT, designed and validated a TT specific survey and outlined the time-cost associated with the recommended survey design. Whilst these pervious chapters clearly aim to inform programmatic work within trachoma elimination programmes the following chapters provide a more robust understanding

of the geographical distribution of TT. In Chapter 5, I explore the community-level spatial heterogeneity of TT with focus on its relationship with the distribution of TF. I then present a method for identifying TT community-level hot spots.

Chapter 5: **Understanding the spatial distribution of trichiasis and its association with trachomatous inflammation follicular**

5.1 Overview

The previous chapter provided a validated design for a trachomatous trichiasis (TT) specific survey. This methodology has been made available for national trachoma control programs through the support of the WHO, Eighth NTD-STAG. This survey design is programmatically useful and has since been implemented in numerous countries. As countries approach elimination targets it will become even more difficult to find the final TT cases. In the following chapters I will present geostatistical methods that may contribute to TT case finding in the context of the end game for trachoma control.

With the availability of the GTMP dataset, I have estimated the burden of TT and designed and validated a TT specific survey. I will now use country specific exerts for the GTMP dataset, namely; Benin, Cote d'Ivoire, DRC, Guinea, Ethiopia, Malawi, Mozambique, Nigeria, Sudan and Uganda to explore the spatial heterogeneity of TF and TT. I will specifically identify the association between community-level TF and TT and how this is influenced by environmental risk factors.

I built the geostatistical models and conducted the analysis presented here. I also wrote the manuscript informed by the information presented in this chapter.

5.2 Background

Trachoma is a blinding disease caused by recurrent ocular *Chlamydia trachomatis* infection which results in chronic inflammation of the tarsal conjunctiva. This is characterised by sub-epithelial follicles, which may meet the definition to be classified as TF [21]. TF is the sign

used to determine whether public health-level interventions against active trachoma are needed [19]. Through repeated reinfection, scarring may develop, eventually causing the eyelashes to turn inward and touch the globe, a state known as TT. Inturned eyelashes that abrade the cornea can result in corneal opacity and blindness [21]. Corrective surgery [74, 77] or epilation [75] are used to manage these TT cases. As ocular chlamydial transmission declines in many countries [83, 202], it becomes important to focus on areas where TT remains a public health problem in the absence of TF.

The natural progression of trachoma, as implicitly conceptualized within WHO recommendations for programme-level interventions, follows that episodes of TF are pre-requisites on the causal pathway to TT, with moderate to high prevalences of TF being a proxy for current transmission of ocular *C. trachomatis,* and TT a proxy for historic transmission. The prevalences of these signs are therefore taken as signals for *C. trachomatis* infection at different time points (TF is current, and TT is historic). Even though these signs are markers of transmission at different timepoints, the driving factors (water, sanitation and hygiene) for community-level TF distribution gradually change and I expect; in areas where antibiotic MDA for trachoma [203] has not yet occurred, current community-level TF (as a proxy measure of trachoma transmission) to be a good predictor of current community-level TT. However, this is not always the case.

Many country programmes [83] have successfully reduced district-level TF prevalence in children aged 1-9-years below the elimination threshold of 5% [73, 133, 204]. However, to eliminate trachoma as a public health problem, district-level TT prevalence must also be reduced below 0.2% in adults aged ≥15-years [133, 162]. Whilst TF is no longer considered a public health problem in these districts, in many, TT still is. This is particularly evident in Nigeria where 94 local government areas (LGAs) in six states mapped through the GTMP

yielded a district-level TF prevalence below the elimination threshold and district-level TT prevalence above the threshold [180-184, 205, 206]. This is attributable to historic rather than contemporary transmission intensity. It is therefore important to better understand what factors influence the development of TT so as to develop more targeted control interventions. Understanding where TT cases are likely to occur could help to guide strategic placement of TT intervention services.

Thanks to the GTMP, there has been an increasing availability of high quality geolocated trachoma and water, sanitation and hygiene (WASH) data. The GTMP was launched in December 2012 with the aim of mapping the global prevalence of trachoma through PBPS's. The GTMP systematically collected trachoma and WASH data across 1,542 districts in 29 countries in areas where control activities including mass drug administration (MDA) had not yet occurred [71]. These data can be used to further our understanding of TT distribution.

In this study, I attempted to identify risk factors that, in addition to TF, might contribute to the variation in community-level TT prevalence. I fitted binomial mixed models with random effects at community-level to GTMP data from ten countries. I then tested for residual spatial correlation and, in countries where this was detected, used geostatistical methods to model the variation in TT prevalence between countries.

5.3 Data

Ten GTMP collaborating countries provided data for this study: Benin, Cote d' Ivoire, Democratic Republic of the Congo (DRC), Ethiopia, Guinea, Malawi, Mozambique, Nigeria, Sudan and Uganda. Data provided were from 15,051 clusters (or communities) within 624 EUs. Individual-level information on the presence or absence of TF and TT, as well as water and sanitation conditions of geolocated households, were provided. Community-level TT

prevalence was calculated as the ratio between the number of adults aged ≥15-years with trichiasis in at least one eye and the number of adults aged ≥15-years examined. Community-level TF prevalence was calculated as the ratio between the number of children aged 1–9 years with TF in at least one eye and the number of children aged 1–9 years examined.

Table 5.1 Summary of GTMP data included in the analysis

Country	No. of communities	TF in children aged 1–9 years			TT in adults aged ≥15-years		
		No. examined	No. positive (%)	No. communities prevalence ≥5%	No. examined	No. positive (%)	No. communities with prevalence ≥0.2%
Benin	213	18,781	1,594 (8.5%)	94 (44.1%)	16,170	254 (1.6%)	89 (41.8%)
Cote d'Ivoire	256	17,658	1,829 (10.4%)	174 (68%)	18,771	39 (0.2%)	26 (10.2%)
DRC	1,023	74,142	7,022 (9.5%)	610 (59.6%)	52,200	1,137 (2.2%)	511 (50%)
Ethiopia	4,480	186,308	40,131 (21.5%)	3,119 (69.6%)	289,230	4,711 (1.6%)	2,154 (48.1%)
Guinea	295	19,488	832 (4.3%)	98 (33.2%)	21,955	66 (0.3%)	53 (18%)
Malawi	1,948	82,185	3,437 (4.2%)	561 (28.8%)	110,815	358 (0.3%)	259 (13.3%)
Mozambique	696	34,602	2,133 (6.2%)	297 (42.7%)	35,895	155 (0.4%)	117 (16.8%)
Nigeria	5,364	337,962	10,070 (3%)	1,105 (20.6%)	371,928	4,815 (1.3%)	2,035 (37.9%)
Sudan	667	33,830	1,394 (4.1%)	172 (25.8%)	40,501	327 (0.8%)	197 (29.5%)
Uganda	109	6,019	183 (3%)	26 (23.9%)	7,445	21 (0.3%)	20 (18.3%)
Total	15,051	810,975	68,625 (8.5%)	6,256 (41.6%)	964,910	11,883 (1.2%)	5,461 (36.3%)

Physical and social environmental factors play an important role in the natural history of trachoma. These factors can accelerate progression to TT (Figure 5.1).

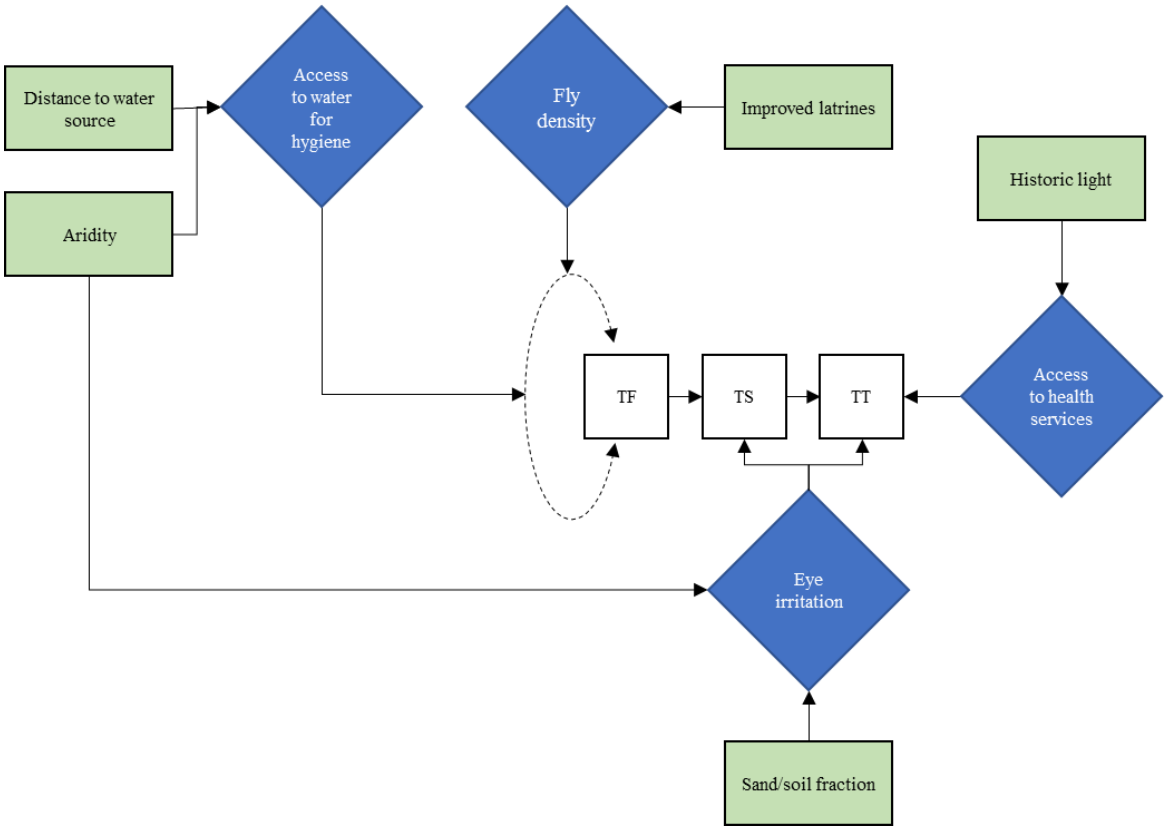


Figure 5.1 Conceptual framework describing the progression of TF to TT in the context of environmental risk factors. Proxy indicators are displayed in green boxes with the mechanism of influence in blue diamonds. Here I suspect that access to water for hygiene and fly density contribute to reoccurrence of TF. This reoccurrence leads to TS and TT. I suspect eye irritation hastens the progress from TS to TT and access to health services influences continued presence of TT.

Facial cleanliness is a well-established association of TF [32, 41, 45-48] and access to water is necessary to facilitate personal hygiene practices. Previous studies have indeed found an association between distance to water and risk for trachoma [32, 50-52]. It has also been reported that there is an association between flies on faces and TF [38, 51, 53-56]. *M. sorbens* prefers to breed on human faeces left exposed on the soil [53, 93]. For this analysis, community-level WASH indicators were created from the GTMP household WASH dataset (Appendix F). The categorization of these indicators were informed by the WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation (JMP) [169]. I then calculated the ratio of access to each categorized WASH indicator.

Previous studies have shown that lower precipitation levels and higher temperatures can lead to an increase in the risk of TF [207]. Therefore, I selected climate-related factors including annual total precipitation, mean temperature, aridity index and potential evapo-transpiration (PET) for this analysis. Gridded maps at 1 km^2 resolution of annual total precipitation and mean temperature were extracted from the WorldClim database [115]. WorldClim averages the climate data for minimum, mean, and maximum temperature and precipitation for 1970-2000. The aridity index and PET raster datasets of 1 km^2 resolution, were obtained from the Consortium for Spatial Information (CGIAR-CSI) [208]. CGIAR-CSI modelled aridity index and PET using the data available from WorldClim as input parameters.

It has been suggested that areas in Sudan with frequent sandstorms result in eye trauma [68]. Irritation of the eyes leads to rubbing with fingers which may accelerate the progression of TT. Hence, in this analysis I consider the proportion of sand in topsoil, as a potential risk factor for TT. These data were obtained from the ISRIC-World Soil Information project included in the Harmonized Soil Map of the World (revised in 2012) [209].

I speculate that access to healthcare and other services are associated with developed infrastructure, and therefore sought an infrastructure indicator. Light density at night has been shown to be correlated with local economic activity and gross production rate at different scales [210, 211]. Night light (NL) emission captured by the Operational Linescan System (OLS) instrument on board a satellite of the Defence Meteorological Satellite Program was used as a proxy measure of poverty across Africa [212, 213]. A gridded map of straight line distances to stable lights, namely NL emissivity > 0, was subsequently produced from the raw NL raster for 1997. This historic year was chosen because I was interested in a measure of infrastructure during the childhood of those likely to have TT at the time of survey data collection, rather than that at the time of the surveys themselves.

All the aforementioned environmental datasets were derived from georeferenced raster files and converted to a standardised resolution of 5 km × 5 km. I considered a 5km buffer a good catchment area for rural communities, considering this as a regular distance people can cover around communities. The georeferenced data were constructed in ArcGIS 10.1 (ESRI, Redlands, CA, USA). When shrinkage of spatial resolution was needed for the 1 km^2 resolution covariates, I estimated the mean value in a 5 km × 5 km window using the aggregate tool available in the Spatial Analyst toolbox of ArcGIS 10.1 (Appendix G).

To identify collinearity among the selected variables, I used the variance inflation factors (VIF) [214], defined as

$$VIF_j = \frac{1}{1-R_j^2}$$

where R_j^2 is the fraction of explained variance in the j-th explanatory variables by the other explanatory variables.

5.4 Model formulation

The models use community-level data, however because I use random effects U_i to represent the unexplained extra binomial variation between communities these are defined as binomial mixed models [215]. I assume communities will have inherent differences which are not explained by community-level variables but are captured in differences in space (i.e. U_i). Let p_i denote the probability of having TT, β_0 is the intercept and U_i is community-level unstructured random effects (let i denote the i-th community). In this model I then fit the following nested binomial mixed models, where γTF_i is the regression coefficient for the effect of TF prevalence on the log-odds of TT:

$$M1: log\ (\frac{p_i}{1-p_i})\ =\ \beta_0 + U_i;$$

$$M2: log\ (\frac{p_i}{1-p_i})\ =\ \beta_0 + \gamma TF_i + U_i;$$

$$M3: log\ (\frac{p_i}{1-p_i})\ =\ \beta_0 + \gamma TF_i + \sum_{j=1} \beta_j d_{ij} + U_i,$$

where in M3; β_j are the regression coefficients of the explanatory variables d_{ij} and d_{ij} are the explanatory variables described in the previous section with ij as the j -th variable within the i-th community. In fitting M3, I also carried out variable selection using a backward stepwise approach, starting from the mixed effects model with all the variables included. The models were fitted in lme4 using the Laplace approximation of the lme4 package [216]. I use the likelihood ratio test to select among the three models defined above.

The likelihood ratio test compares the log likelihoods of the different models using the following formula:

$$likelihood\ ratio\ test\ statistic\ =\ 2(ll(m^2) - ll(m^1))$$

where $ll(m^x)$ is the log likelihood with m^1 being the more restrictive model and m^2 the less restrictive model. The statistic approximately follows a chi-square distribution and statistical significance is determined through considering degrees of freedom equal to the difference in variables between the models.

To assess the presence of residual spatial correlation, I first obtained a point estimate of the community-level unstructured random effects U_i from the best model identified in the previous step, and then computed the empirical semi-variogram. A semi-variogram provides insights

into the rate of decay of spatial autocorrelation in the data by computing the mean squared difference between pairs of residuals as a function of the distance between their associated geographical locations. A flat semi-variogram is interpreted as evidence against the presence of spatial correlation. To test for spatial correlation more formally, I also generated 95% confidence intervals under the assumption of spatial independence. These intervals were obtained by computing semi-variograms on 1,000 randomly permuted point estimates of U_i while holding the geographical locations fixed. For each simulation, data values were randomly allocated to the spatial locations. The empirical variogram was computed for each simulation using the same lags as for the variogram originally computed for the data. The envelops were computed by taking, at each lag, the maximum and minimum values of the variograms for the simulated data.

In cases where I found evidence of spatial correlation, I fitted the model using PrevMap. PrevMap [217] is an R package which implements fitting and spatial prediction of a geostatistical model used in the context of prevalence mapping [218]. Parameters are estimated by Monte Carlo maximum likelihood using importance sampling techniques to approximate the high-dimensional intractable integral that defines the likelihood function.

I fitted the following geostatistical binomial logistic model, in which the function $S(x)$ represents a spatial Gaussian process, and the Gaussian process is assumed to be stationary and isotropic and defined as having a mean of zero, variance of Σ^2 with the correlation function Corr($S(x)$,$S(x')$) being an exponential correlation function with a scale parameter Φ representing the rate of decay in the correlation. In the fitted geostatistical model, Φ and Σ^2 were estimated from the data; initial values for Φ and Σ^2 were derived from the empirical semi-variogram, and maximum likelihood-based estimates of these parameters were then derived from the fitted geostatistical model:

$$M4: log\left(\frac{p_i}{1-p_i}\right) = d'\beta + S(x) + U_i,$$

where d is a vector of covariates with the associated regression coeffiencts β, $S(x)$are spatial random effects that account for spatial variation in TT prevalence between communities not explained by the predictors, and U_i is unstructured random effect representing extra-binomial variation within communities. The script can be found in Appendix H.

The likelihood ratio test was used to compare the binomial mixed model with the geostatistical model. To assess the fit of the geostatistical model the observed TT prevalence values were plotted against the predicted values along with the predictive 95% confidence intervals. Then the percent of observed values that fall within the 95% predictive confidence intervals was calculated.

5.5 Results

The output for the cluster-level multicollinearity test suggests that temperature, precipitation, aridity index, and PET interact with one another (Table 5.2). Since aridity was highly correlated with each of these indicators, I only retained this variable and excluded the remainder.

Table 5.2 Multicollinearity test results for gridded covariates

Variable	VIF
Annual mean temperature	5.4
Annual total precipitation	47.6
Aridity Index	62.3
PET	6.4
Euclidean distance to ground water	1.2
Sand/soil fraction	2.0
Stable Night Light (1997)	1.2
Accessibility	1.3

The strength of association for variables in the full mixed effect model varied between country (Table 5.3). There was very strong evidence of positive association ($P<0.05$) between community-level TF prevalence and TT prevalence in all countries except Guinea and Uganda.

In contrast, there was evidence of a negative association with access to latrines in two of ten countries (DRC and Malawi ($p<0.01$)), with access to improved latrines in Nigeria ($p<0.01$), and with improved water source on property in Nigeria ($p<0.01$). There was evidence of a positive association with access to latrines in three of ten countries (Ethiopia, Mozambique and Nigeria ($p<0.01$)), with access to improved latrines in Uganda ($p<0.05$), and with water source variables in three countries (Ethiopia, Nigeria (water source on property) and Guinea (water source distance more than 30 minutes) ($p<0.05$)). Observed relationships with environmental factors was equally heterogeneous with; negative associations observed for aridity index in DRC, Ethiopia, and Nigeria; positive association observed for aridity index in Sudan; positive associaiton with sand/soil fraction in Benin; negative association with sand/soil fraction in Ethiopia; positive association with night light in DRC, Mozambique and Nigeria; and negative association with night light in Ethiopia.

Table 5.3 Relative increase in odds derived from a multivariate binomial logistic model where community-level prevalence of TT in adults aged ≥15-years is dependent on a 10% increase in community-level prevalence of TF in children aged 1–9 years. Community-level household prevalence of improved sanitation and hygiene facilities as well as gridded covariates were included along with community-level prevalence of TF.

Covariate type	Covariates	Benin		Cote d'Ivoire		DRC		Ethiopia		Guinea	
		OR	p-value	OR	p-value	OR	p-value	OR	p-value	OR	p-value
Trachoma transmission	TF prevalence	1.779	<0.001	1.854	0.010	1.527	<0.001	1.226	<0.001	1.427	0.311
WASH	Latrine defecation	0.596	0.087	0.909	0.421	0.902	<0.001	1.046	<0.001	1.067	0.254
	Improved latrines	1.664	0.109	1.153	0.324	0.961	0.110	0.990	0.489	0.950	0.362
	Improved water source	0.978	0.554	0.925	0.280	0.978	0.067	1.015	0.005	1.022	0.683
	Improved water source on property	1.328	0.191	0.854	0.469	0.948	0.758	1.011	0.801	1.075	0.708
	Water source on property	0.805	0.283	1.522	0.052	0.951	0.618	1.031	0.506	1.083	0.699
	Water source distance more than 30 minutes	0.948	0.105	1.033	0.688	1.015	0.224	1.008	0.170	1.133	0.031
Large scale environmental	Aridity Index	0.854	0.402	0.904	0.729	0.919	0.006	0.881	<0.001	0.903	0.121
	Sand/soil fraction	1.962	0.033	1.444	0.450	1.002	0.977	0.798	<0.001	0.973	0.899
Poverty	Stable night light (1997)	0.293	0.200	10.257	0.371	1.311	0.003	0.807	0.003	0.397	0.414

Covariate type	Covariates	Malawi		Mozambique		Nigeria		Sudan		Uganda	
		OR	p-value	OR	p-value	OR	p-value	OR	p-value	OR	p-value
Trachoma transmission	TF prevalence	1.486	0.001	1.739	<0.001	1.196	<0.001	1.638	<0.001	2.157	0.163
WASH	Latrine defecation	0.835	0.034	1.082	0.001	1.051	<0.001	0.983	0.410	1.104	0.470
	Improved latrines	0.830	0.116	0.973	0.412	0.976	<0.001	0.998	0.963	1.259	0.020
	Improved water source	1.061	0.111	1.019	0.285	1.011	0.078	1.008	0.644	0.979	0.883
	Improved water source on property	1.026	0.942	0.852	0.244	0.944	<0.001	0.870	0.151	0.000	0.055
	Water source on property	0.916	0.792	1.135	0.283	1.023	0.023	1.052	0.628	812.073	0.058
	Water source distance more than 30 minutes	0.992	0.844	1.015	0.446	0.993	0.406	0.981	0.318	1.283	0.263
Large scale environmental	Aridity Index	0.978	0.779	0.902	0.070	0.565	<0.001	2.293	<0.001	2.095	0.400
	Sand/soil fraction	1.112	0.526	1.047	0.551	0.977	0.438	0.950	0.457	0.540	0.226
Poverty	Stable night light (1997)	1.538	0.640	2.160	0.001	2.121	<0.001	1.295	0.265	16.041	0.259

The nested mixed effects models in Table 5.4 show that when TF was added in as a fixed effect, the proportional reduction in variance (taken as one minus the variance explained by M1 over variance explained by M2) ranged from 0.06 (in Nigeria) to 0.42 (in Benin). When environmental risk factors were added, the proportional change in variance (taken as one minus the variance explained by M1 over variance explained by M3) ranged from 0.25 (in Ethiopia) to 0.71 (in Cote d'Ivoire). In all countries, variance continued to decrease as TF and then environmental risk factors were added to the model. Full R outputs can be found in Appendix I.

Table 5.4 Comparison of variance explained by each mixed effects model.

			Null model	TF only model	TF prevalence + risk factors model
Benin	**Variance**		**2.25**	**1.3**	**1.17**
	Proportional reduction	Null model		0.42	0.48
		TF only model			0.10
Cote d'Ivoire	**Variance**		**13.75**	**10.28***	**2.95**
	Proportional reduction	Null model			0.71
		TF only model			
DRC	**Variance**		**1.52**	**1.23**	**1.05**
	Proportional reduction	Null model		0.19	0.31
		TF only model			0.15
Ethiopia	**Variance**		**1.22**	**1.01**	**0.92**
	Proportional reduction	Null model		0.17	0.25
		TF only model			0.09
Guinea	**Variance**		**1.88**	**1.87***	**1.66**
	Proportional reduction	Null model			0.11
		TF only model			
Malawi	**Variance**		**2.17**	**1.96**	**1.73**
	Proportional reduction	Null model		0.1	0.20
		TF only model			0.12
Mozambique	**Variance**		**4.45**	**3.52**	**3.08**
	Proportional reduction	Null model		0.21	0.31
		TF only model			0.13
Nigeria	**Variance**		**1.93**	**1.82**	**0.99**
	Proportional reduction	Null model		0.06	0.49
		TF only model			0.46
Sudan	**Variance**		**1.98**	**1.85**	**1.18**
	Proportional reduction	Null model		0.07	0.40
		TF only model			0.36

The best models selected using the likelihood ratio test are shown in Table 5.5. DRC, Ethiopia and Nigeria maintained the largest number of variables significant at the 5% level. These three countries also had the largest quantities of data available.

Table 5.5 Relative increase in odds derived from a multivariate binomial logistic model where community-level prevalence of TT in adults aged ≥15-years is dependent on a 10% increase in community-level prevalence of TF in children aged 1–9 years. Community-level household prevalence of improved sanitation and hygiene facilities as well as gridded covariates were included along with community-level TF.

Covariate type	Covariates	Benin		Cote d'Ivoire		DRC		Ethiopia		Guinea	
		OR	p-value	OR	p-value	OR	p-value	OR	p-value	OR	p-value
Trachoma transmission	TF prevalence	1.834	<0.001	1.417	0.187	1.566	<0.001	1.227	<0.001		
WASH	Latrine defecation					0.895	<0.001	1.045	<0.001		
	Improved latrines										
	Improved water source							1.014	0.005		
	Improved water source on property										
	Water source on property										
	Water source distance more than 30 minutes									1.102	0.048
Large scale environmental	Aridity Index					0.911	0.001	0.879	<0.001		
	Sand/soil fraction							0.795	<0.001		
Poverty	Stable night light (1997)					1.409	<0.001	0.806	0.003		

Covariate type	Covariates	Malawi		Mozambique		Nigeria		Sudan		Uganda	
		OR	p-value	OR	p-value	OR	p-value	OR	p-value	OR	p-value
Trachoma transmission	TF prevalence	1.481	<0.001	1.721	<0.001	1.193	<0.001	1.640	<0.001		
WASH	Latrine defecation	0.819	<0.001	1.085	<0.001	1.056	<0.001				
	Improved latrines					0.977	<0.001			1.251	0.006
	Improved water source										
	Improved water source on property					0.955	0.001				
	Water source on property					1.018	0.039				
	Water source distance more than 30 minutes										
Large scale environmental	Aridity Index					0.570	<0.001	2.752	<0.001		
	Sand/soil fraction										
Poverty	Stable night light (1997)			2.155	0.001	2.093	<0.001				

Semi-variograms generated with the Pearson's residuals of the best fitting non-spatial binomial model suggest presence of residual spatial correlation in DRC, Ethiopia, Mozambique and Sudan. The 95% confidence intervals generated under the assumption of spatial independence demonstrate spatial correlation in these countries (Figure 5.2).

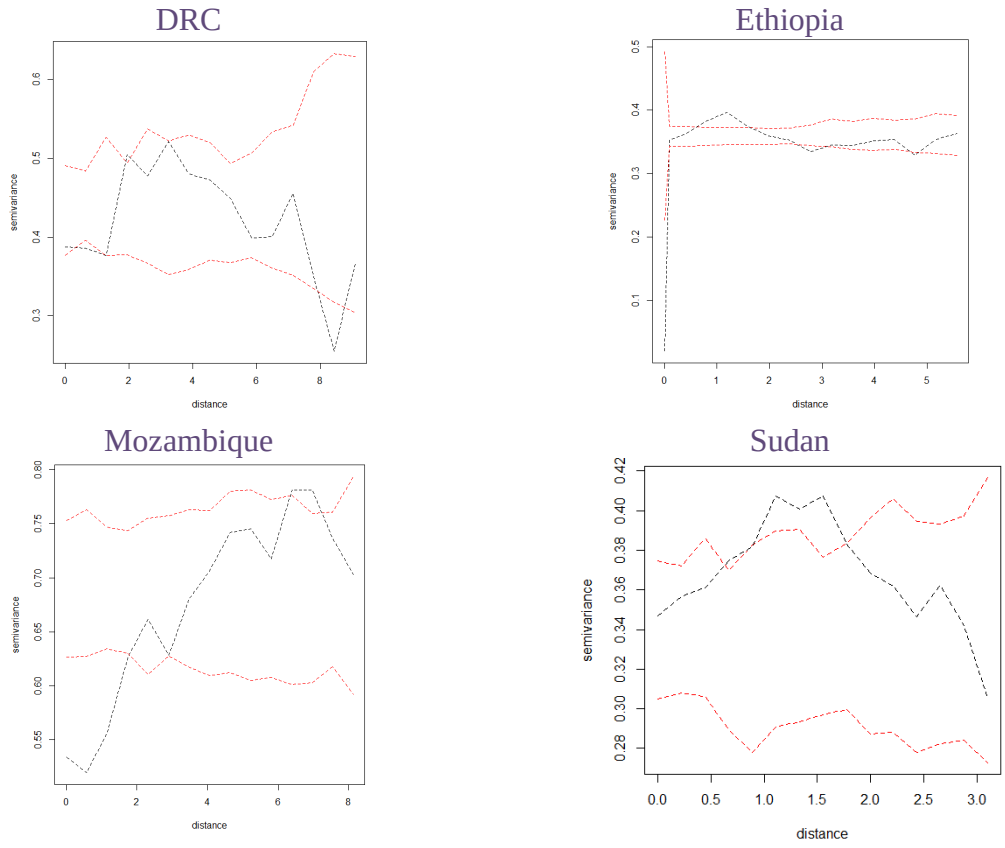


Figure 5.2 Semi-variograms were generated with the Pearson's residuals derived from the best fitting non-spatial mixed methods model. The 95% confidence intervals (red dashed lines) and semi-variogram (black dashed line) created through generating 1,000 simulations are displayed here. The x-axis was determined through taking half the maximum distance between communities and grouping into 15 classes of distance (bins). All distances are in kilometres.

The distance at which spatial correlation fell below 5% (taken as the minimum values of the variograms for the simulated data), also known as the semi-variogram range, spanned from 3.0 km (in Ethiopia) (95% credible interval 1.6-6.0 km) to 14.2 km (in Mozambique) (95% credible interval 3.4-58.9 km), corresponding with a very rapid decline in spatial correlation with distance at larger scales, after accounting for covariates (Table 5.6). The calculated distance

at which spatial correlation fell below 5% corresponds with the plots (Figure 5.2) when accounting for the wide 95% confidence intervals (Table 5.6). The spatial correlation in these four countries is larger than the minimum distance between communities and so I consider there to be evidence of residual spatial correlation.

Table 5.6 Scale of community-level TT prevalence spatial correlation in kilometres when accounting for covariates significant at the 5% level, by country with 95% confidence intervals

Country	scale of spatial correlation	95% confidence intervals
DRC	7.7 km	3.0 - 20.1 km
Ethiopia	3.0 km	1.6 – 6.0 km
Mozambique	14.2 km	3.4 – 58.9 km
Sudan	2.8 km	0.9 - 10.8 km

The large likelihood ratio test statistics lead to small p-values, suggesting that the geostatistical models are a better fit than the best fitting binomial mixed effect models in all four countries (Table 5.7).

Table 5.7 Likelihood ratio test values

Country	Likelihood ratio test statistic	p-value
DRC	2802.412	<0.0001
Ethiopia	12053.222	<0.0001
Mozambique	1805.792	<0.0001
Sudan	1147.988	<0.0001

Whilst the geostatistical models are better fitting than the mixed effect models, they are still very limited. The observed values plotted against the predicted values (Figure 5.3) demonstrate low predictability in the models. In DRC, Ethiopia, Mozambique and Sudan the percent of observed TT prevalence values that fell within the 95% confidence intervals of the predicted values were 12.4%, 38.1%, 1.0%, and 3.4% respectively.

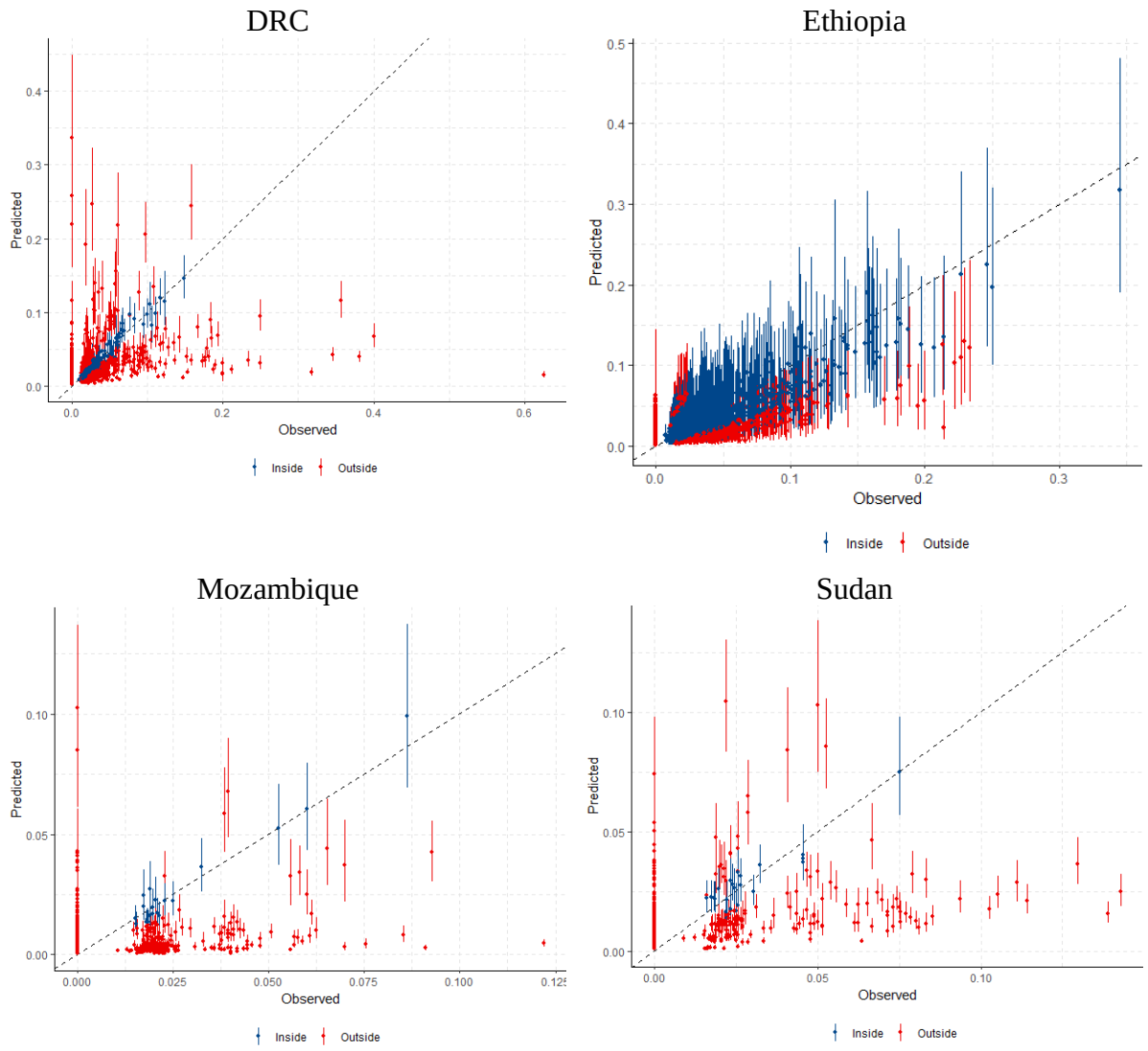


Figure 5.3 Plot of observed vs. predicted TT prevalence values as assessment of fit for the geostatistical models. The blue values fall within the 95% confidence intervals of the predicted values and the red values fall outside of the 95% confidence intervals.

5.6 Discussion

Remarkably few studies have specifically examined TT in the context of environmental risk factors and spatial distribution [67, 68, 146]. Here, I utilise the uniquely expansive and standardised GTMP datasets covering ten countries to demonstrate a large degree of heterogeneity in the relationship between community-level TT and TF, which I use as a proxy for contemporary transmission. The models also suggest that whilst community-level TF prevalence is the strongest predictor of TT, it does not fully explain the variation in community-

level TT prevalence, thus implying that occasionally high TT communities will be found where TF is low. However, associations with other potential risk factors are inconsistent, suggesting that the clinical history of trachoma may vary substantially between settings and emphasising the importance of understanding local context when designing interventions for at-risk populations. It is important to consider that even the better fitting geostatistical models demonstrate a great deal of uncertainty and so whilst informative, these models should not be overly interpreted.

Many studies have identified potential risk factors for the occurrence of TF, including socio-economic, demographic and environmental correlates [32, 51, 207, 219], and have explored its spatial distribution at different geographic scales [11, 37, 52, 220, 221]. The minimal studies that have addressed spatial distribution of TT however were limited in the amount of available data and so had constraints on generalizability of their outcomes [67, 68, 146]. The models presented here, with large datasets across ten countries reached similar conclusions and so provide some additional validation to this previous work.

I observe that whilst a 10% increase in community-level TF prevalence leads, on average, to an increase in the odds for TT of 55%, this varies from as little as 20% for Nigeria to 86% for Cote d'Ivoire. There are several potential explanations for the heterogeneous associations observed between TT and TF. High odds ratios such as those seen for Benin, Cote d'Ivoire and Mozambique, lead me to speculate that progression from TF to TT in these communities is unusually quick, or that trachoma transmission situation has not drastically changed over time. The relatively low odds ratios in Ethiopia and Nigeria, where a 10% increase in community-level TF prevalence is associated with an increase in TT odds of 23% and 20% respectively, suggest either very slow rates of progression, or a reduction in transmission intensity over time. Alternatively, both scenarios could be attributed to historic population movement.

In addition to differing magnitudes of association, for all countries, adjusting for TF leaves a substantial proportion of variance in TT unexplained. For example, the proportion of variance explained by TF ranges from as little as 6% in Nigeria, to 42% in Benin. Environmental covariates, on average, explain an additional 9% (in Ethiopia) to 46% (in Nigeria) of variance. It has been widely observed that dry conditions, which would be consistent with a low aridity index, are risk factors for TF in children [11, 38, 50, 222]. I found this same association for TT in three of the countries, where there was a negative association with aridity. However, I did observe an unexpected positive association between community-level TT prevalence and aridity index in Sudan. Interestingly, a 1997 study in Mali also found a conflicting relationship between TF and TT prevalence [32], with the authors speculating this may be attributed to dry conditions in the north contributing to TF and the humid environment in the south contributing to blinding complications. It has been shown that coinfection with other bacteria [223], such as *Streptococcus pneumoniae* and *Staphylococcus aureus* [66], could affect progression of TT [80]. These coinfections may be more prominent in humid climates.

Latrine defecation, aridity index and night light maintain strong significance in three of the ten countries. This, sometimes counter-intuitively, suggests that hygiene practices, dry/hot climate and historic infrastructure may contribute to increased community-level TT in some settings, but generally they do not. Previous study have clearly shown the association between WASH and TF [32, 51, 207, 219] and so it is not surprising that our models, which account for TF prevalence, generally do not maintain significant associations between TT and WASH. The variation in direction of association may be an artefact of WASH improvements over time or in fact the latrines themselves may contribute to fly population if not appropriately maintained. Further, it is possible that the fly population is not influencing recurrent transmission [60]. Current TT is caused by many previous infections and so areas that historically had poor

WASH may still have TT even though the situation has improved. Other variations in identified covariates can mostly be explained by country context.

There may be several explanations for the inconsistency of association between large-scale indicators and community-level TT seen between countries. This modelling approach does not capture population movement. Migration could certainly play a role in the geographic distribution of TT. It is also important to note that different ethnic groups may have different progression rates to TT. For example, a study in The Gambia found the polymorphism in the TNF-α gene promoter was associated with scarring and more frequent among Mandinkas than other ethnic groups [224].

I observe residual spatial correlation in four countries (DRC, Ethiopia, Mozambique, and Sudan), suggesting in the remaining countries that there are no outstanding large-scale environmental factors influencing the progression to TT. In the geostatistical models, I identified a very rapid decline in spatial correlation with distance at larger scales after accounting for covariates. This suggests that clustering of TT occurs only at small spatial scales.

Whilst these findings are not generalizable across country context, they can provide general direction for where to initiate case finding activities. As has been found in the Guinea Worm eradication experience, active surveillance and case finding will be essential as trachoma elimination nears [135] and these activities become more expensive as prevalence drops [225]. This uniquely large and standardised analysis provides important insight into the variation in community-level TT distribution and identifies substantial variation in the relationship between community-level TF and TT prevalence. However, the many differences seen across counties in these models suggest that it is unwise to rely on modelling on its own when undergoing TT

case finding. These models may serve as a starting point but should not replace current case finding techniques. For some countries important environmental risk factors were identified which can be used to inform targeting of active case finding through providing insight on where TT cases are more likely to be found in these countries. The findings suggest that in some countries it may be possible to inform strategic location of TT management services, improving efficiency in the end-stage of trachoma elimination.

The geostatistical models presented here identified a very rapid decline in spatial correlation, suggesting small clusters. In the next chapter I will look more closely at the clustering of TT and identify communities containing a disproportionate amount of TT (hot spots). I will then follow methods similar to those described here where I will identify large scale environmental covariates associated with these hot spots.

Chapter 6: **Identifying trachomatous trichiasis (TT) hot spots**

6.1 Overview

As estimated in Chapter 3, the number of TT cases requiring management in 2016 was an estimated 2.8 million. The logistical challenge of finding these positive cases is a serious issue country programmes face when planning for and implementing TT interventions. It is necessary to strategically place surgical clinics and resources to ensure those who need surgery have access to it. One way to do this is through identifying and describing TT hot spots, or areas of higher-than-expected prevalence.

Whilst, in Chapter 5, I built geostatistical models identifying a rapid decline in spatial correlation, in this chapter I will further explore the spatial structure of community-level TT through identifying hot spots. I conducted the analysis presented here and built the geostatistical models. I also wrote the manuscript informed by the information presented in this chapter.

6.2 Background

TF is the clinical sign for identifying active trachoma transmission, an outcome of repeated *Chlamydia trachomatis* infection [19]. Recurrent infection leads to chronic inflammation of the tarsal conjunctiva [21]. Over time, scarring can lead to an everted eyelid where eyelashes touch the globe, TT. This trauma may lead to corneal opacity and blindness [21]. Quality surgery can correct the in-turned eyelid, relieving pain and stopping progression of TT [74, 77, 78]. For less severe cases, epilation is often used as a management technique [75]. The logistical challenge of finding positive TT cases is a serious issue country programmes face when planning for and implementing TT interventions. It is necessary to strategically place

surgical clinics and resources to ensure those who need interventions have access to them. The identification of TT hot spots could be used to guide decision making around placement of TT management resources.

Here, the term hot spot is used to describe a group of communities with a TT prevalence higher than expected from spatial randomness. It is well established that *C. trachomatis* is transmitted through close physical contact, with transmission highest between family members and close communities [32-42], indicating clustering at local scales. Given the role of sanitation and hygiene in TF transmission [32, 41, 45-48] and differences in access and practices, between community clustering is also expected. Chronic TF can lead to scarring which may results in TT and so it is sensible to anticipate a similar clustered phenomenon for TT.

Clustering of environmental risk factors or increased TF prevalence may contribute to clustering of high TT prevalence communities. As observed in the previous chapter, adjusting for TF in the binomial mixed models leaves a substantial proportion of variance in TT unexplained. In some countries, this variance was further explained by environmental covariates. It is reasonable to expect that a similar relationship will be observed when TT hot spots are the dependent variable.

Whilst disease distribution is heterogeneous at various scales, few spatial explicit analyses have been conducted for TT. A 2015 study examined the spatial relationship between TT and CO in Nigeria and determined risk associated at both individual and cluster-levels. In this evaluation 14% of variance was attributed to the cluster-level [146]. Whilst there are no publications where TT hot spots are identified, computations have been conducted for other diseases, such as TB where local indictors of spatial association (LISA) indicated spatial clustering of high response to or elevated rate of TB infection and Getis-Ord Gi defined the hot spots [150].

This study seeks to identify community level TT hot spots and determine if these can be attributed to environmental risk factors or current trachoma transmission (TF prevalence). During the end-game of trachoma elimination, TT prevalence is expected to be very low (approaching 0.2% in adults aged ≥15-years) [133, 162], making it difficult to find and offer services to those affected. Hot spot detection and a stronger understanding of spatially distributed risk factors could guide strategies and improve efficiency in case finding.

6.3 Methods

6.3.1 Data

The data used in this analysis were collected through the GTMP [71] and derived from georeferenced raster files converted to a standardised resolution of 5 km by 5 km (Appendix G). These data were described in detail previously (5.3). Briefly, ten of the 29 GTMP countries made their data available for this analysis, namely; Benin, Cote d' Ivoire, Democratic Republic of the Congo (DRC), Ethiopia, Guinea, Malawi, Mozambique, Nigeria, Sudan and Uganda. These GTMP data include information on TT, TF and WASH representing 15,051 communities in 624 EUs. The georeferenced raster files provided data on aridity [208], sand/soil fraction [209] and night light density [210, 211].

6.3.2 Hot spot detection

To test for the existence of spatial heterogeneity and spatial autocorrelation a Moran's I statistics was run. The Moran's I statistic measures spatial autocorrelation, or the similarity between observations as a function of space. Global Moran's I can be used as a cluster detection technique, which quantifies similarities among spatial related areas [124]. The Moran's I

statistics is described with, n = number of spatial united indexed by i and j; Z = variable of interest; W_{ij} = measure of closeness of areas i and j.

$$I = \frac{n\Sigma_i\Sigma_j W_{ij}(Z_i - \bar{Z})(Z_j - \bar{Z})}{(\Sigma_i\Sigma_j W_{ij})\Sigma_k(Z_k - \bar{Z})^2}$$

A Moran's I analysis was performed for each country using the software package GeoDa [226]. Distance weighting was generated as Euclidean distance and k-nearest neighbours (KNN). To mimic a queen selection (Figure 6.1), the K value was taken as 8.

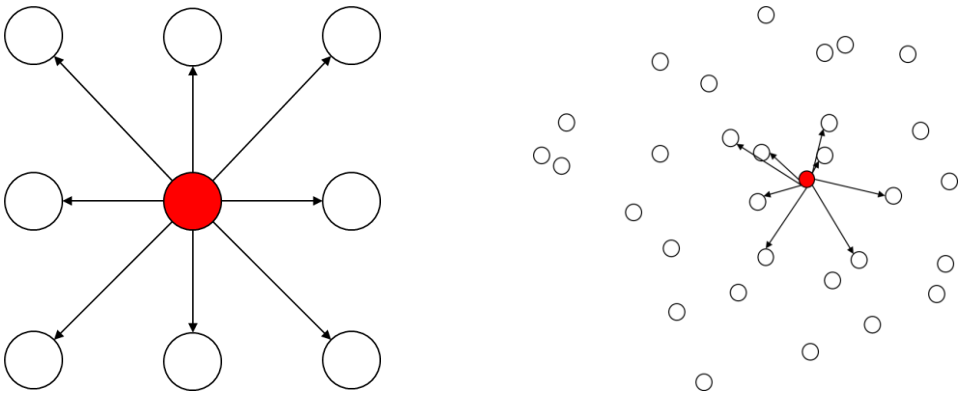


Figure 6.1 Illustration of a k-nearest neighbour of eight

To further understand the discreet spatial structure of TT, a Local Moran's I statistic was used. The local indicators of spatial association (LISA) technique evaluates clustering of individual units [227].

Spatial clustering of disease can be measured through global or local methods [132]. Global Moran's I determine whether clustering is present throughout the study area by providing a single statistical measure of the degree of spatial clustering; identifying if spatial clustering exists. The LISA method defines the location and extent of clustering.

A LISA statistic classification of high-high or low-low, indicate positive spatial autocorrelation and are often considered spatial clusters. High-low and low-high classifications indicate negative spatial autocorrelation and are often considered spatial outliers. A high-high classification for this analysis indicates that the proportion of TT in surrounding communities is high and the value in the specified community is significantly higher than the surrounding communities. A low-low classification in this analysis indicates that the proportion of TT in the surrounding communities is low and the value in the specified community is significantly lower than the surrounding communities. A low-high classification indicates the surrounding communities have a high proportion of TT and the specified community has a significantly lower proportion. Finally, a high-low classification indicates the surrounding communities have a low proportion of TT and the specified community has a significantly higher value. Here I consider communities classified as high-high to be hot spots. Because the data used in this analysis were derived from random sampling, it is unlikely that by random chance only neighbouring homogeneous communities were selected. Therefore, a high prevalence community surrounded by other high prevalence communities may be a signal of on outlier of public health importance. Through identifying these hot spots, case finding could practically start with the hot spot community and then move to the eight nearest neighbour surrounding communities.

A 2015 study by Lietman *et al.* determined that when prevalence of ocular chlamydial infection is decreasing in a district, high prevalence communities will be present and are not expected to interfere with progress [148]. These communities would be considered high-low in this analysis and so are considered as isolated high-prevalence communities that are not of major public health importance. The LISA statistics for community level TT proportion was generated in GeoDa [226].

A boxplot was created to explore the association between TT hot spots and community-level TF prevalence, as a proxy for *C. trachomatis* transmission intensity.

6.3.3 Model formulation

The models use community-level data, however because I use random effects U_i to represent the unexplained extra binomial variation between communities these are defined as binomial mixed models [215]. Let p_i denote the probability of being a TT hot spot (binary high-high communities = 1 and other communities = 0), β_0 is the intercept and U_i is community-level unstructured random effects (let i denote the i-th community). I then fit the following nested binomial mixed models, where $\gamma TF_i$is the regression coefficient for the effect of TF prevalence on the log-odds of TT hot spots:

$$M1: log\ (\frac{p_i}{1-p_i}) = \beta_0 + U_i;$$

$$M2: log\ (\frac{p_i}{1-p_i}) = \beta_0 + \gamma TF_i + U_i;$$

$$M3: log\ (\frac{p_i}{1-p_i}) = \beta_0 + \gamma TF_i + \sum_{j=1} \beta_j d_{ij} + U_i,$$

where in M3; β_j are the regression coefficients of the explanatory variables d_{ij} and d_{ij} are the explanatory variables described in the previous section with ij as the j -th variable within the i-th community.

In fitting M3, I also carried out variable selection using a backward stepwise approach starting from the mixed effect model with all the variables included. The models were fitted in PrevMap using the Laplace approximation of the lme4 package [216]. The likelihood ratio test was used

to test for the significance of each of the variables which were removed until all were significant at 5% level.

6.4 Results

The Moran's I results demonstrate evidence of weak spatial autocorrelation for TT, with the evidence slightly stronger in larger settings, namely Ethiopia and Nigeria. The scatterplots illustrate the autocorrelation of each community, with positive spatial autocorrelation in the upper right and lower left quadrants and negative spatial autocorrelation in the lower right and upper left quadrants (Figure 6.2). The Moran's I statistic is visualized as the regression slope of the plot and is estimated as 0.33 (p-value <0.001 with 9,999 permutations).

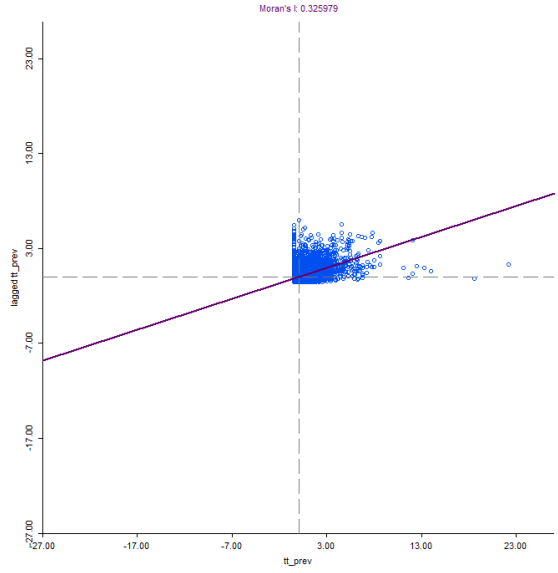


Figure 6.2 Community level TT Moran's I scatter plots, community level TT proportions are on the x axis and spatial lag is on the y axis. The slope of the regression line (purple) is the Moran's I statistic using k nearest neighbour (k value of 8).

The output of the LISA statistic identified 712 communities classified as TT hot spots (high-high), 9,999 permutations and p-value of <0.01 (Figure 6.3).

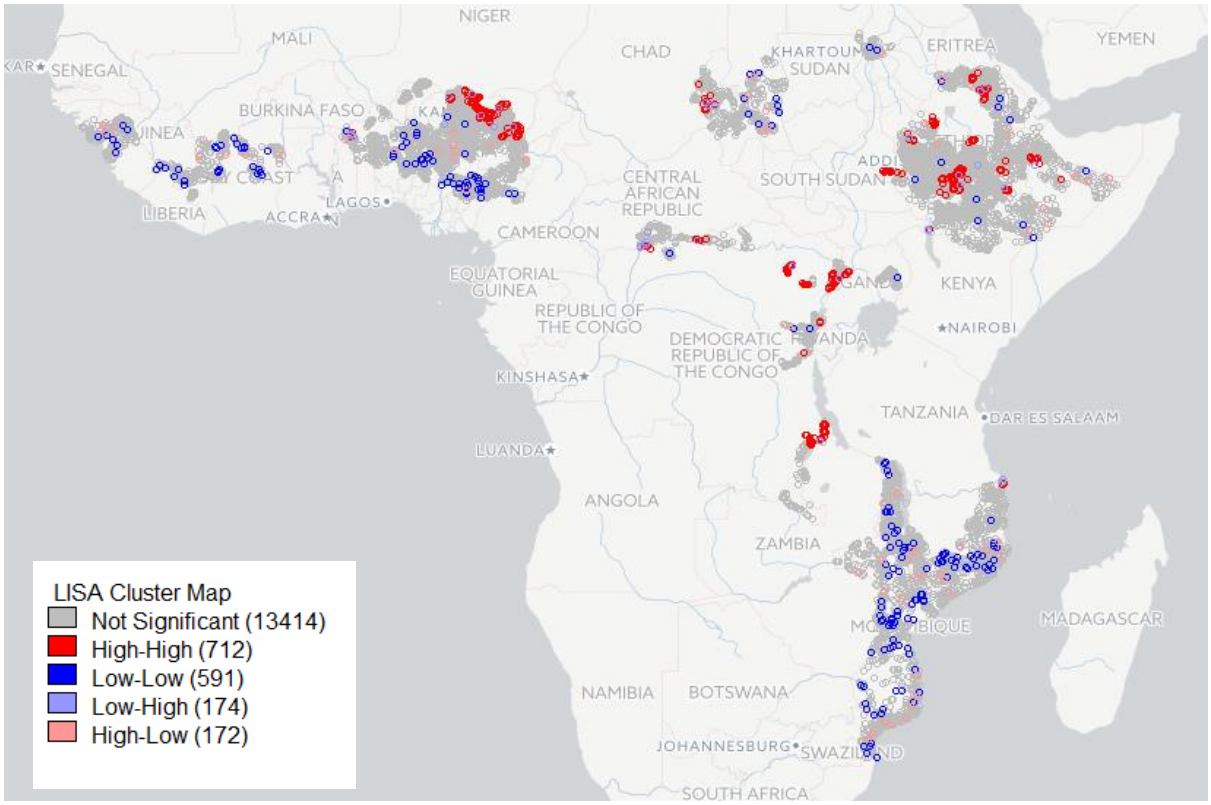


Figure 6.3 LISA TT prevalence results where significance is set to 0.01 with 9,999 permutations.

Zero hot spots were identified in Cote d'Ivoire, Guinea, Malawi and Uganda (Table 6.1). The mean number of persons included in the TT estimate per hot spot is 58. The largest proportion of TT hot spots (where the denominator is the total number of communities included in the survey within each country), were in DRC (0.13), Ethiopia (0.06) and Nigeria (0.05).

Table 6.1 Community hot spots identified through LISA analysis where the community was classified as high-high, with a p-value <0.01

Country	Number of hot spots	Community-level TT prevalence (%)		
		Mean	Range	SD
Benin	4	6.2	3.2-14.0	5.2
Cote d'Ivoire	0	-	-	-
DRC	137	7.5	1.5-62.5	7.2
Ethiopia	272	7.0	1.4-34.5	5.3
Guinea	0	-	-	-
Malawi	0	-	-	-
Mozambique	2	2.3	2.1-2.6	0.3
Nigeria	280	5.6	1.4-21.8	3.7
Sudan	17	3.7	1.9-7.7	2.1
Uganda	0	-	-	-

Generally, communities classified as TT hot spots also seemed to have an increased proportion of TF in comparison to the communities not classified as hot spots (Figure 6.4). This is particularly evident in DRC and Ethiopia. The model presented in this chapter tests for significance to determine association.

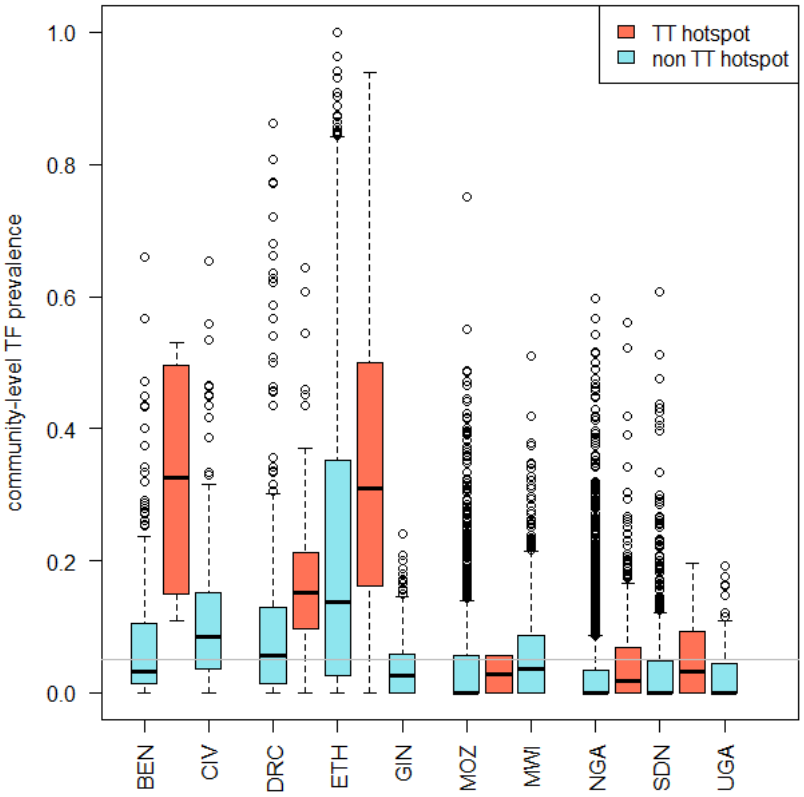


Figure 6.4 Boxplot illustrating the distribution of TF proportion among hot spot and non-hot spot communities

The proportion of identified TT positive individuals living in TT hot spots communities ranged from 21.6% (in Nigeria) to 37.3% (in DRC) and the proportion of communities identified as hot spots ranged from 5.2% (in Nigeria) to 13.3% (in DRC). On average, 24% of cases were found in 6.3% of communities.

Because so few hot spots were identified in many of the countries, only DRC, Ethiopia and Nigeria were retained for the remaining analysis. In the mixed effect model containing all covariates (M3), there was variation in strength of association between countries, with

community-level TF prevalence only significant (P<0.05) in DRC and Ethiopia (Table 6.2). However, aridity index was signficiant in all three countries (P<0.05). In two countries (Ethiopia and Nigeria (P<0.05)) evidence of association was found in a number of WASH indicators (improved water source, latrine defication, and location of water source), sand/soil fraction and night light.

The nested mixed effect models did not show a proportional reduction in variance (taken as one minus the variance explained by M1 over variance explained by M2) when TF was added to the model as a fixed effect. There was also no proportional reduction in variance (taken as one minus the variance explained by M1 over variance explained by M3) when the environmental risk factors were added as fixed effects. In all countries variance was mostly explained by the community-level random effect.

Through a backward stepwise approach variable selection resulted in unique models for each country (Table 6.2). Community-level TF prevalence, aridity index, and night light were the only significant covariates in DRC. In Ethiopia, community-level TF prevalence, latrine defecation, improved water source, aridity index, sand/soil fraction and night light were initially signficiant. However, when unsignificant covariates were removed, the WASH indcators lost their significance. In Nigeria, all covarites were significant other than distance to water.

Table 6.2 Relative increase in odds derived from a multivariate binomial logistic model where TT hot spot classification is dependent on a 10% increase in community-level prevalence of TF in children aged 1–9 years. Community-level household prevalence of improved sanitation and hygiene facilities as well as gridded covariates are included along with community-level TF.

Covariate type	Covariates	DRC			Ethiopia			Nigeria		
		OR	CI (95%)	p-value	OR	CI (95%)	p-value	OR	CI (95%)	p-value
Trachoma transmission	TF prevalence	1.468	1.468-1.469	<0.001	1.276	1.087-1.497	0.0029	1.013	1.013-1.013	<0.001
WASH	Latrine defecation							1.059	1.059-1.059	<0.001
	Improved latrines							0.981	0.98-0.981	<0.001
	Improved water source							1.063	1.062-1.063	<0.001
	Water source distance more than 30 minutes									
Large scale environmental	Aridity Index	0.778	0.778-0.778	<0.001	0.007	0.007-0.007	<0.001	0.332	0.332-0.332	<0.001
	Sand/soil fraction				0.263	0.263-0.263	<0.001	1.393	1.393-1.394	<0.001
Poverty	Stable night light (1997)	1.040	1.039-1.04	<0.001	0.919	0.919-0.919	<0.001	1.188	1.187-1.188	<0.001

In DRC and Ethiopia, the likelihood ratio test statistics were significant when community-level TF alone was included in the model as a fixed effect (Table 6.3). However, in Nigeria both the model with community-level TF alone and the model with additional environmental risk factors showed significant test statistics. To determine which between these two models was the “best fitting” model I examined the AIC value for the two models. The AIC value for the model containing community-level TF along with the significant environmental risk factors had the lowest AIC and so was taken as the “best fitting” model for Nigeria.

Table 6.3 Likelihood ratio test values from the different nested models, by country

Model	DRC			Ethiopia			Nigeria		
	AIC	Likelihood ratio test statistic	p-value	AIC	Likelihood ratio test statistic	p-value	AIC	Likelihood ratio test statistic	p-value
Null	811.6			2055.2			2202.8		
TF prevalence	429.4	384.2	<0.001	863.1	1194.1	<0.001	892.1	1312.6	<0.001
TF + significant environmental covariates	432.5	0.86	0.650	945.9	0	1	888.6	15.5	0.016
TF + all environmental covariates	735.2	0	1	1058.3	0	1	895.6	0	1

6.5 Discussion

If endemic countries can identify hot spots of TT, they can better target interventions to have maximum impact and thus move closer to the GET 2020 targets for eliminating blinding trachoma as a public health problem. Though it is important to note that hot spot detection should supplement and not replace current case finding techniques. In this analysis, half the countries did not have hot spots, yet their prevalences were often above the TT threshold. Whilst higher levels of TT might be expected in communities with higher active transmission, analyses of co-distributions have shown this is not always the case. This is the first multi-country study to explore the use of geospatial techniques to identify clusters of TT and used a comprehensive contemporary database of community-level pre-intervention TT prevalence data collected during the GTMP.

In 2015, Lietman *et al.* published a study where a mathematical model illustrates the geometric shape of ocular chlamydial infection distribution as the disease is disappearing in a district [148]. The model demonstrates a heavy tail post-treatment, suggesting that after treatment most ocular chlamydial infection goes away whilst some lingering cases will remain. If this model holds true in different settings, it provides confidence that a single cross-sectional survey demonstrates a geometric distribution of

infection. When district level prevalence is decreasing, the occasional high prevalence community is not expected to interfere with progress. These occasional communities are not trachoma hot spots.

It is important to distinguish key differences between this 2015 study and the work shown here. Lietman *et al.* aimed to show what happens to ocular chlamydial infection over time after intervention is in place. Ocular chlamydia is a sign of infection, interpreted as ongoing transmission, whilst TT is the morbidity stage of the disease. These are two different indicators and so the modelling is inherently different. Furthermore, a TT hot spot is likely to be fundamentally different before and after intervention. Before the **SAFE** strategy is implemented, communities containing more than expected proportions of TT are likely a result of disease ecology and poverty, whilst after **SAFE** implementation a TT hot spot may identify areas that are worst served by the health system or missed by surgical camps. This post-implementation scenario is very unpredictable and so here I focus on pre-intervention TT hot spots.

These methods loosely align with a 2012 study aiming to identify geospatial hot spots for TB in Beijing [150]. This study implemented a Moran's I statistic to identify locations of spatial autocorrelation and Getis-Ord Gi to identify hot spots. A hot spot was defined as an identified clustered pattern of high response to or elevated rate of TB infection. The Moran's I statistic showed it very unlikely that neighbouring values of TB are the result of random pattern. The peak Z score of the LISA was used to indicate spatial clustering and the clustered patterns identified by Getis-Ord Gi were interpreted as hot spots. These methods identified areas of clustered higher than expected disease. Whilst Lietman *et al.* warn an occasional high prevalence, community is not a hot spot,

here I suggest that a *cluster* of communities with higher than expected proportions (LISA high-high) of TT might be.

This study demonstrates stronger spatial autocorrelation in larger settings (i.e. DRC, Ethiopia and Nigeria), which may be an outcome of larger amounts of data available in these settings. The proportion of identified TT positive individuals living in TT hot spots communities ranged from 21.6% (in Nigeria) to 37.3% (in DRC) and the proportion of communities identified as hot spots ranged from 5.2% (in Nigeria) to 13.3% (in DRC). With an average of 24% of cases found in 6.3% of communities, hot spot communities make an important public health impact.

The model only containing TF prevalence as a fixed effect was the best fit in DRC and Ethiopia, suggesting in these two settings, TF is predictive of TT hot spots. However, in Nigeria, the model including a series of environmental covariates model was proven to be the best fit. These models do not behave as expected, with variation in direction of association between covariates between the countries. This further suggesting that TT hot spots cannot be accurately predicted through proxy indicators. It is possible that the identified hot spots are an artefact of population migration.

I found variation in both TF prevalence and environmental risk factors in all three countries. In DRC, the majority of the hot spots are located in the north-eastern districts of Pawa, Bambu, Nyarembe, and Angumu. Interestingly, these four districts have low TF prevalence and high TT prevalence. Similarly, in Nigeria hot spots were generally identified in areas of low TF prevalence. Whereas, in Ethiopia, the district-level TF and TT prevalences are both high in districts containing hot spots.

Less is known about the trachoma migration dynamics in DRC. Though, in Nigeria, hot spots are concentrated in the northern and north-eastern portion of the country and there appears to be a pattern around the border of Jigawa region and along the southern border of the Zinder region of Niger. The bordering districts in Jigawa (Ringim and Garki) conducted a trachoma survey in 2007 [192] which triggered implementation of the full **SAFE** strategy. Cross-border dynamics could be influencing transmission, explaining low TF prevalence and the high TT may be a reflection of historic transmission.

As demonstrated here, population dynamics are difficult to predict, and the inconsistent association of environmental risk factors further complicate the understanding of TT hot spot distribution. It is additionally, inadvisable to solely rely on district-level TF and TT prevalence for strategizing. This analysis demonstrated that hot spots only account for 25% of cases and the remaining cases are widely distributed. Whilst, simply identifying these hot spots as a standard step in survey data analysis could provide insight into the specific situation in different settings a large portion of cases may be missed.

6.6 Limitations

It is important to acknowledge the limitations of identifying hot spots from trachoma survey data. Trachoma surveys collect data on TT incidents through a two stage random sampling technique, where first communities are randomly selected and then households within those communities [22]. Hence, it is possible, by random chance, a larger proportion of high prevalence communities or households were included in the survey. Our models also do not account for migration patterns, which may play a large role in TT distribution.

Chapter 7: **Summary and discussion of findings**

To provide evidence in guiding strategic use of resources for implementing TT management campaigns the aims of my thesis are to **quantify the current global distribution of TT**, and to **explore spatial heterogeneity in TT at fine spatial scales**. Through this work I have shown that TT distribution is highly spatially heterogeneous. In some settings this heterogeneity is driven by TF prevalence, use of latrines for defecation, aridity index and night light (as a proxy for infrastructure); whilst in most settings it is not possible to identify the driving influences of spatial heterogeneity. For trachoma control programmes looking to target interventions and move towards certification of elimination, it is essential that robust strategies (like the TT specific survey designed as part of this thesis) are used to comprehensively assess burden in a non-biased way across implementation sites. This work also shows that spatial clustering is important, but that hot spots only account for 25% of cases and the remaining cases are widely distributed, so simply targeting hot spots risks missing a large portion of affected individuals.

7.1 Discussion of findings

Trichiasis remains a significant public health problem in many countries, with a global backlog estimated for 2016 at 2.8 million people, 61% of whom live in sub-Saharan Africa. The considerable reduction from the 2009 estimate of 8.2 million [163] is likely the result of a combination of factors. First, there are now more, and better data derived from rigorous surveys. Second, in many countries, there has been an impressive programmatic scale up to manage TT, conducted by a complex network of governments and their partners. Third, there is likely to be an effect on the incidence of TT from the

intensive efforts to reduce active trachoma prevalence in many contexts; such efforts have been ramping up in endemic countries since the World Health Assembly's 1998 commitment to global elimination of trachoma as a public health problem [2]. Teasing out the relative contribution of each of these factors is not possible at the present time, but regardless of cause, the reduction is welcome news for global health.

An understanding of the backlog and distribution of trichiasis cases is necessary to effectively plan for surgical services and other components of management of individuals with trichiasis. To reduce TT prevalence in each district of each endemic country to <0.2% in adults aged ≥15-years, which is the defined elimination threshold for TT [162], at least 2.0 million people will need to have their TT appropriately managed. Precise data on the number of people affected with TT are essential to enable national control programmes to effectively target surgeries, to evaluate ongoing interventions, and to determine if elimination thresholds have been reached. To achieve these goals, it is necessary for robust strategies to be developed and implemented.

Here, I designed and validated such a strategy. The PBPS presented is designed to measure TT with precision against the elimination goals. Based on the evidence shown here, a TT specific survey targeting 30 clusters and examining at least 2,818 adults aged ≥15-years will enable programmes to estimate an expected TT prevalence of 0.2% with absolute precision of 0.2%. Results from the validation exercise suggest this design provides consistent prevalence estimates with reasonable precision for public health decision making.

A report of this survey design and recommendations was published by WHO with the accompanying protocol made available for use in national programmes in 2017 [172].

The outcomes from the use of this survey design thus qualify as evidence of elimination as a public health problem within the dossiers of elimination. It should be noted that this survey is meant to measure public health significance of TT, and not patient management. Programmatic survey approaches are useful in understanding the epidemiological status of diseases within a given area, but there is an inherent risk that patients may not be provided a direct link to appropriate health services. TT graders have been trained to refer positive TT cases to nearby facilities for management, however, currently there is no follow-up mechanism to provide support for patients to arrange for treatment. This strategy is not engaged within the routine health services and so identifying cases does not directly translate to cases receiving treatment.

It is important to note that, whilst this survey was designed for an expected prevalence of 0.2%, the TT prevalence estimates from all four validation districts were higher. This does potentially limit generalizability to very low prevalence settings. In Chapter 5, I illustrate the spatial heterogeneity of TT distribution, which may suggest that as prevalence decreases the design effect may increase. It may be, that after a certain threshold, population-based prevalence surveys using random sampling alone may not be appropriate and other strategies such as purposive sampling should be adopted.

Whilst previous work has identified clustering of TF [36-38, 52], there is limited understanding around the spatial structure of TT. As trachoma control programmes near the elimination targets it becomes difficult to find the “rare” occurring TT case that needs management. The elimination threshold for TT (district-level prevalence >0.2% in adults aged ≥15-years) equates to less than two positive TT cases per 1,000 adult population. Understanding the spatial structure of TT could guide where to conduct case finding activities, hopefully resulting in a more efficient use of resources.

Here I use community-level TF prevalence, information on WASH and large-scale environmental and socio-economic indicators to model the spatial variation in community-level TT prevalence. I modelled the prevalence of TT in a community and investigated its assocation with community-level TF prevalence. I carried out a spatial exploratory analysis using semi-variograms to identify spatial structure in the distribution of community-level TT prevalence. In countries where spatial structure was identified, it was used to inform a geostatistical model constructed with the TT prevalence data and a suite of explanatory risk factors.

This analysis found evidence of an association between community-level TF and TT prevalence. The estimated regression relationship between community-level TF and TT was significant at 5% level in eight out of ten countries. I estimate that a 10% increase in community-level TF prevalence leads to an increase odds for TT ranging from 20% to 86% when accounting for additional risk factors.

Whilst the models suggest community-level TF prevalence is the strongest predictor of TT, TF prevalence does not fully explain the variation in community-level TT prevalence. This implies that high prevalence TT communities will occasionally be found where TF is low. The models found inconsistency in association with other potential risk factors, implying that the clinical history of trachoma varies between settings and emphasising the importance of understanding local context when designing interventions. Furthermore, the geostatistical models identified the scale of spatial correlation to range from 3.0 km (Ethiopia) to 14.2 km (Mozambique). This rapid decline in spatial correlation indicates that clustering of TT occurs only at small spatial scales. The limited spatial correlation implies that a two-stage cluster sampling approach is appropriate for measuring TT, assuming that the appropriate number of

clusters are selected. In Chapter 4, through bootstrapping, I evaluated the stability of a 30-cluster survey. The estimates for 30 clusters and 60 clusters were closely aligned with only slightly wider confidence intervals. This confirms that in settings similar to those included in the validation, the traditional method of 30 randomly selected clusters provides sufficient variation within a district to measure TT prevalence.

In Chapter 3, I estimated 2.8 million TT cases requiring management are outstanding. However, many of these cases are found in generally low prevalence settings. The logistical challenge of finding these positive cases is a serious issue country programmes face when planning for and implementing TT interventions. It is necessary to strategically place surgical clinics and resources to ensure those who need surgery have access to it. One way to do this may be through identifying and describing TT hot spots, or areas of higher-than-expected prevalence.

A LISA statistic was used to classify communities as TT hot spots. A series of mixed effect models where TT hot spots are explained by community-level random effect, community-level random effect with community-level TF prevalence as a fixed effect, and community-level random effect with community-level TF prevalence and environmental covariates as fixed effects were run.

The output of the LISA statistic identified 712 communities classified as hot spots within 118 EUs, with the majority found in DRC, Ethiopia and Nigeria. In these countries 24% of TT cases were found in 6% of the communities (hot spots) and 100% of the hot spot communities were found in EUs above the elimination threshold. Of the positive cases identified, 99% were found in districts above the elimination threshold. The proportion of identified TT positive individuals living in TT hot spot communities

ranged from 21.6% (in Nigeria) to 37.3% (in DRC) and the proportion of communities identified as hot spots ranged from 5.2% (in Nigeria) to 13.3% (in DRC). There is evidence of strong association with community-level TF prevalence in DRC and Ethiopia, but not in Nigeria.

I found variation in both TF prevalence and environmental risk factors in all three countries. In DRC and Nigeria hot spots were generally identified in areas of low TF prevalence. Whereas, in Ethiopia, the district-level TF and TT prevalences are both high in districts containing hot spots. Cross-border dynamics could be influencing transmission in DRC and Nigeria, explaining low TF prevalence and the high TT may be a reflection of historic transmission. Population dynamics are difficult to predict, and the inconsistent association of environmental risk factors further complicate the understanding of TT hot spot distribution.

The large proportion of cases found in these relatively few hot spot communities suggest that the TT survey presented in Chapter 4 may risk either missing these communities or selecting many of these communities in the random sampling potentially biasing the outcomes. However, from a public health perspective and with the evidence shown in the oversampling exercise, this does not appear to be the case in the validation settings.

As more national trachoma control programmes approach the elimination targets it may be advisable to not solely rely on district-level TF and TT prevalence for strategizing. Through conducting surveys with geolocated data and running LISA (as described in Chapter 6), hot spot communities could be identified and serve as the starting point for case finding. Hot spot identification should not replace current case finding techniques

(as seen in Chapter 6 hot spots were only identified in half the countries and only account for 25%), but act as an additional detail useful for choosing where to begin case finding.

My thesis quantified the current global distribution of TT and explored spatial heterogeneity in TT at fine spatial scales. Through this work I have provided evidence and tools for national trachoma control programmes to use as they approach elimination of trachoma as a public health problem and apply for their dossier.

7.2 Future directions for trachoma programs

The future for trachoma elimination is promising. The trachoma community has made huge strides to meet the goals of elimination as a public health problem by 2020. However, there is still work to be done, especially around addressing the TT backlog. The TT specific survey design presented here has since been implemented in five countries and there are plans to expand its implementation to the remaining areas where a TT survey is necessary. The outcomes of these surveys will contribute to updating the TT backlog estimate, which will be calculated annually, following the methods described here, in preparation of the Global Elimination of Trachoma 2020 meetings. The backlog estimate provides a measurable target which is encouraging to funders facilitating the expanse of resources.

Understanding the geostatistical structure of TT provides guidance on where to focus resources for TT management. To bring district-level TT prevalence below the elimination threshold, in many areas it is necessary to conduct case finding activities. The TT hot spot model presented here should be further validated through field testing.

Trachoma can and will be eliminated as a public health problem and the methods and tools provided in this thesis can be used to advance this initiative.

References

1. World Health Organization. WHO Alliance for the Global Elimination of Blinding Trachoma by the year 2020: progress report on elimination of trachoma, 2013. Wkly Epidemiol Rec. 2014;89(39):421-8.

2. Global elimination of blinding trachoma. 51st World Health Assembly; May 16, 1998; Geneva: World Health Assembly; 1998.

3. Ekpo UF, Hurlimann E, Schur N, Oluwole AS, Abe EM, Mafe MA, et al. Mapping and prediction of schistosomiasis in Nigeria using compiled survey data and Bayesian geospatial modelling. Geospat Health. 2013;7(2):355-66. doi: 10.4081/gh.2013.92. PubMed PMID: 23733296.

4. Molla YB, Wardrop NA, Le Blond JS, Baxter P, Newport MJ, Atkinson PM, et al. Modelling environmental factors correlated with podoconiosis: a geospatial study of non-filarial elephantiasis. Int J Health Geogr. 2014;13:24. doi: 10.1186/1476-072X-13-24. PubMed PMID: 24946801; PubMed Central PMCID: PMCPMC4082615.

5. Woodhall DM, Wiegand RE, Wellman M, Matey E, Abudho B, Karanja DM, et al. Use of geospatial modeling to predict Schistosoma mansoni prevalence in Nyanza Province, Kenya. PLoS One. 2013;8(8):e71635. doi: 10.1371/journal.pone.0071635. PubMed PMID: 23977096; PubMed Central PMCID: PMCPMC3743764.

6. Flueckiger RM, Nikolay B, Gelderblom HC, Smith JL, Haddad D, Tack W, et al. Integrating data and resources on neglected tropical diseases for better planning: the NTD mapping tool (NTDmap.org). PLoS neglected tropical diseases. 2015;9(2):e0003400. doi: 10.1371/journal.pntd.0003400. PubMed PMID: 25654670; PubMed Central PMCID: PMCPMC4318578.

7. Cano J, Rebollo MP, Golding N, Pullan RL, Crellen T, Soler A, et al. The global distribution and transmission limits of lymphatic filariasis: past and present. Parasit Vectors. 2014;7:466. doi: 10.1186/s13071-014-0466-x. PubMed PMID: 25303991; PubMed Central PMCID: PMCPMC4197264.

8. Pullan RL, Brooker SJ. The global limits and population at risk of soil-transmitted helminth infections in 2010. Parasit Vectors. 2012;5:81. doi: 10.1186/1756-3305-5-81. PubMed PMID: 22537799; PubMed Central PMCID: PMCPMC3419672.

9. Pullan RL, Gething PW, Smith JL, Mwandawiro CS, Sturrock HJ, Gitonga CW, et al. Spatial modelling of soil-transmitted helminth infections in Kenya: a disease control planning tool. PLoS neglected tropical diseases. 2011;5(2):e958. doi: 10.1371/journal.pntd.0000958. PubMed PMID: 21347451; PubMed Central PMCID: PMC3035671.

10. Magalhaes RJ, Clements AC, Patil AP, Gething PW, Brooker S. The applications of model-based geostatistics in helminth epidemiology and control. Adv Parasitol. 2011;74:267-96. doi: 10.1016/B978-0-12-385897-9.00005-7. PubMed PMID: 21295680; PubMed Central PMCID: PMC3037997.

11. Clements AC, Kur LW, Gatpan G, Ngondi JM, Emerson PM, Lado M, et al. Targeting trachoma control through risk mapping: the example of Southern Sudan. PLoS Negl Trop Dis.

2010;4(8):e799. doi: 10.1371/journal.pntd.0000799. PubMed PMID: 20808910; PubMed Central PMCID: PMC2923154.

12. Sturrock HJ, Gething PW, Clements AC, Brooker S. Optimal survey designs for targeting chemotherapy against soil-transmitted helminths: effect of spatial heterogeneity and cost-efficiency of sampling. Am J Trop Med Hyg. 2010;82(6):1079-87. doi: 10.4269/ajtmh.2010.09-0702. PubMed PMID: 20519603; PubMed Central PMCID: PMC2877414.

13. Clements AC, Bosque-Oliva E, Sacko M, Landoure A, Dembele R, Traore M, et al. A comparative study of the spatial distribution of schistosomiasis in Mali in 1984-1989 and 2004-2006. PLoS Negl Trop Dis. 2009;3(5):e431. doi: 10.1371/journal.pntd.0000431. PubMed PMID: 19415108; PubMed Central PMCID: PMCPMC2671597.

14. Brooker S, Kabatereine NB, Smith JL, Mupfasoni D, Mwanje MT, Ndayishimiye O, et al. An updated atlas of human helminth infections: the example of East Africa. Int J Health Geogr. 2009;8:42. doi: 10.1186/1476-072X-8-42. PubMed PMID: 19589144; PubMed Central PMCID: PMCPMC2714505.

15. Clements AC, Lwambo NJ, Blair L, Nyandindi U, Kaatano G, Kinung'hi S, et al. Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. Trop Med Int Health. 2006;11(4):490-503. doi: 10.1111/j.1365-3156.2006.01594.x. PubMed PMID: 16553932; PubMed Central PMCID: PMC2202922.

16. Zoure HG, Noma M, Tekle AH, Amazigo UV, Diggle PJ, Giorgi E, et al. The geographic distribution of onchocerciasis in the 20 participating countries of the African Programme for Onchocerciasis Control: (2) pre-control endemicity levels and estimated number infected. Parasit Vectors. 2014;7:326. doi: 10.1186/1756-3305-7-326. PubMed PMID: 25053392; PubMed Central PMCID: PMCPMC4222889.

17. Zoure HG, Wanji S, Noma M, Amazigo UV, Diggle PJ, Tekle AH, et al. The geographic distribution of Loa loa in Africa: results of large-scale implementation of the Rapid Assessment Procedure for Loiasis (RAPLOA). PLoS Negl Trop Dis. 2011;5(6):e1210. doi: 10.1371/journal.pntd.0001210. PubMed PMID: 21738809; PubMed Central PMCID: PMC3125145.

18. Solomon AW, Pavluck A, Courtright P, Abdala M, Alemayehu W, Al-Khatib T, et al. Completing the Global Trachoma Map. PLoS neglected tropical diseases. 2016;Submitted.

19. Thylefors B, Dawson CR, Jones BR, West SK, Taylor HR. A simple system for the assessment of trachoma and its complications. Bull World Health Organ. 1987;65(4):477-83. PubMed PMID: 3500800; PubMed Central PMCID: PMC2491032.

20. Palmer SL, Winskell K, Patterson AE, Boubacar K, Ibrahim F, Namata I, et al. 'A living death': a qualitative assessment of quality of life among women with trichiasis in rural Niger. Int Health. 2014;6(4):291-7. doi: 10.1093/inthealth/ihu054. PubMed PMID: 25125577.

21. Mabey DC, Solomon AW, Foster A. Trachoma. Lancet. 2003;362(9379):223-9. doi: 10.1016/S0140-6736(03)13914-1. PubMed PMID: 12885486.

22. Solomon AW, Peeling RW, Foster A, Mabey DC. Diagnosis and assessment of trachoma. Clinical microbiology reviews. 2004;17(4):982-1011. doi: 10.1128/CMR.17.4.982-1011.2004. PubMed PMID: 15489358; PubMed Central PMCID: PMC523557.

23. Harding-Esch EM, Edwards T, Sillah A, Sarr I, Roberts CH, Snell P, et al. Active trachoma and ocular Chlamydia trachomatis infection in two Gambian regions: on course for elimination by 2020? PLoS Negl Trop Dis. 2009;3(12):e573. doi: 10.1371/journal.pntd.0000573. PubMed PMID: 20027217; PubMed Central PMCID: PMCPMC2791206.

24. Pant BP, Bhatta RC, Chaudhary JS, Awasthi S, Mishra S, Sharma S, et al. Control of Trachoma from Achham District, Nepal: A Cross-Sectional Study from the Nepal National Trachoma Program. PLoS neglected tropical diseases. 2016;10(2):e0004462. doi: 10.1371/journal.pntd.0004462. PubMed PMID: 26871898; PubMed Central PMCID: PMCPMC4752456.

25. Cox JT, Mkocha H, Munoz B, West SK. Trachomatous scarring among children in a formerly hyper-endemic district of Tanzania. PLoS neglected tropical diseases. 2017;11(12):e0006085. doi: 10.1371/journal.pntd.0006085. PubMed PMID: 29232692; PubMed Central PMCID: PMCPMC5741261.

26. West SK, Zambrano AI, Sharma S, Mishra SK, Munoz BE, Dize L, et al. Surveillance Surveys for Reemergent Trachoma in Formerly Endemic Districts in Nepal From 2 to 10 Years After Mass Drug Administration Cessation. JAMA ophthalmology. 2017;135(11):1141-6. doi: 10.1001/jamaophthalmol.2017.3062. PubMed PMID: 28973295; PubMed Central PMCID: PMCPMC5710394.

27. Rebollo MP, Bockarie MJ. Shrinking the lymphatic filariasis map: update on diagnostic tools for mapping and transmission monitoring. Parasitology. 2014;141(14):1912-7. doi: 10.1017/S0031182014001231. PubMed PMID: 25225828.

28. Vlaminck J, Fischer PU, Weil GJ. Diagnostic Tools for Onchocerciasis Elimination Programs. Trends Parasitol. 2015;31(11):571-82. doi: 10.1016/j.pt.2015.06.007. PubMed PMID: 26458784.

29. Martin DL, Bid R, Sandi F, Goodhew EB, Massae PA, Lasway A, et al. Serology for trachoma surveillance after cessation of mass drug administration. PLoS Negl Trop Dis. 2015;9(2):e0003555. doi: 10.1371/journal.pntd.0003555. PubMed PMID: 25714363; PubMed Central PMCID: PMCPMC4340913.

30. Grayston JT, Wang SP, Yeh LJ, Kuo CC. Importance of reinfection in the pathogenesis of trachoma. Rev Infect Dis. 1985;7(6):717-25. PubMed PMID: 4070905.

31. Nichols RL, Bobb AA, Haddad NA, McComb DE. Immunofluorescent studies of the microbiologic epidemiology of trachoma in Saudi Arabia. Am J Ophthalmol. 1967;63(5):Suppl:1372-408. PubMed PMID: 5337745.

32. Schemann JF, Sacko D, Malvy D, Momo G, Traore L, Bore O, et al. Risk factors for trachoma in Mali. Int J Epidemiol. 2002;31(1):194-201. PubMed PMID: 11914321.

33. Jones BR. The prevention of blindness from trachoma. Transactions of the ophthalmological societies of the United Kingdom. 1975;95(1):16-33. PubMed PMID: 775692.

34. Dawson CR, Daghfous T, Messadi M, Hoshiwara I, Schachter J. Severe endemic trachoma in Tunisia. Br J Ophthalmol. 1976;60(4):245-52. PubMed PMID: 1276112; PubMed Central PMCID: PMC1017485.

35. Mabey DC, Bailey RL, Ward ME, Whittle HC. A longitudinal study of trachoma in a Gambian village: implications concerning the pathogenesis of chlamydial infection. Epidemiol Infect. 1992;108(2):343-51. PubMed PMID: 1582475; PubMed Central PMCID: PMCPMC2271993.

36. West SK, Munoz B, Turner VM, Mmbaga BB, Taylor HR. The epidemiology of trachoma in central Tanzania. Int J Epidemiol. 1991;20(4):1088-92. PubMed PMID: 1800408.

37. Broman AT, Shum K, Munoz B, Duncan DD, West SK. Spatial clustering of ocular chlamydial infection over time following treatment, among households in a village in Tanzania. Invest Ophthalmol Vis Sci. 2006;47(1):99-104. doi: 10.1167/iovs.05-0326. PubMed PMID: 16384950.

38. Hagi M, Schemann JF, Mauny F, Momo G, Sacko D, Traore L, et al. Active trachoma among children in Mali: Clustering and environmental risk factors. PLoS Negl Trop Dis. 2010;4(1):e583. doi: 10.1371/journal.pntd.0000583. PubMed PMID: 20087414; PubMed Central PMCID: PMC2799671.

39. Barenfanger J. Studies on the role of the famiy unit in the transmission of trachoma. Am J Trop Med Hyg. 1975;24(3):509-15. PubMed PMID: 1155691.

40. Treharne JD. The community epidemiology of trachoma. Rev Infect Dis. 1985;7(6):760-4. PubMed PMID: 4070910.

41. Taylor HR, West SK, Mmbaga BB, Katala SJ, Turner V, Lynch M, et al. Hygiene factors and increased risk of trachoma in central Tanzania. Archives of ophthalmology. 1989;107(12):1821-5. PubMed PMID: 2597076.

42. Bailey R, Osmond C, Mabey DC, Whittle HC, Ward ME. Analysis of the household distribution of trachoma in a Gambian village using a Monte Carlo simulation procedure. Int J Epidemiol. 1989;18(4):944-51. PubMed PMID: 2621031.

43. Hu VH, Holland MJ, Burton MJ. Trachoma: protective and pathogenic ocular immune responses to Chlamydia trachomatis. PLoS Negl Trop Dis. 2013;7(2):e2020. doi: 10.1371/journal.pntd.0002020. PubMed PMID: 23457650; PubMed Central PMCID: PMCPMC3573101.

44. Ramadhani AM, Derrick T, Macleod D, Massae P, Mtuy T, Jeffries D, et al. Immunofibrogenic Gene Expression Patterns in Tanzanian Children with Ocular Chlamydia trachomatis Infection, Active Trachoma and Scarring: Baseline Results of a 4-Year Longitudinal Study. Front Cell Infect Microbiol. 2017;7:406. doi: 10.3389/fcimb.2017.00406. PubMed PMID: 28966918; PubMed Central PMCID: PMCPMC5605569.

45. Cumberland P, Hailu G, Todd J. Active trachoma in children aged three to nine years in rural communities in Ethiopia: prevalence, indicators and risk factors. Trans R Soc Trop Med Hyg. 2005;99(2):120-7. doi: 10.1016/j.trstmh.2004.03.011. PubMed PMID: 15607339.

46. Alene GD, Abebe S. Prevalence of risk factors for trachoma in a rural locality of north-western Ethiopia. East Afr Med J. 2000;77(6):308-12. PubMed PMID: 12858929.

47. Guraksin A, Gullulu G. Prevalence of trachoma in eastern Turkey. Int J Epidemiol. 1997;26(2):436-42. PubMed PMID: 9169182.

48. Roba AA, Wondimu A, Patel D, Zondervan M. Effects of intervention with the SAFE strategy on trachoma across Ethiopia. J Epidemiol Community Health. 2011;65(7):626-31. doi: 10.1136/jech.2009.094763. PubMed PMID: 20693489.

49. King JD, Ngondi J, Kasten J, Diallo MO, Zhu H, Cromwell EA, et al. Randomised trial of face-washing to develop a standard definition of a clean face for monitoring trachoma control

programmes. Trans R Soc Trop Med Hyg. 2011;105(1):7-16. doi: 10.1016/j.trstmh.2010.09.008. PubMed PMID: 21036378.

50. West S, Lynch M, Turner V, Munoz B, Rapoza P, Mmbaga BB, et al. Water availability and trachoma. Bull World Health Organ. 1989;67(1):71-5. PubMed PMID: 2706728; PubMed Central PMCID: PMC2491213.

51. Baggaley RF, Solomon AW, Kuper H, Polack S, Massae PA, Kelly J, et al. Distance to water source and altitude in relation to active trachoma in Rombo district, Tanzania. Trop Med Int Health. 2006;11(2):220-7. doi: 10.1111/j.1365-3156.2005.01553.x. PubMed PMID: 16451347.

52. Polack SR, Solomon AW, Alexander ND, Massae PA, Safari S, Shao JF, et al. The household distribution of trachoma in a Tanzanian village: an application of GIS to the study of trachoma. Trans R Soc Trop Med Hyg. 2005;99(3):218-25. doi: 10.1016/j.trstmh.2004.06.010. PubMed PMID: 15653125.

53. Emerson PM, Bailey RL, Mahdi OS, Walraven GE, Lindsay SW. Transmission ecology of the fly Musca sorbens, a putative vector of trachoma. Trans R Soc Trop Med Hyg. 2000;94(1):28-32. PubMed PMID: 10748893.

54. Abdou A, Nassirou B, Kadri B, Moussa F, Munoz BE, Opong E, et al. Prevalence and risk factors for trachoma and ocular Chlamydia trachomatis infection in Niger. Br J Ophthalmol. 2007;91(1):13-7. doi: 10.1136/bjo.2006.099507. PubMed PMID: 16899525; PubMed Central PMCID: PMC1857562.

55. Amza A, Kadri B, Nassirou B, Stoller NE, Yu SN, Zhou Z, et al. Community risk factors for ocular Chlamydia infection in Niger: pre-treatment results from a cluster-randomized trachoma trial. PLoS Negl Trop Dis. 2012;6(4):e1586. doi: 10.1371/journal.pntd.0001586. PubMed PMID: 22545165; PubMed Central PMCID: PMCPMC3335874.

56. Edwards T, Harding-Esch EM, Hailu G, Andreason A, Mabey DC, Todd J, et al. Risk factors for active trachoma and Chlamydia trachomatis infection in rural Ethiopia after mass treatment with azithromycin. Trop Med Int Health. 2008;13(4):556-65. doi: 10.1111/j.1365-3156.2008.02034.x. PubMed PMID: 18282237.

57. Alemayehu W, Melese M, Fredlander E, Worku A, Courtright P. Active trachoma in children in central Ethiopia: association with altitude. Trans R Soc Trop Med Hyg. 2005;99(11):840-3. doi: 10.1016/j.trstmh.2005.06.013. PubMed PMID: 16107273.

58. Taye A, Alemayehu W, Melese M, Geyid A, Mekonnen Y, Tilahun D, et al. Seasonal and altitudinal variations in fly density and their association with the occurrence of trachoma, in the Gurage zone of central Ethiopia. Ann Trop Med Parasitol. 2007;101(5):441-8. doi: 10.1179/136485907X176544. PubMed PMID: 17550650.

59. Ngondi J, Gebre T, Shargie EB, Graves PM, Ejigsemahu Y, Teferi T, et al. Risk factors for active trachoma in children and trichiasis in adults: a household survey in Amhara Regional State, Ethiopia. Trans R Soc Trop Med Hyg. 2008;102(5):432-8. doi: 10.1016/j.trstmh.2008.02.014. PubMed PMID: 18394663.

60. West SK, Emerson PM, Mkocha H, McHiwa W, Munoz B, Bailey R, et al. Intensive insecticide spraying for fly control after mass antibiotic treatment for trachoma in a hyperendemic setting: a randomised trial. Lancet. 2006;368(9535):596-600. doi: 10.1016/S0140-6736(06)69203-9. PubMed PMID: 16905024.

61. Turner VM, West SK, Munoz B, Katala SJ, Taylor HR, Halsey N, et al. Risk factors for trichiasis in women in Kongwa, Tanzania: a case-control study. International journal of epidemiology. 1993;22(2):341-7. PubMed PMID: 8505194.

62. West SK, Munoz B, Mkocha H, Hsieh YH, Lynch MC. Progression of active trachoma to scarring in a cohort of Tanzanian children. Ophthalmic Epidemiol. 2001;8(2-3):137-44. PubMed PMID: 11471083.

63. Burton MJ, Rajak SN, Hu VH, Ramadhani A, Habtamu E, Massae P, et al. Pathogenesis of progressive scarring trachoma in Ethiopia and Tanzania and its implications for disease control: two cohort studies. PLoS Negl Trop Dis. 2015;9(5):e0003763. doi: 10.1371/journal.pntd.0003763. PubMed PMID: 25970613; PubMed Central PMCID: PMCPMC4430253.

64. Wolle MA, Munoz B, Mkocha H, West SK. Age, sex, and cohort effects in a longitudinal study of trachomatous scarring. Invest Ophthalmol Vis Sci. 2009;50(2):592-6. doi: 10.1167/iovs.08-2414. PubMed PMID: 18936137; PubMed Central PMCID: PMCPMC3820011.

65. Ramadhani AM, Derrick T, Holland MJ, Burton MJ. Blinding Trachoma: Systematic Review of Rates and Risk Factors for Progressive Disease. PLoS Negl Trop Dis. 2016;10(8):e0004859. doi: 10.1371/journal.pntd.0004859. PubMed PMID: 27483002; PubMed Central PMCID: PMCPMC4970760.

66. Burton MJ, Bowman RJ, Faal H, Aryee EA, Ikumapayi UN, Alexander ND, et al. The long-term natural history of trachomatous trichiasis in the Gambia. Invest Ophthalmol Vis Sci. 2006;47(3):847-52. doi: 10.1167/iovs.05-0714. PubMed PMID: 16505016.

67. Schemann JF, Laffly D, Sacko D, Zephak G, Malvy D. Trichiasis and geoclimatic factors in Mali. Trans R Soc Trop Med Hyg. 2007;101(10):996-1003. doi: 10.1016/j.trstmh.2007.05.015. PubMed PMID: 17658570.

68. Salim AR, Sheikh HA. Trachoma in the Sudan. An epidemiological study. Br J Ophthalmol. 1975;59(10):600-4. PubMed PMID: 1191619; PubMed Central PMCID: PMC1017417.

69. Negrel AD, Mariotti SP. Trachoma rapid assessment: rationale and basic principles. Community Eye Health. 1999;12(32):51-3. PubMed PMID: 17492004; PubMed Central PMCID: PMCPMC1706027.

70. Anthony W Solomon AP, Paul Courtright, Agatha Aboe, Liknaw Adamu, Wondu Alemayehu... The Global Trachoma Mapping Project Methodology. Opthalmic Epidemiology. Pending Publication.

71. Solomon AW, Pavluck AL, Courtright P, Aboe A, Adamu L, Alemayehu W, et al. The Global Trachoma Mapping Project: Methodology of a 34-Country Population-Based Study. Ophthalmic epidemiology. 2015;22(3):214-25. doi: 10.3109/09286586.2015.1037401. PubMed PMID: 26158580.

72. Solomon AW PA, Courtright P, Brooker SJ, Chu BK, Dejene M, Emerson PM, Flueckiger RM, Haddad D, Macleod C, McCullagh S, Millar T, Mpyet C, Ngondi J, Rotondo L, Willis R, Foster A. Completing the Global Trachoma Map. under review. 2015.

73. World Health Organization. Report of the 3rd global scientific meeting on trachoma. Johns Hopkins University, Baltimore, MA, 19-20 July 2010: Geneva: World Health Organization; 2010.

74. Burton M, Solomon A. What's new in trichiasis surgery? Community Eye Health. 2004;17(52):52-3. PubMed PMID: 17491821; PubMed Central PMCID: PMC1705745.

75. Rajak SN, Habtamu E, Weiss HA, Bedri A, Gebre T, Genet A, et al. Epilation for trachomatous trichiasis and the risk of corneal opacification. Ophthalmology. 2012;119(1):84-9. doi: 10.1016/j.ophtha.2011.06.045. PubMed PMID: 21975041; PubMed Central PMCID: PMC3694301.

76. Mark R, Foster A, Huber J, Bauer B. Trichiasis surgery for trachoma; the bilamellar tarsal rotation procedure. The World Health Organization.

77. Habtamu E, Wondie T, Aweke S, Tadesse Z, Zerihun M, Zewudie Z, et al. Posterior lamellar versus bilamellar tarsal rotation surgery for trachomatous trichiasis in Ethiopia: a randomised controlled trial. Lancet Glob Health. 2016;4(3):e175-84. doi: 10.1016/S2214-109X(15)00299-5. PubMed PMID: 26774708; PubMed Central PMCID: PMCPMC5075282.

78. Merbs SL, Resnikoff S, Kello AB, Mariotti S, Greene G, West SK. Trichiasis surgery for trachoma. Geneva, Switzerland: The World Health Organization, 2015.

79. Rajak SN, Collin JR, Burton MJ. Trachomatous trichiasis and its management in endemic countries. Survey of ophthalmology. 2012;57(2):105-35. doi: 10.1016/j.survophthal.2011.08.002. PubMed PMID: 22285842; PubMed Central PMCID: PMC3316859.

80. Burton MJ, Bowman RJ, Faal H, Aryee EA, Ikumapayi UN, Alexander ND, et al. Long term outcome of trichiasis surgery in the Gambia. Br J Ophthalmol. 2005;89(5):575-9. doi: 10.1136/bjo.2004.055996. PubMed PMID: 15834088; PubMed Central PMCID: PMC1772648.

81. Rajak SN, Makalo P, Sillah A, Holland MJ, Mabey DC, Bailey RL, et al. Trichiasis surgery in The Gambia: a 4-year prospective study. Invest Ophthalmol Vis Sci. 2010;51(10):4996-5001. doi: 10.1167/iovs.10-5169. PubMed PMID: 20505197; PubMed Central PMCID: PMC3066607.

82. Tadesse D, Montgomery I, Sankar G. HEAD START - an innovative training approach for life-long learning. Community Eye Health. 2017;30(97):14. PubMed PMID: 28603395; PubMed Central PMCID: PMCPMC5461515.

83. World Health Organization. WHO Alliance for the Global Elimination of Blinding Trachoma by the year 2020: progress report on elimination of trachoma, 2017. Wkly Epidemiol Rec. 2017;92(26):357-68.

84. World Health Organization. Global WHO Alliance for the Elimination of Blinding Trachoma by 2020. Wkly Epidemiol Rec. 2012;87(17):161-8.

85. Burton MJ, Frick KD, Bailey RL, Bowman RJ. Azithromycin for the treatment and control of trachoma. Expert Opin Pharmacother. 2002;3(2):113-20. doi: 10.1517/14656566.3.2.113. PubMed PMID: 11829725.

86. West SK, Munoz B, Mkocha H, Gaydos CA, Quinn TC. Number of years of annual mass treatment with azithromycin needed to control trachoma in hyper-endemic communities in Tanzania. J Infect Dis. 2011;204(2):268-73. doi: 10.1093/infdis/jir257. PubMed PMID: 21673038; PubMed Central PMCID: PMCPMC3114471.

87. Jimenez V, Gelderblom HC, Mann Flueckiger R, Emerson PM, Haddad D. Mass drug administration for trachoma: how long is not long enough? PLoS neglected tropical diseases. 2015;9(3):e0003610. doi: 10.1371/journal.pntd.0003610. PubMed PMID: 25799168; PubMed Central PMCID: PMCPMC4370651.

88. Lietman T, Porco T, Dawson C, Blower S. Global elimination of trachoma: how frequently should we administer mass chemotherapy? Nat Med. 1999;5(5):572-6. doi: 10.1038/8451. PubMed PMID: 10229236.

89. Gebre T, Ayele B, Zerihun M, Genet A, Stoller NE, Zhou Z, et al. Comparison of annual versus twice-yearly mass azithromycin treatment for hyperendemic trachoma in Ethiopia: a cluster-randomised trial. Lancet. 2012;379(9811):143-51. doi: 10.1016/S0140-6736(11)61515-8. PubMed PMID: 22192488.

90. Amza A, Kadri B, Nassirou B, Cotter SY, Stoller NE, Zhou Z, et al. A Cluster-Randomized Trial to Assess the Efficacy of Targeting Trachoma Treatment to Children. Clin Infect Dis. 2017;64(6):743-50. doi: 10.1093/cid/ciw810. PubMed PMID: 27956455.

91. Emerson PM, Cairncross S, Bailey RL, Mabey DC. Review of the evidence base for the 'F' and 'E' components of the SAFE strategy for trachoma control. Trop Med Int Health. 2000;5(8):515-27. PubMed PMID: 10995092.

92. West S, Munoz B, Lynch M, Kayongoya A, Chilangwa Z, Mmbaga BB, et al. Impact of face-washing on trachoma in Kongwa, Tanzania. Lancet. 1995;345(8943):155-8. PubMed PMID: 7823670.

93. Hafez M AM. Studies on the ecology of Musca sorbens Wied. in Egypt. Bulletin de la Société Entomologiqe d'Egypte. 1958;42(83):121.

94. Curtis CF, Hawkins PM. Entomological studies of on-site sanitation systems in Botswana and Tanzania. Trans R Soc Trop Med Hyg. 1982;76(1):99-108. PubMed PMID: 6123168.

95. Emerson PM, Lindsay SW, Alexander N, Bah M, Dibba SM, Faal HB, et al. Role of flies and provision of latrines in trachoma control: cluster-randomised controlled trial. Lancet. 2004;363(9415):1093-8. doi: 10.1016/S0140-6736(04)15891-1. PubMed PMID: 15064026.

96. Stocks ME, Ogden S, Haddad D, Addiss DG, McGuire C, Freeman MC. Effect of water, sanitation, and hygiene on the prevention of trachoma: a systematic review and meta-analysis. PLoS Med. 2014;11(2):e1001605. doi: 10.1371/journal.pmed.1001605. PubMed PMID: 24586120; PubMed Central PMCID: PMCPMC3934994.

97. Transforming our world: the 2030 agenda for sustainable development. General Assemgly: Seventieth session: United Nations; 2015.

98. Callahan K, Ogale YP, Palmer SL, Emerson PM, Hopkins DR, Withers PC, Jr., et al. Trachoma control as a vehicle toward international development and achievement of the millennium development goals. PLoS Negl Trop Dis. 2014;8(9):e3137. doi: 10.1371/journal.pntd.0003137. PubMed PMID: 25232728; PubMed Central PMCID: PMCPMC4169372.

99. Cromwell EA, Courtright P, King JD, Rotondo LA, Ngondi J, Emerson PM. The excess burden of trachomatous trichiasis in women: a systematic review and meta-analysis. Trans R Soc Trop Med Hyg. 2009;103(10):985-92. doi: 10.1016/j.trstmh.2009.03.012. PubMed PMID: 19362326.

100. Congdon N, West S, Vitale S, Katala S, Mmbaga BB. Exposure to children and risk of active trachoma in Tanzanian women. Am J Epidemiol. 1993;137(3):366-72. PubMed PMID: 8452144.

101. Sadiq ST, Glasgow KW, Drakeley CJ, Muller O, Greenwood BM, Mabey DC, et al. Effects of azithromycin on malariometric indices in The Gambia. Lancet. 1995;346(8979):881-2. PubMed PMID: 7564674.

102. Marks M, Bottomley C, Tome H, Pitakaka R, Butcher R, Sokana O, et al. Mass drug administration of azithromycin for trachoma reduces the prevalence of genital Chlamydia trachomatis infection in the Solomon Islands. Sex Transm Infect. 2016;92(4):261-5. doi: 10.1136/sextrans-2015-052439. PubMed PMID: 26888658; PubMed Central PMCID: PMCPMC4893086.

103. Fonck K, Kaul R, Kimani J, Keli F, MacDonald KS, Ronald AR, et al. A randomized, placebo-controlled trial of monthly azithromycin prophylaxis to prevent sexually transmitted infections and HIV-1 in Kenyan sex workers: study design and baseline findings. Int J STD AIDS. 2000;11(12):804-11. doi: 10.1258/0956462001915327. PubMed PMID: 11138916.

104. Esrey SA, Potash JB, Roberts L, Shiff C. Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma. Bull World Health Organ. 1991;69(5):609-21. PubMed PMID: 1835675; PubMed Central PMCID: PMCPMC2393264.

105. Porco TC, Gebre T, Ayele B, House J, Keenan J, Zhou Z, et al. Effect of mass distribution of azithromycin for trachoma control on overall mortality in Ethiopian children: a randomized trial. JAMA. 2009;302(9):962-8. doi: 10.1001/jama.2009.1266. PubMed PMID: 19724043.

106. Frick KD, Hanson CL, Jacobson GA. Global burden of trachoma and economics of the disease. Am J Trop Med Hyg. 2003;69(5 Suppl):1-10. PubMed PMID: 14692674.

107. Accelerating work to overcome the global impact of neglected tropical diseases, a roadmap for implementation. Geneva, Switzerland: World Health Organization,, 2012 WHO/HTM/NTD/2012.1.

108. The end in sight: 2020 INSight. International Coalition for Trachoma Control, 2011.

109. The London Declaration. Uniting to Combat NTDs.

110. Graham SL, Barling KS, Waghela S, Scott HM, Thompson JA. Spatial distribution of antibodies to Salmonella enterica serovar Typhimurium O antigens in bulk milk from Texas dairy herds. Preventive veterinary medicine. 2005;69(1-2):53-61. doi: 10.1016/j.prevetmed.2005.01.018. PubMed PMID: 15899296.

111. World Health Organization. Global Data on Visual Impairments 2010. 2012 WHO/NMH/PBD/12.01.

112. Ranson MK, Evans TG. The global burden of trachomatous visual impairment: I. Assessing prevalence. Int Ophthalmol. 1995;19(5):261-70. PubMed PMID: 8864809.

113. Evans TG, Ranson MK. The global burden of trachomatous visual impairment: II. Assessing burden. Int Ophthalmol. 1995;19(5):271-80. PubMed PMID: 8864810.

114. Frick KD, Melia BM, Buhrmann RR, West SK. Trichiasis and disability in a trachoma-endemic area of Tanzania. Archives of ophthalmology. 2001;119(12):1839-44. PubMed PMID: 11735797.

115. Habtamu E, Wondie T, Aweke S, Tadesse Z, Zerihun M, Zewdie Z, et al. Trachoma and Relative Poverty: A Case-Control Study. PLoS Negl Trop Dis. 2015;9(11):e0004228. doi: 10.1371/journal.pntd.0004228. PubMed PMID: 26600211; PubMed Central PMCID: PMCPMC4657919.

116. Habtamu E, Wondie T, Aweke S, Tadesse Z, Zerihun M, Zewudie Z, et al. The Impact of Trachomatous Trichiasis on Quality of Life: A Case Control Study. PLoS Negl Trop Dis. 2015;9(11):e0004254. doi: 10.1371/journal.pntd.0004254. PubMed PMID: 26598937; PubMed Central PMCID: PMCPMC4657886.

117. Frick KD, Basilion EV, Hanson CL, Colchero MA. Estimating the burden and economic impact of trachomatous visual loss. Ophthalmic Epidemiol. 2003;10(2):121-32. PubMed PMID: 12660860.

118. Frick KD, Keuffel EL, Bowman RJ. Epidemiological, demographic, and economic analyses: measurement of the value of trichiasis surgery in The Gambia. Ophthalmic Epidemiol. 2001;8(2-3):191-201. PubMed PMID: 11471088.

119. Tobler W. A computer movie simulating urban growth in the Detroit region. Economic Geography. 1970;46:234-40.

120. Goovaerts P. Geostatistics for Natural Resources Evaluation. New York: Oxford University Press; 1997.

121. Pullan RL, Sturrock HJ, Soares Magalhaes RJ, Clements AC, Brooker SJ. Spatial parasite ecology and epidemiology: a review of methods and applications. Parasitology. 2012;139(14):1870-87. doi: 10.1017/S0031182012000698. PubMed PMID: 23036435; PubMed Central PMCID: PMC3526959.

122. Knox EG. Detection of clusters. Methodologies of Enquiry into Disease Clustering. Elliott, P. ed. London: Health Statistics Unit; 1989. p. 17-22.

123. Cuzick J, Edwards R. Spatial clustering for inhomogeneous populations. Journal of the Royal Statistical Society Series B. 1990;52:73-104.

124. Moran PA. Notes on continuous stochastic phenomena. Biometrika. 1950;37(1-2):17-23. PubMed PMID: 15420245.

125. Diggle PJ. Statistical Analysis of Spatial Point Patterns. London: Arnold Publishers; 2003.

126. Mantel N. The detection of disease clustering and a generalized regression approach. Cancer Res. 1967;27(2):209-20. PubMed PMID: 6018555.

127. Jacquez G, Grimson R, Kheifets L, Waller L, Wartenberg D. Statistics and computing in disease clustering. Statistics in Medicine; University of British Columbia1996. p. 681 - 952.

128. Wartenberg D, Greenberg M. Detecting disease clusters: the importance of statistical power. Am J Epidemiol. 1990;132(1 Suppl):S156-66. PubMed PMID: 2192551.

129. Getis A, Ord JK. Local spatial statistics: An overview. In: Longley P, Batty M, editors. Spatial Analysis: Modelling in a GIS Environment. Cambridge: Geoinformation International; 1996. p. 261-77.

130. Kulldorff M. A spatial scan statistic. Communications in Statistics: Theory and Methods1997. p. 1481-96.

131. Jacquez GM, Greiling DA. Local clustering in breast, lung and colorectal cancer in Long Island, New York. Int J Health Geogr. 2003;2(1):3. PubMed PMID: 12633503; PubMed Central PMCID: PMCPMC151676.

132. Pfeiffer D. Spatial analysis in epidemiology. Oxford ; New York: Oxford University Press; 2008. xii, 142 p. p.

133. World Health Organization. Report of the 2nd global scientific meeting on trachoma. Geneva, 25-27 August, 2003 (WHO/PBD/GET 03.1): Geneva: World Health Organization; 2003.

134. Smith JL, Flueckiger RM, Hooper PJ, Polack S, Cromwell EA, Palmer SL, et al. The geographical distribution and burden of trachoma in Africa. PLoS neglected tropical diseases. 2013;7(8):e2359. doi: 10.1371/journal.pntd.0002359. PubMed PMID: 23951378; PubMed Central PMCID: PMC3738464.

135. Beyene HB, Bekele A, Shifara A, Ebstie YA, Desalegn Z, Kebede Z, et al. Elimination of Guinea Worm Disease in Ethiopia; Current Status of the Disease's, Eradication Strategies and Challenges to the End Game. Ethiop Med J. 2017;55(Suppl 1):15-31. PubMed PMID: 28878428; PubMed Central PMCID: PMCPMC5582630.

136. Lockwood DN, Shetty V, Penna GO. Hazards of setting targets to eliminate disease: lessons from the leprosy elimination campaign. BMJ. 2014;348:g1136. doi: 10.1136/bmj.g1136. PubMed PMID: 24508610.

137. Furst T, Cavaliero A, Lay S, Dayer C, Chan S, Smrekar A, et al. Retrospective active case finding in Cambodia: An innovative approach to leprosy control in a low-endemic country. Acta Trop. 2017;180:26-32. doi: 10.1016/j.actatropica.2017.12.031. PubMed PMID: 29289558.

138. Moura ML, Dupnik KM, Sampaio GA, Nobrega PF, Jeronimo AK, do Nascimento-Filho JM, et al. Active surveillance of Hansen's Disease (leprosy): importance for case finding among extra-domiciliary contacts. PLoS Negl Trop Dis. 2013;7(3):e2093. doi: 10.1371/journal.pntd.0002093. PubMed PMID: 23516645; PubMed Central PMCID: PMCPMC3597486.

139. Shen J, Zhou M, Xu X, Ray A, Zhang G, Yan L. A big challenge in case finding at low endemic situation: analysis on 1462 new leprosy patients detected in China in 2007. Lepr Rev. 2010;81(3):176-83. PubMed PMID: 21067058.

140. Freitas LRS, Duarte EC, Garcia LP. Analysis of the epidemiological situation of leprosy in an endemic area in Brazil: spatial distribution in the periods 2001 - 2003 and 2010 - 2012. Rev Bras Epidemiol. 2017;20(4):702-13. doi: 10.1590/1980-5497201700040012. PubMed PMID: 29267754.

141. Ramos AC, Yamamura M, Arroyo LH, Popolin MP, Chiaravalloti Neto F, Palha PF, et al. Spatial clustering and local risk of leprosy in Sao Paulo, Brazil. PLoS Negl Trop Dis. 2017;11(2):e0005381. doi: 10.1371/journal.pntd.0005381. PubMed PMID: 28241038; PubMed Central PMCID: PMCPMC5344525.

142. Monteiro LD, Martins-Melo FR, Brito AL, Alencar CH, Heukelbach J. Spatial patterns of leprosy in a hyperendemic state in Northern Brazil, 2001-2012. Rev Saude Publica. 2015;49. doi: 10.1590/S0034-8910.2015049005866. PubMed PMID: 26603352; PubMed Central PMCID: PMCPMC4650934.

143. De Souza Dias MC, Dias GH, Nobre ML. The use of Geographical Information System (GIS) to improve active leprosy case finding campaigns in the municipality of Mossoro, Rio Grande do Norte State, Brazil. Lepr Rev. 2007;78(3):261-9. PubMed PMID: 18035777.

144. Tadesse Argaw A, Shannon EJ, Assefa A, Mikru FS, Mariam BK, Malone JB. A geospatial risk assessment model for leprosy in Ethiopia based on environmental thermal-hydrological regime analysis. Geospat Health. 2006;1(1):105-13. doi: 10.4081/gh.2006.285. PubMed PMID: 18686236.

145. Bousema T, Drakeley C, Gesase S, Hashim R, Magesa S, Mosha F, et al. Identification of hot spots of malaria transmission for targeted malaria control. J Infect Dis. 2010;201(11):1764-74. doi: 10.1086/652456. PubMed PMID: 20415536.

146. Smith JL, Sivasubramaniam S, Rabiu MM, Kyari F, Solomon AW, Gilbert C. Multilevel Analysis of Trachomatous Trichiasis and Corneal Opacity in Nigeria: The Role of Environmental and Climatic Risk Factors on the Distribution of Disease. PLoS Negl Trop Dis. 2015;9(7):e0003826. doi: 10.1371/journal.pntd.0003826. PubMed PMID: 26222549; PubMed Central PMCID: PMCPMC4519340.

147. Yayemain D, King JD, Debrah O, Emerson PM, Aboe A, Ahorsu F, et al. Achieving trachoma control in Ghana after implementing the SAFE strategy. Trans R Soc Trop Med Hyg. 2009;103(10):993-1000. doi: 10.1016/j.trstmh.2009.02.007. PubMed PMID: 19286239.

148. Lietman TM, Gebre T, Abdou A, Alemayehu W, Emerson P, Blumberg S, et al. The distribution of the prevalence of ocular chlamydial infection in communities where trachoma is disappearing. Epidemics. 2015;11:85-91. doi: 10.1016/j.epidem.2015.03.003. PubMed PMID: 25979286.

149. Gurjav U, Jelfs P, Hill-Cawthorne GA, Marais BJ, Sintchenko V. Genotype heterogeneity of Mycobacterium tuberculosis within geospatial hotspots suggests foci of imported infection in Sydney, Australia. Infect Genet Evol. 2015. doi: 10.1016/j.meegid.2015.07.014. PubMed PMID: 26187743.

150. Liu Y, Li X, Wang W, Li Z, Hou M, He Y, et al. Investigation of space-time clusters and geospatial hot spots for the occurrence of tuberculosis in Beijing. Int J Tuberc Lung Dis. 2012;16(4):486-91. doi: 10.5588/ijtld.11.0255. PubMed PMID: 22325066.

151. Raso G, Matthys B, N'Goran EK, Tanner M, Vounatsou P, Utzinger J. Spatial risk prediction and mapping of Schistosoma mansoni infections among schoolchildren living in western Cote d'Ivoire. Parasitology. 2005;131(Pt 1):97-108. PubMed PMID: 16038401.

152. Brooker S. Spatial epidemiology of human schistosomiasis in Africa: risk models, transmission dynamics and control. Trans R Soc Trop Med Hyg. 2007;101(1):1-8. doi: 10.1016/j.trstmh.2006.08.004. PubMed PMID: 17055547; PubMed Central PMCID: PMC1975763.

153. Xu M, Cao C, Wang D, Kan B. Identifying environmental risk factors of cholera in a coastal area with geospatial technologies. Int J Environ Res Public Health. 2015;12(1):354-70. doi: 10.3390/ijerph120100354. PubMed PMID: 25551518; PubMed Central PMCID: PMCPMC4306866.

154. Clements AC, Reid HL, Kelly GC, Hay SI. Further shrinking the malaria map: how can geospatial science help to achieve malaria elimination? Lancet Infect Dis. 2013;13(8):709-18. doi: 10.1016/S1473-3099(13)70140-3. PubMed PMID: 23886334.

155. Hay S, Renshaw M, Ochola SA, Noor AM, Snow RW. Performance of forecasting, warning and detection of malaria epidemics in the highlands of western Kenya. Trends Parasitol. 2003;19(9):394-9. PubMed PMID: 12957515; PubMed Central PMCID: PMCPMC3173865.

156. Hay SI, Were EC, Renshaw M, Noor AM, Ochola SA, Olusanmi I, et al. Forecasting, warning, and detection of malaria epidemics: a case study. Lancet. 2003;361(9370):1705-6. doi: 10.1016/S0140-6736(03)13366-1. PubMed PMID: 12767739; PubMed Central PMCID: PMCPMC3164796.

157. Hay SI, Snow RW, Rogers DJ. Predicting malaria seasons in Kenya using multitemporal meteorological satellite sensor data. Trans R Soc Trop Med Hyg. 1998;92(1):12-20. PubMed PMID: 9692138.

158. Hay SI, Snow RW, Rogers DJ. From predicting mosquito habitat to malaria seasons using remotely sensed data: practice, problems and perspectives. Parasitol Today. 1998;14(8):306-13. PubMed PMID: 17040796.

159. Mabaso ML, Craig M, Ross A, Smith T. Environmental predictors of the seasonality of malaria transmission in Africa: the challenge. Am J Trop Med Hyg. 2007;76(1):33-8. PubMed PMID: 17255225.

160. Abeku TA, Hay SI, Ochola S, Langi P, Beard B, de Vlas SJ, et al. Malaria epidemic early warning and detection in African highlands. Trends Parasitol. 2004;20(9):400-5. doi: 10.1016/j.pt.2004.07.005. PubMed PMID: 15324728; PubMed Central PMCID: PMCPMC3173853.

161. Abeku TA, van Oortmarssen GJ, Borsboom G, de Vlas SJ, Habbema JD. Spatial and temporal variations of malaria epidemic risk in Ethiopia: factors involved and implications. Acta Trop. 2003;87(3):331-40. PubMed PMID: 12875926.

162. World Health Organization. Validation of elimination of trachoma as a public health problem. WHO/HTM/NTD/2016.8. Geneva: World Health Organization, 2016.

163. Mariotti SP, Pascolini D, Rose-Nussbaumer J. Trachoma: global magnitude of a preventable cause of blindness. Br J Ophthalmol. 2009;93(5):563-8. doi: 10.1136/bjo.2008.148494. PubMed PMID: 19098034.

164. WHO Alliance for the Global Elimination of Blinding Trachoma by 2020. Report of the 11th meeting of the WHO Alliance for the Global Elimination of Blinding Trachoma. Geneva, Switzerland: 2007 2-4 April 2007. Report No.

165. Hooper PJ, Zoerhoff KL, Kyelem D, Chu B, Flueckiger RM, Bamani S, et al. The effects of integration on financing and coverage of neglected tropical disease programs. Am J Trop Med Hyg. 2013;89(3):407-10. Epub 2013/07/10. doi: 10.4269/ajtmh.13-0018. PubMed PMID: 23836563; PubMed Central PMCID: PMCPMC3771274.

166. Brooker S, Kabatereine NB, Tukahebwa EM, Kazibwe F. Spatial analysis of the distribution of intestinal nematode infections in Uganda. Epidemiol Infect. 2004;132(6):1065-71. PubMed PMID: 15635963; PubMed Central PMCID: PMC1975758.

167. ArcGIS Desktop. 10.3 ed. Redlands, CA: Environmental Systems Research Institute (ESRI); 2014.

168. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2016.

169. Progress on drinking water, sanitation and hygiene: 2017 update and SDG baseline. Geneva: World Health Organization (WHO) and the United Nations Children's Fund (UNICEF), 2017 Licence: CC BY-NC-SA 3.0 IGO.

170. Rajak SN, Habtamu E, Weiss HA, Bedri A, Gebre T, Bailey RL, et al. The clinical phenotype of trachomatous trichiasis in Ethiopia: not all trichiasis is due to entropion. Invest Ophthalmol Vis Sci. 2011;52(11):7974-80. doi: 10.1167/iovs.11-7880. PubMed PMID: 21896855; PubMed Central PMCID: PMC3219424.

171. Burton MJ, Holland MJ, Faal N, Aryee EA, Alexander ND, Bah M, et al. Which members of a community need antibiotics to control trachoma? Conjunctival Chlamydia trachomatis infection load in Gambian villages. Invest Ophthalmol Vis Sci. 2003;44(10):4215-22. PubMed PMID: 14507864.

172. World Health Organization Strategic and Technical Advisory Group on Neglected Tropical Diseases. Design and validation of a trachomatous trichiasis-only survey (WHO/HTM/NTD/PCT/2017.08). Geneva: World Health Organization; 2018.

173. Ngondi J, Reacher M, Matthews F, Brayne C, Emerson P. Trachoma survey methods: a literature review. Bull World Health Organ. 2009;87(2):143-51. PubMed PMID: 19274367; PubMed Central PMCID: PMCPMC2636192.

174. Rajak SN, Habtamu E, Weiss HA, Bedri A, Zerihun M, Gebre T, et al. Why do people not attend for treatment for trachomatous trichiasis in Ethiopia? A study of barriers to surgery. PLoS Negl Trop Dis. 2012;6(8):e1766. doi: 10.1371/journal.pntd.0001766. PubMed PMID: 22953007; PubMed Central PMCID: PMC3429389.

175. Habte D, Gebre T, Zerihun M, Assefa Y. Determinants of uptake of surgical treatment for trachomatous trichiasis in North Ethiopia. Ophthalmic Epidemiol. 2008;15(5):328-33. doi: 10.1080/09286580801974897. PubMed PMID: 18850469.

176. Solomon AW ZM, Kuper H, et al. Trachoma control: a guide for program managers. Geneva: World Health Organization, 2006.

177. Bio AA, Boko PM, Dossou YA, et al. Prevalence of trachoma in northern Benin: results from 11 population-based prevalence surveys covering 26 districts. Ophthalmic Epidemiology. 2017;in press.

178. Kalua K, Chisambi A, Chinyanya D, Kamwendo Z, Masika M, Willis R, et al. Completion of Baseline Trachoma Mapping in Malawi: Results of Eight Population-Based Prevalence Surveys Conducted with the Global Trachoma Mapping Project. Ophthalmic epidemiology. 2016;23(sup1):32-8. doi: 10.1080/09286586.2016.1230224. PubMed PMID: 27726469.

179. Kalua K, Phiri M, Kumwenda I, Masika M, Pavluck AL, Willis R, et al. Baseline Trachoma Mapping in Malawi with the Global Trachoma Mapping Project (GTMP). Ophthalmic Epidemiol. 2015;22(3):176-83. doi: 10.3109/09286586.2015.1035793. PubMed PMID: 26158575; PubMed Central PMCID: PMCPMC4673584.

180. Muhammad N, Mpyet C, Adamu MD, William A, Umar MM, Goyol M, et al. Mapping Trachoma in Kaduna State, Nigeria: Results of 23 Local Government Area-Level, Population-Based Prevalence Surveys. Ophthalmic epidemiology. 2016;23(sup1):46-54. doi: 10.1080/09286586.2016.1250918. PubMed PMID: 27918227.

181. Adamu MD, Mpyet C, Muhammad N, Umar MM, Muazu H, Olamiju F, et al. Prevalence of Trachoma in Niger State, North Central Nigeria: Results of 25 Population-Based Prevalence Surveys Carried Out with the Global Trachoma Mapping Project. Ophthalmic epidemiology. 2016;23(sup1):63-9. doi: 10.1080/09286586.2016.1242757. PubMed PMID: 27918223.

182. Mpyet C, Muhammad N, Adamu MD, Muazu H, Umar MM, Abdull M, et al. Prevalence of Trachoma in Bauchi State, Nigeria: Results of 20 Local Government Area-Level Surveys. Ophthalmic epidemiology. 2016;23(sup1):39-45. doi: 10.1080/09286586.2016.1238945. PubMed PMID: 27846369.

183. Mpyet C, Muhammad N, Adamu MD, Muazu H, Umar MM, Goyol M, et al. Prevalence of Trachoma in Katsina State, Nigeria: Results of 34 District-Level Surveys. Ophthalmic epidemiology. 2016;23(sup1):55-62. doi: 10.1080/09286586.2016.1236975. PubMed PMID: 27775463.

184. Mpyet C, Muhammad N, Adamu MD, Muazu H, Umar MM, Alada J, et al. Trachoma Mapping in Gombe State, Nigeria: Results of 11 Local Government Area Surveys. Ophthalmic epidemiology. 2016;23(6):406-11. doi: 10.1080/09286586.2016.1230633. PubMed PMID: 27726459.

185. Smith JL, Sturrock HJ, Olives C, Solomon AW, Brooker SJ. Comparing the performance of cluster random sampling and integrated threshold mapping for targeting trachoma control, using computer simulation. PLoS Negl Trop Dis. 2013;7(8):e2389. doi: 10.1371/journal.pntd.0002389. PubMed PMID: 23991238; PubMed Central PMCID: PMC3749968.

186. World Health Organization Alliance for the Global Elimination of Trachoma by 2020. Second Global Scientific Meeting on Trachomatous Trichiasis. Cape Town, 4-6 November 2015 (WHO/HTM/NTD/2016.5). Geneva: World Health Organization; 2016.

187. Courtright P, Gass K, Lewallen S, MacArthur C, Pavluck A, Solomon A, et al. Global trachoma mapping project: training for mapping of trachoma (version 3) [Available at: http://www.trachomacoalition.org/node/122]. London: International Coalition for Trachoma Control; 2014.

188. Courtright P, Gass K, Lewallen S, MacArthur C, Pavluck A, Solomon A, et al. Global trachoma mapping project: training for mapping of trachoma (version 2) [Available at: http://www.trachomacoalition.org/node/357]. London: International Coalition for Trachoma Control; 2013.

189. Burton M, Habtamu E, Ho D, Gower EW. Interventions for trachoma trichiasis. Cochrane Database Syst Rev. 2015;(11):CD004008. doi: 10.1002/14651858.CD004008.pub3. PubMed PMID: 26568232; PubMed Central PMCID: PMCPMC4661324.

190. Hajian-Tilaki K. Sample size estimation in epidemiologic studies. Caspian Journal of Internal Medicine. 2011;2(4):289-98. PubMed PMID: PMC3895825.

191. Pavluck A, Chu B, Mann Flueckiger R, Ottesen E. Electronic data capture tools for global health programs: evolution of LINKS, an Android-, web-based system. PLoS neglected tropical diseases. 2014;8(4):e2654. doi: 10.1371/journal.pntd.0002654. PubMed PMID: 24722343; PubMed Central PMCID: PMCPMC3983089.

192. The Trachoma Atlas 2015. Available from: www.trachomaatlas.org.

193. Smith JL, Haddad D, Polack S, Harding-Esch EM, Hooper PJ, Mabey DC, et al. Mapping the global distribution of trachoma: why an updated atlas is needed. PLoS neglected tropical diseases. 2011;5(6):e973. doi: 10.1371/journal.pntd.0000973. PubMed PMID: 21738814; PubMed Central PMCID: PMC3125147.

194. Grothendieck G. sqldf: Perform SQL Selects on R Data Frames. 2014.

195. Canty A, Ripley B. boot: Bootstrap R (S-Plus) Functions. 2015.

196. Tillé Y, Matei A. sampling: Survey Sampling. 2015.

197. Dowle M, Srinivasan A, Short T, Lianoglou S, Saporta R, Antonyan E. data.table: Extension of Data.frame. 2015.

198. Wickham H. The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software. 2011;40(1):1-29.

199. Davison AC, Hinkley DV. Bootstrap methods and their application. Cambridge ; New York, NY, USA: Cambridge University Press; 1997. x, 582 p. p.

200. Solomon AW, Pavluck A, Courtright P, Aboe A, Adamu L, Alemayehu W, et al. The Global Trachoma Mapping Project: methodology of a 34-country population-based study. Ophthalmic Epidemiol. 2015;22(3):214-25.

201. Mahande M, Tharaney M, Kirumbi E, Ngirawamungu E, Geneau R, Tapert L, et al. Uptake of trichiasis surgical services in Tanzania through two village-based approaches. The British Journal of Ophthalmology. 2007;91(2):139-42. doi: 10.1136/bjo.2006.103374. PubMed PMID: PMC1857633.

202. Migchelsen SJ, Sepulveda N, Martin DL, Cooley G, Gwyn S, Pickering H, et al. Serology reflects a decline in the prevalence of trachoma in two regions of The Gambia. Sci Rep. 2017;7(1):15040. doi: 10.1038/s41598-017-15056-7. PubMed PMID: 29118442; PubMed Central PMCID: PMCPMC5678181.

203. Evans JR, Solomon AW. Antibiotics for trachoma. Cochrane Database Syst Rev. 2011;(3):CD001860. doi: 10.1002/14651858.CD001860.pub3. PubMed PMID: 21412875.

204. World Health Organization. Future approaches to trachoma control: report of a global scientific meeting (WHO/PBL/96.56). Geneva: World Health Organization; 1997.

205. Adamu Y, Macleod C, Adamu L, Fikru W, Kidu B, Abashawl A, et al. Prevalence of Trachoma in Benishangul Gumuz Region, Ethiopia: Results of Seven Population-Based Surveys from the Global Trachoma Mapping Project. Ophthalmic epidemiology. 2016;23(sup1):70-6. doi: 10.1080/09286586.2016.1247877. PubMed PMID: 27918248.

206. Mpyet C, Muhammad N, Adamu MD, Muazu H, Umar MM, Goyol M, et al. Prevalence of Trachoma in Kano State, Nigeria: Results of 44 Local Government Area-Level Surveys. Ophthalmic epidemiology. 2017;24(3):195-203. doi: 10.1080/09286586.2016.1265657. PubMed PMID: 28276755.

207. Ramesh A, Kovats S, Haslam D, Schmidt E, Gilbert CE. The impact of climatic risk factors on the prevalence, distribution, and severity of acute and chronic trachoma. PLoS Negl Trop Dis. 2013;7(11):e2513. doi: 10.1371/journal.pntd.0002513. PubMed PMID: 24244768; PubMed Central PMCID: PMC3820701.

208. Zomer RJ, Trabucco A, Bossio DA, Verchot LV. Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. Agriculture, Ecosystems & Environment. 2008;126(1):67-80. doi: http://dx.doi.org/10.1016/j.agee.2008.01.014.

209. ISRIC. World Soil Information, 2013. Soil property maps of Africa at 1km. www.isric.org.

210. Henderson JV, Storeygard A, Weil DN. A Bright Idea for Measuring Economic Growth. The American economic review. 2011;101(3):194-9. doi: 10.1257/aer.101.3.194. PubMed PMID: PMC4112959.

211. Chen X, Nordhaus WD. Using luminosity data as a proxy for economic statistics. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(21):8589-94. doi: 10.1073/pnas.1017031108. PubMed PMID: PMC3102367.

212. Elvidge CD, Safran J, Tuttle B, Sutton P, Cinzano P, Pettit D, et al. Potential for global mapping of development via a nightsat mission. GeoJournal. 2007;69(1):45-53. doi: 10.1007/s10708-007-9104-x.

213. Ebener S, Murray C, Tandon A, Elvidge CC. From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery. International Journal of Health Geographics. 2005;4:5-. doi: 10.1186/1476-072X-4-5. PubMed PMID: PMC549533.

214. Naimi B. R: Uncertainty analysis for SDMs: usdm-package.

215. Jiang J. Linear and Generalized Linear Mixed Models and Their Applications: Springer; 2007.

216. Bates D, Maechler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2015;67(1):1-48. doi: 10.18637/jss.v067.i01.

217. Giorgi E, Diggle PJ. PrevMap: an R package for prevalence mapping. Journal of Statistical Software. 2016.

218. Diggle PJ, Tawn JA, Moyeed RA. Model-Based Geostatistics. Applied Statistics. 1998;47(3):299-350.

219. Sahlu T, Larson C. The prevalence and environmental risk factors for moderate and severe trachoma in southern Ethiopia. J Trop Med Hyg. 1992;95(1):36-41. PubMed PMID: 1740817.

220. Katz J, Zeger SL, Tielsch JM. Village and household clustering of xerophthalmia and trachoma. Int J Epidemiol. 1988;17(4):865-9. PubMed PMID: 3265700.

221. Macharelli CA, Schellini SA, Opromolla PA, Dalben I. Spatial distribution of trachoma cases in the City of Bauru, State of Sao Paulo, Brazil, detected in 2006: defining key areas for improvement of health resources. Revista da Sociedade Brasileira de Medicina Tropical. 2013;46(2):190-5. doi: 10.1590/0037-8682-1632-2013. PubMed PMID: 23559341.

222. Prost A, Negrel AD. Water, trachoma and conjunctivitis. Bull World Health Organ. 1989;67(1):9-18. PubMed PMID: 2650903; PubMed Central PMCID: PMC2491216.

223. Bobo LD, Novak N, Munoz B, Hsieh YH, Quinn TC, West S. Severe disease in children with trachoma is associated with persistent Chlamydia trachomatis infection. J Infect Dis. 1997;176(6):1524-30. PubMed PMID: 9395364.

224. Conway DJ, Holland MJ, Bailey RL, Campbell AE, Mahdi OS, Jennings R, et al. Scarring trachoma is associated with polymorphism in the tumor necrosis factor alpha (TNF-alpha) gene promoter and with elevated TNF-alpha levels in tear fluid. Infection and immunity. 1997;65(3):1003-6. PubMed PMID: 9038309; PubMed Central PMCID: PMC175081.

225. Fitzpatrick C, Sankara DP, Agua JF, Jonnalagedda L, Rumi F, Weiss A, et al. The cost-effectiveness of an eradication programme in the end game: Evidence from guinea worm disease. PLoS Negl Trop Dis. 2017;11(10):e0005922. doi: 10.1371/journal.pntd.0005922. PubMed PMID: 28981510; PubMed Central PMCID: PMCPMC5628789.

226. Anselin L, Syabri I, Kho Y. GeoDa: An Introduction to Spatial Data Analysis. Geographical Analysis. 2006;38(1):5-22. doi: 10.1111/j.0016-7363.2005.00671.x.

227. Anselin L. Local Indicators of Spatial Association - LISA. Geographical Analysis. 1995;27:93-115.

Appendices

Appendix A R script for calculating TT prevalence normalized by age and sex

```
##This script was modified from GTMP script written by Brian Chu and Rebecca Willis
clean <- as.data.frame(read.csv("ISO_raw.csv"))
population<- read.csv("ISO_population.csv")
clean["tt_old"] <- 0
clean$tt_old[clean$TT=="1"] <- 1

##generate age groups for the analysis, update for data available for specific country
library(sqldf)
mydb_tt <- sqldf('SELECT EU, CLUSTER, (CASE
            WHEN AGE BETIEN 15 AND 19 THEN "15_19"
            WHEN AGE BETIEN 20 AND 24 THEN "20_24"
            WHEN AGE BETIEN 25 AND 29 THEN "25_29"
            WHEN AGE BETIEN 30 AND 34 THEN "30_34"
            WHEN AGE BETIEN 35 AND 39 THEN "35_39"
            WHEN AGE BETIEN 40 AND 44 THEN "40_44"
            WHEN AGE BETIEN 45 AND 49 THEN "45_49"
            WHEN AGE BETIEN 50 AND 54 THEN "50_54"
            WHEN AGE BETIEN 55 AND 59 THEN "55_59"
            WHEN AGE BETIEN 60 AND 64 THEN "60_64"
            WHEN AGE BETIEN 65 AND 69 THEN "65_69"
            WHEN AGE BETIEN 70 AND 74 THEN "70_74"
            WHEN AGE BETIEN 75 AND 79 THEN "75_79"
            WHEN AGE >= 80 THEN "80+"
            END) AS AGE_GROUP, SEX,
            COUNT(*) AS RESIDENTS,
            SUM(CASE WHEN tt_old=1 THEN 1 ELSE 0 END) AS tt
            FROM clean
            WHERE AGE >= 15
            GROUP BY EU, CLUSTER, AGE_GROUP, SEX
            ORDER BY EU, CLUSTER, AGE_GROUP ASC, SEX DESC')

##calculate unadjusted and adjusted TT prevalence
ttprev_male <- as.data.frame(sqldf('SELECT mydb_tt.EU, mydb_tt.CLUSTER,
mydb_tt.AGE_GROUP, mydb_tt.SEX, mydb_tt.RESIDENTS, mydb_tt.tt,
population.percent_age_male AS age_Iight
                        FROM mydb_tt
                        LEFT JOIN population
                        ON mydb_tt.AGE_GROUP = population.age_group
                        WHERE mydb_tt.SEX = "1"'))

ttprev_female <- as.data.frame(sqldf('SELECT mydb_tt.EU, mydb_tt.CLUSTER,
mydb_tt.AGE_GROUP, mydb_tt.SEX, mydb_tt.RESIDENTS, mydb_tt.tt,
population.percent_age_female AS age_Iight
                         FROM mydb_tt
                         LEFT JOIN population
                         ON mydb_tt.AGE_GROUP = population.age_group
                         WHERE mydb_tt.SEX = "2"'))
```

```
#append the dataframes
ttprev <- as.data.frame(rbind(ttprev_male, ttprev_female))

#This creates the new column named "prev_unadj" filled with zeros and calculates unadjusted
prevalence for each group
ttprev["prev_unadj"] <- 0;
ttprev$prev_unadj <- (ttprev$tt / ttprev$RESIDENTS)

#generate unadjusted EU-level prevalence for comparison purposes
ttprev_unadjusted_cluster <- aggregate(cbind(RESIDENTS, tt) ~ EU+CLUSTER, data=ttprev,
sum)
ttprev_unadjusted_cluster$cluster_prev_unadj <-
(ttprev_unadjusted_cluster$tt/ttprev_unadjusted_cluster$RESIDENTS)
ttprev_unadjusted <- aggregate(cluster_prev_unadj ~ EU, data=ttprev_unadjusted_cluster, mean)
colnames(ttprev_unadjusted)[colnames(ttprev_unadjusted)=="cluster_prev_unadj"] <-
"ttprev_unadj"

#This creates the new column named "adj_tt" filled with zeros and Iights the data
ttprev["adj_tt"] <- 0;
ttprev$adj_tt <- (ttprev$prev_unadj * ttprev$age_Iight)

#collapse on cluster level and get sum of Iighted prevalence
ttprev_cluster <- aggregate(adj_tt ~ EU + CLUSTER, data = ttprev, sum)

#calculate mean of adjusted cluster prevalences
ttprev_eu <- aggregate(adj_tt ~ EU, data = ttprev_cluster, mean)

a <- merge(ttprev_unadjusted, ttprev_eu, by="EU", all=TRUE)

##bootstrap to generate confidence intervals
dataset <- ttprev_cluster
str(dataset)
dataset$EU <- as.factor(dataset$EU)
dataset$cluster <- as.factor(dataset$CLUSTER)
dataset$cluster_prev <- dataset$adj_tt

#boot statistic function
clustermean <- function(df, i) {
  num_clusters <- nrow(df)
  r <- round(runif(num_clusters, 1, nrow(df))) #nrow(df) allows the analysis to divide by the correct
#clusters
  df2 <- numeric()
  for (i in 1:num_clusters) {
    df2[i] <- df[r[i],]$cluster_prev
  }
  return(mean(df2))
}

#create empty data frame for results
bootResult_tt <- data.frame(EU=character(), bootmean=numeric(), se=numeric(),
ci95_low=numeric(), ci95_high=numeric(), stringsAsFactors=FALSE)

#bootstrap function, looped over each EU
library(boot)
num_reps <- 10000 #Should be at least 1000 but preferably 10000, higher reps the more precise
for (i in 1:nlevels(dataset$EU)) {
  data2 <- subset(dataset, EU==levels(EU)[i])
  b <- boot(data2, clustermean, num_reps)
  m <- mean(b$t)
```

```
  se <- sd(b$t)

#calculate 2.5/97.5 percentiles as Confidence Interval
  q <- quantile(b$t, c(0.025, 0.975))
  ci_loIr <- q[1]
  ci_upper <- q[2]

#write result to data frame
  eu_temp <- as.character(data2$EU[1])
  bootResult_tt[i,] <- c(eu_temp, m, se, ci_loIr, ci_upper)

#histogram of mean bootstrap results with CI
  hist(b$t, breaks=50, main=paste("EU",eu_temp, "- Histogram of mean bootstrap results, n=",
num_reps))
  abline(v=ci_loIr, lty="dashed", col="black" )
  abline(v=ci_upper, lty="dashed", col="black" )
  #abline(v=ci_loIrSE, lty="dashed", col="blue" )
  #abline(v=ci_upperSE, lty="dashed", col="blue" )
}

#merge prevalence and confidence intervals into one table
b <- merge(ttprev_unadjusted, ttprev_eu, by="EU", all=TRUE)
b <- merge(b,bootResult_tt, by="EU", all=TRUE)
results_bootstrap <- merge(b,bootResult_tt, by="EU", all=TRUE)

#write combined results to csv file
write_filename <- "Results.csv"
write.csv(results_bootstrap, write_filename)
```

Appendix B TT backlog sensitivity analysis results

Country	Location	Hygiene			Sanitation				Population	TT backlog estimate
		Basic service	Limited service	No handwashing facility	At least basic	Limited service	Open defecation	Unimproved		
Angola	rural	15%	12%	73%	21%	5%	56%	17%		
	urban	37%	13%	50%	62%	27%	3%	7%		
	total	25%	12%	63%	39%	15%	33%	13%	25,789,024	3,138
Zambia	rural	5%	24%	71%	19%	7%	25%	50%		
	urban	26%	33%	41%	49%	20%	1%	30%		
	total	14%	28%	59%	31%	12%	15%	41%	12,526,314	1,524

Country	Location	Hygiene			Sanitation				Population	TT backlog estimate
		Basic service	Limited service	No handwashing facility	At least basic	Limited service	Open defecation	Unimproved		
Botswana	rural				39%	10%	36%	14%		
	urban				75%	6%	2%	16%		
	total				60%	8%	17%	15%	2,024,904	1,049
Zimbabwe	rural	24%	52%	25%	31%	15%	39%	15%		
	urban	46%	38%	16%	54%	42%	0%	4%		
	total	31%	47%	22%	39%	24%	26%	11%	13,061,239	6,765

Country	Location	Hygiene			Sanitation				Population	TT backlog estimate
		Basic service	Limited service	No handwashing facility	At least basic	Limited service	Open defecation	Unimproved		
Burundi	rural				51%	6%	3%	39%		
	urban				46%	40%	1%	13%		
	total				50%	11%	3%	36%	7,877,728	155
Uganda	rural	6%	22%	72%	17%	9%	7%	67%		
	urban	15%	21%	64%	28%	43%	2%	27%		
	total	8%	22%	71%	19%	14%	6%	60%	34,634,650	680

Country	Location	Hygiene			Sanitation				Population	TT backlog estimate
		Basic service	Limited service	No handwashing facility	At least basic	Limited service	Open defecation	Unimproved		
Namibia	rural	27%	58%	15%	15%	3%	76%	5%		
	urban	62%	28%	9%	55%	21%	20%	4%		
	total	44%	44%	12%	34%	11%	50%	5%	2,113,077	1,094
Zimbabwe	rural	24%	52%	25%	31%	15%	39%	15%		
	urban	46%	38%	16%	54%	42%	0%	4%		
	total	31%	47%	22%	39%	24%	26%	11%	13,061,239	6,765

Country	Location	Hygiene			Sanitation				Population	TT backlog estimate
		Basic service	Limited service	No handwashing facility	At least basic	Limited service	Open defecation	Unimproved		
Iran (Islamic Republic of)	rural				79%	17%	2%	3%		
	total				88%	10%	1%	1%	75,149,669	1,928
	urban				92%	7%	0%	1%		
Pakistan	rural	46%	43%	11%	48%	9%	19%	24%		
	total	60%	31%	8%	58%	8%	12%	22%	207,774,520	5,330
	urban	83%	12%	5%	74%	8%	0%	18%		

Country	Location	Hygiene			Sanitation				Population	TT backlog estimate
		Basic service	Limited service	No handwashing facility	At least basic	Limited service	Open defecation	Unimproved		
Iraq	rural	81%	7%	12%	60%	9%	0%	5%		
	total	91%	4%	5%	54%	10%	0%	4%	19,184,543	492
	urban	95%	2%	2%	51%	11%	0%	3%		
Pakistan	rural	46%	43%	11%	48%	9%	19%	24%		
	total	60%	31%	8%	58%	8%	12%	22%	207,774,520	5,330
	urban	83%	12%	5%	74%	8%	0%	18%		

Country	Location	Hygiene			Sanitation				Population	TT backlog estimate
		Basic service	Limited service	No handwashing facility	At least basic	Limited service	Open defecation	Unimproved		
Mexico	rural	80%	15%	5%	81%	8%	6%	6%		
	total	88%	9%	3%	44%	7%	2%	2%	112,336,538	5,428
	urban	90%	8%	2%	45%	7%	1%	1%		
Guatemala	rural	70%	27%	3%	53%	7%	10%	30%		
	total	77%	21%	3%	67%	9%	6%	18%	11,237,196	543
	urban	83%	14%	2%	81%	10%	1%	8%		

Country	Location	Hygiene			Sanitation				Population	TT backlog estimate
		Basic service	Limited service	No handwashing facility	At least basic	Limited service	Open defecation	Unimproved		
Nauru	total				66%	31%	3%	1%	10,084	1
	urban				66%	31%	3%	1%		
Vanuatu	rural				51%	13%	2%	34%		
	total				53%	18%	2%	27%	515,870	48
	urban				61%	32%	1%	6%		

Appendix C Estimated national-level trichiasis backlogs, 2016, with comparisons to the corresponding estimates for 2009

WHO region	Country	2009 published estimate generated from 2007 data [162]	Source for 2016 estimate (1)	2016 estimate	Year(s) of data collection used for 2016 estimate (2)	Representativeness
Africa Region	Algeria	86,700	Retained previous estimate	86,700		
	Angola	no data	Expert assessment (3)	0		
	Benin	7,600	GTMP [168]	11,782	2014-2015	Estimate based on a backlog calculated from 12 age- and sex-adjusted PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
	Botswana	32,900	Expert assessment (3)	0		
	Burkina Faso	32,800	PBPS – no raw data available	19,443	2005-2015	Estimate based on a backlog calculated from 63 PBPS datasets covering all districts in the country
	Burundi	no data	Expert assessment (3)	0		
	Cameroon	47,200	PBPS – no raw data available	698	2012-2014	Estimate based on a backlog calculated from 1) 35 (70%) age- and sex-adjusted PBPS datasets, plus 2) 15 (30%) PBPS datasets, covering all districts in the country where evidence indicates trachoma may be a public health problem
			PBPS – raw data available	10,743	2010-2014	
	Central African Republic	1,000	PBPS – no raw data available	6,539	2011	Estimate based on a backlog calculated from 8 (53%) PBPS datasets, security concerns prevented the remaining 7 (47%)districts in the country where evidence indicates trachoma may be a public health problem from being surveyed
	Chad	34,300	GTMP	24,597	2014-2015	Estimate based on a backlog calculated from 1) 43 (74%) age- and sex-adjusted PBPS datasets, plus 2) 14 (24%) PBPS datasets. At the time of publication 1 (2%)
			PBPS – no raw data available	23,024	2002	

WHO region	Country	2009 published estimate generated from 2007 data [162]	Source for 2016 estimate (1)	2016 estimate	Year(s) of data collection used for 2016 estimate (2)	Representativeness
						district where evidence indicates trachoma may be a public health problem remained to be surveyed
	Congo	no data	GTMP	0	2015	Estimate based on a backlog calculated from 1 age- and sex-adjusted PBPS dataset covering all districts in the country where evidence indicates trachoma may be a public health problem
	Côte d'Ivoire	59,900	GTMP	1,216	2015	Estimate based on a backlog calculated from 10 age- and sex-adjusted PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
	Democratic Republic of the Congo	no data	GTMP[169]	33,333	2014-2016	Estimate based on a backlog calculated from 30 (75%) age- and sex-adjusted PBPS datasets, security concerns prevented 1 (3%) district from being surveyed. At the time of publication an additional 9 district (23%) where evidence indicates trachoma may be a public health problem remained to be surveyed.
	Eritrea	42,000	GTMP	774	2014	Estimate based on a backlog calculated from 1) 2 (5%) age- and sex-adjusted PBPS datasets, plus 2) 36 (92%) PBPS datasets, security concerns prevented the remaining 1 (3%) district in the country where evidence indicates trachoma may be a public health problem from being surveyed
			PBPS – no raw data available	24,793	2006-2014	
	Ethiopia	1,272,600	Expert assessment [170-174] (4)	693,037	2012-2016	Expert assessment based on a backlog calculated from 1) 196 (92%) age- and sex-adjusted PBPS datasets, security concerns prevented the remaining 16 districts (8%) in the country where evidence indicates trachoma may be a public health problem from being surveyed, minus 2) programmatic TT surgery output from 2012-2016
	Gambia	10,500	PBPS – no raw data available	0	2016	PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem

WHO region	Country	2009 published estimate generated from 2007 data [162]	Source for 2016 estimate (1)	2016 estimate	Year(s) of data collection used for 2016 estimate (2)	Representativeness
	Ghana	3,000	PBPS – no raw data available	1,379	2007-2008	Estimate based on a backlog calculated from 23 PBPS datasets, covering all districts in the country where evidence indicates trachoma may be a public health problem
	Guinea	25,100	GTMP	5,523	2014-2016	Estimate based on a backlog calculated from 31 age- and sex-adjusted PBPS datasets, covering all districts in the country where evidence indicates trachoma may be a public health problem
			PBPS – raw data available	24,302	2011-2013	
	Guinea-Bissau	16,400	PBPS – raw data available	21,255	2005	Estimate based on a backlog calculated from 11 PBPS datasets, covering all districts in the country where evidence indicates trachoma may be a public health problem
	Kenya	306,800	PBPS – no raw data available	30,195	2004-2012	Estimate based on a backlog calculated from 1) 3 (20%) age- and sex-adjusted PBPS datasets, plus 2) 12 (80%) PBPS datasets, covering all districts in the country where evidence indicates trachoma may be a public health problem
			PBPS – raw data available	21,363	2004	
	Malawi	33,400	GTMP [175, 176]	13,446	2013-2015	Estimate based on a backlog calculated from 1) 30 (91%) age- and sex-adjusted PBPS datasets, plus 2) 3 (9%) PBPS datasets, covering all districts in the country where evidence indicates trachoma may be a public health problem
			PBPS – no raw data available	1,128	2012	
			PBPS – raw data available	817	2014	
	Mali	67,600	Expert assessment (5)	13,852		Estimate based on a backlog calculated from 55 PBPS datasets, covering all districts in the country
	Mauritania	2,500	PBPS – no raw data available	1,556	2004-2013	Estimate based on a backlog calculated from 29 PBPS datasets, covering all districts in the country where evidence indicates trachoma may be a public health problem
	Mozambique	60,500	GTMP [177]	18,817	2013-2014	Estimate based on a backlog calculated from 1) 99 age- and sex-adjusted PBPS datasets, covering all districts in the country where evidence indicates trachoma may be a public health problem
			PBPS – raw data available	96	2015	

WHO region	Country	2009 published estimate generated from 2007 data [162]	Source for 2016 estimate (1)	2016 estimate	Year(s) of data collection used for 2016 estimate (2)	Representativeness
	Namibia	6,100	Expert assessment (3)	0		
	Niger	59,600	PBPS – no raw data available	40,529	2009-2014	Estimate based on a backlog calculated from 49 PBPS datasets, covering all districts in the country where evidence indicates trachoma may be a public health problem
	Nigeria	627,300	GTMP [178-182]	193,951	2013-2014	Estimate based on a backlog calculated from 1) 294 (66%) age- and sex-adjusted PBPS datasets, plus 2) 103 (23%) PBPS datasets, security concerns prevented the remaining 47 (11%) districts in the country where evidence indicates trachoma may be a public health problem from being surveyed
			PBPS – no raw data available	90,865	2003-2014	
			PBPS – raw data available	3,800	2014-2015	
	Senegal	129,800	GTMP	12,707	2014	Estimate based on a backlog calculated from 1) 20 (43%) age- and sex-adjusted PBPS datasets, plus 2) 27 (57%) PBPS datasets, covering all districts in the country where evidence indicates trachoma may be a public health problem
			PBPS – no raw data available	30,905	2000-2012	
			PBPS – raw data available	3,077	2015	
	South Sudan	no data	Expert assessment	81,093		Estimate provided by the Ministry of Health, South Sudan. To date, only a quarter of South Sudan has been mapped – this figure is likely to be an under-estimate for the country
	Togo	2,900	PBPS – no raw data available	318	2009	Estimate based on a backlog calculated from 3 PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
	Uganda	610,600	GTMP	680	2014	Estimate based on a backlog calculated from 1) 39 (89%) age- and sex-adjusted PBPS datasets, plus 2) 5 (11%) PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
			PBPS – no raw data available	14,559	2008-2012	
			PBPS – raw data available	65,329	2013-2016	
	United Republic of Tanzania	214,800	GTMP [183, 184]	8,096	2014	Estimate based on a backlog calculated from 97 age- and sex-adjusted PBPS datasets, covering all districts in the

WHO region	Country	2009 published estimate generated from 2007 data [162]	Source for 2016 estimate (1)	2016 estimate	Year(s) of data collection used for 2016 estimate (2)	Representativeness
			PBPS – raw data available	63,157	2004-2016	country where evidence indicates trachoma may be a public health problem
	Zambia	8,500	GTMP	1,524	2015	Estimate based on backlog calculated form 1) 7 (70%) age- and sex- adjusted PBPS datasets, plus 2) 3 (30%) PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
			PBPS – no raw data available	2,168	2010-2015	
	Zimbabwe	44,100	GTMP	6,765	2014-2015	Estimate based on a backlog calculated from 16 age- and sex-adjusted PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
Mediterranean Region	Afghanistan	83,100	Retained previous estimate	83,100		
	Djibouti	3,900	Expert assessment (6)	75		
	Egypt	35,400	GTMP	35,362	2015	Estimate based on a backlog calculated from 1) 4 (80%) age- and sex-adjusted PBPS datasets, plus 2) 1 (20%) PBPS dataset
			PBPS – no raw data available	35,400	2015	
	Iran (Islamic Republic of)	49,300	Expert assessment (3)	0		

WHO region	Country	2009 published estimate generated from 2007 data [162]	Source for 2016 estimate (1)	2016 estimate	Year(s) of data collection used for 2016 estimate (2)	Representativeness
	Iraq	43,900	Expert assessment (3)	0		
	Libya	13,200	Expert assessment (7)	33,400		
	Morocco	6,400	PBPS – no raw data available	0	2015	Estimate based on a backlog calculated from PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
	Oman	600	PBPS – no raw data available	600	2005	PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
	Pakistan	71,700	GTMP	5,330	2015	Estimate based on a backlog calculated from 1) 42 (47%) age- and sex-adjusted PBPS datasets, plus 2) 31 (35%) PBPS datasets, security concerns prevented 15 (17%) districts from being surveyed. An additional 1 (1%) district where evidence indicates trachoma may be a public health problem remains to be surveyed.
			PBPS – no raw data available	23,420	2012	
	Somalia	10,300	Retained previous estimate	10,300		
	Sudan	528,100	GTMP [185]	22,508	2014-2015	Estimate based on a backlog calculated from 1) 109 (76%) age- and sex-adjusted PBPS datasets, plus 2) 22 (15%) PBPS datasets, security concerns prevented 12

WHO region	Country	2009 published estimate generated from 2007 data [162]	Source for 2016 estimate (1)	2016 estimate	Year(s) of data collection used for 2016 estimate (2)	Representativeness
			PBPS – no raw data available	6,751	2006-2013	(8%) districts from being surveyed. At the time of publication an additional 1 (1%) district where evidence indicates trachoma may be a public health problem remained to be surveyed.
			PBPS – raw data available	36,982	2007-2010	
	Yemen	270,800	GTMP	5,821	2013-2015	Estimate based on a backlog calculated from 1) 42 (53%) age- and sex-adjusted PBPS datasets, security concerns prevented the remaining 38 districts in the country where evidence indicates trachoma may be a public health problem from being surveyed
Region of the Americas	Brazil	58,000	Retained previous estimate	58,000	2003-2006	PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
	Colombia	no data	GTMP	48	2015	Estimate based on a backlog calculated from 1) 3 (25%) age- and sex-adjusted PBPS datasets, plus 2) 6 (50%) PBPS datasets. At the time of publication an additional 3 (25%) districts where evidence indicates trachoma may be a public health problem remained to be surveyed.
			PBPS – no raw data available	23	2003-2009	
	Guatemala	30	PBPS – no raw data available	543	2011	Estimate based on a backlog calculated from 2 PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
	Mexico	20	Expert assessment (3)	0		
South-East Asia Region	India	443,000	Retained previous estimate	443,000		
	Nepal	138,800	PBPS – no raw data available	25,240	2001-2012	Estimate based on a backlog calculated from 1) 30 (64%) age- and sex-adjusted PBPS datasets, plus 2) 17

WHO region	Country	2009 published estimate generated from 2007 data [162]	Source for 2016 estimate (1)	2016 estimate	Year(s) of data collection used for 2016 estimate (2)	Representativeness
			PBPS – raw data available	3,252	2002-2009	(36%) PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
Western Pacific Region	Australia	1,100	Retained previous estimate	1,100		
	Cambodia	29,200	GTMP [186]	4,999	2014-2015	Estimate based on a backlog calculated from 14 age- and sex-adjusted PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
	China	2,330,600	PBPS – no raw data available	71,280		PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
	Fiji	800	PBPS using the methods of the GTMP [187] and subsequent investigation [188]	0	2013	Estimate based on a backlog calculated from 1 age- and sex-adjusted PBPS dataset covering all districts in the country where evidence indicates trachoma may be a public health problem
	Kiribati	100	GTMP	69	2015	Estimate based on a backlog calculated from 1 age- and sex-adjusted PBPS dataset covering all districts in the country where evidence indicates trachoma may be a public health problem
	Lao People's Democratic Republic	900	GTMP [189]	630	2013-2015	Estimate based on a backlog calculated from 16 age- and sex-adjusted PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
	Myanmar	65,800	Retained previous estimate	65,800		
	Nauru	0	Expert assessment (3)	0		

WHO region	Country	2009 published estimate generated from 2007 data [162]	Source for 2016 estimate (1)	2016 estimate	Year(s) of data collection used for 2016 estimate (2)	Representativeness
	Papua New Guinea	5,800	GTMP [190]	156	2015	Estimate based on a backlog calculated from 7 age- and sex-adjusted PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
	Solomon Islands	500	GTMP [191]	44	2013	Estimate based on a backlog calculated from 1) 3 (38%) age- and sex-adjusted PBPS datasets, plus 2) 5 (63%) PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
			PBPS – no raw data available	8	2012	
	Vanuatu	200	GTMP [192]	48	2014	Estimate based on a backlog calculated from 1 age- and sex-adjusted PBPS datasets covering all districts in the country where evidence indicates trachoma may be a public health problem
	Viet Nam	210,000	Expert assessment (6) [108]	100,000		

1) Unless otherwise specified, "Retained previous estimate" refers to the 2009 estimate by Mariotti et al.[163]

2) Individuals examined in these surveys were men and women aged ≥15-years

3) Health ministry reports that there is no evidence of trichiasis being a public health problem, or that evidence indicates that trichiasis is not a public health problem

4) Estimate provided by the Federal Ministry of Health of Ethiopia; determined by calculating the backlog indicated by the most recent population-based prevalence survey in each trachoma-endemic district and subtracting from it the number of individuals with trichiasis managed by the health system since those surveys

5) Estimate provided by the Ministère de la Santé, Mali; determined by calculating the backlog indicated by the most recent population-based prevalence survey in each trachoma-endemic district and subtracting from it the number of individuals with trichiasis managed by the health system since those surveys

6) Estimate derived from information provided at the 2010 WHO Alliance for GET2020 meeting

7) Estimate informed by data from a Rapid Assessment of Avoidable Blindness (RAAB) survey

Appendix D A trichiasis specific survey pilot test protocol

LIST OF ABBREVIATIONS

TF	Trachomatous inflammation follicular
TT	Trachomatous trichiasis
UIG	Ultimate intervention goal
WHO	World Health Organization
MoH	Ministry of Health
ICTC	International Coalition for Trachoma Control
EU	Evaluation Unit

Summary

A TT specific survey has been designed with the intention of measuring trachomatous trichiasis (TT) with precision suitable for measurement against the WHO elimination criteria. This survey needs to be validated in the field and so a pilot test in Cameroon, Chad, Uganda and Tanzania is recommended.

The pilot will involve a 2 stage sampling of one evaluation unit (district) from each country where all individuals living in the selected households in half of the selected clusters will be examined for signs of TT based on the WHO Simplified Trachoma Grading Scheme and the 40-years-and-older population will be examined in the remaining clusters. When TT is found in an eye, the eyelid of that eye will be everted and inspected for trachomatous scarring (TS) based on the WHO Simplified Trachoma Grading Scheme. The mobile phone number of those with positive cases will be collected. This will help to ensure follow-up with these individuals. Data will be collected and stored using the LINKS electronic data collection system and GPS coordinates will be collected at each selected household.

All survey activities will be implemented by national control programmes as part of their activities, and will allow estimation of TT prevalence for programmatic purposes. However, data collection will also include oversampling for the purpose of validating the survey design. Findings will directly inform cost-effective design of national trachoma elimination programmes, and may be used to further refine WHO guidelines for the control of trachoma.

Background and Rationale

Trachoma is a leading cause of preventable blindness worldwide and is endemic in 51 countries [1]. Prolonged conjunctival infection with Chlamydia trachomatis leads to inflammatory response, trachomatous follicular (TF), which results in scarring of the upper sub-tarsal conjunctiva. The highest rates of TF are found in young children with a decrease occurring around the school-age [2-4]. Over time, repeat infection and additional scarring can cause the eyelid to turn inward, resulting in lashes rubbing against the cornea. This painful state is called trachomatous trichiasis (TT) and can damage the cornea, leading to vision impairment or blindness [5]. TT incidence increase with age, being rare in children. Whilst this disease is painful and debilitating, there is a strategy to stop its progress. The "SAFE" strategy involves (S) surgery to correct trichiasis, (A) mass drug administration of azithromycin and improved sanitation and hygiene, focusing on (F) facial cleanliness and (E) environmental improvement [6].

The World Health Assembly Resolution 51.11 of 1998 targets the elimination of trachoma as a public health problem by the year 2020 [6]. A series of global trachoma scientific meetings (GTSM) resulted in generating the ultimate intervention goal for TF as, less than 5% prevalence and TT as less than 1 case unknown to the health system per 1,000 population per health district [7].

Because trichiasis is the blinding stage of trachoma, appropriate management of individuals with trichiasis is the priority of every trachoma elimination programme. Obtaining reasonably precise data on TT prevalence helps programmes to plan surgical services, monitor progress, and assess whether or not the TT prevalence components of

the elimination goals has been successfully accomplished. Until now, the prevalences of TF in 1-9 year-olds and TT in adults (or the whole population) have usually been measured at the same time, as has been done at baseline for >1500 districts worldwide as part of the Global Trachoma Mapping Project. However, as programmes evolve, there are four broad scenarios where a TT specific survey is necessary.

1. If a pre-intervention survey finds TF below 5% and TT above 0.1%, an impact survey to again measure the prevalence of TF is not indicated.
2. If a pre-intervention survey finds TF above 30% and TT above 0.1% the district will undergo a minimum of 5 years of intervention before an impact assessment to again measure the prevalence of TF. During this time a programme may want to re-evaluate TT prevalence rather than wait the full 5+ years to make adjustments to the surgery strategy.
3. If a post-intervention impact assessment finds TF below 5% and TT above 0.1%, further work to manage cases of TT should be implemented, and an attempt then made to re-measure the TT prevalence.
4. If a previous survey finds TT below 0.1%, but the quality of the survey is in question and expert opinion suggests TT should be re-evaluated.

Objectives

The purpose of this survey is to estimate the prevalence of trichiasis within a population with precision against the WHO threshold of 1/1,000 total population.

Methods

The survey will be carried out utilizing existing electronic data collection system, LINKS. The preferred system is LINKS because many programmes are already familiar with the android tool. Best practices for data management will be used. This includes regularly calculating descriptive statistics and generating point maps during the data collection process. This ensures high data quality at the completion of the survey. The data will be stored on a secure server which is regularly backed up, as this greatly reduces the risk of data loss. Age and sex adjusted prevalence calculations will be provided to the Ministry of Health (MOH) for review. Upon the approval from the MOH it is recommended that the national programme share the results with WHO to facilitate global monitoring of progress and assistance with alignment of resources for intervention with country-specific needs.

Survey Design

Survey Sites

Inclusion criteria

An evaluation unit (EU) should typically be the normal administrative unit for health care management, consisting of a population between 100,000-250,000 persons [8]. In situations where population size is very large, country-specific protocols should be followed to create smaller evaluation units.

Within the EU, trichiasis cases should be suspected. The basis for this suspicion can come from results of previous trachoma surveys, evidence of cases within

healthcare facilities or confirmation of individual cases by healthcare professionals in the field.

Exclusion criteria

EUs are excluded from the study if they are not "suspected" to be endemic to trachoma or if previous surveys accepted by national programmes estimate a TT prevalence of less than 0.1%.

Survey Population

Inclusion criteria

All individuals living in the selected households in half the selected clusters and individuals 40-years-and-older in the remaining clusters are eligible for the study.

Exclusion criteria

Individuals who do not provide consent are not eligible for the study.

Recruitment Procedure

Sample size

I assume intervention reduces TT uniformly across the whole population. I further assume the 40+ age group constitutes about 17% (this value was derived from reviewing national population pyramids) of the population and is responsible for 85% of cases of TT. If these assumptions are correct, then a prevalence of 0.1% or less in

the whole population would correspond to a prevalence of 0.5% or less in the 40+ age group (0.1% x 85% divided by 17%).

The mean design effect was calculated from the 313 surveys. I chose to use 2 as our design effect, as this value covers a minimum of 80% of the EUs in the extreme scenarios.

Therefore, our null hypothesis states the prevalence is 0.5%.

$n = (Z^2 \times P(1 - P))/e^2$

z = standard normal deviate corresponding to 95% confidence intervals

P = hypothesised prevalence

e = precision

$$765 = (1.96^2 \text{ x } 0.005 (1 - 0.005)) / 0.005^2$$

$$765*2 = 1,529$$

Because we are using this survey as a validation exercise, we will over sample.

A sample of 1,529 adults 40-years-and-old is the minimum sample. So, we will sample the 40-years-and-older population in half the selected clusters and all individuals 1-year-and-older in the remaining clusters and increase the number of clusters (60 clusters).

Sample selection

The survey should be representative of the entire EU and so two stage sampling strategy is required.

First, a list of clusters (villages) should be selected through a random sampling procedure using probability proportional to population size.

For this step it will be necessary to obtain a list of all villages and their population in the EU.

1529/(30×1.5)=34

Sample size = 1529

Households to visit in one day = 30

Average number of 40+ per household = 1.5

Minimum number of clusters = 34

**because we are using this survey to validate the sampling we will survey 60 clusters in an effort to oversample suitable for computer simulations*

Second, households should be randomly selected within communities. All eligible individuals who live in the selected households will be included in the sample. If eligible individuals are not home at the time of the survey, the team should return to the household at a time when the individuals are expected to be home. It is important to avoid convenience sampling in these scenarios.

For this step it will be necessary to obtain a list of all households within the selected villages.

If there is no list of households, compact segment sampling will be used.

Questionnaire

The survey team should ask the patient for basic demographic information. GPS data should be collected at every household.

Clinical examination

Examining an individual from trichiasis involves looking at the eye in its primary position of gaze, using a 2.5 x magnifying loupe on the examiner's head, and illumination, which can be either sunlight or a household torch. The eye is examined from the front, and from the side. In some individuals, where there is uncertainty as to whether there may be eyelashes touching the eyeball (which may occur, for example, if the iris is very dark in colour), or where there is uncertainty as to whether there may be evidence of epilation, the upper eyelid may need to be gently pushed upwards by exerting very mild pressure on the skin overlying the upper orbital margin, without eversion of the eyelid, in order to expose the free edge of the eyelid. This examination takes less than ten seconds per eye and is simple and painless.

In individuals who are found to have trichiasis, questions about previous management are asked. The eyelid is then everted to allow examination of the tarsal conjunctiva for evidence of trachomatous scarring (TS). Eversion requires the examiner to ask the subject to look down, grasp the eyelashes of the upper eyelid between thumb and forefinger, pull the eyelid very gently downwards, then flip the outer edge upwards over

the point of the index or little finger of the other hand. Such a procedure causes mild discomfort in children, but in adults with trichiasis, the eversion manoeuvre temporarily relieves the discomfort of eyelashes abrading the eyeball and is extremely well tolerated.

All Graders should have completed an objective structured clinical examination workshop to ensure grader quality.

An objective structured clinical examination (OSCE) for trichiasis grading

Station 1: The grader trainee will examine the grader trainer, following the standard sequence for examination

The trainer will evaluate whether the trainee follows the correct sequence for examination, which is

1. Grader puts on the loop and cleans hands
2. Patient sits and looks straight
3. With illumination the grader examines the eyes to see if the lid margin and lashes are visible
4. Grader slightly lifts upper eyelid to better examine the lid margin and position of the lashes
5. Grader asks patient to move the eyeball to the right and the left to see if any eyelashes come in contact with the cornea

Stations 2, 3 and 4: The grader trainee will demonstrate his or her ability to identify trichiasis and the absence of trichiasis in a series of photographs

Grader should be able correctly identify the following:

- neither trichiasis nor entropion
- upper lid trichiasis with upper lid entropion
- upper lid major trichiasis without entropion (>5 eyelashes touching the globe)
- upper lid minor trichiasis without entropion (1-5 eyelashes touching the globe)
- evidence of epilation

When the trainee identifies an abnormal eye in a photograph, he or she will be expected to ask the grader trainer the appropriate questions about previous management of trichiasis, acting as if the eye pictured belonged to the trainer.

Station 5: The grader trainee will demonstrate his or her ability to evert both right and left eyelids of a normal individual and examine the conjunctivae for trachomatous scarring.

Ethical considerations

No personal identifiers are collected in the electronic database. Personal identifiers collected on the paper consent form will remain in-country. Written informed consent will be obtained from all individuals participating in the project. The survey team will be trained to not coerce persons to provide consent. The consent form will be explained to all participants in their local language.

Informed consent/assent

Introduction

Greetings, my name is _____.I work with the [Ministry of Health or Implementing Partner Name]. We are visiting select households in your community and other communities in your district to find out about trachoma control. The purpose of this survey is to provide information to guide planning of efficient surgical campaigns. Understanding where trichiasis cases are located will help ensure that all individuals with trichiasis who would like surgery receive surgery.

Procedures

First, I would like to talk with you for about 10 minutes and ask a few standard questions. I am going to ask you some questions about yourself.

I will then look at the eyelids of everyone aged 40 year and above in your household to see if they have trichiasis. The eyes will be visually examined from the front, and from the sides. If there is uncertainty as to whether there may be eyelashes touching the eyeball, or where there is uncertainty as to whether there may be evidence of epilation, the upper eyelid may need to be gently pushed upwards by exerting very mild pressure on the skin. This examination takes less than 10 seconds per eye and is very simple and painless.

If trichiasis is found, questions about previous management will be asked. The eyelid will then be everted to allow examination of the tarsal conjunctiva for evidence of trachomatous scarring. The examiner will grasp the eyelashes of the upper eyelids between thumb and forefinger, pull the eyelid very gently downwards, then flip the

outer edge upwards over the point of the index or little finger of the other hand. This procedure temporarily relieves the discomfort of eyelashes abrading the eyeball and is extremely well tolerated in adults.

Risks

Examination of eyes for trichiasis is simple and painless. Everting the eyelid is very well tolerated in adults and often relieves discomfort.

Benefits

Any persons found with trichiasis will be told where they can receive free eyelid surgery. Information from this survey will help the programme in planning ways of making sure every person has access to surgery.

Privacy

Only the people leading the evaluation will see the results of your tests. Your name will not be given to anyone. Your name will not be collected in the evaluation interview – only on this consent form. Village test results will be shared with the district health officer and other health officials and programmes only for the purpose of following up with needed surgery. Information will be kept private as allowed by law. Your name will not appear in any report that comes from this evaluation.

Right not to participate or to stop participating

You are free to choose to be a part of this evaluation. Even if you agree to be in this evaluation, you may stop at any time. If you decide not to be in this evaluation, you and your family will not lose any benefits.

Whom to contact about the evaluation

If you have any questions, feel that you have become sick because of this evaluation, or want to stop being in the evaluation, you may call _____ of the _____ at _____.

Consent to participate in the evaluation

By signing below, I agree that I have read this consent form or someone has explained it to me. I have had all my questions about the evaluation answered. I agree to participate in the evaluation.

Participant (or parent/guardian)	Signature or thumb print
Name (print)	
Date / Place	

Signature of witness if participant is unable to read

By signing below, I confirm that this consent form was read in its entirety to the participant. The participants had his/her questions answered and gave his/her thumb print freely.

Witness	Signature
Name (print)	
Date / Place	

Data collection tools

Before starting the questionnaire, obtain informed consent from participants. No personal identifiers are collected in the electronic database. Personal identifiers collected on the paper consent form will remain in-country.

Location Variables	
Name of District	User entered text
EU ID	User entered text
Cluster ID	User entered text
Recorder ID	User entered text
Name of Recorder	Linked to RECORDER_CODE
Date and time record started	Automated timestamp
Date and time record completed	Automated timestamp
Capture GPS data—latitude	Automated capture of coordinates
Capture GPS data—longitude	Automated capture of coordinates
Capture GPS data—altitude	Automated capture of coordinates
Capture GPS data—accuracy	Automated capture of coordinates

Demographic Variables		
Examination status	1	Yes (with consent)
	2	Absent
	3	Refused
	4	Other
Age (years)	User entered integer (must be over 39 years)	
Sex	1	Male
	2	Female

Clinical Examination Variables (Right Eye)		
TT presence in right eye (upper lid)	0	Sign absent
	1	Sign present
	2	Not able to grade
If TT presence in right eye (upper lid) is present then	0	Sign absent
	1	Sign present

TS presence in right eye	2	Not able to grade
TT presence in right eye (lower lid)	0	Sign absent
	1	Sign present
	2	Not able to grade
If TT presence in right eye (lower lid) is present then TS presence in right eye	0	Sign absent
	1	Sign present
	2	Not able to grade

Known to the System Variables (if TT positive in right eye)		
Have you ever been offered surgery by a health worker to correct the trichiasis (in-turned eyelashes) in this eye?	1	Yes, a health worker informed me and offered me surgery, and I had surgery
	2	Yes, a health worker informed me and offered me surgery and I accepted the offer but I did not get surgery
	3	Yes, a health worker informed me and offered me surgery, but I declined it
	4	No health worker informed me and offered me surgery
Have you ever been offered epilation by a health workers to correct the trichiasis (in-turned eyelashes) in this eye?	1	Yes
	2	No
	3	Don't know

Clinical Examination Variables (Left Eye)		
TT presence in left eye (upper lid)	0	Sign absent
	1	Sign present
	2	Not able to grade
If TT presence in left eye (upper lid) is present then TS presence in left eye	0	Sign absent
	1	Sign present
	2	Not able to grade
TT presence in left eye (lower lid)	0	Sign absent
	1	Sign present
	2	Not able to grade
If TT presence in left eye (lower lid) is present then TS presence in left eye	0	Sign absent
	1	Sign present
	2	Not able to grade

Known to the System Variables (if TT positive in Left eye)		
Have you ever been offered surgery by a health worker to correct the trichiasis (in-turned eyelashes) in this eye?	1	Yes, a health worker informed me and offered me surgery, and I had surgery
	2	Yes, a health worker informed me and offered me surgery and I accepted the offer but I did not get surgery
	3	Yes, a health worker informed me and offered me surgery, but I declined it
	4	No health worker informed me and offered me surgery
Have you ever been offered epilation by a health workers to correct the trichiasis (in-turned eyelashes) in this eye?	1	Yes
	2	No
	3	Don't know

1. Progress report on elimination of trachoma, 2013, in weekly epidemiological record. 2014, World Health Organization: Geneva, Switzerland. p. 421-428.
2. west, S.K., et al., The epidemiology of trachoma in central Tanzania. Int J Epidemiol, 1991. 20(4): p. 1088-92.
3. Schwab, L., et al., The epidemiology of trachoma in rural Kenya. Variation in prevalence with lifestyle and environment. Study Survey Group. Ophthalmology, 1995. 102(3): p. 475-82.
4. Zerihun, N., Trachoma in Jimma zone, south western Ethiopia. Trop Med Int Health, 1997. 2(12): p. 1115-21.
5. Mabey, D.C., A.W. Solomon, and A. Foster, Trachoma. Lancet, 2003. 362(9379): p. 223-9.
6. Global elimination of blinding trachoma, in 51st World Health Assembly. 1998, World Health Assembly: Geneva.
7. Report of the 3rd global scientific meeting on trachoma. 2010, Geneva: World Health Organization: Johns Hopkins University, Baltimore, MA, 19-20 July 2010.
8. Solomon AW, Z.M., Kuper H, et al, Trachoma control: a guide for programme managers. 2006, World Health Organization: Geneva.

Appendix E R script for TT survey design validation

TT Age Group Prevalence *modified from GTMP code written by R. Willis and B. Chu

```
clean <- as.data.frame(read.csv("TT dataset.csv"))

population<- read.csv("population dataset.csv")


clean["tt_old"] <- 0

clean$tt_old[clean$resident_left_eye_upper_tt=="1"] <- 1

clean$tt_old[clean$resident_left_eye_loIr_tt=="1"] <- 1

clean$tt_old[clean$resident_right_eye_upper_tt=="1"] <- 1

clean$tt_old[clean$resident_right_eye_loIr_tt=="1"] <- 1


clean["tt_true"] <- 0

clean$tt_true[clean$resident_left_eye_upper_tt=="1" & clean$resident_scarring_left_eye=="1"] <-
1

clean$tt_true[clean$resident_left_eye_loIr_tt=="1" & clean$resident_scarring_left_eye=="1"] <- 1

clean$tt_true[clean$resident_right_eye_upper_tt=="1" & clean$resident_scarring_right_eye=="1"]
<- 1

clean$tt_true[clean$resident_right_eye_loIr_tt=="1" & clean$resident_scarring_right_eye=="1"] <-
1

clean$tt_true[clean$resident_left_eye_upper_tt=="1" & clean$resident_scarring_left_eye=="2"] <-
1

clean$tt_true[clean$resident_left_eye_loIr_tt=="1" & clean$resident_scarring_left_eye=="2"] <- 1

clean$tt_true[clean$resident_right_eye_upper_tt=="1" & clean$resident_scarring_right_eye=="2"]
<- 1

clean$tt_true[clean$resident_right_eye_loIr_tt=="1" & clean$resident_scarring_right_eye=="2"] <-
1


###Source census-specific code to generate age groups for the analysis

library(sqldf)

mydb_tt <- sqldf('SELECT cluster_eu, cluster_cluster, (CASE

            WHEN resident_age BETIEN 15 AND 19 THEN "15_19"

            WHEN resident_age BETIEN 20 AND 24 THEN "20_24"

            WHEN resident_age BETIEN 25 AND 29 THEN "25_29"
```

```
            WHEN resident_age BETIEN 30 AND 34 THEN "30_34"

            WHEN resident_age BETIEN 35 AND 39 THEN "35_39"

            WHEN resident_age BETIEN 40 AND 44 THEN "40_44"

            WHEN resident_age BETIEN 45 AND 49 THEN "45_49"

            WHEN resident_age BETIEN 50 AND 54 THEN "50_54"

            WHEN resident_age BETIEN 55 AND 59 THEN "55_59"

            WHEN resident_age BETIEN 60 AND 64 THEN "60_64"

            WHEN resident_age BETIEN 65 AND 69 THEN "65_69"

            WHEN resident_age BETIEN 70 AND 74 THEN "70_74"

            WHEN resident_age BETIEN 75 AND 79 THEN "75_79"

            WHEN resident_age >= 80 THEN "80+"

            END) AS AGE_GROUP, resident_sex,

            COUNT(resident_id) AS RESIDENTS,

            SUM(CASE WHEN tt_old=1 THEN 1 ELSE 0 END) AS tt_old,

            SUM(CASE WHEN tt_true=1 THEN 1 ELSE 0 END) AS tt_true

            FROM clean

            WHERE resident_examined = 1 OR (resident_examined = 2 AND
resident_left_eye_upper_tt is not NULL )AND resident_age >= 15

            GROUP BY cluster_eu, cluster_cluster, AGE_GROUP, resident_sex

            ORDER BY cluster_eu, cluster_cluster, AGE_GROUP ASC, resident_sex DESC')

data.frame(mydb_tt)


###Calculating unadjusted and adjusted TT prevalence from exported file

ttprev_male <- as.data.frame(sqldf('SELECT mydb_tt.cluster_eu, mydb_tt.cluster_cluster,
mydb_tt.AGE_GROUP, mydb_tt.resident_sex, mydb_tt.RESIDENTS, mydb_tt.tt_old,
mydb_tt.tt_true, population.percent_age_male AS age_Iight

                        FROM mydb_tt

                        LEFT JOIN population

                        ON mydb_tt.AGE_GROUP = population.age_group

                        WHERE mydb_tt.resident_sex = "1"'))

ttprev_female <- as.data.frame(sqldf('SELECT mydb_tt.cluster_eu, mydb_tt.cluster_cluster,
mydb_tt.AGE_GROUP, mydb_tt.resident_sex, mydb_tt.RESIDENTS, mydb_tt.tt_old,
mydb_tt.tt_true, population.percent_age_female AS age_Iight

                         FROM mydb_tt
```

```
                        LEFT JOIN population

                        ON mydb_tt.AGE_GROUP = population.age_group

                        WHERE mydb_tt.resident_sex = "2"'))

ttprev <- as.data.frame(rbind(ttprev_male, ttprev_female)) # Append the dataframes


# This creates the new column named "prev_unadj" filled with zeros & then does the unadjusted prevalence calculation for that group

ttprev["prev_unadj"] <- 0;

ttprev$prev_unadj <- (ttprev$tt_true / ttprev$RESIDENTS)


#Now collapse on cluster level and get sum of Iighted prevalence aggregated by age group

ttprev_age_group <- aggregate(prev_unadj ~ AGE_GROUP + cluster_eu, data = ttprev, sum)

data.frame(ttprev_age_group)
```

TT Cluster Prevalence *modified from GTMP code written by R. Willis and B. Chu

```
clean <- as.data.frame(read.csv("TT dataset.csv"))

population<- read.csv("population dataset.csv")


clean["tt_old"] <- 0

clean$tt_old[clean$resident_left_eye_upper_tt=="1"] <- 1

clean$tt_old[clean$resident_left_eye_loIr_tt=="1"] <- 1

clean$tt_old[clean$resident_right_eye_upper_tt=="1"] <- 1

clean$tt_old[clean$resident_right_eye_loIr_tt=="1"] <- 1


clean["tt_true"] <- 0

clean$tt_true[clean$resident_left_eye_upper_tt=="1" & clean$resident_scarring_left_eye=="1"] <- 1

clean$tt_true[clean$resident_left_eye_loIr_tt=="1" & clean$resident_scarring_left_eye=="1"] <- 1

clean$tt_true[clean$resident_right_eye_upper_tt=="1" & clean$resident_scarring_right_eye=="1"] <- 1
```

```
clean$tt_true[clean$resident_right_eye_loIr_tt=="1" & clean$resident_scarring_right_eye=="1"] <-
1

clean$tt_true[clean$resident_left_eye_upper_tt=="1" & clean$resident_scarring_left_eye=="2"] <-
1

clean$tt_true[clean$resident_left_eye_loIr_tt=="1" & clean$resident_scarring_left_eye=="2"] <- 1

clean$tt_true[clean$resident_right_eye_upper_tt=="1" & clean$resident_scarring_right_eye=="2"]
<- 1

clean$tt_true[clean$resident_right_eye_loIr_tt=="1" & clean$resident_scarring_right_eye=="2"] <-
1



###Source census-specific code to generate age groups for the analysis

library(sqldf)


mydb_tt <- sqldf('SELECT cluster_eu, cluster_cluster, (CASE

            WHEN resident_age BETIEN 15 AND 19 THEN "15_19"

            WHEN resident_age BETIEN 20 AND 24 THEN "20_24"

            WHEN resident_age BETIEN 25 AND 29 THEN "25_29"

            WHEN resident_age BETIEN 30 AND 34 THEN "30_34"

            WHEN resident_age BETIEN 35 AND 39 THEN "35_39"

            WHEN resident_age BETIEN 40 AND 44 THEN "40_44"

            WHEN resident_age BETIEN 45 AND 49 THEN "45_49"

            WHEN resident_age BETIEN 50 AND 54 THEN "50_54"

            WHEN resident_age BETIEN 55 AND 59 THEN "55_59"

            WHEN resident_age BETIEN 60 AND 64 THEN "60_64"

            WHEN resident_age BETIEN 65 AND 69 THEN "65_69"

            WHEN resident_age BETIEN 70 AND 74 THEN "70_74"

            WHEN resident_age BETIEN 75 AND 79 THEN "75_79"

            WHEN resident_age >= 80 THEN "80+"

            END) AS AGE_GROUP, resident_sex,

            COUNT(resident_id) AS RESIDENTS,

            SUM(CASE WHEN tt_old=1 THEN 1 ELSE 0 END) AS tt_old,

            SUM(CASE WHEN tt_true=1 THEN 1 ELSE 0 END) AS tt_true

            FROM clean
```

```
        WHERE resident_examined = 1 OR (resident_examined = 2 AND
resident_left_eye_upper_tt is not NULL )AND resident_age >= 15

        GROUP BY cluster_eu, cluster_cluster, AGE_GROUP, resident_sex

        ORDER BY cluster_eu, cluster_cluster, AGE_GROUP ASC, resident_sex DESC')

data.frame(mydb_tt)

###Calculating unadjusted and adjusted TT prevalence from exported file


ttprev_male <- as.data.frame(sqldf('SELECT mydb_tt.cluster_eu, mydb_tt.cluster_cluster,
mydb_tt.AGE_GROUP, mydb_tt.resident_sex, mydb_tt.RESIDENTS, mydb_tt.tt_old,
mydb_tt.tt_true, population.percent_age_male AS age_Iight

                FROM mydb_tt

                LEFT JOIN population

                ON mydb_tt.AGE_GROUP = population.age_group

                WHERE mydb_tt.resident_sex = "1"'))


ttprev_female <- as.data.frame(sqldf('SELECT mydb_tt.cluster_eu, mydb_tt.cluster_cluster,
mydb_tt.AGE_GROUP, mydb_tt.resident_sex, mydb_tt.RESIDENTS, mydb_tt.tt_old,
mydb_tt.tt_true, population.percent_age_female AS age_Iight

                 FROM mydb_tt

                 LEFT JOIN population

                 ON mydb_tt.AGE_GROUP = population.age_group

                 WHERE mydb_tt.resident_sex = "2"'))


ttprev <- as.data.frame(rbind(ttprev_male, ttprev_female)) # Append the dataframes


# This creates the new column named "prev_unadj" filled with zeros & then does the unadjusted
prevalence calculation for that group

ttprev["prev_unadj"] <- 0;

ttprev$prev_unadj <- (ttprev$tt_true / ttprev$RESIDENTS)


#Generate unadjusted EU-level prevalence for comparison purposes

ttprev_unadjusted_cluster <- aggregate(cbind(RESIDENTS, tt_true) ~ cluster_eu+cluster_cluster,
data=ttprev, sum)

ttprev_unadjusted_cluster$cluster_prev_unadj <-
(ttprev_unadjusted_cluster$tt_true/ttprev_unadjusted_cluster$RESIDENTS)
```

```
ttprev_unadjusted <- aggregate(cluster_prev_unadj ~ cluster_eu, data=ttprev_unadjusted_cluster, mean)

colnames(ttprev_unadjusted)[colnames(ttprev_unadjusted)=="cluster_prev_unadj"] <- "ttprev_unadj"



# This creates the new column named "adj_tt" filled with zeros & then does the Iighting

ttprev["adj_tt"] <- 0;

ttprev$adj_tt <- (ttprev$prev_unadj * ttprev$age_Iight)



#Now collapse on cluster level and get sum of Iighted prevalence

ttprev_cluster <- aggregate(adj_tt ~ cluster_eu + cluster_cluster, data = ttprev, sum)

data.frame(ttprev_cluster)



#Scatterplot

plot(ttprev_cluster$cluster_cluster, ttprev_cluster$adj_tt, xlab = "cluster ID", ylab = "TT prevalence")
```

TT EU level prevalence *modified from GTMP code written by R. Willis and B. Chu

```
clean <- as.data.frame(read.csv("TT dataset.csv"))

population<- read.csv("population dataset.csv")



clean["tt_old"] <- 0

clean$tt_old[clean$resident_left_eye_upper_tt=="1"] <- 1

clean$tt_old[clean$resident_left_eye_loIr_tt=="1"] <- 1

clean$tt_old[clean$resident_right_eye_upper_tt=="1"] <- 1

clean$tt_old[clean$resident_right_eye_loIr_tt=="1"] <- 1



clean["tt_true"] <- 0

clean$tt_true[clean$resident_left_eye_upper_tt=="1" & clean$resident_scarring_left_eye=="1"] <- 1

clean$tt_true[clean$resident_left_eye_loIr_tt=="1" & clean$resident_scarring_left_eye=="1"] <- 1

clean$tt_true[clean$resident_right_eye_upper_tt=="1" & clean$resident_scarring_right_eye=="1"] <- 1
```

```
clean$tt_true[clean$resident_right_eye_loIr_tt=="1" & clean$resident_scarring_right_eye=="1"] <-
1

clean$tt_true[clean$resident_left_eye_upper_tt=="1" & clean$resident_scarring_left_eye=="2"] <-
1

clean$tt_true[clean$resident_left_eye_loIr_tt=="1" & clean$resident_scarring_left_eye=="2"] <- 1

clean$tt_true[clean$resident_right_eye_upper_tt=="1" & clean$resident_scarring_right_eye=="2"]
<- 1

clean$tt_true[clean$resident_right_eye_loIr_tt=="1" & clean$resident_scarring_right_eye=="2"] <-
1

###Source census-specific code to generate age groups for the analysis

library(sqldf)


mydb_tt <- sqldf('SELECT cluster_eu, cluster_cluster, (CASE

        WHEN resident_age BETIEN 15 AND 19 THEN "15_19"

        WHEN resident_age BETIEN 20 AND 24 THEN "20_24"

        WHEN resident_age BETIEN 25 AND 29 THEN "25_29"

        WHEN resident_age BETIEN 30 AND 34 THEN "30_34"

        WHEN resident_age BETIEN 35 AND 39 THEN "35_39"

        WHEN resident_age BETIEN 40 AND 44 THEN "40_44"

        WHEN resident_age BETIEN 45 AND 49 THEN "45_49"

        WHEN resident_age BETIEN 50 AND 54 THEN "50_54"

        WHEN resident_age BETIEN 55 AND 59 THEN "55_59"

        WHEN resident_age BETIEN 60 AND 64 THEN "60_64"

        WHEN resident_age BETIEN 65 AND 69 THEN "65_69"

        WHEN resident_age BETIEN 70 AND 74 THEN "70_74"

        WHEN resident_age BETIEN 75 AND 79 THEN "75_79"

        WHEN resident_age >= 80 THEN "80+"

        END) AS AGE_GROUP, resident_sex,

        COUNT(resident_id) AS RESIDENTS,

        SUM(CASE WHEN tt_old=1 THEN 1 ELSE 0 END) AS tt_old,

        SUM(CASE WHEN tt_true=1 THEN 1 ELSE 0 END) AS tt_true

        FROM clean

        WHERE resident_examined = 1 OR (resident_examined = 2 AND
resident_left_eye_upper_tt is not NULL )AND resident_age >= 15
```

```
            GROUP BY cluster_eu, cluster_cluster, AGE_GROUP, resident_sex

            ORDER BY cluster_eu, cluster_cluster, AGE_GROUP ASC, resident_sex DESC')

data.frame(mydb_tt)

###Calculating unadjusted and adjusted TT prevalence from exported file


ttprev_male <- as.data.frame(sqldf('SELECT mydb_tt.cluster_eu, mydb_tt.cluster_cluster,
mydb_tt.AGE_GROUP, mydb_tt.resident_sex, mydb_tt.RESIDENTS, mydb_tt.tt_old,
mydb_tt.tt_true, population.percent_age_male AS age_Iight

                        FROM mydb_tt

                        LEFT JOIN population

                        ON mydb_tt.AGE_GROUP = population.age_group

                        WHERE mydb_tt.resident_sex = "1"'))


ttprev_female <- as.data.frame(sqldf('SELECT mydb_tt.cluster_eu, mydb_tt.cluster_cluster,
mydb_tt.AGE_GROUP, mydb_tt.resident_sex, mydb_tt.RESIDENTS, mydb_tt.tt_old,
mydb_tt.tt_true, population.percent_age_female AS age_Iight

                         FROM mydb_tt

                         LEFT JOIN population

                         ON mydb_tt.AGE_GROUP = population.age_group

                         WHERE mydb_tt.resident_sex = "2"'))


ttprev <- as.data.frame(rbind(ttprev_male, ttprev_female)) # Append the dataframes


# This creates the new column named "prev_unadj" filled with zeros & then does the unadjusted
prevalence calculation for that group

ttprev["prev_unadj"] <- 0;

ttprev$prev_unadj <- (ttprev$tt_true / ttprev$RESIDENTS)


#Generate unadjusted EU-level prevalence for comparison purposes

ttprev_unadjusted_cluster <- aggregate(cbind(RESIDENTS, tt_true) ~ cluster_eu+cluster_cluster,
data=ttprev, sum)

ttprev_unadjusted_cluster$cluster_prev_unadj <-
(ttprev_unadjusted_cluster$tt_true/ttprev_unadjusted_cluster$RESIDENTS)

ttprev_unadjusted <- aggregate(cluster_prev_unadj ~ cluster_eu, data=ttprev_unadjusted_cluster,
mean)
```

```
colnames(ttprev_unadjusted)[colnames(ttprev_unadjusted)=="cluster_prev_unadj"] <-
"ttprev_unadj"


# This creates the new column named "adj_tt" filled with zeros & then does the Iighting

ttprev["adj_tt"] <- 0;

ttprev$adj_tt <- (ttprev$prev_unadj * ttprev$age_Iight)


#Now collapse on cluster level and get sum of Iighted prevalence for histrograms comparing the 15
and 40 age groups turn on appropriate cluster name

ttprev_cluster <- aggregate(adj_tt ~ cluster_eu + cluster_cluster, data = ttprev, sum)

ttprev_cluster15 <- aggregate(adj_tt ~ cluster_eu + cluster_cluster, data = ttprev, sum)

#ttprev_cluster40 <- aggregate(adj_tt ~ cluster_eu + cluster_cluster, data = ttprev, sum)


#Mean of adjusted cluster prevalences to get EU prevalence

ttprev_eu <- aggregate(adj_tt ~ cluster_eu, data = ttprev_cluster, mean)

data.frame(ttprev_eu)
```

TT Bootstrap comparing 60 and 30 clusters *modified from GTMP code written by R. Willis and B. Chu

```
#Load, transform data

dataset <- ttprev_cluster

str(dataset)

dataset$cluster_eu <- as.factor(dataset$cluster_eu)

dataset$cluster_cluster <- as.factor(dataset$cluster_cluster)

dataset$cluster_prev <- dataset$adj_tt


#Boot statistic function = mean of XX random clusters

clustermean <- function(df, i) {

  num_clusters <- 60

  r <- round(runif(num_clusters, 1, nrow(df))) #nrow(df) allows the analysis to divide by the correct
# clusters

  df2 <- numeric()
```

```
  for (i in 1:num_clusters) {
    df2[i] <- df[r[i],]$cluster_prev
  }
  return(mean(df2))
}

#create empty data frame for results
bootResult_tt60 <- data.frame(cluster_eu=character(), bootmean=numeric(), se=numeric(),
ci95_low=numeric(), ci95_high=numeric(), stringsAsFactors=FALSE)

#Bootstrap function, looped over each EU
library(boot)
num_reps <- 10000
for (i in 1:nlevels(dataset$cluster_eu)) {
  data2 <- subset(dataset, cluster_eu==levels(cluster_eu)[i])
  b <- boot(data2, clustermean, num_reps)
  m <- mean(b$t)
  se <- sd(b$t)

  #calculate 2.5/97.5 percentiles as Confidence Interval
  q <- quantile(b$t, c(0.025, 0.975))
  ci_loIr60 <- q[1]
  ci_upper60 <- q[2]

   #write result to data frame
  eu_temp <- as.character(data2$cluster_eu[1])
  bootResult_tt60[i,] <- c(eu_temp, m, se, ci_loIr, ci_upper)

}

#Boot statistic function = mean of XX random clusters
clustermean <- function(df, i) {
```

```
  num_clusters <- 30

  r <- round(runif(num_clusters, 1, nrow(df))) #nrow(df) allows the analysis to divide by the correct
# clusters

  df2 <- numeric()

  for (i in 1:num_clusters) {

    df2[i] <- df[r[i],]$cluster_prev

  }

  return(mean(df2))

}



#create empty data frame for results

bootResult_tt30 <- data.frame(cluster_eu=character(), bootmean=numeric(), se=numeric(),
ci95_low=numeric(), ci95_high=numeric(), stringsAsFactors=FALSE)



#Bootstrap function, looped over each EU

library(boot)

num_reps <- 10000

for (i in 1:nlevels(dataset$cluster_eu)) {

  data2 <- subset(dataset, cluster_eu==levels(cluster_eu)[i])

  b2 <- boot(data2, clustermean, num_reps)

  m <- mean(b2$t)

  se <- sd(b2$t)



  #calculate 2.5/97.5 percentiles as Confidence Interval

  q <- quantile(b$t, c(0.025, 0.975))

  ci_loIr30 <- q[1]

  ci_upper30 <- q[2]



  #write result to data frame

  eu_temp2 <- as.character(data2$cluster_eu[1])

  bootResult_tt30[i,] <- c(eu_temp2, m, se, ci_loIr, ci_upper)
```

```
}
data.frame(bootResult_tt30)


#histogram of mean bootstrap results with CI
hist(b$t,xlab = "TT prevalence", breaks=50, main="Budaka 40+",col = 'blue')
hist(b2$t, breaks=50,main="overlap",border = 'red',add=T)
abline(v=ci_loIr60, lty="solid", col="blue" )
abline(v=ci_upper60, lty="solid", col="blue" )
abline(v=ci_loIr30, lty="dashed", col="red" )
abline(v=ci_upper30, lty="dashed", col="red" )
legend("topright",c("60 clusters", "30 clusters"),fill=c("blue","red"))
box()
```

2 stage simulation

```
################40pls simulation###################################


require(sampling)
###pull dataset into R
population<- read.csv("population dataset.csv") ##update!!
df <- read.csv ("TT dataset 40pls.csv")  ## update with “true” dataset from EU prevalence code!!



###create a function
library(data.table)
simulate <- function(tt_prev) {


###create cluster list
cluster <-1:60


###create cluster subset
selected_clusters <- c(sample(cluster,size = 30, replace = FALSE))


###using selected clusters create subset of households
cluster_subset <- subset(df,df$cluster_cluster %in% selected_clusters)
cluster_subset <- cluster_subset[order(cluster_subset$cluster_cluster,cluster_subset$household_id),]


###create list of unique household_ids from the selected clusters
list<- subset(cluster_subset,select = c(cluster_cluster,household_id))
list<- unique(list)
order(list$cluster_cluster,list$household_id)
###randomly select the households within the selected clusters
library(plyr)
count<-
household_subset <-ddply(list,.(cluster_cluster),function(x) x[sample(nrow(x),20),])
```

```
###subset from the original dataset using the selected households and clusters

clean <- subset(df,df$household_id %in% household_subset$household_id)


###calculate prevalence


###update dataset with 5 year age brackets

library(sqldf)

mydb_tt <- sqldf('SELECT cluster_eu, cluster_cluster, (CASE

            WHEN resident_age BETIEN 40 AND 44 THEN "40_44"

            WHEN resident_age BETIEN 45 AND 49 THEN "45_49"

            WHEN resident_age BETIEN 50 AND 54 THEN "50_54"

            WHEN resident_age BETIEN 55 AND 59 THEN "55_59"

            WHEN resident_age BETIEN 60 AND 64 THEN "60_64"

            WHEN resident_age BETIEN 65 AND 69 THEN "65_69"

            WHEN resident_age BETIEN 70 AND 74 THEN "70_74"

            WHEN resident_age BETIEN 75 AND 79 THEN "75_79"

            WHEN resident_age >= 80 THEN "80+"

            END) AS AGE_GROUP, resident_sex,

            COUNT(resident_id) AS RESIDENTS,

            SUM(CASE WHEN tt_old=1 THEN 1 ELSE 0 END) AS tt_old,

            SUM(CASE WHEN tt_true=1 THEN 1 ELSE 0 END) AS tt_true

            FROM clean

            WHERE resident_examined = 1 OR (resident_examined = 2 AND
resident_left_eye_upper_tt is not NULL )AND resident_age >= 15

            GROUP BY cluster_eu, cluster_cluster, AGE_GROUP, resident_sex

            ORDER BY cluster_eu, cluster_cluster, AGE_GROUP ASC, resident_sex DESC')


###Calculating unadjusted and adjusted TT prevalence from clean dataset

ttprev_male <- as.data.frame(sqldf('SELECT mydb_tt.cluster_eu, mydb_tt.cluster_cluster,
mydb_tt.AGE_GROUP, mydb_tt.resident_sex, mydb_tt.RESIDENTS, mydb_tt.tt_old,
mydb_tt.tt_true, population.percent_age_male AS age_Iight

                        FROM mydb_tt
```

```
                        LEFT JOIN population

                        ON mydb_tt.AGE_GROUP = population.age_group

                        WHERE mydb_tt.resident_sex = "1'"))


ttprev_female <- as.data.frame(sqldf('SELECT mydb_tt.cluster_eu, mydb_tt.cluster_cluster,
mydb_tt.AGE_GROUP, mydb_tt.resident_sex, mydb_tt.RESIDENTS, mydb_tt.tt_old,
mydb_tt.tt_true, population.percent_age_female AS age_Iight

                         FROM mydb_tt

                         LEFT JOIN population

                         ON mydb_tt.AGE_GROUP = population.age_group

                         WHERE mydb_tt.resident_sex = "2'"))


ttprev <- as.data.frame(rbind(ttprev_male, ttprev_female)) # Append the dataframes




### This creates the new column named "prev_unadj" filled with zeros & then does the unadjusted
prevalence calculation for that group

ttprev["prev_unadj"] <- 0;

ttprev$prev_unadj <- (ttprev$tt_true / ttprev$RESIDENTS)


###Generate unadjusted EU-level prevalence for comparison purposes

#ttprev_unadjusted_cluster <- aggregate(cbind(RESIDENTS, tt_true) ~ cluster_eu+cluster_cluster,
data=ttprev, sum)

#ttprev_unadjusted_cluster$cluster_prev_unadj <-
(ttprev_unadjusted_cluster$tt_true/ttprev_unadjusted_cluster$RESIDENTS)

#ttprev_unadjusted <- aggregate(cluster_prev_unadj ~ cluster_eu, data=ttprev_unadjusted_cluster,
mean)

#colnames(ttprev_unadjusted)[colnames(ttprev_unadjusted)=="cluster_prev_unadj"] <-
"ttprev_unadj"


### This creates the new column named "adj_tt" filled with zeros & then does the Iighting

ttprev["adj_tt"] <- 0;

ttprev$adj_tt <- (ttprev$prev_unadj * ttprev$age_Iight)
```

```
###Now collapse on cluster level and get sum of Iighted prevalence for histrograms comparing the 15 and 40 age groups turn on appropriate cluster name

ttprev_cluster <- aggregate(adj_tt ~ cluster_eu + cluster_cluster, data = ttprev, sum)

#ttprev_cluster15 <- aggregate(adj_tt ~ cluster_eu + cluster_cluster, data = ttprev, sum)

#ttprev_cluster40 <- aggregate(adj_tt ~ cluster_eu + cluster_cluster, data = ttprev, sum)


###Mean of adjusted cluster prevalences to get EU prevalence

ttprev_eu <- aggregate(adj_tt ~ cluster_eu, data = ttprev_cluster, mean)

#data.frame(ttprev_eu)


}


r<- do.call(rbind,replicate(10000,simulate(),simplify = FALSE))


m40 <- mean(r$adj_tt)

se40 <- sd(r$adj_tt)


#calculate 2.5/97.5 percentiles as Confidence Interval

q <- quantile(r$adj_tt, c(0.025, 0.975))

ci_loIr40 <- q[1]

ci_upper40 <- q[2]



#write result to data frame

eu_temp2 <- c(m40,se40,ci_loIr40,ci_upper40)

data.frame(eu_temp2)


#histogram of meresults with CI

hist(r$adj_tt,xlab = "TT prevalence", breaks=50, main="'Monduli 40+",col = 'blue')

box()


###################################15pls#####################################
```

```
require(sampling)

###pull dataset into R

population<- read.csv("population dataset.csv") ##update!!

df <- read.csv ("TT dataset 15pls.csv")  ## update with “true” dataset from EU prevalence code!!


###create a function

library(data.table)

simulate <- function(tt_prev) {



  ###create cluster list

  cluster <-1:60


  ###create cluster subset

  selected_clusters <- c(sample(cluster,size = 30, replace = FALSE))


  ###using selected clusters create subset of households

  cluster_subset <- subset(df,df$cluster_cluster %in% selected_clusters)

  cluster_subset <-
cluster_subset[order(cluster_subset$cluster_cluster,cluster_subset$household_id),]


  ###create list of unique household_ids from the selected clusters

  list<- subset(cluster_subset,select = c(cluster_cluster,household_id))

  list<- unique(list)

  order(list$cluster_cluster,list$household_id)



  ###randomly select the households within the selected clusters

  library(plyr)

  household_subset <-ddply(list,.(cluster_cluster),function(x) x[sample(nrow(x),20),])


  ###subset from the original dataset using the selected households and clusters
```

```
clean <- subset(df,df$household_id %in% household_subset$household_id)


###calculate prevalence


###update dataset with 5 year age brackets
library(sqldf)
mydb_tt <- sqldf('SELECT cluster_eu, cluster_cluster, (CASE
            WHEN resident_age BETIEN 15 AND 19 THEN "15_19"
              WHEN resident_age BETIEN 20 AND 24 THEN "20_24"
              WHEN resident_age BETIEN 25 AND 29 THEN "25_29"
              WHEN resident_age BETIEN 30 AND 34 THEN "30_34"
              WHEN resident_age BETIEN 35 AND 39 THEN "35_39"
              WHEN resident_age BETIEN 40 AND 44 THEN "40_44"
              WHEN resident_age BETIEN 45 AND 49 THEN "45_49"
              WHEN resident_age BETIEN 50 AND 54 THEN "50_54"
              WHEN resident_age BETIEN 55 AND 59 THEN "55_59"
              WHEN resident_age BETIEN 60 AND 64 THEN "60_64"
              WHEN resident_age BETIEN 65 AND 69 THEN "65_69"
              WHEN resident_age BETIEN 70 AND 74 THEN "70_74"
              WHEN resident_age BETIEN 75 AND 79 THEN "75_79"
              WHEN resident_age >= 80 THEN "80+"
              END) AS AGE_GROUP, resident_sex,
              COUNT(resident_id) AS RESIDENTS,
              SUM(CASE WHEN tt_old=1 THEN 1 ELSE 0 END) AS tt_old,
              SUM(CASE WHEN tt_true=1 THEN 1 ELSE 0 END) AS tt_true
              FROM clean
              WHERE resident_examined = 1 OR (resident_examined = 2 AND
resident_left_eye_upper_tt is not NULL )AND resident_age >= 15
              GROUP BY cluster_eu, cluster_cluster, AGE_GROUP, resident_sex
              ORDER BY cluster_eu, cluster_cluster, AGE_GROUP ASC, resident_sex DESC')


###Calculating unadjusted and adjusted TT prevalence from clean dataset
```

```
  ttprev_male <- as.data.frame(sqldf('SELECT mydb_tt.cluster_eu, mydb_tt.cluster_cluster,
mydb_tt.AGE_GROUP, mydb_tt.resident_sex, mydb_tt.RESIDENTS, mydb_tt.tt_old,
mydb_tt.tt_true, population.percent_age_male AS age_Iight

                              FROM mydb_tt

                              LEFT JOIN population

                              ON mydb_tt.AGE_GROUP = population.age_group

                              WHERE mydb_tt.resident_sex = "1"'))


  ttprev_female <- as.data.frame(sqldf('SELECT mydb_tt.cluster_eu, mydb_tt.cluster_cluster,
mydb_tt.AGE_GROUP, mydb_tt.resident_sex, mydb_tt.RESIDENTS, mydb_tt.tt_old,
mydb_tt.tt_true, population.percent_age_female AS age_Iight

                               FROM mydb_tt

                               LEFT JOIN population

                               ON mydb_tt.AGE_GROUP = population.age_group

                               WHERE mydb_tt.resident_sex = "2"'))


  ttprev <- as.data.frame(rbind(ttprev_male, ttprev_female)) # Append the dataframes




  ### This creates the new column named "prev_unadj" filled with zeros & then does the unadjusted
prevalence calculation for that group

  ttprev["prev_unadj"] <- 0;

  ttprev$prev_unadj <- (ttprev$tt_true / ttprev$RESIDENTS)


  ###Generate unadjusted EU-level prevalence for comparison purposes

  #ttprev_unadjusted_cluster <- aggregate(cbind(RESIDENTS, tt_true) ~ cluster_eu+cluster_cluster,
data=ttprev, sum)

  #ttprev_unadjusted_cluster$cluster_prev_unadj <-
(ttprev_unadjusted_cluster$tt_true/ttprev_unadjusted_cluster$RESIDENTS)

  #ttprev_unadjusted <- aggregate(cluster_prev_unadj ~ cluster_eu, data=ttprev_unadjusted_cluster,
mean)

  #colnames(ttprev_unadjusted)[colnames(ttprev_unadjusted)=="cluster_prev_unadj"] <-
"ttprev_unadj"


  ### This creates the new column named "adj_tt" filled with zeros & then does the Iighting
```

```
  ttprev["adj_tt"] <- 0;

  ttprev$adj_tt <- (ttprev$prev_unadj * ttprev$age_Iight)


  ###Now collapse on cluster level and get sum of Iighted prevalence for histrograms comparing the
15 and 40 age groups turn on appropriate cluster name

  ttprev_cluster <- aggregate(adj_tt ~ cluster_eu + cluster_cluster, data = ttprev, sum)

  #ttprev_cluster15 <- aggregate(adj_tt ~ cluster_eu + cluster_cluster, data = ttprev, sum)

  #ttprev_cluster40 <- aggregate(adj_tt ~ cluster_eu + cluster_cluster, data = ttprev, sum)



  ###Mean of adjusted cluster prevalences to get EU prevalence

  ttprev_eu <- aggregate(adj_tt ~ cluster_eu, data = ttprev_cluster, mean)

  #data.frame(ttprev_eu)


}


r2<- do.call(rbind,replicate(10000,simulate(),simplify = FALSE))



m <- mean(r2$adj_tt)

se <- sd(r2$adj_tt)


#calculate 2.5/97.5 percentiles as Confidence Interval

q <- quantile(r2$adj_tt, c(0.025, 0.975))

ci_loIr <- q[1]

ci_upper <- q[2]



#write result to data frame

eu_temp <- c(m,se,ci_loIr,ci_upper)

data.frame(eu_temp)
```

```
#histogram of meresults with CI

hist(r2$adj_tt,xlab = "TT prevalence", breaks=50, main="Monduli 40+",col = 'blue')

box()


#histogram of mean bootstrap results with CI

hist(r2$adj_tt, xlab="TT prevalence",breaks=50, xlim = c(0.001,0.05), ylim = c(0,900),
main="Monduli 2 stage comparing results from 15+ and 40+ with 30 clusters",col = 'blue')

hist(r$adj_tt, breaks=50,main="overlap",border = 'red',add=T)

legend("topright",c("15 and older", "40 and older"),fill=c("blue","red"))

box()
```

Appendix F GTMP WASH indicators

Responses are classified as improved are marked ■.

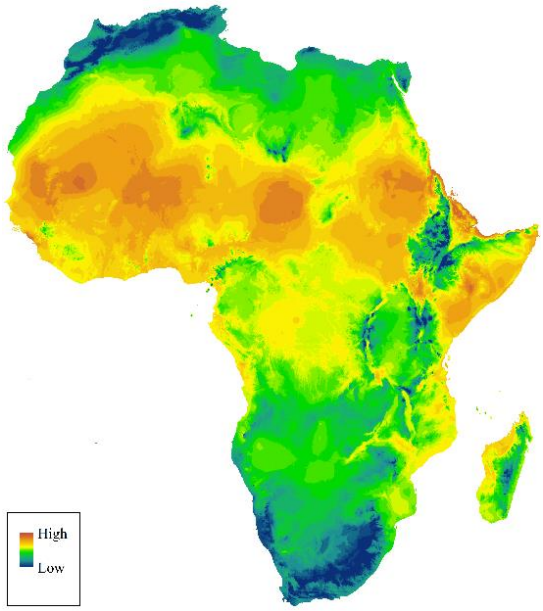
S1) Where do you and other adults in the household usually defecate? number of households reporting that adults living in the household defecate in either a shared public latrine or a private latrine / number of households enrolled in the survey	1) Shared or public latrine	■
	2) Private latrine	■
	3) No structure, outside near the house	
	4) No structure, in the bush or field	
	9) Other	
S2) Improved latrine: What kind of toilet facility do the adults in the household use? Observed. number of households where an improved latrine is observed / number of households enrolled in the survey	1) Flush/pour flush to piped sewer system	■
	2) Flush/pour flush to septic tank	■
	3) Flush/pour flush to pit latrine	■
	4) Flush/pour flush to open drains	
	5) Flush/pour flush to unknown place	■
	6) Ventilated improved pit latrine (VIP)	■
	7) Pit latrine with slab	■
	8) Pit latrine without slab/open pit	
	9) Composting toilet	■
	10) Bucket	
	11) Hanging toilet/hanging latrine	
	12) No facilities or bush or field	
	99) Other	
H3) Improved Water source: In the dry season, what is the main source of water used by your household for washing faces? number of households reporting to access improved water sources / number of households enrolled in the survey	1) Piped water into dwelling	■
	2) Piped water into yard/pot	■
	3) Public tap/standpipe	■
	4) Tubewell/borehole	■
	5) Protected dug well	■
	6) Unprotected dug well	
	7) Protected spring	■
	8) Unprotected spring	
	9) Rainwater collection	■
	10) Water vendor	
	11) Surface water (e.g. river, dam, lake, canal)	
	99) Other	

H4) Wash near: Washing water: If you collected water there to bring back to the house, how long does it take to go there, get water, and come back? number of households reporting access to improved water sources and those water sources are on the premises / number of households enrolled in the survey	0) All face washing done at the water source	
	1) Water source in the yard	
	2) Less than 30 minutes	
	3) Between 30 minutes and 1 hour	
	4) More than 1 hour	

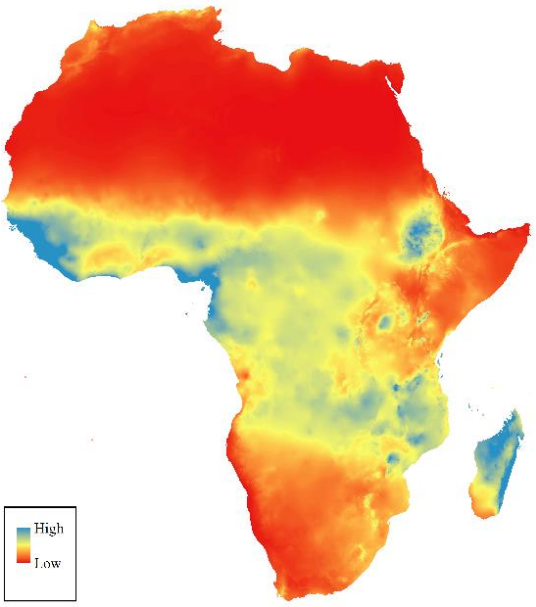
Appendix G Raster images of environmental risk factors

5 kilometre gridded climate raster maps

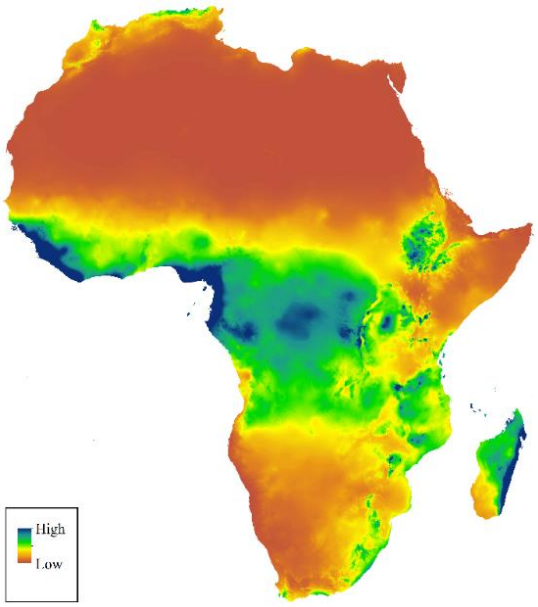
Annual mean temperature



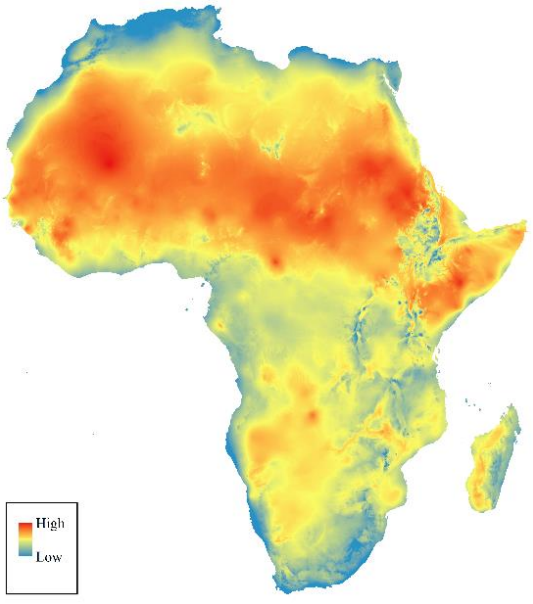
Annual mean precipitation



Aridity index

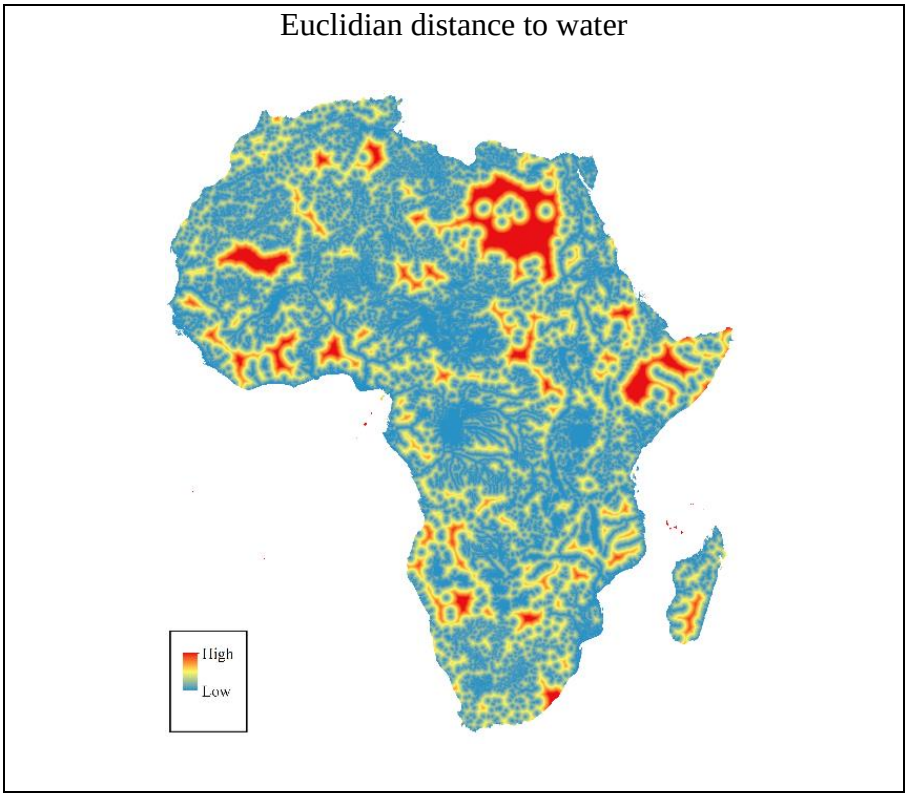


PET



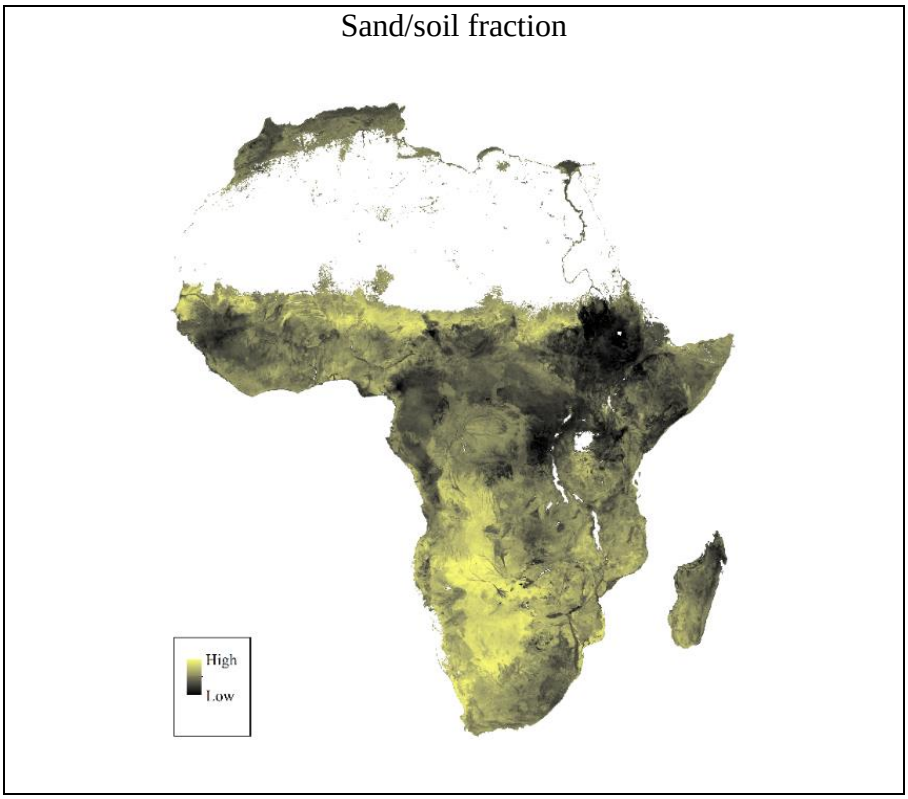
5 kilometre gridded Euclidean distance to water raster map

Euclidian distance to water

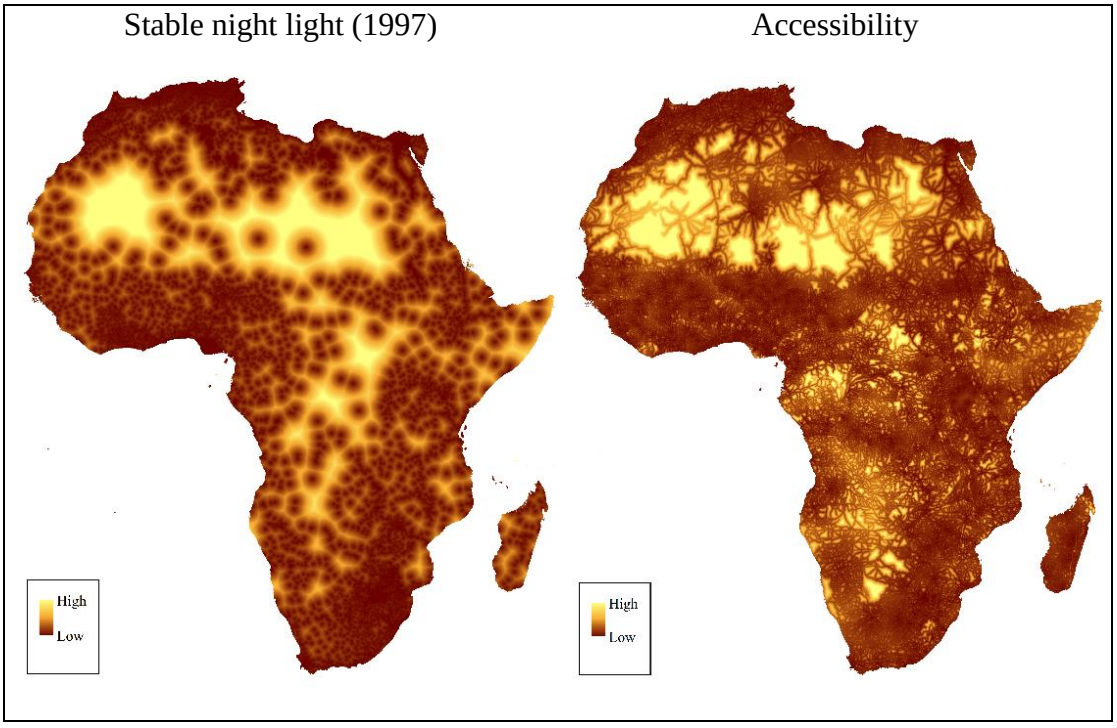


5 kilometre gridded soil composition sand fraction raster map

Sand/soil fraction



5 kilometre gridded remoteness raster map



Appendix H R script TT spatial models

```
rm(list=ls())
library(PrevMap)
library(rgdal)
library(lme4)
library(geoR)


df <- as.data.frame(read.csv("cluster_full.csv")) #cluster level data set

#subset by country
data<- subset(df, Country == "BEN") #change this depending on what country you are interested in

##cluster level random effect
#To measure effect size run the model with cluster level random effect first with no covariates, then with TF, then with all

REnull <- glmer(cbind(tt_pos,tt_examined-tt_pos) ~ (1 | cluster_id), family = binomial("logit"), data = data, control=glmerControl(optimizer="bobyqa"))
summary(REnull)

RETF <- glmer(cbind(tt_pos,tt_examined-tt_pos) ~ tf_prev + (1 | cluster_id), family = binomial("logit"), data = data, control=glmerControl(optimizer="bobyqa"))
summary(RETF)

REall <- glmer(cbind(tt_pos,tt_examined-tt_pos) ~ tf_prev + S1 + S2 + imp_un_water + water_yard  +  ai + sand_5cm_s + stlight97_s + (1 | cluster_id), family = binomial("logit"), data = data, control=glmerControl(optimizer="bobyqa"))
summary(REall)

#only include significant indicators identified in the REall model
REsig <- glmer(cbind(tt_pos,tt_examined-tt_pos) ~ tf_prev + sand_5cm_s + (1 | cluster_id), family = binomial("logit"), data = data, control=glmerControl(optimizer="bobyqa"))
summary(REsig)

##extract the residuals of random effect

#conversion of the coordinates from LONG-LAT into Web Mercator
#the distance between locations is now measured in kilometers
coords.ll <- SpatialPoints(data[,c("LNG","LAT")],CRS("+init=epsg:4236"))
coords.web <- SpatialPoints(coords.ll,CRS("+init=epsg:3857"))
data$web_x <- coordinates(coords.web)[,1]/1000
data$web_y <- coordinates(coords.web)[,2]

#only include REsig indicators
data2 <- data[complete.cases(data[,c("tt_pos","tf_prev", "sand_5cm_s", "cluster_id", "web_x","web_y")]),]

REsig <- glmer(cbind(tt_pos,tt_examined-tt_pos) ~ tf_prev + sand_5cm_s + (1 | cluster_id), family = binomial("logit"), data = data2, control=glmerControl(optimizer="bobyqa"))
summary(REsig)

#this line extracts the random effects from the mixed model
rand.eff <- ranef(REsig)$cluster_id[,1]

#calculate the maximum distance between locations
points <- subset(data2, select = c(web_x, web_y))
```

```
dist <- dist(points[,-1])
maxd <- max(dist)

coords <- as.matrix(data2[, c("web_x", "web_y")])

vari <- variog(coords=data2[,c("web_x","web_y")],data=rand.eff, uvec = seq(0,maxd/2,
length=15))

vari.fit <- variofit(vari, ini.cov.pars = c(0.5,1.0), cov.model = "matern", fix.nugget = FALSE,
nugget = 0 , fix.kappa = TRUE, kappa = 0.5)

vari.sim <- variog.mc.env(obj.variog = vari, cords = data2 [,c("web_x","web_y")], data =
rand.eff,nsim = 1000)

matplot(vari$u,cbind(vari.sim$v.upper, vari$v, vari.sim$v.lower), type="l", col = c(2,1,2),lty =
c("dashed", "dashed", "dashed"), xlab="distance",ylab="semivariance")

#matplot(lines(vari.fit), type = "l", add=TRUE, lwd=2)

##geostatistical modelling
#now estimate the geospatial model using Monte Carlo maximum likelihood (use the best model
from above)

# 1 is the variance of the residual spatial correlation
# 10 is the scale of the spatial correlation in meters
par2 <- c(REsig@beta,1,10)

#This function defines the options for the MCMC algorithm used in the Monte Carlo maximum
likelihood method.
c.mcmc <- control.mcmc.MCML(n.sim=10000,burnin=2000,thin=8,h= (1.65)/(nrow(data2)^(1/6)))
str(c.mcmc)

#only include indicators from REsig model
fit.MCML3 <- binomial.logistic.MCML(tt_pos ~ tf_prev + sand_5cm_s, units.m=~tt_examined,
coords=~web_x+web_y, par0=par2,control.mcmc = c.mcmc, fixed.rel.nugget = 0,kappa=0.5,
start.cov.pars = 10,data=data2, method = "nlminb")

summary.fit <-summary(fit.MCML3)
fit.MCML3$log.lik

#scale of the spatial correlation
exp(summary.fit$cov.pars[2,1])*log(20) # km

#95% confidence interval
exp(summary.fit$cov.pars[2,1]+c(-1,1)*qnorm(0.975)*summary.fit$cov.pars[2,2])*log(20) # km
```

Appendix I Binomial mixed effect model R outputs

Benin	
Null model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 601.6 608.3 -298.8 597.6 211 Scaled residuals: Min 1Q Median 3Q Max -0.5974 -0.4871 -0.4253 0.3697 0.9761 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 2.253 1.501 Number of obs: 213, groups: cluster_id, 213 Fixed effects: Estimate Std. Error z value Pr(>|z|) (Intercept) -5.0649 0.1863 -27.19 <2e-16 ***
TF only model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 565.1 575.2 -279.6 559.1 210 Scaled residuals: Min 1Q Median 3Q Max -0.9528 -0.5326 -0.4362 0.4484 1.2856 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.296 1.138 Number of obs: 213, groups: cluster_id, 213 Fixed effects: Estimate Std. Error z value Pr(>|z|) (Intercept) -5.4023 0.1896 -28.494 < 2e-16 *** tf_prev 6.0655 0.8843 6.859 6.94e-12 ***
TF prevalence + risk factors model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + S1 + S2 + imp_un_water + water_yard + ai + sand_5cm_s + stlight97_s + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 566.2 599.8 -273.1 546.2 203 Scaled residuals: Min 1Q Median 3Q Max -1.0134 -0.5338 -0.3864 0.4153 1.7015 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.176 1.084 Number of obs: 213, groups: cluster_id, 213

Cote d'Ivoire	
Null model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 198.2 205.3 -97.1 194.2 254 Scaled residuals: Min 1Q Median 3Q Max -0.10289 -0.07618 -0.07288 -0.06605 0.46770 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 13.75 3.708 Number of obs: 256, groups: cluster_id, 256 Fixed effects: Estimate Std. Error z value Pr(>|z|) (Intercept) -9.413 1.174 -8.018 1.07e-15 ***
TF only model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 198.7 209.4 -96.4 192.7 253 Scaled residuals: Min 1Q Median 3Q Max -0.15848 -0.09210 -0.08337 -0.07566 0.59333 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 10.28 3.206 Number of obs: 256, groups: cluster_id, 256 Fixed effects: Estimate Std. Error z value Pr(>|z|) (Intercept) -9.400 1.310 -7.177 7.12e-13 *** tf_prev 3.487 2.646 1.318 0.187
TF prevalence + risk factors model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + S1 + S2 + imp_un_water + water_yard + ai + sand_5cm_s + stlight97_s + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 199.6 235.1 -89.8 179.6 246 Scaled residuals: Min 1Q Median 3Q Max -0.4829 -0.1842 -0.1238 -0.1007 2.2171 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 2.953 1.718 Number of obs: 256, groups: cluster_id, 256

DRC	
Null model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 3018.5 3028.3 -1507.2 3014.5 1021 Scaled residuals: Min 1Q Median 3Q Max -0.7402 -0.5966 -0.1094 0.3914 2.0896 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.521 1.233 Number of obs: 1023, groups: cluster_id, 1023 Fixed effects: Estimate Std. Error z value Pr(>|z|) (Intercept) -4.40726 0.06655 -66.23 <2e-16 ***
TF only model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 2864.7 2879.5 -1429.4 2858.7 1020 Scaled residuals: Min 1Q Median 3Q Max -1.4608 -0.5579 -0.2798 0.4170 2.7763 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.234 1.111 Number of obs: 1023, groups: cluster_id, 1023
TF prevalence + risk factors model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + S1 + S2 + imp_un_water + water_yard + ai + sand_5cm_s + stlight97_s + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 2795.7 2845.0 -1387.8 2775.7 1013 Scaled residuals: Min 1Q Median 3Q Max -1.4488 -0.5453 -0.2834 0.4139 3.0121 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.048 1.024 Number of obs: 1023, groups: cluster_id, 1023

Ethiopia	
Null model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 12605.3 12618.1 -6300.7 12601.3 4478 Scaled residuals: Min 1Q Median 3Q Max -0.8142 -0.6160 -0.5088 0.4562 1.7023 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.217 1.103 Number of obs: 4480, groups: cluster_id, 4480 Fixed effects: Estimate Std. Error z value Pr(>|z|) (Intercept) -4.53562 0.03073 -147.6 <2e-16 ***
TF only model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 12182.0 12201.2 -6088.0 12176.0 4477 Scaled residuals: Min 1Q Median 3Q Max -1.0466 -0.5887 -0.4671 0.4728 2.5096 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.006 1.003 Number of obs: 4480, groups: cluster_id, 4480
TF prevalence + risk factors model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + S1 + S2 + imp_un_water + water_yard + ai + sand_5cm_s + stlight97_s + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 12029.9 12094.0 -6004.9 12009.9 4470 Scaled residuals: Min 1Q Median 3Q Max -1.1047 -0.5939 -0.4334 0.4923 2.6740 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 0.919 0.9587 Number of obs: 4480, groups: cluster_id, 4480

Guinea	
Null model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 341.2 348.6 -168.6 337.2 293 Scaled residuals: Min 1Q Median 3Q Max -0.4189 -0.2980 -0.2843 -0.2674 1.4111 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.876 1.37 Number of obs: 295, groups: cluster_id, 295 Fixed effects: Estimate Std. Error z value Pr(>|z|) (Intercept) -6.5984 0.2888 -22.84 <2e-16 ***
TF only model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 342.3 353.4 -168.2 336.3 292 Scaled residuals: Min 1Q Median 3Q Max -0.3994 -0.2972 -0.2825 -0.2619 1.4218 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.865 1.366 Number of obs: 295, groups: cluster_id, 295
TF prevalence + risk factors model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + S1 + S2 + imp_un_water + water_yard + ai + sand_5cm_s + stlight97_s + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 349.3 386.2 -164.7 329.3 285 Scaled residuals: Min 1Q Median 3Q Max -0.4981 -0.3145 -0.2774 -0.2230 1.9768 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.664 1.29 Number of obs: 295, groups: cluster_id, 295

Malawi	
Null model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 794.9 803.9 -395.4 790.9 694 Scaled residuals: Min 1Q Median 3Q Max -0.4681 -0.2907 -0.2684 -0.2300 1.4442 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 2.17 1.473 Number of obs: 696, groups: cluster_id, 696 Fixed effects: Estimate Std. Error z value Pr(>|z|) (Intercept) -6.2583 0.2033 -30.79 <2e-16 ***
TF only model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 783.9 797.5 -388.9 777.9 693 Scaled residuals: Min 1Q Median 3Q Max -0.5216 -0.2926 -0.2615 -0.2220 1.6484 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.956 1.399 Number of obs: 696, groups: cluster_id, 696
TF prevalence + risk factors model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + S1 + S2 + imp_un_water + water_yard + ai + sand_5cm_s + stlight97_s + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 785.0 830.4 -382.5 765.0 686 Scaled residuals: Min 1Q Median 3Q Max -0.5639 -0.3152 -0.2625 -0.1940 3.4738 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.733 1.316 Number of obs: 696, groups: cluster_id, 696

Mozambique	
Null model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 1854.8 1865.9 -925.4 1850.8 1946 Scaled residuals: Min 1Q Median 3Q Max -0.2187 -0.1822 -0.1730 -0.1613 1.6067 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 4.449 2.109 Number of obs: 1948, groups: cluster_id, 1948 Fixed effects: Estimate Std. Error z value Pr(>|z|) (Intercept) -7.2466 0.2278 -31.8 <2e-16 ***
TF only model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 1821.3 1838.1 -907.7 1815.3 1945 Scaled residuals: Min 1Q Median 3Q Max -0.6170 -0.1956 -0.1771 -0.1625 2.2368 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 3.522 1.877 Number of obs: 1948, groups: cluster_id, 1948
TF prevalence + risk factors model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + S1 + S2 + imp_un_water + water_yard + ai + sand_5cm_s + stlight97_s + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 1810.5 1866.2 -895.2 1790.5 1938 Scaled residuals: Min 1Q Median 3Q Max -0.6033 -0.2180 -0.1814 -0.1462 2.3079 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 3.084 1.756 Number of obs: 1948, groups: cluster_id, 1948

Nigeria	
Null model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 13437.7 13450.9 -6716.8 13433.7 5362 Scaled residuals: Min 1Q Median 3Q Max -0.6629 -0.4916 -0.4384 0.4427 1.5356 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.929 1.389 Number of obs: 5364, groups: cluster_id, 5364 Fixed effects: Estimate Std. Error z value Pr(>|z|) (Intercept) -5.13003 0.03822 -134.2 <2e-16 ***
TF only model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 13312.0 13331.8 -6653.0 13306.0 5361 Scaled residuals: Min 1Q Median 3Q Max -0.9981 -0.4917 -0.4343 0.4481 1.4304 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.821 1.35 Number of obs: 5364, groups: cluster_id, 5364
TF prevalence + risk factors model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + S1 + S2 + imp_un_water + water_yard + ai + sand_5cm_s + stlight97_s + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 11898.1 11963.9 -5939.0 11878.1 5354 Scaled residuals: Min 1Q Median 3Q Max -1.1294 -0.4989 -0.2747 0.3235 4.6856 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 0.998 0.999 Number of obs: 5364, groups: cluster_id, 5364

Sudan	
Null model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 1264.6 1273.6 -630.3 1260.6 665 Scaled residuals: Min 1Q Median 3Q Max -0.5020 -0.4079 -0.3798 0.6002 1.0650 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.975 1.405 Number of obs: 667, groups: cluster_id, 667 Fixed effects: Estimate Std. Error z value Pr(>|z|) (Intercept) -5.436 0.134 -40.57 <2e-16 ***
TF only model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 1244.6 1258.1 -619.3 1238.6 664 Scaled residuals: Min 1Q Median 3Q Max -0.7947 -0.4005 -0.3658 0.5623 1.1083 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.845 1.358 Number of obs: 667, groups: cluster_id, 667
TF prevalence + risk factors model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + S1 + S2 + imp_un_water + water_yard + ai + sand_5cm_s + stlight97_s + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 1158.5 1203.5 -569.3 1138.5 657 Scaled residuals: Min 1Q Median 3Q Max -0.8231 -0.4473 -0.2898 0.2510 3.0878 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 1.18 1.086 Number of obs: 667, groups: cluster_id, 667

Uganda	
Null model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 117.9 123.3 -56.9 113.9 107 Scaled residuals: Min 1Q Median 3Q Max -0.5515 -0.4583 -0.4264 -0.3920 4.0309 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 0 0 Number of obs: 109, groups: cluster_id, 109 Fixed effects: Estimate Std. Error z value Pr(>|z|) (Intercept) -5.7650 0.2186 -26.38 <2e-16 ***
TF only model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 119.1 127.1 -56.5 113.1 106 Scaled residuals: Min 1Q Median 3Q Max -0.5732 -0.4563 -0.4187 -0.3738 3.6379 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 0 0 Number of obs: 109, groups: cluster_id, 109
TF prevalence + risk factors model	Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] Family: binomial (logit) Formula: cbind(tt_pos, tt_examined - tt_pos) ~ tf_prev + S1 + S2 + imp_un_water + water_yard + ai + sand_5cm_s + stlight97_s + (1 | cluster_id) Data: data Control: glmerControl(optimizer = "bobyqa") AIC BIC logLik deviance df.resid 119.9 146.8 -49.9 99.9 99 Scaled residuals: Min 1Q Median 3Q Max -0.8797 -0.4156 -0.3128 -0.2073 3.5966 Random effects: Groups Name Variance Std.Dev. cluster_id (Intercept) 0 0 Number of obs: 109, groups: cluster_id, 109