Malawi ICOSA Impact Monitoring Baseline & FY1 Results

Contents

1. Executive summary

Malawi is endemic for both *S. haematobium* and *S. mansoni*. Annual mass treatment of enrolled school children with praziquantel (PZQ) began in 2009, with the support of the Malawi Government, the WHO and World Vision. Treatment was initially carried out in 10 districts and scaled up to 18 districts by 2011, however, treatment and geographical coverage were limited due to availability of PZQ and financial restrictions.

Additional support to the National Schistosomiasis and Soil Transmitted Helminths (STH) Control Programme was provided with the UK Department for International Development (DFID) and the SCI under the Integrated Control of Schistosomiasis in Sub Saharan Africa (ICOSA) project in 2011. This initial five year support aims to assist scale-up to enable treatment of all school-age children in Malawi, through the provision of PZQ and funds to support essential programme activities such monitoring and evaluation.

The health impact of the national control programme on the infection and morbidity of schistosomiasis (SCH) and STH can be monitored through longitudinal surveys in combination with cross-sectional surveys. The monitoring surveys will include important baseline data collection and a series of follow-up surveys. A longitudinal survey design requires baseline data collection from schools prior to the initiation of large-scale distribution of praziquantel and albendazole or mebendazole through the school-based platform. Follow up surveys will be conducted immediately prior to subsequent rounds of treatment for the life of the programme to monitor the impact of the health intervention. Mapping of the country was conducted in 2012 and determined that all mapped districts were moderate to low prevalence and thus, in line with WHO recommendations for preventive chemotherapy, annual mass drug administration should be conducted every two years (biennial).

Baseline evaluation was carried out in March 2012 and lasted 5 weeks with the 1st follow up being done in March 2014 prior to the mass drug administration in April 2014. The sentinel sites are in the districts of Balaka, Blantyre, Chiradzulu, Lilongwe, Mwanza, N. and S. Mzimba, Neno, Ntcheu and Ntchisi in Malawi.

2. Field methodology

2.1. Study Design

Evaluation used concurrent 'cohort' (longitudinal) and 'cross-sectional' studies. The longitudinal study followed a cohort of randomly sampled primary Standard 1, 2 and 3 (roughly equivalent to 6, 7 and 8-year-olds, although there is variation in the ages of children in each standard in Malawi and generally in Africa) recruited at baseline, and measured again at follow-up; consequently, the age of children in the cohort study increased over time. The cross-sectional study recruits new Standard 1 pupils every year. The aim of the longitudinal study was to monitor prevalence, intensity and morbidity over the course of PCT rounds, while the aim of the cross-sectional study was to monitor levels of transmission. Decreased prevalence and intensity in the Standard 1 children (who we assume have not received treatment previously), could be due to a 'halo-effect' of treatment i.e. a reduction in the force of infection.

2.2. Sample Size

Statistical analysis by SCI has shown that to detect a reduction in *S. haematobium* intensity of 65% with 80% power requires an achieved sample of 22 schools, and 50 pupils per school. Assuming a 60% drop-out rate within each school over the entire course of the study, this implies a baseline sample of 2,750 pupils i.e. **22 schools, 125 pupils per school**. The schools monitored were randomly selected from a list of all schools in the appropriate districts. The baseline sample of 125 pupils per school was split roughly evenly between the three lowest Standards: a minimum of 40 Standard 1; 40 Standard 2 and 40 Standard 3. In the 1st follow-up data was only collected for those pupils that participated in the baseline study and no attempt to 'top-up' sample sizes to compensate for drop out was made. For the cross-sectional study, 40 new Standard 1s were recruited at the 1st follow-up.

2.3. Survey Methods

During the baseline survey, cohorts of 125 children from Standards 1, 2 and 3 (aged approximately 6, 7 and 8 years) were randomly selected in each of the schools and enrolled into the study. The children were measured for the following indicators:

i. *S. haematobium*: eggs per 10ml of urine
ii. *S. mansoni* and soil-transmitted helminths: number of eggs per gram of faeces using the Kato-Katz method (4 slides over 2 days).
iii. Haematuria: urine dipsticks
iv. Anaemia: finger prick blood sample and the use of a Hemacue to measure haemoglobin concentration
v. Wasting and stunting: height and weight
vi. Age and sex.

This group of selected children, now in standards 2, 3, and 4, as well as a new group of 40 Standard 1 children, were re-tested to measure the same indicators during the 1st follow-up. WHO approved standard operating procedures for sample collection were used at each time-point.

2.4. Method of sentinel site selection

SCI's protocol is to monitor only in those districts where prevalence of schistosomiasis is moderate or high i.e. SCI does not monitor in non-endemic or low prevalence districts where a full control program is not implemented. All districts except Mzuzu City surveyed in the mapping in February 2012 were determined to have moderate prevalence of *S. haematobium,* and consequently all districts except Mzuzu City were included in the selection of sites to be monitored for this species (see Table 1).

S. mansoni infection was more focal and only present at moderate prevalence in Chiradzulu, Blantrye Rural, Lilongwe City and Lilongwe Rural East. Following district stratification by *S. mansoni* infection, such that the number of schools selected for *S. mansoni* monitoring, reflected the frequency of moderate risk areas in the monitoring areas, 22 schools were selected that would be monitor *S. haematobium* infection with a subset of 9 schools which also monitor *S. mansoni* infection. Due to the low prevalence of STH's, STH infection was only monitored in those schools where the Kato-Katz slides were already prepared for *S. mansoni*.

Mapping Area	*S. mansoni* prevalence (%) (95% confidence interval)	*S. mansoni* Risk	*S. haematobium* prevalence (%) (95%CI)	S.h. risk
Balaka	0.59 (0.00-1.49)	Low	14.55 (10.21-18.89)	Moderate
Blantyre City	1.37 (0.41-2.32)	Low	10.17(6.83-13.51)	Moderate
Blantyre Rural	7.28 (0.42-14.15)	Moderate	26.32 (18.98-33.67)	Moderate
Chiradzulu	6.46 (0.00-15.67)	Moderate	31.76 (21.81-41.71)	Moderate
Lilongwe City	8.24 (4.23-12.24)	Moderate	7.64 (4.28-11.0)	Moderate
Lilongwe Rural East	13.50 (2.46- 24.54)	Moderate	9.56 (4.58-14.55)	Moderate
Lilongwe Rural West	2.00 (0.01-3.99)	Low	11.97 (5.85-18.10)	Moderate
Mwanza & Neno (joint analysis)	1.90 (0.00-4.78)	Low	17.76 (6.52-29.00)	Moderate
Mzimba North	0.51(0.00-1.08)	Low	7.90 (2.44-13.36)	Moderate
Mzimba South	0.68 (0.00-1.77)	Low	10.54 (5.15-15.93)	Moderate
Mzuzu City	2.83 (0.50-5.17)	Low	2.17 (0.55-3.79)	Low

Table 1. District prevalence from 2012 mapping

3. Issues from the field

The protocol was for *S. haematobium* detection was for the sample to be first assessed for blood in the urine using haemastix. There were some issues with data collection in the field where sometimes a field wasn't filled in with a zero to indicate where a negative result should have been entered. For instance, if a child had no entry for haemastix, but had a negative entry for the urine filtration result, then it is likely that the child tested negative with haemastix but this was not recorded correctly. Consequently, we had to use some logic decisions when deciding on the likely *S. haematobium* infection status and infection intensity. The sequential rules we used to decide on infection status were:

1. If no information was recorded for haemastix or urine filtration then we did not assign an infection status and took the data to be truly missing.
2. If the number of eggs recorded in the urine filtration was zero then the infection status was taken to be negative. No reference was made to the haemastix result.
3. If the number of eggs recorded in the urine filtration was greater than zero then the infection status was taken to be positive. No reference was made to the haemastix result.
4. If the haemastix result was recorded (positive or negative) but the urine filtration result was missing we assumed that the urine filtration was negative but mistakenly was not recorded as such.

And the sequential rules we used to decide on infection intensity (ep10ml) were:

1. If infection status was missing then infection intensity was also missing
2. If infection status was negative (0) then infection intensity was 0.

3. If infection status was positive then infection intensity = 10 * number of eggs recorded/volume of urine recorded.

During follow-up, finding the children assessed at baseline proved challenging. In some cases the schools had closed or the children dropped out. As the teams had only two days to sample each school, follow-up rates were low for some schools. Further details are provided in later sections.

4. Data cleaning notes

4.1.Timeline

	Baseline	Y1 data	Y2 data	Y3 data	Y4 data
Data collection data	March - 12	March - 14			
Data entry began	April - 12	April - 14			
Data entry finished	Aug - 13	May - 14			
Raw data received by SCI	Aug- 13	May – 14			
Data cleaning began	Oct - 13	June – 14			
Data cleaning completed	Aug - 14	Aug - 14			
Analysis began	Aug - 14	Aug - 14			
Analysis completed	Sept - 14				

4.2.Cleaning information

- The baseline data was initially received but due to either a bug in the system or a virus, it became corrupt and could not be extracted. It was not until JW travelled to Malawi June 2013 that all the data was sorted through and finally scanned to FF to be sent for entering by DAMCO.
- As discussed above, for the baseline data it was difficult to interpret missing data compared to negative data.
- There were 2630 unique individuals recorded other variable information below.
- We changed baseline volumes that were recorded as greater than 10ml to 10ml as the teams recorded total volume rather than volume filtered.
- Follow up data was received in good time and in the correct format. There were some issues with the ID's of the schools that needed to be sorted but description is included in the information sheet.
- An outline of the data and all other issues can be found in a word doc with locations of raw, cleaned and the code used for this *'Malawi sentinel sites code and data information.docx'.*

4.3.Additional information

We used Anthroplus (http://www.who.int/growthref/tools/en/) software within R in order to estimate each child's BMI and relative height, weight and BMI for their age and gender, expressed as a z-score. As per the WHO recommendations, we excluded observations of height and/or weight outside of normal expected ranges for the child's age and gender (i.e. greater than 6 standard deviations from the mean in either direction; 15 observations were excluded). We followed WHO definitions of stunting and wasting:

Stunted	=	1	when z-score of height-for-age < -2
	=	0	otherwise
Wasted	=	1	when z-score of weight-for-height < -2
	=	0	otherwise

We also used the Haemoglobin measurements in order to determine whether a child was anaemic. As per the WHO recommendations (http://www.who.int/vmnis/indicators/haemoglobin.pdf) we used the below definition of anaemia. No adjustment was made to account for altitude.

Gender and age	Limit below which child was defined as anaemic (g/l)
Children 6 - 59 months of age	110
Children 5 - 11 years of age	115
Children 12 - 14 years of age	120
Non-pregnant women (15 years of age and above)	120
Men (15 years of age and above)	130

5. Description of dataset

5.1. Number of children and basic characteristics sampled - Baseline

Below is a table of the total number of children who were sampled at each school. The final number used for the various analyses differs due to the incorrect recording of information as discussed above or missing information. Overall we captured information from 1523 children at *S. haematobium* schools and 1119 in the schools monitored for both *S. haematobium* and *S. mansoni* during baseline. Overall there were 221 more boys sampled than girls.

Baseline information

District	School Name	Monitor for Sh	Monitor for Sm	M&E ID code	Total at Baseline
Balaka	Mgoloka FP	Yes	No	001-021	124
Blantyre City	Young Ambassadors	Yes	No	002-021	107
Lilongwe Rural West	Mguwata	Yes	No	007-021	120
Lilongwe Rural West	Likuni Boys Primary	Yes	No	007-022	125
Lilongwe Rural West	Msokoneza JP	Yes	No	007-023	125
Mwanza	Mphete	Yes	No	008-021	94
Neno	Lisungwi	Yes	No	008-022	124
Mzimba South	Kanyerere JP	Yes	No	010-021	103
Mzimba South	Etchiyeni	Yes	No	010-022	124
Mzimba South	Kasangazi	Yes	No	010-023	122
Ntcheu	Kambuku I JP	Yes	No	012-023	120
Ntcheu	Njolomole	Yes	No	013-021	122
Ntcheu	Kamtsitsi JP	Yes	No	013-022	113
	Total 13				**1,523**
Blantyre Rural	Mbame	Yes	Yes	003-022	125
Blantyre Rural	Chaweta	Yes	Yes	003-021	125
Chiradzulu	Namachete	Yes	Yes	004-021	125
Lilongwe City	Amazing Grace Pvt	Yes	Yes	005-021	124
Lilongwe Rural East	Mlombwa	Yes	Yes	006-021	125

Lilongwe Rural East	Chimanazo	Yes	Yes	006-022	125
Lilongwe Rural East	Ngala	Yes	Yes	006-023	125
Ntchisi	Chikhota	Yes	Yes	012-021	120
Ntchisi	Mpamila	Yes	Yes	012-022	125
	Total 9				**1119**

5.2. Number of children and basic characteristics sampled – Follow up

The table below displays the total children followed for the cohort and the cross sectional in each school and the diagnostics carried out. The final number used for the various analyses differs due to the incorrect recording of information as previously mentioned above or due to missing information. Overall we followed up 1458 students in the cohort and 846 new students from the cross-sectional. Mean age for standards did increase as you move through the standards, however, the range of ages in all standards is roughly the same for the cohort. Once again there was a higher number of boys in both samples (cohort =136, cross sectional=39).

Follow up numbers in cohort and cross section and total numbers successfully collected for diagnostic.

School	Cohort				Cross Sectional			
	Total	Dipstick	Sm	Sh	Total	Dipstick	Sm	Sh
Amazing Grace	39	34	0	34	40	39	0	39
Chaweta Lea	71	71	64	71	40	40	0	40
Chikhota	51	32	51	32	38	38	38	38
Chimanazo	71	65	71	65	39	39	38	39
Etchiyeni	88	88	0	88	41	41	0	41
Kambuku I	55	55	0	55	43	43	0	43
Kamtsitsi	70	69	0	69	40	40	0	40
Kanyerere	73	73	0	73	34	34	0	34
Kasangazi	76	76	0	76	42	42	0	42
Likuni Boys	87	87	0	87	40	40	0	40
Lisungwi	75	71	0	71	40	40	0	40
Mbame	46	46	46	46	39	39	39	39
Mgoloka FP	94	92	0	92	43	42	0	42
Mguwata	64	64	0	64	40	40	0	40
Mlombwa	45	5	0	5	40	10	2	10
Mpamila	30	28	30	28	39	37	39	37
Mphete	68	68	0	68	30	30	0	30
Msokoneza	72	72	0	72	41	41	0	41
Namachete	90	90	90	90	39	39	39	39
Ngala	42	9	14	9	38	6	38	6
Njolomole	82	81	0	81	40	40	0	40
Young Ambass	69	68	0	68	20	20	0	20
Total	**1458**	**1344**	**366**	**1344**	**846**	**780**	**233**	**780**

6. Statistical Methodology

We provide in the results section a table of descriptive statistics for each schistosomiasis and STH species, firstly overall and then split into schools monitored for *S. haematobium* and schools monitored for *S. mansoni* (with three schools being in both tables). The tables show:

Statistic	Description and notes
N schools included in analysis	The number of schools analysed for the focal species
N pupils with data	The number of pupils analysed for the focal species
Prevalence (and 95% CI)	The arithmetic mean prevalence across all pupils with data for the focal species. 95% confidence intervals were calculated using the 'survey' package in R, and the 'logit' method to account for the binary structure of the data. No adjustment was made for the proportion of children sampled in a school, or the proportion of schools sampled in the study.
Mean intensity (and 95% CI)	The arithmetic mean intensity of infection across all pupils with data for that species. 95% confidence intervals were calculated using the 'survey' package in R, assuming a Normal error distribution. No adjustment was made for the proportion of children sampled in a school, or the proportion of schools sampled in the study.
Proportion of schools with infection	Each school was classified as either having infection for the focal species or not having infection for the focal species. The proportion of schools with infection was then calculated.
School prevalence range	The prevalence was calculated for each school separately using all pupils with data for the focal species. The table shows the prevalence in the schools with the lowest and highest prevalence respectively.
Proportion of pupils heavily infected (and 95% CI)	The arithmetic mean prevalence of heavy infections (defined in the data cleaning notes) across all pupils with data for the focal species. 95% confidence intervals were calculated using the 'survey' package in R, with no adjustment made for the proportion of children sampled in a school, or the proportion of schools sampled in the study.

Overall prevalence by school was calculated using the 'survey' package in R. Each school was analysed separately using the 'logit' method to account for the binary structure of the data. During the calculation of 95% confidence intervals, no adjustment was made for the proportion of children sampled in a school.

Intensity by school was calculated using the 'lmer' package in R. A single generalised negative binomial linear model was fitted where school was included as a fixed factor. When back-transformed to the natural scale, the model produces the arithmetic mean of all infection intensities within a school and the associated 95% confidence intervals. No adjustment was made for the proportion of children sampled in a school.

We investigated gender and age differences in *S. haematobium* and *S. mansoni* prevalence separately, including only schools that were monitored for that species. Additionally, to ensure generality of results, we included only age and sex combinations with at least 20 pupils from at least 5 schools. We fitted a binomial mixed model with fixed terms of age (as a factor), sex and the associated interaction; school was fitted as a random effect. The significance of fixed effect terms was tested using log-likelihood ratio tests comparing models without and without the respective term. Where the age by sex interaction was non-significant, the significance of the main effect terms was assessed in a separate model with the interaction removed.

The significance of changes in schistosomiasis between baseline and follow-up was assessed using generalised linear models. For prevalence, a logistic regression with binomial errors was used, while for infection intensity

the errors were assumed to follow a negative binomial distribution. In all models age, sex, year and school were included as fixed factors and the significance of the year term was assessed using log-likelihood ratio tests comparing models with and without the year term.

Drop-out was investigated through the use of a logistic regression on the baseline data only, where the response variable was whether or not the pupil dropped out of the study. As in the previous models, age, sex, school were included as fixed factors. However, in the drop-out models we included an additional term of infection status as a binary variable or infection intensity as a linear term and the significance of these terms was assessed using log-likelihood ratio tests comparing models with and without the infection term.

7. Results

7.1. Summary table of descriptive statistics at Baseline

Baseline parasite and prevalence summary table

All schools - Baseline	*Any schistosome infection*	*S. haematobium*	*S. mansoni*	*Ascaris*	*Hookworm*	*Trichuris*
N schools included in analysis	22	22	9	9	9	9
N pupils with data	2613	2600	1091	1091	1091	1091
Prevalence (and 95% CI)	**10.50%** (6.7 - 16.2)	**10%** (6.4 - 15.3)	**1.90%** (0.6 - 5.6)	**0.10%** (0 - 0.6)	**0.20%** (0 - 1.1)	**0%** (0 - 0)
Mean intensity (epg or eggs/10ml) (and 95% CI)	n/a	4.3 (1 - 7.6)	1.9 (0 - 3.7)	0.1 (0 - 0.3)	0.3 (0 - 1)	0 (0 - 0)
Proportion of schools with infection	86.40%	81.80%	44.40%	11.10%	0%	11.10%
School prevalence range	0% - 36.3%	0% - 31.3%	0% - 10.9%	0% - 0.8%	0% - 1.7%	0% - 0%
Proportion of pupils heavily infected (and 95% CI)	n/a	**0.0%** (0-0)	**0.0%** (0-0)	**0.0%** (0-0)	**0.0%** (0-0)	**0.0%** (0-0)

All schools	*Stunting*	*Wasting*	*Anaemia*
N schools included in analysis	22	22	22
N pupils with data	2573	2559	2630
Prevalence (and 95% CI)	**0.10%** (0 - 0.3)	**0.30%** (0.1 - 0.5)	**25.80%** (20.7 - 31.7)
Proportion of schools with positive	13.6%	31.8%	100%
School positive range	0%-0.9%	0%-1%	3.3%-56.7%

7.2. Summary table of descriptive statistics at Follow Up - Cohort data

Follow up parasite and prevalence summary table

All schools – Follow up	*Any schistosome infection*	*S. haematobium*	*S. mansoni*	*Ascaris*	*Hookworm*	*Trichuris*
N schools included in analysis	22	22	7	7	7	7
N pupils with data	1379	1344	366	364	365	363
Prevalence (and 95% CI)	**5.20%** (2.9 - 9.3)	**5.40%** (2.9 - 9.7)	**0.30%** (0.1 - 1.3)	**0%** (0 - 0)	**0.80%** (0.2 - 3.6)	**0%** (0 - 0)
Mean intensity (epg or eggs/10ml) (and 95% CI)	n/a	1.6 (0.4 - 2.8)	0.1 (0 - 0.3)	0 (0 - 0)	0.2 (0 - 0.5)	0 (0 - 0)
Proportion of schools with infection	54.50%	54.50%	14.30%	0%	28.60%	0%
School prevalence range	0% - 27.5%	0% - 44.4%	0% - 1.1%	0% - 0%	0% - 7.7%	0% - 0%
Proportion of pupils heavily infected (and 95% CI)	n/a					

All schools – Follow Up	*Stunting*	*Wasting*	*Anaemia*
N schools included in analysis	21	21	22
N pupils with data	1398	1398	1384
Prevalence (and 95% CI)	0.10% (0 - 1)	0.60% (0.2 - 1.5)	28% (19.9 - 37.8)
Proportion of schools with positive	4.80%	19%	100%
School positive range	0% - 3%	0% - 4.8%	2.9% - 70.6%

7.3. Summary table of descriptive statistics at Follow Up - Cross-sectional data

All schools - FY Cross Sectional	*Any schistosome infection*	*S. haematobium*	*S. mansoni*	*Ascaris*	*Hookworm*	*Trichuris*
N schools included in analysis	22	22	7	7	7	7
N pupils with data	814	780	233	233	233	233
Prevalence (and 95% CI)	6.90% (3.1 - 14.5)	7.20% (3.2 - 15.2)	0% (0 - 0)	0% (0 - 0)	0.90% (0.1 - 6.4)	0% (0 - 0)
Mean intensity (epg or eggs/10ml) (and 95% CI)	n/a	2.1 (0.2 - 4)	0 (0 - 0)	0 (0 - 0)	0.4 (0 - 1.3)	0 (0 - 0)
Proportion of schools with infection	45.50%	45.50%	0%	0%	14.30%	0%
School prevalence range	0% - 50%	0% - 66.7%	0% - 0%	0% - 0%	0% - 100%	0% - 0%
Proportion of pupils heavily infected (and 95% CI)	n/a					

All schools - FY Cross Sectional	*Stunting*	*Wasting*	*Anaemia*
N schools included in analysis	21	21	22
N pupils with data	689	689	843
Prevalence (and 95% CI)	0.70% (0.1 - 5)	5.40% (1.8 - 15.1)	31.80% (23.4 - 41.6)
Proportion of schools with positive	4.80%	28.60%	90.90%
School positive range	0% - 33.3%	0% - 90%	0% - 76.3%

7.4. Parasitological results by species

We provide here a brief summary of results for each species, and associated graphs. For the baseline data, we present the results for each school, whereas for the follow-up data we present the results for the cohort and cross-sectional data separately. Detailed examination of differences in schistosomiasis over time is in the subsequent section (section 8), and detailed tables can be found in the associated spreadsheet which also includes information on intensity, nutrition and anaemia '*Malawi results.xlsx*'.

The graphs below for each species show A. baseline figures followed by B. follow up cohort and cross sectional for each species.

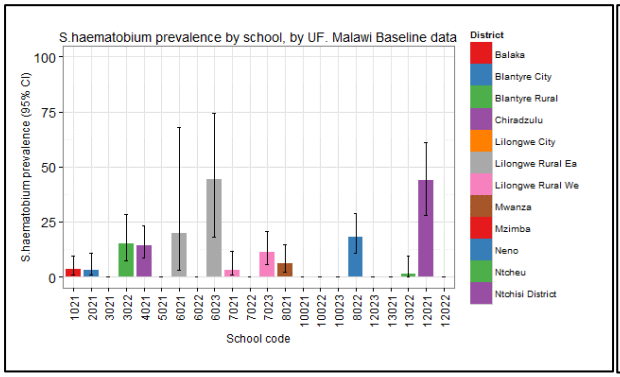
(i) **Bar charts** show the prevalence of infection at each school with associated 95% confidence intervals. For schistosomiasis, the bars are coloured by district. For STH's the bars are coloured by district with the colours from left to right on the graph corresponding to the key read from top to bottom.

(ii) **Dot plots** show the intensity of infection for each child as points on the graph, with the mean and associated 95% confidence intervals overlaid in black. The points for each school are shown in different colours to enable easy discrimination between schools.

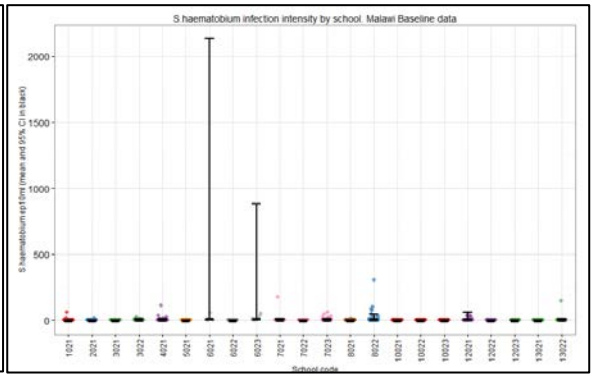
7.4.1. S. haematobium

At baseline, there were four schools where no *S. haematobium* was found and two schools where prevalence was greater than 25%. Mean intensity was highest in schools with higher prevalence however schools in the south of the country in Chiradzulu and Blantyre Rural which have comparable prevalence's to Lilongwe Rural East, Neno and Ntchisi (Central and Central/South Malawi) were found to have greater mean intensity 18.8 eggs/10ml, (95%CI 7.7-46.1) and mean =30.8 eggs/10ml (95%CI 12.3-76.8) respectively. Mwanza and Ntchisisi district both had one school which reports an increase in prevalence among the children who were followed up, however all schools show a reduction in intensity. As expected, the prevalence in cross sectional Standard 1 children was higher in all schools.

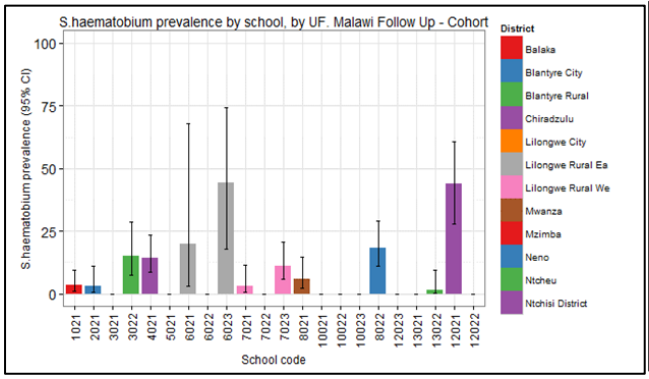
A. (i)



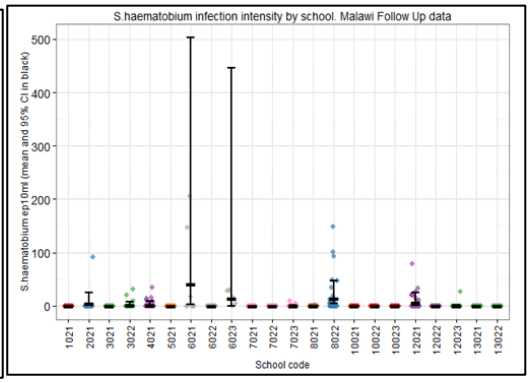
(ii)



B. (i)



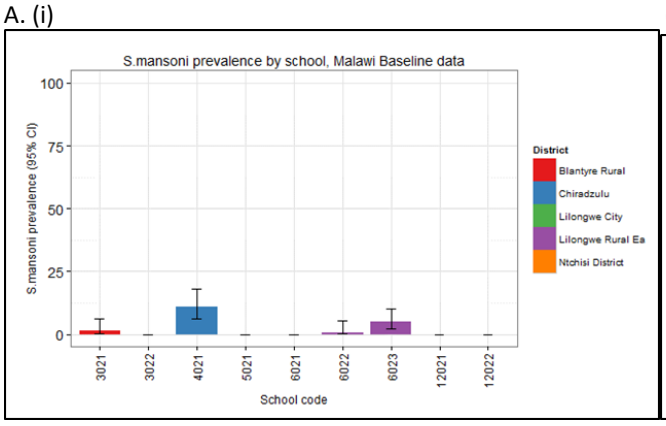
(ii)



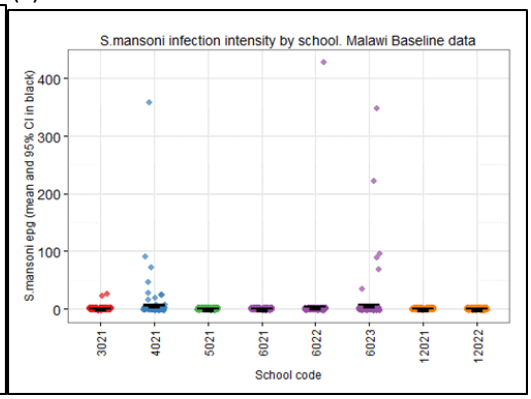
7.4.2. S. mansoni

At baseline and follow-up, prevalence of *S. mansoni* was much lower than prevalence of *S. haematobium* with only four schools at baseline having any positive children, and only one school with prevalence greater than 10%. At follow up, only one child was assessed as being positive for *S. mansoni*. The positive child at follow-up was in the cohort data and therefore there were no children in the cross-sectional data positive for *S. mansoni* at follow-up; consequently we have not included a graph here showing all zero prevalence.

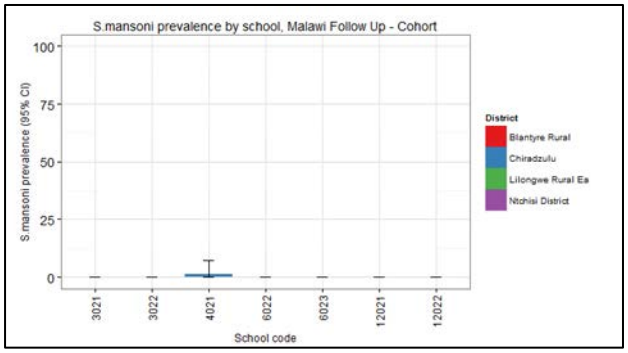
A. (i)



(ii)



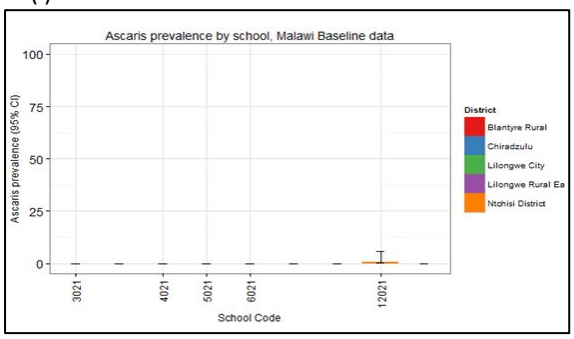
B. (i)



7.4.3. A. lumbricoides

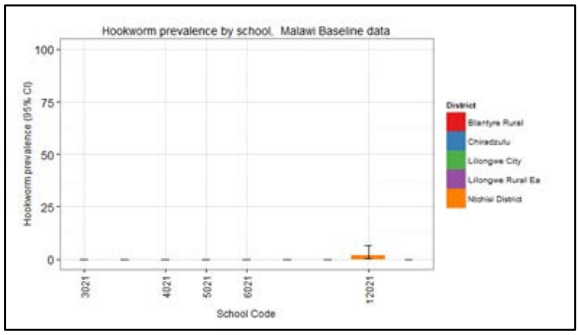
There was only one *A. lumbricoides* infection recorded as baseline, and no *A. lumbricoides* infection recorded at follow-up. Consequently we do not include graphs at follow-up here.

A. (i)

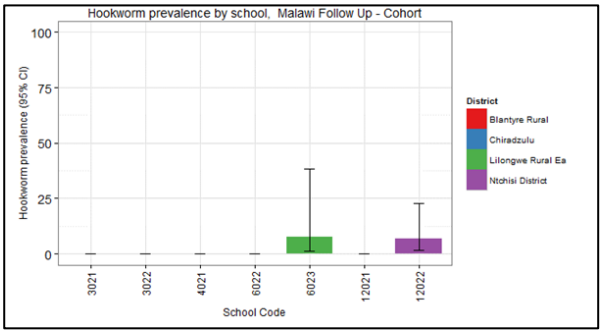


7.4.4. Hookworm

A. (i)



B. (i)



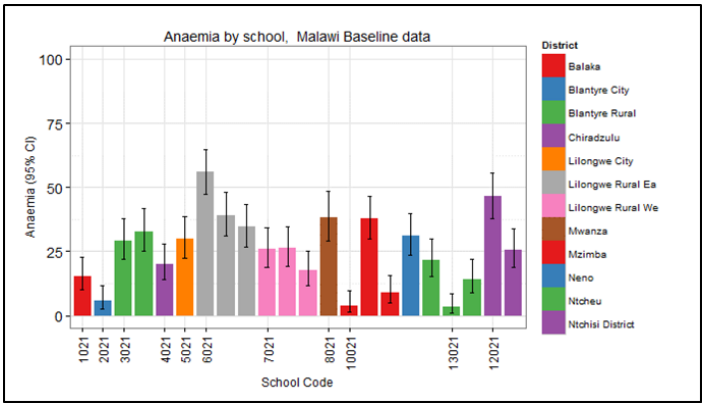
Hookworm showed a slight increase from baseline to follow up, with two infections detected at baseline and five detected at follow-up. However, due to the very low prevalence and intensity to begin with this may be simply due to sampling variability of daily egg presence in stool.

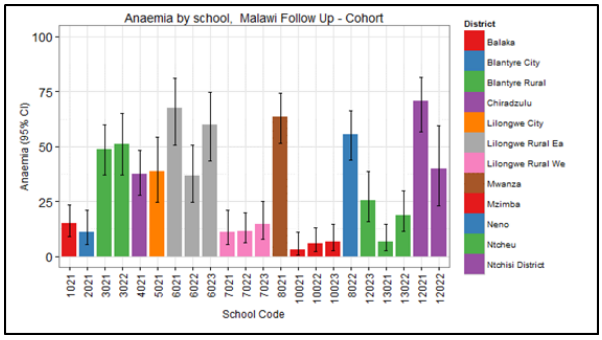
7.4.5. T. trichiura

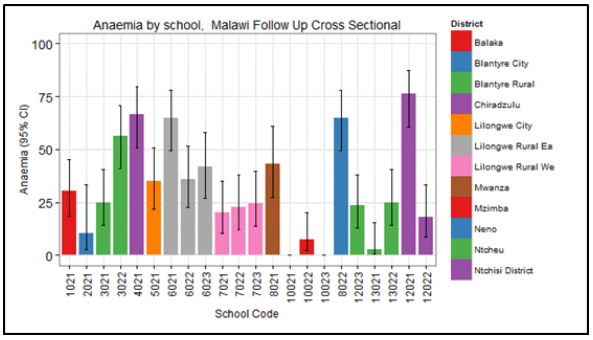
There were no *T. trichiura* infections detected at either baseline or follow-up. Consequently, we have not included graphs here.

7.5. Anaemia

At baseline, anaemia was present in all schools with prevalence ranging from 3.3% to 56%. Unfortunately, anaemia appears to have increased in all schools from baseline to follow up. Prevalence in the new Standard one students is slightly higher than the cohort children with prevalence of 31% and 28% respectively.




Anaemia by school, Malawi Follow Up - Cohort
Anaemia (95% CI)
100
75
50
25
0
1021
2021
3021
3022
4021
5021
6021
6022
6023
7021
7022
7023
8021
10021
10022
10023
8022
12023
13021
13022
12021
12022
School Code
District
Balaka
Blantyre City
Blantyre Rural
Chiradzulu
Lilongwe City
Lilongwe Rural Ea
Lilongwe Rural We
Mwanza
Mzimba
Neno
Ntcheu
Ntchisi District


Anaemia by school, Malawi Follow Up Cross Sectional
Anaemia (95% CI)
100
75
50
25
0
1021
2021
3021
3022
4021
5021
6021
6022
6023
7021
7022
7023
8021
10021
10022
10023
8022
12023
13021
13022
12021
12022
School Code
District
Balaka
Blantyre City
Blantyre Rural
Chiradzulu
Lilongwe City
Lilongwe Rural Ea
Lilongwe Rural We
Mwanza
Mzimba
Neno
Ntcheu
Ntchisi District

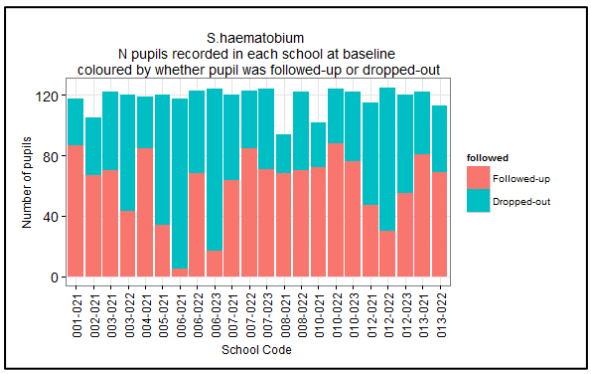
8. Investigating schistosomiasis prevalence and intensity over time

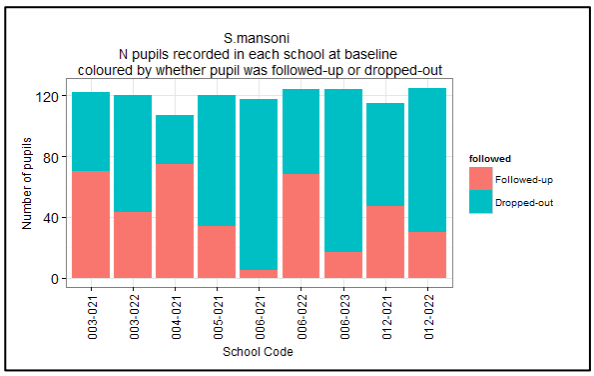
We investigated the changes in schistosomiasis over time using four separate datasets. Each dataset contained information from 22 schools and included only pupils where data for at least one schistosomiasis species was recorded.

1. The 'baseline' dataset contained details all off pupils sampled at baseline (standards 1-3), with an additional flag to indicate whether they were followed-up two years later (1st follow-up) or dropped-out of the study. Analysis of the baseline datasets enabled us to explore differences at baseline between those children that dropped-out and those children that were followed-up in the study. The baseline dataset contained information from 2,630 pupils.
2. The 'cohort' dataset contained details of all pupils sampled (standards 1-3) at baseline, and a subset of these same children at 1st follow-up in standards 2-5. Analysis of the cohort dataset enabled us to investigate changes in schistosomiasis prevalence among those children that were treated as part of the MDA. The cohort dataset contained information from 2,630 pupils at baseline, and 1,372 pupils at follow-up.
3. The 'followed-up' dataset is a subset of the cohort dataset and contains only information on those children that were measured both at baseline and at 1st follow-up. Comparison of the cohort and followed-up dataset enabled us to investigate those children that dropped-out of the study between baseline and follow-up. The followed-up dataset contained information from 1,372 pupils at baseline and at follow-up.
4. The 'cross-sectional' dataset contains details of pupils in standard 1 at baseline and standard 1 at follow-up and consequently are not the same pupils. Note that pupils in standards 1 at baseline were included in all four datasets. The cross-sectional dataset enabled us to investigate changes in schistosomiasis prevalence among children who have not been previously treated as part of the MDA, and may therefore serve as an indication of prevalence in the wider community. The cross-sectional dataset contained information from 870 pupils at baseline at 804 pupils at follow-up.

8.1. Investigating drop-outs between baseline and follow-up

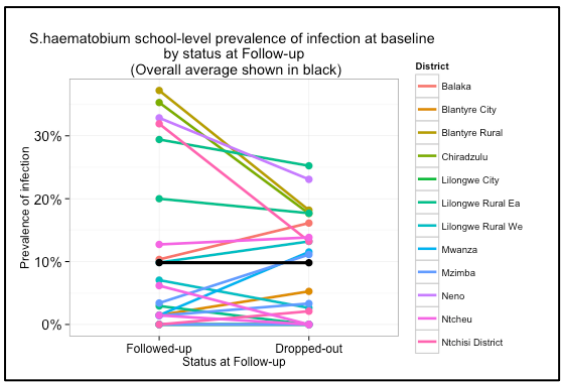
Due to *S. mansoni* being monitored at a subset of schools only, we investigated dropouts for each species separately. The graphs below show the dropout rates for each school separately. Overall, the drop-out rate was higher than expected for both species of schistosomiasis. 48% of those pupils monitored for *S. haematobium* dropped out the study between baseline and follow-up, and 64% of those pupils monitored for *S. mansoni* dropped out of the study. However, as the graphs show, there was considerable variation between schools in drop-out rate. The sample size calculation prior to data collection assumed a 60% drop-out over the course of the study and not in a single year. Consequently the results may suffer from a lack of power to detect differences in prevalence and intensity in only those children followed-up.

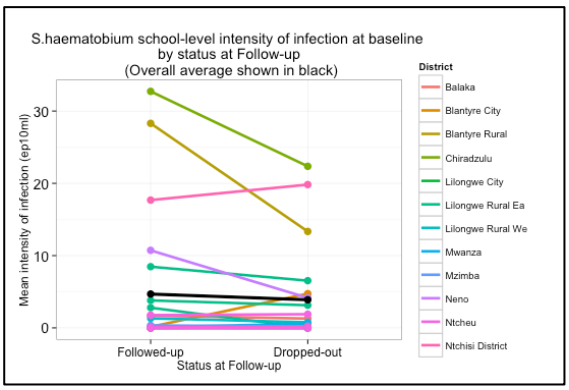




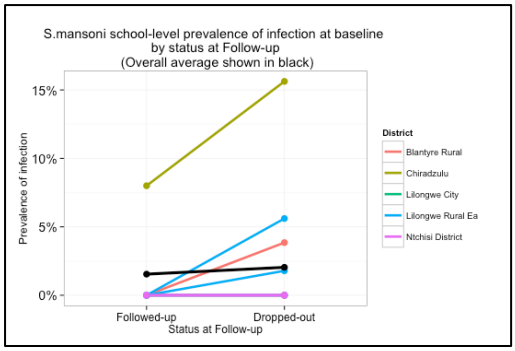
The table below details prevalence and mean intensity for each schistosomiasis species at baseline, and is split by whether the pupils were also recorded at follow-up or dropped out of the study. The graphs in the associated spreadsheet show the difference in prevalence and infection intensity for each school, with the overall differences shown in black. Note that there were a number of schools with no infections recorded and consequently there are a number of observations on the horizontal line at zero.

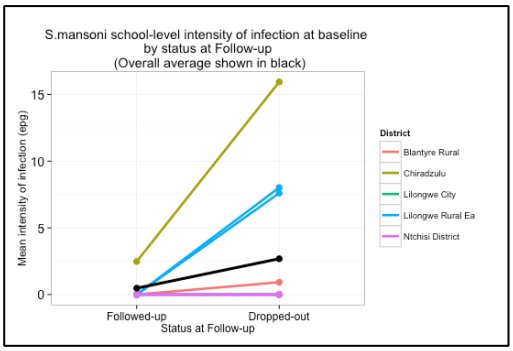
Drop-out comparison	Status at FY1	Schistosomiasis Species	
		Sh	Sm
# pupils at baseline	followed-up	1352	389
	dropped-out	1243	686
# pupils infected at baseline	followed-up	133	6
	dropped-out	122	14
Prevalence at baseline (95% CI)	followed-up	9.84% (5.67% - 16.54%)	1.54% (0.28% - 7.92%)
	dropped-out	9.81% (6.29% - 15.0%)	2.04% (0.76% - 5.37%)
% higher dropped-out than followed-up		-0.3%	32.5%
p-value of difference		0.025	0.024
Mean ep10ml/epg at baseline (95% CI)	followed-up	4.68 (0.32 - 9.05)	0.47 (0 - 1.34)
	dropped-out	3.89 (1.21 - 6.57)	2.69 (0 - 5.52)
% higher dropped-out than followed-up		-16.9%	472.3%
p-value of difference		0.327	0.008





For *S. haematobium* the prevalence of infection was slightly, but significantly (p = 0.025), higher in those pupils that were followed-up compared to those pupils that dropped out, but there was no evidence that infection intensity differed between those children who were followed-up and dropped-out (p=0.327). Consequently, our data does not suggest that those pupils that were most infected were more likely to drop-out of the study.





In contrast to *S. haematobium*, there was evidence that children with *S. mansoni* infection were more likely to drop-out of the study than be followed up. 70% of the 20 pupils that were infected at baseline dropped out of the study, compared to 64% of the uninfected pupils at baseline, and formal statistical analysis showed that the prevalence (p=0.024) and mean intensity of infection (p=0.008) was higher in those children measured at baseline that later dropped-out of the study. However these numbers are based on very small numbers of infected individuals and consequently apparent large percentage differences may be caused by a very small number of pupils. We will consider this finding in further detail when discussing changes over time in *S. mansoni.*

8.2. Changes in *S. haematobium* between baseline and follow-up

The table below shows the changes in *S. haematobium* prevalence and intensity between baseline and follow-up for each of the three datasets containing baseline and follow-up data. For those children that received MDA following baseline data collection, there was more than a 40% decrease in infection prevalence and more than a 60% decrease in infection intensity for both the cohort and followed-up dataset. However, these differences were significant for the cohort dataset but marginally non-significant for the followed-up dataset. Given that the magnitude of the effect is similar in both models, the non-significance in the second model is likely due to reduced power in the second model due to the smaller samples size when we consider only those pupils that were followed-up.

For those pupils in standard 1 only (cross-sectional data), there was a non-significant 4% reduction in prevalence but a significant 58% reduction in infection intensity. Consequently there is some evidence that the program may be having effects in the wider community beyond that due to treatment of school-aged children.

S. haematobium over time	Survey	Dataset		
		Cohort	Followed-up	Cross-sectional
# pupils	baseline	2595	1296	862
	F1	1315	1296	767
# schools with infection	baseline	18	17	13
	F1	12	12	10
# pupils infected	baseline	255	119	67
	F1	69	69	57
Prevalence (95% CI)	baseline	9.83% (6.31% - 14.99%)	9.18% (5.18% - 15.77%)	7.77% (4.10% - 14.23%)
	F1	5.25% (2.83% - 9.52%)	5.32% (2.85% - 9.72%)	7.43% (3.41% - 15.44%)
Percentage change		-46.6%	-42.0%	-4.4%
p-value of difference		0.005	0.068	0.147

Infection Intensity - Mean ep10ml (95% CI)	baseline	4.30 (0.97 - 7.63)	4.36 (0 - 8.79)	4.99 (0.36 - 9.61)
	F1	1.55 (0.36 - 2.74)	1.57 (0.36 - 2.78)	2.09 (0.26 - 3.93)
Percentage change		-64.0%	-64.0%	-58.0%
p-value of difference		0.008	0.098	<0.001

8.3. Changes in *S. mansoni* between baseline and follow-up

In the follow-up study, there was only one child of 354 examined that tested positive for *S. mansoni*, and consequently drawing robust conclusions from the data is not straightforward. There was a significant 85% reduction in *S. mansoni* prevalence (p<0.001) when all pupils in the cohort were examined. However, the 83% reduction in prevalence observed when analysis was restricted to followed-up children only was non-significant (p=0.484) which, given that the magnitude of the effect is similar in the cohort and followed-up data, may be due to a three-fold reduction in the sample size at baseline when analysis is restricted only to those pupils that were followed-up.

In the cross-sectional study of standard 1 pupils only, no infections were observed at follow-up compared to 8 infections at baseline. Consequently there is evidence that the program may be having effects in the wider community beyond that due to treatment of school-aged children.

***S. mansoni* over time**	**Survey**	**Dataset**		
		Cohort	**Followed-up**	**Cross-sectional**
# pupils	baseline	1075	333	351
	F1	354	333	228
# schools with infection	baseline	4	1	3
	F1	1	1	0
# pupils infected	baseline	20	6	8
	F1	1	1	0
Prevalence (95% CI)	baseline	1.86% (0.64% - 5.29%)	1.80% (0.34% - 8.82%)	2.28% (0.78% - 6.37%)
	F1	0.28% (0.06% - 1.35%)	0.30% (0.06% - 1.53%)	0.00% n/a
Percentage change		-84.8%	-83.3%	-100.0%
p-value of difference		<0.001	0.475	<0.001
Infection Intensity - Mean ep10ml (95% CI)	baseline	1.89 (0 - 3.78)	0.56 (0 -1.56)	4.09 (0.09% - 8.08%)
	F1	0.10 (0 - 0.278)	0.11 (0 - 0.30)	0.00 n/a
Percentage change		-94.6%	-80.6%	-100.0%
p-value of difference		<0.001	0.484	<0.001

Investigating potential impact of non-random drop-out

To gain an understanding of the potential impact of the non-random drop-out from the study between baseline and follow-up we performed some exploratory analysis. The single pupil that tested positive for *S.*

mansoni at follow-up did not test positive at baseline. However, this pupil was in school code 4021 that had the highest recorded *S. mansoni* prevalence at baseline where 11 pupils tested positive.

Of the 20 children infected at baseline, only 6 were identified at follow-up, although none of these children tested positive for *S. mansoni*. Consequently, a 'worst-case' scenario is that all 14 of the pupils that dropped-out remained infected at follow-up. Additionally, of the 1,055 children uninfected at baseline only 383 (36%) were followed-up. A reasonable estimate might be that if all of these 1,055 children had been followed-up then we might have observed a total of 3 new infections (1 pupil recorded as being newly infected divided by 36%). This population-level calculation is likely an overestimate as the single pupil that tested positive at follow-up came from a school with 72% follow-up so we might expect prevalence at follow-up to be closer to 1.4 children (1* 1/72%). Therefore, an estimated 'worst-case' scenario is for 17 (14 + 3) infected pupils at follow-up if all pupils remained in the study. Although this 15% reduction in prevalence is a lot smaller than the 85% observed, it provides some confidence that the reduction in prevalence observed was not simply due to non-random drop out from the study. Furthermore, if we arbitrarily assume a 50% reduction in prevalence in the children infected as baseline then we would expect 10 pupils to be infected at follow-up which equates to an overall reduction in prevalence of 50%. Additional evidence of a true drop in prevalence comes from the cross-sectional data where there was a significant reduction in prevalence and where issues of drop-out are not of concern.

9. Discussion

In these analyses we found evidence of positive impact of MDA within those children treated and in the wider community for both species of schistosomiasis.

Drop out from the study between baseline and follow-up was larger than expected, at 48% for *S. haematobium*, and 64% for *S. mansoni* compared to an expected 60% dropout over the course of the study. If these dropout rates were to be maintained for the next follow-up then we would expect only 27% of the original sample of *S. haematobium* to be recorded and just 13% of the original population recorded for *S. mansoni.* A high drop-out rate will have two likely effects: firstly, if the most heavily infected children are most likely to drop out of the study then an overall reduction in prevalence and intensity may be observed simply due to the drop-out. Although there are statistical methods to deal with non-random dropout, the analysis is not in any way trivial. Furthermore, analysis of this sort is always an approximation and a full dataset is always preferable. Secondly, the smaller sample sizes will lead to a reduction in power to be able to detect a true difference of a constant amount. Consequently, it is more likely that no significant difference in prevalence and intensity will be detected, regardless of whether or not there was a true effect in the population. An alternative to cohort sampling is cross-sectional sampling where the children selected each year have the same characteristics in each year (i.e. the same school, standard and gender) but are not the same children, and therefore dropout is not an issue. Cohort sampling is more appropriate when the objective is to assess the efficacy of treatment in individual children, whereas cross-sectional sampling may be more appropriate when the objective of sampling is assessing impact in the community. Given that the efficacy of PZQ has been clearly demonstrated previously, cross-sectional sampling may be more appropriate than cohort sampling for measuring programmatic impact. SCI is currently in the process of transitioning to cross-sectional sampling for those programs that have not yet started monitoring, and will be developing sampling strategies for those countries such as Malawi that have already performed cohort sampling.

For *S. haematobium* we found evidence of over 40% decrease in prevalence and over a 60% decrease in infection intensity between baseline and follow-up, despite prevalence at baseline likely being underestimated. These differences were significant for the cohort data but the differences were not significant when the data was restricted only to those pupils that were followed up, likely due to the smaller sample sizes in the latter sample. However, in the cross-sectional data there was no significant difference in prevalence over time, although there was a significant reduction in infection intensity. Importantly, there was no evidence

that the pupils with the heaviest infection were more likely to drop out of the study between baseline and follow-up.

Similarly to *S. haematobium*, differences seen in *S. mansoni* between baseline and follow-up in the cohort data was significant whereas similar differences observed in the follow-up data was not significant, likely due to the differences in sample sizes. However, only one infected individual was observed at follow-up compared to 6 pupils at baseline that were later followed-up. We also performed some calculations to demonstrate that the non-random drop-out observed for this species was likely not the sole cause of the reduction in prevalence observed. Further evidence of the impact of the program comes from the cross-sectional pupils where the number of infected pupils dropped from eight to zero, with no corresponding issues of drop-out.