Figure 1. Logic Model for the effects of community deworming.
Reproduced with permission from Taylor-Robinson 2015 [6].



Table 1. Characteristics of the base trials and the long-term follow-up studies

|  |  |  |
| --- | --- | --- |
| **Study ID****(versions)** | **Base trial** | **Follow-up study** |
| **Study ID** | **Country** | **Population** | **Number randomized****(clusters)** | **Intervention** | **Population** | **Data collection** | **Sample size** | **Timing** | **Difference in deworming exposure between study groups** |
| **Baird series**(2010, 2011a, 2011b,2012, 2015, 2016) | Miguel & Kremer 2004 | Kenya | School children aged between 6 and 18 years1 | 32 5652(75) | Deworming3 every 6 months at school, plus health promotion | Adults aged 19 to 26 years who participated in the base trial as children | Questionnaire survey | 5084 | 9 to 11 years after base trial started | 2.4 additional years of deworming |
| **Ozier series**(2011, 2014, 2015, 2016) |  |  |  |  |  | Children aged 8 to 15 years who now attend the base trial schools, but were too young at the time of the trial to have participated | Field survey  | 21 309 for height and weight;2371 for cognitive assessment | 11 to 12 years after base trial started | Exposure to the ‘spill-over’ effects of deworming during the first year of life |
| **Croke 2014**(2014) | Alderman 2006 | Uganda | Pre-school children aged 1 to 7 years | 27 995(50) | Deworming4 every 6 months at child health days (CHD) | Children aged 6 to 16 years who live in the area of the base trial and may have participated as children | Large-scale survey unrelated to base trial | 763 | 10 to 11 years after base trial started | Less than 2 additional doses of deworming tablets  |

1 In Miguel & Kremer 2004 girls aged 13 years or older were not intended to receive the drug intervention due to potential drug teratogenicity. However, some did receive deworming treatment.

2 Miguel & Kremer 2004 was a quasi-randomized trial utilizing sequential allocation.
3 In Miguel & Kremer 2004 deworming medication was given as albendazole every 6 months (600 mg in 1998 and 400 mg in 1999) plus praziquantel at 40 mg/kg annually. It is estimated that 72% of children in the intervention and 5% of children in control groups received this.
4 In Alderman 2006 deworming medication was given as albendazole 400 mg every 6 months. It is estimated that the deworming coverage increased from 21.7% before the intervention started in 2000 to 65.8% in 2003 in the intervention group, and from 23.9 to 34.6% in the control group (according to a cluster survey of households in all parishes, including 750 households in each group).

Table 2. Risk of bias assessments for the base trials

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Study ID** | **Selection bias** | **Reporting and detection bias** | **Attrition bias** | **Other biases** |
| **Sample selection** | **Confounding** | **Blinding of outcome assessors** | **Blinding of data analysis** | **Contamination** | **Co-intervention** |
| **Miguel 2004** | **HIGH RISK**- Systematic allocation (non-random)- Subsamples described as “random” but no details given | **UNCLEAR RISK**Groups broadly similar according to comparison of variables at baseline, but missing data to assess and confirm it | **HIGH RISK**Not blinded | **UNCLEAR RISK**Blinding not described | **HIGH RISK**- No clusters were lost- Considerable missing data for all outcomes. | **LOW RISK**- Deworming coverage of 5% in the control group- Transfer rate into a different school between 2% and 8%, with similar proportions among the 3 groups | **HIGH RISK**- Worm prevention education through regular public health lectures, wall charts, and training of teachers1- Other school-based interventions simultaneously in 27/75 project schools |
| **Alderman 2006** | **LOW RISK**Cluster randomised controlled trial | **LOW RISK**Balanced baseline characteristics | **HIGH RISK**Not blinded | **HIGH RISK**Not blinded | **LOW RISK**Two clusters were lost | **HIGH RISK**Children dewormed in 2003: 65.8% in intervention group, 34.6% in the control group | **LOW RISK** None |

1 Some may view this as part of the intervention, but current global policy advocates drug distribution, not intensive school health education

Table 3. Risk of bias assessments of the long-term follow-up studies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Study ID** | **Selection bias** | **Reporting and detection bias** | **Attrition bias** | **Selective reporting** |
| **Sample selection** | **Confounding** | **Blinding of outcome assessors** | **Blinding of data analysis** |
| **Baird series** | **LOW RISK**- Computer generated random sampling from the eligible population, stratified by school, grade, and gender | **UNCLEAR RISK**- Age and academic performance prior to base trial appeared similar, but other potential confounders not presented- Uncertain risk of confounding due to the quasi-randomised design of the base trial | **LOW RISK**- Outcome assessors were unaware of how treatment would be defined in the analysis | **HIGH RISK**- Not blinded | **LOW RISK**- 2/75 clusters not included in the analysis.- Effective tracking rate of 82.7% | **HIGH RISK**- No a-priori analytic plan- Multiple significance testing- Inconsistency of outcome reporting over time- Post-hoc sub-group analyses presented as main results in the abstract - Important findings of no effect not reported in abstract |
| **Ozier series** | **LOW RISK**- Computer generated random sampling from eligible population1 | **UNCLEAR RISK**- Data on potential confounders are not provided separately for intervention and control groups- Only two cohorts (of seven) contain relevant randomized comparisons. Additional analyses of the whole sample are at uncertain risk of confounding due to secular trends 1 | **LOW RISK**- Outcome assessors were unaware of how treatment would be defined in the analysis | **HIGH RISK**- Not blinded | **UNCLEAR RISK**- Around 28% of sample excluded as they had migrated into the area after the base trial.- Migration out of the area, which would represent missing data, is not well quantified.- 2/75 clusters not included in the analysis. | **UNCLEAR RISK**- No a-priori analytic plan- Important finding of no effect on height not reported in abstract until the 2016 version. Data on weight not reported at all. |
| **Croke 2014** | **UNCLEAR RISK**-Selection of villages described as ‘random‘ but methods not specified -Selection of households within villages by systematic selection | **UNCLEAR RISK****-** Some confounders (access to water and private education) appear unbalanced | **LOW RISK**- Data were collected through a larger survey conducted for other reasons and unrelated to the base study | **HIGH RISK**- Not blinded | **HIGH RISK**- 28/50 clusters not included in analysis- Numeracy and literacy test outcomes available for 710/763 children (6.9% missing data)- Potential migration out of the area not addressed | **UNCLEAR RISK**- No a-priori analytic plan |

1 Ozier series: Of the seven annual cohorts, none of the children born in 1995 or 1996 lived in areas with active deworming programmes in the first year of life, whilst all the children born in 2001 did. Analyses across all seven cohorts therefore represent a mixture of randomised and observational data.

Table 4. Assessment of selective reporting in Baird 2011a

|  |  |  |
| --- | --- | --- |
| **Policy important domains** | **Abstract** | **Tables and appendices** |
| **Number of outcomes reported as a beneficial** | **Number of outcomes reported as no effect**  | **Number of outcomes reported with P < 0.05** | **Number of outcomes reported with P >0.05** |
| **Nutritional status** | 0 | 0 | 0 | 3 |
| **Physical well-being** | 1 | 0 | 2 | 0 |
| **School enrolment and attendance** | 1 | 0 | 1 | 2 |
| **School performance and tests of cognition** | 1 | 0 | 1 | 6 |
| **Economic productivity** | 6 | 0 | 13c | 19c |

a p < 0.1 and > 0.05.

b p < 0.1 and > 0.05.

c Economic productivity measured in hours worked (7 subgroups); missed days (4 subgroups); occupational subgroups (12); wage subsamples/derivative measures (9)

Table 5. Summary of effects reported in the long-term follow-up studies

|  |  |  |
| --- | --- | --- |
| **Policy important domains** | **Reported outcomes**(unit of measurement) | **Effect size**(95% CI) |
| **Baird series** | **Ozier series** | **Croke 2014** |
| **Nutritional status** | **Body mass index** (kg/m2) | **0.02 kg/m2 higher**(0.07 lower to 0.11 higher) | - | - |
| **Height** (cm) | **0.11 cm shorter** (0.65 shorter to 0.43 taller) | **0.20 cm taller1**(0.39 shorter to 0.80 taller) | - |
| **Haemoglobin** (g/dL) | **0.10 g/dL higher2,3**(0.06 lower to 0.27 higher) | - | - |
| **Physical well-being** | **Self-reported health status4**(% rated as ‘very good’) | **4.0 % more**(0.4 more to 7.6 more) | - | - |
| **Poor health in the past month**(work days missed) | **0.11 days fewer5**(0.38 fewer to 0.17 more) | - | - |
| **School enrolment and attendance** | **School enrolment**(%) | **-** | - | **1.86 % higher**(0.72 lower to 4.44 higher) |
| **School enrolment** (total years) | **0.29 years more** (0.00 more to 0.58 more) | - | - |
| **Secondary school attendance** (%)  | **3.0 % higher** (4.0 lower to 10.0 higher) | - | - |
| **School performance and tests of cognition** | **Had to repeat at least one grade**(%) | **6.3 % higher** (2.7 higher to 9.9 higher) | - | - |
| **Passed secondary school entrance exam**(%) | **5.0 % higher**(1.2 lower to 11.2 higher) | - | - |
| **Raven’s matrices test score6**(normalized scores, %) | **1.1 % lower7**(10.7 lower to 8.5 higher) | **22.0 % higher**(6.4 higher to 37.6 higher) | - |
| **English vocabulary test score**(normalized scores, %)  | **7.6 % higher**8(3.4 lower to 18.6 higher) | **16.1 % higher**(3.1 lower to 35.3 higher) | **16.4 % higher**(17.74 lower to 50.54 higher) |
| **Math score**(normalized scores, %) | - | - | **30.1 % higher**(0.81 lower to 61.0 higher) |
| **Economic productivity** | **Hours worked per week**(hours) | **1.58hours more** (0.50 fewer to 3.66 more) | - | - |
| **Monthly earnings (waged employment plus self-employed earnings)** | **226 higher9**(1162 lower to 1614 higher) |  |  |
| **Monthly earnings (waged employment only)** | **26.9 % more**(9.9 % more to 43.9 % more) | - | - |

1 Ozier also reports height-for-age and stunting, which are consistent with the findings for height.
2 Baird 2011a reported control group estimate of 126.1 and coefficient estimate of 1.03 but no unit of measure, and we asume they used gr/L (SI units); we we report this outcome as gr/dL

3 Findings on haemoglobin are not reported in the Baird 2016 version, but in Baird 2011a and 2011b.

4 The Baird series also report the proportion of women who had experienced a miscarriage, which was lower in the intervention group. It is excluded from this table as it seems a spurious outcome to present in isolation without measuring a large range of other potential health outcomes.

5 Findings on work days missed due to poor health in the past month are not reported in the Baird 2016 version, but in Baird 2011a. In Baird 2011b, this outcome is reported for the out-of-school subsample only.

6 Ozier used the 12 questions set B of the Raven's Progressive Matrices. Baird give no further details on the questions used for assessing the Raven’s matrice test score.

7 Findings on Raven’s matrices test score are not reported in the Baird 2016 version, but in Baird 2011a only.

8 Findings on English vocabulary test score are not reported in the Baird 2016 version, but in Baird 2011a, 2011b and 2012.

9 The unit of this outcome is not reported, although we could assume it is the local currency.

Table 6. Outcome appraisal of all outcomes reported in the abstract of Baird 2016

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outcomes reported in the abstract** | **Evidence base for stated effect** | **Effect present in whole sample?1** | **Effect robust to adjustment for multiple inference?2** | **Effect consistent across related outcomes?3** |
| **Men** | “stay enrolled for more years of primary school” | Men from intervention areas had higher total years enrolled in primary school (P < 0.05). | Yes | No | No | No statistically significant difference in the total number of school grades attained (P > 0.1), and adults from intervention areas more likely to have repeated at least one grade (P < 0.01).  |
| “work 17% more hours each week” | Men from intervention areas worked more hours in the past week (P < 0.05). | No | No | - | - |
| “spend more time in non-agricultural self-employment” | A borderline effect on hours worked in non- agricultural self-employment in men (P < 0.1). | Yes (P < 0.05) | Remains borderline | No | No statistically significant difference in monthly non-agricultural earnings (P > 0.1). |
| “spend more time in manufacturing” | Men from intervention areas had a higher manufacturing job indicator (P < 0.05). | Yes  | No | No | No statistically significant effect on hours worked in waged employment (P > 0.1), and no statistically significant difference in monthly non-agricultural earnings (P > 0.1). |
| “miss one fewer meal per week” | Men from intervention areas had eaten more meals the previous day (P < 0.01). | Yes | Yes | - | - |
| **Women** | “one quarter more likely to have attended secondary school.” | Women from intervention areas had higher secondary school attendance (P < 0.05). | No | No | No | No statistically significant difference in the number of school grades attained (P > 0.1). |
| “reallocate time from traditional agriculture into cash crops” | Women from intervention areas had a higher ‘grows cash crop’ indicator (P < 0.05). | Yes | No | - | - |
| “reallocate time from traditional agriculture into non-agricultural self-employment” | Women from intervention areas worked more hours in non-agricultural self-employment in the last week (P < 0.05). | Yes | No | No | No statistically significant difference in monthly non-agricultural earnings (P > 0.1). |

1 The sub-group analysis by sex was not introduced until the third edition of the Baird series and so is considered post-hoc. We considered the effect to be present in the whole sample if P < 0.05 for both sexes combined.
2 The authors of the Baird series conducted adjustments for multiple inference. We considered the effect robust to adjustment if the FDA q-value < 0.05.
3 With so many outcomes presented, we considered whether the effects of related outcomes consistently suggested benefit.