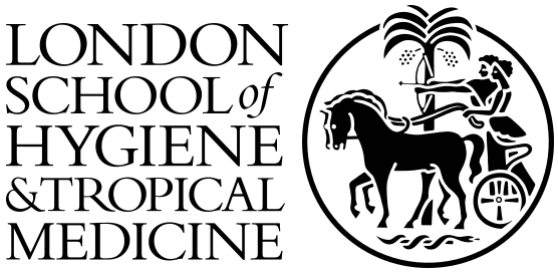


Heterogeneity in cluster randomised trials of azithromycin mass drug administration for trachoma control

Tansy Edwards

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy

of the
University of London

August 2021

Department of Clinical Research

Faculty of Infectious Tropical Diseases

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Self-funded with partial salary support from the MRC International Statistics and Epidemiology Group Programme Grant (MRC and FCDO (formerly DFID); grant MR/ K012126/1)

Research group affiliation:
MRC International Statistics and Epidemiology Group

Declaration

I, Tansy Edwards, 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.

Signature: Date: 10 August 2021

Abstract

Background

A cluster randomized trial (CRT) provides an ideal framework to evaluate mass drug administration (MDA) intervention strategies for control of trachoma, a leading infectious cause of preventable blindness. Heterogeneity is a crucial consideration in the design and analysis of CRTs, as it can occur as clustering of the outcome and clustering of not receiving treatment (non-participation during MDA).

Methods

Data from a CRT of MDA interventions for trachoma control in The Gambia (NCT00792922) were used to investigate clustering of, and risk factors for non-participation. Simulation studies investigated implications of non-participation occurring independently of baseline infection status (homogeneously, analogous to MDA coverage<100%) and amongst those with infection at baseline (heterogeneously), on power to detect pre-specified effect sizes in intention-to-treat (ITT) analysis in CRTs. An ITT analysis evaluated *population-level effectiveness* of azithromycin MDA on a secondary outcome of the CRT in The Gambia of all-age all-cause mortality. A pragmatic bootstrapping approach, simultaneously adjusting for clustering of both mortality and non-participation to minimise bias, was used to estimate a complier average causal effect (CACE), as an indication of *efficacy* in those who receive treatment.

Results

Non-participation clustered repeatedly in the same households over three annual MDA rounds.

Increased numbers of clusters in CRTs are required to allow for non-participation amongst individuals infected at baseline. The ITT rate ratio for two annual MDA rounds versus no MDA on all-age all-cause mortality was 1.11 (95% CI: 0.85-1.44). The CACE rate ratio was 1.36 (0.86-2.79).

Conclusions

CRT design should include consideration of likely prevalence of non-participation amongst infected individuals, rather than relying solely on expected coverage, during mass treatment rounds. Results were inconclusive about whether large-scale azithromycin MDA could yield important reductions in mortality in The Gambia. Estimating efficacy from CRTs of MDA interventions without bias remains a challenge.

Acknowledgements

I am indebted to my supervisors, Professor Robin Bailey and Professor Elizabeth Allen, for their intellectual insight and their patience.

I am very grateful for permission to use data from the Partnership of Rapid Elimination of Trachoma trial conducted in The Gambia, the opportunities to participate in fieldwork, and for the support of several academic colleagues involved in the PRET trial and the development of the work in this thesis; Professor David Mabey, Dr Emma Harding-Esch, Professor Neal Alexander and Professor Peter Smith.

I am so thankful for the support of my family, Mum, Steve, Dad, Jimmy, Ria and Lucien.

The members of my office also deserve a mention for their unending encouragement; Dave, Jackie, Andrea and Daniel. I would especially like to thank Jaya for her infectious enthusiasm and camaraderie.

Contents

List of Figures

List of Tables

List of Abbreviations

ANCOVA	Analysis of covariance
AT	As treated analysis
CACE	Complier average causal effect
CI	Confidence interval
CRT	Cluster randomised trial
EA	Enumeration area
EBA	Eligible for treatment but absent during MDA
EBU	Eligible for treatment but status unknown
GET2020	Global Elimination of Trachoma by 2020
GPS	Global positioning system
HH	Household
ICC	Intra-cluster correlation
ITT	Intention-to-treat
LRT	Likelihood ratio test
LSHTM	London School of Hygiene & Tropical Medicine
MAR	Missing at random
MCAR	Missing completely at random
MDA	Mass drug administration
NECP	National Eye Care Program, in The Gambia
NTDs	Neglected tropical diseases
PNT	Present but not treated during MDA
PP	Per-protocol
PRET	Partnership for Rapid Elimination of Trachoma

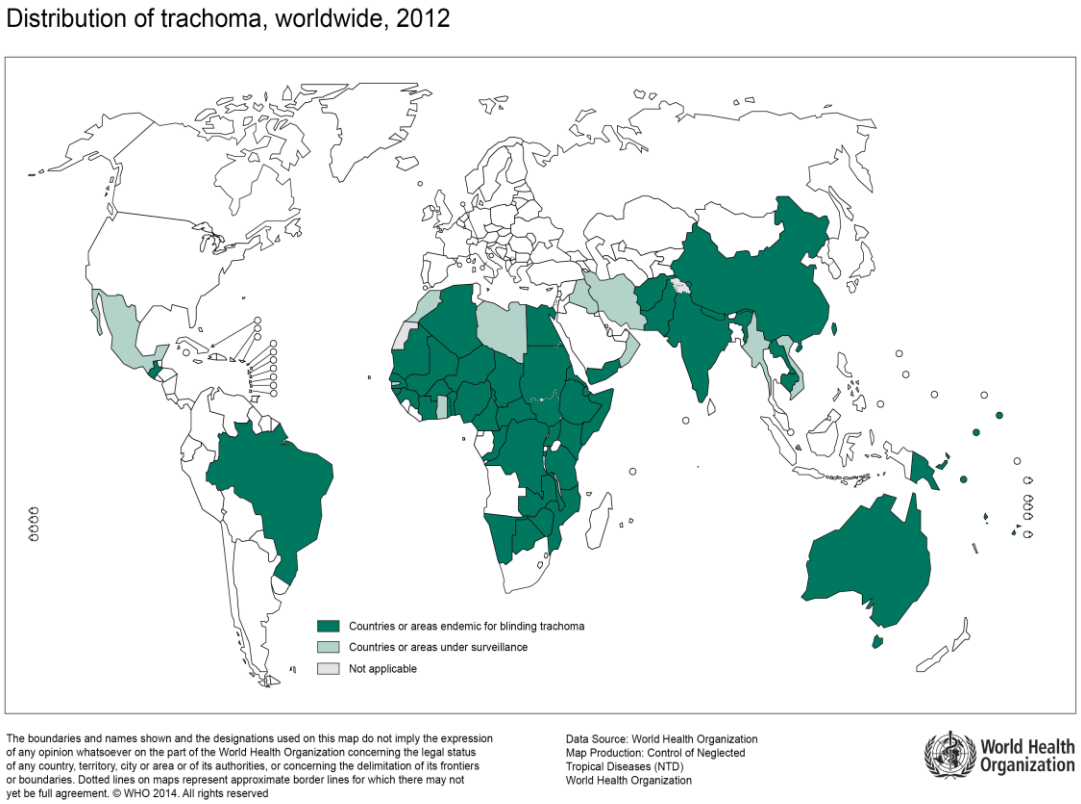
RR	Rate ratio
SAFE	**S**urgery for in-turned eyelashes, **A**ntibiotics for treatment of C. *trachomatis* infection, **F**acial cleanliness and **E**nvironmental improvement for trachoma control
SET	Settlement
TANA	Trachoma Amelioration in Northern Amhara
TF	Trachomatous inflammation, follicular
TI	Trachomatous inflammation, intense
TS	Trachomatous conjunctival scarring
TT	Trachomatous trichiasis
WHO	World health organisation

1. Introduction

1.1. Natural History of Trachoma

Trachoma is the leading infectious cause of blindness worldwide[1-3]. Trachoma is endemic in poorer and more remote parts of 51 countries in Africa, Asia, Central and South America, Australia and the Middle East, with the highest burden in Africa (Figure 1). The World Health Organisation (WHO) report that more than 230 million people live in endemic areas and could be at risk of trachoma[3].

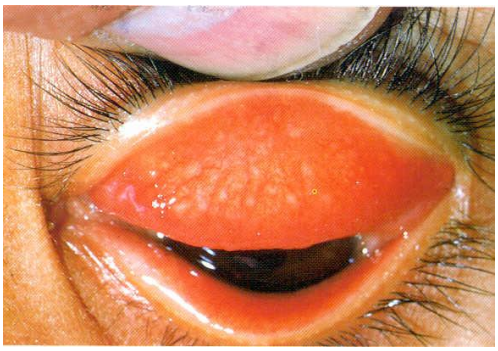
Figure 1. Worldwide Distribution of Trachoma



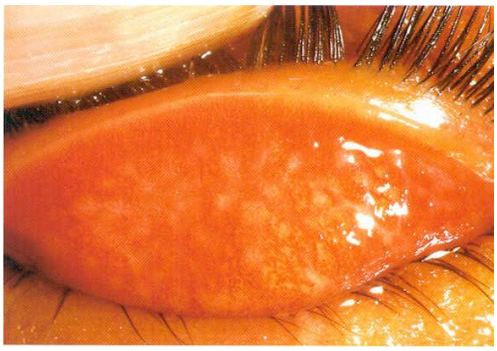
Source: World Health Organisation. http://www.who.int/trachoma/epidemiology/en/. Accessed: 14 June 2018.

Trachoma is caused by the bacterium *Chlamydia trachomatis*. Evidence of chlamydial infection is identified via laboratory testing of ocular swabs taken from the upper tarsal conjunctiva (inside of the upper eyelid)[4]. Clinical signs of disease are identified by visual inspection of the upper tarsal conjunctiva (Figure 2). Two grades are used for clinically active trachoma; **TF**: trachomatous inflammation, follicular and **TI**: trachomatous inflammation, intense[5]. Clinically active trachoma (**AT**) is defined as the presence of TF and, or TI.

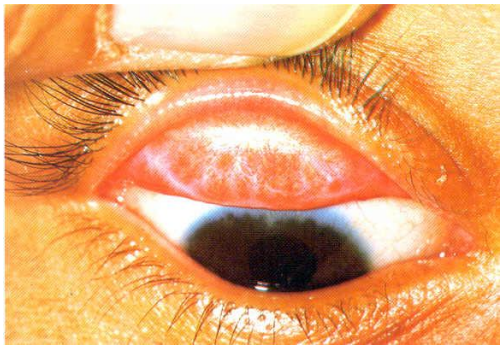
Figure 2. Clinical signs of trachoma



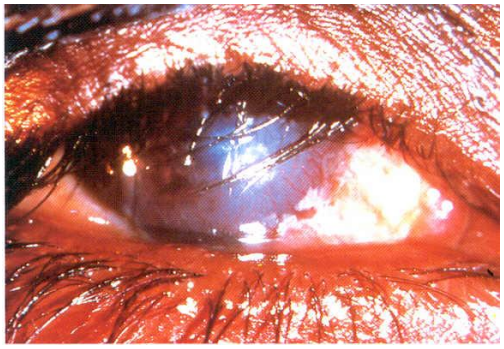
TF: trachomatous inflammation, follicular is the presence of five or more follicles of at least 0·5 mm diameter in the central part of the upper tarsal conjunctiva



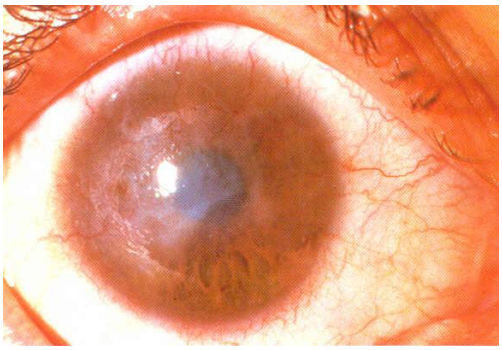
TI: trachomatous inflammation, intense is pronounced inflammatory thickening of the upper tarsal conjunctiva obscuring more than half the normal deep tarsal vessels



TS: trachomatous conjunctival scarring is the presence of easily visible scars in the tarsal conjunctiva caused by repeated inflammation



TT: trachomatous trichiasis as at least one eyelash rubbing on the eyeball, or evidence of recent removal of in-turned eyelashes



CO: corneal opacity is easily visible corneal opacity over the pupil, so dense that at least part of the pupil margin is blurred when viewed through the opacity

Source: World Health Organisation. Department of Neglected Tropical Diseases. Simplified grading card. SAFE documents[6]*. Accessed 14 June 2018.*

Repeated ocular bacterial infections with *Chlamydia trachomatis* infection in childhood cause progression between clinical stages of disease (Figure 2). Persistent TF and TI lead to trachomatous conjunctival scarring (**TS**) which eventually causes the eyelashes to turn inwards so that they can scratch the eyeball. This stage of disease is trachomatous trichiasis (**TT**). TT causes damage to the cornea that can eventually lead to corneal opacity, vision loss and

irreversible blindness. Even when active trachoma has disappeared from younger individuals in endemic communities, TT can persist as public health problem for decades[2].

It should be acknowledged that TF is an imperfect indicator of *C. trachomatis* infection; TF can be present when infection is not and vice-versa[7]. In a recent meta-analysis[8], there was strong positive correlation between TF and *C. trachomatis* infection prior to MDA (Pearson's correlation coefficient r = 0.92, 95% CI 0.83 to 0.96, p<0.001) but weaker post MDA (r = 0.60, 95% CI 0.25 to 0.81, p=0.003). Correlation between TI and infection was lower pre-MDA than for TF and not present post-MDA. TF remains the field measurement of choice in the absence of a suitably inexpensive and reliable diagnostic test that can be used in the field.

A disease of poverty[9], trachoma is spread through close social contact and overcrowding, facilitated by poor sanitation and poor access to clean water[1,4], as these factors enable infectious secretions caused by trachoma to be transmitted between children with unclean faces or interact in close contact. Trachoma clusters within communities and households where hygiene practices are also likely to be clustered. Women bear more of the burden of active trachoma and trichiasis than men, thought to be as a result of increased exposure through caring for children[2].

Loss of vision, occurring secondary to trichiasis, typically occurs in adulthood. In some of the most severely affected trachoma endemic areas, trichiasis is found in children and teenagers[3,10,11]. Vision loss related disability incurs stigma and economic hardship for affected individuals, families and communities.

1.2. Trachoma Control and Elimination

The World Health Organisation (WHO) endorsed control strategy for trachoma is SAFE (**S**urgery for in-turned eyelashes, **A**ntibiotics for treatment of C. *trachomatis* infection, **F**acial cleanliness and **E**nvironmental improvement)[12]. The WHO alliance for the Global Elimination of Trachoma by 2020 (GET2020) was launched in 1996 to support implementation of the SAFE strategy in affected countries[13].

The simplified grading system for trachoma (Figure 2) provides a field tool for rapid assessment of the prevalence of trachoma. Confirmation of the prevalence of *C. trachomatis* infection requires a more complex and costly process; eye swabs need to be kept cold in the field and during transportation to a laboratory for analysis. Laboratory resources required are not always available in affected countries and if not available, swabs must be shipped internationally. Rapid

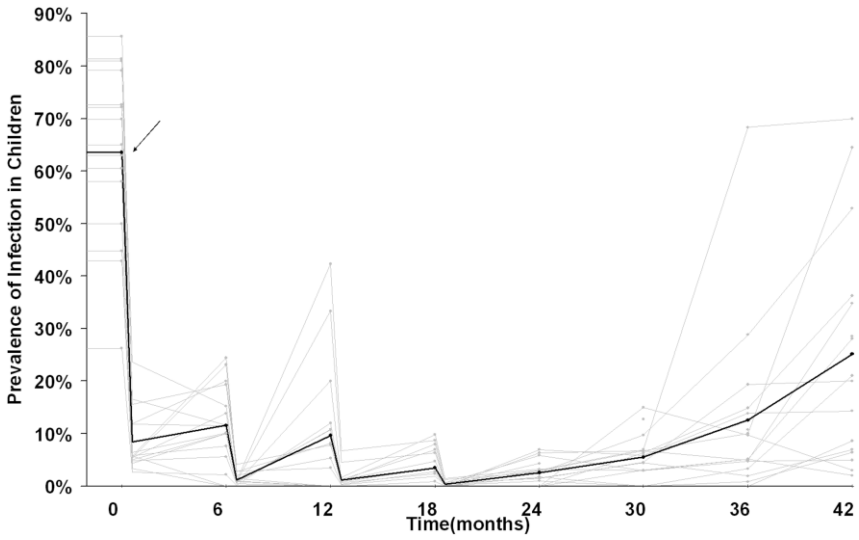
assessments of prevalence based on the simplified grading tool enables more timely decision making for implementation of interventions.

Targets for elimination of trachoma as a public health problem are defined for trichiasis and active trachoma separately, as less than 1 case of trichiasis per 1000 total population in all districts of an endemic country, and a reduction in the prevalence of TF in children aged 1–9 years to less than 5%, again in all districts within a country[14].

The mainstay of trachoma control is treatment with the broad-spectrum antibiotic azithromycin in order to clear *C. trachomatis* infection in individuals[1]. Azithromycin is delivered via community-wide, or mass drug administration (MDA), typically to all community members aged ≥1 year with the exception of pregnant women[15]. MDA is considered to be a cost-effective intervention when azithromycin is donated and more cost-effective than a screen and treat type approach[15,16]. MDA is a convenient way to reach a large number of infected individuals and can also offer additional benefit through *indirect* effects (for example herd effects in untreated individuals in treated communities), thereby reducing risk of further transmission[17,18]. Although local elimination is possible with MDA[19-22], *C. trachomatis* infection can re-emerge after MDA is discontinued [23-26], especially in high endemicity settings. A simple example of a return of infection is illustrated in **Figure 3**, showing the prevalence of infection in 16 villages in Ethiopia over a period of 42 months after one round of MDA at baseline. The mean of village-level prevalence falls between baseline and 18 months follow-up and then starts to steadily rise again.

Current recommendations for MDA are based on district-level prevalence estimates of TF in children aged 1-9 years, where a district is considered as an administrative population unit of approximately 100,000-250,000 people[27]. In districts with prevalence of at least 10%, WHO recommends treatment of entire communities with annual MDA for at least 3 consecutive years, then reassessment of prevalence[27,28], advocating coverage targets of at least 80%. If TF prevalence is 5-9.9%, one year of MDA plus F&E is recommended followed by impact assessments. MDA is not recommended if the prevalence is less than 5%. Facial cleanliness and environmental improvements are recommended in all scenarios where trachoma persists.

Figure 3. Reduction of *C. trachomatis* infection after MDA and return post-MDA in Ethiopia



Grey lines: prevalence of infection in children aged 1-5 years old, in each of 16 villages
Black line: mean of the village-level prevalence of infection over time
Black arrow: timing of MDA
Source: Figure 2 from Lakew T, House J, Hong KC, Yi E, Alemayehu W, Melese M, et al. (2009) Reduction and Return of Infectious Trachoma in Severely Affected Communities in Ethiopia. PLoS Negl Trop Dis 3(2): e376.

1.3. Mass Drug Administration Strategies

Optimal delivery strategies for MDA are needed for different prevalence settings to make the most efficient use of resources available through identification of who best to treat and how often[29,30]. Once prevalence is closer to zero and moving on a trajectory towards achieving global elimination of trachoma, it is of interest to know how MDA can be used to prevent a return to previous levels that sustained transmission and importantly, when MDA can be discontinued.

1.3.1. Frequency

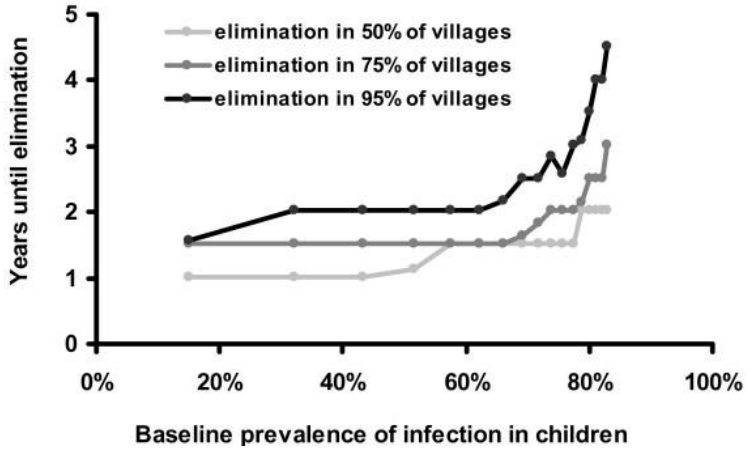
In high prevalence settings (hyper-endemic, prevalence of TF>30%), three years of annual treatment to all community members, as currently recommended, may be insufficient to meet elimination targets. Options for increased frequency include more than one MDA round per year, either to all community members or MDA focussed on children who are at greater risk of infection.

Observational data from 71 hyper-endemic communities in Tanzania suggested more than seven years of annual mass treatment may be required to eliminate infection in the setting of that study[31]. Modelling studies predicted up to five years of biannual treatment of all community members could be required for elimination in higher endemicity settings[32,33] (**Figure 4**).

In a CRT in hyper-endemic areas of Ethiopia, four mass treatment rounds distributed biannually (twice per year), to all individuals, led to a lower prevalence of infection in children aged 1-5 years old at 24 months follow-up, compared to two mass treatment rounds distributed annually to all individuals[20], but in another trial with four annual and seven biannual MDA rounds with treatment offered to all, the prevalence of infection in children aged 0-9 years in each arm at 42 months follow-up was similar[34]. Beyond 24 months, the mean of village-level prevalence of infection remained low (<5%) in both arms up to 42 months follow-up, in the CRT with longer treatment duration (**Figure 5.**).

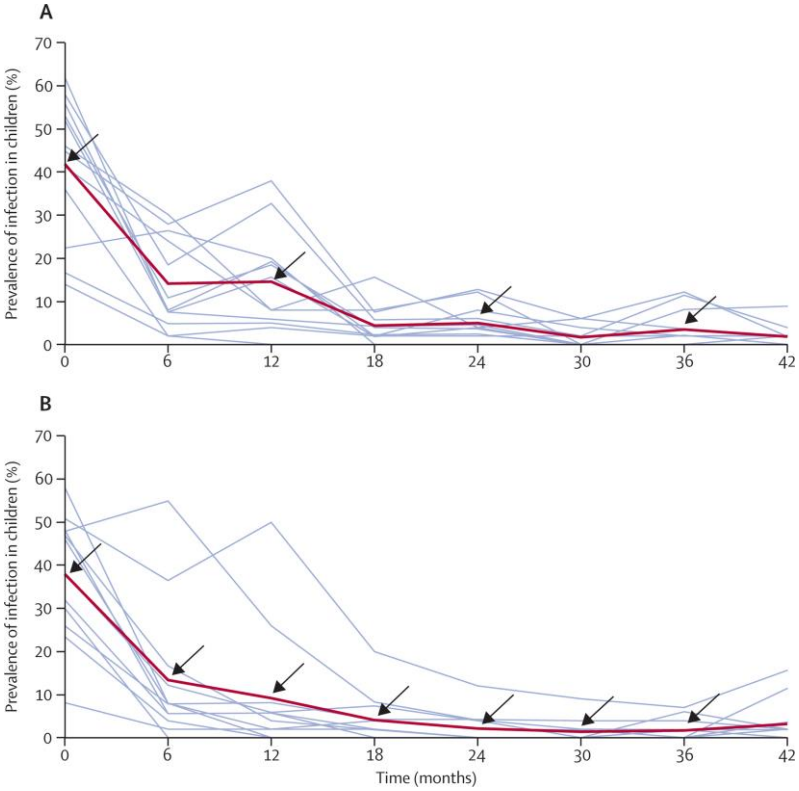
In a CRT comparing six biannual MDA rounds with treatment only offered to those aged 0-12 years old, to three annual MDA rounds with treatment offered to all[35], the prevalence of *C. trachomatis* infection in children aged 0-5 years old in the arm with biannual treatment of children was non-inferior to prevalence in the arm with annual treatment of everyone, at 36 months follow-up (**Figure 6**). Non-inferiority was also seen for the prevalence of infection in adults. The authors of this study suggest that biannual MDA to children only, could require fewer antibiotics and that it may be logistically simpler to only treat children during MDA rounds.

Figure 4. Predicted time to elimination with biannual treatment of community members aged ≥1 year old



*Source: Figure 6 of Ray KJ, Porco TC, Hong KC, Lee DC, Alemayehu W, Melese M, Lakew T, Yi E, House J, Chidambaram JD, Whitcher JP, Gaynor BD, Lietman TM. (2007). A rationale for continuing mass antibiotic distributions for trachoma. BMC Infectious Diseases **7**:91*

Figure 5. Prevalence of infection in children aged 0-9 years in Ethiopian communities with either annual or biannual MDA offered to *all* individuals



A: annual MDA, B: twice-yearly MDA

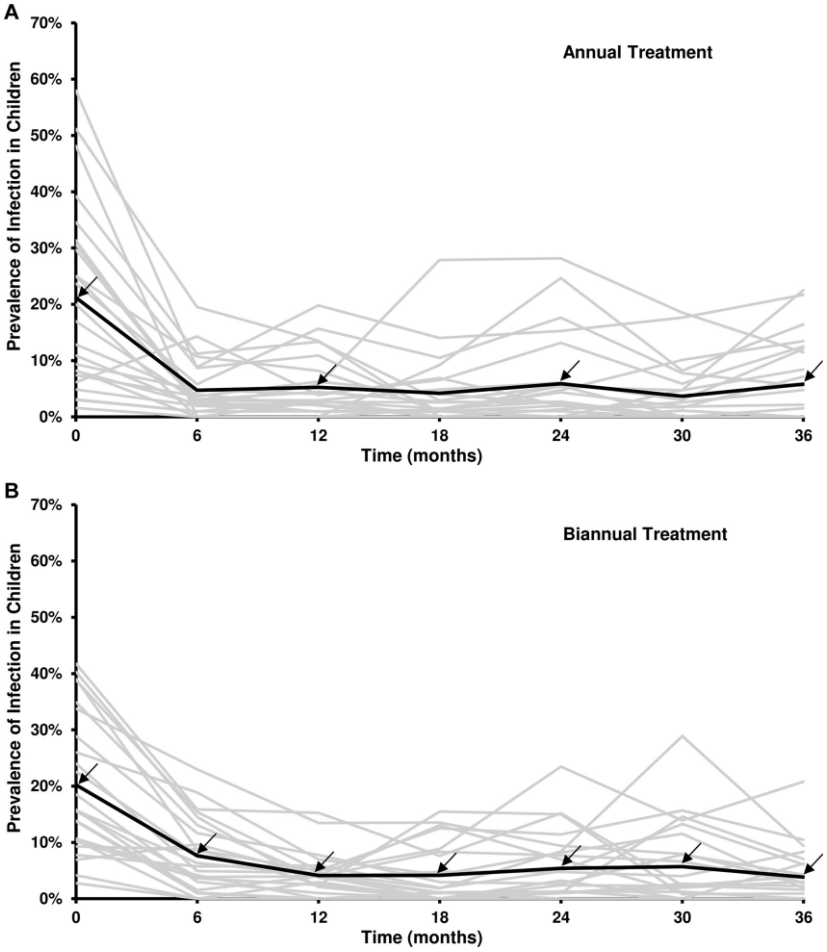
Grey lines: village-level prevalence of infection in children aged 0-9 years old

Red line: mean of village-level prevalence of infection

Black arrow: timing of MDA

Source: Figure 2 of Gebre T, Ayele B, Zerihun M, Genet A, Stoller NE, Zhou Z, House JI, Yu SN, Ray KJ, Emerson PM, Keenan JD, Porco TC, Lietman TM, Gaynor BD. (2012) Comparison of annual versus twice-yearly mass azithromycin treatment for hyperendemic trachoma in Ethiopia: a cluster-randomised trial. Lancet 379(9811): 143-51.

Figure 6. Prevalence of infection in children aged 0-5 years in communities with either *annual MDA for all* or *biannual MDA for children* aged 0-12 years in Niger



Grey lines: village-level prevalence of infection in children aged 0-5 years old
Black line: mean prevalence of infection
Black arrow: timing of MDA
Source: Figure 2 of Amza A, Kadri B, Nassirou B, Cotter SY, Stoller NE, Zhou Z, Bailey RL, Mabey DC, Porco TC, Keenan JD, Gaynor BD, West SK, Lietman TM. (2016). A Cluster-Randomized Trial to Assess the Efficacy of Targeting Trachoma Treatment to Children. Clin Infect Dis. 15;64(6):743-750.

Typically, endemicity levels for trachoma are considered as low (hypo-endemic, prevalence of TF<20%), medium (meso-endemic, prevalence of TF of 20-30%) or high (hyper-endemic, prevalence of TF≥30%[27]. In low and medium prevalence settings, less frequent MDA could provide a move towards elimination targets at a similar rate of progress to annual MDA and require fewer resources.

Modelling studies and small field studies of just one or two communities have suggested that MDA could be required less than once per year[19,22,32,36]. Simulations were used to explore what happens to prevalence of infection over time if an early MDA stopping rule (or MDA graduation rule) was in place, in contrast to three years of annual MDA provided at baseline, year one and

year two, for different prevalence settings[32]. Under the graduation rule, MDA would be discontinued early during a three-year period of annual MDA if the prevalence of infection falls below 5%, in interim surveys conducted between annual MDA rounds. It would be expected that the prevalence of infection after three years follow-up in communities that stopped receiving MDA, would be comparable to the prevalence in communities that continued MDA. Thus, an MDA graduation strategy could substantially reduce usage of antibiotics and resources for implementation.

Data from three field trials to test this hypothesis are now available[37-39]. A CRT in The Gambia where clusters were randomly allocated to an MDA graduation rule or three years of annual MDA, MDA was discontinued based on prevalence of infection six months after the baseline MDA round in the MDA stopping rule arm[39]. By the time the study started, the baseline prevalence of infection and TF in all communities was very low (0.8% and 7% respectively). As the prevalence of TF was already below the MDA treatment threshold based on TF at the time of the baseline survey, the results may not be applicable to other hypo-endemic settings with TF just above the MDA threshold.

Two CRTs with the same design were conducted in Tanzania. One CRT included communities with an overall prevalence of infection of just over 20% at baseline, ranging from <5% up to almost 45%, with an overall prevalence of TF of around 30% at baseline[38]. In the second CRT, the overall prevalence of infection was around 5% at baseline and the prevalence of TF was around 12% at baseline[37]. Neither trial provided evidence that MDA can be stopped before the recommended three-year period of annual MDA in a hypo- or meso-endemic area.

1.3.2. Coverage

Coverage is calculated as the percentage of individuals receiving azithromycin treatment during MDA relative to community census data. High coverage is considered important to prevent re-emergence of trachoma post MDA[30]. During MDA rounds, the recommended minimum target for coverage is 80%[27].

Superiority of enhanced efforts to achieve a higher coverage target of 90%, compared to standard efforts to achieve the minimum 80% target, were assessed in CRTs in three different settings. Each of the three CRTs had a 2x2 factorial design to allow for simultaneous evaluation of MDA frequency strategies. In Niger, evaluation of MDA coverage strategies was stratified by randomly allocated frequency strategy (clusters randomly allocated to six biannual MDA rounds, with the first MDA following the baseline survey, where azithromycin was given to children aged

0-12 years[40] and clusters randomly allocated to three annual MDA rounds with treatment for all [41]). Although there was some evidence that an enhanced coverage strategy reduced the prevalence of *C. trachomatis* infection in children aged 0-5 years over time, more quickly than the standard strategy in clusters where only children were treated, there was no evidence of an improved reduction in prevalence with the enhanced strategy in the final survey data[40]. In clusters with community-wide treatment of all, there was no evidence of an improved rate of decline or difference in infection at the end of the study in children aged 0-5 years old[41]. In Tanzania, where MDA was offered to all community members for three years, there was no evidence of a difference in the prevalence of *C. trachomatis* infection in children aged 0-5 years between enhanced and standard coverage strategy arms[38,39]. In Niger and Tanzania, baseline prevalence of infection was around 20% in children aged 0-5 years old. As previously mentioned, the prevalence of *C. trachomatis* infection in children aged 0-5 years in The Gambia was very low, less than 1% overall at the end of the study so a coverage comparison was not meaningful.

While results of CRTs evaluating MDA coverage strategies did not provide compelling evidence of a need to recommend enhanced coverage above the current target of 80%, coverage during MDA rounds remains an important consideration[29]. Not receiving treatment during MDA for trachoma is typically referred to as non-participation, rather than non-compliance, reflecting the choice of individuals to take treatment when it is offered to their community. Non-participation could hinder trachoma control efforts if it occurs systematically and persistently in high prevalence settings and if it occurs in enough infected individuals to allow for sustained transmission in any prevalence setting.

1.4. Impact of azithromycin MDA on TF

Field evaluations of MDA strategies described above included analyses based on an objective laboratory confirmed outcome of *C. trachomatis* infection, to obtain a clearer picture of the impact of MDA strategies on the infectious causative agent of trachoma. Evidence generated from these field trials can inform future research directions and decision making for trachoma control, with a view to breaking transmission of infection and meeting elimination targets.

In areas identified for large-scale distribution of azithromycin, decision-making is still based on prevalence of TF from rapid assessment surveys. Azithromycin MDA also leads to a reduction in the prevalence of TF but as already highlighted, TF may persist where there is no longer any *C. trachomatis* infection. Decision-making based on TF when implementing MDA strategies based on results for an outcome of infection may lead to continuation of MDA where it perhaps isn't

needed, if TF persists at a prevalence level above the threshold in the absence of infection in communities.

In addition, the 10% threshold for prevalence of TF is typically based on a point estimate for district level prevalence, calculated with some adjustment for within-district correlation as a result of population-based prevalence sampling (PBPS) of clusters and individuals[42], rather than the upper bound of a 95% CI for prevalence. So, it may be that for some districts with a borderline point estimate of prevalence just under 10%, the upper bounds of the 95% CI for prevalence could be more than 10%. Therefore, these districts could need three years of annual MDA but not be deemed eligible for it.

1.5. Rigorous evaluation of MDA interventions for trachoma control

1.5.1. Cluster randomised trials

Since MDA is delivered on a community-wide basis, a cluster randomized trial (CRT) provides the most scientifically rigorous framework to collect field data to evaluate mass drug administration (MDA) intervention strategies for trachoma control[43].

A CRT of an MDA intervention enables a *pragmatic* evaluation of the *population-level effectiveness* of MDA delivered at community level[43], based on an intention-to-treat (ITT) analysis that includes all available data regardless of whether communities or individuals participated according to the trial protocol or randomisation schedule. Any observed *population-level effectiveness* can be due to both *direct* and *indirect* effects of the intervention. A *direct* effect of azithromycin would be clearing infection in those who take it. The overall effect of azithromycin MDA in a CRT could also be influenced by any *indirect* beneficial effects of the intervention, such as a herd effect in untreated individuals in treated communities, or, influenced by potential negative effects on prevalence of the outcome, such as the community-wide delivery method not reaching the target population and or the intervention not being wanted by, or acceptable to, some community members. A herd effect in the context of MDA for trachoma, or other NTDs with MDA control, could be observed as a reduction in prevalence of infection in older children and adults in communities where say only younger children are offered treatment during an MDA round, compared to prevalence in older children and adults in communities not offered MDA[17]. In other words, a herd effect is when there is some protection afforded to untreated individuals in treated clusters.

1.5.2. Heterogeneity in the design and analysis of CRTs of MDA for trachoma control

Trachoma is known to cluster within communities[44-46] and the implementation of MDA rarely leads to uniform coverage everywhere[47,48]. Heterogeneity is a crucial consideration in the design and analysis of CRTs. Heterogeneity in a CRT of MDA interventions for trachoma can occur in a number of ways; as clustering of the outcome due to the infectious nature of trachoma and clustering of risk factors that enable transmission, clustering of not receiving treatment during MDA (non-participation) and in particular, clustering of not receiving treatment amongst those at higher risk of the outcome. These examples of clustering can occur at each level of the hierarchical population structure within a trial cluster, e.g. at household level within clusters[49,50].

The validity of design and analysis assumptions is dependent on taking appropriate account of possible sources of heterogeneity, as clustering generates additional variability in the data compared to data without clustering. Variability in the data has implications for precision of any effects estimated from the data and power for hypothesis testing. Therefore, the presence of heterogeneity influences several stages of design and analysis of a CRT.

During the design stage, sampling of clusters and individuals from a study area and randomisation of clusters requires consideration of the hierarchical population structure, in order to achieve representative samples from the study area, balance between trial arms and to limit the risk of contamination or unwanted additional variability between trial clusters. Samples size calculations need to incorporate information about potential clustering in the data in relation to the hierarchical data structure and sampling process. Ignoring additional variability in the data due to clustering will mean reduced power.

During the analysis stage, any clustering in the outcome data needs to be accounted for with appropriate statistical methodology, otherwise variability in the data will be under-estimated with the consequence that the statistical significance of findings will be over-estimated.

Clustering of non-participation could occur at cluster level as differential prevalence of non-participation between trial clusters. Within trial clusters, there could also be clusters of non-participation amongst households. It would be hoped that randomisation would lead to balance in the distribution of cluster-level non-participation between arms in a trial. If cluster-level non-participation is not associated with the prevalence of infection at cluster-level, or non-participation occurs in uninfected individuals, the influence of non-participation on results could be negligible. If non-participation occurs in infected individuals, the impact of treatment will be reduced in those trial clusters and the corresponding treatment arm. Systematic variability in

non-participation associated with a higher risk of infection, could introduce bias in any between-arm comparisons even in an ITT analysis[51]. Also, lower than expected reductions in prevalence in treated clusters could lead to smaller effect sizes than expected and a loss of power to detect pre-specified effect sizes[52]. In CRTs of MDA frequency strategies, there are multiple opportunities for non-participation to occur and for variability in the extent of non-participation and clustering of non-participation over time[53].

1.6. Critical review of published literature: CRTs of azithromycin MDA for trachoma

A systematic literature search was undertaken to identify CRTs evaluating the impact of azithromycin MDA on trachoma outcomes. Information was extracted that related to reporting and handling of heterogeneity in the data during the design and analysis stages of the CRTs.

1.6.1. Systematic search strategy

Preliminary investigations found that not all cluster randomised trials are specifically described as such in the title or abstract. Identifying a cluster unit of randomisation may only be apparent from longer descriptions in a full-text article. Therefore, PubMed (www.pubmed.org) was used to search for randomised trials for trachoma, regardless of intervention or unit of randomisation. The search term combination used was “Randomized Controlled Trials” as a Mesh topic or “Randomized Controlled Trial” as a publication topic, in combination with trachoma as a Mesh term. The combined search terms were thus; (("Trachoma"[Mesh]) AND ("Randomized Controlled Trial" [Publication Type] OR "Randomized Controlled Trials as Topic"[Mesh])).

Where a title or abstract indicated eligible outcomes and interventions, full text articles were retrieved to obtain more accurate information on randomisation and hence avoid excluding studies with a cluster unit of randomisation. Full text articles of other sub-studies or follow-up studies of eligible cluster randomised trials were obtained for complete information through the original search results or through a search of articles citing eligible CRTs or using the clinicaltrials.gov NCT identifier.

Other databases were considered but were found to sub-optimal in relation to PubMed; ClinicalTrials.gov and Current Controlled Trials (too few results); EMBASE and MEDLINE (Ovid: returned around half the number of results as PubMed); SCIRUS (limited search choices and results) and Google Scholar (not specific enough, many returned results referred to an article that only cited a randomised trial).

The search was updated on 16 June 2018 and returned 135 results. Studies were included if the outcomes included TF or *C. trachomatis* infection and interventions included mass treatment of azithromycin delivered at community level.

Fourteen studies of azithromycin MDA interventions for trachoma were identified, including two studies with a longitudinal design where clusters were randomly selected for MDA for follow-up and then control clusters were randomly selected for follow-up, rather than selection of all clusters prior to randomisation[18,23] (Table 1). CRTs were considered as distinct trials based on comparisons of interventions for trachoma to answer specific research questions. For example, the Trachoma Amelioration in Northern Amhara (TANA) trial in Ethiopia (NCT00322972) had a large framework with six randomisation arms, in order to encompass three CRTs evaluating the impact of different interventions on trachoma outcomes.

Table 1: CRTs of MDA interventions for trachoma

	Trial, lead author (year)	**Intervention arms**	**Trachoma Outcomes**	**Sample size for outcome evaluation**	**Aims**
1	ASANTE Tanzania West (2017)[54] Ervin (2016)[55]	Annual MDA if prevalence of TF in children aged 1-9 years old≥5% and prevalence of infection >1% vs Annual MDA with same criteria plus surveillance and treatment of newcomers and travellers.	Proportion of communities with a prevalence of *C. trachomatis* infection <1% at 24 months follow-up	52 clusters. 100 children aged 1-9 years.	To determine whether the proportion of communities with a prevalence of *C. trachomatis* infection <1% in children aged 1-9 years at 24 months follow-up is different between arms.
2	PRET Niger Amza (2018)[41] Oldenburg (2018)[40] Amza (2016)[35] Amza (2012)[56]	2x2 factorial design; 1: 3 x annual MDA to all ages vs biannual (twice-yearly) MDA to children aged 6 months to 12 years (6 biannual rounds). 2: Treatment coverage 80-90% vs >90% (results *not reported*)	TF and C. *trachomatis* infection in children aged 0-5 years. C. *trachomatis* infection in those aged ≥15 years after three years follow-up.	48 clusters. 100 children aged 0-5 years, 40 aged ≥15 years per cluster.	To determine whether prevalence of infection at 36 months in the biannual child treatment arm is non-inferior to the annual MDA for all strategy. To determine whether prevalence of infection at 36 months is different between coverage arms
3	PRET Ziada Tanzania Yohannen (2013)[37]	2 x annual MDA rounds to all ages vs MDA graduation rule (stop if prevalence of infection was <5% in children aged 0-5 years halfway between scheduled MDA rounds)	TF and C. *trachomatis* infection in children aged 0-5 years at 18 months follow-up.	16 clusters. 100 children aged 0-5 years.	To determine whether prevalence of infection at 36 months under the graduation rule is non-inferior to an annual MDA strategy in communities

	Trial, lead author (year)	Intervention arms	Trachoma Outcomes	Sample size for outcome evaluation	Aims
					with prevalence of TF between 10-20% in children <5 years old.
4	PRET Tanzania West (2013)[38] Harding-Esch (2010)[49] Stare (2011)[57]	2x2 factorial design; 1: 3 annual MDA rounds (all ages) vs MDA cessation rule (stop if prevalence of infection was <5% in children aged 0-5 years halfway between scheduled MDA rounds) 2: Treatment coverage 80-90% vs >90%	TF and C. *trachomatis* infection in children aged 0-5 years after three years follow-up.	32 clusters. 100 children aged 0-5 years.	To determine whether prevalence of infection at 36 months is different in the two coverage arms in communities with prevalence of TF >20%.
5	PRET The Gambia Harding-Esch (2013)[39] Harding-Esch (2010)[49] Stare (2011)[57]	2x2 factorial design; 3 annual MDA rounds (all ages) vs MDA cessation rule (stop if prevalence of infection was <5% in children aged 0-5 years halfway between scheduled MDA rounds) Treatment coverage 80-90% vs >90%	TF and C. *trachomatis* infection in children aged 0-5 years after three years follow-up.	48 clusters. 100 children aged 0-5 years.	To determine whether prevalence of infection at 36 months under the graduation rule is non-inferior to an annual MDA strategy To determine whether prevalence of infection at 36 months is different between coverage arms
6	TANA: MDA frequency 1 year, Ethiopia House (2009)[17]	Quarterly MDA for one year in children aged 1-10 years vs no MDA (delayed for one year)	C. *trachomatis* infection those aged ≥11 years at 12 months follow-up.	32 clusters. 60 children aged 1-10 years, 60 aged ≥11 years.	To determine whether there is a protective herd effect in untreated individuals aged ≥11 years.

	Trial, lead author (year)	Intervention arms	Trachoma Outcomes	Sample size for outcome evaluation	Aims
			C. *trachomatis* infection those aged 1-10 years at 12 months follow-up.		
7	TANA: MDA frequency 4 years, Ethiopia Gebre (2012)[34]	4 rounds of annual MDA vs biannual MDA for 4 years (7 MDA rounds) to all ages.	C. *trachomatis* infection in children aged 0-9 years at 18, 30 and 42 months follow-up.	24 clusters. 60 children aged 0-9 years, 60 aged ≥10 years.	To determine whether biannual (twice-yearly) MDA is more likely to eliminate infection than annual MDA, over a 3.5-year period.
8	TANA: MDA and latrine provision, Ethiopia Stoller (2011)[58]	MDA to all ages at baseline. Latrine provision vs no latrine provision	*C. trachomatis* infection in children aged 0–9 years at 24 months follow-up.	24 clusters. 60 children aged 0-9 years.	To investigate the effect of intensive latrine promotion on reduction and emergence of infection with ocular *C. trachomatis* after mass treatment with antibiotics
9	MDA frequency, Ethiopia Melese (2008)[20]	Two rounds of annual MDA vs 4 rounds of biannual MDA (all ages)	*C. trachomatis* infection in children aged 1-5 years at 24 months follow-up.	16 clusters. Unclear but would appear to be all children aged 1-5 years.	To determine whether biannual (twice-yearly) MDA is more likely to eliminate infection than annual MDA over a two-year period.
10	Single MDA and re-emergence, Ethiopia Chidambaram (2006) [23]	Randomly selected communities from a fixed number of clusters of	*C. trachomatis* infection in children aged 1-5 years at 24 months follow-up.	8 treated villages and 15 control villages	To investigate re-emergence of infection two years after MDA.

	Trial, lead author (year)	Intervention arms	Trachoma Outcomes	Sample size for outcome evaluation	Aims
		communities offered a single MDA round (all ages). Randomly selected control communities from with same clusters of communities		All children aged 0-5 years	
11	Single MDA and herd effect, Ethiopia. Chidambaram (2004)[18]	Randomly selected communities from a fixed number of clusters of communities offered a single MDA round (all ages). Randomly selected control communities from with same clusters of communities	*C. trachomatis* infection in ≤18 months old children at 6 months follow-up	8 treated villages and 8 control villages All children aged 0-18 months	To investigate an indirect herd effect in untreated children who were ineligible for treatment during MDA
12	MDA and health education, Ethiopia Cumberland (2008)[59] Edwards (2008)[45] Edwards (2006)[60]	1: Radio (control) 2: Radio + MDA 3: Radio + MDA + IEC 4: Radio + MDA + IEC + video broadcasts	*C. trachomatis* infection and active trachoma in 3-9 year-olds at 36 months follow-up.	40 clusters. 50 children aged 3-9 years per cluster.	To determine whether health education, in additional to MDA, can provide a further reduction in prevalence of trachoma
13	MDA and fly control, Tanzania West (2006)[61]	Single MDA at baseline vs Single MDA at baseline plus insecticide spraying for fly control.	*C. trachomatis* infection in children aged 1-7 years at 6 months follow-up and TF at 6 and 12 months follow-up	16 clusters. All children under 8 years old within clusters, longitudinal follow-up.	To determine whether intensive insecticide spraying after MDA could reduce trachoma and infection

	Trial, lead author (year)	Intervention arms	Trachoma Outcomes	Sample size for outcome evaluation	Aims
14	Targeted MDA, Nepal Holm (2001)[62]	1: MDA to children 1-10 vs 2: Targeted treatment to children with active trachoma and all household or family members	*C. trachomatis* infection and active trachoma in 1-7 year-olds at 6 months follow-up.	12 clusters. All children aged 1-7 years old.	To compare the effectiveness of a mass treatment strategy to a targeted treatment strategy

PRET = Partnership for Rapid Elimination of Trachoma; TANA = Trachoma Amelioration in Northern Amhara, MDA = mass drug administration, IEC = printed health education materials

Below, an overview of the importance of heterogeneity within each key stage of design and analysis is given, followed by a summary of results from the literature review.

For one of the 14 studies (trial 2 in Table 1), the main effects of MDA interventions on trachoma outcomes were published as three separate reports resulting in 16 references for 14 studies. Hence, the number of reference links given below when summarised design and analysis aspects may be more than the number of trials reported to include the given aspect, where relevant.

1.6.2. Sampling

The sampling frame for trial clusters will depend on the hierarchical population structure in the study area. Population units could take a number of nested forms within an administrative and, or a geographical hierarchy and could be contiguous[43,63]. For example, groups of households within communities, with communities forming larger villages and then groups of villages forming sub-districts within districts. The terminology for the population unit at each level of the hierarchy will vary by country. Trachoma is known to cluster at household and community level[49,50] and so it is reasonable to assume that clustering of trachoma will also occur at each level of the population hierarchy within districts.

The sampling frame of the hierarchical population structure allows for CRTs conducted in districts earmarked for MDA, to include a sampling process for clusters that provides a representative sample of the district. Wider applicability of trial findings to other districts can then be considered. If MDA is distributed throughout the entire district during the trial period, the results at the end of the trial may provide an indication of future control efforts required in the district. CRTs with a convenience or purposive sampling process will provide more widely applicable evidence in areas similar to those meeting inclusion criteria for the trial.

Knowledge of the hierarchical population structure is important, to be able to identify the most appropriate unit to serve as trial clusters, to minimise potential risks of contamination (receipt of interventions in control clusters), and to be able to take appropriate account of clustering of the outcome during the analysis stage. Choice of trial cluster for randomisation and data collection could correspond to the unit at which delivery of community-wide MDA is centralised.

In seven of the 14 trials in the review, a random process was used to select a representative sample of clusters at district-level[17,18,23,34,39,58,62]. In four trials, a sample of clusters with a pre-specified level trachoma prevalence was identified[35,37,38,54]. In two trials, clusters were randomly

selected from areas within districts based on feasibility of implementing the intervention[59,61]. How representative trial clusters might be at district level was unclear for one trial[20].

Given that trachoma clusters in households, additional variability could be present in the data if more than one child per household is included in a survey of trachoma outcomes. In one trial, households were randomly sampled within clusters before all children in the eligible age range were examined[59]. In nine trials, a simple random sample of children within a pre-specified age range was taken from a community census[17,34,35,37-39,54,58,62]. In these trials, more than one child per household could have been included by chance, or deliberately, if the number of children in the target age range in the community happened to be less than or equal to the number of children to be sampled. In three trials, all children within a target age range for whom consent was given were examined[18,23,61] and in one trial, the sampling process was not explicitly stated but is assumed to be all children within a target age range[20]. So, heterogeneity in trachoma outcome data could have been present at household level in all 14 CRTs of MDA for trachoma control.

1.6.3. Randomisation

Heterogeneity in the prevalence of disease between communities, contamination risk (the risk of individuals in control arm clusters receiving the intervention) due to geographical contiguity and small numbers of clusters are important considerations during randomisation of trial clusters. A simple randomisation process may not achieve adequate balance, with a risk of bias when evaluating intervention effects[43,63,64]. Covariate-balance may be desirable for the prevalence of the outcome at baseline, or measures or factors associated with the outcome[65,66]. It may also be necessary to reassure stakeholders that the distribution of interventions is fair across geographical or political areas[43].

Restricted randomisation of clusters (e.g. stratification, matching, applying covariate-based constraints) can reduce variance and achieve balance with respect to known potential confounders[64,67]. However, there is a potential for bias if there too much restriction. In order to assess the validity of possible random allocations with acceptable balance, a validity matrix can be constructed that displays the probability of each pair of clusters being allocated to the same intervention. If the probabilities are not close to 0.5, the restriction criteria should be re-considered[43,63,64]. It is recommended that covariates used to impose restrictions in the randomisation are accounted for in the analysis[68].

Four trials allocated population units to interventions throughout a wider study area, not just trial clusters[17,34,39,58]. Restricted randomisation was used in the form of stratification for three trials[35,39,61]. The stratification variable was not adjusted for in one of these trials[39], but was in the other two trials[35,61]. Stratification plus matching was used for one trial[59] without adjustment for stratification or matching variables in the analysis. Constrained randomisation for covariate-balance was used in three trials[37,38,54], without specification of whether covariate-balance utilised a continuous measure or a binary or categorical variable. Although it was specified for these three trials that the covariates were adjusted for in the analysis, it is also not known whether a continuous or categorical variable corresponding to the one used in the restriction process was used in the adjusted analysis.

Investigations of the validity and extent of restriction of the restricted randomisations were not described for the seven trials that has a restricted randomisation process. Allocation to intervention arm was by simple randomisation for five trials[17,20,34,58,62] and for two trials, clusters were randomly sampled to receive MDA and then control clusters were randomly sampled from the study area[18,23].

Although 50% of the CRTs included in the literature review utilised some form of restricted randomisation, no trial reports mentioned investigation of the extent of restriction that resulted from the process, or of the validity of the set of acceptable randomisations. Generating a complete set, or even a large number, of acceptable randomisations meeting balance requirements and then generating a validity matrix can be computationally intensive. It is possible that researchers also underestimate the importance of checking whether the balance requirements are too restrictive. Statistical software packages for covariate-constrained randomisation packages called "cvcrand" in both R[69] and Stata[70] have been published this year (2018). These packages will generate a list of random allocations that meet pre-specified balance criterion and randomly select one allocation to use in the trial, but do not produce a validity matrix.

A restricted random sampling approach to selection of trial clusters from a larger sample of available clusters has been proposed by Kraschnewski *et al* (2010)[71], to identify how many random samples of trial clusters meet inclusion criteria, from all possible samples of the required number of clusters, so that a random sample selection made from those meeting inclusion criteria. The authors claim that this approach can lead to a representative and unbiased sample of study clusters. However, the authors did not go as far as producing or suggesting a validity matrix to evaluate the extent of restriction of their eligibility criteria. .

1.6.4. Sample size

Commonly used CRT sample size formulae require assumed values for desired levels of power and significance, the expected summary measure of the outcome at follow-up in each of the control and intervention arms, the extent of heterogeneity (or clustering) of the outcome and cluster size, as the number of individuals per cluster[43]. The chosen formula will depend on the summary measure appropriate for the outcome, e.g. proportions, means or rates and the measure of effect of interest, e.g. a difference or a ratio of summary measures. Heterogeneity in the outcome between clusters can be quantified using estimates of intra-cluster correlation (ICC), between-cluster variance, a coefficient of variation or a design effect[43]. An example of a sample size formula for the number of clusters per arm for a binary outcome, to detect a difference between arms of the proportion with the outcome, is given in **Figure 7**.

Figure 7. Number of clusters per arm for a binary outcome

$$c = 1 + (z_{\frac{\alpha}{2}} + z_{\beta})^2 \left[\frac{\left[\frac{\pi_0(1-\pi_0)}{n} + \frac{\pi_1(1-\pi_1)}{n} + k^2(\pi_0^2 + \pi_1^2) \right]}{(\pi_0 + \pi_1)^2} \right]$$

where,

c = number of clusters per arm

$z_{\alpha/2}$ = standard normal distribution value corresponding to the upper tail value of $\alpha/2$ for a significant probability of a difference < α on a two-sided test.

z_{β} = standard normal distribution value corresponding to the upper tail value of β for power of 100(1-β)%.

π_0 = true proportion with the outcome in the control arm (e.g. proportion of individuals with *C. trachomatis* infection)

π_1 = true proportion with the outcome in the intervention arm

k = coefficient of variation of the true proportions between clusters in each arm

Source: Hayes JH, Moulton LH. Cluster Randomised Trials: Chapman & Hall/CRC Press. Taylor and Francis Group.; 2009

In this formula, heterogeneity in the prevalence of the outcome is accounted for using the coefficient of variation. A desirable detectable effect size is linked to the summary measure of the outcome in each arm at follow-up. In the above example, the effect size could be a difference in proportions. The expected efficacy of treatment in those who are treated (e.g. the percentage of individuals whose *C. trachomatis* infection is cleared after taking azithromycin) and the expected percentage of non-participation (not taking treatment when offered it) are not specific

components of the sample size formulae. However, presumed levels of efficacy and non-participation can be incorporated into assumptions about detectable effect sizes and summary measures in each arm at follow-up. In the above example, expected values of π_0 and π_1 could be based on how different the true values are expected to be at follow-up, after consideration of how efficacy and non-participation may influence the prevalence of the outcome in the intervention arm at follow-up and consideration of an effect size that would be of public health importance.

In all but two trials out of 14 in the literature review, primary analyses were based on comparing the prevalence of *C. trachomatis* infection between arms at follow-up. The primary analysis for one trial had a binary outcome based on whether or not clusters had a prevalence of infection ≤1% at follow-up and the sample size was based on test of a difference in proportion of clusters with the outcome[54]. Another trial had a primary analysis that evaluated re-emergence of infection post MDA although did include a second analysis comparing prevalence between arms at follow-up although no sample size justification was reported[23].

Four trials in the literature review were funded by a single grant award to conduct trials in three countries with a design harmonised as much as possible (PRET; trials 2-5 in Table 1) and resulted in six published reports of main effects of MDA interventions on trachoma outcomes. Sample size calculations for these trials were based on a continuous outcome measure of cluster-level prevalence of *C. trachomatis* infection, normalised via a square root transformation, rather than a binary outcome such as in the example above[35,37-41]. Heterogeneity was accounted for based on an assumed value for the standard deviation on the normalised scale, although an assumed mean value for prevalence was not reported for the control arm to correspond to the detectable difference reported. Although all four trials used the same sample size approach, not all trial reports specifically mentioned the square root transformation.

Three trials conducted as part of the TANA grant award (Table 1) and a CRT prior to the TANA trials, conducted by one of the partners in the PRET trials, reported sample size calculations in a similar way based on a continuous measure of cluster level prevalence of infection[17,20,34,58]. For these trials however, there was no mention of a square root transformation in any of the three main trial reports.

Account of potential clustering in the data was not reported for the sample size calculation for one trial[59] and for a further three trials, a sample size calculation was not reported at all[18,23,61,62].

There was no mention of coverage, non-participation or efficacy, in reporting of sample size calculations for any of the trials in the literature review, so it is unclear whether these aspects were considered during trial design.

Typically, two levels are considered in the data hierarchy of a CRT; cluster (level 2) and individual (level 1). In **Figure 7** above, these two levels are considered with account of heterogeneity between clusters. In a trachoma CRT setting, there could be three-levels of data hierarchy in which heterogeneity in the outcome could occur; cluster (level 3), household (level 2) and individual (level 1). Ignoring household-level heterogeneity at level two in sample size calculations could mean a study has reduced power, in trials where the sampling process could result in more than one individual per household. Applications of three-level sample size calculations for CRTs are starting to appear appearing in the literature[72-75] including development of freely available software[73]. A three-level hierarchy was not considered in reported sample size estimations for any of the trials.

1.6.5. Analysis of trachoma outcomes

Analysis options for CRTs are to analyse data at cluster-level or individual-level[43]. Individual-level analysis can incorporate data measured at each level of hierarchy in the sampling frame, e.g. individual, household and cluster or individuals and cluster, and therefore account for heterogeneity in the data at each level above individuals. Analytical approaches for an individual-level analysis could be to use logistic regression for a binary outcome for infection status, and there are a variety of options for adjusting for between-cluster and even between-household variation in the data in individual-level analyses[43], e.g. random effects regression, generalised estimating equations or some robust standard error adjustment for clustering such as Huber-White. A cluster-level analysis accounts for between-cluster variability in the data with the use of cluster summary measures to obtain treatment effects and conduct hypothesis testing[43]. In high prevalence settings, individual-level analyses may offer little additional power compared to cluster-level analyses that yield population averaged results for intervention effects[17,20,76]. In lower prevalence settings, it could be of interest to learn more about how trachoma clusters within households and communities to understand where and how trachoma persists. Individual-level analyses that adjust for heterogeneity in outcome data at both levels two and three in a three-level hierarchy could mean more realistic estimates of effect sizes and reduced risk of Type I error. If a CRT is small, with 15 clusters per arm or less, a cluster-level analysis is recommended[43].

It was of interest to determine, from published reports of CRTs, whether data were analysed at cluster or individual level, which methods were used to account for clustering in the data and whether the analytical approach followed what would be the expected approach based on the sample size assumptions.

Of the 14 trials, 12 had a primary analysis comparing prevalence of trachoma outcomes between arms at follow-up, data from eight trials were summarised and analysed at cluster-level and data from four trials were analysed at individual level.

Of the four trials with sample size calculations based on normalised cluster-level prevalence of *C. trachomatis* infection via a square root transformation, the primary analysis matched the design with linear regression modelling of cluster-level transformed prevalence for three trials[35,37,38,40,41]. For one trial, negative binomial regression was used instead to allow for a high occurrence of clusters with zero cases of infection at follow-up[39].

For the other four trials with a sample size calculation based on cluster-level prevalence as a continuous outcome measure, expressed in a similar way to the trials mentioned above, but with no specific mention of a square root transformation, a variety of analysis approaches were used. If parametric methods are used for comparisons between arms, there could be deviations from an assumption of normally distributed data with a relatively small number of analysis units. A comparison between arms was made using ANCOVA with adjustment for baseline prevalence for one trial[58]. For another trial, a non-parametric Wilcoxon rank sum test was used for comparison between arms[17]. For another trial, linear regression of cluster-level prevalence data was carried out using a robust regression technique that excluded outliers based on Cook's distance[20]. Pooled linear regression was used in another trial to account for survey data collected during repeated surveys[34]. The trial reports do not address whether the assumption of normality was reasonable.

For all four trials with individual-level data analysis, adjustments for heterogeneity in the outcome data were included. For two trials, a Huber-White robust standard error adjustment[18,62] was made to logistic regression models to account for clustering of the outcome between trial clusters. For another trial, generalised estimating equation logistic regression was used[61] and for the remaining trial, three-level random effects logistic regression was used to account for clustering at household and trial cluster level[59].

Heterogeneity in trachoma outcome data was not typically quantified in trial reports; heterogeneity measures for trachoma outcomes using between-unit variance estimates or design effects were only reported at baseline for two trials[49] and at follow-up for one trial[45]. However, eight publications for seven trials included raw data for numbers of individuals examined and numbers of individuals with the outcome, for each trachoma survey[17,20,23,34,39-41,58].

Household variation in the data is only measurable and analysable where the sampling process allows data collection from more than one child per household. Four trials incorporated full population sampling of children by design[18,20,23,61] but based on raw data reported, it is likely that a further three trials included almost all children with consent, as the denominators in each cluster were less than the planned random sample of 60 children within the required age range[17,34,58]. One trial included random sampling of households from within clusters then included all children within the required age range[59]. This was the only trial to specifically account for clustering of trachoma at household level in the analysis. Five trials randomly sampled children within a pre-specified age range from community lists of children[35,37-39,62].

With typically small sample sizes in CRTs, adjustment for baseline covariates may account for chance imbalance between arms in baseline prevalence of trachoma outcomes, which is known to be correlated with prevalence at follow-up, and therefore improve precision of intervention effects[43]. Nine trials adjusted for baseline prevalence in the analysis of follow-up data[20,34,35,37-41,54,58,61].

An ITT approach was used in the primary analysis of trachoma outcomes for all studies apart from one CRT for which receipt of interventions was captured at cluster level but not individual level and an as-treated analysis was performed at cluster level, due to such broad deviation from the allocation schedule[59].

1.6.6. Reporting and handling of coverage and individual participation during MDA

It is plausible that if disease outcomes are correlated within communities and households, then treatment receipt during mass treatment rounds is also correlated within these units[51]. As previously noted, treatment receipt during MDA rounds is commonly referred to as participation in the trachoma literature. During MDA rounds for trachoma control, there are expected minimum coverage targets for each community-wide distribution and it is implicitly assumed that any non-participation will occur homogeneously. That is, non-participation is independent of risk of the outcome. Even homogeneous non-participation could reduce power of a trial, if it

is assumed during trial design that there will be full participation. Heterogeneous non-participation, where those at higher risk of the outcome do not participate during MDA could also have a notable impact on power.

Randomised trial analyses primarily use an intention-to-treat (ITT) analysis approach, assuming interventions were received as allocated. In case of deviation from the allocation, common alternatives are per-protocol (PP, excluding individuals or units who do not receive treatment), or as-treated (analysing data according to receipt of treatment, regardless of allocation to intervention). These alternative analysis populations can lead to biased estimates of intervention effects[77,78], if those who did not receive treatment are those whose risk of the outcome is correlated with a reason for not receiving treatment. Excluding such individuals from the analysis will mean that analysis groups are subject to imbalance with respect to important confounders and therefore not comparable.

Non-compliance (or non-participation in a trachoma setting) with treatment during MDA in CRTs can reduce power to detect ITT effects[79]. It is possible, through adjustment for factors associated with non-compliance, to regain power[80] but bias could still be a problem[78]. Applications of models that account simultaneously for heterogeneity in both outcome and compliance data in CRTs are emerging in the literature[51,79,81], with computation possibly now afforded using specialist software (e.g. MPlus[82]). These applications are considered part of the Complier Average Causal Effect (CACE) methodological framework which draws on latent variable techniques to estimate treatment effects with a lower risk of bias[78]. Certain assumptions are required that can pose challenges to the validity of application of these methods to CRTs, including assumptions that there is no clustering of outcome and treatment data. Proposed solutions are also emerging[79,81,83], mainly for normally distributed outcomes and computational options for non-normal outcome data are still extremely limited.

A recent review of reporting and handling of individual-level receipt of treatment in CRTs highlighted that treatment receipt is often under-reported, with substantial variation in the definition, reporting and handling of treatment receipt in trial analyses[84], despite recommendations of the Consolidated Standards of Reporting Trials statement[85].

Of the 14 trials, a mean of cluster level coverage with either a standard deviation or 95% confidence interval (CI) was reported for six trials[17,34,35,37-41]. For five trials, only an overall percentage of coverage was reported[18,23,54,61,62]. Coverage data was unavailable for one trial[59]. Raw data were available for coverage in each trial cluster for two trials[20,58]. Clustering of

treatment receipt within households during MDA was reported using intra-cluster correlation (ICC) values for three trials (trials 3, 4 and 7 in Table 1)[47,48]. For trial 7, a univariate analyses of factors associated with treatment receipt was also reported based on data collected as part of the main trial[47].

No information was provided from any trials about how non-participation occurred according to infection status prior to MDA rounds, so there is no indication from existing reports about whether non-participation may occur more, or less frequently amongst individuals infected at baseline.

In the analyses of trachoma outcomes, as already noted, the primary analysis was according to ITT for 13 CRTs, although for one trial the ITT comparison based on cluster level allocation of MDA was adjusted for individual treatment receipt during MDA[61]. For two trials, secondary analyses were conducted based on as-treated receipt of azithromycin fitted as a linear effect of cluster level coverage[38,39]. In the report of a third trial with an as-treated secondary analysis, treatment receipt was analysed as "mean antibiotic coverage" so it is unclear exactly how coverage was included in a regression model[37].

1.6.7. Summary

Following a review of published reports of CRTs of MDA interventions for trachoma control, with a focus on how heterogeneity is reported and handled during trial design and analysis, a number of gaps in the literature were identified and allowed identification of objectives for this thesis.

Sampling and Randomisation:

User-written software packages still appear to lack functionality to generate a validity matrix for restricted randomisation, that would enable relaxation or adjustment of the restriction criterion as necessary. Full functionality of an accessible restricted randomisation process could be beneficial if alternative MDA delivery strategies were to be allocated based on fairness of distribution, in areas earmarked for large-scale roll-out of azithromycin MDA, if there are plans for an embedded CRT.

User-written software to enable a restricted random sampling process, with full functionality to assess the validity of acceptable random samples, could be useful to inform representative sampling across large geographical area in future trachoma impact surveys, following global mapping of trachoma to inform the need for SAFE interventions[86,87].

Sample size:

The most recently published CRTs of MDA for trachoma control, included the CRT that provided the motivation of this thesis (described in chapter 2), had sample size calculations based on simulations of a continuous, normalised prevalence of *C. trachomatis* infection, via a square root transformation. For future CRTs of MDA interventions, where it is likely that household clustering of trachoma is expected to be present in the data, the formulaic approach such as that shown in **Figure 7** could be developed further to allow for an additional level of clustering in the data.

Definition and reporting and handling of non-participation:

At the time of initiation of the work in this thesis, there had been little exploration of clustering of non-participation in individual MDA rounds or over multiple MDA rounds in the same trial clusters, or exploration of risk factors for non-participation in the context of trachoma. Historically, the focus of research was on methods for accurate measurement of coverage[88].

CRT analyses are typically based on an ITT analysis population, with little apparent attention paid to how non-participation, or efficacy (clearance of infection in those who take azithromycin), could pose challenges for trial design and analysis. Based on information reported in publications of CRTs of MDA for trachoma, the implications of non-participation and efficacy for CRT power and sample size would appear to have been ignored.

Appropriate analysis methods to account for coverage and treatment receipt in individuals are also unclear for CRTs of MDA interventions.

1.7. Aims and objectives

1.7.1. Overall aim

To investigate the influence of heterogeneity in outcome data and non-participation on the design and analysis of cluster randomised trials to evaluate azithromycin mass treatment interventions for trachoma control.

1.7.2. Specific objectives

i) To investigate heterogeneity in, and determinants of, non-participation during repeated annual MDA rounds, using data from a CRT of MDA interventions for trachoma in The Gambia.

ii) To investigate the impact of non-participation, in relation to baseline prevalence of *C. trachomatis* infection, on power to detect effects in CRTs of mass treatment interventions, using simulation studies.

iii) To investigate the impact of azithromycin on all-age all-cause mortality, as a secondary outcome in a CRT of MDA interventions for trachoma in The Gambia, accounting for heterogeneity in mortality and non-participation during MDA rounds.

1.8. Ethical Approval

Approval for this PhD project was given by the Research Ethics Committee of the London School of Hygiene & Tropical Medicine (Reference number: 6080, dated 25 November 2011).

1.9. Funding

Part-time salary funding for my PhD studies was provided by the MRC Tropical Epidemiology Group (UK Medical Research Council (MRC) and the UK Department for International Development (DFID); grant codes G7508177 and MR/K012126/1).

2. Motivating data: The Partnership for Rapid Elimination of Trachoma (PRET) trial in The Gambia

The PRET collaborative project included three 2x2 factorial CRTs in The Gambia, Tanzania and Niger, comparing alternative frequency and coverage strategies for MDA to current WHO recommendations, with respect to impact on trachoma, infection and cost-effectiveness (clinicaltrials.gov NCT00792922). The Gambia represents a low prevalence setting, with Tanzania and Niger representing medium-high prevalence settings. These CRTs were the first randomised field trials to investigate the impact of MDA coverage strategies specifically and to investigate MDA frequency and coverage strategies simultaneously. The three CRTs had harmonised protocols for primary data collection, although there were some country-endemicity-specific differences in implementation.

The PRET trial in The Gambia was led by London School of Hygiene and Tropical Medicine and is the motivating dataset for this thesis. The Gambia is a low-endemicity setting that could feasibly reach the 2020 elimination target for trachoma[89]. This chapter describes the design of the CRT in the Gambia and the data available.

A 2x2 factorial design allowed for alternative MDA frequency and coverage strategies to be included, for simultaneous evaluation of two intervention approaches over a three-year period. For MDA frequency, an MDA graduation, or stopping, rule (SR) strategy for MDA was compared to annual MDA for three years according to WHO recommendations. For MDA coverage, enhanced efforts to aim for at least 90% coverage were compared to standard efforts that aim for 80% coverage. Six-monthly surveys provided a framework to monitor census data and the prevalence of *C. trachomatis* infection and TF over a three-year period. The first MDA round took place soon after the baseline census and trachoma surveys (baseline mass treatment) and the final survey was at 36 months follow-up.

2.1. My role in the PRET trial in The Gambia

I was the trial statistician for PRET The Gambia, joining the trial team after design of the trial and approval of the protocol. I performed the sampling of clusters, randomisation of clusters to interventions and analyses of baseline, interim and final survey data. I contributed to a number of peer-reviewed manuscripts of study findings (chapter 2.9).

2.2. Trial Aims

- To investigate if the prevalence of *C. trachomatis* infection and TF are non-inferior at three years follow-up in the MDA stopping rule arm, compared to the WHO annual MDA frequency arm.
- To investigate if the prevalence of *C. trachomatis* infection and TF are different at three years follow-up between standard and enhanced coverage arms.

2.3. Interventions

The number of clusters and intervention groups generated with the 2x2 factorial design is shown in Table 2. The four intervention groups were thus;

- *Standard-WHO*: standard coverage for 3 consecutive annual MDA rounds (WHO frequency)
- *Enhanced-WHO*: enhanced coverage for 3 consecutive annual MDA rounds
- *Standard-SR*: standard coverage in communities continuing MDA, discontinuation of MDA if prevalence falls below 5%
- *Enhanced-SR*: enhanced coverage in communities continuing MDA, discontinuation of MDA if prevalence falls below 5%

Table 2. Number of clusters included in random allocation to intervention strategies

		Frequency		Total
		Annual MDA (WHO)[a]:	Graduation rule[b]:	
Coverage	Standard	12	12	24
	Enhanced	12	12	24
Total		24	24	48

[a] 3 consecutive annual MDA rounds

[b] discontinuation of MDA if prevalence of C. *trachomatis* infection falls below 5% in 0-5 year old children in a cluster

2.4. Outcomes

The two primary outcomes were *C. trachomatis* infection and TF in children aged 0-5 years. Secondary outcomes were mortality, enlarged spleen as a proxy measure of malaria morbidity measured in children and mortality in adults.

2.5. Sampling and Randomisation

PRET took place in four districts earmarked for azithromycin distribution; Lower and Central Baddibu in the North Bank Region and Foni Bintang and Foni Kansala in the Western Region, in the South Bank region of the River Gambia.

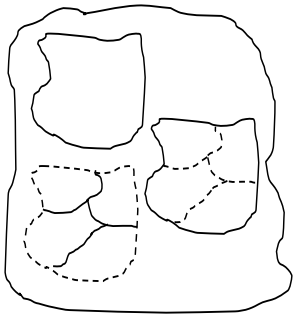
Population structure in The Gambia: Within each district there are two smaller population units; enumeration areas (EAs) and settlements. An EA is a census tract intended to have a population size of approximately 600-800 individuals. The relationship between EAs and settlements varies within districts in one of three ways; 1) an EA and settlement represent the same population unit; 2) an EA comprises of two or more settlements; 3) two or more EAs together form a settlement (Figure 8, top). The unit of randomisation was EA.

Sampling of clusters: Sampling of EAs was stratified by district to enable a representative sample of EAs at district-level and at area level, based on EAs defined in the national census of 2003. Twelve EAs were randomly selected per district such that only one EA per larger settlement was selected (Figure 8, middle).

Randomisation of clusters to interventions: All EAs (approximately 100) in the four districts were randomised simultaneously to one of the four groups. Randomisation was stratified by district such that three EAs per district were allocated to each of the four intervention groups, for 12 EAs per group in total. Further restrictions were applied such that: all EAs within multiple-EA settlements received the same allocation, to avoid contamination in study clusters and for logistic simplicity (Figure 8, bottom); allocation was balanced by district; allocation was balanced throughout the entire study area. Stratified by district, the randomisation process allocated EAs within multiple-EA settlements to the same allocation at the time of allocation of a study EA and then the remaining non-study EAs to one of the four groups, again with multiple-EA settlements assigned to the same group.

Sampling of individuals: A full census was conducted at six-monthly intervals between May 2008 and May 2011 inclusive (seven surveys). Trachoma surveys were conducted after each census round. TF and *C. trachomatis* infection were measured in a random selection of 100 children aged 0-5 years (Figure 9) from each EA census. In three randomly selected EAs per group (12 EAs in total), all children aged 0-9 were examined. Within-household correlation could occur in the data if more than one child per household was selected. This could occur by chance or by necessity in smaller EAs. Each individual, household and EA had a unique identifier number.

Figure 8 Population Units in The Gambia



Population Unit Boundaries:

Outer solid line=district boundary

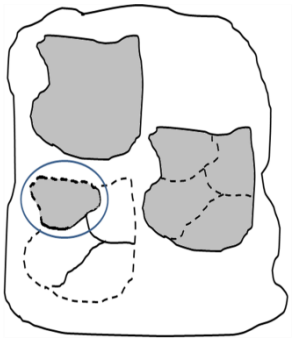
Solid line within district=EA boundary

Dashed line within district=settlement boundary

top left: single EA-settlement unit

bottom left: multiple-EA settlement

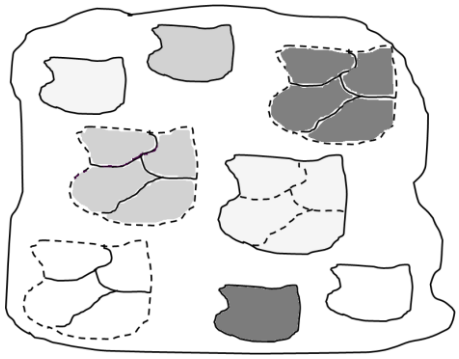
bottom right: multiple-settlement EA



Restricted Random Sampling:

From within each district, a settlement can only be represented once to avoid additional variability in the data; if an EA was randomly selected as a trial cluster from a multiple-EA settlement, the other EAs in the same settlement cannot be selected.

Shaded areas represent acceptable cluster sample selection



Restricted Randomisation of Entire Study Area:

The trial took place in 48 EAs but the entire district was to receive azithromycin MDA according to the randomisation schedule.

Within districts, settlements must receive the same allocation to avoid contamination with a fair, balanced overall allocation within those districts.

Shading represents a possible allocation to four groups

2.6. Sample size

Twenty-four clusters per arm were estimated to have at least 80% power to detect non-inferiority of the MDA stopping rule compared to three years of annual MDA within a margin of 8%, (i.e. an absolute difference in prevalence of 8%) at three years follow-up, in children aged 0-5 years old, assuming a two-tailed alpha of 0.05.[57].

2.7. Available data

Census data: Population structure of EAs and settlements in each district prior to the baseline survey in 2008. Seven 6-monthly census rounds began with a baseline survey in May-June 2008, then at 6, 12, 18, 24, 30 and finally 36 months follow-up in May-June 2011. Census data included a complete list of all members of all households in the 48 study EAs, including deaths, movement

and new additions during the study period. Household level data were recorded for water and latrine access, education of household head and recall of community health education in the EA.

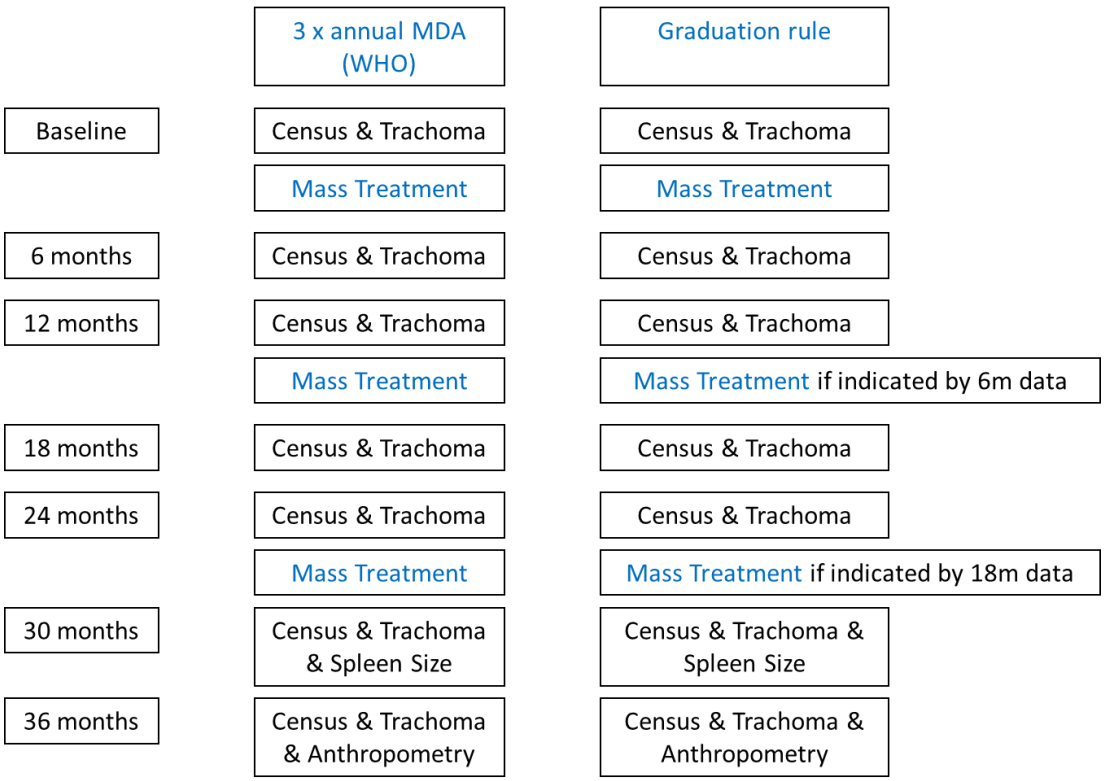
Trachoma data: *C. trachomatis* infection, TF and whether the child had nasal discharge or ocular discharge from seven surveys conducted after each census round. Detailed field and laboratory procedures are published[44].

Treatment data: Individual-level treatment receipt recorded against the census list for each MDA round in each cluster. Treatment coverage was calculated relative to EA census data for each treatment round.

Secondary outcome data: spleen size, as a proxy for malaria morbidity, in children aged 0-5 years at 30 months follow-up, anthropometry measurements in children aged 1-4 years at 36 months follow-up, all-age all-cause mortality as captured in census data during the study period (**Figure 9**).

Additional data: Global positional system (GPS) co-ordinates for study households.

Figure 9 Trial surveys and data collection



2.8. Data summary

2.8.1. Participants

The resident population of children aged 0-5 years was around 7000 and around 30,000 for individuals of all ages, at each census time point. The median number of enumerated children aged 0-5 years per household at each survey was 2 or 3 with an inter-quartile range of 1-4 or 1-5. Trachoma outcomes were measured for approximately 5000 children aged 0-5 years old per survey.

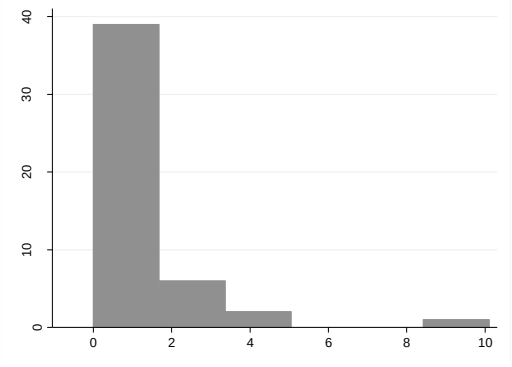
2.8.2. Distribution of *C. trachomatis* infection, TF and non-participation at baseline

At baseline, the overall prevalence (95% CI) of TF and *C. trachomatis* infection was 6.3% (5.6-7.0%) and 0.8% (0.5-1.0%). Cluster-level prevalence data were approximately exponentially distributed (Figure 10). An analysis of factors associated with TF and *C. trachomatis* infection at baseline using random effects logistic regression[49], confirmed that there was significant clustering of TF at EA and household level (between-household variation was 1.11 (standard error (SE) 0.34), between-EA variation 1.10 (SE 0.32). Due to sparse data, it was not possible to obtain similar measures of clustering for infection. Just over half of the infections (n=21, 54%) were detected in nine EAs in one district in the South Bank region. Infection was detected in two EAs in the other South Bank region district and in the North Bank regions, in one EA in one district and three EAs in the other district. Fourteen EAs contained one or more *C. trachomatis* infections at baseline, and ten at 36 months[39].

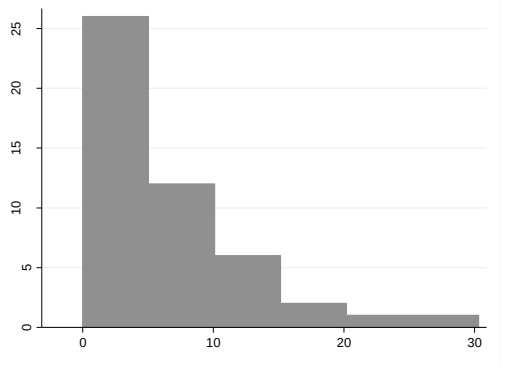
2.8.3. Treatment coverage

EA-level treatment coverage in children under 10 years old ranged from 63%-98% in the standard coverage arm and 63%-99% in the enhanced coverage arm, at baseline. Similar ranges were observed at 12 and 24 months and in individuals of all ages[39]. Cluster level prevalence of non-participation also appeared to be exponentially distributed (Figure 10). A detailed analysis of participation during MDA rounds is the focus of the next chapter.

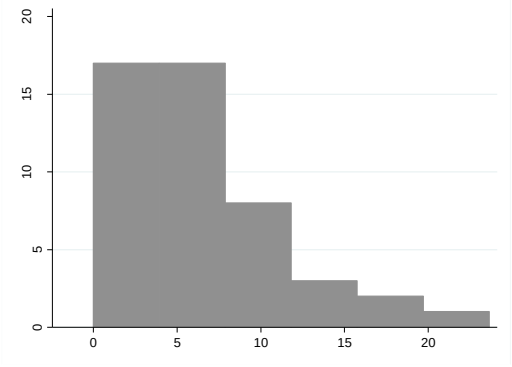
Figure 10. Cluster-level prevalence of infection, TF and non-participation at baseline in 48 trial clusters



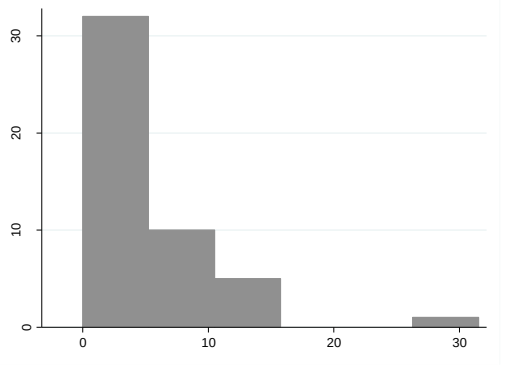
Prevalence of C. trachomatis infection at baseline in children aged 0-5 years old



Prevalence of TF at baseline in children aged 0-5 years old



Percent non-participation per EA at baseline in children aged 1-9 years old



Percent non-participation per EA at baseline in examined children aged 0-5 years old

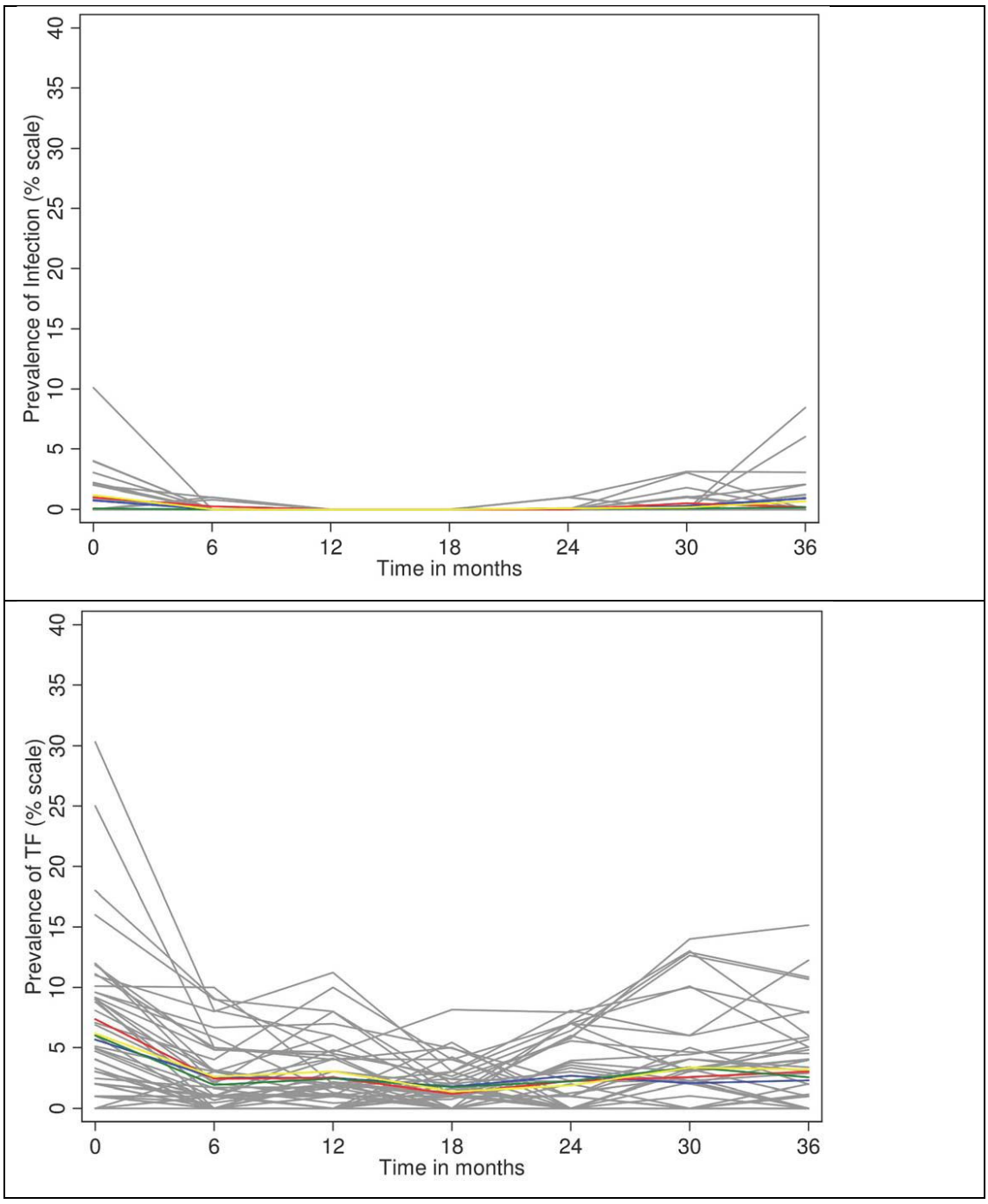
2.8.4. MDA Graduation

For communities allocated to the MDA graduation, or stopping rule arm, MDA was to be discontinued if the prevalence of C. *trachomatis* infection had fallen below 5% in children aged years old children in a cluster, based on interim trachoma surveys at 6 and 18 months follow-up. As there were no children with infection in the graduation arm at 6-months follow-up or at 18-months follow-up, MDA only occurred once at baseline in the SR arm.

2.8.5. Prevalence of *C. trachomatis* infection and TF at follow-up

The prevalence of infection was reduced to zero in all clusters by 12 months follow-up (Figure 11) and remained below 1% at the end of the trial at 36 months follow-up[39]. In this low endemicity setting on target for elimination by 2020, the prevalence of *C. trachomatis* infection was negligible and the prevalence of TF well below the mass treatment threshold, after only one MDA round.

Figure 11. Prevalence of *C. trachomatis* infection (top) and TF (bottom) during the trial.



Grey lines: percent prevalence for each cluster in each survey. Coloured lines: cluster mean percent prevalence by intervention arm (blue: standard coverage-annual MDA, red: enhanced coverage-annual MDA, green: standard-graduation rule (baseline MDA only), yellow: enhanced-graduation rule (baseline MDA only).

Source: Figures 4 and 5 of Harding-Esch EM, Sillah A, Edwards T, Burr SE, Hart JD, Joof H, et al. (2013) Mass Treatment with Azithromycin for Trachoma: When Is One Round Enough? Results from the PRET Trial in The Gambia. PLoS Negl Trop Dis 7(6): e2115.

2.9. My contributions to peer-reviewed publications

I am a co-author on four peer-reviewed publications arising from the main results of the PRET study, for which I conducted data analysis and contributed to writing the papers, especially sections related to methods and results. Two papers relate to trachoma outcomes and two to secondary outcomes.

1: Burr SE, Hart J, Edwards T, Harding-Esch EM, Holland MJ, Mabey DC, Sillah A, Bailey RL. Anthropometric indices of Gambian children after one or three annual rounds of mass drug administration with azithromycin for trachoma control. BMC Public Health. 2014 Nov 18;14:1176.

2: Hart JD, Edwards T, Burr SE, Harding-Esch EM, Takaoka K, Holland MJ, Sillah A, Mabey DC, Bailey RL. Effect of azithromycin mass drug administration for trachoma on spleen rates in Gambian children. Trop Med Int Health. 2014 Feb;19(2):207-11.

3: Harding-Esch EM, Sillah A, Edwards T, Burr SE, Hart JD, Joof H, Laye M, Makalo P, Manjang A, Molina S, Sarr-Sissoho I, Quinn TC, Lietman T, Holland MJ, Mabey D, West SK, Bailey R; Partnership for Rapid Elimination of Trachoma (PRET) study group. Mass treatment with azithromycin for trachoma: when is one round enough? Results from the PRET Trial in the Gambia. PLoS Negl Trop Dis. 2013 Jun 13;7(6):e2115.

4: Harding-Esch EM, Edwards T, Mkocha H, Munoz B, Holland MJ, Burr SE, Sillah A, Gaydos CA, Stare D, Mabey DC, Bailey RL, West SK; PRET Partnership. Trachoma prevalence and associated risk factors in the Gambia and Tanzania: baseline results of a cluster randomised controlled trial. PLoS Negl Trop Dis. 2010 Nov 2;4(11):e861.

2.10. Ethical approval for the trial

Ethical approval for the PRET trial was obtained from the Research Ethics Committee of the London School of Hygiene & Tropical Medicine (LSHTM), UK; The Gambia government/Medical Research Council (MRC) Joint Ethics Committee, The Gambia; the Johns Hopkins Institutional Review Board; and the Tanzanian National Institute for Medical Research. Oral consent was obtained from the village leaders and written (thumbprint or signature) consent from the child's guardian at the time of examinations, which was signed by an independent witness. The research was done in accordance with the declaration of Helsinki.

The PRET trial in The Gambia was registered with ClinicalTrials.gov NCT00792922

2.11. Trial Funding

The PRET trials were funded by a grant from the Bill and Melinda Gates Foundation, awarded to John's Hopkins University, Baltimore, USA.

I received salary support for my statistical role in the trial, through a sub-contract awarded to the London School of Hygiene & Tropical Medicine.

3. Non-participation during azithromycin mass treatment rounds in PRET The Gambia: Heterogeneity and risk factors

The work for this chapter has been published following peer-review. I designed the experiment, performed the analyses and wrote the paper. The full citation is,

Edwards T, Allen E, Harding-Esch EM, Hart J, Burr SE, Holland MJ, Sillah A, West SK, Mabey D, Bailey R. Non-participation during azithromycin mass treatment for trachoma in The Gambia: heterogeneity and risk factors. PLoS Neglected Tropical Diseases. 2014 Aug 28;8(8):e3098.

RESEARCH PAPER COVER SHEET

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

PLEASE NOT THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Tansy Edwards
Principal Supervisor	Professor Robin Bailey
Thesis Title	Heterogeneity in cluster randomised trials of mass treatment for trachoma control

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	PLoS Neglected Tropical Diseases		
When was the work published?	2014		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion		n/a	
Have you retained the copyright for the work?*	Yes, (CC BY 4.0)	Was the work subject to academic peer review?	Yes

If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I designed the experiment, performed the analyses and wrote the paper.

Student Signature: **Date: 14/08/2018**

Supervisor Signature: **Date: 14/08/2018**

3.1. Introduction

3.1.1. Importance of non-participation during MDA for trachoma control

The success of MDA for NTDs is thought to depend heavily on adequate population coverage in affected areas and participation amongst those offered treatment [88,90-92]. With increasing provision of MDA for trachoma, prevalence is expected to fall so that endemic areas will, over time, become low prevalence settings on a trajectory towards the endgame of elimination [93]. In such settings, MDA participation amongst those at highest risk of infection is important. If spatial clusters, or hotspots, of non-participation occur during MDA and correlate with hotspots of infection, it could be speculated that reservoirs of infection could remain to facilitate continued transmission. This would in turn increase the time needed to reach elimination goals. Identification of factors associated with persistent non-participation in low prevalence settings could therefore provide important clues about how to minimise non-participation. Determining whether infected individuals are amongst non-participators in previous annual MDAs may also provide information regarding the importance of non-participation in low prevalence areas and the potential need for resources to improve participation.

3.1.2. Factors associated with non-participation during MDA for trachoma control

C. *trachomatis* infection, follicular trachoma (TF) and non-participation with azithromycin MDA have all been found to cluster within communities and also within households [10,44-47,50,94]. Limited data on non-participation in trachoma control suggest that non-participation is associated mainly with household level decision-making factors, related to knowledge and awareness of trachoma control and also mode of delivery (for example, perception of community drug distributors). A case-control study in Tanzania found household level risk factors such as guardians of children reporting poorer health, increased burden due to family health, more children per household and younger guardians [95]. At community level, enhanced effort to increase coverage during implementation of MDA was successful in achieving higher participation rates. Studies in Nigeria and South Sudan identified prior household head knowledge of trachoma control and prior notification of MDA as factors associated with better participation but no association with age or gender [96,97]. In a cluster randomised trial (CRT) in Ethiopia, women and younger children were more likely to be non-participators [47]. In the PRET trial in Tanzania, non-participation in the post-baseline MDA was not associated with infection status prior to MDA in four trial clusters [50].

This chapter investigates non-participation in 1-9 year old children in the PRET study districts in The Gambia during three annual MDAs. The purpose of the work presented in this chapter is to quantify non-participation amongst children aged 1-9 years during PRET, to identify factors

associated with non-participation of different types at child, household and community level, to investigate the presence of heterogeneity of non-participation at household and, or community level and determine if any observed household or community heterogeneity is spatially clustered.

3.2. Methods

3.2.1. Study Design

A cross-sectional analysis of participation during each of the three MDA rounds conducted within approximately one month after baseline, year one and year two census surveys was conducted. Longitudinal data collected from repeated census surveys were also used to investigate persistent non-participation.

For each MDA round, treatment receipt and eligibility were categorised according to one of the following categories, using treatment data recorded on census forms by the National Eye Care Program (NECP) treatment distribution teams:

- ineligible (not added to cohort at time of MDA, deceased before MDA, moved elsewhere;
- present not treated (PNT): eligible and present during MDA but not treated;
- treated;
- eligible for treatment but absent (EBA): eligible for treatment as a resident but absent on the treatment day;
- eligible for treatment but status unknown (EBU): treatment status not recorded.

Care was taken to account for population movement during PRET and to ensure that treatment eligibility and receipt was captured appropriately for those who moved.

3.2.2. Participants

The focus of this study on children aged 1-9 years because prevalence of TF is used for control decision making in WHO recommendations for trachoma control [98].

The size of this study was determined by the PRET sample size which was based on power to detect effects of the MDA strategies on the primary outcomes of *C. trachomatis* infection and follicular trachoma in children aged 0-5 [39,44]. All available data were used in the analysis for this study.

For the cross-sectional analysis of non-participation at each time point, children residing in treated communities who were either PNT, treated or EBA were eligible for inclusion, that is, children eligible to receive treatment and with known treatment status. The number of resident children aged 1-9 years in the PRET was 9777 immediately prior to the baseline treatment in 48 treated EAs. In the 24 treated communities at year one and year two, there were 5504 and 6086 eligible children, respectively (Table 7).

3.2.3. Treatment Distribution

A central treatment station was set up in each community during MDAs. Adults aged 14 years or above received 1g of azithromycin, with height used as a surrogate for weight for children's dosing on the basis of 20 mg/kg. Treatment was directly observed by the treatment team and the number of tablets or ml of suspension recorded.

NECP staff attended the initial training workshop for the PRET trial. Prior to each MDA, treatment team leaders received training about recording treatment status on census forms from the trial coordinator and about dosing from NECP. Team leaders trained their team. Data review and feedback took place throughout MDAs. Communities were sensitised to MDA by the trial field team before fieldwork started. During the census prior to treatment, the study was again explained to households, and the expected dates for examination and treatment teams' visits were explained. Supervisory field visits were conducted by the NECP to ensure appropriate distribution. Treatment team members were given per diems to cover food and accommodation for days spent in the field, as a single payment at the end of the fieldwork based on the expected number of days needed.

3.2.4. Outcomes

The two binary outcomes which were analysed for each MDA round were 1) PNT versus treated and 2) EBA versus treated.

3.2.5. Other data

It was of interest to investigate the effects of covariates measured at EA, household and child level. The following data were collected as per the standardised trial procedures.

EA level covariates included coverage delivery strategy allocation, location on either the North or South river bank and the corresponding district, EA type (single settlement, multi-settlement, or segment of a settlement) and EA population size (small: <600, medium: 600-800, large: >800 individuals).

For households, covariates included size (small: <11, medium: 11-16, large: >16 individuals), latrine access, time to primary water source, recall of community health education, years of education of household head, a diagnosis of TF for a child aged 0-5 years in the household during the survey immediately prior to the MDA and treatment status of the household head. It was hypothesised that a recent diagnosis of TF in a child in the household could be associated with subsequent treatment choice either through knowledge of TF and appropriate treatment.

Child level covariates were gender, age, participation in a previous ocular examination survey and treatment status at previous MDAs.

Latitude and longitude coordinates were measured for each household using a GPS eTrex® handheld device [99] .

3.2.6. Analysis methods

Data were analysed using Stata[70] and SaTScan™ [100] and maps were produced using Quantum GIS [101].

The number (%) of children treated, PNT or EBA was summarised overall and by characteristics of interest for each MDA. The same analysis approach was used for each binary outcome. Firstly, the data from all 48 treated EA at baseline were used to conduct univariate random effects logistic regression in order to identify individual associations with each outcome. EA level random intercept terms were included in all models. A household level random intercept was also included for EBA versus treated comparisons but not for the PNT analysis due to the relatively low prevalence of PNT. Factors associated with the outcome, identified by a likelihood ratio test (LRT) p-value of <0.1 in univariate analyses, were included in a step-wise model building approach to obtain a final multivariate model. A LRT p-value of <0.1 was used for inclusion and exclusion in the final model. Coverage delivery allocation was included in all multivariate models *a priori,* since by design, the enhanced allocation was intended to increase participation. The same final multivariate models were fitted to the year one and two MDA data for validation. In addition, treatment status at previous treatment rounds was added to each of these final models *a priori*. Tests for interaction in final models were pre-planned between coverage allocation and other covariates if coverage allocation was associated with the non-participation outcome. Intra-cluster correlation coefficients (ICCs) with corresponding 95% confidence intervals were obtained from final multivariable models.

Considering the study districts in areas north and south of the River Gambia separately (Figure 12), spatial point patterns were investigated using Kulldorf's scan statistic [102] for MDA round (baseline, year one and year two) to test whether PNT and EBA cases were randomly distributed over space compared to treated children and identify the location of any significant spatial clusters. Within SatScan software, a circular window is moved systematically throughout the geographic space to identify clusters by centring the window on each household location with a window size of 0% to 50% of study population to allow detection of small and large clusters. Clusters containing more than 50% of the population are ignored. A LRT test for a Poisson based model was conducted for each location and size of scanning window to test the hypothesis of an increased rate of non-participator type compared with the distribution outside. P-values corresponding to the most likely and secondary clusters are determined using Monte Carlo replications of the dataset.

Spatial clusters of PNT and EBA children were identified and added to maps showing the location of children and their treatment status. The locations of infected children at year three are shown on the map for the year two MDA for visual inspection (Figure 13, Figure 14).

3.3. Results

3.3.1. Extent of non-participation and heterogeneity

Participation was high overall during each MDA (Table 7). The overall prevalence of non-participation at baseline was 6.2% with 1.0% of children PNT and 5.2% of children EBA. Over the three MDAs, the percentage of EBA children appeared to increase and the percentage of PNT children decrease. By year two, overall non-participation increased to 10.4% (paired t-test of EA summary data $p<0.01$) due to the increase in EBA children. Reductions in PNT non-participation were not significant.

Of 1626 households eligible for treatment in the 24 communities participating in all three MDAs, one household (0.1%) had PNT children in all three MDAs and 34 (2.1%) had EBA children in all three MDAs. Persistent EBA households were generally larger and within EA comprising of multiple settlements. The persistent PNT household was further from water, without latrine access and with a household head with no recall of health education or education. Households with non-participating children had either PNT or EBA children, not a mixture of non-participator types.

3.3.2. Associations with non-participation

Treatment status was captured such that it was possible to distinguish between untreated individuals eligible for treatment as present in the community on the day of the treatment team visit and those resident but absent during the visit. To determine whether to proceed with an analysis of non-participation as a binary outcome or consider PNT and EBA children as separate outcome categories, random effects logistic regression was used to test for differences whilst accounting for any EA and household clustering of cases.

PNT and EBA children differed by district (p=0.001), household size (<0.001), household head years of education (0.046), latrine access (p=0.006) and household head treatment status (p<0.001), after adjustment for between-EA variation (Table 7). It was of interest to compare each type of non-participator to treated children. Due to the small numbers of PNT children, use of multi-level multinomial regression analysis techniques was not possible. Thus, analyses proceeded with the two binary comparisons as stated, by comparing each type of non-participator to treated children (PNT versus treated and EBA versus treated) using random effects logistic regression. In univariate analyses (Table 3), categories of household head treatment status were combined with other categories if zero cases of PNT were observed (for numbers in each category, see Table 7).

The final multivariate model for being PNT versus treated at baseline included coverage allocation, time to water, household size, household head treatment status and district (Table 4). Children residing in a medium or large household compared to small (p<0.001) and within 15 minutes of primary water source (p<0.001) were less likely to be PNT. A child was more likely to be PNT if the household head was untreated (p<0.001). An association with district was found (p=0.002), due to a difference between the study districts south of the river. There was no effect apparent of coverage allocation (p=0.842). A TF diagnosis in a child aged 0-5 years old in the household during the baseline examination round prior to MDA was linked to reduced odds of being PNT in the univariate analysis (Table 3). After accounting for household size and coverage allocation during model building, a recent TF diagnosis in the household was no longer associated with PNT non-participation (p=0.163). Recent TF diagnoses were more common in larger households where PNT non-participation was less likely and so these covariates were likely to be explaining the same variation in the data.

The same final model was applied to the year one and year two data, adding the child's previous treatment status (Table 4). For these follow-up MDAs, the fixed term for district was removed due to zero PNT cases in study districts north of the river. Treatment status one year previously

was an important predictor of non-participation at both years one and two, with children who were PNT at the previous round being more likely to be PNT again the following year (baseline treatment status at year one MDA: p=0.034, year one treatment status at year two MDA: p=0.032, Table 4). Treatment status at baseline was not associated with being PNT at year two (p=0.656).

The final multivariate model for being EBA versus treated at baseline (Table 5) suggested EBA non-participation was more likely for children who were not part of the baseline ocular examination (p<0.001), aged 3-5 or 1-2 years compared to 6-9 years (p<0.001). Also for children whose household head was also EBA compared to treated, who resided in households further from water (p=0.018) and possibly for those whose household head could not recall community health education (p=0.060). Coverage allocation was not associated with being EBA (p=0.166). Children who were EBA at each previous round were more likely to be EBA at later time points (Table 5). Results also suggest that children who were ineligible at both previous treatment rounds were more likely to be EBA at year two.

There was evidence of notable clustering of non-participation types in ICCs from final multivariate models. In the EBA versus treated comparisons ICCs suggested substantially more variation was present between households within EAs, than between EAs (Table 5). ICCs from PNT models at EA level were closer to the ICCs estimated at household level for EBA children. A possible explanation for this is that variability does exist between households for PNT status (i.e. clustering at household level is more prominent than EA clustering) but the very low prevalence of PNT non-participation meant that between-household variation was undeterminable in this dataset.

Spatial coordinates were unavailable for 11 households, excluding 23 children from spatial analyses. Spatial clusters of PNT and EBA children were detected at baseline in study areas on each side of the river bank (Table 6, Figure 13, Figure 14). No PNT children were reported in year one or year two in the northern river bank districts. Spatial clusters of PNT and EBA children reduced in size in each subsequent MDA and by year two, clusters included either single households or a small group of households, possibly representing compounds.

Seventeen cases of *C. trachomatis* infection in annually treated communities at year three were found close to the northern and southern Senegalese borders, over which trachoma is endemic and MDA has not taken place. These infections were detected amongst ineligible or treated children during the three prior MDAs, apart from one child near the northern border who was

persistently EBA during the MDAs. Two cases were located in an EA with households within a year two EBA cluster on the southern side of the river (Figure 14). In the two EAs with households in this spatial cluster, approximately 15% of children aged 1-9 years old were EBA during the year two MDA.

Trachoma surveys were conducted in a random sample of 100 children from each cluster. At baseline, 38 children aged 0-5 years old had *C. trachomatis* infection and 37 were known to be treated in the in the post-baseline MDA round. The remaining child had a missing treatment status. There were no infections detected immediately prior to the year one MDA. One infection was detected in the MDA arm at year two and that child was treated.

3.4. Discussion

In this large study of non-participation during azithromycin MDA from a low prevalence trachoma setting, we demonstrate evidence of heterogeneity of non-participation in children aged 1-9 years, particularly at household level, in line with studies in higher prevalence settings [47,48]. We also observed persistent non-participation over time in annual MDAs, as seen elsewhere in a CRT setting [95].

Geographical clustering of non-participation represents a new finding and we found two different types of non-participators. We found some evidence, though not statistically significant, of an association between infection and non-participation during a previous MDA, however, the prevalence of infection and TF in children aged 0-5 years old at the end of PRET was below a level requiring any SAFE interventions. Detection of infection in communities close to untreated higher prevalence areas [39], relatively high EBA rates in those communities during the previous MDA and literature from The Gambia and elsewhere linking travel with re-infection [103,104], together suggest the observed infections resulted from exposure to untreated persons. Travel plans could have been set prior to notification of MDA timing and so unrelated to intentional non-participation, although intentional decisions to miss treatment is a possibility.

Household level covariates were associated with greater likelihood of being PNT and EBA. Household head non-participation and their type of non-participation predicted PNT and EBA status in children, implying household decision making with respect to MDA participation behaviour. The finding that children in households further from their primary water source were more likely to be PNT or EBA is probably indicative of some other unmeasured risk factor, for example, marginalisation within the community due to either household head or community leader choice, or a mixture of the two. Active trachoma has also been found to be associated

with lower socio-economic status (SES) and isolation of household from the community [105] so access to, or participation in, trachoma control activities could also be affected by these unmeasured factors. Smaller household size was important for predicting PNT status but not EBA, compared to treated children, which could represent some effect of lower SES. Recent migration into the community could also mean less access to community decision making and activities. Participation in a previous TF examination survey could be indicative of increased awareness and acceptance of control activities in annually treated communities, however, a proxy effect cannot be concluded in case of potential bias introduced by households more willing to take part in all control and assessment activities. Results from the Gambian setting suggest that enhanced efforts to increase coverage of mass treatment programs, by means of an extra treatment team visit to the community do not improve participation, in contrast to the PRET trial conducted in Tanzania [95].

Extent and implications of missing data for treatment status:

The outcome in this analysis was treatment status in children aged 1-9 years old, categorised as treated, PNT or EBA). Children in this age group with missing treatment status were excluded from the analysis of each time point (missing at baseline: 403/10180 (4.0%); year 1: 146/5650 (2.6%); Year 2: 175/6261 (2.8%)). These were the only exclusions of data points due to EBU status.

In all MDA rounds, missing treatment status in adults was more common than in children aged 1-9 years old. It is thought that this is because the 1-9 years age group were considered the most important target group to receive azithromycin, as the group in whom trachoma prevalence determines decision making for MDA distribution, and that missing treatment status is more likely to correspond to untreated rather than treated status. Treatment status of the head of the household of each child was included as a covariate as a proxy for household level decision making about receiving treatment during MDA, in the analysis of each time point. In order to include as much available data as possible, household head treatment status was initially categorised as treated, present not treated, eligible but absent, ineligible or as a final category of eligible but unknown (EBU) to account for the group with unknown (missing) treatment status.

Due to data sparsity, treatment status amongst household heads was re-categorised as i) treated, or ii) untreated or EBU, for the multivariable analysis of PNT (present not treated) versus treated in children aged 1-9 years old at baseline and year two. Although data were also sparse for PNT vs treated outcomes at year one amongst categories of household head

treatment, the analysis did include household head EBU as a separate category. Both household head untreated and EBU status categories were associated with increased odds of PNT versus untreated status in children aged 1-9 years old. In the analysis of EBA (eligible but absent) versus treated status in children, data sparsity led to a combined category of ineligible and EBU for household head treatment status at baseline (under a hypothesis that both groups were most likely untreated). In the year one and year two analyses, household head EBU status was associated with increased odds of EBA versus untreated status in children aged 1-9 years old, as was a household head status of EBA. These results correspond to the hypothesis that adults with missing treatment status were untreated and household-level decision making determines treatment status is children.

Treatment status of the children included in each analysis at previous MDA rounds was also included as a covariate to investigate associations with treatment status in a current round and past behaviour with regard to receiving treatment. In these analyses it was possible to include children eligible for treatment but with a missing treatment status in a previous round, as an EBU category. Eligible but unknown previous treatment status was not associated with either PNT or EBA status in a current MDA round.

Less than 5% of children aged 1-9 years olds had missing treatment status in each analysis. Exclusions of children with unknown treatment status could have led to bias in the results if missing treatment status was associated with one of the outcomes (PNT or EBA) and an important risk factor. Two extreme examples of the occurrence of bias and the implications on results would be;

- underestimation of strength and magnitude of association: if the children with missing status in baseline univariable analyses were almost all PNT and residing in households without a recent TF diagnosis, the previous odds ratio for lower odds (0.39, p=0.025) of PNT in households with a recent TF diagnosis would have been much closer to zero (further from the null effect) with a much smaller p-value.
- estimation of effect in opposite direction: if the children with missing status in baseline univariable analyses were in fact all EBA and residing in households less than 15 minutes from water, the odds ratio for lower odds (0.58, p=0.010) of EBA associated with residing less than 15 minutes from a water source would change direction to show increased, rather than reduced, odds of being EBA for households closer to their water source, with strong evidence of this association.

It is likely, however, with such large sample sizes available for analysis at each time point and

with data representative of all clusters, any impact of bias on the results was negligible and there were plausible explanations for the findings of the analyses conducted. It is not suggested that the purely hypothetical scenarios of potential bias given above, occurred.

Studies of MDA participation in Africa for onchocerciasis and lymphatic filariasis, other NTDs for which control is through mass community-wide treatment, have also linked non-participation to household level decision making factors, for example, a perception of low disease risk or lack of family or household support [106-108]. The Gambia has relatively high vaccination coverage [109], elimination of trachoma by 2020 is attainable [89] and non-participation was higher in the districts south of the river where the prevalence of TF was consistently lower during PRET [39]. It is perhaps plausible therefore that a household level decision based on a perceived lack of need for treatment could apply in this low prevalence setting, although we do not have data from each community to assess this. Reasons for being EBA in this setting could be logistic and independent of participation choices, for example, population movement and travel where children are sent away for weaning which is common practice in The Gambia or farming related activities.

We found a geographical effect on non-participation and trachoma outcomes[39]. Infections did occur in one part of the study area with notable EBA non-participation at the previous MDA, however, even if all PNT and EBA children at the year two MDA had been found to have infection and TF, the overall prevalence of each outcome at year three would have been less than 5% and thus below MDA continuation thresholds. Similarly, there was no evidence that non-participation occurs more frequently in children infected prior to MDA rounds. Therefore, for the Gambian national trachoma control program, efforts and resources to address non-participation are not required. It might be the case that this finding translates to other low prevalence settings with high coverage of MDA.

3.5. Conclusions

For national control programs in low and medium prevalence settings, heterogeneous non-participation linked to increased risk of infection could present challenges for elimination efforts. Links between infection and non-participation in prior MDA rounds could result in prevalence levels meeting criteria for continued MDA at the time of impact assessment. Identification of hotspots of infection and non-participation, along with modifiable risk factors for non-participation could take place during impact assessment following repeated MDA. The results could then aid control program managers working towards elimination goals in low and

medium prevalence settings, by enabling them to target delivery resources for continued MDA, to improve coverage in areas with a greater threat of continued transmission.

Although the impact of non-participation was likely to be negligible in the communities included in the PRET trial in The Gambia, it is plausible that non-participation could have non-negligible effects during the analysis of a CRT in a higher prevalence setting, especially if non-participation occurs in those infected prior to MDA. The next chapter will explore, via simulation studies, the impact of non-participation on power to detect effects in CRTs of azithromycin MDA for trachoma and consider applications to CRTs of other NTDs with MDA as the main control measure.

Table 3. Univariate analysis of associations with each non-participation type

		PNT vs Treated[a] N=9272		EBA vs Treated[b] N=9678	
Characteristic		OR (95 CI)	LRT p-value[c]	OR (95 CI)	LRT p-value[c]
Coverage	Standard	1		1	
	Enhanced	0.93 (0.25-3.53)	0.916	0.58 (0.34-0.99)	0.051
Bank	South	1		1	
	North	0.18 (0.04-0.85)	0.016	1.28 (0.74-2.22)	0.375
District	Foni Bintang	1		1	
	Foni Kansala	9.43 (1.65-54.2)	0.002	0.52 (0.25-1.09)	0.288
	Lower Baddibu	0.33 (0.03-3.34)		0.86 (0.41-1.82)	
	Central Baddibu	1.62 (0.25-10.7)		1.00 (0.48-2.07)	
EA type	Multiple-SET	1		1	
	Multiple-EA	0.42 (0.10-1.84)	0.289	1.23 (0.66-2.28)	0.793
	Single EA-SET	0.25 (0.03-2.02)		1.15 (0.51-2.59)	
EA population size[d]	Small	1		1	
	Medium	2.19 (0.43-11.3)	0.386	0.62 (0.33-1.18)	0.292
	Large	3.23 (0.59-17.7)		0.65 (0.33-1.28)	
Household size[e]	Small	1		1	
	Medium	0.43 (0.26-0.69)	<0.001	0.93 (0.67-1.30)	0.921
	Large	0.22 (0.12-0.40)		0.97 (0.69-1.37)	
Latrine access	No	1		1	
	Yes	0.54 (0.26-1.10)	0.106	1.13 (0.69-1.85)	0.640
Time to water	> 15 mins	1		1	
	< 15 mins	0.48 (0.28-0.80)	0.005	0.58 (0.38-0.87)	0.010
Recall of health education	No	1		1	
	Yes	1.25 (0.71-2.23)	0.443	0.70 (0.50-0.98)	0.037
Years of education of head	<1 year	1		1	
	≥1 year	1.31 (0.64-2.68)	0.472	0.91 (0.48-1.71)	0.764
Gender	Male	1		1	
	Female	0.76 (0.50-1.15)	0.188	1.11 (0.90-1.37)	0.322
Age (years)	6-9	1		1	
	3-5	1.00 (0.62-1.61)	0.396	1.02 (0.80-1.30)	<0.001

		PNT vs Treated[a] N=9272		EBA vs Treated[b] N=9678	
Characteristic		OR (95 CI)	LRT p-value[c]	OR (95 CI)	LRT p-value[c]
	1-2	1.39 (0.83-2.31)		1.65 (1.27-2.15)	
Recent TF diagnosis in HH	No	1		1	
	Yes	0.39 (0.15-0.99)	0.025	0.94 (0.61-1.46)	0.786
Previous TF exam	Yes	1		1	
	No	1.46 (0.92-2.30)	0.101	2.85 (2.17-3.75)	<0.001
Household head treatment status[f]	Treated	1		1	
	PNT	-	<0.001	1.42 (0.47-4.34)	0.001
	EBA	-		2.68 (1.68-4.29)	
	Untreated or EBU	3.85 (2.38-6.22)		-	
	Ineligible or EBU	-		0.85 (0.33-2.17)	

PNT = present not treated, EBA = eligible but absent during MDA, EBU = eligible during MDA but treatment status unknown, HH = household, EA = enumeration area, SET = settlement.

[a] EA level random effect included in logistic regression model.

[b] EA and household (HH) random effects included in logistic regression model.

[c] LRT = likelihood ratio test of overall association, comparing models with and without characteristic of interest.

[d] EA size: small <600, medium 600-800, large ≥800

[e] HH size: small <11, medium 11-16, large ≥17

[f] some re-grouping of categories necessary due to zero events in some categories.

Table 4. Multivariate models for PNT versus treated children

		Baseline N = 9272		Year one N = 5131		Year two N = 5479	
Characteristic		OR (95 CI)	LRT p-value[a]	OR (95 CI)	LRT p-value[a]	OR (95 CI)	LRT p-value[a]
Coverage	Standard	1		1		1	
	Enhanced	1.14 (0.32-4.00)	0.842	2.17 (0.16-29.0)	0.557	1.02 (0.08-12.8)	0.988
District	Foni Bintang	1					
	Foni Kansala	9.66 (1.72-54.1)	0.002				
	Lower Baddibu	0.42 (0.04-4.14)					
	Central Baddibu	1.33 (0.21-8.91)					
Household population size [b]	Small	1		1		1	
	Medium	0.45 (0.28-0.73)	<0.001	0.94 (0.38-2.33)	0.375	0.81 (0.25-2.63)	0.370
	Large	0.23 (0.13-0.42)		1.78 (0.68-4.64)		0.43 (0.12-1.48)	
Time to water	> 15 mins	1		1		1	
	< 15 mins	0.37 (0.22-0.62)	<0.001	7.01 (1.05-47.1)	0.019	0.09 (0.03-0.30)	<0.001
Household head treatment status[c]	Treated	1		1		1	
	PNT	-	<0.001	-	<0.001	-	<0.001
	EBA	-		-		-	
	Ineligible	-		1.99 (0.24-16.2)		-	
	EBU	-		16.0 (4.11-62.6)		-	
	PNT or EBA			36.2 (16.4-80.0)		-	
	Untreated or EBU	3.90 (2.38-6.40)		-		12.4 (4.57-33.6)	
Baseline treatment status[c]	Treated	-		1		1	
	PNT	-		40.2 (4.73-341.8)	0.034	3.43 (0.17-67.8)	0.656

		Baseline N = 9272		Year one N = 5131		Year two N = 5479	
Characteristic		OR (95 CI)	LRT p-value[a]	OR (95 CI)	LRT p-value[a]	OR (95 CI)	LRT p-value[a]
	EBA	-		1.01 (0.13-9.28)		-	
	Ineligible			0.94 (0.32-2.72)		-	
	Eligible-unknown			1.68 (0.45-6.25)		-	
	EBA, ineligible, EBU			-		0.80 (0.25-2.52)	
Year one treatment status[c]	Treated	-				1	
	PNT	-				11.7 (1.27-108.6)	0.032
	EBA	-				6.19 (1.44-26.5)	
	Ineligible or EBU					2.43 (0.67-8.79)	
ICC (EA)		0.43 (0.25-0.63)		0.62 (0.33-0.84)		0.60 (0.26-0.86)	

Models include an EA level random effect

PNT = present not treated, EBA = eligible but absent during MDA, EBU = eligible during MDA but treatment status unknown, HH = household, EA = enumeration area, SET = settlement.

[a] LRT = likelihood ratio test of overall association, comparing models with and without characteristic of interest.

[b] HH size: small <11, medium 11-16, large ≥17

[c] some re-grouping required due to zero PNT children in some categories.

Table 5. Multivariate models for **EBA** versus treated children

		Baseline N = 9678		Year one N = 5459		Year two N = 6064	
Characteristic		OR (95 CI)	LRT p-value[a]	OR (95 CI)	LRT p-value[a]	OR (95 CI)	LRT p-value[a]
Coverage	Standard	1		1		1	
	Enhanced	0.62 (0.32-1.24)	0.166	0.53 (0.28-1.17)	0.132	0.77 (0.47-1.26)	0.314
Water access	>15 mins	1		1		1	
	< 15 mins	0.59 (0.38-0.91)	0.018	2.27 (1.22-4.22)	0.007	0.69 (0.46-1.04)	0.076
Recall of health education	No	1		1		1	
	Yes	0.72 (0.51-1.02)	0.060	0.91 (0.59-1.41)	0.597	1.24 (0.90-1.70)	0.191
Age (years)	6-9	1		1		1	
	3-5	1.82 (1.39-2.39)	<0.001	2.57 (1.79-3.69)	<0.001	1.51 (1.16-1.97)	0.001
	1-2	2.99 (2.23-4.02)		3.62 (2.49-5.27)		1.62 (1.21-2.17)	
Previous TF exam	Yes	1		1		1	
	No	4.47 (3.29-6.07)	<0.001	2.09 (1.48-2.96)	<0.001	1.35 (1.04-1.75)	0.026
Household head treatment status[b]	Treated	1		1		1	
	PNT	1.49 (0.48-4.63)	0.001	0.53 (0.03-8.33)	<0.001	0.51 (0.08-3.03)	<0.001
	EBA	2.82 (1.73-4.59)		3.43 (1.71-6.88)		4.11 (2.59-6.51)	
	Ineligible	-		3.30 (1.59-6.82)		3.27 (1.03-10.4)	
	EBU	-		3.09 (1.04-9.19)		2.23 (1.23-4.02)	
	Ineligible or EBU	0.95 (0.36-2.53)		-		-	
Baseline treatment status	Treated			1		1	
	PNT			1.88 (0.40-8.75)	<0.001	0.66 (0.12-3.77)	0.011
	EBA			3.97 (2.41-6.52)		1.72 (1.05-2.83)	

		Baseline N = 9678		Year one N = 5459		Year two N = 6064	
Characteristic		OR (95 CI)	LRT p-value[a]	OR (95 CI)	LRT p-value[a]	OR (95 CI)	LRT p-value[a]
	Ineligible			1.03 (0.70-1.50)		1.48 (1.13-1.94)	
	EBU			0.98 (0.37-2.62)		1.81 (0.94-3.52)	
Year one treatment status	Treated					1	
	PNT					0.90 (0.30-2.73)	<0.001
	EBA					7.56 (5.20-11.0)	
	Ineligible					1.54 (1.15-2.06)	
	EBU					0.77 (0.22-2.72)	
ICC (EA)		0.15 (0.09-0.24)		0.08 (0.03-0.18)		0.05 (0.02-0.11)	
ICC (HH)		0.53 (0.46-0.60)		0.51 (0.42-0.59)		0.38 (0.31-0.45)	

Models include EA and household random effects.

PNT = present not treated, EBA = eligible but absent during MDA, EBU = eligible during MDA but treatment status unknown, HH = household, EA = enumeration area, SET = settlement.

[a] LRT = likelihood ratio test of overall association, comparing models with and without characteristic of interest.

[b] At baseline, re-grouping of household head treatment status was required due to zero EBA in the ineligible household head status group.

[c] some re-grouping required due to zero PNT children in some categories.

Table 6. Spatial clusters of non-compliance

Round	Study districts	Type	Clusters	Radius (km)	p-value
Baseline	North River Bank	PNT	1	3.13	<0.001
		EBA	1	6.27	<0.001
			2	0.062	0.001
			3	0.048	0.002
			4	0	0.010
	South River Bank	PNT	1	7.43	<0.001
		EBA	1	3.64	<0.001
			2	0.054	<0.001
			3	0	0.001
			4	0.22	0.010
Year one	North[a]	EBA	1	0.85	<0.001
	South	PNT	1	4.80	<0.001
		EBA	1	0	<0.001
			2	0.12	0.001
Year two	North [a]	EBA	1	0.079	<0.001
			2	0.080	0.027
	South	PNT	1	0	<0.001
			2	0	0.002
			3	0	0.002
			4	0	0.0002
		EBA	1	0.25	<0.001
			2	0.026	0.001
			3	0.35	0.013

PNT = present not treated, EBA = eligible but absent during MDA.

[a] no PNT cases at year one or year two in districts north of the River Gambia

Table 7. Treatment status amongst children aged 1-9 years eligible for treatment at each time point

		Baseline Treatment Status N = 9777				Year one N = 5504				Year two N = 6086			
Characteristic		N	Treated	PNT	EBA	N	Treated	PNT	EBA	N	Treated	PNT	EBA
Total		9777	9178 (93.8)	99 (1.0)	505 (5.2)	5504	5086 (92.4)	45 (0.8)	373 (6.8)	6086	5457 (89.6)	22 (0.4)	607 (10.0)
Coverage	Standard	4793	4436 (92.6)	48 (1.0)	309 (6.4)	2598	2379 (91.6)	23 (0.9)	196 (7.5)	2928	2590 (88.5)	9 (0.3)	329 (11.2)
	Enhanced	4984	4737 (95.1)	51 (1.0)	196 (3.9)	2914	2715 (93.2)	22 (0.8)	177 (6.1)	3162	2871 (90.8)	13 (0.4)	278 (8.8)
Bank	North	4784	4461 (93.2)	47 (1.0)	276 (5.8)	2799	2672 (95.5)	0 (0)	127 (4.5)	3013	2777 (92.2)	0 (0)	236 (7.8)
	South	4993	4712 (94.4)	52 (1.0)	229 (4.6)	2705	2414 (89.2)	45 (1.7)	246 (9.1)	3073	2680 (87.2)	22 (0.7)	371 (12.1)
District	Foni Bintang	2111	1981 (93.8)	6 (0.3)	124 (5.9)	1092	1007 (92.2)	5 (0.5)	80 (7.3)	1259	1107 (87.9)	6 (0.5)	146 (11.6)
	Foni Kansala	2882	2731 (94.8)	46 (1.6)	105 (3.6)	1613	1407 (87.2)	40 (2.5)	166 (10.3)	1814	1573 (86.7)	16 (0.9)	225 (12.4)
	Lower Baddibu	2099	1979 (94.3)	4 (0.2)	116 (5.5)	1190	1151 (96.7)	0 (0)	39 (3.3)	1224	1130 (92.3)	0 (0)	94 (7.7)
	Central Baddibu	2685	2482 (92.4)	43 (1.6)	160 (6.0)	1609	1521 (94.5)	0 (0)	88 (5.5)	1789	1647 (92.1)	0 (0)	142 (7.9)
EA type	Multiple-SET	6036	5647 (93.6)	83 (1.4)	306 (5.1)	3337	3045 (91.3)	17 (0.5)	275 (8.2)	3703	3339 (90.2)	9 (0.2)	355 (9.6)
	Multiple-EA	2402	2263 (94.2)	12 (0.5)	127 (5.3)	1611	1515 (94.0)	25 (1.6)	71 (4.4)	1754	1544 (88.0)	9 (0.5)	201 (11.5)
	Single EA-SET	1339	1263 (94.3)	4 (0.3)	72 (5.4)	556	526 (94.6)	3 (0.5)	27 (4.8)	629	574 (91.3)	4 (0.6)	51 (8.1)
EA population size [a]	Small	2341	2188 (93.5)	8 (0.3)	145 (6.2)	1757	1638 (93.2)	2 (0.1)	117 (6.6)	786	704 (89.6)	0 (0)	82 (10.4)
	Medium	3426	3241 (94.6)	34 (1.0)	151 (4.4)	1009	963 (95.4)	6 (0.6)	40 (4.0)	1658	1476 (89.0)	8 (0.5)	174 (10.5)
	Large	4010	3744 (93.4)	57 (1.4)	209 (5.2)	2738	2485 (90.8)	37 (1.3)	216 (7.9)	3642	3277 (90.0)	14 (0.4)	351 (9.6)

		Baseline Treatment Status N = 9777				Year one N = 5504				Year two N = 6086			
Characteristic		N	Treated	PNT	EBA	N	Treated	PNT	EBA	N	Treated	PNT	EBA
Household population size[b]	Small	3200	2981 (93.1)	53 (1.7)	166 (5.2)	1609	1491 (92.7)	12 (0.7)	106 (6.6)	1259	1144 (90.8)	6 (0.5)	109 (8.7)
	Medium	3118	2936 (94.2)	29 (0.9)	153 (4.9)	1828	1703 (93.2)	18 (1.0)	107 (5.9)	1816	1640 (90.3)	9 (0.5)	167 (9.2)
	Large	3459	3256 (94.1)	17 (0.5)	186 (5.4)	2067	1892 (91.5)	15 (0.7)	160 (7.7)	3011	2673 (88.8)	7 (0.2)	331 (11.0)
Latrine access	No	900	847 (94.1)	11 (1.2)	42 (4.7)	605	555 (91.7)	7 (1.2)	43 (7.1)	714	638 (89.3)	4 (0.6)	72 (10.1)
	Yes	8877	8326 (93.8)	88 (1.0)	463 (5.2)	4899	4531 (92.5)	38 (0.8)	330 (6.7)	5372	4819 (89.7)	18 (0.3)	535 (10.0)
Water access	> 15 mins	1497	1350 (90.2)	39 (2.6)	108 (7.2)	757	718 (94.9)	3 (0.4)	36 (4.7)	835	729 (87.3)	11 (1.3)	95 (11.4)
	< 15 mins	8280	7823 (94.5)	60 (0.7)	397 (4.8)	4747	4368 (92.0)	42 (0.9)	337 (7.1)	5251	4728 (90.0)	11 (0.2)	512 (9.8)
Recall of education program	No	6612	6191 (93.6)	54 (0.8)	367 (5.6)	3821	3514 (92.0)	26 (0.7)	281 (7.3)	4266	3831 (89.8)	20 (0.5)	415 (9.7)
	Yes	3165	2982 (94.2)	45 (1.4)	138 (4.4)	1683	1572 (93.4)	19 (1.1)	92 (5.5)	1820	1626 (89.3)	2 (0.1)	192 (10.6)
Years of education of head	<1 year	9204	8632 (93.8)	89 (1.0)	483 (5.2)	5169	4789 (92.6)	39 (0.8)	341 (6.6)	5728	5152 (89.9)	15 (0.3)	561 (9.8)
	≥1 year	573	541 (94.4)	10 (1.8)	22 (3.8)	335	297 (88.6)	6 (1.8)	32 (9.6)	358	305 (85.2)	7 (2.0)	46 (12.8)
Gender	Male	5063	4752 (93.9)	58 (1.1)	253 (5.0)	2887	2663 (92.2)	23 (0.8)	201 (7.0)	3134	2835 (90.4)	15 (0.5)	284 (9.1)
	Female	4714	4421 (93.8)	41 (0.9)	252 (5.3)	2617	2423 (92.6)	22 (0.8)	172 (6.6)	2952	2622 (88.9)	7 (0.2)	323 (10.9)
Age (years)	6-9	3998	3775 (94.4)	38 (1.0)	185 (4.6)	2274	2153 (94.6)	17 (0.8)	104 (4.6)	2531	2320 (91.7)	9 (0.4)	202 (8.0)
	3-5	3591	3379 (94.1)	34 (1.0)	178 (5.0)	1888	1735 (91.9)	15 (0.8)	138 (7.3)	2019	1796 (89.0)	7 (0.4)	216 (10.7)
	1-2	2188	2019 (92.3)	27 (1.2)	142 (6.5)	1342	1198 (89.2)	13 (1.0)	131 (9.8)	1540	1345 (87.3)	6 (0.4)	189 (12.3)
Baseline treatment status	Treated	-	-	-	-	4340	4059 (93.5)	30 (0.7)	251 (5.8)	3873	3552 (91.7)	13 (0.3)	308 (8.0)
	PNT	-	-	-	-	35	29 (82.8)	3 (8.6)	3 (8.6)	35	31 (88.6)	2 (5.7)	2 (5.7)
	EBA	-	-	-	-	244	195 (79.9)	1 (0.4)	48 (19.7)	236	198 (83.9)	0 (0)	38 (16.1)
	Ineligible	-	-	-	-	758	690 (91.0)	5 (0.7)	63 (8.3)	1832	1590 (86.8)	7 (0.4)	235 (12.8)
	EBU	-	-	-	-	127	113 (89.0)	6 (4.7)	8 (6.3)	114	90 (79.0)	0 (0)	24 (21.1)

		Baseline Treatment Status N = 9777				Year one N = 5504				Year two N = 6086			
Characteristic		N	Treated	PNT	EBA	N	Treated	PNT	EBA	N	Treated	PNT	EBA
Year one treatment status	Treated	-	-	-	-	-	-	-	-	4683	4320 (92.3)	11 (0.2)	352 (7.5)
	PNT	-	-	-	-	-	-	-	-	50	41 (82.0)	3 (6.0)	6 (12.0)
	EBA	-	-	-	-	-	-	-	-	283	188 (66.4)	3 (1.1)	92 (32.5)
	Ineligible	-	-	-	-	-	-	-	-	1028	871 (84.7)	5 (0.5)	152 (14.8)
	EBU	-	-	-	-	-	-	-	-	46	41 (89.1)	0 (0)	5 (10.9)
TF in HH prior to treatment	No	8435	7897 (93.6)	94 (1.1)	444 (5.3)	5173	4768 (92.1)	45 (0.9)	360 (7.0)	5717	5114 (89.5)	22 (0.4)	581 (10.2)
	Yes	1342	1276 (95.1)	5 (0.4)	61 (4.6)	331	318 (96.1)	0 (0)	13 (3.9)	373	347 (93.0)	0 (0)	26 (7.0)
Previous TF exam	Yes	4827	4594 (95.2)	33 (0.7)	200 (4.1)	3561	3331 (93.5)	23 (0.7)	207 (5.8)	4386	3967 (90.5)	16 (0.4)	403 (9.2)
	No	4950	4579 (92.5)	66 (1.3)	305 (6.2)	1943	1755 (90.4)	22 (1.1)	166 (8.5)	1704	1494 (87.7)	6 (0.4)	204 (12.0)
Household head treatment status	Treated	8678	8189 (94.4)	70 (0.8)	419 (4.8)	4815	4529 (94.0)	13 (0.3)	273 (5.7)	5294	4825 (91.1)	7 (0.1)	462 (8.7)
	PNT	161	124 (77.0)	28 (17.4)	9 (5.6)	47	21 (44.7)	25 (53.2)	1 (2.1)	48	39 (81.2)	7 (14.6)	2 (4.2)
	EBA	626	560 (89.5)	1 (0.2)	65 (10.4)	262	218 (83.2)	2 (0.8)	42 (16.0)	417	324 (77.7)	0 (0)	93 (22.3)
	Ineligible	33	33 (100)	0 (0)	0 (0)	255	211 (82.7)	1 (0.4)	43 (16.9)	65	48 (73.9)	0 (0)	17 (26.1)
	EBU	279	267 (95.7)	0 (0)	12 (4.3)	125	107 (85.6)	4 (3.2)	14 (11.2)	262	221 (84.4)	8 (3.1)	33 (12.6)

Data are n (%).

PNT = present not treated, EBA = eligible but absent during MDA, EBU = eligible during MDA but treatment status unknown, HH = household, EA = enumeration area, SET = settlement.

[a] EA size: small <600, medium 600-800, large ≥800

[b] HH size: small <11, medium 11-16, large ≥17

Figure 12. Map of The Gambia showing study districts on the North and South sides of the River Gambia (dark grey: study districts, pale grey: remaining districts)

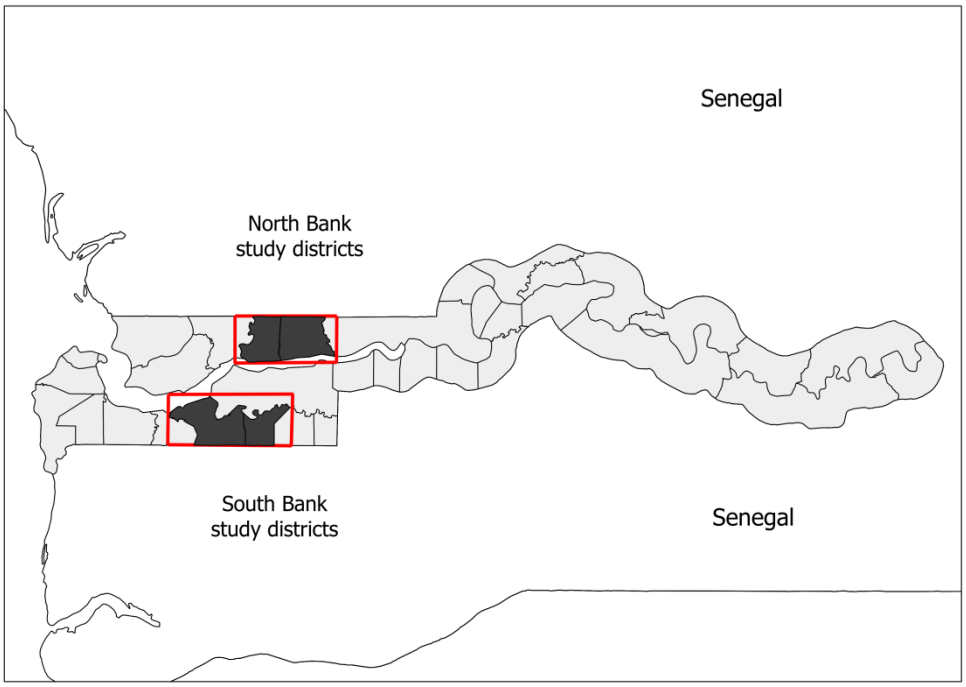
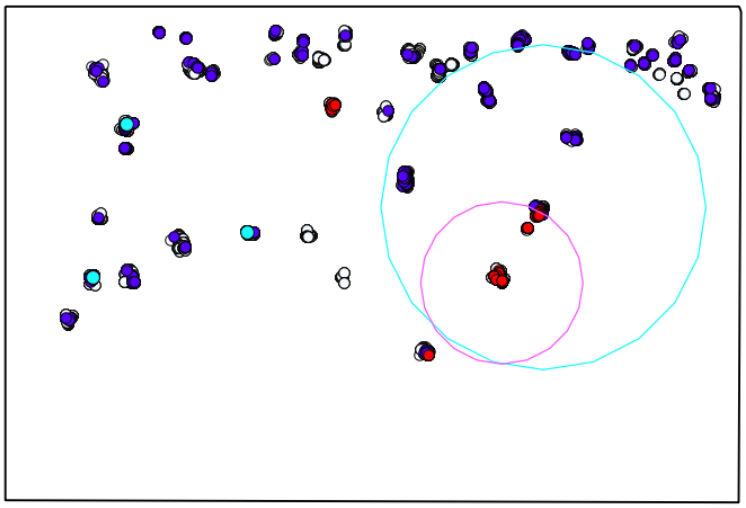
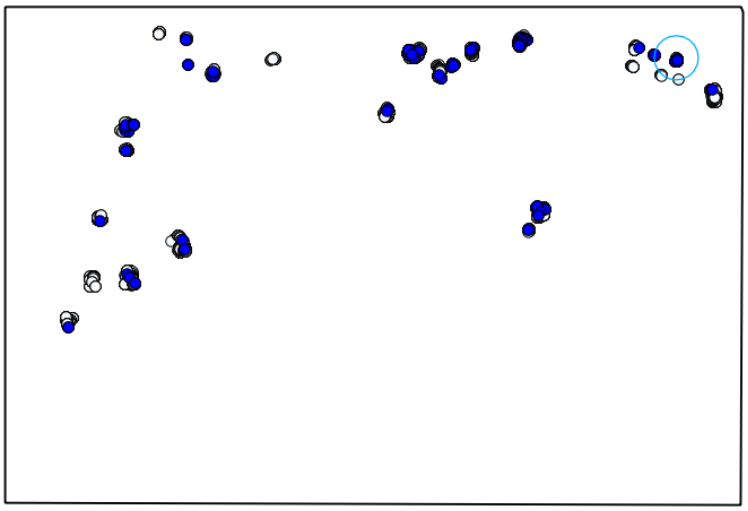


Figure 13**.** Geographical clusters of PNT and EBA non-participation in northern study districts (A: baseline treatment round, B: year one, C: year two)

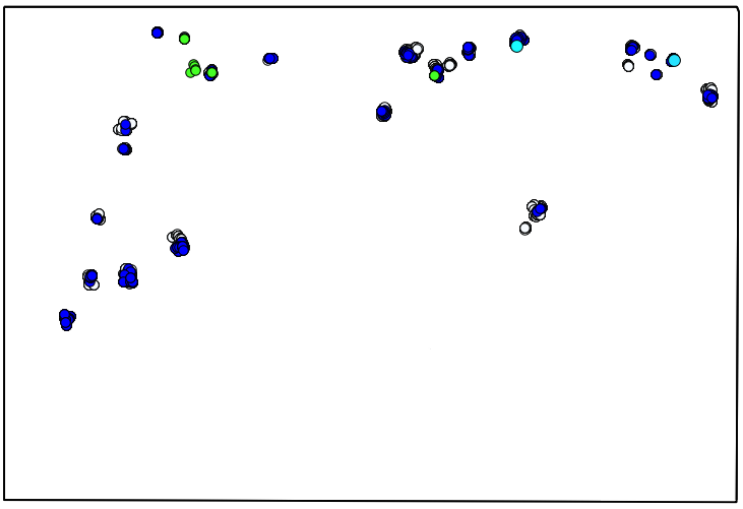
A: Baseline



B: Year one



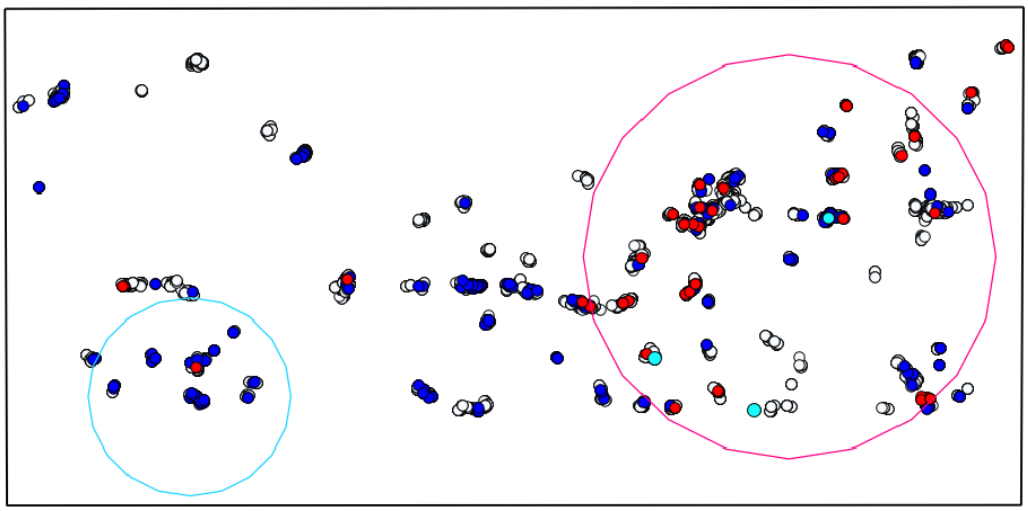
C: Year two



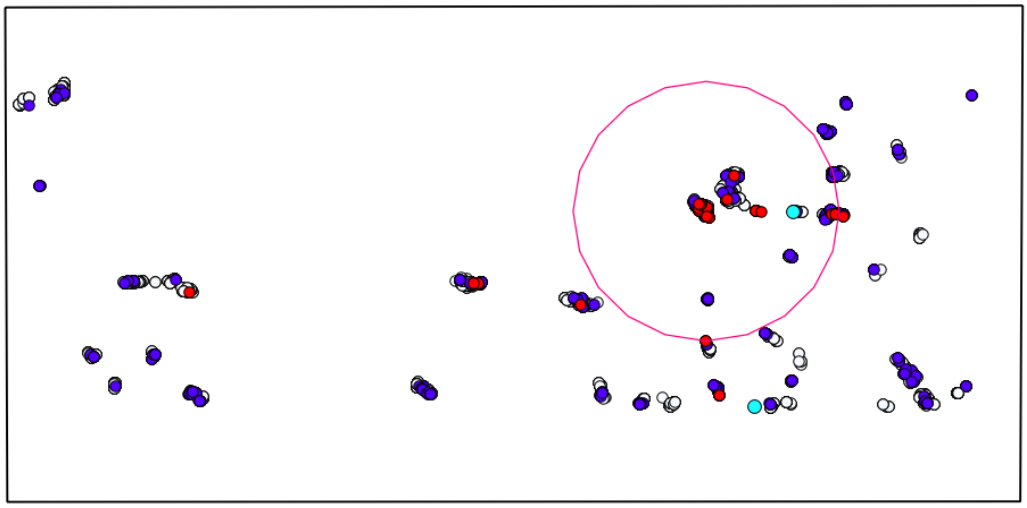
Treated (grey), PNT (red), EBA (blue), PNT cluster (pink), EBA cluster (light blue). No PNT children at year one or year two in study districts north of the river. Children aged 0-5 years with C*t* infection at year three (green)

Figure 14. Location of PNT and EBA children aged 1-9 years by HH and spatial clusters of HHs with PNT and EBA children in southern study districts (A: baseline treatment round, B: year one, C: year two)

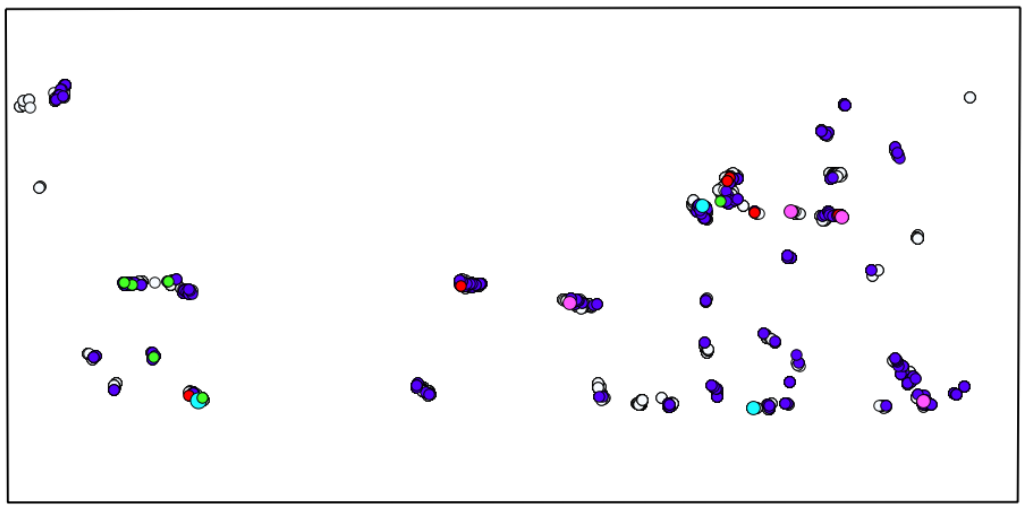
A: Baseline



B: Year one



C: Year two



Treated (grey), PNT (red), EBA (blue), PNT cluster (pink), EBA cluster (light blue). Children aged 0-5 years with *C. trachomatis* infection at year three (green)

4. Implications of non-participation and efficacy on power for design of cluster randomised trials evaluating mass drug administration (MDA) interventions: application to trachoma control with azithromycin MDA

4.1. Introduction

Mass drug administration (MDA) is the mainstay of control for five neglected tropical diseases is; trachoma, onchocerciasis, lymphatic filariasis, schistosomiasis and soil-transmitted helminths[110]. When individuals are eligible but absent, or choose not to take the treatment offered, during community- or school-wide MDA, this has been referred to as non-adherence, non-compliance, or in the context of trachoma, non-participation[111,112]. Non-participation during MDA can hinder control and elimination efforts by maintaining reservoir of infection. Typically, the term *coverage* is used to describe the percentage of the eligible target population who receive treatment during MDA rounds. Reports of coverage do not commonly distinguish between individuals infected or uninfected at the time of the MDA (i.e. by infections status at baseline). Indeed, the reason for administering the drugs irrespective of infection status may be that this status is not easy to ascertain. Nevertheless, it is the non-participation amongst infected individuals that is of most concern during MDA and any evaluation of MDA interventions.

The optimal trial design for evaluation of MDA interventions that are randomised and delivered to clusters of individuals, rather than randomisation of individuals to intervention or no intervention, is a cluster randomised trial (CRT)[43], due to the likelihood of clustering of the outcome(s) and because community- or school-wide delivery is considered a convenient and cost-effective way to reach the target population for treatment[113]. There is strong evidence of clustering of *Chlamydia trachomatis* infection, signs of disease and non-participation during azithromycin MDA for trachoma, within households and communities[45,49,50,89,94,114].

Non-compliance is known to reduce power to detect intention-to-treat (ITT) effects in both individually and cluster randomised trials[51,79,115]. A CRT comparing an MDA intervention with non-participation amongst infected individuals during MDA, to an arm with no MDA, will likely have a smaller than hypothesised effect in the intervention arm as more infection will remain at follow-up. This could then lead to a smaller than hypothesised effect size, that the trial may not be adequately powered to detect. In the context of NTDs, non-compliance, non-adherence or non-participation might be used interchangeably to refer to those eligible for MDA but who do not receive the treatment offered.

Coverage targets for azithromycin MDA rounds for trachoma are at least 80% according to WHO recommendations[27] based on prevalence of signs of disease, namely follicular trachoma (TF), therefore it follows that some non-participation is to be expected (up to 20% non-participation overall if coverage is at least 80%). In addition, although a single dose of azithromycin is known to be an effective treatment for clearance of *C. trachomatis* infection, it may not successfully clear infection in every individual who receives it. That is, the efficacy in individuals, or percentage of treated individuals whose infection is cleared, may be less than 100%. Modelling studies based on field data from a CRT in Tanzania have suggested that azithromycin clears *C. trachomatis* infection in around 70% of individuals in practice[116]. Therefore, it is feasible that lower efficacy than expected will also mean a loss of power.

There was no mention of sample size adjustment for non-participation or efficacy during the design of CRTs of azithromycin MDA interventions for trachoma control published between 2001-2018[17,18,20,23,34,35,37-41,45,49,54-62]. Of note is that given that non-participation and efficacy are of greater concern in infected rather than uninfected individuals at the time the MDA is taking place, there do not appear to be any published data of the extent of non-participation amongst infected individuals following surveys for *C. trachomatis* infection. Thus, little is known about a plausible prevalence of non-participation by infection status in relation to overall coverage.

In this study, drawing on the example of a trachoma-specific context of evaluation of MDA interventions in a CRT, simulation studies were used to

- highlight the loss of power for ITT analysis in a CRT in the presence of non-participation during MDA interventions
- highlight the distinction between i) non-participation occurring amongst both infected and uninfected individuals as overall non-participation (or, overall coverage) and ii) non-participation amongst infected individuals in particular
- demonstrate that sample size adjustment in terms of number of clusters could be either over-estimated or under-estimated dependent on whether non-participation assumptions are based on coverage or non-participation amongst infected individuals specifically.
- demonstrate that efficacy assumptions can also be incorporated during trial design

4.2. Simulation strategy

4.2.1. Simulated Trial Design

A two-arm superiority CRT with equal allocation of clusters to receive either one round of MDA or no MDA. The primary outcome was *C. trachomatis* infection (binary) at follow-up, also measured at baseline. All individuals within clusters were assumed eligible to receive treatment. The number of individuals per cluster was fixed at 100, based on published trials of MDA interventions for trachoma control[35,38,39].

4.2.2. Overview

There were three parts to the generation and analysis of simulated data.

Firstly, under assumptions of **no** non-participation and 100% treatment efficacy in treated individuals for a reference trial scenario, the minimum number of clusters was established to provide at least 80% power to detect an ITT effect, based on comparing prevalence of infection at follow-up between arms. Fixed parameter assumptions were applied for baseline prevalence of infection in all clusters, a two-sided alpha of 5% and 100 individuals per cluster. This minimum number of clusters was used for all subsequent simulations that introduced non-participation into trials with the same assumed level of efficacy.

Second, non-participation was introduced at an individual level within MDA clusters into simulated trial datasets, based on typical coverage assumptions. In this part, there was no correlation or dependence between non-participation and infection status at baseline, which is the same as assuming that non-participation occurs independently of infection status. The implication of this is that the prevalence of non-participation (equivalent to 100% - percent coverage) in a cluster is the same (within sampling error) as the prevalence of non-participation amongst infected individuals in that cluster, and the same as the prevalence of non-participation

amongst uninfected individuals in that cluster. Results of this part illustrate the loss of power in the presence of non-participation.

The third part allows for positive or negative associations between non-participation and infection status within clusters. Results of this part illustrate when there could be over- or under-adjustment to the number of clusters required in the presence of unequal distribution of non-participation by infection status, whilst also allowing for less than 100% efficacy during trial design.

Simulations for parts two and three were repeated assuming efficacy in treated individuals of 85%, 75%, 65%, based on published estimates[116].

4.2.3. Generating two-level data

A approach to simulating two-level CRT data was taken, based on assuming a distribution for cluster level data and then expanding the data to have two levels (cluster and individual).

The rationale for this approach was based on published analysis of data from 75 clusters in Tanzania with a mean prevalence of *C. trachomatis* infection of 15%[117] and other simulation studies of trachoma prevalence survey data[118]. These publications suggest it is reasonable to assume that cluster level prevalence data for trachoma are exponentially distributed, where the mean value for cluster level prevalence is up to 15%. A mean cluster level prevalence of infection of 15% was selected here as a plausible level of infection in areas with high prevalence of signs of disease and therefore eligible for MDA[35]. Therefore, cluster level prevalence of infection at baseline was generated as an exponentially distributed variable, truncated at 0 and 100 with a mean of 15%.

In parts two and three that incorporated non-participation, cluster level data were generated based on correlated exponential variables; one for prevalence of infection at baseline and one for prevalence of overall non-participation in that cluster, with a fixed correlation parameter of 0.5. A copula approach was used[119-121] that transformed normal random variables into exponential random variables and again, values were truncated at 0 and 100. Non-participation data from the PRET CRT in The Gambia suggests an exponential distributional assumption for the cluster level prevalence of non-participation is reasonable (chapter 2.8.3).

4.2.4. Determining number of clusters required (part one)

The mean cluster level prevalence of infection at baseline remained at 15% in all simulation studies.

In part one, a consistent process was applied to establish the minimum number of clusters required under each efficacy assumption. That is, the minimum number of clusters required for at least 80% power was determined under an assumption for 100% efficacy first of all. The steps of this process, to obtain more precise numbers of clusters than applying a sample size formula, are detailed in **Box 1**. This minimum number of clusters was then used in simulation study parts two and three that incorporated non-participation under an assumption of 100%. Similarly, for efficacy of 85%, 75%, 65% (like four sub-studies within parts two and three of the simulation study).

Individual level data were created for infection status at follow-up based on infection status at baseline, arm (MDA vs no MDA) and efficacy. For example, under an assumption of 85% for example, only 85% of infected individuals at baseline in the MDA arm were classified as uninfected at follow-up, the remaining 15% would still be infected at follow-up due to lack of efficacy. This is applied probabilistically with rounding to the nearest integer.

4.2.5. Estimating power with non-participation independent of infection status (equal distribution of non-participation amongst infected and uninfected individuals: part two)

In part two, a binary variable was generated for individual-level non-participation status during MDA (0=participator, 1=non-participator). Under an assumption of independence between baseline infection status and non-participation, this variable was created by random assignment of non-participation status amongst individuals in a cluster. Thus, within sampling error, the cluster mean prevalence of non-participation amongst infected individuals equals the cluster mean prevalence of non-participation amongst infected individuals, and both equal the overall cluster mean prevalence of non-participation assigned to that cluster during the initial generation of correlated cluster level data.

In part two, step 5) of the process described in **Box 1**, non-participation affects infection status at follow up, such that

- Infected individuals who are participators in the MDA arm do not have infection at follow-up, according to the fixed efficacy assumption previously described
- Infected individuals who are non-participators in the MDA arm, and individuals in the no MDA arm, have infection at follow-up
- Uninfected individuals in both arms are uninfected at follow-up

The same other steps 6), 7) and 8) in **Box 1** were repeated to estimate power to detect an ITT effect in 1,000 trials for 100% efficacy and each fixed value of cluster mean prevalence value of non-participation from 1-20% inclusive (1,000 trials for each of 20 values of participation). The process was then repeated for efficacy of 85%, 75%, 65% (a further 3 x 20 x 1,000 trials). All simulation parameters are summarised in **Table 8**.

Power estimated from each set of 1,000 trials was plotted and tabulated against mean cluster level prevalence, with a line illustrating loss of power as non-participation increases by efficacy.

4.2.6. Examining power with non-participation dependent on infection status (equal and unequal distribution of non-participation amongst infected and uninfected individuals: part three)

Part three allowed for the percentage of non-participation amongst infected individuals within a cluster to be fixed at each value of 1-20% inclusive (**Table 8**), subject to a given overall value of non-participation in each cluster. This was repeated under an assumption of 100% efficacy and an assumption of overall mean cluster level prevalence of non-participation of 1-20% inclusive (20 x 20 x 1,000 trials). After assigning non-participation status amongst a fixed percentage of infected individuals, the remainder of non-participators in a cluster were assigned amongst the uninfected individuals at baseline.

Again, the simulations were repeated for efficacy assumptions of 85%, 75%, 65% (a further 3 x 20 x 20 x 1,000 trials).

Power was estimated from each set of 1,000 trials as before. For part 3, contour plots were used to show how power changes by overall mean cluster level prevalence of non-participation from generation of cluster level data and fixed percentage of non-participation amongst infected individuals at baseline. Contours of different colours were used to show power distinctions with cut-off values of 90%, 85%, 82.5% and 80%.

4.2.7. Extension to part three: investigating the impact of increasing the number of clusters

For an assumed level of 85% efficacy, part three simulations were extended and repeated with increased numbers of clusters, to demonstrate power for ITT analysis assuming 10% overall non-participation (90% coverage) and 20% non-participation (80% coverage), allowing for 1-20% non-participation amongst infected individuals at baseline.

4.3. Results

4.3.1. Minimum number of clusters to achieve at least 80% power, by efficacy

Each set of simulations in parts two and three required a fixed sample size based on 100 per cluster and the same number of clusters for each efficacy scenario, where the number of clusters was based on a trial with no non-participation and a minimum target for "starting power" of 80% to detect an ITT effect.

The starting power (number of clusters) with a minimum number of clusters for each value of efficacy was 89.8 (14), 84.7 (18), 86.6 (26) and 83.0 (36) for fixed assumed values of efficacy of 100%, 85%, 75% and 65%, respectively (**Table 9**). It is noted that simulated data output for the mean cluster prevalence of infection at baseline and follow-up was as expected (**Table 9**).

The smallest number of clusters was selected that provided power of at least 80% and closest to 80%. There is noticeable variability in the starting power under each efficacy assumption (**Table 9**). The simulations used exactly the same code, with automated incorporation of the efficacy assumption. To confirm that the variability observed is simply due to *discreteness of the binomial distribution*, the simulations were repeated in two ways; i) estimating starting power for all integer values of efficacy between 60-100%, for different minimum targets for starting power of 80%, 90%, 95% and ii) with 25, 50, 75, and 100 individuals per cluster to rule out cluster size as an explanatory factor.

The least oscillating variability in starting power by efficacy was observed when the starting power target was at least 95%, that is, designing a trial with the minimum number of clusters to have at least 95% power for an ITT analysis but no adjustment for non-participation (up to 2%; 95-97% starting power in simulated data, highest line in **Figure 15**). More variability was observed as the minimum target level for power decreased; for a target of 90% (middle line in **Figure 15**), starting power in simulated data ranged from 90-95% and for a target of 80% (bottom line in **Figure 15**), 80-89%. The oscillations in each line observed in **Figure 15,** arising due to discreteness of the binomial distribution, are not dissimilar to the oscillating patterns due to discreteness of the binomial distribution as explained by Agresti (1998: e.g Figure 4, 2003: e.g. Figure 1)[122,123] and Brown *et al* (2001)[124], in their detailed explorations of exact versus approximate interval estimation for binomial proportions, especially with proportions close to 0

or 1 and small denominators. There was little evidence that cluster size played a role in variability of starting power in simulated data (**Table 10**).

4.3.2. Power: Non-participation assumed to be independent of infection status (part two)

The results of these simulations demonstrate that power will soon fall and fall below a minimum target level of 80% in the presence of non-participation, where non-participation has been ignored when determining the required number of clusters. The variability in starting power by efficacy here serves a useful purpose of highlighting that, if a CRT design ignoring non-participation results in only just enough power to meet this minimum level of 80%, even very small amounts of non-participation in infected individuals will mean a CRT is underpowered for an ITT analysis (**Table 11, Figure 16**). Also, that if a CRT does not include any sample size adjustment for non-participation in infected individuals but is well powered and only very small amounts of non-participation in infected individuals occur, designing a trial to have 90% power rather than 80% may allow for negligible chance non-participation in infected individuals, however, this is not guaranteed.

For example,

- the results for 85% efficacy where starting power was 84.7% show just 2% non-participation overall (and in infected individuals), on average, would mean the ITT analysis is underpowered.
- the results for 100% efficacy where starting power was 89.1% show that there might still be at least 80% power for an ITT analysis if there is very little non-participation, i.e. less than 5% non-participation, in infected individuals on average.

These results do not imply these exact percentages are allowable under these efficacy assumptions in all CRT scenarios – rather simply demonstrate that non-participation should not be ignored, as it cannot be guaranteed that a trial design that does not take non-participation into account will be able to tolerate even small amounts of non-participation in infected individuals.

The summary of simulated data in **Table 11** also shows that, on average, across each set of 1,000 trials for each simulation parameter of efficacy and non-participation that the cluster mean percentage of non-participation amongst infected and uninfected individuals is equal to within 0.3% and equal to the fixed parameter for mean cluster prevalence of non-participation to within 0.6%, as expected.

4.3.3. Power: non-participation dependent on infection status (part three)

Here, prevalence of non-participation amongst infected individuals was considered for a range of 1-20%, within an overall prevalence of non-participation at cluster level at each of 1-20% also.

In part three, where simulations allow for unequal (or heterogeneous) non-participation by infection status, the average simulated data parameters differed slightly from the fixed simulation parameters, again due to discreteness of the binomial distribution. In simulated data, the percentage of non-participation amongst infected individuals is *up to 3% lower* than the simulation parameter for prevalence of non-participation in infected individuals; examples for 100% and 85% efficacy are given in **Table 12** around values of prevalence that were shown to lead to a fall in power below 80% in the previous section (**Table 11, Figure 16**). This discreteness

due to working at the extremes of the binomial distribution and is explained in more detail at the end of this section.

Contour plots allow illustration of a shift in power via a change in contour colour as non-participation in infected individuals increases along the x-axis, according to overall non-participation shown on the y-axis, with red indicating a shift to power less than 80%, for each assumed value of efficacy (**Figure 17**, **Figure 18**). The purpose of the contour plots is to highlight the distinction between non-participation amongst infected individuals, amongst uninfected individuals and amongst all individuals (overall non-participation).

The implication of the discreteness described above is that the contour shift to red (<80% power) is observed *up to 3% points higher along the x-axis* than expected according to the simulation parameters. The overall conclusions of the simulation studies are unaffected by this discreteness.

For the example of 100% efficacy, in the previous section, starting power was relatively high at 89.1% and power dropped below 80% when prevalence of non-participation was 6%, overall and amongst infected individuals, assuming independence and so equal distribution by infection status (**Table 11, Figure 16**). In the contour plot (**Figure 17**), this value of 6% amongst infected individuals is shown as a grey vertical dashed line. Without the observed discreteness in simulated data, the shift to a red contour would begin around 6% on the x-axis. In this plot, this shift change begins at 9% on the x-axis, 3% higher than 6%.

The solid grey line shows where the percentage of non-participation amongst infected individuals was expected to be equal to overall non-participation in the simulated data according to simulation parameters. Above and below this line, the distribution is unequal; above the line, the non-participation amongst infected individuals is less than non-participation in uninfected individuals; below it, greater. So, a contour is red when the prevalence of non-participation is high enough amongst infected indiviudals for notable loss of power. Where the contour is green, yellow or orange, there maybe up to 20% non-participation overall but not enough non-participation amongst infected individuals to lose power for primary analysis.

The dark blue line represents when the percentage of non-participation in infected individuals was first at least as high as the percentage of non-participation in uninfected individuals in the *simulated data*. Without the observed discreteness, this dark blue line would be observed to lie approximately on top of the solid grey line. The dark blue solid circle represents the simulation parameter required to achieve the mean prevalence of non-participation amongst infected individuals in simulated data that matches the value of the placement of the vertical dashed line (at 9% rather than the expected 6% for 100% efficacy, in **Figure 17, Table 12**).

The light blue line represents when non-participation amongst infected individuals in simulated data is within 0.5% of the simulation parameter. The light blue solid circle represents the simulation parameter required to achieve the mean prevalence of non-participation amongst infected individuals in simulated data at the value of the placement of the vertical dashed line that is within 0.5% of the simulation parameter (at 8% rather than 6% for 100% efficacy in **Figure 17, Table 12**).

For 85% efficacy, with a starting power closer of 85%, a shift to a red contour expected in a vertical pattern around 2% on the x-axis is observed at around 4-5% in the contour plot, also highlighted by the two blue dots (**Figure 18**, **Table 12**). For 75% efficacy, a shift expected at 5% is seen around 5-7%. For 65% efficacy, with the starting power closest to 80% at 83%, the expected shift at 3% does begin around 3%; here only small amounts of non-participation will lead to a reduction in power below 80% if there was only just over 80% power to start with.

The small difference between simulated data and fixed simulation parameters for prevalence of non-participation in infected individuals is explained by both low prevalence of infection at baseline and low prevalence of non-participation (discreteness at the extreme ends of the binomial distribution). The prevalence of infection in all clusters is mostly very low, given that the mean is 15% of an exponential distribution (in one example set of 1,000 trials, on average, 60% of clusters had a prevalence of infection less than 15%). The prevalence of infection out of 100 per cluster forms the denominator for prevalence of non-participation amongst infected individuals. For example, a prevalence of infection of 13% in a cluster is a denominator of 13 infected individuals; 10% non-participation amongst infected individuals is a non-integer value (13x0.1=1.3). Given that an integer value (either 1/13=7.6%, 2/13=15.4%) is needed to assign each individual a status for baseline infection and non-participation, in order to then assign and analyse follow-up infection prevalence, some rounding is required when assigning a number of infected individuals as non-participators. The closest approximation to the fixed simulation parameter for prevalence of non-participation amongst infected individuals in simulated data was obtained when rounding non-integer values up to the nearest integer for cluster level prevalence data (generated as non-integer values as part of the copula approach of generating correlated exponential random variables after drawing correlated bivariate normal random variables) and round to the nearest integer for number of non-participators out of the number of infected individuals.

Simulations were conducted of larger cluster sizes of 200-1000, in increments of 100, for an example scenario of 100% efficacy, 10% non-participation overall and 10% non-participation amongst infected individuals. In part two, the maximum difference between the simulation parameter of 10% non-participation and the prevalence of non-participation amongst infected individuals in simulated data was 0.6%. The smallest sample size of individuals per cluster found to provide a maximum difference of less than 1% in part three where the differences of up to 3% occurred, was 900. Thus, to avoid the discreteness observed in simulated two-level data with a rare binary exposure conditional on another rare binary exposure in this study, extreme cluster sizes are required.

4.3.4. Retention of power with equal and unequal distribution of non-participation amongst infected and uninfected individuals

With 85% efficacy as an example, selected because power was only just above the desirable minimum of 80% in simulated trials with no non-participation, it is easier to highlight how power could be preserved in the presence of non-participation with an increase in the number of clusters.

If expected coverage during MDA was 90% on average, investigators may choose to account for 10% non-participation. This would allow for up to 10% non-participation in infected individuals and in this example, an increase from 18 to 24 clusters retained power for ITT analysis for up to

10% non-participation in infected individuals (**Figure 19**, **middle panel**). This can be seen as a vertical pattern where the area of the contour plot becomes red around 10-12% on the x-axis.

If 90% coverage was an over-estimate of the percentage of non-participation amongst infected individuals and this percentage was say only 5%, the addition of 6 clusters would be an over-adjustment and potentially be a waste of resources. This can be seen in the plot for values of less than 10% on x-axis.

If 90% coverage was in fact an under-estimate of the percentage of non-participation amongst infected individuals and this percentage was say 15%, the addition of 6 clusters would not be enough to retain power for ITT analysis. This can be seen in the plot for values of more than 10% on the x-axis, ignoring the extreme scenario in the bottom right hand corner where overall non-participation is less than 3%. Where overall non-participation is this low, up to 20% non-participation in infected individuals makes little sense in reality; results are only presented to complete the 20x20 grid shown in the contour plots.

The minimum target coverage for azithromycin MDA for trachoma control is 80%. Allowing for 20% overall non-participation corresponding to 20% non-participation in infected individuals required an increase from 18 (no adjustment for non-participation) to 36 clusters (**Figure 19, bottom panel**), doubling the number of clusters required for at least 80% power. Should the distribution of non-participation be such that much less non-participation occurs in infected individuals in reality, this would be a huge over-estimate of the number of clusters required.

4.4. Discussion

The results of these simulation studies emphasise the importance of accounting for both non-participation and efficacy of an intervention in those who receive it, during the design of a CRT. Not doing so means there is a risk that the trial will be underpowered for primary analyses. These findings are applicable to evaluation of MDA interventions for neglected tropical diseases, other diseases such as malaria[125] and CRTs of vaccine interventions where there is non-participation within clusters randomised to receive vaccination[126]. That is, to CRTs of interventions where there could be individual level non-participation (or non-adherence) to interventions within clusters or where the intervention may not have the direct benefit intended in all of those who receive it.

There is an important distinction to be made during design, between expected levels of coverage in all individuals and expected levels of coverage amongst *infected* individuals. If based on coverage of 80%, an adjustment is made for 20% non-participation, the increase in number of clusters will be an *over-adjustment* if the % of infected individuals who do not take treatment is in fact lower than overall non-participation of 20%. This larger sample size would be unnecessary and may not be affordable or would be a waste of resources.

If the mean cluster percentage of non-participation amongst all individuals (100% - percentage coverage) is due to more non-participation amongst infected than uninfected individuals, an adjustment based on expected coverage could lead to an *underestimation* of the number of clusters required and an underpowered trial, again wasteful.

In the absence of any context-specific information, using coverage in all individuals could be a conservative adjustment but investigators will need to draw on all expertise and data available to make a reasonable judgement about whether this would be conservative or not.

Assuming 90% power rather than 80% power during design may go some way to retain power for ITT analyses, should there only be small amounts of non-participation amongst infected individuals, but this cannot be wholly relied upon.

Increases in the number of individuals per cluster could also be considered, but this may not be practical if there are limited numbers of eligible individuals available within a cluster for enrolment and measurement of the outcomes, for example, the number of children aged 0-5 years old in a remote village.

Whilst there is some focus on investigating risk factors for non-participation during MDA rounds in the field of each of the five NTDs with MDA control[112,127,128], to identify those more likely not to receive treatment, there is extremely limited information about whether non-participation is occurring in individuals with infection prior to MDA specifically. It is possible that risk factors for infection correlate with risk factors for non-participation and thus non-participation in infected individuals is a risk to power of any CRTs of MDA interventions in these settings.

There do not appear to be any published reports of the extent of non-participation in MDA rounds amongst infected individuals following a baseline survey for *C. trachomatis* infection, so little is known about the prevalence of non-participation amongst those infected prior to an MDA round. One paper based on data from Tanzania reported no evidence of an association between infection status at baseline and participation (p=0.09) but did not report the percentage of NP amongst infected and uninfected children[50]. It is possible this is also the case for other diseases since analyses of non-participation can be focussed instead of potentially causal characteristics of non-participators[112,129]. In the absence of a statistical association, non-participation amongst infected individuals could still be at a level that could lead to a non-negligible loss of power in ITT analyses.

These results do not provide a one-size-fits-all approach for adjusting for non-participation during MDA in the design stage of a CRT, nor could they as each CRT will be context-specific in terms of the disease and interventions applied in each arm, requiring a different set of parameter assumptions in sample size calculations. A simple superiority design with a single round of MDA in one arm versus no MDA in the other arm was selected for demonstration here and a pragmatic approach to generating two-level data was used in simulations, based on published studies of the likely distribution of cluster level data for trachoma outcomes. It would also be wise to consider the implications of non-participation and efficacy for other trial designs such as those with an active control arm, multiple arms or a non-inferiority design.

The approach to generation of two-level data used here, based on distributional assumptions for cluster level data was simpler than an alternative approach that could have been based on definition of specific quantities of correlation or heterogeneity in hierarchical data. Incorporation of measures such as intra-cluster correlation (ICC) or coefficients of variation for both infection and non-participation would add unnecessary complexity by introducing more potentially variable simulation parameters. A truncated exponential distribution was used based

on a trachoma specific context[130,131]. Higher prevalence contexts or other diseases may require a different distributional assumption.

The sampling variation in starting power and small differences observed between prevalence of non-participation amongst infected individuals in simulated data and the fixed simulation parameters were inconvenient in that they added a level of complexity to interpretation. The contour plots were intended to be a simple visual tool illustrating the impact on power of heterogenous non-participation by infection status. The conclusions of this study hold clear regardless of differences of up to 3% between simulated data and simulation parameters for prevalence of non-participation in infected individuals. Interestingly, these aspects of the results highlight something not yet apparent in the literature relating to simulation of CRT data with a binary outcome and a binary exposure variable conditional on another binary exposure. Any approach to generate two-level data with dependent binary exposures would very likely also be subject to discreteness of the binomial distribution highlighted in detail in the literature[122-124], especially working with the parameters at the extremes of the binomial distribution. It can be noted that, with respect to a simulation-based approach to CRT sample size investigations, only the simpler approach in part two is required as non-participation in infected individuals will lead to a loss of power. The sampling variability in starting power when trying to achieve a level of just 80% did serve a purpose to illustrate the impact (or not) of very small amounts of non-participation.

There are guidelines on selection of number of replications for a simulation study[132], intended where the study purpose is to obtain a specified level of precision of some measures of effect. As not applicable here, 1,000 replications were deemed adequate to demonstrate the impact of non-participation on power. In a test for one simulation study, an increase to 100,000 replications did not improve granularity and smoothness of shifts in contour plots, or reduce the impact of discreteness observed.

Other alternative analysis methods may be more statistically appropriate than an unpaired *t-test* to compare cluster level prevalence data, such as regression with adjustment for baseline covariates[133], particularly in light of many clusters with low or zero prevalence. However, a simple *t-test* was an adequate method to demonstrate the key message that power for analysis is lost if non-participation and efficacy are ignored during design.

Reinfection post-MDA or post-baseline in the no MDA arm was not specifically addressed as a parameter in the simulation studies, either as transmission post MDA within treated clusters or from neighbouring untreated communities. Reinfection post MDA could can be considered analogous to additional non-participation when estimating the cluster mean prevalence of infection at follow-up during trial design. Substantial transmission in the no MDA arm could mean that design assumptions of baseline prevalence may not hold.

4.5. Conclusions

It is unlikely that all individuals will receive treatment during MDA, even if those delivering treatment actively seek out all those eligible for treatment. Any non-participators will be amongst those infected at baseline, the status of which is not always known at the time. It is also possible that treatment will not clear infection in 100% of infected individuals who receive it.

Both non-participation in infected individuals and imperfect efficacy of treatment should be considered during the design stage of a CRT, to ensure adequate sample size and power for primary comparisons between arms at follow-up.

In the context of MDA control for infectious diseases where the intention is to clear infection, it is important to think beyond likely coverage amongst all individuals in a cluster (ignoring infection status) and include an adjustment for the likely extent of non-participation amongst those infected at baseline. Relying on overall coverage could lead to an over- or under-adjustment in sample size.

This chapter considered design implications of correlated non-participation and outcome data on power to detect effects in CRTs. In the next chapter, an analysis of the impact of MDA on a secondary outcome in the PRET trial will include consideration of clustering of the outcome and non-participation

Box 1. Determination of the minimum number of clusters for at least 80% power in the absence of non-participation

For a fixed number of clusters;

1) Cluster level data for prevalence of infection at baseline were generated according to an exponential distribution with a mean of 15%, rounding up to the next integer for the prevalence in each cluster
2) Each cluster was randomly assigned one of two arms (no MDA or MDA)
3) The dataset was then expanded to have 100 individuals per cluster
4) A binary variable was generated for infection status at baseline (0=not infected, 1=infected) so that the number of infected individuals in each cluster matched the cluster level prevalence generated in the first step
5) A binary variable was generated for infection status at follow-up (0=not infected, 1=infected) such that
    i) Infected individuals in the MDA arm do not have infection at follow-up, according to the fixed efficacy assumption (for example, if efficacy is 100% then all infected individuals are uninfected at follow-up, if efficacy is 85%, then probabilistically (with rounding to the nearest integer) 15% of infected individuals remain infected at follow-up, in each cluster)
    ii) Infected individuals in the no MDA arm have infection at follow-up
    iii) Uninfected individuals in both arms are uninfected at follow-up
6) Simulations generated 1,000 such trials and an unpaired *t-test* of cluster level prevalence at follow-up between arms was conducted in an ITT analysis, where clusters were analysed according to MDA or no MDA allocation.
7) The power afforded by this fixed number of clusters was calculated as the percentage of trials in which the *t-test* p-value was ≤ 0.05
8) This process was repeated with different fixed numbers of clusters to determine the minimum number of clusters achieving at least 80% power
9) This process was repeated for efficacy assumptions of 100%, 85%, 75% and 65%.

Table 8. Simulation parameters

Key parameter	Method of estimation and initial value	Sensitivity analysis
Baseline mean prevalence of *C. trachomatis* infection: cluster level	Exponential distribution for mean cluster level prevalence: 15%, truncated so that all values were between 0-100.	Not applicable
Non-participation: cluster level	Exponential distribution assumed for mean cluster level prevalence value of 1%, truncated so that all values were between 0-100, correlated with cluster level prevalence of baseline infection	Truncated exponential distribution assumed for mean cluster level prevalence values from 2-20% inclusive
Non-participation: individual level	Independence between non-participation and baseline infection status; equal distribution of non-participation in infected and uninfected individuals, achieved through random assignment of individuals within a cluster to non-participators, with a cap corresponding to the number of non-participators in that cluster	Equal and unequal (heterogeneous) non-participation status amongst infected individuals at baseline, for values of the percentage of infected individuals who were non-participators was equal to each of 1% to 20% inclusive. Assign remaining non-participators amongst individuals not infected at baseline in a cluster until the number of non-participators matches each cluster level prevalence of overall non-participation.
Efficacy in individuals (effective treatment of infection in those who take it)	100%	85%, 75%, 65%

Table 9. Minimum number of clusters for at least 80% power for trials with <u>no</u> non-participation and 100 individuals per cluster

				Simulated data			
Cluster mean baseline infection prevalence (%)	Assumed value of efficacy (%)	Expected cluster mean prevalence at follow-up in MDA arm*	Minimum number of clusters (both arms) for at least 80% power from 1,000 simulations	ITT power from an unpaired *t-test*	Cluster mean baseline infection prevalence: both arms	Cluster mean follow-up infection prevalence: no MDA arm	Cluster mean follow-up infection prevalence: MDA arm
15	100	0	14	89.1	15.6	15.6	0.0
15	85	2.25	18	84.7	15.6	15.7	2.3
15	75	3.75	26	86.6	15.5	15.4	3.7
15	65	5.25	36	83.0	15.4	15.4	5.4

* $\pi_1 = \pi_0\left[1-\left(\frac{EFF\%}{100}\right)\right]$

ITT = Intention-to-treat analysis comparing prevalence between arms in a cluster level analysis

Table 10. Power to detect an ITT effect, by efficacy, cluster size and number of clusters in a CRT with no non-participation

Baseline infection mean cluster prevalence = 15%		Power from 1000 simulations		
Efficacy	Cluster size (equal)	Number of clusters and power for unpaired *t-test* of cluster level data to detect ITT effect		
			14	**12**
100	25		86.4	75.9
	50		88.2	77.4
	75		86.7	79.3
	100		**87.5**	76.9
		20	**18**	**16**
85	25	85.1	79.2	71.6
	50	84.3	78.1	74.8
	75	84.8	80.2	69.0
	100	**83.6**	78.9	70.4
			26	**24**
75	25		82.8	79.8
	50		81.0	77.4
	75		82.8	78.6
	100		**84.0**	79.9
		40	**38**	**36**
65	25	87.0	80.4	76.9
	50	89.1	80.0	77.8
	75	86.0	80.9	79.8
	100	89.4	**82.9**	76.9

Each experiment based on 1,000 simulated trial datasets. ITT = intention-to-treat

Bold type highlights minimum number of clusters required to provide at least 80% power

Results from primary simulation studies with 100 per cluster; 100% efficacy: 14 clusters 89.1% power; 85% efficacy: 18 clusters 83.0% power; 75% efficacy: 26 clusters 86.6% power; 65% efficacy: 36 clusters 82.1% power

No evidence that variability in power of at least 80% by efficacy (**82.9-87.5%**) is attributable to cluster size.

Table 11. Impact of non-participation in infected individuals on power by baseline prevalence and efficacy: non-participation independent of baseline infection status (equal distribution)

Fixed parameters			Simulated Data		
Risk %	Efficacy	Mean cluster overall non-participation %	Cluster mean % NP amongst infected individuals	Cluster mean % NP amongst uninfected individuals	Power: ITT
15	100	0	-	-	89.1
15	100	1	1.5	1.6	89.4
15	100	2	2.5	2.5	89.1
15	100	3	3.6	3.5	87.8
15	100	4	4.4	4.4	82.2
15	100	5	5.8	5.6	82.8
15	**100**	**6**	**6.3**	**6.5**	**76.7**
15	85	0	-	-	84.7
15	85	1	1.6	1.6	83.8
15	**85**	**2**	**2.5**	**2.5**	**79.7**
15	85	3	3.6	3.6	79.0
15	75	0	-	-	86.6
15	75	1	1.6	1.6	88.7
15	75	2	2.5	2.5	85.6
15	75	3	3.5	3.5	84.1
15	75	4	4.5	4.5	82.1
15	**75**	**5**	**5.6**	**5.5**	**79.4**
15	65	0	-	-	83.0
15	65	1	1.5	1.6	82.3
15	65	2	2.5	2.5	80.4
15	**65**	**3**	**3.6**	**3.5**	**79.4**

NP = non-participation. Data are displayed in full in **Figure 1.**

Data show when power for ITT analyses first falls below the pre-specified minimum level of 80%, given the cluster mean percentage of non-participation amongst infected individuals (**bold type**).

Results show that in the simulated data, the cluster mean prevalence of non-participation amongst infected and uninfected individuals is approximately equal on average and equal to the fixed parameter.

Table 12. Examples of equal and unequal non-participation by infection status and power by efficacy: 100% and 85% efficacy

Simulation (fixed) parameters			Simulated data		
Efficacy	% mean overall non-participation	% non-participation amongst infected	mean % non-participation amongst infected	Mean % non-participation amongst uninfected	Power: ITT
100	6	4	2.3	8.2	87.3
100	6	5	3.4	7.8	88.2
100	*6*	*6*	*4.1*	7.7	85.2
100	6	7	4.8	7.3	85.8
100	**6**	**8**	**5.8**	7.1	82.8
100	**6**	**9**	**7.1**	6.9	80.2
100	6	10	8.6	6.7	81.3
100	6	11	**9.0**	6.4	78.3
100	*10*	*10*	*8.8*	12.1	76.8
85	2	2	0.6	3.1	83.6
85	2	3	1.3	2.9	83.4
85	**2**	**4**	**2.2**	2.6	80.0
85	**2**	**5**	**3.3**	2.4	81.0
85	2	6	**4.0**	2.2	79.5

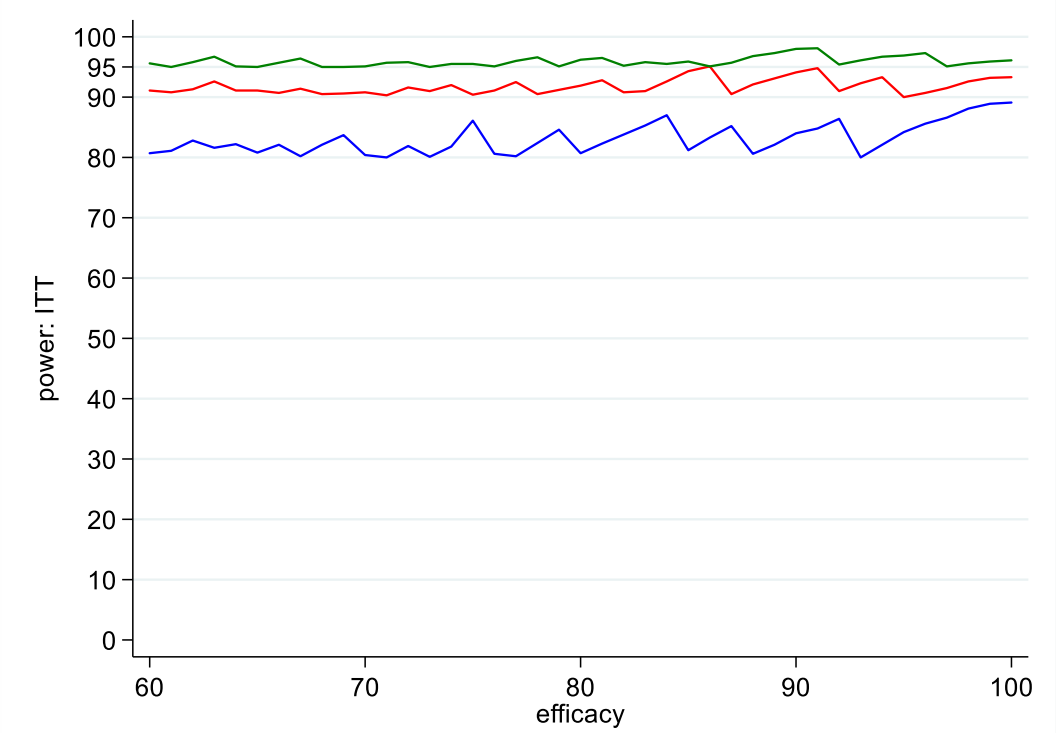
In simulated data, prevalence of non-participation amongst infected individuals is 4.1% for simulation parameters of 6% overall and amongst infected; 8.8% for simulation parameters of 10%

In simulated data, prevalence on non-participation amongst infected individuals is within 0.5% (5.8% versus 6%) for simulation parameters of 6% overall non-participation and 8% non-participation amongst infected for 100% efficacy (**Figure 3)**. For 85% efficacy, within 0.5% (2.2%) for fixed parameters of 2% and 4% (**Figure 4**).

In simulated data, prevalence of non-participation if infected is first at least the value of prevalence if uninfected for simulation parameters of 6% overall and 9% amongst infected for 100% efficacy (3% higher than expected, **Figure 3**). For 85% efficacy, for parameters of 2% and 5% (again 3% higher than expected, **Figure 4**)

Bold type: in simulated data, power falls below 80% for 9% non-participation amongst infected, rather than expected value of 6% for 100% efficacy (**Figure 2**) and for 4% rather than 2% for 85% efficacy (**Figure 2**)

Figure 15. Starting power by efficacy in simulated data, with the minimum number of clusters required to detect an ITT effect in a CRT with no non-participation with three levels of minimum starting power (95%, 90%, 80%)

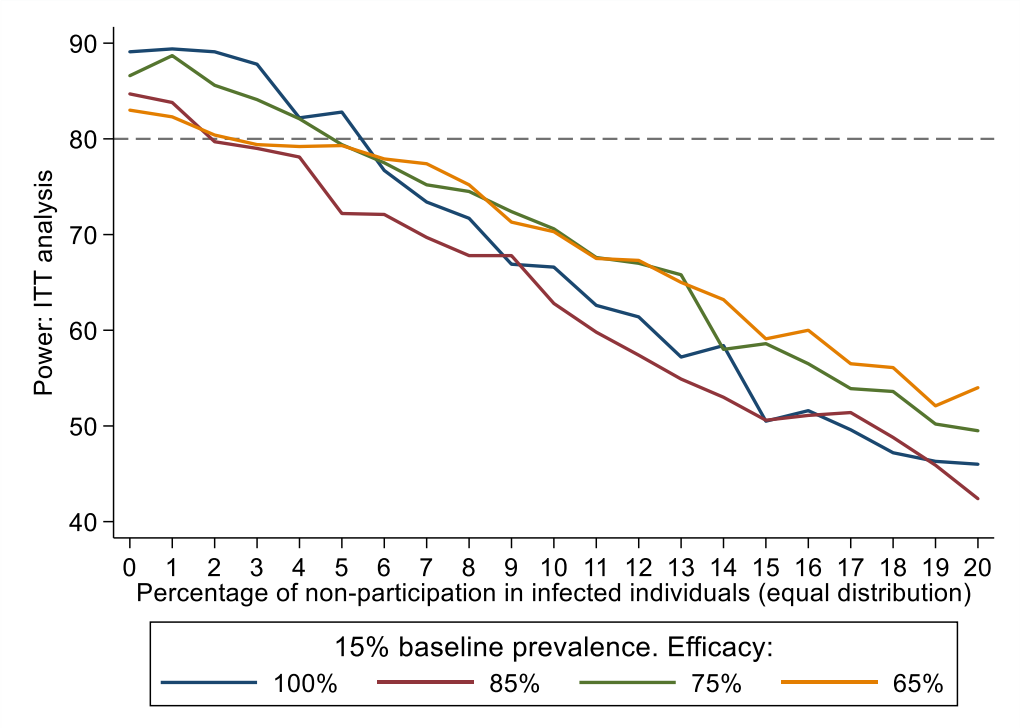


Green: starting power in simulated data for a minimum number of clusters to achieve at least 95% power

Red: starting power in simulated data for a minimum number of clusters to achieve at least 90% power

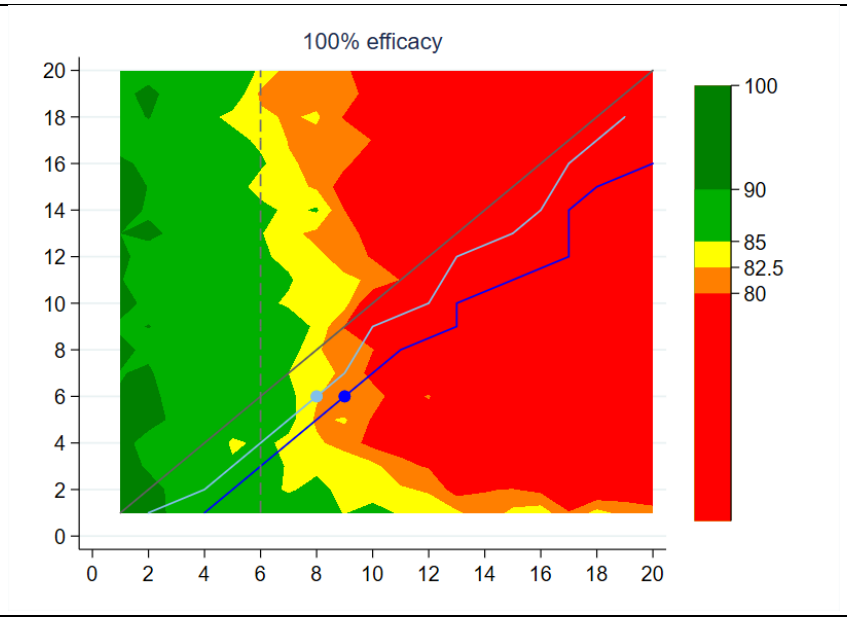
Blue: starting power in simulated data for a minimum number of clusters to achieve at least 80% power

Figure 16. Impact of non-participation on power by non-participation with equal distribution by baseline infection status, by efficacy



Simulated trials include a fixed number of clusters (**Table 2**) based on power of at least 80% with no sample size adjustment for non-participation during design; starting power with no non-participation is shown at zero on the horizontal axis.

Figure 17. Impact of non-participation on power by non-participation with equal and unequal distribution by baseline infection status: 100% efficacy



The number of clusters remains fixed as the minimum number of clusters required for at least 80% power with no non-participation (from part 1). Cluster mean overall non-participation in all individuals is shown on the y-axis based (simulation parameter). Cluster mean percentage of non-participation in infected individuals is shown on the x-axis (simulation parameter).

Grey dashed line: where power drops below 80% for non-participation amongst infected individuals when there is equal distribution as in **Figure 2** (at 6%). Without the small differences observed between the simulated data and simulated parameters, a shift to a red contour would occur to the right of this line, instead, this shift to red is observed around 9% indicated by the dark blue solid circle (**Table 5**).

Grey solid line: indicates where simulation parameters specify equal distribution of non-participation between infected and uninfected individuals; with unequal distribution above and below this line. Above this line, the prevalence of non-participation amongst infected individuals is less than the prevalence of non-participation in uninfected individuals; below it, greater.

Dark blue line: without the small differences in simulated data and simulation parameters (discreteness), this would lie approximately on top of the grey solid line. This line shows where non-participation amongst infected individuals is at least as high as non-participation in uninfected individuals in simulated data; the solid circle represents the simulation parameter required to achieve the prevalence of non-participation amongst infected individuals in simulated data that matches the value of the placement of the vertical dashed line

Lighter blue line: indicates where non-participation amongst infected individuals in simulated data is within 0.5% of the fixed simulation parameter value; the solid circle represents the simulation parameter required for prevalence of non-participation amongst infected individuals in simulated data within 0.5% of the simulation parameter of the vertical dashed line.

There is still power for an ITT analysis for less than 9% non-participation in infected individuals (corresponding to 6% under the simulation parameters, **Figure 2**). Along the bottom of the plot, (overall non-participation of 1%), results suggest power for up to 20% non-participation in infected individuals; an anomaly with little meaning as there is so little non-participation overall, non-participation in infected individuals cannot actually reach the percentages indicated on the x-axis.

Figure 18. Impact of non-participation on power by non-participation with equal and unequal distribution by baseline infection status: 85%, 75% and 65% efficacy

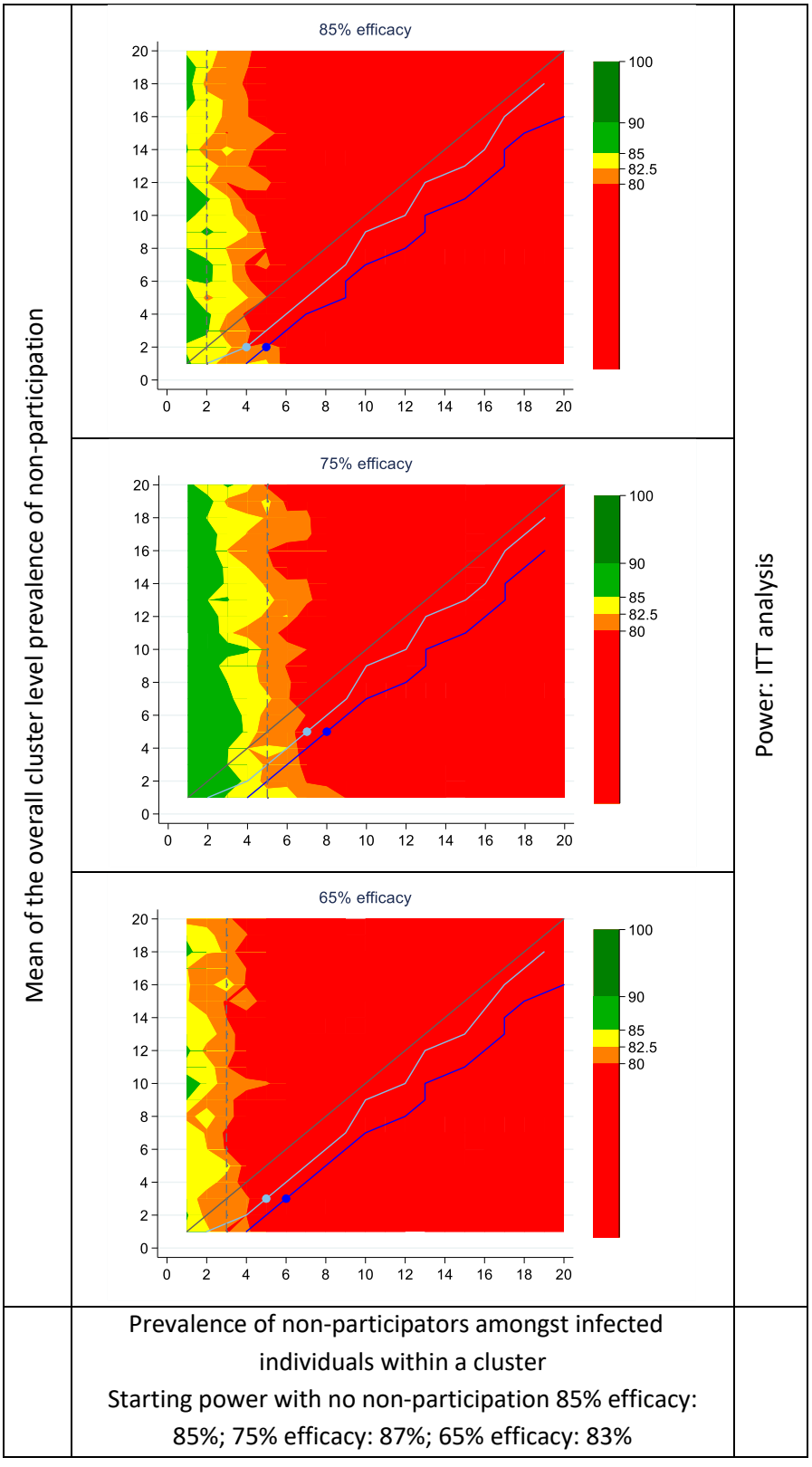
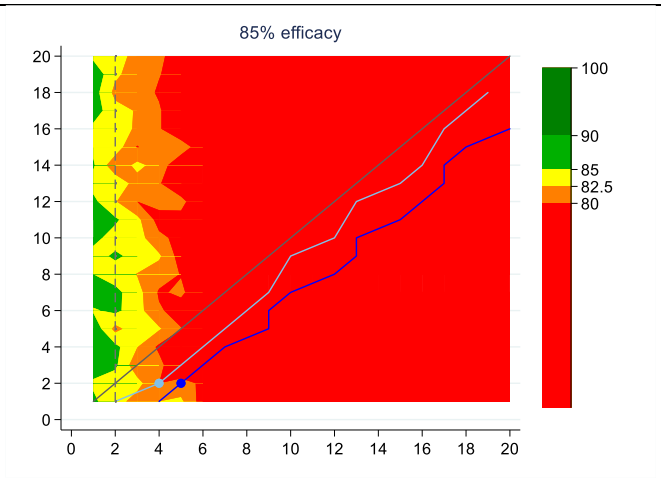
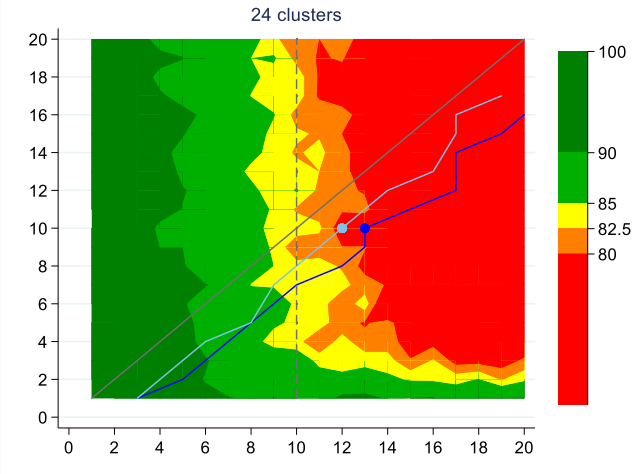


Figure 19. Power for ITT analysis with increased numbers of clusters: 85% efficacy

No adjustment for non-participation (18 clusters)

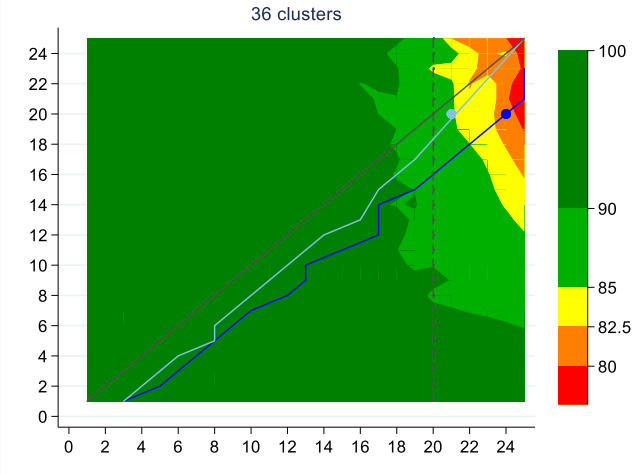


Increase in trial size to 24 clusters (adjust for 90% coverage, 10% non-participation amongst infected)



Under-adjustment if more non-participation occurs in more infected individuals than uninfected individuals; over-adjustment if fewer than 10% of infected individuals do not receive treatment

Increase in trial size to 36 clusters (adjust for 80% coverage, 20% non-participation amongst infected)



Large over-adjustment ff non-participation occurs in a relatively small percentage of infected individuals

5. Evaluating the impact of azithromycin on mortality in the PRET trial in The Gambia

5.1. Introduction

Azithromycin, a broad-spectrum antibiotic, has been shown to be efficacious in treating *Chlamydia trachomatis* infection. During mass drug administration (MDA) rounds with azithromycin, a single dose of azithromycin is offered to all members of communities in endemic areas, with the exception of pregnant women and children under 6 months, according to World Health Organisation (WHO) guidelines[98].

The gold standard approach to evaluate MDA strategies, where MDA interventions are delivered to entire communities is a cluster randomised trial (CRT). A CRT of an MDA intervention could be considered *pragmatic* in that it will measure the *population-level effectiveness* of MDA[43], quantified via an intention-to-treat (ITT) analysis. In the context of the PRET trial for trachoma, *effectiveness* will primarily be due to the *efficacy* of the therapeutic treatment as a *direct* effect on the outcome in those who take it (i.e. clearing infection). However, e*ffectiveness* will also be a combination of any *indirect* effects of the intervention, such as potential herd effects in untreated individuals in treated communities, the ability of the delivery method to reach the target population and acceptability of the intervention to community members.

Both *effectiveness* and *efficacy* are of interest and value to public health policy makers[116], but the ITT analysis may not provide a reliable indication of *efficacy*. Estimating *effectiveness* and *efficacy* without bias, in the presence of non-compliance, is challenging.

Non-compliance will reduce power and introduce bias in an ITT analysis, especially if there is clustering of non-compliance[51,79,134]. *Effectiveness* will be underestimated if not all of the target population for treatment receive it because any beneficial effect in a treated arm will be diluted, and especially so if non-compliance is associated with risk of poorer health. Common analysis approaches to estimate *efficacy* in those who receive treatment are per-protocol (PP), where participants who do not comply are excluded from a between-arm analysis and as-treated (AT), where data are analysed according to actual treatment receipt. Neither of these approaches will provide an unbiased effect of treatment in those who take it because the randomisation is broken; groups of individuals being compared are no longer subject to the properties of the randomisation that ensure balance with respect to confounders of the treatment effect on the outcome. Selection bias will occur if treatment receipt or missing treatment status is associated with better or worse outcomes, meaning that those who receive treatment are not representative of individuals allocated to receive that treatment[78,135]. Within a CRT framework,

selection bias is more likely in the presence of clustering of non-compliance to treatment and clustering of the outcome[79].

An alternative approach to estimate *efficacy*, that limits the impact of selection bias, is to estimate a complier average causal effect (CACE). The CACE is the effect of treatment in compliers and is obtained by comparing the outcome in observed compliers in the treated arm of a trial, to the outcome in an assumed comparable group of would-be compliers the untreated arm[78,136]. However, methods for inference using CACE in CRTs are complex[51,79] and especially so for outcome data that do not follow a normal distribution.

Mass distribution of a broad-spectrum antibiotic on a large geographical scale to trachoma endemic communities has generated much interest in whether azithromycin MDA is beneficial for other health outcomes, such as mortality, malaria morbidity and anthropological outcomes. Whilst CRTs of MDA interventions for trachoma may provide a framework to evaluate the impact of MDA on secondary outcomes, such trials will be underpowered if there was no pre-specified level of power, to detect a pre-specified effect, on a secondary outcome during the design of the trial. In the previous chapter, simulation studies highlighted that any non-participation (non-compliance) will mean even less power.

In this chapter, ITT, PP, AT and CACE estimates will be obtained for the effect of azithromycin on mortality using data from the PRET trial in The Gambia. CACE estimates with corresponding 95% confidence intervals will be obtained using a pragmatic bootstrapping approach to account for clustering of the outcome, clustering of non-compliance and missing treatment status, which could also be clustered. Although the PRET trial in The Gambia was not designed specifically to evaluate the impact of azithromycin MDA on mortality, mortality was a pre-specified secondary endpoint and there is currently much public health interest in whether there is a potential benefit for mortality[137,138].

5.2. Methods

5.2.1. Participants and cohort definition

Although census data of individuals of all ages are available for the PRET trial communities from six-monthly census rounds for a period of three years (baseline, 6, 12, 18, 24, 30 and 36 months), a two-year period open cohort for mortality is analysed here starting from the year one (12 months) census round and ending at the 36 months census round. At baseline, all clusters received MDA but because the MDA stopping rule was met early in the trial based on prevalence of *C. trachomatis* infection[39] (chapter 2), there is a two-year follow-up period where half of the

trial clusters received annual MDA and half did not, according to a randomised allocation. To estimate *effectiveness* and *efficacy* of azithromycin MDA to all community members on mortality, this two-year period of follow-up provided a convenient opportunity to apply the methodology to estimate a CACE of *efficacy*, as the simplest application requires an intervention arm in which no interventions are given.

The two-year open cohort included individuals who were resident in the cohort in the year one census round. Time at risk was calculated from the date of year one census until death, end of the study (36 months census) or until an individual moved out of the study area. The open cohort included individuals who were resident at year one or who became residents any time after the year one census, up to and including the penultimate census round at 30 months follow-up. If the date of death was unavailable for deceased individuals, the date of death was assumed to be halfway between the time point where the individual was reported deceased and the date of the previous census or treatment round in their EA (interval censoring).

5.2.2. Outcomes

The primary outcomes of the PRET Gambia trial were follicular trachoma and *C. trachomatis* infection in children aged 0-5 years. A secondary outcome of PRET in The Gambia is analysed here; all-cause mortality in all individuals aged ≥1 year old. All-cause mortality in children aged 1-4 years is also reported, given that the impact of MDA on child mortality is of current public health interest[137,138].

5.2.3. Sample size and power

The sample size of 48 clusters with a random sample of 100 children aged 0-5 per cluster was estimated based on hypothesis testing of the primary outcomes of PRET GM which were *C. trachomatis* infection and follicular trachoma[39,57].

Census data provided a mean of 870 individuals per cluster, with a mean of 170 per cluster aged 1-4 years old. Although mortality was pre-specified as a secondary outcome, there was no pre-specified mortality reduction with pre-specified power during the design of the trial.

5.2.4. Analysis Populations

Intention-to-Treat (ITT): included all clusters and individuals comparing arms based on random allocation of clusters to MDA or no MDA arms.

Per-protocol (PP): included all clusters in both arms but excluded individuals in the MDA arm who were not treated or whose treatment status was missing.
Complier average causal effect (CACE): included all clusters allocated to MDA or no MDA with a comparison between arms of individuals observed to receive treatment and an assumed comparable group of individuals in the no MDA arm who would have taken treatment if offered it.

Actual treatment receipt (AT): included all clusters in both arms, comparing all untreated individuals in both arms to treated individuals in the MDA arm. Individuals with missing treatment status were excluded from this analysis.

5.2.5. Analysis Methods

Data were analysed using Stata version 14 Special Edition[70].

Characteristics of EAs, households and resident individuals in the study area at the start of the open-cohort (at year one follow-up of the PRET trial), were summarised by arm.

Cluster-level mortality and non-compliance data were summarised to illustrate heterogeneity in the data. Mortality rates and corresponding 95% CIs were calculated overall, by arm and by district.

ITT, PP and AT analyses were carried out using Poisson regression models to take account of variable time at risk, with a robust standard error (SE) adjustment to account for clustering and no *a priori* adjustments.

Defining compliance status for the CACE analysis

In previous chapters, treatment receipt during an MDA round has been referred to as participation. In this chapter, compliance will be used instead as terminology for treatment receipt, in place of participation, in line with published literature about CACE analyses.

During the two-year cohort period, two MDA rounds occurred in the MDA arm and individuals could have been eligible for 0, 1 or 2 treatments based on their time of entry to and exit from the cohort. Therefore, an individual's compliance status could be *complete* (received all treatments for which they were eligible, or were ineligible for all treatments), *partial* (eligible for 2 treatments and only received 1 treatment) or *none* (eligible for at least one treatment and didn't receive any treatments).

The primary compliance status variable for all individuals in the MDA arm had three categories; i) *complete*, ii) *none* and *partial* and iii) missing treatment status. This categorisation separates out individuals fully compliant with treatment in each MDA round during the two-year follow-up period, accounting for eligibility. Complete compliance is assumed if an individual receives all treatments for which they were eligible and so includes those who were ineligible for treatment. This analysis was repeated in the 1-4 year old age group. Sensitivity analyses were carried out using alternative compliance status variables (see below).

Missing treatment status is included as a separate compliance category, rather than excluding individuals with missing treatment status. Exclusion of those with missing treatment status from the analysis will introduce bias if treatment status is not missing completely at random (MCAR) or not missing at random (MAR)[135]. MCAR is where there is no association between missingness and other data in the dataset and MAR is where there is some systematic missingness, but it can be explained by another variable in the dataset. It is plausible that those with missing treatment status could be more likely to be non-participators, for example, those too sick to present for treatment or to receive a visit from the treatment team. It cannot be assumed that treatment status data are MCAR or MAR.

CACE analysis

Treatment status is observed in the MDA arm and so, for the MDA arm, it is possible to define compliance categories (as described above) and to calculate mortality rates for each compliance category, as well as overall. In the no MDA arm, only overall data are observed that provide an overall mortality rate for the no MDA arm (**Figure 21**).

The CACE rate ratio is intended to be a comparison of rates between comparable groups of compliers in each arm. As compliance status is not observed in the no MDA arm, it is not possible to directly calculate the rate in compliers in the no MDA arm. Some assumptions are required to allow estimation of the mortality rate in a comparable would-be group of compliers in the no MDA arm.

First of all, compliance status is considered to be a pre-randomisation, baseline characteristic in individuals that captures the underlying compliance behaviour of an individual. For example, whether or not someone would take a dose of azithromycin when offered it is something inherent regardless of whether they are randomized to a treatment arm, or not. In the methodological literature, this is referred to as *principle compliance* status[78,139]. In the context

of an individually randomised trial, the randomisation process should result in balanced proportions of individuals with each underlying compliance status as a baseline characteristic, in each arm (or a balanced proportion of person-time contribution in the case of a rate outcome). Under these assumptions, it is then assumed that amongst non-compliers in each arm who do not receive any treatment, the mortality rate is the same because randomisation has led to comparable groups of non-compliers. Analogous to this, it is assumed that for the group in this trial with missing treatment status, there is balance in the person-time contribution from those with unknown treatment status and the same mortality rate, in each arm.

From these assumptions of the same mortality rate and person-time contribution for non-compliers and those with missing treatment status in each arm, the person-time in years and number of deaths can be derived for these categories in the no MDA arm. Then subtracting these values from the total person-time and total number of deaths in the no MDA arm, the person-years, number of deaths and mortality rate can be obtained for an assumed comparable group of would-be compliers in the no MDA arm. Thus, a CACE RR comparing rates in compliers between arms can be calculated.

This approach, as described by Sommer & Zeger (1991) and Little *et al* (2009)[77,78], ignores any clustering of mortality rates and compliance status however, hence the need for a pragmatic analysis technique than can account for clustering in the data. A two-step bootstrapping procedure, similar to the methods of Opondo *et al* [140] was used to re-estimate the ITT effect and to estimate the CACE of azithromycin treatment in individuals on mortality. The ITT effect was re-estimated to check that the application of the bootstrapping approach worked as intended and gave similar results to the regression analysis. Random bootstrap samples were drawn with replacement within the strata of intervention arm, EA (trial cluster) and compliance status, in order to account for between-cluster (EA) variation in the outcome and compliance status. The sampling distribution from 10,000 replications was simulated to obtain the RR and corresponding 95% CI, by taking the median value as the RR and the 2.5^{th} and 97.5^{th} percentiles as the 95% CI bounds.

Sensitivity analyses for compliance status were done for all-age all-cause mortality for two reasons; 1) to consider the implications of assuming mortality rates in partial compliers were the same as in non-compliers; first by re-running the CACE analysis with partial compliance grouped with complete compliance, then with partial compliance as a separate category of compliance (**Table 16**); 2) to consider the validity of the assumption of comparable mortality

rates in non-participators in each arm. This analysis used baseline treatment information to categorise individuals in the no MDA arm as compliers or non-compliers, whilst keeping the compliance status for individuals in the MDA arm as per the primary compliance variable (complete compliance versus partial and non-compliance). This analysis assumed that compliance at baseline in the no MDA arm would be the same in subsequent MDA rounds, had treatment been available. Similar mortality rates in non-compliers in each arm would support the assumptions made in the CACE analysis based on the primary compliance variable.

To further describe all-age, all-cause mortality within the PRET cohort, Poisson regression with a robust SE adjustment and *a priori* adjustment for arm, coverage allocation, sex and age group (1-4, 5-14, 15-29, ≥30 years) was used to investigate associations with mortality at household and cluster level. Household level factors of interest were whether the household had a latrine and whether the household's water source was within a 15-minute walk. Cluster level factors of interest were distance from a health centre (<5km, ≥5km) and EA type (single settlement, part of a larger settlement or made up of several settlements).

5.3. Results

5.3.1. Participants

Measured characteristics of EAs and households were balanced by arm (**Table 13**). The distribution of sex and timing of entry into the census for individuals included in the analysis also appeared to be balanced.

5.3.2. Mortality rates

In total, 41802 individuals were included in the two-year cohort who contributed 75152.7 person-years at risk. With 554 deaths, the mortality rate was 7.4 (95% CI: 6.8 – 8.0) per 1000 person-years (**Table 14**). Mortality rates were similar in the MDA arm and no MDA arm. The overall mortality rate in children aged 1-4 years old was 4.7 per 1000 person-years (95% CI: 3.6 – 6.1), with similar rates in each arm (**Table 14**).

There was strong evidence of between-cluster variation in mortality rates ($p<0.001$) and some possible, weak evidence of clustering of mortality rates in children aged 1-4 years ($p=0.079$, **Table 15**).

5.3.3. Treatment status

Coverage amongst individuals aged ≥1 year old in the MDA arm of the two-year open cohort, eligible to receive treatment (resident at the time of an MDA round), was at least 85% during each MDA round (**Figure 20**). Not all individuals in the MDA arm were eligible in each round based on time of entry to the cohort; a decreasing percentage of being ineligible for treatment corresponded to an increase in the proportion of individuals in the cohort being treated in each subsequent MDA round. Between 7-10% of individuals in the MDA arm of the open cohort did not receive treatment in each round. Excluding those ineligible to receive treatment, these proportions would be higher due to more non-compliance in older ages.

Based on observed compliance in the MDA arm over the year one and year two treatment rounds, 82% of individuals in the two-year open cohort were completely compliant (received all treatments for which they were eligible or were ineligible in both rounds, **Figure 20**). Partial and non-compliers comprised 16% of the cohort and treatment status was unknown for one or more rounds in 2%.

Non-compliance over the two-year annual treatment period varied between clusters (**Table 15**) with median cluster-level percentages of non-compliance similar to overall percentages. All clusters received MDA at baseline and individuals who were eligible to receive treatment could

be classified as compliers, non-compliers or missing treatment status based on that one round of treatment. Although the median cluster level percentage of non-compliance in the baseline MDA round was slightly higher in the MDA arm compared to the no MDA arm (9.5% versus 6.7%, **Table 15**), the difference in distribution of non-compliance between arms was not significantly different (Wilcoxon rank sum test p=0.125).

Although the overall percentage of missing treatment data was low (<5%), with the majority of clusters having no or very little missing data, some clusters had more than 20% missing data (**Table 15**). There did not appear to be a different distribution of missing treatment data between arms for the baseline MDA round (Wilcoxon rank sum test p=0.650). Amongst those with unknown treatment status in the MDA arm, the two-year mortality rate was very high; 215 per 1000 person-years with missing treatment status data more common in older individuals. Missing treatment status could be more likely in those who were not treated and particularly so if the reason for not receiving treatment was being to unwell to attend a central distribution point for treatment in the community or to receive a visit from the treatment.

In children aged 1-4 years, there were no deaths amongst those with missing treatment status in the MDA arm.

5.3.4. Effectiveness of azithromycin MDA

There was no evidence of a difference in mortality rates between the MDA and no MDA arms in ITT analyses of all ages or in children aged 1-4 years old (**Table 17**, **Figure 22**). A benefit of MDA would be shown by an ITT rate ratio of less than 1, with the largest contribution being a direct effect of benefit in those who receive treatment. The ITT rate ratio was just above one (RR=1.11). If two annual azithromycin MDA rounds can be effective in reducing mortality in this setting, this RR could suggest that those who would benefit most from treatment in the MDA arm did not receive it. Alternatively, it could be that the distribution of mortality rates in the absence of treatment and underlying causes of mortality were not balanced by arm, which is a possibility with a small number of clusters per arm. Similarly, underlying non-compliance behaviour patterns associated with underlying cause of mortality may not have been balanced between arms.

Sensitivity analyses using random effects Poisson regression and negative binomial regression, as alternative approaches to account for clustering in the data, gave the same results to 1 decimal place.

Similar results were obtained for the subgroup of children aged 1-4 years old, with a rate ratio closer to one.

5.3.5. Efficacy of azithromycin in individuals

If azithromycin MDA provides a direct benefit of a reduction in mortality, it would be expected that a PP and AT analysis would give a RR less than 1. Results of the PP and AT analysis were almost identical; a significant 50% reduction in all-age all-cause mortality (**Table 17**, **Figure 22**).

Results for children aged 1-4 years old were similar but less pronounced; the PP analysis suggested a 7% reduction in mortality and the AT analysis, a reduction of 22%. Neither of these effects were significant (**Table 17**, **Figure 22**).

If the assumptions made during a CACE analysis are valid and there is a true benefit of treatment, it would generally be expected that a CACE RR would suggest a larger beneficial effect size than the ITT RR, but perhaps not as large as a PP or AT analysis because the CACE analysis is supposedly less subject to selection bias. The CACE RR for all-age all-cause mortality was 1.36 (95% CI: 0.86 – 2.79, **Table 17**).

The sensitivity analyses considering compliers as complete or partial compliers gave a CACE RR of 1.26 (95% CI: 0.88 – 2.02) and with partial compliance as a distinct compliance category gave a CACE RR of 1.36 (0.86 – 2.72). These results possibly suggest that partial compliers are not so dissimilar to compliers. The complete compliance group also includes individuals who were not eligible to receive any treatment during their time at risk; a group who comply 100% based on eligibility but who do not have the opportunity to receive treatment. The mortality rate in ineligible individuals in the MDA arm was similar to the mortality rate in complete compliers eligible to receive at least one treatment when summarised as per **Figure 21**, ignoring clustering (0.002 versus 0.004).

In the MDA arm, 68% of non-compliers at baseline complied in future MDA rounds and 15% of baseline compliers were did not comply in the two-year follow-up period. Comparing compliers in the MDA arm based on observed primary compliance in the two-year open cohort period, to compliers in the no MDA arm based on observed compliance during the baseline MDA round gave a CACE RR of 0.48 (95% CI: 0.38 – 0.60); an effect similar to AT and PP effects. In this analysis, complier groups in each arm are also unlikely to be comparable.

Mortality rates were found to differ between districts, with one district having higher adult mortality than each of the others (**Table 18**). In the southern study area, with the district with the highest adult mortality rate, more heterogeneous population/settlement structures were strongly associated with increased mortality (**Table 18**). Overall non-compliance was higher in the southern districts compared to the northern districts (*chi-squared test* p<0.001 ignoring clustering; 70% vs 83% complete compliance in those eligible for at least one treatment).

The sampling distributions of bootstrapped values of the ITT and CACE RRs for all-age, all-cause mortality were normally distributed (**Figure 23**) but the 95% CI for the ITT RR was less precise compared to Poisson regression results.

In children aged 1-4 years old, the CACE RR based on primary compliance status was very close to the ITT effect (**Table 17**, **Figure 22**). In the much smaller sample of children aged 1-4 years old, the sampling distribution for the CACE RR was skewed to the right with some extreme high values (**Figure 23**).

5.4. Discussion

There was no evidence of effectiveness of azithromycin MDA on mortality. There was no clear, unbiased indication of efficacy, that is, if there is an effect of azithromycin in those who take it, as suggested by an as-treated analysis, there is no unbiased way of quantifying how beneficial it could be against all-cause mortality.

The sample size available from the PRET trial would likely only be able to provide enough power to detect very large, unrealistic effect sizes in this setting. This was an analysis of a secondary outcome in a study that was not intentionally underpowered, but not intentionally powered for an outcome of all-cause mortality either. An underpowered trial can be at risk of a false negative result (type II error), a false-positive (type I error) or an exaggerated effect size where a true effect is detected[141].

All-age, all-cause mortality was low (7.4 per 1,000 person-years) and the PRET trial did not have pre-specified power, for a pre-specified effect size, for a difference in mortality between arms. In addition, although treatment status was only missing for 2% of individuals in the MDA arm, the overall mortality rate in those with missing treatment status was very high in the MDA arm (215 per 1,000 person-years). This meant that the CACE analysis had to have three compliance categories; participators, non-participators and missing treatment status and the bootstrap method could allow for multiple compliance categories. Neither the ITT or CACE provided

evidence of a difference in mortality between arms, while per-protocol and as-treated analyses did show a significant reduction in mortality. The ITT and CACE rate ratios comparing MDA to no MDA were not in the direction of benefit (1.11 and 1.36 respectively).
Baseline compliance was not a reliable predictor of future compliance for everyone, despite previous findings of repeat compliance behaviour (chapter 3). If it had been, mortality rates in non-compliers in the MDA arm and non-compliers in the no MDA arm could have been compared to check the assumption of comparable rates in non-compliers in each arm.

If the significant AT and PP findings were due to a direct therapeutic effect of the drug in those who receive it, it would be expected that ITT and CACE results would be in the direction of benefit, even if not significant. If individuals at lower underlying risk of mortality are more likely to actively access health care provisions and actively engage with health care activities on offer, a healthy complier effect might explain AT and PP findings, especially if the (unmeasured) underlying risk of mortality differed between arms. Although again, one might expect to see the ITT and CACE RRs in the same direction of effect even if not significant.

Imbalance between arms is more likely in small randomised trials. It may be the case that the distribution of clustering of compliance status (including missing status), clustering of underlying mortality risk and clustering of compliance status (including unknown status) associated with mortality in the MDA arm, is not comparable to what would have been observed in the no MDA arm, or at least we cannot assume it would be. Such imbalance between arms would lead to bias in both ITT and CACE results and a violation of a key assumption in a CACE analysis; that the mortality rate in a would-be group of non-compliers in the no MDA arm is the same as the observed mortality rate in non-compliers in the MDA arm. Similarly, for missing treatment status and any other category of compliance status other than complier.

Although the randomisation was stratified by district, and therefore by northern vs southern study areas, this may not have eliminated any bias resulting from over-or under-representation of one district, nor achieved balance between arms or by district or overall. Thus, it is plausible that between-cluster variation of underlying mortality risk, non-compliance and mortality associated with compliance status could have led to bias in the CACE RR.

It thought likely that the pragmatic approach to estimating efficacy did not manage to limit bias and that the association between missing treatment status and increased mortality is likely to have introduced substantial bias.

Although data were available for an extra year prior to the period of no MDA in on arm, a three-year analysis period starting from the baseline census does not fit well within a CACE analysis framework and would have only provided a small increase in power. Instead, sensitivity analyses of compliance status were used to incorporate baseline treatment information for those in the two-year open cohort.

Results reported here are in contrast with results from other CRTs in Africa. Data from a trial in Ethiopia designed to evaluate the impact of MDA on trachoma, without a pre-specified level of power to detect a pre-specified effect on child mortality, showed a 50% reduction in all-cause mortality in children aged 1-9 years old (ITT odds ratio=0.51, 95% CI: 0.29 – 0.90, p=0.020)[137], comparing clusters receiving one to four rounds of MDA over the course of a year to clusters not receiving MDA. In another CRT of azithromycin MDA for trachoma control in Niger, without a pre-specified level of power to detect a pre-specified effect, mortality in children aged 6-59 months was compared between clusters receiving six biannual rounds of MDA over three years to clusters receiving three rounds of MDA over three years (ITT rate ratio=0.81, 95% CI: 0.66 – 1.00, p=0.07)[142]. In a recent CRT, specifically designed to evaluate the impact of azithromycin MDA on child mortality, mortality rates in children aged 1-59 months were compared between clusters receiving four six-monthly rounds of MDA over two years, to clusters receiving four rounds of placebo treatment, in three countries. Results from Niger, Tanzania and Malawi combined showed a reduction in mortality of 13.5% (95% CI: 6.7 to 19.8%)[138].

Data from 24 clusters in the trial in Ethiopia receiving MDA at baseline were used for an as-treated analysis of all-cause and infectious-cause mortality in different age groups (12 clusters: MDA to all aged ≥1 year at baseline and 12 clusters: MDA to all aged ≥1 year at baseline plus latrine construction). The authors attempted to reduce the risk of bias with a matched analysis by household, to adjust for household-level confounders of the effect of treatment receipt on mortality, and reported a significant reduction in mortality in treated children aged 1-5 years old compared to untreated children[47]. Although the analysis of the Ethiopian data included measures to limit bias, residual confounding of the as-treated treatment effect is likely. The authors of the Ethiopian trial themselves, highlight that the magnitude of the effect detected in children aged 1-5 years old is unrealistically high and the published report does not mention whether there were any missing data for treatment receipt.

The as-treated results from Ethiopia and The Gambia may have arisen due to an actual benefit of azithromycin in those who receive it, but there is no way of knowing if there is a true benefit or what the magnitude of a true effect could be, from the data available.

This study illustrated a pragmatic bootstrapping procedure to estimate the *efficacy* of azithromycin treatment using data from a CRT of azithromycin MDA for trachoma control, intended to limit the impact of selection bias whilst accounting for heterogeneity in the outcome, non-participation and missing treatment status. This approach is preferable to other methods subject to bias, such as PP and AT analyses and could provide an alternative to modelling approaches to estimate *efficacy* such as those reported by Lui et al (2014)[116] when applied to field data. However, the CACE analysis was also subject to bias when applied to the PRET data.

Other reasons for bias and unreliability in the analyses could be misclassification of treatment status during MDA rounds, errors in census data and for the CACE analysis, that the bootstrapping approach to account for non-compliance and clustering is inadequate for a rare outcome, an inadequate sample size and small compliance categories. Bias in a CACE analysis can be reduced through adjustment for strong predictors of compliance[143], but such adjustments require complex computation methods and are not possible within the pragmatic two-step bootstrapping approach. Limitations of bootstrapping methods for analysis of clustered data for correlated rare binary events are documented[144]. Opondo *et al*[140], who proposed this bootstrapping application to data from CRTs, report an outcome with an overall prevalence of 19% and overall non-participation in the intervention arm of 58%, without a missing data compliance category. In their analysis, 1,000 replications were adequate for a normal sampling distribution for the CACE and for identical values for the ITT point estimate and 95% confidence bounds from regression and bootstrapping methods, to two decimal places. Missing data is acknowledged to be a major difficulty when trying to estimate a CACE[143]. Imputation of missing treatment status was not considered as any approach to do so, based on the limited covariate data or assuming missing treatment status was always non-participation, would likely introduce further bias.

The mortality rate observed in here in children aged 1-4 years old of 4.7 per 1000 person-years (3.6 – 6.1) would appear to fit with published data from The Gambia, if previously observed declines in mortality continued into the PRET trial period. Demographic surveillance data from a geographic area north of the River Gambia, where half of the PRET clusters were located, showed large declines in mortality in 1-4 year olds over the 10 year period prior to the PRET trial[145]. Previously high mortality in this age group in The Gambia was attributed to diseases such as measles, malaria, diarrhoea and pneumonia[145] but it is thought that good immunisation rates and malaria control measures have contributed to the decline[145-147]. Within children aged 1-4

years old, it is the children under two who are at higher risk of death, with risk factors thought to include not only infectious diseases but also maternal mortality and residing in more remote areas with limited access to appropriate health care[147]. Data from the recent CRT specifically designed to evaluate the impact of azithromycin MDA on child mortality, suggest that the largest benefit is in children under six months of age[138]. Despite azithromycin MDA being a relatively low-cost large-scale intervention when azithromycin is donated[113], an unfeasibly large trial would be required to demonstrate an impact of azithromycin MDA on under-five all-cause mortality, through some direct effect against infectious causes of deaths in under-twos, in The Gambia.

The CACE analysis approach illustrated here could be more widely applicable to CRTs for other neglected tropical diseases (NTDs) with MDA control (schistosomiasis, lymphatic filiariasis, soil transmitted helminths and onchocerciasis)[148], where there is a period of follow-up with no MDA in one randomisation arm. It could also be applicable to vaccine trials where clusters are allocated to vaccination or no vaccination, but individuals are not randomised within vaccinated clusters[126].

5.5. Conclusions

The PRET trial data have not provided any conclusive information about whether large-scale azithromycin MDA would lead to important reductions in mortality in The Gambia. Estimating efficacy in a CRT of MDA intervention, with precision and without bias, remains a challenge. The pragmatic bootstrapping approach to estimate the effect of azithromycin in those who receive it requires reliable compliance and census data.

Table 13. Characteristics of the two-year open cohort

Characteristic		MDA	No MDA
Number of EA		24	24
Bank	South	12	12
	North	12	12
District	South: district 1	6	6
	South: district 2	6	6
	North: district 1	6	6
	North: district 2	6	6
EA type	Multiple-SET	12 (50.0)	13 (54.2)
	Multiple-EA	9 (37.5)	7 (29.2)
	Single EA-SET	3 (12.5)	4 (16.7)
EA size at 12 months census: small <11, medium 11-16, large ≥17	Small	13 (54.2)	10 (41.7)
	Medium	6 (25.0)	11 (45.8)
	Large	5 (20.8)	3 (12.5)
Number of Households at 12 months census		1560	1543
HH size: small <11, medium 11-16, large ≥17	Small	964 (61.8)	995 (64.4)
	Medium	402 (25.8)	365 (23.7)
	Large	194 (12.4)	183 (11.9)
Latrine access	No	197 (12.6)	124 (8.0)
	Yes	1363 (87.4)	1419 (92.0)
Time to water	≥15 minutes	217 (13.9)	200 (13.0)
	< 15 minutes	1343 (86.1)	1343 (87.0)
Distance from health centre	<5 km	1163 (74.6)	1203 (78.0)
	≥5 km	397 (25.5)	340 (22.0)
Number of residents in open-cohort		21664	20138
Sex	Male	10000 (46.2)	9399 (46.7)
	Female	11664 (53.8)	10739 (53.3)
Time of entry to census: census data collection started at 0 months; two-year cohort time at risk started at 12 months	0 months	16210 (74.8)	16009 (79.5)
	6 months	1000 (4.6)	1153 (5.7)
	12 months	1746 (8.1)	1034 (5.1)
	18 months	859 (4.0)	788 (3.9)
	24 months	1409 (6.5)	606 (3.1)
	30 months	440 (2.0)	548 (2.7)

Data are n (%). EA = enumeration area, SET = settlement, MDA = mass drug administration.

Table 14. All-cause mortality rates for the two-year open cohort

		All-ages	1-4 years
All	N	41802	8192
	Person-years	75152.7	11291.2
	Deaths	554	53
	Rate per 1000 person-years (95% CI)	7.4 (6.8 – 8.0)	4.7 (3.6 – 6.1)
MDA arm	N	21664	4245
	Person-years	38664.0	5864.9
	Deaths	299	28
	Rate per 1000 person-years (95% CI)	7.7 (6.9 – 8.7)	4.8 (3.3 – 6.9)
No MDA arm	N	20138	3947
	Person-years	36488.7	5434.0
	Deaths	255	25
	Rate per 1000 person-years (95% CI)	7.0 (6.2 – 7.9)	4.6 (3.1 – 6.8)
South			
District 1	N	9381	1742
	Person-years	17098.5	2438.0
	Deaths	108	11
	Rate per 1000 person-years (95% CI)	6.3 (5.2 – 7.6)	4.5 (2.5 – 8.1)
District 2	N	13217	2423
	Person-years	23548.8	3373.3
	Deaths	245	23
	Rate per 1000 person-years (95% CI)	10.4 (9.2 – 11.8)	6.8 (4.5 – 10.3)
North			
District 1	N	8695	1781
	Person-years	15516.6	2393.8
	Deaths	97	9
	Rate per 1000 person-years (95% CI)	6.3 (5.1 – 7.6)	3.8 (2.0 – 7.2)
District 2	N	10509	2246
	Person-years	18987.9	3086.0
	Deaths	104	10
	Rate per 1000 person-years (95% CI)	5.5 (4.5 – 6.6)	3.2 (1.7 – 6.0)

Table 15. Summary of treatment status and mortality data

	Age (years)	
	All	1-4
N	41802	8192
MDA arm, N	21664	4245
Missing treatment status in MDA arm, n (%)	508 (2.3)	90 (2.1)
Number (%) of deaths amongst those with missing treatment status in MDA arm over two treatment rounds	140 (27.9)	0/90 (0)
Median (range) of cluster level percent partial or non-compliance in MDA arm over two treatment rounds	15.4 (6.1 – 30.8)	14.1 (4.0 – 25.7)
Median (range) of cluster level percent non-compliance in MDA arm over two treatment rounds	4.9 (0.7 – 11.7)	5.5 (0.0 – 13.7)
Median (range) of cluster level percent missing treatment status in MDA arm over two treatment rounds	2.0 (0.6 – 6.7)	1.6 (0 – 7.8)
Median (range) of cluster level percent non-compliance in the MDA arm at baseline MDA round	9.3 (1.4 – 18.5)	6.5 (0.6 – 18.0)
Median (range) of cluster level percent missing treatment status in the MDA arm at baseline MDA round	0.8 (0 – 23.9)	1.7 (0 – 19.8)
Median (range) of cluster level percent non-compliance in the no MDA arm at baseline MDA round	6.7 (0 – 25.9)	5.9 (0 – 31.3)
Median (range) of cluster level percent missing treatment status in the no MDA arm at baseline MDA round	1.0 (0 – 38.3)	3.0 (0 – 39.8)
Median (range) of cluster level mortality rate per 1000 person-years	6.7 (2.2 – 24.2)	4.0 (0 – 21.1)
Mortality clustering p-value ITT[a]	<0.001	0.079

[a] p-value from likelihood ratio test of between-cluster variation from null random effect Poisson regression model

Table 16. Categorisation of compliance status for CACE analysis

	Treated Status (+ treated, - eligible but not treated)							
Arm	Baseline	Year 1	Year 2	Classification	Primary compliance categorization [a]	Sensitivity compliance 1[b]	Sensitivity compliance 2 [c]	Sensitivity compliance 3 [g]
Two rounds of treatment, starting at year one:								
MDA arm		+	+	Complete	Complier	Complier	Complier	
		+	-	Partial	Non-complier	Complier	Partial	
		-	+	Partial	Non-complier	Complier	Partial	
		-	-	None	Non-complier	Non-complier	Non-complier	
		m	+ (or -)	Missing	Missing	Missing	Missing	
		+ (or -)	m	Missing	Missing	Missing	Missing	
		m	m	Missing	Missing	Missing	Missing	
No MDA arm				n/a	Unknown	Unknown	Unknown	
Three rounds of treatment [d]:								
MDA arm	+	+	+	Complete	Complier	Complier	Complier	Complier
	-	+	+	Partial	Non-complier [e]	Complier	Partial [e]	Complier
	+	+	-	Partial	Non-complier	Complier	Partial	Non-complier
	-	+	-	Partial	Non-complier	Complier	Partial	Non-complier
	+	-	+	Partial	Non-complier	Complier	Partial	Non-complier
	-	-	+	Partial	Non-complier	Complier	Partial	Non-complier
	+	-	-	Partial	Non-complier	Complier	Partial [f]	Non-complier
	-	-	-	None	Non-complier	Non-complier	Non-complier	Non-complier
No MDA arm	+			Complete	Unknown	Unknown	Unknown	Complier
	-			None	Unknown	Unknown	Unknown	Non-complier

Grey text indicates sensitivity analyses not done

[a] complier category includes only complete compliance; compliance based on three-year information not different enough to warrant a sensitivity analysis as partial compliance based on not receiving baseline treatment but complying in year one and two can reasonbly be considered as complete compliance for the two-year follow-up period of interest

[b] complier category includes complete and partial compliance (at least one treatment in year one or year two when eligible for two treatments); no difference between compliance variables based on two-year or three-year information

[c] includes partial compliance as a separate category; before, when partial compliance was included with complier or non-complier, it was assumed that rates for partial compliers were the same as for compliers or non-compliers respectively

[d] missing categorisations as for two-year period (not shown)

[e] a different categorisation compared to considering just two rounds of treatment; reasonable to consider this group as compliers since compliance was complete for the two-year follow-up period of interest

[f] another different categorisation compared to considering just two rounds of treatment; reasonable to consider this group as non-compliers since there was no compliance during the two-year follow-up period of interest

[g] allows for classification in the no MDA arm as compliers, non-compliers or missing treatment based on observed data at baseline and assumes individuals would have the same compliance status at subsequent rounds if offered treatment.

Table 17. All-cause mortality rate ratios for two-year open cohort by analysis population

		Mass drug administration (MDA)												No MDA			Rate Ratio (95% CI)
		Compliers			Non-compliers [c]			Missing treatment			Total			Total			
		N	PY	D	N	PY	D	N	PY	D	N	PY	D	N	PY	D	
All ages	ITT[a]	17738	31516.3	109	3425	6523.2	50	501	651.0	140	21664	38690.5	299	20138	36513.7	255	1.11 (0.85 – 1.44)
	PP[a]	17738	31516.3	109										20138	36513.7	255	0.50 (0.37 – 0.66)
	AT (C vs NC)[a]	17738	31516.3	109	3425	6523.2	50							20138	36513.7	255	0.49 (0.38 – 0.62)
	CACE[b c]																1.36 (0.86 – 2.79)
1-4 years	ITT[a]	3589	4879.2	20	566	865.7	8	90	120.0	0	4245	5864.9	28	3947	5434.0	25	1.04 (0.56 – 1.94)
	PP[a]	3589	4879.2	20	566	865.7	8							3947	5434.0	25	0.93 (0.50 – 1.72)
	AT (C vs NC)[a]	3589	4879.2	20	566	865.7	8										0.78 (0.46 – 1.34)
	CACE[b c]																1.05 (0.51 – 2.88)

N = number of individuals, PY = person-years, D = number of deaths, C = compliers, NC = non-compliers

[a] Poisson regression with robust SE

[b] Bootstrap procedure with 10,000 replications

[c] non-compliers include partial compliers

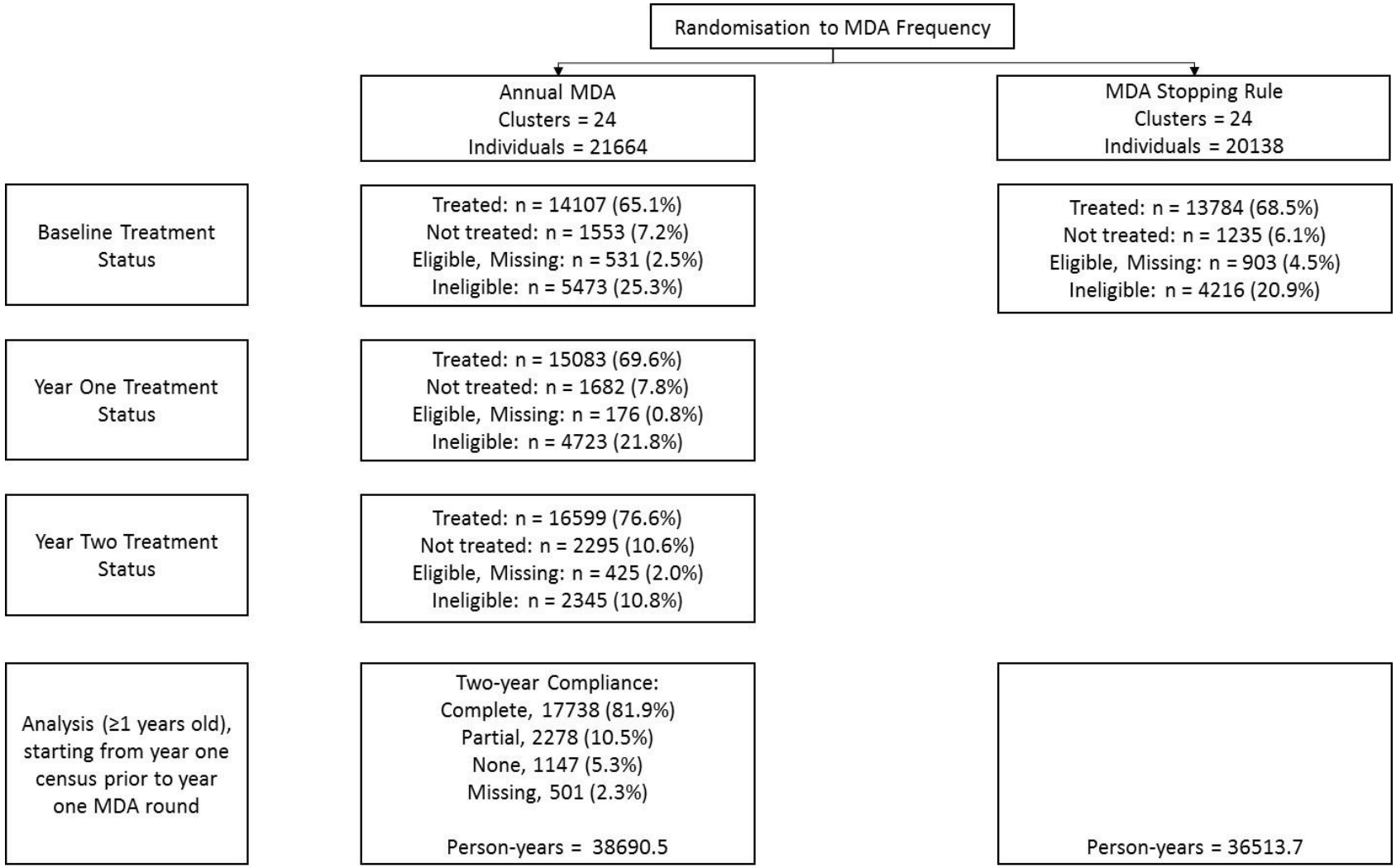
Table 18. Associations with two-year mortality in the two-year open cohort

Age	District	All: RR (95% CI)	All: RR (95% CI)	South bank area: RR (95% CI)	South bank area: RR (95% CI)
All	Foni Bintang	1		1	1
	Foni Kansala	**1.58 (1.25 – 2.01)**			
	Lower Baddibu	1.06 (0.80 – 1.41)		-	-
	Central Baddibu	0.96 (0.72 – 1.28)		-	-
1-4	Foni Bintang		1	1	1
	Foni Kansala		1.50 (0.67 – 3.37)	1.66 (0.12 – 0.66)	1.57 (0.70 – 3.53)
	Lower Baddibu		0.84 (0.32 – 2.20)	-	-
	Central Baddibu		0.72 (0.21 – 2.38)	-	-
5-14	Foni Bintang		1	1	1
	Foni Kansala		1.00 (0.37 – 2.73)		1.06 (0.39 – 2.89)
	Lower Baddibu		1.37 (0.46 – 4.08)	1.10 (0.41 – 2.95)	-
	Central Baddibu		1.06 (0.42 – 2.71)	-	-
15-29	Foni Bintang		1	1	1
	Foni Kansala		**1.86 (0.98 – 3.54)**	**2.03 (1.06 – 3.89)**	**1.96 (1.03 – 3.70)**
	Lower Baddibu		0.85 (0.31 – 2.33)	-	-
	Central Baddibu		0.39 (0.13 – 1.17)	-	-
30+	Foni Bintang		1	1	1
	Foni Kansala		**1.60 (1.15 – 2.21)**	**1.75 (1.27 – 2.40)**	**1.66 (1.25 – 2.22)**
	Lower Baddibu		1.10 (0.74 – 1.63)	-	-
	Central Baddibu		1.09 (0.74 – 1.60)	-	-
Distance from Health Centre	<5 km	-	-	1	1
	≥5 km	-	-	**1.24 (1.02 – 1.51)**	1.12 (0.87 – 1.45)

Age	District	All: RR (95% CI)	All: RR (95% CI)	South bank area: RR (95% CI)	South bank area: RR (95% CI)
EA type	Single EA-SET	-	-	-	1
	Multiple-SET within an EA	-	-	-	**1.55 (1.23 – 1.96)**
	Multiple-EA within a settlement	-	-	-	**1.40 (1.05 – 1.85)**

Adjusted for arm, coverage allocation, sex. All EAs on the north bank were <5km from a health centre. Distance to health centre and EA type were associated with mortality in districts but only in the southern study area. After adjustment for distance in the southern districts, there was no effect of EA type. For multi-settlement EAs, 9/14 (64%) were ≥5 km from a health centre; as were 1/6 (17%) multi-EA settlements and 1/4 (25%) single EAs; higher mortality in multi-settlement EAs could be the same increased mortality seen for distance from a health facility, so distance may still be an explanatory factor or a proxy for isolation or poverty.

Figure 20. Two-year open cohort: participant flow



Complete compliance = receiving all treatments for which the individual was eligible.

Overall coverage in each round amongst those eligible: baseline MDA arm = 87.1%; baseline no MDA arm = 86.6%; year one MDA arm = 89.0%; year two MDA arm = 85.9%

Figure 21. Obtaining the complier average causal effect (CACE) of treatment on all-age all-cause mortality, ignoring clustering

		Complier (C)	Non-complier (NC)	Missing treatment (MT)	Total
MDA	N	17738	3425	501	21664
	Deaths (D)	109	50	140	299
	Person-years (PY)	31516.3	6523.2	651.0	38690.5
	Rate	0.003	0.008	0.215	0.008
No MDA	N	←	←	←	20138
	Deaths (D)	D = 255 – 132.1 – 49.2 = 73.7	D = rate x PY = 0.008 x 6156.2 = 49.2	D = rate x PY = 0.215 x 614.4 = 132.1	255
	Person-years (PY)	PY = 36513.7 – 614.4 – 6156.2 = 29743.1	same proportion of person-time contribution as MDA arm; PY = (6523.3/38690.5) x 36513.7 = 6156.2	same proportion of person-time contribution as MDA arm; PY = (651.0/38690.5) x 36513.7 = 614.4	36513.7
	Rate	Rate = 73.7 / 29743.1 = 0.002	same as rate MDA arm = 0.008	same as rate as MDA arm = 0.215	0.007

CACE Rate Ratio
= 0.003 / 0.002 = 1.50

ITT Rate Ratio
= 0.008/0.007 = 1.14

Observed data values shown in black text, assumed values shown in blue text.

Figure 22. All-cause mortality rate ratios and 95% CIs for two-year open cohort

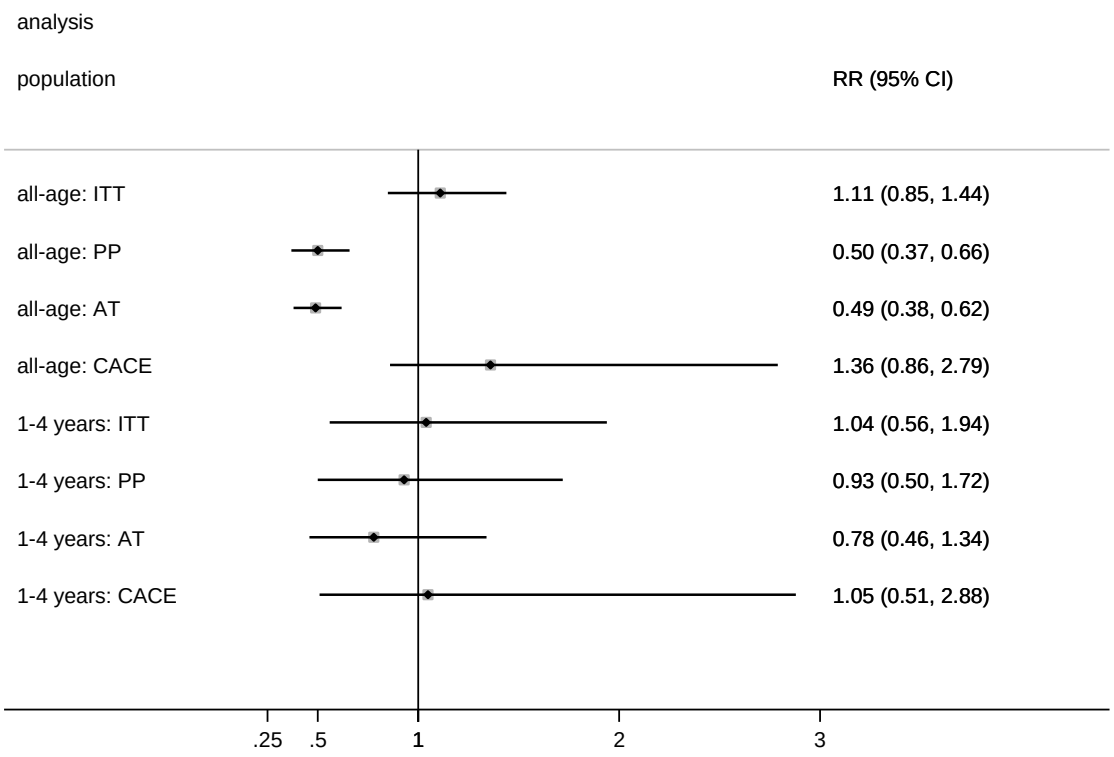


Figure 23. Distribution of bootstrapped ITT mortality rate ratios

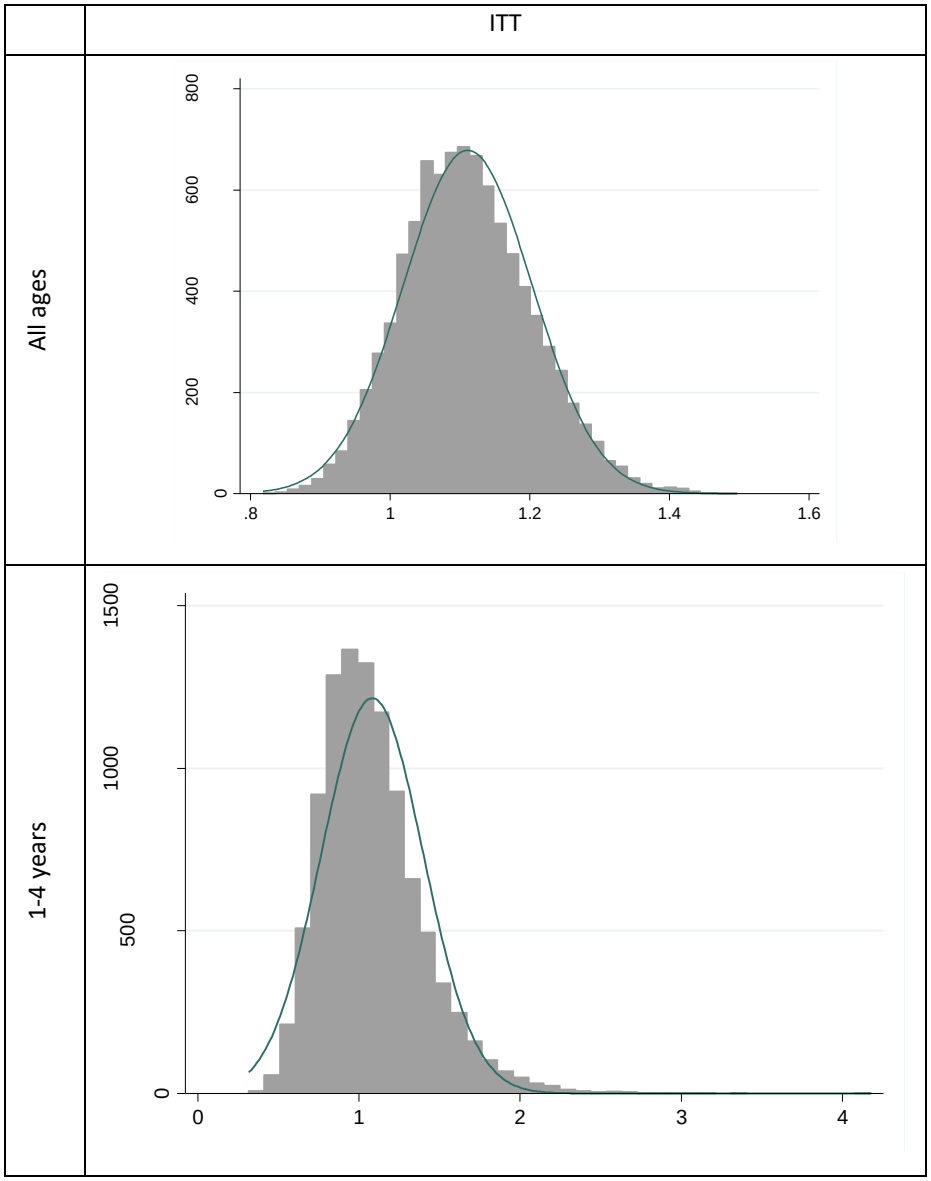
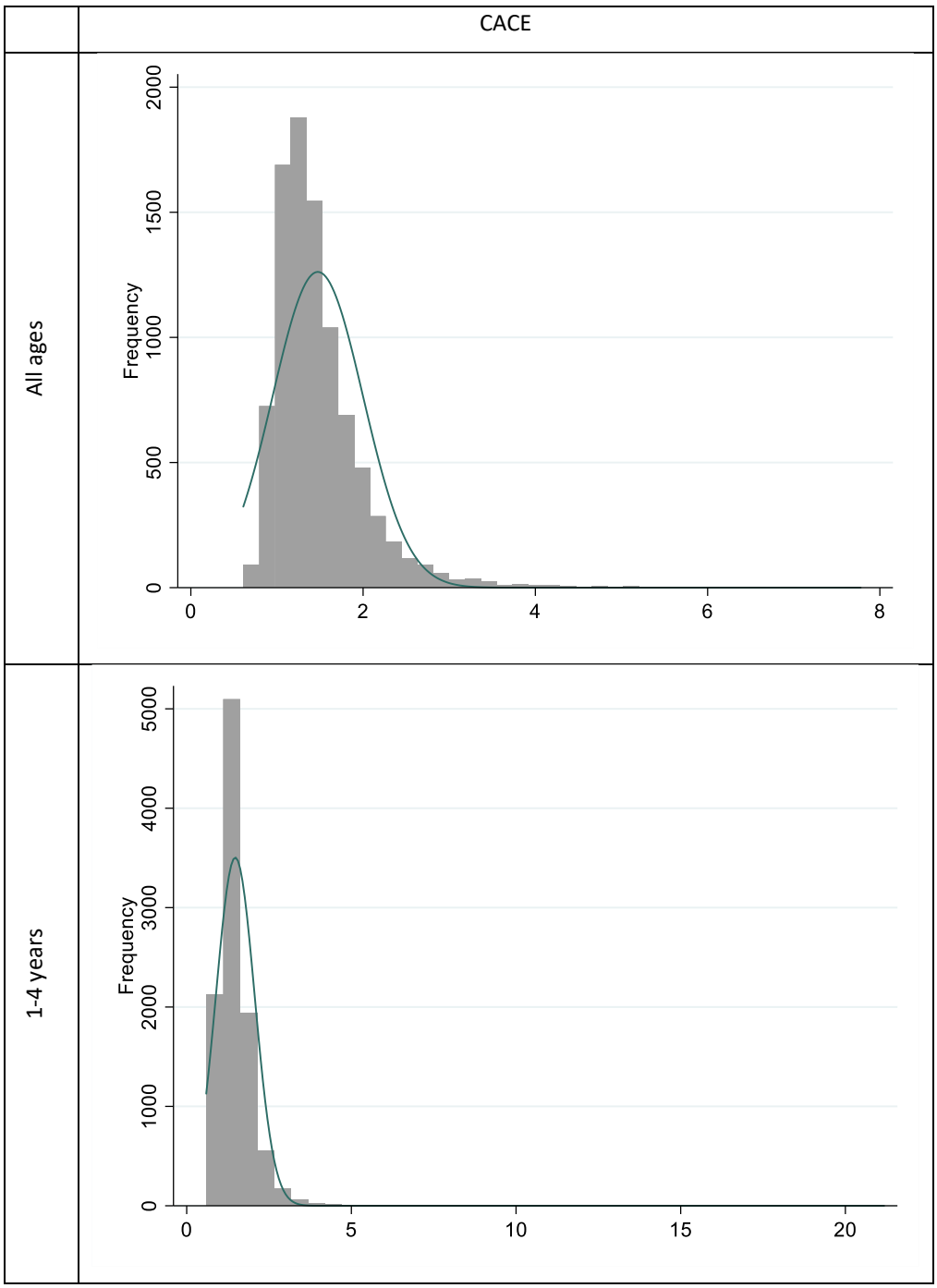


Figure 24. Distribution of bootstrapped CACE mortality rate ratios



6. Discussion

A detailed review of published CRTs of MDA interventions for trachoma control identified how heterogeneity in trachoma outcomes, coverage and individual receipt of azithromycin during MDA rounds was reported and handled during the design and analysis of the included CRTs.

The implications of heterogeneity in outcome data and heterogeneity in non-participation data, for the design and analysis of CRTs to evaluate MDA interventions for trachoma control, were explored in this thesis.

In this chapter, findings of the work in this thesis are summarised in relation to trachoma control, and for the design and analysis of future CRTs of mass treatment for trachoma and other diseases.

6.1. Summary of findings

Untreated individuals during MDA for trachoma control could hinder elimination efforts by maintaining a source of transmission and re-infection, especially if they have infection prior to MDA or are at a greater risk of exposure to infected individuals. Over the course of three annual MDA rounds in the hypo-endemic trachoma setting of The Gambia, non-participation increased significantly amongst children aged 1-9 years old from 6% to 10%, with strong evidence that non-participation was spatially heterogeneous and occurred repeatedly in some of the same households and individuals (chapter 3).

Non-participation occurred in individuals who were recorded as present but not treated (PNT) or eligible but absent (EBA) during each MDA round. Each non-participation status could include individuals for whom a choice was made not to receive treatment. Comparing children aged 1-9 years old of each type of non-participator to children who participated (received treatment during MDA), it was found that predictors of non-participation were mainly at household level. For both PNT and EBA, increased time to water and non-participation of the household head were risk factors, larger household size was a risk factor for being PNT and non-inclusion in a previous trachoma examination survey and younger age were risk factors for EBA (chapter 3). There were no quantitative or qualitative interviews of participators or non-participators to provide confirmation information, but based on the findings of this analysis, it is hypothesised that non-participators could be those in families or residential groups who may engage less with community health activities. There was no evidence of lower non-participation in communities where enhanced coverage efforts took place. Non-participation could have been due to activities away from the community, such as farming or trading and independent of risk of

poorer health. If a lack of engagement with community-based health activities is associated with ostracization or geographical distance from central community health activities, it is plausible that there is an association with poor health in general. Visits to untreated communities in endemic areas during the time of an MDA round poses a risk of infection and then re-transmission to the treated community. Clustering of persistent non-participation, if it occurs in relation to any community health activities, could hinder control efforts for a number of infectious diseases.

Power was low for investigating statistical evidence of an association between non-participation and subsequent risk of infection at follow-up as only 24 children had infection at the end of the study. Infections did occur in children residing in communities close to untreated endemic areas with higher rates of EBA non-participation. The low numbers of infections also made it difficult to assess how much non-participation may occur amongst children with *C. trachomatis* infection prior to an MDA. Information from other settings about the probability of non-participation amongst infected individuals is also lacking, probably because investigations of risk factors for non-participation have been focussed on possible causal effects. As infection status, confirmed by laboratory testing, is unknown in an MDA round following soon after a trachoma survey, infection status itself was not considered a plausible causative factor in the analyses performed, given that individuals did not know their infection status and could not make a treatment decision based on infection status, even though it is possible that risk factors for infection could also be risk factors for non-participation. In the clusters of the PRET trial in The Gambia, only 38 children were infected at baseline, of whom 37 were subsequently treated with one child having a missing treatment status. At 24 months, one child was infected in communities subsequently offered MDA and that child had a missing treatment status (chapter 3).

Levels of infection and TF were both very low at the end of the PRET trial in The Gambia and non-participation was also relatively low (≤10%), so non-participation in this setting was considered unlikely to influence trachoma control. In fact, elimination targets for TF in children aged 1-9 years old in The Gambia have since been considered to have been reached[149] (chapter 3).

Untreated individuals during MDA also pose a risk of loss of power and bias in cluster randomised trial evaluations[51,79]. Despite low prevalence and a lack of evidence of a link between non-participation and infection status in the hypo-endemic setting of The Gambia, these results raise an important question about how non-participation occurring amongst infected individuals can affect pre-specified power in a CRT of MDA interventions. In settings

with more infection, there is more opportunity for non-participation to occur amongst infected individuals (heterogeneous non-participation).

It would appear from the literature review that little attention is paid to non-participation and efficacy of azithromycin in individuals who receive treatment, during trial design (chapter 1.6) Simulation studies were used to explore the impact of non-participation on power to detect pre-specified effect sizes of a difference in prevalence of infection between arms in CRTs of MDA versus no MDA. In order to understand the relationship between overall non-participation, non-participation amongst individuals infected at baseline and power, simulation studies were based initially on trials accounting for efficacy, but with no account of non-participation in sample size calculations. Then, in trials with an adjustment for overall non-participation included in sample sizes, i.e. trials powered to allow for coverage<100%, the impact of different percentages of infected individuals not receiving treatment on power was explored, to determine how many additional clusters might be necessary to retain a pre-specified level of power (chapter four). Given that population sizes may limit the possibility of increasing sample size by increasing the number of individuals per cluster and that CRT power is most reliant on the number of clusters, only increases in the number of clusters were considered to boost the sample size.

Based on the design of the simulation studies conducted here (chapter four), results show that non-participation amongst individuals infected at baseline has important implications for study power, when the cluster mean percentage of non-participation amongst individuals infected at baseline is greater than the cluster mean percentage of non-participation overall (or greater than the cluster mean percentage of non-participation amongst infected individuals). Coverage targets for azithromycin MDA for trachoma control are at least 80% so simulation studies included parameters for non-participation of up to 20% non-participation overall and up to 20% non-participation amongst those infected at baseline. The most conservative approach when estimating the number of clusters required for a CRT with MDA coverage of 80% would mean allowing for 20% non-participation when making assumptions about likely effect sizes. If, on average at cluster level, 20% of infected individuals do not receive treatment during MDA, the increase in the number of clusters required to retain the pre-specified level of power could be very large; at least a 100% increase in, or double, the number of clusters compared to a trial with no adjustment for non-participation. This has huge resource implications, however, if the cluster mean percentage of non-participation amongst individuals infected at baseline is notably less than the cluster mean prevalence of non-participation overall or amongst uninfected individuals, fewer additional clusters may be required. If, for example, the cluster mean level coverage is 80% (20% non-participation), but the cluster mean percentage of non-participation

amongst infected individuals is around 10%, it could be the case that sample sizes need only account for 90% coverage (meaning an increase of around a 50% increase in clusters) to retain power.

In the context of trachoma, there are apparently no published data on the prevalence of non-participation during MDA rounds amongst individuals infected at baseline, although data from CRTs do exist to evaluate this. Such existing data could be used to quantify cluster mean non-participation in infected individuals, in relation to the cluster mean coverage.

After a therapeutic benefit of azithromycin for *C. trachomatis* infection in individuals was confirmed, azithromycin MDA became the mainstay of trachoma control and there is no doubt that this reduces the community level prevalence of *C. trachomatis* infection and TF. In reality, the *efficacy*, or percentage of individuals whose infection is cleared with a single dose of azithromycin, is unlikely to be 100%.

An ITT analysis of CRT data evaluates the *population-level effectiveness* of azithromycin MDA on the outcome, ignoring any individual-level non-participation. Future efforts to control trachoma or other health outcomes via azithromycin MDA would benefit from an accurate indication of *efficacy*, i.e. the effect of azithromycin in those who actually take it[116]. *Efficacy* is very difficult to estimate reliably, without bias, within a randomised controlled trial framework because individual receipt of treatment (compliance, or participation) may not be balanced between arms and may be linked to an increased risk of the outcome. In CRTs, there is the added complexity that clusters are randomised and not individuals. In an analysis of CRT data based on actual treatment receipt, there could be confounding of the effect of treatment on the outcome in children due to factors at individual, familial and cluster level.

Unbiased analytical approaches to estimate *efficacy* in CRTs, involve estimating efficacy based on observed data in the treatment arm and assumed effects in the untreated arm, as a way to estimate effects in observed and assumed compliers. Analytical approaches for continuous normally distributed outcome variables are growing in the literature, that can simultaneously account for clustering of the outcome, clustering of non-participation and factors associated non-participation. However, they require powerful software packages that can handle complex structural equation modelling of instrumental variables. The methodology is not easily understandable to, or applicable by, a general audience and it is highly unlikely that models fitted to relatively rare binary or rate outcomes would even converge.

A pragmatic two-step bootstrapping approach to obtain a complier average causal effect (CACE) was investigated, as way of using CRT data to estimate the *efficacy* of azithromycin in individuals on mortality with limited bias, whilst accounting for heterogeneity in the outcome and treatment data. This method utilised observed data in participators and non-participators in the MDA arm and overall data in the no MDA arm. Assuming similar mortality rates and person-time contributions for non-participators in the no MDA arm, to non-participators in the MDA arm, it was possible to compare the mortality rate in observed participators in the MDA arm, to the mortality rate in an assumed comparable group of participators in the no MDA arm (chapter 5).

The mortality data from PRET in The Gambia provided the opportunity to consider an application of this analytical approach to a heterogeneous health related outcome, as a pragmatic approach to an unbiased analysis of efficacy that could simultaneously account for clustering of mortality and non-participation.

The ITT and CACE analyses of all-age all-cause mortality did not provide any conclusive information about whether large-scale azithromycin MDA can lead to important reductions in mortality in The Gambia (chapter 5). All-age, all-cause mortality was low (7.4 per 1,000 person-years) and the PRET trial did not have pre-specified power, for a pre-specified effect size, for a difference in mortality between arms. In addition, although treatment status was only missing for 2% of individuals in the MDA arm, the overall mortality rate in those with missing treatment status was very high in the MDA arm (215 per 1,000 person-years). In addition to clustering of mortality and non-participation, there was significant clustering of missing treatment status, highlighting a new source of heterogeneity in the analysis of all-age all-cause mortality. A substantial contribution of deaths in those with missing treatment status in the MDA arm meant that this was also assumed to be the case for an assumed comparable group in the no MDA arm. This led to a substantial reduction in power to compare mortality between compliers (participators). Neither the ITT or CACE results provided evidence of a difference in mortality between arms, while per-protocol and as-treated analyses did show a significant reduction in mortality. The ITT and CACE rate ratios comparing MDA to no MDA were not in the direction of benefit (1.11 and 1.36 respectively). It thought that the association between missing treatment status and increased mortality is likely to have introduced bias in the CACE analysis and that the pragmatic approach to estimating efficacy, applied in this situation, did not adequately reduce any potential bias due to clustering of missing data, mortality or non-participation.

Estimating *efficacy* in a CRT of MDA interventions, with precision and without bias, remains a methodological challenge. Complete and accurate compliance data are required for this

pragmatic bootstrapping approach to estimate efficacy, especially as it is not possible to adjust for factors associated with non-participation with this method.

6.2. Research in context of other published literature

6.2.1. Factors associated with non-compliance during MDA

At the time of publication of the analysis of participation data in The Gambia in chapter 3 (Edwards *et* al, 2014[4]), this analysis of risk factors for non-participation in children in multiple MDA rounds in the same setting was one of very few studies in the context of trachoma; it appears to be the first study of geographical (spatial) clustering of repeated non-participation during azithromycin MDA and the first study to report on the different types of non-participators (PNT and EBA). Other studies of repeated non-participation and risk factors for non-participation in the context of trachoma were conducted using data from a parallel PRET CRT in Tanzania[5-7], also drawing similar conclusions around household level decision making with regard to participation. One of the Tanzanian studies also investigated factors associated with those whose participation status changed in repeated annual MDA rounds[6]. Results were suggestive of similar findings to increased odds of non-participation in each MDA round in The Gambia, with respect to hypotheses of geographical or social isolation and acceptance of public health interventions; in Tanzania, households with a change from participation in the first MDA to non-participation in the second annual MDA, were more likely to live further from the distribution site, have a guardian born outside the village with short-term residency and be assigned to a male community treatment assistant.

Prior to 2014, there was some focus on coverage during MDA rounds of azithromycin for trachoma by Cromwell and colleagues[8-10], however this was related more to reliable estimation and monitoring of coverage during MDA rounds, than to investigating individual, household and community risk factors in detail. The analysis in chapter 3 followed on from similar studies of risk factors for non-participation for other NTDs with MDA control, for example, lymphatic filariasis, onchocerciasis[11-15]. These studies also linked household level decision making to non-participation during MDA. Subsequent work by other researchers includes;

- an analysis of factors associated with receipt of treatment less than 3 times over the course of six annual MDA rounds of azithromycin for trachoma in Ethiopia, which found that non-participation was higher in older age groups which is likely similar to The Gambia[16].
- a modelling study of compliance behaviour over time in MDA rounds to predict the impact of this on elimination targets for soil-transmitted helminths[17]

- a systematic review by Shuford *et al* (2016)[18] of 112 studies of MDA for the five NTDs with MDA control (soil-transmitted helminth infections (hookworm, ascariasis, and trichuriasis), lymphatic filariasis, schistosomiasis, onchocerciasis, and trachoma) on reporting of coverage (those who receive the drug) and compliance (those who actually consume the drug).

An important gap in the literature exists regarding the relationship between infection status and non-participation. It has not yet been possible to study the impact of knowledge of infection status on non-participation as there is not yet a reliable rapid field test for C. *trachomatis* infection. What is needed is analysis of the extent of non-participation by infection status prior to MDA in different country and prevalence settings; there are unpublished data in existence from trials and surveys from multiple countries. In the Gambia, there were so few infections before MDA that non-participation in infected individuals would have had little to no bearing on trachoma control or the impact of MDA on an outcome of infection in a CRT analysis.

6.2.2. Consideration of non-compliance and efficacy during sample size determination for cluster randomised trials

There was no mention of sample size adjustment for non-participation or efficacy during the design of CRTs of azithromycin MDA interventions for trachoma control published between 2001-2018, included in the literature review in chapter 1[19-40]. A limited number of publications were identified that referred to consideration of non-participation (non-compliance) in relation to sample size estimation for CRTs;

- a systematic review of sample size methodology for cluster randomised trials by Rutterford, Copas and Eldridge (2015)[41]
- two publications by Lui and Chang in 2011[42,43] proposing sample size methods for cluster randomised trials accounting for non-compliance, for CRTs with either a non-inferiority and superiority design.

The systematic review of sample size methodology for CRTs also highlights that increased sample sizes are required to allow for non-compliance, that intervention effects are typically estimated in the presence of non-compliance in pragmatic CRTs and includes the proposed methods of Lui and Chang. The publications of Lui and Chang were the only ones included in the review relating to non-compliance and no other papers could be found relating to non-compliance during sample size estimation and design of CRTs. A crude approach to sample size adjustment for non-participation in infected individuals and efficacy, could be to consider a range of assumed average values for each of these parameters and use these assumptions to

adjust the expected outcome (e.g. prevalence, rate) in the intervention arm, when applying the simpler formulae of Hayes and Moulton[44] for example. It could be the case that this crude approach was considered during design of one or more of the CRTs for trachoma included in the literature review, there was just no mention of it in the publications.

In their formulae to calculate sample size for CRTs with non-compliance, Lui and Chang[42,43] incorporate efficacy in principle compliers and the proportion of non-compliers, two parameters described as part of the CACE estimation methods described in chapter 5. Both of these parameters were also considered during my simulation studies.

The simulation studies conducted as part of this thesis highlight the importance of a distinction between coverage in all individuals and coverage (participation) amongst infected individuals for sample size adjustment, because the latter group are the most relevant in terms of risk of reducing the impact of treatment. The gap in the literature relating to non-compliance in infected individuals, mentioned above, could be reduced by analysis of existing data for trachoma and other NTDs. In the PRET CRT in The Gambia, consumption of azithromycin was directly observed; community members were invited to a central location where field workers provided the single dose of azithromycin and recorded treatment status against a census list generated as part of the study. The systematic review by Shuford *et al* (2016)[18] highlighted that there can be a difference between coverage (those who receive the drug) and compliance (those who actually consume the drug) in MDA rounds. Whilst this is thought to be a negligible issue in the PRET CRT in The Gambia (coverage is analogous to compliance as compliance (participation) was directly observed), the conclusions of the systematic review are that data capture of coverage versus compliance is a potentially an underestimated issue of importance. Coverage versus compliance could therefore represent two required simulation parameters in simulation-based sample size estimation, although incorporation of directly observed treatment status in trial design and conduct would avoid a need for this additional layer of complexity.

The published literature still includes a greater breadth of methodology for accounting for non-compliance during analysis of cluster randomised trials, rather than design which is a relatively neglected topic. In a systematic review about reporting of non-adherence in cluster randomised trials, there was a summary of how many publications included an analysis of efficacy based on either a per-protocol or as-treated analysis, but no mention was made of sample size in the review[45].

6.2.3. Estimating efficacy of azithromycin in individuals

An application of the pragmatic bootstrap approach to estimation of the CACE of azithromycin for a primary outcome of C. *trachomatis* infection or other secondary outcomes, in CRTs of MDA for trachoma, does not appear to have been published by other researchers. The method did not work well for a very rare outcome and a rare compliance category, although it appears to have worked very well in a CRT with 20% prevalence of the outcome and approximately 50% non-compliance[46]. It is not yet known for which range of values for the prevalence of infection and non-participation, the bootstrap method for estimation of the CACE works well. This is the subject of post-doctoral simulation studies.

There are datasets in existence with higher prevalence of C. *trachomatis* infection at follow-up than in the PRET trial in The Gambia, for which it might be of interest to apply the bootstrap approach to estimation of the CACE. For example, the prevalence of infection was around 10% at 12-months follow-up in the control arm of a CRT in Ethiopia, where 24 communities were randomised to either annual MDA (control is baseline MDA only) or biannual MDA (intervention is MDA at baseline and 6-months post baseline)[31]. Treatment status (compliance) categories would be more complex than in a trial with a control arm of no MDA or an intervention administered just once, but available control arm data would allow investigation of the observed proportion of non-compliers at baseline and sensitivity analyses could include consideration of partial compliers in the biannual arm as either compliers or non-compliers.

There are secondary outcomes of interest in CRTs of azithromycin MDA for trachoma control that include mortality, malaria and nutritional status. For these outcomes, there would be a clear interest from investigators to estimate the efficacy in individuals in an unbiased analysis, as well as effectiveness from an ITT analysis.

Efficacy for an outcome of mortality is of particular interest, given the results of three studies suggesting a reduction in child mortality in ITT analyses in children in communities receiving azithromycin MDA compared to children in communities that did not[47]. In this published pooled analysis of the three studies, authors conclude that azithromycin MDA is a potential tool against child mortality in sub-Saharan Africa. The pooled rate ratio suggested a 14% reduction in the rate of child mortality (IRR = 0.86, 95% CI: 0.78 – 0.94) across three trials conducted in four countries. One of these trials was the MORDOR trial (Niger, Malawi, Tanzania)[48], in which clusters were randomised to either biannual azithromycin to children aged <12 years, or biannual placebo to the same age group (placebo control arm). The ITT effect was a mortality rate reduction of 14% (95% CI: 7 to 20%) over a two-year period. The

other two trials were CRTs of azithromycin MDA with trachoma as a primary outcome and mortality as a secondary outcome, conducted prior to the MORDOR trial;

- PRET trial (Niger)[49]: This trial did not have a control arm of placebo or no MDA, instead the comparison was biannual azithromycin to annual azithromycin (active control arm). The ITT effect was a mortality rate reduction of 19% (95% CI: 0 to 34%) in children aged 6-59 months.
- TANA trial (Ethiopia)[50]: This trial had a control arm of 12 communities with no MDA (nothing control arm), compared to 36 intervention communities receiving either annual MDA to all individuals, biannual MDA to all, or quarterly MDA to children aged 1-9 years (12 clusters allocated to each frequency of MDA combined into one intervention arm for the mortality analysis). The ITT effect was reduced odds of mortality of 49% (95% CI: 10 to 71%) in children aged 1-9 years.

These trials, with adequate power to detect an ITT effect, provide further opportunities to apply the bootstrap approach to estimate efficacy in individuals with respect to mortality.

Other examples of available datasets for analysis of secondary outcomes are; the PRET trial in Niger (annual vs biannual azithromycin MDA for three years with trachoma as the primary outcome), which included measurement of secondary outcomes of malaria parasitaemia and nutritional status at 36 months[51-53]; the PRET trial in The Gambia (three annual MDA rounds versus one at baseline), which included measurement of anthropometric indices at 36 months[54] and spleen size as a proxy for malaria at 30 months follow-up[55]. Sample sizes for these trials were based on primary outcomes for trachoma, not any secondary outcomes. Where an ITT effect was not detected, there will likely be too little power to detect a CACE.

Any application of the bootstrap method to other datasets, however, would need to include confirmation of the ability of the approach to reproduce the observed ITT effect accurately and normality of the sampling distribution for bootstrapped CACE values, in order to establish the likely validity of the approach for a specific dataset. The available data for control arms varies in the existing datasets from no MDA at all, to a reduced number of doses of azithromycin (active control arm), to a placebo treatment. Some control and intervention arms include more than one dose of azithromycin or placebo. It is possible to apply the bootstrap approach to each of these datasets but there will need to careful consideration of how compliers and non-compliers are categorised, whether treatment status was measured accurately and the extent of missing data. These datasets should also allow some investigation of whether the assumptions made during estimation of the CACE hold.

6.3. Strengths

A detailed analysis of spatial clustering, heterogeneity in non-participation and factors associated with non-participation was possible due to a large dataset of children aged 1-9 years old, plus information on their family members, even though the overall prevalence of non-participation was not high (chapter 3).

The simulation studies in chapter four highlight how important it is to consider non-participation and efficacy during CRT design of MDA interventions, especially given the apparent lack of attention paid to non-participation and efficacy during design of published CRTs for trachoma. The simulation studies also highlighted the lack of published information about non-participation in relation to individuals infected at baseline, i.e. those most at risk of infection at follow-up and those who pose a risk of transmission to uninfected individuals. This information could be informative, not only in trial design, but also for trachoma control in programmatic settings.

A study of the impact of heterogeneous non-participation specifically on power to detect effects in CRTs of MDA interventions does not appear to have been done before for trachoma or other neglected tropical infectious diseases with MDA interventions, in terms of implications for trial design and sample size. Published methodological literature appears to have rather been focussed on complex analysis methods to account for non-compliance in CRTs, after it has occurred and how to account for factors associated with non-compliance to try to regain lost power[79,136]. This work in this thesis may be informative to those planning CRTs who wish to ensure adequate power to estimate *population-level effectiveness* and attempt a secondary unbiased analysis of *efficacy*.

The analysis of mortality illustrated a pragmatic approach to estimate *efficacy* of azithromycin in those who take it, accounting for clustering of the outcome, non-participation and missing exposure data. This analytical approach is more accessible than complex structural equation modelling and is relatively straight forward to implement. This approach provides an opportunity to evaluate *efficacy* from existing field data (or *field efficacy*), where the prevalence of trachoma is higher or for other infectious diseases with mass treatment control interventions (chapter 5).

6.4. Limitations

Although the study of factors associated with non-participation was well powered, investigations were limited to a small number of factors measured as part of a CRT to evaluate trachoma outcomes. Only a small number of household level measures were available such as size, access to a latrine and water and also inclusion of children in trachoma surveys by way of random sampling as part of the trial design. It was not possible to accurately measure hypothesised factors such as poverty, ostracization, or to interview a selection of individuals to determine reasons for non-participation (chapter 3).

The simulation studies in chapter four were not able to produce precise results that would quantify exactly how much non-participation to allow for in the design of a CRT of MDA interventions, or a clear rule for how many additional clusters to include to account for non-participation. These studies did highlight however how non-participation can be considered in relation to infection status at baseline and how it could be taken into account during the design stage.

The simulation study of the impact of non-participation used a simple example of allocation of clusters to either MDA or no MDA, with a superiority design to test for a difference in prevalence of infection between arms. More complex designs that include multiple MDA rounds, MDA delivered at a different frequency between arms or a non-inferiority comparison between arms were not explored in simulation studies. However, a simple MDA vs no MDA superiority design was considered adequate to highlight the importance of including provision for non-participation on power to detect effects in CRTs, regardless of study design.

Missing data for treatment status was a major problem in the analysis of the *efficacy* of azithromycin for all-age all-cause mortality, especially as it was clustered and associated with the outcome (chapter 5). In the absence of missing compliance status data and a link between missing status and the outcome, it is not known from the analyses conducted how precise results from the pragmatic bootstrapping approach could be. Similarly, how precise this approach could be for trachoma outcomes with relatively low prevalence post-MDA is unknown. The prevalence of trachoma outcomes in the PRET trial in The Gambia very low at the time of the final follow-up survey (<1% for infection and below the 10% MDA threshold for TF), so the pragmatic approach to estimating the *field efficacy* of azithromycin was not applied to data for trachoma in children from the PRET trial. Extensive simulation studies could explore the usefulness of this approach as a method to estimate *efficacy* in CRTs of MDA interventions, in terms of bias and precision, for relatively rare outcomes and expected coverage levels. It could be the case that

the pragmatic bootstrapping approach can only work well if the prevalence of the outcome at follow-up post-intervention and the prevalence non-participation are large enough. Simulation studies could also investigate sample sizes required for precise results for an application of this approach to CRTs of mass treatment and contrast this to sample sizes required to detect expected effects in ITT analyses.

6.5. Wider applicability of results

MDA is the mainstay of control of five neglected tropical diseases (trachoma, schistosomiasis, lymphatic filariasis, soil transmitted helminths, onchocerciasis)[148]. Mass delivery of interventions to entire communities or schools also occurs for other infectious diseases, for example, malaria. Some vaccine studies utilise a cluster randomised design with no individual randomisation of individuals within clusters. It is plausible that any infectious diseases studied using a CRT design can be clustered, along with uptake of interventions to those offered it during MDA rounds. It is also possible that in field studies for these infectious diseases or interventions, there will be some missing data for participation status during MDA and that missingness may, or may not be correlated with baseline infection status and risk of the outcome at follow-up.

Therefore, the types of heterogeneity observed in the PRET trial in The Gambia and the implications of heterogeneity explored in this thesis could apply to any CRTs used to evaluate MDA or vaccine interventions.

6.6. Possible Expansion

A number of opportunities arise for expansion of the work in this thesis.

As already highlighted, data from existing CRTs of trachoma could be analysed to explore the extent of non-participation amongst individuals previously included in trachoma surveys with testing for *C. trachomatis* infection. It is likely that similar data also exist from CRTs of MDA interventions for other NTDs that could inform the design of any future CRTs for the five NTDs with MDA control.

Simulation studies could be conducted to further investigate implications of heterogeneity and methods to deal with heterogeneity in CRTs;

In relation to the work presented in this thesis, the implications of non-participation on power to detect ITT effects could be explored for a new CRT for trachoma control to be conducted in Ethiopia (https://www.lshtm.ac.uk/research/centres-projects-groups/stronger-safe), led by

researchers from London School of Hygiene and Tropical Medicine. The proposed CRT will investigate MDA delivery strategies as well as enhanced facial cleanliness and environmental intervention components of the SAFE strategy for trachoma control.

It would also be useful to investigate the validity of pragmatic bootstrapping approach more formally, with simulation studies, to determine when or if this approach can be reliably applied in low prevalence settings where non-participation is also relatively low due to coverage targets of at least 80%.

The CACE analysis to estimate *efficacy* of azithromycin in individuals could be applied to a recent CRT of azithromycin MDA that was specifically designed and powered for an ITT evaluation of child mortality[138].

6.7. Overall conclusions

Where non-participation is clustered but does not occur in individuals at risk of *C. trachomatis* infection, there is unlikely to be a detrimental effect on trachoma control or power for a CRT of MDA interventions, and efforts to improve coverage and individual participation during MDA rounds are not required.

Existing trial data in medium and high prevalence settings should be analysed to establish the extent of correlation between cluster level prevalence of infection and non-participation, and the extent of non-participation amongst individuals with infection in surveys conducted prior to MDA rounds. Further research could inform whether resources are required to increase participation, by investigating spatial hotspots of infection and non-participation.

Non-participation amongst individuals with infection at baseline in relation to coverage needs to be taken into account during the design of any CRTs of MDA intervention, to ensure an adequate number of clusters are included to retain a desired level of power. The additional number of clusters required to allow for non-participation will depend on the coverage and also whether the cluster mean percentage of infected individuals who do not receive treatment exceeds the cluster mean percentage of uninfected individuals who do not receive treatment (or overall cluster mean prevalence of non-participation).

A novel pragmatic bootstrapping approach to estimate the *efficacy* of treatment in those who receive it during MDA rounds, that simultaneously accounts for clustering of outcome data and

non-participation, may not be enough to provide adequate precision or sufficiently limit bias, especially in the presence of incomplete compliance data.

References

1. Burton MJ. Trachoma: an overview. *British medical bulletin* 2007; **84**(1): 99.
2. Taylor HR, Burton MJ, Haddad D, West S, Wright H. Trachoma. *The Lancet* 2014; **384**(9960): 2142-52.
3. World Health Organization. Trachoma. http://www.who.int/trachoma/epidemiology/en/ (accessed 14 June 2018). 2018.
4. Mabey DCW, Solomon AW, Foster A. Trachoma. *The Lancet* 2003; **362**(9379): 223-9.
5. Thylefors B, Dawson CR, Jones BR, West S, Taylor HR. A simple system for the assessment of trachoma and its complications. *Bulletin of the World Health Organization* 1987; **65**(4): 477.
6. World Health Organization, Department of Control of Neglected Tropical Diseases. Trachoma simplified grading card. SAFE documents. http://www.who.int/trachoma/resources/SAFE_documents/en/ (accessed 14 June 2018). 1987.
7. See CW, Alemayehu W, Melese M, et al. How reliable are tests for trachoma?--a latent class approach. *Investigative Ophthalmology & Visual Science* 2011; **52**(9): 6133-7.
8. Ramadhani AM, Derrick T, Macleod D, Holland MJ, Burton MJ. The relationship between active trachoma and ocular Chlamydia trachomatis infection before and after mass antibiotic treatment. *PLoS Neglected Tropical Diseases* 2016; **10**(10): e0005080.
9. Habtamu E, Wondie T, Aweke S, et al. Trachoma and relative poverty: a case-control study. *PLoS Neglected Tropical Diseases* 2015; **9**(11): e0004228.
10. Edwards T, Smith J, Sturrock HJ, et al. Prevalence of trachoma in unity state, South Sudan: results from a large-scale population-based survey and potential implications for further surveys. *PLoS Neglected Tropical Diseases* 2012; **6**(4): e1585.
11. Ngondi J, Reacher MH, Matthews FE, et al. Risk factors for trachomatous trichiasis in children: cross-sectional household surveys in Southern Sudan. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2009; **103**(3): 305-14.
12. Solomon AW. Trachoma control: a guide for programme managers: World Health Organization; 2006.
13. World Health Organization. Report of the first meeting of the WHO Alliance for the Global Elimination of Trachoma, 1997.
14. World Health Organization. WHO Alliance for the Global Elimination of Trachoma by 2020: progress report on elimination of trachoma, 2014–2016. *World Health Organisation Weekly epidemiological record* 2017; **92**: 357-68.
15. Evans JR, Solomon AW. Antibiotics for trachoma. *The Cochrane Library* 2011.
16. Baltussen RM, Sylla M, Frick KD, Mariotti SP. Cost-effectiveness of trachoma control in seven world regions. *Ophthalmic epidemiology* 2005; **12**(2): 91-101.
17. House JI, Ayele B, Porco TC, et al. Assessment of herd protection against trachoma due to repeated mass antibiotic distributions: a cluster-randomised trial. *The Lancet* 2009; **373**(9669): 1111-8.
18. Chidambaram JD, Melese M, Alemayehu W, et al. Mass antibiotic treatment and community protection in trachoma control programs. *Clinical infectious diseases* 2004; **39**(9): e95-7.
19. Gaynor BD, Miao Y, Cevallos V, et al. Eliminating trachoma in areas with limited disease. *Emerging infectious diseases* 2003; **9**(5): 596.
20. Melese M, Alemayehu W, Lakew T, et al. Comparison of annual and biannual mass antibiotic administration for elimination of infectious trachoma. *Journal of the American Medical Association* 2008; **299**(7): 778-84.
21. Gill DA, Lakew T, Alemayehu W, et al. Complete elimination is a difficult goal for trachoma programs in severely affected communities. *Clinical Infectious Diseases* 2008; **46**(4): 564-6.

22. Solomon AW, Harding-Esch E, Alexander NDE, et al. Two doses of azithromycin to eliminate trachoma in a Tanzanian community. *New England Journal of Medicine* 2008; **358**(17): 1870-1.
23. Chidambaram JD, Alemayehu W, Melese M, et al. Effect of a single mass antibiotic distribution on the prevalence of infectious trachoma. *Journal of the American Medical Association* 2006; **295**(10): 1142-6.
24. Gebre T, Ayele B, Zerihun M, et al. Comparison of annual versus twice-yearly mass azithromycin treatment for hyperendemic trachoma in Ethiopia: a cluster-randomised trial. *The Lancet* 2011.
25. Lakew T, Alemayehu W, Melese M, et al. Importance of coverage and endemicity on the return of infectious trachoma after a single mass antibiotic distribution. *PLoS Neglected Tropical Diseases* 2009; **3**(8): e507.
26. Lakew T, House J, Hong KC, et al. Reduction and return of infectious trachoma in severely affected communities in Ethiopia. *PLoS Neglected Tropical Diseases* 2009; **3**(2): e376.
27. World Health Organization. Report of the 3rd Global Scientific Committee on Trachoma, Johns Hopkins University, Baltimore, MA, 19–20 July 2010, 2010.
28. International Coalition for Trachoma Control. Trachoma Action Planning. A planning guide for the national elimination of blinding trachoma. http://www.trachomacoalition.org/sites/default/files/content/resources/files/ICTC%20TAP%20planning%20guide%20eng.pdf (accessed 14 August 2018), 2015.
29. Hollingsworth TD. Counting Down the 2020 Goals for 9 Neglected Tropical Diseases: What Have We Learned From Quantitative Analysis and Transmission Modeling? *Clinical Infectious Diseases* 2018; **66**(suppl_4): S237-S44.
30. Lietman T, Porco T, Dawson C, Blower S. Global elimination of trachoma: how frequently should we administer mass chemotherapy? *Nature medicine* 1999; **5**(5): 572-6.
31. 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. *Journal of Infectious Diseases* 2011; **204**(2): 268.
32. Ray KJ, Lietman TM, Porco TC, et al. When can antibiotic treatments for trachoma be discontinued? Graduating communities in three African countries. *PLoS Neglected Tropical Diseases* 2009; **3**(6): e458.
33. Ray KJ, Porco TC, Hong KC, et al. A rationale for continuing mass antibiotic distributions for trachoma. *BMC Infectious Diseases* 2007; **7**: 91.
34. Gebre T, Ayele B, Zerihun M, et al. Comparison of annual versus twice-yearly mass azithromycin treatment for hyperendemic trachoma in Ethiopia: a cluster-randomised trial. *The Lancet* 2012; **379**(9811): 143-51.
35. Amza A, Kadri B, Nassirou B, et al. A cluster-randomized trial to assess the efficacy of targeting trachoma treatment to children. *Clinical infectious diseases* 2016; **64**(6): 743-50.
36. Burton MJ, Holland MJ, Makalo P, et al. Profound and sustained reduction in Chlamydia trachomatis in the Gambia: a five-year longitudinal study of trachoma endemic communities. *PLoS Neglected Tropical Diseases* 2010; **4**(10): e835.
37. Yohannan J, Munoz B, Mkocha H, et al. Can we stop mass drug administration prior to 3 annual rounds in communities with low prevalence of trachoma?: PRET Ziada trial results. *JAMA Ophthalmology* 2013; **131**(4): 431-6.
38. West SK, Bailey R, Munoz B, et al. A randomized trial of two coverage targets for mass treatment with azithromycin for trachoma. *PLoS Neglected Tropical Diseases* 2013; **7**(8): e2415.
39. Harding-Esch EM, Sillah A, Edwards T, et al. Mass treatment with azithromycin for trachoma: when is one round enough? Results from the PRET Trial in the Gambia. *PLoS Neglected Tropical Diseases* 2013; **7**(6): e2115.
40. Oldenburg CE, Amza A, Kadri B, et al. Comparison of Mass Azithromycin Coverage Targets of Children in Niger: A Cluster-Randomized Trachoma Trial. *The American journal of tropical medicine and hygiene* 2018; **98**(2): 389-95.

41. Amza A, Kadri B, Nassirou B, et al. Effectiveness of expanding annual mass azithromycin distribution treatment coverage for trachoma in Niger: a cluster randomised trial. *British journal of ophthalmology* 2018; **102**(5): 680-6.
42. Solomon AW, Pavluck AL, Courtright P, et al. The Global Trachoma Mapping Project: Methodology of a 34-Country Population-Based Study. *Ophthalmic Epidemiology* 2015; **22**(3): 214-25.
43. Hayes JH, Moulton LH. Cluster Randomised Trials: Chapman & Hall/CRC Press. Taylor and Francis Group.; 2009.
44. Harding-Esch EM, Edwards T, Mkocha H, et al. Trachoma Prevalence and Associated Risk Factors in The Gambia and Tanzania: Baseline Results of a Cluster Randomised Controlled Trial. *PLoS neglected tropical diseases* 2010; **4**(11): e861.
45. Edwards T, Harding-Esch EM, Hailu G, et al. Risk factors for active trachoma and Chlamydia trachomatis infection in rural Ethiopia after mass treatment with azithromycin. *Tropical medicine & international health : TM & IH* 2008; **13**(4): 556-65.
46. Hägi M, Schémann JF, Mauny F, et al. Active trachoma among children in Mali: Clustering and environmental risk factors. *PLoS Neglected Tropical Diseases* 2010; **4**(1): e583.
47. Keenan JD, Ayele B, Gebre T, et al. Childhood mortality in a cohort treated with mass azithromycin for trachoma. *Clinical infectious diseases* 2011; **52**(7): 883.
48. Ssemanda EN, Munoz B, Harding-Esch EM, et al. Mass Treatment with Azithromycin for Trachoma Control: Participation Clusters in Households. *PLoS neglected tropical diseases* 2010; **4**(10): e838.
49. Harding-Esch EM, Edwards T, Mkocha H, et al. Trachoma prevalence and associated risk factors in the gambia and Tanzania: baseline results of a cluster randomised controlled trial. *PLoS Neglected Tropical Diseases* 2010; **4**(11): e861.
50. Ssemanda EN, Munoz B, Harding-Esch EM, et al. Mass treatment with azithromycin for trachoma control: participation clusters in households. *PLoS Neglected Tropical Diseases* 2010; **4**(10).
51. Jo B, Asparouhov T, Muthén BO. Intention-to-treat analysis in cluster randomized trials with noncompliance. *Statistics in medicine* 2008; **27**(27): 5565-77.
52. Jo B. Statistical power in randomized intervention studies with noncompliance. *Psychological methods* 2002; **7**(2): 178.
53. Ssemanda EN, Mkocha H, Levens J, Munoz B, West SK. Community mass treatment with azithromycin for trachoma: Factors associated with change in participation of children from the first to the second round. *Clinical Epidemiology and Global Health* 2013.
54. West SK, Munoz B, Mkocha H, et al. Treating village newcomers and travelers for trachoma: Results from ASANTE cluster randomized trial. *PloS one* 2017; **12**(6): e0178595.
55. Ervin A-M, Mkocha H, Munoz B, et al. Surveillance and azithromycin treatment for newcomers and travelers evaluation (asante) trial: design and baseline characteristics. *Ophthalmic epidemiology* 2016; **23**(6): 347-53.
56. Amza A, Kadri B, Nassirou B, et al. Community risk factors for ocular Chlamydia infection in Niger: pre-treatment results from a cluster-randomized trachoma trial. *PLoS Neglected Tropical Diseases* 2012; **6**(4): e1586.
57. Stare D, Harding-Esch E, Munoz B, et al. Design and baseline data of a randomized trial to evaluate coverage and frequency of mass treatment with azithromycin: the Partnership for Rapid Elimination of Trachoma (PRET) in Tanzania and The Gambia. *Ophthalmic Epidemiology* 2011; **18**(1): 20-9.
58. Stoller NE, Gebre T, Ayele B, et al. Efficacy of latrine promotion on emergence of infection with ocular Chlamydia trachomatis after mass antibiotic treatment: a cluster-randomized trial. *International health* 2011; **3**(2): 75.
59. Cumberland P, Edwards T, Hailu G, et al. The impact of community level treatment and preventative interventions on trachoma prevalence in rural Ethiopia. *International Journal Epidemiology* 2008; **37**(3): 549-58.

60. Edwards T, Cumberland P, Hailu G, Todd J. Impact of health education on active trachoma in hyperendemic rural communities in Ethiopia. *Ophthalmology* 2006; **113**(4): 548-55.
61. West SK, Emerson PM, Mkocha H, et al. Intensive insecticide spraying for fly control after mass antibiotic treatment for trachoma in a hyperendemic setting: a randomised trial. *The Lancet* 2006; **368**(9535): 596-600.
62. Holm SO, Jha HC, Bhatta RC, et al. Comparison of two azithromycin distribution strategies for controlling trachoma in Nepal. *Bulletin World Health Organisation* 2001; **79**(3): 194-200.
63. Moulton LH. Covariate-based constrained randomization of group-randomized trials. *Clinical Trials* 2004; **1**(3): 297.
64. Sismanidis C, Moulton LH, Ayles H, et al. Restricted randomization of ZAMSTAR: a 2× 2 factorial cluster randomized trial. *Clinical Trials* 2008; **5**(4): 316.
65. Raab GM, Butcher I. Balance in cluster randomized trials. *Statistics in Medicine* 2001; **20**(3): 351-65.
66. Ivers NM, Halperin IJ, Barnsley J, et al. Allocation techniques for balance at baseline in cluster randomized trials: a methodological review. *Trials* 2012; **13**(1): 120.
67. Carter B, Hood K. Balance algorithm for cluster randomized trials. *BMC Medical Research Methodology* 2008; **8**(1): 65.
68. Wright N, Ivers N, Eldridge S, Taljaard M, Bremner S. A review of the use of covariates in cluster randomized trials uncovers marked discrepancies between guidance and practice. *Journal of clinical epidemiology* 2015; **68**(6): 603-9.
69. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.; 2013.
70. StataCorp. Stata Statistical Software. College Station, Texas; 2013.
71. Kraschnewski JL, Keyserling TC, Bangdiwala SI, et al. Optimized probability sampling of study sites to improve generalizability in a multisite intervention trial. *Preventing chronic disease* 2010; **7**(1): A10.
72. Heo M, Leon AC. Statistical power and sample size requirements for three level hierarchical cluster randomized trials. *Biometrics* 2008; **64**(4): 1256-62.
73. Spybrook J, Bloom H, Congdon R, Hill C, Martinez A, Raudenbush S. Optimal Design for Longitudinal and Multilevel Research: Documentation for the Optimal Design Software Version 3.0. Available from www.wtgrantfoundation.org or from sitemaker.umich.edu/group-based. 2011.
74. Teerenstra S, Lu B, Preisser JS, Van Achterberg T, Borm GF. Sample Size Considerations for GEE Analyses of Three-Level Cluster Randomized Trials. *Biometrics* 2010.
75. Teerenstra S, Moerbeek M, Van Achterberg T, Pelzer BJ, Borm GF. Sample size calculations for 3-level cluster randomized trials. *Clinical Trials* 2008; **5**(5): 486-95.
76. Murray D. Design and Analysis of Group-Randomized Trials. New York, NY: Oxford University Press; 1998.
77. Sommer A, Zeger SL. On estimating efficacy from clinical trials. *Statistics in medicine* 1991; **10**(1): 45-52.
78. Little RJ, Long Q, Lin X. A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. *Biometrics* 2009; **65**(2): 640-9.
79. Jo B, Asparouhov T, Muthén BO, Ialongo NS, Brown CH. Cluster randomized trials with treatment noncompliance. *Psychological methods* 2008; **13**(1): 1.
80. Becque T, White I. Regaining power lost by non-compliance via full probability modelling. *Statistics in medicine* 2008; **27**(27): 5640-63.
81. Connell AM. Employing complier average causal effect analytic methods to examine effects of randomized encouragement trials. *The American journal of drug and alcohol abuse* 2009; **35**(4): 253-9.

82. Muthén LK, Muthén BO. Mplus User's Guide. Sixth Edition. Los Angeles, CA, 1998-2010.
83. Guo T. Causal effects in randomized trials in the presence of partial compliance: Breastfeeding effect on infant growth: MCGILL UNIVERSITY; 2009.
84. Agbla SC, DiazOrdaz K. Reporting non-adherence in cluster randomised trials: A systematic review. *Clinical Trials* 2018: 1740774518761666.
85. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. *British Medical Journal* 2012; **345**.
86. Courtright P, Rotondo LA, MacArthur C, et al. Strengthening the links between mapping, planning and global engagement for disease elimination: lessons learnt from trachoma. *The British journal of ophthalmology* 2018.
87. Solomon AW, Pavluck AL, Courtright P, et al. The Global Trachoma Mapping Project: Methodology of a 34-Country Population-Based Study. *Ophthalmic Epidemiology* 2015; **22**(3): 214-25.
88. Baker MC, Krotki K, Sankara DP, et al. Measuring treatment coverage for neglected tropical disease control programs: analysis of a survey design. *American journal of epidemiology* 2013; **178**(2): 268-75.
89. Harding-Esch EM, Edwards T, Sillah A, et al. Active trachoma and ocular Chlamydia trachomatis infection in two Gambian regions: on course for elimination by 2020? *PLoS Neglected Tropical Diseases* 2009; **3**(12): e573.
90. Worrell C, Mathieu E. Drug coverage surveys for neglected tropical diseases: 10 years of field experience. *American journal of tropical medicine and hygiene* 2012; **87**(2): 216-22.
91. World Health Organization. Sustaining the drive to overcome the global impact of neglected tropical diseases. Second WHO report on neglected tropical diseases, 2013.
92. Kyelem D, Biswas G, Bockarie MJ, et al. Determinants of success in national programs to eliminate lymphatic filariasis: a perspective identifying essential elements and research needs. *The American journal of tropical medicine and hygiene* 2008; **79**(4): 480.
93. World Health Organization. Accelerating work to overcome the global impact of neglected tropical diseases: A roadmap for implementation. http://www.who.int/neglected_diseases/NTD_RoadMap_2012_Fullversion.pdf (accessed 14 June 2018), 2012.
94. Blake IM, Burton MJ, Bailey RL, et al. Estimating household and community transmission of ocular Chlamydia trachomatis. *PLoS Neglected Tropical Diseases* 2009; **3**(3): e401.
95. Ssemanda EN, Levens J, Mkocha H, Munoz B, West SK. Azithromycin mass treatment for trachoma control: risk factors for non-participation of children in two treatment rounds. *PLoS Neglected Tropical Diseases* 2012; **6**(3): e1576.
96. Cromwell EA, King JD, McPherson S, et al. Monitoring of mass distribution interventions for trachoma in Plateau State, Nigeria. *PLoS Neglected Tropical Diseases* 2013; **7**(1): e1995.
97. Cromwell EA, Ngondi J, Gatpan G, et al. Estimation of population coverage for antibiotic distribution for trachoma control: a comparison of methods. *International health* 2009; **1**: 182-9.
98. Solomon AW. Trachoma control: a guide for programme managers: World Health Organization; 2006.
99. Garmin Limited. eTrex®. Garmin International Inc,Olathe, Kansas, USA; 2004.
100. Kulldorff M. SaTScan™ User Guide, version 9.0. 2010.
101. QGIS Development Team. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org. 2013.
102. 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.

103. Burton MJ, Holland MJ, Makalo P, et al. Re-emergence of Chlamydia trachomatis infection after mass antibiotic treatment of a trachoma-endemic Gambian community: a longitudinal study. *The Lancet* 2005; **365**(9467): 1321-8.
104. Shah NA, House J, Lakew T, et al. Travel and implications for the elimination of trachoma in ethiopia. *Ophthalmic Epidemiology* 2010; **17**(2): 113-7.
105. Montgomery MA, Desai MM, Groce NE, Elimelech M. Relationship between distance to social gathering facilities and risk of trachoma for households in rural Tanzanian communities. *Social science & medicine (1982)* 2011; **73**(1): 1-5.
106. Nuwaha F, Okware J, Ndyomugyenyi R. Predictors of compliance with community-directed ivermectin treatment in Uganda: quantitative results. *Tropical medicine & international health : TM & IH* 2005; **10**(7): 659-67.
107. Yirga D, Deribe K, Woldemichael K, Wondafrash M, Kassahun W. Factors associated with compliance with community directed treatment with ivermectin for onchocerciasis control in Southwestern Ethiopia. *Parasites & vectors* 2010; **3**: 48.
108. Njomo D, Amuyunzu-Nyamongo M, Mukoko D, Magambo J, Njenga S. Socioeconomic factors associated with compliance with mass drug administration for lymphatic filariasis elimination in Kenya: Descriptive study results. *Annals of Tropical Medicine and Public Health* 2012; **5**(2): 103.
109. World Health Organization. WHO vaccine-preventable diseases: monitoring system. 2013 global summary. http://apps.who.int/immunization_monitoring/globalsummary/coverages?c=GMB (accessed 18 August 2018), 2013.
110. Solomon AW, Engels D, Bailey RL, et al. A diagnostics platform for the integrated mapping, monitoring, and surveillance of neglected tropical diseases: rationale and target product profiles. *PLoS Negl Trop Dis* 2012; **6**(7): e1746.
111. Edwards T, Allen E, Harding-Esch EM, et al. Non-participation during azithromycin mass treatment for trachoma in The Gambia: heterogeneity and risk factors. *PLoS Negl Trop Dis* 2014; **8**(8): e3098.
112. Hardwick RA-OX, Truscott JA-O, Oswald WA-O, et al. Individual adherence to mass drug administration in neglected tropical disease control: A probability model conditional on past behaviour. *PLoS Negl Trop Dis* 2021; (1935-2735 (Electronic)).
113. Harding-Esch E, Jofre-Bonet M, Dhanjal JK, et al. Costs of testing for ocular Chlamydia trachomatis infection compared to mass drug administration for trachoma in the Gambia: application of results from the PRET study. *PLoS Neglected Tropical Diseases* 2015; **9**(4): e0003670.
114. Harding-Esch E, Edwards T, Sillah A, et al. Risk factors for active trachoma in The Gambia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2008; **102**(12): 1255-62.
115. Moerbeek MA-O, Schie SV. What are the statistical implications of treatment non-compliance in cluster randomized trials: A simulation study. (1097-0258 (Electronic)).
116. Liu F, Porco TC, Mkocha HA, et al. The efficacy of oral azithromycin in clearing ocular chlamydia: mathematical modeling from a community-randomized trachoma trial. *Epidemics* 2014; **6**: 10-7.
117. Rahman SA, West SK, Mkocha H, et al. The distribution of ocular Chlamydia prevalence across Tanzanian communities where trachoma is declining. (1935-2735 (Electronic)).
118. Smith JL, Sturrock Hj Fau - Olives C, Olives C Fau - Solomon AW, Solomon Aw Fau - Brooker SJ, Brooker SJ. Comparing the performance of cluster random sampling and integrated threshold mapping for targeting trachoma control, using computer simulation. (1935-2735 (Electronic)).
119. Feiveson A. Power by simulation. *Stata J* 2002; **2**(2): 107-24.
120. Barthelme S. Copulas made easy. http://www.r-bloggers.com/copulas-made-easy/; 2011.

121. Smart F. Generate rank correlated variables. http://www.econometricsbysimulation.com/2012/06/how-to-generate-variables-that-are-rank.html; 2012.
122. Agresti A. Dealing with discreteness: makingexact'confidence intervals for proportions, differences of proportions, and odds ratios more exact. *Statistical Methods in Medical Research* 2003; **12**(1): 3-21.
123. Agresti A, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician* 1998; **52**(2): 119-26.
124. Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. *Statistical science* 2001: 101-17.
125. Newby G, Hwang J, Koita K, et al. Review of mass drug administration for malaria and its operational challenges. *The American journal of tropical medicine and hygiene* 2015; **93**(1): 125-34.
126. Halloran ME, Struchiner CJ, Longini Jr IM. Study designs for evaluating different efficacy and effectiveness aspects of vaccines. *American journal of epidemiology* 1997; **146**(10): 789-803.
127. Brieger WR, Okeibunor JC, Abiose AO, et al. Characteristics of persons who complied with and failed to comply with annual ivermectin treatment. *Trop Med Int Health* 2012; **17**(7): 920-30.
128. Brieger WR, Okeibunor JC, Abiose AO, et al. Compliance with eight years of annual ivermectin treatment of onchocerciasis in Cameroon and Nigeria. *Parasites & vectors* 2011; **4**: 152.
129. Edwards T, Allen E, Harding-Esch EM, et al. Non-Participation during Azithromycin Mass Treatment for Trachoma in The Gambia: Heterogeneity and Risk Factors. *PLOS Neglected Tropical Diseases* 2014; **8**(8): e3098. doi: 10.1371/journal.pntd.0003098. eCollection 2014 Aug.
130. 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.
131. Rahman SA, West SK, Mkocha H, et al. The Distribution of Ocular Chlamydia Prevalence across Tanzanian Communities Where Trachoma Is Declining. *PLoS Negl Trop Dis* 2015; **9**(3): e0003682.
132. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Statistics in medicine* 2006; **25**(24): 4279-92.
133. Hayes R, Moulton LH. Cluster Randomised Trials: Chapman & Hall/CRC Press; 2009.
134. Ye C, Beyene J, Browne G, Thabane L. Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study. *BMJ open* 2014; **4**(6): e005362.
135. Bhaskaran K, Smeeth L. What is the difference between missing completely at random and missing at random? *International journal of epidemiology* 2014; **43**(4): 1336-9.
136. Sagarin BJ, West SG, Ratnikov A, Homan WK, Ritchie TD, Hansen EJ. Treatment noncompliance in randomized experiments: Statistical approaches and design issues. *Psychological methods* 2014; **19**(3): 317.
137. Porco TC, Gebre T, Ayele B, et al. Effect of Mass Distribution of Azithromycin for Trachoma Control on Overall Mortality in Ethiopian Children. *JAMA: the journal of the American Medical Association* 2009; **302**(9): 962.
138. Keenan JD, Bailey RL, West SK, et al. Azithromycin to Reduce Childhood Mortality in Sub-Saharan Africa. *New England Journal of Medicine* 2018; **378**(17): 1583-92.
139. Frangakis CE, Rubin Db Fau - Zhou X-H, Zhou XH. Clustered encouragement designs with individual noncompliance: bayesian inference with randomization, and application to advance directive forms. (1465-4644 (Print)).
140. Opondo C, Halliday K, Witek-McManus S, Allen E. Estimating intervention effect in cluster randomised controlled trials with non-compliance. Trials; 2017: BIOMED CENTRAL LTD 236 GRAYS INN RD, FLOOR 6, LONDON WC1X 8HL, ENGLAND; 2017.

141. Button KS, Ioannidis JP, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience* 2013; **14**(5): 365.
142. O'Brien KS, Cotter SY, Amza A, et al. Childhood Mortality After Mass Distribution of Azithromycin: A Secondary Analysis of The PRET Cluster-Randomized Trial in Niger. *The Pediatric infectious disease journal* 2018.
143. Jo B. Model misspecification sensitivity analysis in estimating causal effects of interventions with non-compliance. *Stat Med* 2002; **21**(21): 3161-81.
144. Field CA, Welsh AH. Bootstrapping clustered data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2007; **69**(3): 369-90.
145. Jasseh M, Webb EL, Jaffar S, et al. Reaching Millennium Development Goal 4–The Gambia. *Tropical Medicine & International Health* 2011; **16**(10): 1314-25.
146. Scott S, Odutola A, Mackenzie G, et al. Coverage and timing of children's vaccination: an evaluation of the expanded programme on immunisation in The Gambia. *PloS one* 2014; **9**(9): e107280.
147. Scott S, Kendall L, Gomez P, et al. Effect of maternal death on child survival in rural West Africa: 25 years of prospective surveillance data in The Gambia. *PloS one* 2017; **12**(2): e0172286.
148. Solomon AW, Engels D, Bailey RL, et al. A diagnostics platform for the integrated mapping, monitoring, and surveillance of neglected tropical diseases: rationale and target product profiles. *PLoS Neglected Tropical Diseases* 2012; **6**(7): e1746.
149. Burr SE, Sillah A, Sanou AS, et al. Cross-sectional surveys of the prevalence of follicular trachoma and Trichiasis in the Gambia: has elimination been reached? *PLoS Neglected Tropical Diseases* 2016; **10**(9): e0004906.

Appendices

Appendix 1. Ethical approval confirmation

OBSERVATIONAL/INTERVENTIONS RESEARCH ETHICS COMMITTEE

25 November

Tansy Edwards

Dear Tansy

Study Title:	**Analysis of data and design issues from PRET Gambia**
LSHTM ethics ref:	**6080**
Department:	**Epidemiology and Population Health**

Thank you for your application of 10 November for the above research, which has now been considered by the Chair on behalf of the Committee.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

The final list of documents reviewed and approved is as follows:

Document	*Version*	*Date*	
LSHTM ethics application	n/a	25/11/11	
Protocol	V1.0	25/11/11	
Information Sheet	V1.0	25/11/11	
Consent form	V1.0	25/11/11	

After ethical review

Any subsequent changes to the application must be submitted to the Committee via an E2 amendment form.

Yours sincerely,

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

OBSERVATIONAL/INTERVENTIONS RESEARCH ETHICS COMMITTEE

Professor Andrew J Hall
Chair