



# **The McGovern-Dole International Food for Education and Child Nutrition Program**

---

## **Health Interventions and their Educational and Health Outcomes in Developing Countries: A Systematic Review and Meta-Analysis**

Prepared by The QED Group, LLC

**Dr. Bechir Rassas  
Dr. Edgar Ariza-Nino  
Dr. Katia Peterson**



## Table of Contents

<b>1.0. INTRODUCTION.....</b>	<b>1</b>
<b>1.1. Background.....</b>	<b>1</b>
<b>1.2. Rationale for Selection.....</b>	<b>2</b>
1.2.1. Health Interventions and the MGD Results Framework.....	2
1.2.2. Health Interventions Considered: Causal Pathways and Outcomes.....	2
1.2.2.1. Malaria.....	3
1.2.2.2. Water and Sanitation for Health.....	5
1.2.2.3. Deworming.....	6
<b>1.3. Organization of the Report.....</b>	<b>6</b>
<b>2.0. OBJECTIVE AND METHODOLOGY.....</b>	<b>7</b>
<b>2.1. Objective.....</b>	<b>7</b>
<b>2.2. Methodology.....</b>	<b>7</b>
2.2.1. Outcomes Considered.....	7
2.2.2. Geographic Coverage.....	8
2.2.3. Timeframe.....	8
2.2.4. Target Groups.....	8
2.2.5. Study Language.....	8
2.2.6. Search Sources.....	8
2.2.7. Evidence Considered and Estimation Methods.....	9
2.2.7.1. Screening Criteria.....	9
2.2.7.2. Exclusion Criteria.....	11
2.2.7.3. Statistical Analysis Methodology.....	11
2.2.7.4. Limitations of the Analysis.....	12
2.2.7.4.1. Assessment of Publication Bias.....	12
2.2.7.4.2. Assessment of Risk of Bias in Included Studies.....	13
2.2.7.4.3. Heterogeneity and Stratified Analysis.....	13
<b>3.0. MALARIA.....</b>	<b>15</b>
<b>3.1. Introduction.....</b>	<b>15</b>
<b>3.2. Findings.....</b>	<b>15</b>
3.2.1. Effect on Educational Outcomes.....	16
3.2.2. Effect on Health Outcomes.....	17
<b>3.3. Conclusions.....</b>	<b>18</b>
<b>3.4. Limitations of the findings.....</b>	<b>19</b>
<b>3.5. Future Research Directions.....</b>	<b>20</b>
<b>4.0. WATER AND SANITATION FOR HEALTH.....</b>	<b>20</b>
<b>4.1. Introduction.....</b>	<b>20</b>
<b>4.2. Findings.....</b>	<b>21</b>
4.2.1. Water and Sanitation for Health Interventions and Educational outcomes.....	22
4.2.1.1. Overall Effect Size.....	22
4.2.1.2. Pooled Effect Sizes by Intervention for All Educational Outcomes.....	23
4.2.1.3. Pooled Effect Sizes by Intervention and Individual Educational Outcome.....	24
4.2.2. Water and Sanitation for Health Interventions and Health Outcomes.....	24
<b>4.3. Conclusions.....</b>	<b>27</b>
<b>4.4. Limitation of the Findings.....</b>	<b>28</b>
<b>4.5. Future Research Directions.....</b>	<b>28</b>
<b>5.0. DEWORMING.....</b>	<b>29</b>
<b>5.1. Introduction.....</b>	<b>29</b>
<b>5.2. Findings.....</b>	<b>30</b>

<b>5.3. Conclusions .....</b>	<b>34</b>
<b>5.4 Limitation of Findings .....</b>	<b>34</b>
<b>5.4.1. Further Discussion on Limitations: should deworming policies be re-evaluated?.....</b>	<b>35</b>
5.4.1.1 <i>Deworming: Not a Panacea?</i> .....	35
5.4.1.2. Has the Meta-Analysis “Stacked the Deck” Against Deworming? .....	35
5.4.1.3. Deworming: a Cornerstone for Neglected Tropical Disease Control .....	35
5.4.1.4. Insufficient Length of Follow-up.....	36
5.4.1.5. Epidemiological Externalities.....	36
5.4.1.6. Long-Term Impact.....	37
5.4.1.7. Cost-effectiveness.....	37
<b>5.5. Future Research Directions.....</b>	<b>38</b>
<b>REFERENCES.....</b>	<b>39</b>
<b>ANNEXES.....</b>	<b>56</b>
<b>Annex 1: Major Characteristics of the Studies Included in the Meta-Analyses.....</b>	<b>56</b>
<b>Annex 2: MGD Results Framework.....</b>	<b>59</b>
<b>Annex 3: Analysis of Publication Bias.....</b>	<b>61</b>
<b>Annex 4. Technical Data Used for Analysis: Forest Plots, Funnel Plots, Egger’s Tests and Detailed Statistics.....</b>	<b>65</b>
Annex 4.1: Malaria Interventions, Educational Outcomes.....	65
Annex 4.2: Malaria Interventions, Health Outcomes.....	89
Annex 2.3: WASH Interventions, Educational Outcomes.....	101
Annex 2.4: WASH Interventions, Health Outcomes.....	135

## **1.0. Introduction**

### **1.1. Background**

The McGovern – Dole International Food for Education and Child Nutrition Program (MGD), one of the Foreign Agricultural Service’s leading food assistance programs, helps support education, child development and food security in low-income, food-deficit countries throughout the world. The program is named in honor of former Ambassador and U.S. Senator George McGovern and former U.S. Senator Robert Dole for their efforts to encourage a global commitment to school feeding and child nutrition.

The key objective of the MGD program is to improve literacy of primary school-age children, especially for girls. By providing school meals, teacher training and related support, MGD projects help enhance school enrollment and academic performance. The program also funds supplementary activities that promote children’s health and nutrition in an effort to further support children’s school enrollment, attendance, and capacity to benefit from the educational instruction received.

The MGD program was first authorized in the Farm Security and Rural Investment Act of 2002 (P.L. 107-171). The 2014 Farm Bill reauthorized the program through 2018. USDA is currently funding 29 McGovern – Dole projects in 23 low-income, food-deficit countries throughout the world. McGovern – Dole projects are implemented world wide by non-profit charitable organizations, cooperatives, the United Nations World Food Program and other international organizations.

The present study is part of a broader evaluation and research effort to: (1) support the MGD program’s ability to use rigorous evidence, evaluation and research in strategic decision-making to improve program outcomes; and (2) help the program identify key gaps in the knowledge base on what interventions are successful in improving literacy and reducing hunger. This study builds on three research efforts: a thorough intervention mapping analysis of the MGD program over a five-year period (2009-2013); a comprehensive annotated bibliography of the programmatic and policy topics of relevance to MGD program interventions; and a proposal for selecting research topics for three systematic reviews of the international literature on the impact of education program interventions in developing countries with particular relevance to the MGD program.

The first topic selected for systematic review focused on assessing the effects of school feeding interventions on educational outcomes. The present systematic review and meta-analysis considers health interventions and their educational and health outcomes.

## 1.2. Rationale for Selection

### 1.2.1. Health Interventions and the MGD Results Framework

The rationale for selecting health interventions and educational and health outcomes is fourfold. First, a primary MGD objective is “improved... student health and nutrition” (McGovern-Dole Program, 2009). According to the MGD theory of change, increased use of health and dietary practices leads to improved literacy of school-age children through reduced health-related absences and therefore improved student attendance.

Second, the 2009-2013 MGD intervention mapping analysis indicates that between one-third and one-half of all MGD programs included a health and nutrition component over the past five years (Table 1.1).

**Table 1.1: MGD Programs Targeting Health and Nutrition Outcomes: Average 2009-2013**

Results Framework Outcome	Programs Targeting Outcome (percent)
Improved Knowledge of Health and Hygiene Practices	42
Increased Knowledge of Safe Food Preparation and Storage Practices	52
Improved School Infrastructure	52
Increased Access to Clean Water and Sanitation Services	40
Increased Access to Preventative Health Interventions	29

*Source: Intervention mapping analysis*

Third, the literature on health offers experimental and quasi-experimental evidence from which it is possible to draw conclusions about what programs are likely to work, as measured by their impact on educational and health outcomes.

Fourth, from this growing body of literature, it is possible to sketch a reasonable consensus on some of these outcomes, draw some lessons learned and their policy implications, and identify areas for further investigation to help close the evaluation gap.

### 1.2.2. Health Interventions Considered: Causal Pathways and Outcomes

Based on a thorough literature review and an annotated bibliography prepared as part of a broader research effort to support MGD’s ability to identify what interventions are most successful in improving literacy and reducing hunger. The annotated bibliography was based on a set of research questions with relevance to the MGD theory of change, using systematic search

for published information to locate as much existing material on these research questions as possible. Of the health programs considered, three major interventions were selected for in-depth analysis: malaria; water and sanitation for health; and deworming. The rationale for selecting each the three interventions is detailed below, together with its causal pathways and outcomes.

### *1.2.2.1. Malaria*

Malaria, a serious disease caused by a parasite that can infect a certain type of mosquito which feeds on humans. In the human body, the parasites multiply in the liver, and then infect red blood cells. If not treated immediately, malaria can quickly become life-threatening by disrupting the blood supply to vital organs. Symptoms of malaria include fever, chills, headache, sweats, fatigue, nausea and vomiting. The symptoms usually appear between 10 and 15 days after the mosquito bite.

According to the latest United Nations Millennium Development Goals Report (United Nations, 2015), malaria continues to pose a major public health challenge, with an estimated 214 million cases and 472,000 deaths globally in 2015. The disease is still endemic in 97 countries and territories around the world. According to UNICEF, an estimated 3.3 billion people are at risk of malaria, of which 1.2 billion are at high risk. In high-risk areas, more than one malaria case occurs per 1000 population. Malaria kills a child somewhere in the world every 30 seconds. It infects 350-500 million people each year -- killing 1 million, mostly children, in Africa (UNICEF, 2013).

The vast majority of malarial infections in children are uncomplicated, febrile episodes from which they make an apparent complete recovery when treated. Young children bear a considerable burden in terms of malaria morbidity and mortality (World Health Organization, 2005). For example, malaria is an important cause of anemia (Geerligs et al. 2003; Kassebaum, et al., 2014; Menendez et al., 2000; Ekvall, 2003; Price et al. 2011; Quintero et al., 2011; Korenromp et al., 2004; Ehrhardt et al.; 2006). Anemia and associated co-morbidities are most concentrated among pre-school children, but school-age children also suffer from their effects, resulting in school absenteeism. Chronic anemia is linked to increase infection. Prolonged and repeated illness may result in school absences for significant lengths of time. School attendance can be affected when other members of the family become ill with malaria; girls in particular may be kept at home to help out. The adverse effects on schooling are likely to go far beyond the number of days lost per year, as absenteeism increases failure rates, repetition of school years, and dropout rates – all of which can hinder efforts to improve literacy rates and stall the progress of education systems (Ennosso et al., 1988; Trape et al., 1993; Brooker et al., 2000; Bundy et al., 2000).

Repeated malaria infection has been found to directly impact a child's opportunity and ability to learn (Sachs & Malaney, 2002; Fernando et al., 2006; Bundy, 2011; Ennosso et al., 1988; Trape et

al., 1993; Brooker et al., 2000). Malaria has been hypothesized to have lifelong negative effects on learning ability and cognitive development due to repeated missed days of school and general overall poor health (Rowland et al., 1977; Schiff et al., 1996; Grantham-McGregor, 1991). For example, children who are repeatedly infected with malaria are found to have poorer overall health and nutritional status than children who are not infected. Poor nutrition-specifically low levels of micronutrients-directly impair brain development.

In consideration of both the direct and indirect consequences of malaria on young children, combating malaria is a priority for many governments and donor organizations. There are still many questions about which malaria interventions have the best cost-benefit. The Copenhagen Consensus Center is a think tank that is devoted to uncovering the smartest solutions for the world's biggest problems. Specifically, the Copenhagen Consensus seeks to uncover the cost-benefit of 'smart and sustainable' solutions<sup>1</sup>.

The 2012 Copenhagen Consensus ranked 30 possible interventions, including education for girls, malaria prevention and treatment, rural water supply, microfinance, and HIV combination prevention in order to best cost-benefit ratio. Guided predominantly by consideration of economic costs and benefits, malaria combination treatment was ranked as the second best intervention overall. This decision was based on the finding that the cost-benefit ratio was not only one of the best returns among infectious disease interventions but also one of the best returns consistently seen across the globe:

“Thus spending \$300 million a year on The Subsidy for Malaria Combination Treatment would prevent 300,000 child deaths, with benefits, put in economic terms, that are 35 times higher than the costs. This analysis suggests it is one of the best returns on health that could be made globally” (Copenhagen Consensus, 2012).

Based on the results from the 2012 Copenhagen Consensus, it is clear that the question of whether school based malaria interventions has moved beyond 'should we intervene?' to 'which intervention should we use?'. There is promising emerging evidence that school based malaria interventions coupled with water and sanitation programs (WASH) and Neglected Tropical Diseases (NTDs) (i.e. de-worming) may not only improve children's lives but their communities. The Bill and Melinda Gates Foundation is actively funding proposals through its “Grand Challenges” (Round 14) mechanism that evaluates these types of combination interventions<sup>2</sup>.

---

<sup>1</sup> Studies are conducted by more than 100 economists from internationally renowned institutions, including seven Nobel Laureates, to advise policymakers and philanthropists on how to spend their money most effectively. The goal of the Copenhagen Consensus project is to set priorities among a series of proposals for confronting the greatest global challenges. For more information, <http://www.copenhagenconsensus.com>

<sup>2</sup> <http://gcgh.grandchallenges.org/challenge/new-ways-working-together-integrating-community-based-intervention-round-14>

### *1.2.2.2. Water and Sanitation for Health*

Water and sanitation for health (WASH), also referred to as water supply and sanitation, has two major dimensions: (1) improved sanitation facilities, defined by the WHO/UNICEF joint monitoring program for water supply and sanitation as one that hygienically separates human excreta from human contact; and (2) improved drinking-water source, defined as one that, by nature of its construction or through active intervention, is protected from outside contamination. An improved drinking water source is defined as a facility or delivery point that protects water from external contamination – particularly fecal contamination. This includes piped water into a dwelling, plot, or yard; public tap or standpipe; tube-well or borehole; protected spring; and rainwater collection. An improved sanitation facility is one that hygienically separates human excreta from human contact (WHO/UNICEF, 2015).

According to WHO and UNICEF, more than 32 percent of the world's population (2.4 billion people) lacked improved sanitation facilities, and 663 million people still used unimproved drinking water sources in 2015 (United Nations, 2015). Improved access to safe water and sanitation services and improved hygiene practices are critical in the prevention and care of 16 of the 17 neglected tropical diseases, including trachoma, soil-transmitted helminths (intestinal worms) and schistosomiasis or bilharzia. Neglected tropical diseases affect more than 1.5 billion people in 149 countries, causing blindness, disfigurement, permanent disability and death (United Nations, 2015).

The United Nations estimates that more than 340,000 children under five (almost 1,000 per day) die annually from diarrheal diseases due to poor sanitation, poor hygiene, or unsafe drinking water (United Nations, 2015). Nearly 1 million deaths per year from diarrheal diseases alone could be prevented by improved water, sanitation and hygiene. Poor water, sanitation and hygiene are major contributors to neglected tropical diseases such as schistosomiasis and trachoma, which affect more than 1.5 billion people every year.

Poor water, sanitation and hygiene conditions do not affect only child health; they also have deleterious effects on educational performance. Their impact on school attendance, learning and cognitive development has been documented (see, for instance, Freeman et al., 2011; Blanton et al., 2007; O'Reilly et al., 2008; Mwanri et al., 2001; Talaat et al., 2011; UNICEF, 2010; Dreibelbis, 2013; Mathegana et al., 2001; WHO, 2002). The practice of open defecation is also linked to a higher risk of stunting – or chronic malnutrition – which affects 161 million children worldwide, leaving them with cognitive damage that affects learning for pre-school and school-age children (CDC, n.d.).

Children who lack access to improved water, sanitation and hygiene are also more likely to contract intestinal-worm infections (Prüss-Üstün A. et al., 2008). As discussed in the next section, intestinal-worm infections resulting from poor water, sanitation and hygiene can cause



diarrhea, anemia and similar health effects, with negative implications on enrolment and attendance, reduced class repetition, and increased educational attainment.

### *1.2.2.3. Deworming*

According to the World Health Organization, approximately 2 billion people are infected with soil-transmitted helminths worldwide (World Health Organization, 2015). Caused by different species of parasitic worms, soil-transmitted helminth infections are transmitted by eggs present in human feces, which contaminate the soil in areas where sanitation is poor. Over 270 million preschool-age children and over 600 million school-age children live in areas where these parasites are intensively transmitted, and are in need of treatment and preventive interventions.

The Copenhagen Consensus 2008 ranked “deworming and other nutrition programs in school” as the sixth best intervention overall. In the Copenhagen Consensus 2012, “deworming of school children to improve educational and health outcomes” was ranked fourth among 16 priority interventions (Copenhagen Consensus, 2012).

Deworming programs are relatively easy to implement in school settings. Teachers need only a few hours of training to understand the rationale for deworming, and to learn how to give out the pills and keep a record of their distribution (Deworm the World, 2010).

WHO’s global target is to eliminate morbidity due to soil-transmitted helminthiases in children by 2020. This would be obtained by regularly treating at least 75 percent of the children in endemic areas (an estimated 873 million).

Soil-transmitted helminth infections can cause a range of symptoms, including intestinal manifestations (diarrhea and abdominal pain), general malaise, and weakness. Hookworms cause chronic intestinal blood loss that can have adverse effects on anemia status, growth, and physical development (Crampton, 2000; de Silva et al., 2003; Dossa et al., 2001; Garg et al., 2005; Awathi et al., 2000; Nga et al., 2009; Sur et al., 2005; Le et al., 2007). They also impair the nutritional status of children, with a significant impact on educational outcomes (Bethony et al., 2006; Sakti et al., 1999; Callender et al., 1998; Simeon et al., 1995; Miguel & Kremer, 2004; Stephenson et al., 1993). Since the most disadvantaged school children -- such as girls and the poor -- often suffer most from ill-health and malnutrition, they would gain the most from deworming. (Bundy et al., 2009; Taylor-Robinson, 2012; World Bank, 2011; World Bank, 2015).

## **1.3. Organization of the Report**

This report contains five sections, including this introduction. The next section describes the objective of the study and its methodology. Sections 3-5 present an in-depth discussion of the

empirical evidence derived from the three major health interventions conducted in school settings: malaria, water and sanitation for health, and deworming. Based on a separate systematic review and meta-analysis, each section presents major findings, followed by summary and conclusions, limitations of the findings for each intervention, and implications for possible future research. Detailed technical data used to derive findings are provided as annexes to the report.

The deworming investigation relies on an existing meta-analysis in the Cochrane Collaboration review series (Taylor-Robinson et al., 2012) and the debate on the impact of deworming that followed its publication. The other two meta-analyses (malaria, and water and sanitation for health) were conducted specifically for this study.

## **2.0. Objective and Methodology**

### **2.1. Objective**

The purpose of the present three systematic reviews and meta-analyses is to investigate the likely causal impact of malaria, water and sanitation for health, and deworming interventions on educational and health outcomes for pre-school and primary-school-age children, and their implications for possible future research directions.

### **2.2. Methodology**

#### **2.2.1. Outcomes Considered**

Studies that investigate malaria, water and sanitation for health, and deworming interventions in relation to educational and health outcomes are considered. Based on the analysis in Section 1.2.2 and a detailed annotated bibliography prepared prior to these meta-analyses, educational outcomes include school participation (enrollment, attendance/absenteeism, dropout, and repetition); learning achievement (standardized math test scores, and standardized language test scores); and cognitive development (verbal fluency, memory, and reasoning). Major health outcomes include: anemia/hemoglobin status, and incidence of malaria (for malaria); and presence of E.coli, number of sick days, and number of sick students (for water and sanitation for health). The outcomes for deworming are those used in the Taylor-Robinson et al. meta-analysis: weight gain, height gain, hemoglobin level, and physical well-being (Taylor-Robinson, 2012).

Literacy has not been used as an outcome measure in the literature under consideration because it has proved to be a complex and dynamic concept, continuing to be interpreted and defined in a multiplicity of ways. As such, literacy has expanded from a simple process of acquiring basic cognitive skills, to using these skills in ways that contribute to socio-economic development, to developing the capacity for social awareness and critical reflection as a basis for personal and social change. Reflecting this complexity, UNESCO defines literacy as “a set of tangible skills — particularly the cognitive skills of reading and writing,” and “the ability to use reading, writing and numeracy skills for effective functioning and development of the individual and the

community” (UNESCO, 2006). It should, however, be noted that the multi-dimensional nature of literacy in this definition is captured in at least two of the three categories of outcome measures (learning achievement, and cognitive development) used in the literature reviewed for this study.

### **2.2.2. Geographic Coverage**

Only studies pertaining to developing countries are included.<sup>3</sup>

### **2.2.3. Timeframe**

The literature search was mainly, but not exclusively, based on studies published in 2000-2015. Studies conducted before 2000, but published in 2000-2015 were included. Earlier studies considered as pioneers and/or especially relevant were also considered.

### **2.2.4. Target Groups**

Pre-primary and primary-school-age children are the focus of the investigation.<sup>4</sup>

### **2.2.5. Study Language**

Studies are not excluded on the basis of language.

### **2.2.6. Search Sources**

The studies reviewed for the malaria and water and sanitation for health meta-analyses were identified through a systematic search. The search covered both general and specialist sources pertaining to education, economics, nutrition and health. They included electronic sources and journals, websites of research centers and gray publications (unpublished studies, including studies found through the World Bank, and the Abdul Latif Jameel Poverty Action Lab at MIT). Citation tracking and examination of the body of work of relevant influential authors were used to identify studies meeting the inclusion criteria used in these reviews. Electronic searches were conducted on papers cited in other papers already included in this review as well as cross-checking of references cited in other meta-analysis papers that included health interventions in a school setting. Citation searches were also conducted using Google Scholar for related systematic reviews and relevant impact evaluations. Such impact evaluations and systematic reviews (and the citations therein) were screened for relevance using the screening criteria described below.

---

<sup>3</sup> Developing countries are characterized as such based on the classification used in the International Monetary Fund World Economic Outlook for 2014.

<sup>4</sup> The malaria and WASH meta-analyses focused exclusively on interventions conducted in school settings. The deworming meta-analysis conducted by Taylor-Robinson (2012) extended coverage to children recruited from communities and health facilities.

## 2.2.7. Evidence Considered and Estimation Methods

### 2.2.7.1. Screening Criteria

Only the empirical literature that contains the most rigorous evidence using the strongest methodology for identifying causal impacts was considered. Impact evaluations quantify the effects of programs on individuals, households, and communities. They show whether the changes observed are indeed due to the program intervention and not to other factors (Khandker et al., 2010). Impact evaluations are “analyses that measure the net change in outcomes for a particular group of people that can be attributed to a specific program using the best methodology available, feasible and appropriate to the evaluation question that is being investigated and to the specific context” (International Initiative for Impact Evaluation, 2008). They “compare the outcomes of a program against a counterfactual that shows what would have happened to beneficiaries without the program. Unlike other forms of evaluation (such as ‘performance evaluations’), they permit the attribution of observed changes in outcomes to the program being evaluated” (World Bank, n.d.).

Attribution is different from association between the intervention and outcomes that may have been affected by other contextual factors. Evaluating the impact of an intervention hinges on a fundamental question: What would the situation have been if the intervention had not taken place. While descriptive monitoring leaves ample room for differing interpretations of how much the identified change can be attributed to the intervention, impact evaluations rely on more sophisticated methods to disentangle the net gains from that intervention.

Impact evaluations range from randomized designs to quasi-experimental models. There is consensus that experimental design is the best evaluation method. This method is used to determine what would have been the outcomes had the beneficiaries not participated in the program, in which beneficiaries (called intervention or treatment group) are randomly selected from a set of communities with similar characteristics. Subjects not randomly selected for the intervention form a counterfactual (called comparison or control group). Randomized controlled trials (RCTs), the gold standard by which scientific evidence is evaluated, can be either double-blind trials, an experimental procedure in which neither the subjects nor the experimenters know which subjects are in the test and control groups during the actual course of the experiments; or single-blind trials, an experimental procedure in which the experimenters but not the subjects know the makeup of the test and control groups during the course of the experiments. The control may be a standard practice, a placebo, or no intervention at all.

Ideally, all variables in an experiment will be controlled. In such a controlled experiment, if all the controls work as expected, it is possible to conclude that the results of the experiment are due to the effect of the variable being tested. More generally, experimental design enables the investigator to make claims of the following nature: The two situations were identical until the

intervention was introduced. Since the intervention is the only difference between the two situations, the new outcome was caused by that intervention.

Quasi-experimental designs are used when all the necessary requirements to control influences of extraneous variables cannot be met, most particularly when randomization is not possible for political, ethical, or logistical reasons. When the subjects cannot be randomly assigned to either the experimental or the control group, or when the researcher cannot control which group will get the treatment, participants do not all have the same chance of being in the control or the experimental groups, or of receiving or not receiving the treatment.<sup>5</sup>

While RCTs have pre-test and post-test data for randomly assigned intervention and control groups, quasi-experimental design studies develop a counterfactual using a comparison group which has not been created by randomization. To develop the counterfactual, quasi-experimental studies use statistical techniques to create a comparison group that is matched with the intervention group in socioeconomic and other characteristics, or to adjust for differences between the two groups that might otherwise lead to inaccurate estimates. The goal of such statistical techniques is to simulate a randomized controlled trial.<sup>6</sup> Quasi-experimental methods include the following:

- **Difference-in-Difference (or Double Difference):** An increasingly popular method to estimate causal relationships, this technique compares the before-and-after difference for a group receiving the intervention to the before-after difference for those who did not.
- **Matched comparisons:** An analysis in which subjects in a treatment group and a comparison group are made comparable with respect to extraneous factors by individually pairing study subjects with the comparison group subjects.
- **Instrumental variables:** Have been used primarily in economic research, but have increasingly appeared in epidemiological studies. They are used to control for confounding and measurement error in observational studies, allowing for the possibility of making causal inferences with observational data and can adjust for both observed and unobserved confounding effects.
- **Judgmental matching of comparison groups:** A statistical method that involves creating a comparison group by finding a match for each person or site in the treatment group based on the researcher's judgment about what variables are important.
- **Propensity score matching:** Statistically creating comparable groups based on an analysis of the factors that influenced people's propensity to participate in a given program. The most

---

<sup>5</sup> Following the literature, the event for which an estimate of the causal effect is sought is called *treatment*. The *outcome* is what will be used to measure the effect of the treatment. The treatment and control groups do not necessarily need to have the same pre-intervention conditions. The two groups may well have different characteristics. However, many of those characteristics can reasonably be assumed to remain constant over time or at least over the course of an evaluation.

<sup>6</sup> For details on all these evaluation methods, see for instance Khandker et al., 2010; and Gertler et al., 2011.

common implementation of propensity score matching is one-to-one or pair matching, in which pairs of treated and untreated subjects are formed, such that matched subjects have similar values of the propensity score.

- Regression discontinuity: An analysis used to estimate program impacts in situations in which candidates are selected for treatment based on whether their value for a numeric rating exceeds a designated threshold or cut-off point. The analysis consists of comparing the outcomes of individuals below the cut-off point with those above the cut-off point.

### *2.2.7.2. Exclusion Criteria*

Studies that did not meet the inclusion criteria listed above (including studies that did not have a control group) were not considered.

### *2.2.7.3. Statistical Analysis Methodology*

Data in the studies reviewed were analyzed through meta-analysis.<sup>7</sup> Meta-analysis is the statistical combination of results from those separate studies. It can be used to generalize from the sample of studies based on different assumptions about the distribution of effects. Such a combination yields an overall effect size, a statistic (a quantitative measure) that summarizes the effectiveness of the interventions compared with their control interventions.<sup>8</sup>

The Comprehensive Meta-Analysis software, a computer program for meta-analysis, was used to estimate effect sizes. The random effects meta-analysis methodology was used to derive estimates.<sup>9</sup> Unlike the fixed-effect meta-analysis, which assumes that the treatment effect is

---

<sup>7</sup> According to the Campbell Collaboration -- an international research network that produces systematic reviews of the effects of social interventions in crime and justice, education, international development, and social welfare -- the objective of a systematic review is to “sum up the best available research on a specific question. This is done by synthesizing the results of several studies. A systematic review uses transparent procedures to find, evaluate and synthesize the results of relevant research. Procedures are explicitly defined in advance, in order to ensure that the exercise is transparent and can be replicated... Studies included in a review are screened for quality, so that the findings of a large number of studies can be combined.” (Higgins 2014). This definition applies to any technical research topic. For instance, the U.S. Department of Health and Human Services defines the systematic review as “a critical assessment and evaluation of all research studies that address a particular clinical issue. The researchers use an organized method of locating, assembling, and evaluating a body of literature on a particular topic using a set of specific criteria.” (<http://effectivehealthcare.ahrq.gov/index.cfm/glossary-of-terms/?pageaction=showterm&termid=70;>; accessed 5/9/2015).

<sup>8</sup> The effect size is a generic term for the estimate of effect of treatment for a study. It is a dimensionless measure of effect that is typically used for continuous data when different scales are used to measure an outcome and is usually defined as the difference in means between the intervention and control groups divided by the standard deviation of the control or both groups, where the standard deviation is defined as the spread or dispersion of a set of observations, calculated as the average difference from the mean value in the sample. (See, for instance, Cochrane Community, <http://community.cochrane.org/>; accessed 5/9/2015).

<sup>9</sup> This selection follows the international development meta-analysis literature (see, for instance, Taylor-Robinson, 2012, the deworming meta-analysis reviewed as part of this study). More generally, when studies are gathered from the published literature, especially when those studies are characterized by methodological diversity and involve diverse groups of subjects, the random effects model is a more plausible match. Methodological diversity creates heterogeneity (i.e., variation across studies) through biases variably affecting the results of the different studies. The

common across all studies and that differences in study findings are due to sampling error, or chance, only (Riley et al., 2011), random-effects meta-analysis estimates the average effect across studies, allowing for differences due to both chance and other factors which affect estimates -- such as study location, characteristics of the target population and length or intensity of the treatment. For this reason, the random-effects confidence interval in random-effects meta-analysis is wider than that estimated in a fixed-effect meta-analysis, reflecting a more conservative estimate as a result of the additional uncertainty around the estimate.

Study weights are also more balanced under the random-effects model than under the fixed-effect model. Under the fixed-effects model, it is assumed that the true effect size for all studies is identical, and the only reason the effect size varies between studies is sampling error (error in estimating the effect size). Therefore, when assigning weights to the different studies under the fixed-effect model it is assumed that we can largely ignore the information in the smaller studies because we have better information about the same effect size in the larger studies. By contrast, our objective under the random-effects model is not to estimate one true (“fixed”) effect, but to estimate the mean of a distribution of effects to ensure that all these effect sizes are represented in the summary estimate.<sup>10</sup>

#### *2.2.7.4. Limitations of the Analysis*

##### *2.2.7.4.1. Assessment of Publication Bias*

The presence of bias in the extracted data for the malaria and WASH interventions<sup>11</sup> was evaluated graphically by using the funnel plot and Egger’s regression tests (Egger et al., 1997). To reduce publication bias (a situation that, for instance, may lead journals to prefer studies with positive effects), the search was broadened to the non-published “grey literature” that included conference proceedings, technical reports, dissertations, and theses. However, no attempt was made to assess publication bias through sensitivity analysis for outliers (defined as any study which differed markedly from the overall pattern) or through imputation of missing studies by using “trim and fill” analysis (Duval & Tweedie, 2000) -- a sensitivity analysis method that extends beyond the scope of this study.

---

random-effects estimate and its confidence interval address the question of the average intervention effect in those studies (see, for instance, Borenstein, 2010, Higgins, 2014; Alison 2010).

<sup>10</sup> This is equivalent to saying that we cannot discount a small study by giving it a very small weight (the way we would in a fixed-effect analysis). Since our objective is to estimate the mean effect in a range of studies -- and we do not want that overall estimate to be overly influenced by any one of them -- we cannot give too much weight to a very large study (the way we would in a fixed-effect analysis) and give too little weight to the estimate provided by a small study because that estimate contains information about an effect that no other study has estimated (See, for instance, <http://www.meta-analysis.com/downloads/Meta-analysis%20Fixed-effect%20vs%20Random-effects%20models.pdf>; accessed 6/10/2015).

<sup>11</sup> A standard assessment of publication bias, risk of bias in the included studies, and heterogeneity for deworming was conducted in the Taylor-Robinson et al. (2012) meta-analysis.

Another method of assessing the potential for publication bias is to calculate the “fail-safe N,” the number of studies whose effect size is zero or negative that would be needed to increase the P-value for the meta-analysis to above 0.05 (or any other selected threshold). However, the Cochrane Handbook for Systematic Reviews of Interventions notes that “this and other methods are not recommended for use in Cochrane reviews” (Higgins et al., 2014). (For additional information on publication bias, see Annex 3; for detailed funnel plots and Egger’s regression texts associated with each pooled effect size estimated in the malaria and WASH meta-analyses, see Annex 4.)

#### 2.2.7.4.2. Assessment of Risk of Bias in Included Studies

The Cochrane Collaboration recommends a specific tool for assessing risk of bias in each included study and across studies. The assessment consists of a judgment and a support for that judgment for each entry in a “risk of bias” table, where each entry addresses a specific feature of the study. The judgment for each entry involves assessing the risk of bias as “low risk,” “high risk,” or “unclear risk,” with the last category indicating either lack of information or uncertainty over the potential for bias. Assessment of risk of bias includes sequence generation (checking for possible selection bias), allocation concealment (checking for possible selection bias), blinding in RCTs (checking for possible performance and detection bias), incomplete outcome data (checking for possible attrition bias through withdrawals, dropouts or protocol deviations), selective reporting bias, and other sources of bias.

As for publication bias, a detailed assessment of risk of bias for each study included in the meta-analysis is beyond the scope of this investigation.

#### 2.2.7.4.3. Heterogeneity and Stratified Analysis

We addressed heterogeneity in the malaria and WASH meta-analyses by use of random-effects meta-analysis (see Section 2.2.7.3) and predefined subgroup analyses. Heterogeneity is used to describe the variation in, or diversity of, participants, interventions, and measurement of outcomes across a set of studies. In a statistical sense, it is used to describe the degree of variation in the effect estimates from a set of studies. It is also used to indicate the presence of variability among studies beyond the amount expected due solely to chance. Heterogeneity in meta-analysis is measured by  $I^2$ , a statistical expression of the inconsistency of the results in the studies reviewed. For example, a meta-analysis with  $I^2 = 0$  means that all variability in effect size estimates is due to sampling error within studies. On the other hand, a meta-analysis with  $I^2 = 50$  means that half of the total variability among effect sizes is caused not by sampling error, but by true heterogeneity between studies. According to the Cochrane Handbook (Higgins, 2014) a rough guide to the interpretation of  $I^2$  is as follows:

- 0% to 40%: might not be important;
- 30% to 60%: may represent moderate heterogeneity;



- 50% to 90%: may represent substantial heterogeneity;
- 75% to 100%: considerable heterogeneity.

We visually examined the forest plots from the meta-analyses to look for any obvious heterogeneity among studies in terms of the size or the direction of treatment effect. A forest plot is a graphical representation of the individual results of each study included in a meta-analysis, together with the combined meta-analysis result. The plot also allows researchers to see the heterogeneity among the results of the studies.

We used the  $I^2$  statistic test to quantify the level of heterogeneity among the studies in each analysis. We explored the identified heterogeneity by subgroups of participants, treatments, and outcomes. (Forest plots and  $I^2$  statistics for all interventions and outcomes measured can be found in Annex 4.) The stratified analysis focused on individual outcomes by intervention; outcome category and individual outcomes within each category; and gender, when data were available. Further stratified analyses to control for certain treatment sub-categories and experimental samples are beyond the scope of this study. These include the effect of the following moderators<sup>12</sup> and their impact:

- Study design and quality: RCTs vs. quasi-experimental design; for RCTs, masking of participants and outcome assessors, unit and method of allocation, and exclusion of participants after randomization or proportion of losses after follow-up; working papers vs. published papers; and quasi-experimental design method (for major quasi-experimental design methods, see Section 2.2.7.1).
- Geographic location of study population
- Rural and urban location
- Socio-economic status as defined in each study
- Age of children
- Grade of children
- Study duration
- Sample size and power analysis

---

<sup>12</sup> Statistically, a moderating variable is one that affects the direction and/or strength of the relation between dependent and independent variables.

## 3.0. Malaria

### 3.1. Introduction

Section 1.2.2.1 described how the malaria parasite can infect a certain type of mosquito which feeds on humans and how the malaria infection can cause death if not treated immediately. It also summarized the pathways through which malaria affects educational and health outcomes.

Chloroquine (or chloroquine phosphate) is an antimalarial medicine that can be prescribed for adults and children of all ages. It is a relatively well-tolerated medicine that can be used for either prevention or treatment. Intermittent preventive therapy or intermittent preventive treatment (IPT) is a public health intervention aimed at treating and preventing malaria episodes in pregnant women, infants, children, and schoolchildren.

This section presents a meta-analysis of malaria interventions on educational and health outcomes in school settings. The major characteristics of the studies used in the meta-analysis are provided as Annex 1.

The studies included in this review focus on three intervention strategies: (1) chloroquine prevention and treatment given to all children without any time restriction and regardless of whether they are infected or not; (2) intermittent preventive treatment (IPT) to treat all children for malaria at regular intervals during the transmission season, regardless of whether they are infected or not; and (3) intermittent screening and treatment (IST), where children are tested on every scheduled visit and treated only if they are infected.

The outcomes considered in the studies reviewed are of two types: (1) educational outcomes (student absences, their sustained attention in the classroom, and their performance in language and math tests); and (2) health outcomes: (anemia/hemoglobin status, and incidence of malaria). Hemoglobin is a protein in the red blood cells that carries oxygen to the body's organs and tissues and transports carbon dioxide from the organs and tissues back to the lungs. Anemia is a condition in which school children feel tired and weak because they do not have enough healthy red blood cells to carry adequate oxygen to the tissues. Anemia can have many different causes, including vitamin deficiency and chronic diseases.

### 3.2. Findings

This section presents the major effects of malaria interventions on educational and health outcomes. The two categories of outcomes are presented in turn. The next section (Section 3.2.1) first describes the effects of malaria interventions on the combined educational outcomes (school absences, student attention, language proficiency, and math skills). The effects of each of those

four outcomes are then separately assessed. Section 3.2.2 describes the effects of malaria interventions on the combined health outcomes (anemia/hemoglobin status, and incidence of malaria). The effects of each of those two outcomes are then separately analyzed.

For clarity and ease of presentation, the detailed findings are based on a series of tables derived from the forest plots and associated data presented as Annex 2 which, together with Annex 1, includes detailed statistics of effect sizes such as standard errors, t-values, degrees of freedom, confidence intervals, statistical significance, heterogeneity statistics, funnel plots and Egger’s tests.. The detailed findings are followed by summary and conclusions, limitations of those findings, and implications for future research.

### 3.2.1. Effect on Educational Outcomes

#### **Finding 3.1: Malaria prevention and treatment in school-settings have an overall positive effect on the combined educational outcomes considered**

Table 3.1 illustrates the overall effect of chloroquine and IPT/IST interventions on absenteeism, attention levels, and test scores for language and math. Overall, both sets of interventions had a positive effect on the four selected outcomes but the most of that effect is attributed to chloroquine interventions.

**Table 3.1: Effect of Malaria Prevention and Treatment Interventions on Educational Outcomes**

Outcome	Intervention and effect					
	Chloroquine		IPT/IST		Total	
	Effect estimate	Effect sizes	Effect estimate	Effect sizes	Effect estimate	Effect sizes
Absence	0.260	4			0.260	4
Attention			-0.118 (*)	6	-0.118 (*)	6
Language scores	0.408 (***)	17	0.176 (*)	17	0.288 (***)	34
Math scores	0.490 (***)	17	0.028	6	0.365 (***)	23
Total	0.429 (***)	38	0.074	29	0.276 (***)	67
IPT=Intermittent preventive treatment; IST=Intermittent screening and treatment						
Positive sign (+) favors intervention; negative sign (-) favors control						
(***) Significant at 99% level; (**) Significant at 95% level; (*) Significant at 90% level						

**Finding 3.2: Improvements in math and language scores are solely attributed to chloroquine.**

Chloroquine was the only intervention that demonstrated a statistically significant effect on improving math and language scores. Although the effect is statistically significant, the intervention has small to medium effect<sup>13</sup> of 0.429 (Table 3.1). Both intermittent preventive treatment and intermittent screening and treatment interventions were not found to improve math or language scores at a statistically significant level.

**Finding 3.3: Chloroquine has no effect on absenteeism or attention levels**

Table 3.1 illustrates that chloroquine has no effect on school absences (estimate not statistically different from zero).

**Finding 3.4: IPT/IST interventions had a small effect on attention levels**

Table 3.1 shows that intermittent preventive treatment and intermittent screening and treatment have a small effect on student attention levels. However, this finding was statistically significant at a 90% level.

**Finding 3.5: Chloroquine has a much greater effect on language and math indicators than IPT/IST**

As illustrated in Table 3.1, chloroquine had a greater (and statistically significant) effect language and math (0.408 and 0.490, respectively) scores than IPT/IST (0.028 and 0.074, respectively) and those effects were not statistically significant.

**3.2.2. Effect on Health Outcomes**

**Finding 3.5: Malaria prevention and treatment in school settings have an overall positive effect on the combined health outcomes considered and that effect is stronger than its corresponding effect on educational outcomes**

Overall, chloroquine and IPT/IST interventions had a much greater effect on health outcomes than educational outcomes (effect estimates of 0.507 and 0.276 respectively (see Table 3.2)).

---

<sup>13</sup> Effect size magnitudes are typically interpreted on the basis of rules of thumb suggested by Cohen (1988). According to Cohen, an effect size of about 0.20 is considered “small,” of about 0.50 is considered “medium,” and of about 0.80 is considered “large.” Although these guidelines are broad categorizations, it has become standard practice for researchers to use them when interpreting effect size estimates. Thus, if the means for the treatment and control groups do not differ by 0.2 standard deviations or more, the difference is “trivial” or very small even if it is statistically significant.

**Table 3.2: Effect of Malaria Prevention and Treatment Interventions on Health Outcomes**

Outcome	Intervention and effect					
	Chloroquine		IPT/IST		Total	
	Effect estimate	Effect sizes	Effect estimate	Effect sizes	Effect estimate	Effect sizes
Anemia/ Hemoglobin	0.382 (***)	2	0.097 (*)	7	0.221 (***)	9
Malaria morbidity	0.778 (***)	1	0.610 (***)	14	0.623 (***)	15
Total	0.504 (***)	3	0.508 (***)	21	0.507 (***)	24
IPT=intermittent preventive treatment; IST=Intermittent screening and treatment						
Positive sign (+) favors intervention; negative sign (-) favors control						
(***) Significant at 99% level; (**) Significant at 95% level; (*) Significant at 90% level						

**Finding 3.6: Chloroquine interventions and IPT/IST interventions have nearly identical effects on combined health outcomes.**

As shown in Table 3.2, the effect of chloroquine (0.504) on combined health outcomes is nearly identical to the effect of intermittent preventive treatment and intermittent screening and treatment (0.508). Both effects are significant at the 99% level.

**Finding 3.7: Chloroquine and IPT/IST have a much greater effect on malaria morbidity than anemia/hemoglobin levels.**

As detailed in Table 3.2, the combined effect of chloroquine and IPT/IST on decreasing malaria morbidity (0.623) is larger than their combined effects on increasing anemia/hemoglobin (0.221) levels. This finding applies not only to their combined effects, but also when analyzed separately. The effects of IPT/IST on reducing malaria morbidity is found to have an estimated effect of 0.610 but the effect on anemia/hemoglobin levels is only 0.097. Similarly, the effect of chloroquine on reducing malaria morbidity is 0.779 and only 0.382 on anemia/hemoglobin levels.

### 3.3. Conclusions

#### *Educational outcomes*

- Chloroquine interventions demonstrated the greatest impact (versus IPT/IST) on math and language test scores. Chloroquine demonstrated no impact on attendance rates.
- Neither school absences nor student attention levels are affected by chloroquine prevention and treatment or by IPT/IST.

### *Health outcomes*

- Although both chloroquine and IPT/IST interventions demonstrated a positive effect on anemia/hemoglobin levels, chloroquine had a greater effect size and was statistically significant at a 99% versus IPT/IST at 90% statistical significance level.
- Chloroquine and IPT/IST interventions have a much greater impact on the reduction of malaria morbidity than on anemia and hemoglobin levels. Notably, chloroquine has a much smaller effect size than IPT/IST on malaria morbidity.

### **3.4 Limitations of the findings**

- A significant proportion of the studies had small samples (the smallest sample was in Mali involving 296 students assigned to three distinct trial groups). The small sample sizes limits the precision of treatment effects. Furthermore, many of the studies have such wide confidence intervals that effects sizes are not statistically different from zero. Small sample sizes may artificially ‘deflate’ a program’s real effectiveness.
- There is limited experimental evidence—as illustrated by the small effect sizes—on the benefits of school-based malaria interventions. The impact of school-based malaria interventions can vary widely depending on the intensity of malaria transmission. Furthermore, there is no reliable information on what threshold of malaria transmission yields the best cost-benefit.
- There is a lack of geographic diversity among the studies. Coupled with small sample sizes, a lack of diversity reduces the external validity of existing evidence. This has direct implications on the generalizability of findings to different populations of students, contexts, treatment variations, and outcomes measured.
- There is a paucity of information on the cost and cost-effectiveness<sup>14</sup> and cost-benefit of malaria prevention and treatment through school based programs. Only one study was found that contained a detailed cost analysis of an IST intervention (Drake et al., 2011).<sup>15</sup>
- There is a dearth of evidence on the long-term effects of school based malaria interventions. Only one study (Cutler et al., 2010) extended the malaria literature by investigating the effects of childhood exposure to malaria eradication on educational attainment and economic status in adulthood.<sup>16</sup>

---

<sup>14</sup> For a definition of cost and cost-effectiveness, see Section 2 of this report

<sup>15</sup> The financial cost of IST per child screened was estimated at \$6.61 (in 2010 dollars). Key contributors to cost were salary costs (36 percent) and malaria rapid diagnostic tests (22 percent). Almost half the intervention cost consisted of redeployment of existing resources, including health worker time and use of hospital vehicles. The study concluded that school-based IST is a relatively expensive malaria intervention in the current context, but reducing the complexity of delivery can result in considerable savings in the cost of intervention.

<sup>16</sup> The investigation used data from a large scale eradication program that drastically reduced malaria in India over a short period in the 1950s. Comparing outcomes, at a point in time, for individuals in birth cohorts born before and after the eradication era in areas with varying pre-eradication malaria prevalence, the study found that males

### 3.5 Future Research Directions

- There is a great need to re-evaluate proven and promising interventions **at scale**, and over a **longer time period (5+ years)**, in order to generate evidence on impact and cost-effectiveness (or cost-benefit).
- Future malaria research on the benefits of school-based malaria interventions should be expanded to include how their impact varies according to causal mechanisms and intensity of malaria transmission.
- To yield more robust results and enhance the generalizability of findings to different populations of students, contexts, treatment variations, and outcomes measured, future research should include interventions involving greater geographic diversity and larger samples.
- Additional studies are needed to overcome the current scarcity of information on the cost and cost-effectiveness of malaria prevention and treatment. Such studies would help policymakers' resource allocation efficiency when prioritizing interventions.
- Future research should extent its scope to the effects of childhood exposure to malaria eradication on educational attainment and economic status in adulthood. Taking into consideration the long-term effects of malaria prevention and treatment would not only capture the full benefits of malaria prevention and treatment, but would also refine the cost and cost-effectiveness analysis of this intervention and enhance policymakers' resource allocation efficiency.
- Process evaluations and operations research studies are critical to scale up and reproducibility. There is a dearth of information about why interventions work and why they don't work. This is a critical gap with policy implications that must be addressed.

## 4.0. Water and Sanitation for Health

### 4.1. Introduction

As described in Section 1.2.2.2 , WASH interventions consist of improved sanitation facilities to separate human excreta from human contact, and an improved drinking-water source to protect participants from fecal and other outside contamination.

---

exposed to malaria eradication in early childhood had higher per capita household consumption as adults, and the effects for men were larger than those for women in most specifications. The study did not find any evidence of increased educational attainment for men and mixed evidence for women, a result that may have reflected the trade-off between schooling and labor. Similar long-term study in other geographic and socio-economic settings may shed additional light on those and other relationships.

This section presents a meta-analysis of WASH interventions on educational and health outcomes in school settings. The major characteristics of the studies used in the meta-analysis are summarized in Annex 1. The studies included in this review focus on four intervention strategies: health promotion (HP); water supply (WS); water treatment (WT); and sanitation (SAN), consisting mainly of latrine construction. The impact of the interventions is measured through educational outcomes (school enrollment, absences, and dropouts), and health outcomes (presence of E.coli, number of sick days, and number of sick students).

## 4.2. Findings

WASH interventions are defined in the studies as those for hand washing promotion, water quality, water supply, sanitation, or any combination thereof. Control refers to study participants who have continued with usual practices. Hand washing promotion is any intervention -- such as group discussion, songs, pictorial stories, and dramas -- that promotes adoption of, or increased practice of, hand washing. Hand washing includes water, wash basins, soap, and drying devices. Sanitation refers to any intervention to introduce or expand the provision or use of facilities for urination or defecation. Water quality is any intervention to improve the microbiological quality of drinking water. Water supply refers to any intervention to provide a new or improved water supply or improved distribution such as installation of a new hand pump or school connection or both.

Due to data availability, educational outcomes are limited to school enrollment, student absences, and dropout rates. Health outcomes will be limited to the presence of E. coli, number of sick days for students, and the number of sick students.

This section first analyzes the impact of WASH interventions on educational outcomes. The analysis is followed by their impact on health outcomes. The next section (Section 4.2.1) first describes the effects of WASH interventions on the combined educational outcomes (school enrollment, absences and dropouts). The effects of each of those three outcomes are then separately presented. The following section (Section 4.2.2) first describes the effects of WASH interventions on the combined health outcomes (presence of E. coli, number of sick days, and number of sick students). The effects of each of those three outcomes are then separately analyzed.

In Section 3.2, the detailed findings are based on a series of tables derived from the forest plots and associated data presented as Annex 4 which, together with Annex 3, includes detailed statistics of effect sizes such as standard errors, t-values, degrees of freedom, confidence intervals, statistical significance, heterogeneity statistics, funnel plots and Egger's tests. The detailed findings are followed by summary and conclusions, limitations of those findings, and implications for future research.



## 4.2.1. Water and Sanitation for Health Interventions and Educational outcomes

### 4.2.1.1. Overall Effect Size

#### **Finding 4.1: The overall effect of WASH interventions on the combined educational outcomes is positive, but very small**

Table 4.1 shows the estimated combined mean impact of WASH interventions on educational outcomes (school enrollment, absences and dropouts)<sup>17</sup>. The overall effect size is 0.039 standard deviations, with a 99% confidence interval of (0.028, 0.050),<sup>18</sup> indicating that the impact of WASH interventions on educational outcomes as measured by the difference in outcomes between the treatment group and control group after the interventions is positive. As indicated by (\*\*\*) in the table, this difference is statistically significant at the 99% level.<sup>19</sup> Although positive, the effect size estimate is, at 0.039, very small.<sup>20</sup>

<b>Table 4.1: Overall Effect Size Estimate of WASH Interventions on Educational Outcomes (school enrollment, absences and dropouts)</b>				
Estimate	Standard Error	P-value	95% C.I.L	95% C.I.U
0.039 (***)	0.006	0.00	0.028	0.050
(***) Significant at 99% level; (**) Significant at 95% level; (*) Significant at 90% level				
Number of effect sizes: 26				

The standard error<sup>21</sup> is used to weigh effect sizes when combining studies, so that large studies are considered more important than small studies in the overall analysis.

<sup>17</sup> Due to lack of data, other outcomes such as attendance, learning achievements and cognitive development could not be included in the meta-analysis.

<sup>18</sup> A confidence interval is a range of values such that there is a specified probability that the value of a parameter lies within that range. In our example, we are 99% confident that the 0.039 standard deviation falls between 0.028 and 0.050. Note that the significance level is reflected in the P-value as follows: P-value <0.01 means statistical significance at the 99% level; P-value <0.05 means statistical significance at the 95% level; P-value <0.1 means statistical significance at the 90% level.

<sup>19</sup> A null hypothesis is the statement that WASH interventions have *no* impact on educational outcomes. For a null hypothesis to be rejected as false (i.e., that WASH interventions *do have* an impact on educational outcomes), the result has to be identified as being statistically significant (i.e., unlikely to have occurred due to sampling error alone or, equivalently, due to the unrepresentativeness of the sample). The probability of rejecting the null hypothesis (in this case rejecting the hypothesis that WASH interventions have no impact on educational outcomes) given that it is true, is most often set at 0.05 (95%), but can also be set at 0.01 (99%) or 0.10 (10%). Put differently, to determine whether a result is statistically significant at a given level, a researcher has to calculate a P-value, which is the probability of observing an effect given that the null hypothesis is true. The null hypothesis is rejected if the P-value is lower than the significance level -- which is the case here since the P-value (0.000) is lower than the significance level (0.01).

<sup>20</sup> As explained in Section 3.2.1, an effect size of about 0.20 is considered small, of about 0.50 is considered medium, and of about 0.80 is considered large.

<sup>21</sup> Standard error is a statistical term that measures the accuracy with which a sample represents a population. In statistics, if the sample mean deviates from the actual mean of a population, this deviation is the standard error.

The standard error of effect size is calculated differently for each type of effect size, but it generally requires only knowing the study's sample size or the number of observations in each group.

**Finding 4.2: The overall effect size is considerably higher for girls than for boys**

While the effect size for girls (0.044) is statistically significant, the effect size for boys is not statistically different from zero. (For details, see Annex 4.)

*4.2.1.2. Pooled Effect Sizes by Intervention for All Educational Outcomes*

**Finding 4.3: The effect size for all educational outcomes combined is higher for all WASH interventions combined than for subsets of interventions implemented separately**

Table 4.2 describes the mean effect size of WASH interventions on educational outcomes (school enrolment, absences and dropouts) on WASH interventions conducted in a single form or in combination. The combination of hand washing promotion, water treatment, sanitation, and water supply (HP, WT, SAN and WS) has the highest effect (0.328), followed by sanitation alone (0.037). A combination of hand washing promotion and water treatment (HP and WT) or a combination of hand washing promotion, water treatment and sanitation (HP, WT and SAN) has no effect on educational outcomes (their effects of 0.120 and 0.091, respectively, are not statistically significant). This result underlines the critical role of water supply in WASH interventions. Although not all effects could be estimated by gender for lack of data, Table 4.2 suggests that this conclusion applies to both girls and boys.<sup>22</sup>

**Table 4.2: Mean Effect Size of WASH Interventions on Educational Outcomes (school enrollment, absences and dropouts), by Intervention Category**

Intervention	Effect size		
	Boys	Girls	Total
HP and WT	(^)	0.193	0.120
HP, WT and SAN	(^)	0.124	0.091
HP, WT, SAN and WS	(^)	(^)	0.328 (***)
SAN	(^)	0.041 (***)	0.037 (***)

(\*\*\*) Significant at 99% level; (\*\*) Significant at 95% level; (\*) Significant at 90% level  
 (^) Effect sizes too few to estimate effect sizes separately  
 HP: hand washing promotion; WT: water treatment; SAN: sanitation (latrines); WS: water supply

<sup>22</sup> Rigorous gender-disaggregated WASH studies are in very short supply. For instance, a systematic review (Dickson et al., 2012) to identify and synthesize evidence of the impact of separate toilets for girls on their enrolment and attendance in schools could not find any evidence either for or against the impact of separate toilets for girls on their educational outcomes.

### 4.2.1.3. Pooled Effect Sizes by Intervention and Individual Educational Outcome

**Finding 4.4: The overall effect of WASH interventions on school enrollment is positive for both boys and girls, and is higher for girls than for boys**

The overall effect of WASH interventions on school enrollment (Table 4.3) is 0.033. Notably, the effect is nearly 40 percent higher for girls (0.037) than for boys (0.027).

**Table 4.3: Mean Effect Size of WASH Interventions on School Enrollment**

Intervention	Effect size		
	Boys	Girls	Total
All interventions (^)	0.027 (**)	0.037 (***)	0.033 (***)

(\*\*\*) Significant at 99% level; (\*\*) Significant at 95% level; (\*) Significant at 90% level  
(^ ) Effect sizes too few to estimate effect sizes by intervention

**Finding 4.5: WASH interventions have a positive effect on school absences and dropout rates**

WASH interventions have a combined positive effect of 0.180 on school absences. School dropout is reduced (a mean difference of 0.047)<sup>23</sup> through sanitation programs. (For details, see Annex 2.) This result is important given that there are high dropout rates among girls in developing countries and measures that enable girls to continue attendance in educational environments are essential to the promotion of gender parity and empowerment in those countries.

### 4.2.2. Water and Sanitation for Health Interventions and Health Outcomes

**Finding 4.6: When considered in combination, WASH interventions appear to have no effect on student health, but the effect varies when subsets of those combinations or single interventions are analyzed separately**

The combination of WASH interventions had no effect on student health (a very small and not statistically significant effect of 0.067). The only positive and significant effect (0.281) was through hand washing promotion interventions. When sanitation interventions were added to hand washing promotion and water treatment (HP, WT and SAN), the mean difference in effect (-0.239) becomes negative and statistically significant. This indicates that the experimental intervention influenced the outcome in favor of the control group, rather than the treatment

<sup>23</sup> The mean difference (more correctly, ‘difference in means’) is a standard statistic that measures the absolute difference between the mean value in two groups in a clinical trial. It estimates the amount by which the experimental intervention changes the outcome on average compared with the control (Higgins, 2014).

group.<sup>24</sup> Interestingly, when a water supply intervention is added (HP, WT, SAN and WS), the effect (0.106) of the combination of WASH interventions is no longer statistically different from zero.

**Table 4.4: Mean Effect Size of WASH Interventions on Health Outcomes (presence of E. coli, number of sick days, and number of sick students)**

Intervention	Effect size (^)
HP	0.281 (***)
HP and WT	-0.041
HP, WT and SAN	-0.239 (**)
HP, WT, SAN and WS	0.106
All interventions	0.067
(***) Significant at 99% level; (**) Significant at 95% level; (*) Significant at 90% level	
(^) Effect sizes too few to estimate effect sizes by gender	
HP: hand washing promotion; WT: water treatment; SAN: sanitation (latrines); WS: water supply	

**Finding 4.7: The addition of latrines to intervention schools has a negative effect on health as measured by E. coli contamination, especially for girls**

Table 4.5 sheds some light on the unexpected results depicted in Table 4.4 and summarized in the previous finding. Hygiene promotion and water treatment combinations (HP and WT) do not appear to reduce the risk of E. coli presence (a non-statistically significant effect of -0.087). However, the addition of new latrines (HP, WT and SAN) to intervention schools increases E. coli contamination on students' hands (a much larger and statistically significant effect of -0.524). It is important to note that the overall effect of WASH interventions involving the addition of latrines have a negative and statistically significant effect on health (-0.267) as measured by the risk of E. coli contamination.

<sup>24</sup> Positive values in the tables favor the treatment group and negative values favor the comparison or control group.

**Table 4.5: Mean Effect Size of WASH Interventions on the Presence of E.coli**

Intervention	Effect size		
	Boys	Girls	Total
HP and WT	(^)	(^)	-0.087
HP, WT and SAN	(^)	(^)	-0.524 (***)
All interventions	-0.045	-0.469 (***)	-0.267 (**)

(\*\*\*) Significant at 99% level; (\*\*) Significant at 95% level; (\*) Significant at 90% level  
(^) Effect sizes too few to estimate effect sizes by gender  
HP: handwashing promotion; WT: water treatment; SAN: sanitation (latrines); WS: water supply

Table 4.5 also shows significant interaction by gender. Although there is no demonstrable effect that these interventions have on males in comparison with children in the control schools (effect not statistically different from zero), there does appear to be a risk of E. coli infection among females (-0.469), suggesting that *efforts to increase usage of school latrines by constructing new facilities may pose a risk to children* in the absence of sufficient hygiene behavior change, daily provision of soap and water, and other body cleansing materials. Such complementary interventions are all the more critical due to the central role of sanitation in public health as reflected, for instance, in the poll of readers of the *British Medical Journal* in which sanitation was voted the greatest advance in public health in the last century (Mozynski, 2008). WHO and UNICEF go even further, stating that “without WASH (water, sanitation and hygiene), sustainable development is impossible” (WHO/UNICEF, 2015b).

**Finding 4.8: WASH interventions have a positive effect on student health when measured by the number of sick students**

WASH interventions did not demonstrate an impact on decreasing the number of sick days among school children (0.054). However, WASH interventions did have a statistically significant impact on decreasing the number of sick students (0.250) (See Annex 4).

**Finding 4.9: Hand washing and water treatment interventions may not be sustainable**

A sustainability evaluation of 55 pilot primary schools two and half years after the implementation of a hand washing and water treatment intervention in Kenya (Sabori et al., 2011) revealed that program activities were not successfully sustained in any of the 55 pilot schools. Another study in Pakistan (Luby et al., 2009) revealed a similar conclusion. A systematic review (Vindigni et al., 2011) of hand washing studies in community, school and health-care settings concluded that none of the studies reviewed was able to definitively document long-term behavior change, thereby challenging the sustainability of the various interventions.

### 4.3. Conclusions

- This meta-analysis provides evidence that WASH interventions have an overall positive effect on educational outcomes and that effect is higher for girls than for boys.
- The overall effect of WASH interventions on school enrollment is positive for both boys and girls, but is higher for girls.
- WASH interventions for which data are available reduce school absences and dropout rates.
- The effect size for all educational outcomes combined is higher for all WASH interventions combined than for subsets of interventions implemented separately. Neither hand washing promotion with water treatment, nor combined hand washing promotion, water treatment, and sanitation without water supply interventions, had any effect on educational outcomes.
- Water supply is a determinant factor in the success of WASH activities. The effect size and its statistical significance increase dramatically when water supply is added to other WASH interventions.
- When considered in combination, WASH interventions appear to have no effect on student health, but the effect varies when subsets of those combinations or single interventions are analyzed separately.
- Although WASH interventions may have a positive impact on health when measured by the number of sick students, the addition of latrines to intervention schools has a negative effect on health as measured by *E. coli* contamination, especially for girls. This result points to the conclusion that constructing new latrines may pose a risk to children in the absence of sufficient hygiene behavior change, daily provision of soap and water, and other body cleansing materials. To remedy this situation, the World Health Organization issued guidelines for water, sanitation, and hygiene implementation in schools in low-cost settings (Adams et al., 2009; Byford, 2014). It is expected that implementation of those guidelines will result in improved WASH interventions.
- There is no evidence to support the sustainability of school-based WASH interventions. Furthermore, there are no qualitative studies that investigated why school-based WASH interventions are not sustainable.
- The evidence summarized above shows that much remains to be known about the impact of school-based WASH interventions on educational and health outcomes. However, what we know provides moderate quality evidence<sup>25</sup> that WASH interventions should continue to be supported.

---

<sup>25</sup> As detailed in Section 5.2, moderate quality evidence is defined as evidence suggesting that WASH interventions probably improve educational and health outcomes.

#### 4.4. Limitation of the Findings

- WASH intervention studies are expensive to conduct because they often require large sample sizes to illustrate even minimal effects and require long term study time frames. Consequently, there is very little rigorous evidence on the educational and health benefits of WASH interventions.
- Studies were limited not only in number but also geographic coverage. Consequently, the reliability of generalizing findings to other contexts is severely restricted.
- There is very limited data on adherence and attrition.
- There is little evidence on the sustainability of WASH interventions in school settings. Sustainability was investigated in only two studies, which demonstrated that program activities were not sustained but it was not explored as to ‘why’.
- All studies reviewed have been of a short-term nature and no long-term impact information is available.
- There is limited systematic documentation on intervention processes and implementation. As such there is little evidence as to ‘why’ and ‘how’ WASH interventions succeed or fail.

#### 4.5. Future Research Directions

- The evaluation of WASH interventions at scale and subsequent effects on educational and health outcomes among pre-school and school-age children is required.
- Sanitation has been hailed as the greatest advance in public health in the last century. However, available evidence shows that efforts to increase usage of school latrines by constructing new facilities may have no effect on E. coli reduction among boys and may increase E. coli contamination among girls. However, research is needed on whether and under what conditions WHO guidelines for water, sanitation, and hygiene implementation in schools in low-cost settings would result in improved school-age children’s literacy and health.
- Qualitative data is needed as to why and how WASH interventions increase enrollment rate more among girls than boys.
- Further research on the mechanisms of action of WASH programs especially related to scale up is desperately needed. Detailed descriptions of program processes and implementation features would help explain the direction as well as the magnitude of program results.
- Qualitative data is conspicuously lacking as to why WASH programs struggle to be sustainable.
- Further evidence is needed on the cost-benefit of school-based WASH interventions targeting pre-school and school aged children.

## 5.0. Deworming

### 5.1. Introduction

As described in Section 1.2.2.3, pre-school and school-age children infected with parasitic worms are physically, nutritionally, and cognitively impaired. To control soil-transmitted helminth infections, WHO recommends health and hygiene education, provision of adequate sanitation, and periodic medicinal treatment. The WHO recommended medicines (albendazole and mebendazole) are effective, inexpensive and easy to administer by teachers and other non-medical personnel. They have also been used in millions of people with few and minor side-effects.

This section analyzes in detail the methodology and results Taylor-Robinson et al. used in a deworming meta-analysis performed for the Cochrane Collaboration review series, and also presents the subsequent debate on the impact of deworming that followed its publication. The meta-analysis is based on a series of studies that investigated the effects of deworming drugs for geohelminth worms, administered at health facilities, schools, and communities.

**Table 5.1: Major Characteristics of the Studies Included in the Deworming Meta-Analysis**

**Studies included in the review:** Randomized and quasi-randomized controlled trials.

**Location:** The included trials were undertaken in 23 different countries: Bangladesh (four trials); Ethiopia (two trials); Haiti (two trials); India (five trials); Indonesia (two trials); Jamaica (two trials); Kenya (five trials); South Africa (two trials); Vietnam (three trials); Zanzibar (two trials); Benin, Botswana, Cameroon, Guatemala, Java, Malaysia, Nigeria, Philippines, Sierra Leone, Tanzania, Uganda, Zaire (one trial in each); China, Philippines and Kenya (one multicenter trial).

**Population:** Children aged 16 years or less. Children were recruited from school populations in 20 trials, communities in 16 trials, and in health facilities or by health workers in six trials. Thirty-five trials were based on mass targeted treatment of an unselected population. Fourteen trials were conducted in populations where worms were of high prevalence or intensity, 10 in populations with moderate prevalence and low intensity, and 11 in populations with low prevalence and low intensity. Seven trials studied children screened and selected on the basis of high worm loads.

**Intervention:** Deworming drugs for geohelminth worms, administered at health facilities, schools, and communities. Investigation of effects after a single dose, and after multiple doses.

**Control groups:** Placebo or no treatment was used as a control in the majority of studies. Other studies used vitamin A, vitamin C, or calcium powder. There were 13 trials where both the treatment and control group received nutritional supplementation: multi-nutrient; vitamin B; iron; vitamin A; or child health package.

**Effects of interventions:**

- Major outcome measures: weight; height; hemoglobin; psychometric tests of cognition; measures of physical well-being (Harvard Step Test); school attendance.
- The effects were grouped into trials where children were screened for infection; and trials treating whole populations (a single dose of deworming drug, after multiple doses with follow up for up to a year, and after multiple doses with follow up of one year or more).

Source: Taylor-Robinson et al. 2012



## 5.2. Findings

As described in Table 5.1 and listed in Table 5.2 through Table 5.5, the effect of deworming is measured through six outcomes: school attendance, weight gain, height gain, hemoglobin level, physical well-being, and cognition. Interventions are divided into four categories: (1) a single dose of deworming drug given to children infected with worms in population screened for intestinal helminths (parasitic worms); (2) a single dose of deworming drug given to all children living in an endemic area; (3) multiple doses of deworming drug given to all children (follow-up for up to a year); and (4) multiple doses of deworming drug given to all children (follow-up for over a year). The four intervention categories and their effects are described in turn below. The findings are presented in the form of tables derived from data in Taylor-Robinson et al. (Taylor-Robinson et al., 2012).

### **Finding 5.1: A single dose of deworming drug given to children infected with worms in population screened for intestinal helminths may improve children's weight and hemoglobin status, but the evidence base is small**

Table 5.2 shows that weight gain attributed to a single dose of deworming drug given to children infected with worms in population screened for intestinal helminths increased, together with hemoglobin level. However, the meta-analysis notes that the effect on weight gain (0.58 kg mean difference)<sup>26</sup> is based only on three trials covering 149 participants, and the effect on hemoglobin level (0.37 g/dl mean difference) is based on two trials covering 108 participants. The effect on cognition cannot be determined because it is based on very low-quality evidence<sup>27</sup> (two trials, one of which did not report the outcome and the second reported improvement in only 3 out 10 tests of cognitive function).

---

<sup>26</sup> As noted earlier, the mean difference estimates the amount by which the experimental intervention changes the outcome on average compared with the control.

<sup>27</sup> The quality of evidence used by the authors is based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group, an informal collaboration of people with an interest in addressing the shortcomings of present grading systems in health care. GRADE has developed a transparent approach to grading quality of evidence and strength of recommendations. Many international organizations have provided input into the development of the approach and have started using it. For example, clinical actions are likely to differ depending on whether one concludes that the evidence that a specific drug reduces the risk of stroke in patients is convincing (high quality) or that it is unconvincing (low quality). Similarly, guidelines that recommend that patients with a given health condition should be treated may suggest that patients should definitely be treated, implying that treatment is warranted in all patients, or that patients should probably be treated, implying that treatment may not be warranted in all patients. Using the GRADE system, the meta-analysis classifies outcomes as follows: high quality evidence means deworming improves the outcome under consideration; moderate quality evidence means deworming probably improves the outcome; low quality evidence means deworming may improve the outcome; very low quality evidence means we do not know whether deworming improves the outcome.

**Table 5.2: Effect of a Single Dose of Deworming Drug Given to Children Infected with Worms in Population Screened for Intestinal Helminths**

Outcome	Effect			
	Improves	Probably improves	May improve	Do not know whether it improves
School attendance	--	--		--
Weight gain			X	
Height gain	--	--	--	--
Hemoglobin level	--	--	X	--
Physical well being	--	--	--	--
Cognition				X

Notes: (1) Assessing the evidence using GRADE: high quality evidence = deworming improves the outcome under consideration; moderate quality evidence = deworming probably improves the outcome; low quality evidence = deworming may improve the outcome; very low quality evidence = we do not know whether deworming improves the outcome.  
(2) (–) means outcomes not measured in the studies included in the review; (X) measures the effect of the interventions on the corresponding outcome.  
Source: Authors’ compilation based on information in Taylor-Robinson et al. 2012

**Finding 5.2: A single dose of deworming drug given to all children living in an endemic area may have a positive effect on physical well-being and cognition. There is minimal evidence on increasing hemoglobin levels. There is no evidence that support deworming for weight gain.**

The effects of de-worming on physical well-being (as measured by the Harvard Step Test<sup>28</sup>) is estimated at a mean difference of 6 in two trials covering individuals in one high-prevalence infection area in Kenya (Table 5.3). The first trial reported no effect or a negative effect on cognition. However, this study did not report the actual data. The second trial, covering 1,361 participants, reported no effects on physical well being or cognition. The effects of deworming on hemoglobin levels were studied across three trials totaling 1,005 participants. The mean difference between hemoglobin levels was not statistically significant in any of the trials. The effect of deworming on weight gain indicated no effect in seven of nine trials.

<sup>28</sup> The Harvard step test is a test of aerobic fitness. It has been found to be a good measurement of general fitness by measuring a person's ability to return to a normal heart rate after a strenuous exercise. The more quickly the heart rate returns to resting, the better shape the person is in.

**Table 5.3: Effect of a Single Dose of Deworming Drug Given to All Children Living in an Endemic Area**

Outcome	Effect			
	Improves	Probably improves	May improve	Do not know whether it improves
School attendance	--	--	--	--
Weight gain				X
Height gain	--	--	--	--
Hemoglobin level (*)				
Physical well being			X	
Cognition			X	

Notes: (1) Assessing the evidence using GRADE: high quality evidence = deworming improves the outcome under consideration; moderate quality evidence = deworming probably improves the outcome; low quality evidence = deworming may improve the outcome; very low quality evidence = we do not know whether deworming improves the outcome.  
(2) (–) means outcomes not measured in the studies included in the review; (X) measures the effect on the corresponding outcome.  
(3) (\*) It probably has no effect on hemoglobin levels: In three trials, meta-analysis of hemoglobin difference was not statistically significant.

Source: Authors' compilation based on information in Taylor-Robinson et al. 2012

**Finding 5.3: Multiple doses of deworming drug given to all children (follow-up for up to a year) may have little or no effect on weight gain, hemoglobin level, cognition, and school attendance**

The effects of multiple doses of deworming drug given to all children (follow-up for up to a year) are described in Table 5.4. Deworming increased weight gain in one trial in a high-prevalence location, decreased weight in one trial in a low-prevalence area, but had no effect elsewhere. The effects on hemoglobin levels in four trials totaling 807 participants, a meta-analysis calculated a mean difference of only 0.01 g/dl intervention groups. Another study utilized formal testing to measure various aspects of intellectual development (i.e. cognition) across three intervention trials (30,571 participants; 75 clusters and 571 individually randomized participants). Deworming had no effect on cognition. In two trials (30,243 participants; 75 clusters and 243 individually randomized participants), deworming had a small effect on increasing attendance but only by four percent compared to the control group.

**Table 5.4: Effect of Multiple Doses of Deworming Drug Given to All Children (follow-up for up to a year)**

Outcome	Effect			
	Improves	Probably improves	May improve	Do not know whether it improves
School attendance			X	
Weight gain			X	
Height gain	--	--	--	--
Hemoglobin level			X	
Physical well being			X	
Cognition			X	

Notes: (1) Assessing the evidence using GRADE: high quality evidence = deworming improves the outcome under consideration; moderate quality evidence = deworming probably improves the outcome; low quality evidence = deworming may improve the outcome; very low quality evidence = we do not know whether deworming improves the outcome.  
(2) (–) means outcomes not measured in the studies included in the review; (X) measures the effect on the corresponding outcome.

Source: Authors' compilation based on information in Taylor-Robinson et al. (2012)

**Finding 5.4: Multiple doses of deworming drug given to all children (follow-up for over a year) may improve weight, hemoglobin status, and cognition, but the effect on height and school attendance is not known.**

The effects of multiple doses of deworming drug given to all children (follow-up for over a year) are depicted in Table 5.5. Deworming increased weight gain in one early trial in a low-prevalence location. However, this effect was not reproduced in two subsequent trials in the same location, or in higher-prevalence locations. Two trials measured de-worming and the effects on hemoglobin levels and subsequent intellectual development outcomes. Neither study reported an effect deworming had on hemoglobin levels and subsequent intellectual development effects. However, it is not advisable to compare these two trials directly as they measured different outcomes related to intellectual development. There was only one trial that measured the effect of multiple deworming on school attendance. There was a slight difference between treatment and control group at five percent.

**Table 5.5: Effect of Multiple Doses of Deworming Drug Given to All Children (follow-up for over a year)**

Outcome	Effect			
	Improves	Probably improves	May improve	Do not know whether it improves
School attendance				X
Weight gain			X	
Height gain				X
Hemoglobin level			X	
Physical well being			X	
Cognition			X	
Notes: (1) Assessing the evidence using GRADE: high quality evidence = deworming improves the outcome under consideration; moderate quality evidence = deworming probably improves the outcome; low quality evidence = deworming may improve the outcome; very low quality evidence = we do not know whether deworming improves the outcome.				
(2) (–) means outcomes not measured in the studies included in the review; (X) measures the effect on the corresponding outcome.				
Source: Authors' compilation based on information in Taylor-Robinson et al. (2012)				

### 5.3. Conclusions

- A single dose of deworming drug given to children infected with worms in population screened for intestinal helminths *shows some promise*. It may improve weight and hemoglobin status, but the evidence base for subsequent outcomes from improved weight and hemoglobin levels is very limited.
- The administration of deworming medication in settings without intestinal helminth screening and in endemic areas, *may* have an impact on weight gain, physical well-being, and cognition. Currently *there is insufficient evidence to recommend this strategy due to a limited number of studies and small sample sizes*.
- De-worming medications had no effect on hemoglobin levels.
- De-worming had a minimal effect on school attendance (5%).
- Since the results are based on a limited number of countries and settings, they are difficult to generalize to other locations.

### 5.4 Limitation of Findings

- Nearly all of the evidence that supports de-worming in school settings comes from trials with small sample sizes and conducted over a relatively short time period.

- The generalizability of the results is restricted due to limited geographic diversity among the studies.
- Measurements of cognition and intellectual development are not consistent between studies, thus restricting comparability of results.

#### **5.4.1. Further Discussion on Limitations: should deworming policies be re-evaluated?**

##### *5.4.1.1 Deworming: Not a Panacea?*

Some authors (Garner, 2012; Hawkes, 2013) have interpreted the conclusions in Taylor-Robinson et al. (2012) as suggesting that the benefits of routine deworming policies may need to be reevaluated<sup>29</sup>. Garner et al state in their 2012 study that “Deworming schoolchildren to rid them of intestinal helminths seems a good idea in theory, but the evidence for it just doesn’t stack up. We want policy makers to look at the evidence and the message and consider if deworming is as good as it is cracked up to be” (Garner, 2012). Hawkes related a similar belief in the BMJ article entitled “Deworming Debunked”, “Deworming has been hailed as a panacea: a simple, cheap, and effective way of improving growth, raising brain power, and improving the educational and employment prospects of millions of children. Not if you read the latest revision of the Cochrane review on the subject, published in July this year by a team from the Liverpool School of Tropical Medicine” (Hawkes, 2013).

##### **5.4.1.2. Has the Meta-Analysis “Stacked the Deck” Against Deworming?**

The Cochrane Collaboration is a well-respected source of information for evidence-based decision-making. Systematic reviews and meta-analyses are the cornerstones of evidence for decision-making. The Cochrane review method has strict selection criteria study inclusion in order to ensure that bias is minimized objectivity is increased. However, systematic reviews and meta-analysis are not perfect. There are many research groups who have pointed out several limitations of the meta-analysis, including inadequate consideration of environmental and pathological factors; insufficient length of trial follow-up; oversight of epidemiological externalities; marginalization of long-term impact; and omission of cost-effectiveness considerations (Hotez et al., 2012; Bundy et al., 2013; Baird et al., 2014; Ozier, 2014; Ahuja, 2015). All of these factors are important issues in de-worming interventions.

##### *5.4.1.3. Deworming: a Cornerstone for Neglected Tropical Disease Control*

---

<sup>29</sup> Based on seven experiments, McEwan (2014) also concludes that the mean effect size of deworming drugs is close to zero (0.013) and not statistically significant.

Hotez et al. argue that deworming should remain a cornerstone for neglected tropical disease control despite the conclusions of the meta-analysis because the investigation did not take into account four essential dimensions of deworming:

1. The five major soil-transmitted intestinal worms should not be treated as a single group because they are each quite different in the type of disease and pathology they produce and their nutritional effects on their human hosts.
2. Not all intestinal worms respond to the same deworming medication.
3. Only moderate and heavy intestinal helminth infections typically cause measurable disease.
4. The ability to detect a health improvement from deworming may also depend on whether children in a given area simultaneously suffer from low nutritional intake or if they are co-infected with other pathogens (Hotez et al., 2012).

#### *5.4.1.4. Insufficient Length of Follow-up*

Bundy et al. made clear that in 18 out of 42 studies, the duration of follow-up was six months or less, and four studies reported a month or less of follow-up (Bundy et al., 2013). Neglecting time factor consideration dilutes the potential impact of the intervention and dramatically affects the power of these studies to detect meaningful differences in the outcomes they aim to document. For example, sustained blood loss and inflammation due to worm infection of the intestinal tract has cumulative consequences that can be measured only over relatively long periods of follow-up.

#### *5.4.1.5. Epidemiological Externalities*

An externality is an effect of a decision by one party on another party whose interests were not taken into account when the decision was made. Epidemiological externalities, or “spillover effects,” occur when treatment of an easily accessible portion of the population benefits even those who remain untreated.<sup>30</sup>

Several studies (Kremer & Miguel, 2004; Baird et al., 2011; Baird et al., 2014; Ozier, 2014; Ajuja et al., 2015) have shown that deworming helps break the cycle of transmission. These studies have proven that treating children for parasitic worms, benefits untreated children in the same school. Furthermore, this benefit ‘spills over’ to nearby schools and children in those schools benefit from lower worm load and improved attendance at school.<sup>31</sup>

---

<sup>30</sup> This free benefit is analogous to the “herd immunity” or “community immunity” benefit from vaccination: When a critical portion of a community is immunized against a contagious disease, most members of the community are protected against that disease because there is little opportunity for an outbreak.

<sup>31</sup> Such an outcome has been used to argue that deworming tablets should be a priority for free distribution. One study showed that when free deworming was replaced with a low cost-sharing fee, treatment declined by 80 percent. In addition, sicker children were no more likely to pay for the drugs than their healthier schoolmates, suggesting

Externalities from deworming interventions may follow a newly discovered second pathway through its effects on malaria and HIV infection. Research has shown that worms may worsen malaria (Druilhe et al., 2015; Shapiro et al., 2005) and exacerbate HIV transmission (Finchman et al., 2013; Walson et al., 2008; Walson et al., 2009).

#### *5.4.1.6. Long-Term Impact*

A long-term study implemented in western Kenya in 1998-2001 (Baird et al., 2011) tracked the students who participated in the original deworming program over the following decade. Researchers collected data on health, educational attainment, living standards, and employment status of students from the first study – who, by that time, were between 19- and 26-years old. Analysis of that information showed that deworming improved self-reported health, increased total schooling and mean hours worked. Among other benefits, treatment also led to shifts into more lucrative employment (from food crops to cash crops in agriculture, and from low-skilled casual labor to better-paid, full-time jobs in fields such as manufacturing), and improved living standards.

#### *5.4.1.7. Cost-effectiveness*

Cost-effectiveness analysis helps to identify interventions that use resources most efficiently. Cost-effectiveness is an evaluation method that examines the costs relative to the outcomes, or results, of interventions. The cost-effectiveness analysis uses a specific outcome measure that must be common among the alternatives being considered.

Cost analysis of deworming programs was performed using data from six interventions that have been rigorously tested through randomized evaluations in Madagascar, Kenya and Malawi (Kremer and Miguel, 2007; J-PAL, 2011; J-PAL, 2012). Using cost projections for a large-scale treatment program, J-PAL estimates that deworming costs \$4.55 per Disability Adjusted Life Year (DALY) averted, indicating that the deworming program was “highly cost-effective.”<sup>32</sup>

As demonstrated in Table 5.6, using the same data the analysis also found that school-based deworming is one of the most cost-effective means of increasing school attendance. At nearly 14 additional school years gained per \$100, deworming ranks second only to information on returns to education provided to parents in Madagascar (about 20 additional school years), way ahead of unconditional cash transfer for girls in Malawi (0.02 years), conditional cash transfer for girls’

---

that the fees did not direct treatment to those who needed it most. As summarized in J-PAL (2011), “charging small fees in an attempt to balance access and ‘sustainability’ may be the worst of both worlds, as small fees raise little revenue, but dramatically reduce access to important products for the poor.”

<sup>32</sup> Disability Adjusted Life Years (DALYs) are a common measure of the burden of disease, expressed as years of life lost to illness and premature death. The World Health Organization considers an intervention to be “highly cost-effective” if it costs less than the national GDP per capita for each DALY averted (the relevant threshold for Kenya was \$1,560 in 2009).



attendance in Malawi (0.09 years), Merit scholarships for girls in Kenya (0.27 years), and free primary school uniforms in Kenya (0.71 years).

It is important to note that this calculation is based on a small-scale deworming program through an NGO; since a larger-scale program would have a lower cost per child, deworming is likely to be more cost-effective than indicated in Table 5.6. It is equally important to note that future income gains accruing to treated children would enhance the cost-effectiveness of deworming even further: The long-term study estimates that the initial investment in deworming generates a return<sup>33</sup> of more than 80 percent per year through higher earnings.

**Table 5.6: Cost-effectiveness of deworming through primary schools in Kenya (additional school years gained per \$100 expenditure)**

Intervention	Additional school years gained
Information on returns to education, for parents (Madagascar)	20.7
Deworming through primary schools (Kenya)	13.9
Free primary school uniforms (Kenya)	0.71
Merit scholarships for girls (Kenya)	0.27
Conditional cash transfer for girls' attendance (Malawi)	0.09
Unconditional cash transfer for girls (Malawi)	0.02

Source: Kremer & Miguel, 2007; J-PAL Policy Bulletin, 2012; J-PAL Policy Bulletin, 2011

## 5.5. Future Research Directions

- Further research is needed on the effectiveness of de-worming interventions in various prevalence settings. Moreover, evidence is needed on the cost-benefit of deworming at various prevalence levels.
- There is a need for a cost effectiveness study comparing deworming interventions to other interventions that target reduced infection rates and improved hemoglobin levels.

<sup>33</sup> The return on investment is the amount of money earned as a percentage of the total value of the assets invested.

## References

Note: (\*) denotes study included in the meta-analysis

Adams, J., Bartram, J., Chartier, Y., & Sims, J. (2009). Water, Sanitation and Hygiene Standards for Schools in Low-cost Settings.

Alderman, H., & Bundy, D. (2012). School Feeding Programs and Development: Are We Framing the Question Correctly?. *The World Bank Research Observer*, 27(2), 204-221.

Adelman, S., Gilligan, D. O., & Lehrer, K. (2007). How effective are food-for-education programmes. *2020 Focus Brief on the World's Poor and Hungry People*.

(\*) Adukia, A. (2014). Sanitation and education. *Cambridge, MA: Harvard University*.

Ahuja, A., Baird, S., Hicks, J. H., Kremer, M., Miguel, E., & Powers, S. (2015). *When Should Governments Subsidize Health? The Case of Mass Deworming* (No. 21148). National Bureau of Economic Research, Inc.

Aiken, A. M., Davey, C., Hargreaves, J. R., & Hayes, R. J. (2015). Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a pure replication. *International journal of epidemiology*, dyv127.

Alderman, H., & Bundy, D. (2012). School Feeding Programs and Development: Are We Framing the Question Correctly?. *The World Bank Research Observer*, 27(2), 204-221.

Allison, P. D. (2010). *Survival analysis using SAS: a practical guide*. Sas Institute.

Awasthi, S., Pande, V. K., & Fletcher, R. H. (2000). Effectiveness and cost-effectiveness of albendazole in improving nutritional status of pre-school children in urban slums. *Indian pediatrics*, 37(1), 19-30.

Baird, S., Hicks, J. H., Kremer, M., & Miguel, E. (2014). Worms at Work: Public finance implications of a child health investment. *University of California at Berkeley, -mimeo*.

- Baird, S., Hicks, J. H., Kremer, M., & Miguel, E. (2011). Worms at work: long-run impacts of child health gains. *Berkeley: University of California at Berkeley.*
- (\*) Barger, B., Maiga, H., Traore, O. B., Tekete, M., Tembine, I., Dara, A., ... & Djimde, A. A. (2009). Intermittent preventive treatment using artemisinin-based combination therapy reduces malaria morbidity among school-aged children in Mali. *Tropical Medicine & International Health, 14*(7), 784-791.
- Bethony, J., Brooker, S., Albonico, M., Geiger, S. M., Loukas, A., Diemert, D., & Hotez, P. J. (2006). Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. *The Lancet, 367*(9521), 1521-1532.
- Blanton, E., Ombeki, S., Oluoch, G. O., Mwaki, A., Wannemuehler, K., & Quick, R. (2010). Evaluation of the role of school children in the promotion of point-of-use water treatment and handwashing in schools and households—Nyanza Province, Western Kenya, 2007. *The American journal of tropical medicine and hygiene, 82*(4), 664-671.
- Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97-111.
- (\*) Bowen, A., Ma, H., Ou, J., Billhimer, W., Long, T., Mintz, E., ... & Luby, S. (2007). A cluster-randomized controlled trial evaluating the effect of a handwashing-promotion program in Chinese primary schools. *The American journal of tropical medicine and hygiene, 76*(6), 1166-1173.
- (\*) Brooker, S., & Halliday, K. (2015). Impact of malaria control and enhanced literacy instruction on educational outcomes among school children in Kenya.

- (\* ) Brooker, S., Inyega, H., Estambale, B., Njagi, K., Juma, E., Jones, C., ... & Jukes, M. (2013). Impact of malaria control and enhanced literacy instruction on educational outcomes among Kenyan school children: a multi-sectoral, prospective, randomised evaluation.
- Brooker, S., Guyatt, H., Omumbo, J., Shretta, R., Drake, L., & Ouma, J. (2000). Situation analysis of malaria in school-aged children in Kenya—what can be done?. *Parasitology today*, 16(5), 183-186.
- Bundy, D. A., Walson, J. L., & Watkins, K. L. (2013). Worms, wisdom, and wealth: why deworming can make economic sense. *Trends in parasitology*, 29(3), 142-148.
- Bundy, D. A. (2011). *Rethinking School Health: A key component of education for all*. World Bank Publications.
- Bundy, D. A., Kremer, M., Bleakley, H., Jukes, M. C., & Miguel, E. (2009a). Deworming and development: Asking the right questions, asking the questions right. *PLoS Neglected Tropical Diseases*, 3(1), e362.
- Bundy, D., Burbano, C., Grosh, M., Gelli, A., Jukes, M., & Drake, L. (2009b). Rethinking school feeding. *Social safety nets, child development and the education*.
- Bundy, D. A. P., Lwin, S., Osika, J. S., McLaughlin, J., & Pannenberg, C. O. (2000). What Should Schools Do About Malaria?. *Parasitology Today*, 16(5), 181.
- Byford, T. (2014). Water, sanitation and hygiene standards for schools in low-cost settings. *International Journal of Environmental Studies*, 71(3), 409-410.
- Callender, J. E., Walker, S. P., Grantham-McGregor, S. M., & Cooper, E. S. (1998). Growth and development four years after treatment for the Trichuris dysentery syndrome. *Acta paediatrica*, 87(12), 1247-1249.

Centers for Disease Control (n.d). Global WASH-Related Diseases and Contaminants.

[http://www.cdc.gov/healthywater/wash\\_diseases.html](http://www.cdc.gov/healthywater/wash_diseases.html) (accessed November 27 2015).

(\*). Clarke, S. E., Jukes, M. C., Njagi, J. K., Khasakhala, L., Cundill, B., Otido, J., ... & Brooker, S. (2008). Effect of intermittent preventive treatment of malaria on health and education in schoolchildren: a cluster-randomized, double-blind, placebo-controlled trial. *The Lancet*, 372(9633), 127-138.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

Cohn, L. D., & Becker, B. J. (2003). How Meta-Analysis Increases Statistical Power. *Psychological Methods*, 8(3), 243-253.

Conn, K. (2014). *Identifying Effective Education Interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations* (Doctoral dissertation, Columbia University).

Copenhagen Consensus (2012). Expert Panel Findings.

[http://www.copenhagenconsensus.com/sites/default/files/outcome\\_document\\_updated\\_1105.pdf](http://www.copenhagenconsensus.com/sites/default/files/outcome_document_updated_1105.pdf). Retrieved November 27 2015.

Croke, K. (2014). The long run effects of early childhood deworming on literacy and numeracy: Evidence from Uganda. *Harvard School of Public Health, -mimeo*.

CROMPTON, D. W. T. (2000). The public health importance of hookworm disease. *Parasitology*, 121(S1), S39-S50.

Cutler, D., Fung, W., Kremer, M., Singhal, M., & Vogl, T. (2010). Early-life malaria exposure and adult outcomes: Evidence from malaria eradication in India. *American Economic Journal: Applied Economics*, 72-94.

Davey, C., Aiken, A. M., Hayes, R. J., & Hargreaves, J. R. (2015). Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a

- statistical replication of a cluster quasi-randomized stepped-wedge trial. *International journal of epidemiology*, dyv128.
- De Silva, N. R. (2003). Impact of mass chemotherapy on the morbidity due to soil-transmitted nematodes. *Acta tropica*, 86(2), 197-214.
- Deworm the World (2010). School-Based Deworming: A Planner's Guide to Proposal Development for National School-Based Deworming Programs.
- Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Jama*, 263(10), 1385-1389.
- Dickson, K., Freeman, M., & Javidi, L. (2011). *What impact does the provision of separate toilets for girls at schools have on their primary and secondary school enrolment, attendance and completion?: A systematic review of the evidence*. London, UK: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Drake, T. L., Okello, G., Njagi, K., Halliday, K. E., Jukes, M. C., Mangham, L., & Brooker, S. (2011). Cost analysis of school-based intermittent screening and treatment of malaria in Kenya. *Malar J*, 10, 273.
- Dreibelbis, R., Greene, L. E., Freeman, M. C., Saboori, S., Chase, R. P., & Rheingans, R. (2013). Water, sanitation, and primary school attendance: A multi-level assessment of determinants of household-reported absence in Kenya. *International Journal of Educational Development*, 33(5), 457-465.
- Druilhe, P., Tall, A., & Sokhna, C. (2005). Worms can worsen malaria: towards a new means to roll back malaria?. *Trends in parasitology*, 21(8), 359-362.

- Duval, S., & Tweedie, R. (2000). A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95(449), 89-98.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, 315(7109), 629-634.
- Ehrhardt, S., Burchard, G. D., Mantel, C., Cramer, J. P., Kaiser, S., Kubo, M., ... & Mockenhaupt, F. P. (2006). Malaria, anemia, and malnutrition in African children—defining intervention priorities. *Journal of Infectious Diseases*, 194(1), 108-114.
- Ekvall, H. (2003). Malaria and anemia. *Current opinion in hematology*, 10(2), 108-114.
- Erinoso, A. O., & Bamgboye, E. A. (1988). Sickness absenteeism in a Nigerian polytechnic. *African journal of medicine and medical sciences*, 17(1), 57-61.
- Evans, D., & Ghosh, A. (2008). Prioritizing educational investments in children in the developing world.
- Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological methods*, 17(1), 120.
- Fernando, D., DE SILVA, D. A. M. A. N. I., Carter, R., Mendis, K. N., & Wickremasinghe, R. (2006). A randomized, double-blind, placebo-controlled, clinical trial of the impact of malaria prevention on the educational attainment of school children. *The American Journal of Tropical Medicine and Hygiene*, 74(3), 386-393.
- Fincham, J. E., Markus, M. B., & Adams, V. J. (2003). Could control of soil-transmitted helminthic infection influence the HIV/AIDS pandemic. *Acta tropica*, 86(2), 315-333.

- (\*) Freeman, M. C., Clasen, T., Dreifelbis, R., Saboori, S., Greene, L. E., Brumback, B., ... & Rheingans, R. (2014). The impact of a school-based water supply and treatment, hygiene, and sanitation programme on pupil diarrhoea: a cluster-randomized trial. *Epidemiology and Infection*, 142(02), 340-351.
- (\*) Freeman, M. C., Greene, L. E., Dreifelbis, R., Saboori, S., Muga, R., Brumback, B., & Rheingans, R. (2012). Assessing the impact of a school-based water treatment, hygiene and sanitation programme on pupil absence in Nyanza Province, Kenya: a cluster-randomized trial. *Tropical Medicine & International Health*, 17(3), 380-391.
- Garg, R., Lee, L. A., Beach, M. J., Wamae, C. N., Ramakrishnan, U., & Deming, M. S. (2002). Evaluation of the integrated management of childhood illness guidelines for treatment of intestinal helminth infections among sick children aged 2–4 years in western Kenya. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 96(5), 543-548.
- Garner, Paul (2012). Deworming: not all it's cracked up to be?  
<http://blogs.plos.org/speakingofmedicine/2012/07/18/should-deworming-policies-in-the-developing-world-be-reconsidered/> (accessed 6/18/2015)
- Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2011). *Impact evaluation in practice*. World Bank Publications.
- (\*) Greene, L. E., Freeman, M. C., Akoko, D., Saboori, S., Moe, C., & Rheingans, R. (2012). Impact of a school-based hygiene promotion and sanitation intervention on pupil hand contamination in Western Kenya: a cluster randomized trial. *The American journal of tropical medicine and hygiene*, 87(3), 385-393.



- (\*) Garn, J. V., Greene, L. E., Dreibelbis, R., Saboori, S., Rheingans, R. D., & Freeman, M. C. (2013). A cluster-randomized trial assessing the impact of school water, sanitation, and hygiene improvements on pupil enrollment and gender parity in enrollment. *Journal of water, sanitation, and hygiene for development: a journal of the International Water Association*, 3(4).
- Grantham-McGregor, S. M., Powell, C. A., Walker, S. P., & Himes, J. H. (1991). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study. *The Lancet*, 338(8758), 1-5.
- Geerligs, P. D. P., Brabin, B. J., & Eggelte, T. A. (2003). Analysis of the effects of malaria chemoprophylaxis in children on haematological responses, morbidity and mortality. *Bulletin of the World Health Organization*, 81(3), 205-216.
- (\*) Guinan, M., McGuckin, M., & Ali, Y. (2002). The effect of a comprehensive handwashing program on absenteeism in elementary schools. *American journal of infection control*, 30(4), 217-220.
- Hall, A., Horton, S., & de Silva, N. (2009). The costs and cost-effectiveness of mass treatment for intestinal nematode worm infections using different treatment thresholds. *Public Library of Science Neglected Diseases*, 3(3), 31.
- Hall, A., Hewitt, G., Tuffrey, V., & De Silva, N. (2008). A review and meta-analysis of the impact of intestinal worms on child growth and nutrition. *Maternal & child nutrition*, 4(s1), 118-236.
- (\*) Halliday, K. E., Okello, G., Turner, E. L., Njagi, K., Mcharo, C., Kengo, J., ... & Brooker, S. J. (2014). Impact of intermittent screening and treatment for malaria among school children in Kenya: a cluster randomised trial.
- (\*) Hawkes, N. (2013). Deworming debunked. *BMJ*, 346, e8558.

- Higgins, J. P. T. (2014). Green S. Cochrane handbook for systematic reviews of interventions Version 5.1. 0 [updated March 2011]. The Cochrane Collaboration, 2011. [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
- Hotez, Peter J., Neeraj Mistry and Alan Fenwick (2012). Deworming should remain an essential cornerstone for NTD control. <http://blogs.plos.org/speakingofmedicine/2012/07/18/should-deworming-policies-in-the-developing-world-be-reconsidered/> (accessed 6/18/2015)
- International Initiative for Impact Evaluation (3ie) (2008) *Principles for Impact Evaluation*. 3ie, New Delhi.
- J-PAL Policy Bulletin (2012). “Deworming: A Best Buy for Development.” Cambridge, MA: Abdul Latif Jameel Poverty Action Lab.
- J-PAL Policy Bulletin (2011). “The Price is Wrong.” Cambridge, MA: Abdul Latif Jameel Poverty Action Lab.
- Kassebaum, N. J., Jasrasaria, R., Naghavi, M., Wulf, S. K., Johns, N., Lozano, R., ... & Murray, C. J. (2014). A systematic analysis of global anemia burden from 1990 to 2010. *Blood*, 123(5), 615-624.
- Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2010). *Handbook on impact evaluation: quantitative methods and practices*. World Bank Publications.
- Korenromp, E. L., Armstrong-Schellenberg, J. R., Williams, B. G., Nahlen, B. L., & Snow, R. W. (2004). Impact of malaria control on childhood anaemia in Africa—a quantitative review. *Tropical Medicine & International Health*, 9(10), 1050-1065.
- Kremer, M., & Miguel, E. (2007). The Illusion of Sustainability. *The Quarterly journal of economics*, 122(3), 1007-1065.

- Le, H. T., Brouwer, I. D., Nguyen, K. C., Burema, J., & Kok, F. J. (2007). The effect of iron fortification and de-worming on anaemia and iron status of Vietnamese schoolchildren. *British Journal of Nutrition*, 97(05), 955-962.
- Legovini, A. (2010). Development Impact Evaluation Initiative: A World Bank–Wide Strategic Approach to Enhance Development Effectiveness. Draft Report to the Operational Vice Presidents, World Bank, Washington, DC.
- Luby, S. P., Agboatwalla, M., Bowen, A., Kenah, E., Sharker, Y., & Hoekstra, R. M. (2009). Difficulties in maintaining improved handwashing behavior, Karachi, Pakistan. *The American journal of tropical medicine and hygiene*, 81(1), 140-145.
- Mathew, K., Zachariah, S., Shordt, K., Snel, M., Cairncross, S., Biran, A., & Schmidt, W. P. (2009). The sustainability and impact of school sanitation, water and hygiene education in southern India. *Waterlines*, 28(4), 275-292.
- Mathegana, M., Chauke, L., & Otieno, F. (2001). Improvement of environmental health and hygiene practices-case study in the Northern Province. *Water Science & Technology*, 44(6), 109-117.
- McGovern-Dole International Food for Education and Child Nutrition Program. 74 FR 13072, Mar. 26, 2009
- McEwan, P. J. (2014). Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments. *Review of Educational Research*, 0034654314553127.
- Menendez, C., Fleming, A. F., & Alonso, P. L. (2000). Malaria-related anaemia. *Parasitology today*, 16(11), 469-476.

- Miguel, E., & Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 159-217.
- Mozynski, P. (2008). BMA says inadequate sanitation is a global crisis. *BMJ: British Medical Journal*, 336(7636), 117.
- Mwanri, L., Worsley, A., & Masika, J. (2001). School and anaemia prevention: Current reality and opportunities—A Tanzanian case study. *Health promotion international*, 16(4), 321-331.
- Nga, T. T., Winichagoon, P., Dijkhuizen, M. A., Khan, N. C., Wasantwisut, E., Furr, H., & Wieringa, F. T. (2009). Multi-Micronutrient–Fortified Biscuits Decreased Prevalence of Anemia and Improved Micronutrient Status and Effectiveness of Deworming in Rural Vietnamese School Children. *The Journal of nutrition*, 139(5), 1013-1021.
- O'reilly, C. E., Freeman, M. C., Ravani, M., Migele, J., Mwaki, A., Ayalo, M., ... & Quick, R. (2008). The impact of a school-based safe water and hygiene programme on knowledge and practices of students and their parents: Nyanza Province, western Kenya, 2006. *Epidemiology and Infection*, 136(01), 80-91.
- (\*). Ozier, O. W. (2014). Exploiting externalities to estimate the long-term effects of early childhood deworming. *World Bank Policy Research Working Paper*, (7052).
- (\*). Patel, M. K., Harris, J. R., Juliao, P., Nygren, B., Were, V., Kola, S., ... & Quick, R. (2012). Impact of a hygiene curriculum and the installation of simple handwashing and drinking water stations in rural Kenyan primary schools on student health and hygiene practices. *The American journal of tropical medicine and hygiene*, 87(4), 594-601.
- Price, R. N., Simpson, J. A., Nosten, F., Luxemburger, C., Hkirjaroen, L., ter Kuile, F. E. I. K. O., ... & White, N. J. (2001). Factors contributing to anemia after uncomplicated

falciparum malaria. *The American journal of tropical medicine and hygiene*, 65(5), 614-622.

Prüss-Üstün A, Bos R, Gore F, Bartram J (2008). Safer water, better health: costs, benefits and sustainability of interventions to protect and promote health. World Health Organization, Geneva. [http://apps.who.int/iris/bitstream/10665/43840/1/9789241596435\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/43840/1/9789241596435_eng.pdf). (accessed November 27 2015).

- Quintero, J. P., Siqueira, A. M., Tobón, A., Blair, S., Moreno, A., Arévalo-Herrera, M., ... & Valencia, S. H. (2011). Malaria-related anaemia: a Latin American perspective. *Memórias do Instituto Oswaldo Cruz*, *106*, 91-104.
- Rowland, M. G., Cole, T. J., & Whitehead, R. G. (1977). A quantitative study into the role of infection in determining nutritional status in Gambian village children. *British journal of nutrition*, *37*(03), 441-450.
- Riley, R. D., Higgins, J. P., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *Bmj*, *342*.
- Saboori, S., Mwaki, A., Porter, S., Okech, B., Freeman, M., & Rheingans, R. (2011). Sustaining school hand washing and water treatment programmes: lessons learned and to be learned. *Waterlines*, *30*(4), 298-311.
- Sachs, J., & Malaney, P. (2002). The economic and social burden of malaria. *Nature*, *415*(6872), 680-685.
- Sakti, H., Nokes, C., Hertanto, W., Hendratno, S., Hall, A., & Bundy, D. A. (1999). Evidence for an association between hookworm infection and cognitive function in Indonesian school children. *Tropical Medicine & International Health*, *4*(5), 322-334.
- (\*) Shapiro, A. E., Tukahebwa, E. M., Kasten, J., Clarke, S. E., Magnussen, P., Olsen, A., ... & Brooker, S. (2005). Epidemiology of helminth infections and their relationship to clinical malaria in southwest Uganda. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *99*(1), 18-24.
- Shiff, C., Checkley, W., Winch, P., Premji, Z., Minjas, J., & Lubega, P. (1996). Changes in weight gain and anaemia attributable to malaria in Tanzanian children living under

- holoendemic conditions. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 90(3), 262-265.
- Simeon, D. T., Grantham-McGregor, S. M., & Wong, M. S. (1995). Trichuris trichiura infection and cognition in children: results of a randomized clinical trial. *Parasitology*, 110(04), 457-464.
- (\* ) Simwaka, B. N., Simwaka, K., & Bello, G. (2009). Retrospective analysis of a school-based malaria treatment programme demonstrates a positive impact on health and education outcomes in Mangochi district, Malawi. *Journal of development effectiveness*, 1(4), 492-506.
- Stephenson, L. S., Latham, M. D., & Adams, E. J. (1995). Weight Gain of Kenyan School Children Infected with Hookworm, Trichuris Trichiura and Ascaris Lumbricoides Is Improved Following Once and Twice Yearly Treatment with Albendazole. *Pediatrics*, 95(6), 895-895.
- Sur, D., Saha, D. R., Manna, B., Rajendran, K., & Bhattacharya, S. K. (2005). Periodic deworming with albendazole and its impact on growth status and diarrhoeal incidence among children in an urban slum of India. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 99(4), 261-267.
- (\* ) Talaat, M., Afifi, S., Dueger, E., El-Ashry, N., Marfin, A., Kandeel, A., ... & El-Sayed, N. (2011). Effects of hand hygiene campaigns on incidence of laboratory-confirmed influenza and absenteeism in schoolchildren, Cairo, Egypt. *Emerg Infect Dis*, 17(4), 619-625.

(\*) Taylor-Robinson, D. C., Maayan, N., Soares-Weiser, K., Donegan, S., & Garner, P. (2012).

Deworming drugs for soil-transmitted intestinal worms in children: effects on nutritional indicators, haemoglobin and school performance. *The Cochrane Library*.

Tipton, E. (2014). Small sample adjustments for robust variance estimation with meta-regression.

Trape, J. F., Lefebvre-Zante, E., Legros, F., Druilhe, P., Rogier, C., Bouganali, H., & Salem, G. (1993). Malaria morbidity among children exposed to low seasonal transmission in Dakar, Senegal and its implications for malaria control in tropical Africa. *The American journal of tropical medicine and hygiene*, 48(6), 748-756.

UNICEF. Malaria. December 24 2013.

[http://www.unicef.org/search/search.php?q\\_en=malaria&go.x=13&go.y=9](http://www.unicef.org/search/search.php?q_en=malaria&go.x=13&go.y=9); accessed 9/26/2015.

UNICEF. (2010). Raising clean hands: Advancing learning, health, and participation through WASH in schools. *New York: UNICEF*.

UNESCO (2006). Understandings of Literacy (Chapter 6) in Education for All Monitoring Report . [http://www.unesco.org/education/GMR2006/full/chapt6\\_eng.pdf](http://www.unesco.org/education/GMR2006/full/chapt6_eng.pdf), (accessed 9/23/2015).

United Nations (2015). The Millennium Development Goals Report 2015; United Nations: New York, NY, USA.

Vindigni, S. M., Riley, P. L., & Jhung, M. (2011). Systematic review: handwashing behaviour in low-to middle-income countries: outcome measures and behaviour maintenance. *Tropical Medicine & International Health*, 16(4), 466-477.



- Walson, J. L., Herrin, B. R., & John-Stewart, G. (2009). Deworming helminth co-infected individuals for delaying HIV disease progression. *Cochrane Database Syst Rev*, 3(3).
- Walson, J. L., Otieno, P. A., Mbuchi, M., Richardson, B. A., Lohman-Payne, B., Macharia, S. W., ... & John-Stewart, G. C. (2008). Albendazole treatment of HIV-1 and helminth co-infection: a randomized, double blind, placebo-controlled trial. *AIDS (London, England)*, 22(13), 1601.
- White, H. (2010). A contribution to current debates in impact evaluation. *Evaluation*, 16(2), 153-164.
- White, H. (2009). Theory-based impact evaluation: principles and practice. *Journal of development effectiveness*, 1(3), 271-284.
- World Bank (2015). [http://web.worldbank.org/archive/website01213/WEB/0\\_C-101.HTM](http://web.worldbank.org/archive/website01213/WEB/0_C-101.HTM) (accessed 8/3/2015)
- World Bank (2011). New Reasons Why School-Based Deworming is Smart Development Policy <http://blogs.worldbank.org/education/rethinking-deworming>; (accessed 8/3/2015)
- World Bank. (n.d.). The Development Impact Evaluation (DIME) Initiative, Project Document, World Bank, Washington, D.C.
- World Food Program. (2013). The State of School Feeding 2013. World Food Program, Via C.G. Viola, 68-70, Rome 00148, Italy World Health Organization (2015a). “Soil-transmitted helminth infections” Fact sheet N°366. (<http://www.who.int/mediacentre/factsheets/fs366/en/>, accessed June 19, 2015).
- World Health Organization. (2002). *The world health report 2002: reducing risks, promoting healthy life*. World Health Organization.

World Health Organization. (2005). Malaria control today: current WHO recommendations. *Geneva: WHO*.

World Health Organization/UNICEF (2015a). <http://www.wssinfo.org/definitions-methods>. Accessed November 27 2015.

World Health Organization/UNICEF (2015b) Lack of sanitation for 2.4 billion people is undermining health improvements. Joint WHO/UNICEF News Release. June 30 2015.

World Health Organization (2015). Deworming to combat the health and nutritional impact of helminth infections. <http://www.who.int/elena/titles/deworming/en/> (accessed June 19, 2015).

## Annexes

### Annex 1: Major Characteristics of the Studies Included in the Meta-Analyses

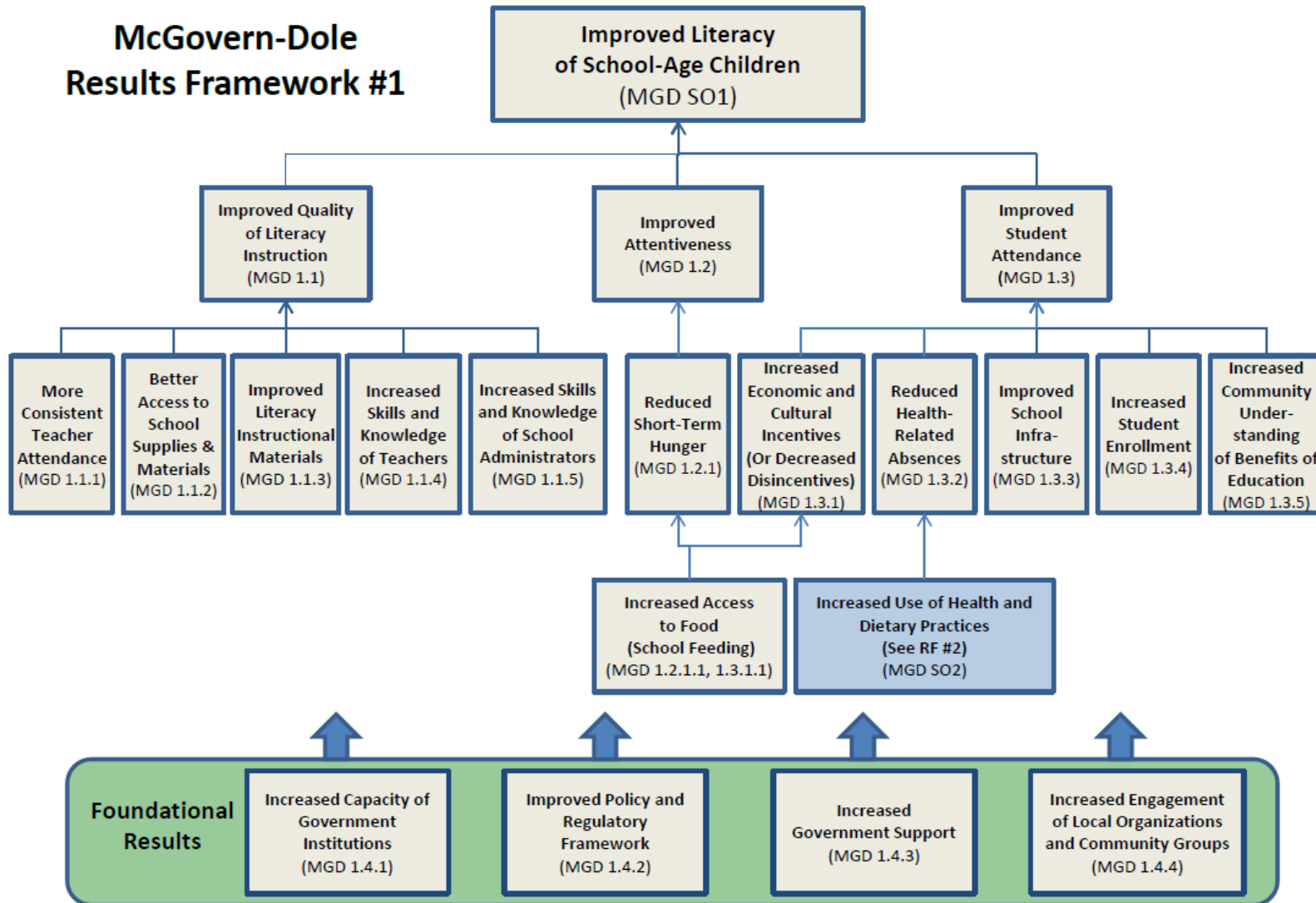
Author	Title	Type of Study	Location	Age/Grade	Sample	Intervention	Outcome measures
<b>Water and Sanitation for Health (WASH)</b>							
Adukia, 2014	Sanitation and Education	Multivariate analysis	India	Primary school-age children	Annual administrative school-level data	Presence of unisex latrines and sex-specific latrines Duration: academic years 2002 through 2006	Enrollment, attendance, and dropout rates
Freeman et al., 2012	Assessing the impact of a school-based water treatment, hygiene and sanitation programme on pupil absence in Nyanza Province, Kenya: a cluster-randomized trial	Cluster randomized trial; multivariate analysis	Kenya, Nyanza Province	primary school-age children	Public primary schools randomly assigned to three groups; 198 schools selected out of 1,084 in 4 districts because they had over 25 pupils per latrine. Out of those, 135 were randomly selected for the study; 5,989 children supplied absence information	135 schools were randomly assigned to 1 of 3 study arms after baseline evaluation: G1: hygiene promotion and water treatment (HP & WT); G2: HP & WT plus sanitation (latrines); G3: the control group which received all interventions at the conclusion of the study 2,015 pupils in G1; 2,008 in G2; 2,013 in G3	Attendance
Freeman, 2013	The impact of a school-based water supply and treatment, hygiene, and sanitation programme on pupil diarrhoea: a cluster-randomized trial	Cluster randomized trial; Multivariate regression model	Nyanza province, Kenya	School-age children attending school	135 schools with nearby dry-season water source and 50 schools without nearby water; school size varies (100 to 900 pupils); surveys: 25 pupils randomly selected from register of grades 4–8	In the water-available group 135 schools were randomly allocated into one of three intervention arms of 45 schools each: G1: hygiene promotion and water treatment (HP&WT); G2: HP&WT plus sanitation (latrines). G3: control group. 50 water-scarce schools randomly assigned to two equal groups getting (1) water supply (WS) improvement plus HP&WT, and latrines or (2) control-school group	Incidence of diarrhea

Author	Title	Type of Study	Location	Age/Grade	Sample	Intervention	Outcome measures
Garn et al., 2013	A cluster-randomized trial assessing the impact of school water, sanitation, and hygiene improvements on pupil enrollment and gender parity in enrollment.	Randomized control trial	Nyanza Province, Kenya	Primary school-age children	Schools divided into 2 groups based on access to water supply during the dry season; 2 separate randomized controlled trials; 135 randomly selected schools, stratified by district.	135 selected sample schools were assigned to three groups of 45 each: G1: HP&WT; G2: HP&WT and SAN (latrines); and G3: control group. Enrollment data at pre-intervention (2007) and two following years (2008, 2009).	Enrollment
Greene et al., 2012	Impact of a School-Based Hygiene Promotion and Sanitation Intervention on Pupil Hand Contamination in Western Kenya: A Cluster Randomized Trial	Randomized control trial	Four districts of Nyanza Province in Western Kenya	Primary school-age children	135 public primary schools in 3 random groups	Schools randomly assigned to three groups: hygiene promotion and water treatment (HP&WT); HP&WT plus latrines (HP&WT&SAN); and a control group. Hand rinse samples were analyzed for E coli presence at a university laboratory	Hand contamination with E Coli
Talaat et al., 2011)	Effects of Hand Hygiene Campaigns on Incidence of Laboratory-confirmed Influenza and Absenteeism in Schoolchildren, Cairo, Egypt	Randomized control trial	Cairo, Egypt	Primary school-age children	60 elementary schools out of a total of 725 schools in Cairo: 30 in intervention and 30 in control group	Children in intervention schools were required to wash hands twice daily, and health messages were provided through entertainment activities. School nurses collected nasal swabs from students with influenza-like-illness. Duration: 12-weeks	Attendance; diarrhea; conjunctivitis; influenza-like illnesses
<b>Malaria</b>							
Barger et al., 2009	Intermittent preventive treatment using artemisinin-based combination therapy reduces malaria morbidity among school-aged children in Mali	Randomized control trial	Mali, Kollo district	School-age children (6–13 years)	296 school children	Intermittent preventive treatment; students received 2 full treatment doses, 2 months apart; IPT or placebo. Duration: 11 months	Anemia

Author	Title	Type of Study	Location	Age/Grade	Sample	Intervention	Outcome measures
Brooker et al., 2015	Impact of malaria control and enhanced literacy instruction on educational outcomes among school children in Kenya: a multisectoral, prospective, randomised evaluation	Factorial cluster randomized trial	Kenya, Southern coast	Classes 1 and 5	5,233 children in 101 government primary schools 24 months.	Intermittent Screening and Treatment of malaria Duration: 24 months	Anemia; school scores
Clarke et al., 2008	Effect of intermittent preventive treatment of malaria on health and education in schoolchildren: a cluster-randomised, double-blind, placebo-controlled trial	Randomized, double blind control trial	Western Kenya	School-age children	30 primary schools with 6,768 children; 3,535 children IPT; 3,223 placebo	Schools randomly assigned to treatment or dual placebo. Intermittent preventive treatment children received 3 treatments at 4-month intervals. Duration: 12 months	Anemia; school scores, sustained attention
Fernando et al., 2006	A Randomized, Double-blind, Placebo-Controlled Clinical Trial of the Impact of Malaria Prevention on the Educational Attainment of School Children	Randomized, double blind control trial	Sri Lanka, southern region	Grades 1 to 5	587 children; grades 1-5 in 4 schools; and residents in the area were randomly assigned to chloroquine (n=295) or placebo (n=292)	At weekly school visits, one chloroquine tablet or placebo given to each child after a meal under the direct supervision of a research assistant or the teacher. Duration: 9 months.	Language and mathematics scores; attendance
Halliday et al., 2014	Impact of Intermittent Screening and Treatment for Malaria among School Children in Kenya: A Cluster Randomised Trial	Double cluster randomized trial.	Kenya, southern coast	Grades 1 and 5	5,233 children in 101 government primary schools	Schools randomly assigned to 4 equal sets of school groups; intermittent screening and treatment for malaria. Duration: 24 months	Anemia; sustained attention; language and arithmetic scores

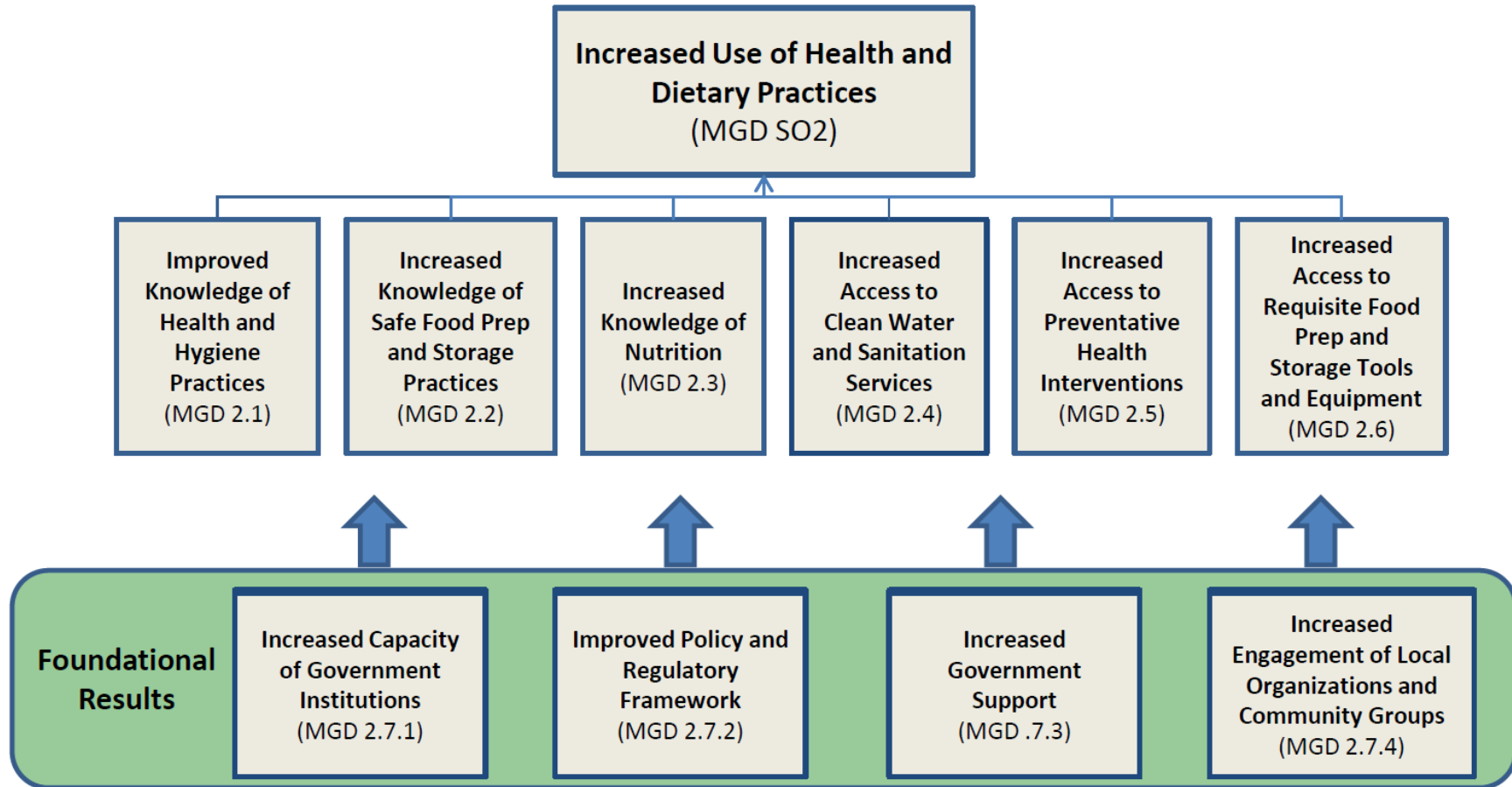
Annex 2: MGD Results Framework

**McGovern-Dole  
Results Framework #1**



**A Note on Foundational Results:** These results can feed into one or more higher-level results. Causal relationships sometimes exist between foundational results.

# McGovern-Dole Results Framework #2



**A Note on Foundational Results:** These results can feed into one or more higher-level results. Causal relationships sometimes exist between foundational results.

### Annex 3: Analysis of Publication Bias

Publication bias refers to the selective publication of studies with a particular outcome --- the greater likelihood that studies with positive results will be published, with the result that most treatments tend to be less effective in practice than the research suggests (see, for instance, Dickersin 1990 or Ferguson et al. 2012). Small studies are at the greatest risk of being lost because, with small samples, only very large effects are likely to be significant and those with small and moderate effects are likely to be unpublished. Large studies are likely to be published regardless of statistical significance.

Funnel plots and Egger tests (Egger et al., 1997) enable the quantification of publication bias. Funnel plots provide a graphical depiction of publication bias, based on the rationale that small studies are more likely to be unreported than large studies, a phenomenon referred to as the “file drawer problem.” The y-axis, showing the standard error corresponding to sample size, is inverted with large studies measured at the top (see funnel plots below). The asymmetry in the plot, as highlighted by the lack of small sample studies which report findings below the average effect at the vertical line, suggests evidence for publication bias.

In the absence of publication bias the studies will be distributed symmetrically throughout the scatter plot. In the possible presence of bias, the bottom of the plot would tend to show a higher concentration of studies on one side of the plot than the other. The funnel plot can also be used to identify outliers -- observations that are numerically distant from the rest of the data. Identification of outliers in meta-analysis can be used to conduct sensitivity analysis (with and without outliers).

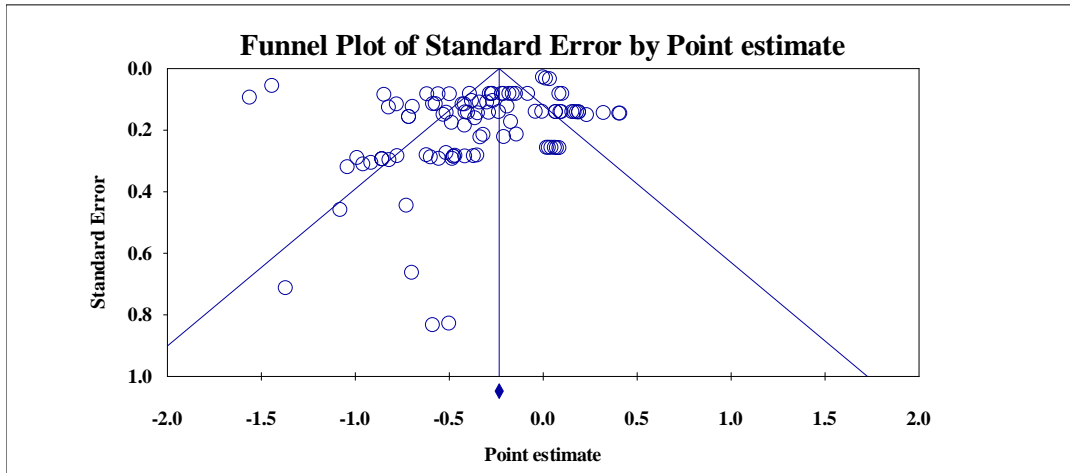
Given the difficulties in accurately assessing asymmetry by visual inspection, statistical tests are recommended. The most widely used statistical test is Egger’s test. Egger’s test is based on two variables: (i) normalized effect estimate (meta-analysis estimate divided by its standard error), and (ii) precision (reciprocal of the standard error of the estimate). The test is based on a simple linear regression to test for intercept  $\beta_0=0$ ; i.e., the null hypothesis that intercept  $b=0$  (or the null hypothesis that there is no funnel plot asymmetry). In this case, the regression line will run through the origin. If the intercept  $b$  deviates from zero (the origin), the deviation provides a measure of asymmetry -- the larger the deviation from zero, the larger the asymmetry. (It is for this reason that Egger’s test is also referred to as “Egger’s test of the intercept.”)

The following two plots are from a biased and unbiased analysis, as reflected in their corresponding funnel plots and Egger’s test statistics.

Example of a biased analysis (effect of malaria interventions on educational and health outcomes for all children described in this study):



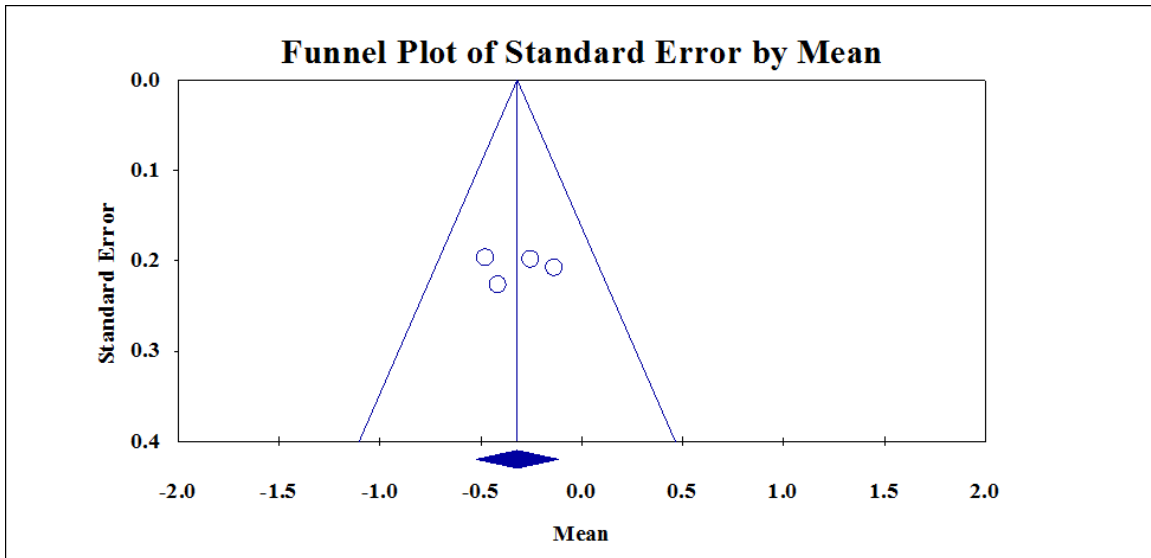
- The effect sizes are not symmetrically distributed
- The Egger's test shows that the intercept (at -1.43674) is statistically different from zero (P-value = 0.01991)



<b>Egger's regression intercept</b>	
Intercept	-1.46374
Standard error	0.61815
95% lower limit (2-tailed)	-2.69093
95% upper limit (2-tailed)	-0.23655
t-value	2.36792
df	95.00000
P-value (1-tailed)	0.00996
P-value (2-tailed)	0.01991

Example of an unbiased analysis (effect of WASH intervention on school absences for girls described in this study):

- The effect sizes are symmetrically distributed
- The Egger’s test shows that the intercept (at -0.45117) is not statistically different from zero (P-value = 0.96228)



Egger's regression intercept	
Intercept	-0.45117
Standard error	8.45175
95% lower limit (2-tailed)	-36.81611
95% upper limit (2-tailed)	35.91377
t-value	0.05338
df	2.00000
P-value (1-tailed)	0.48114
P-value (2-tailed)	0.96228

Assessing publication bias involves: (1) broadening the search to the non-published “grey literature” to reduce the bias; and (2) conducting sensitivity analysis. The present meta-analysis has made every attempt to minimize the publication bias by conducting a thorough search for non-published studies that included conference proceedings, technical reports, dissertations, and theses. Despite this effort, the funnel plots and Egger’s tests presented in Annexes 4-7 indicate that publication bias could not always be eliminated.

Assessing publication bias can also be conducted through imputation of missing studies by using “trim and fill” analysis -- a sensitivity analysis method that extends beyond the scope of this

study. Another method of assessing the potential for publication bias is to calculate the “fail-safe N,” the number of studies whose effect size is zero or negative that would be needed to increase the P-value for the meta-analysis to above 0.05. However, the Cochrane Handbook for Systematic Reviews of Interventions notes that “this and other methods are not recommended for use in Cochrane reviews” (Higgins et al. 2014).

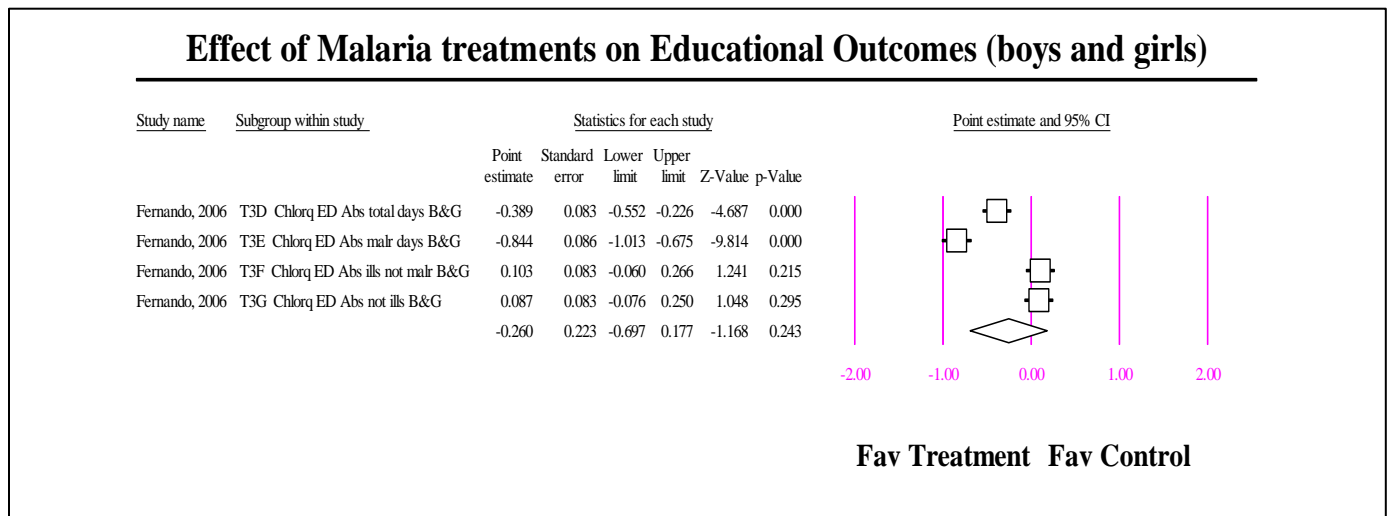
It is very important to note, however, that the presence of publication bias means that the pooled effect sizes may be overestimated and the response ratio effect size estimated by trim and fill corresponds to a reduction in average effect size. Since the effect sizes estimated in this meta-analysis are (when statistically significant) consistently “very small” to “medium,” the trim and fill analysis are expected to make those effect sizes even smaller --- with no major implications on the conclusions and learning agenda presented in this study.

## Annex 4. Technical Data Used for Analysis: Forest Plots, Funnel Plots, Egger's Tests and Detailed Statistics

Data in this annex were used to derive the findings in Section 3.0 (empirical evidence for malaria interventions), Section 4.0 (empirical evidence for WASH interventions) and Annex 1 (analysis of publication bias). Annexes 2.1-2.4, which served as a basis for constructing the tables in Section 3.0 and Section 4.0, provide detailed statistics of effect sizes, including standard errors, t-values, degrees of freedom, confidence intervals, statistical significance, heterogeneity statistics, funnel plots and Egger's tests. Number of studies in the statistical tables below refers to the number of effect sizes, not the number of studies themselves.

### Annex 4.1: Malaria Interventions, Educational Outcomes

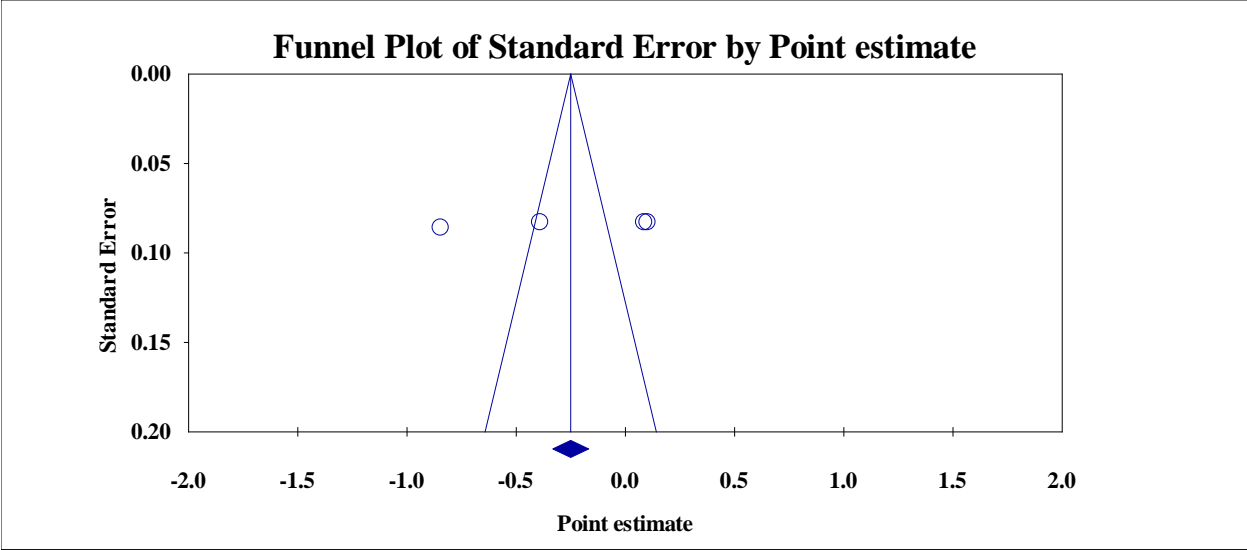
#### Pooled Effect Sizes of Chloroquine Interventions on Educational Outcomes All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	(0.260)	-	0.223

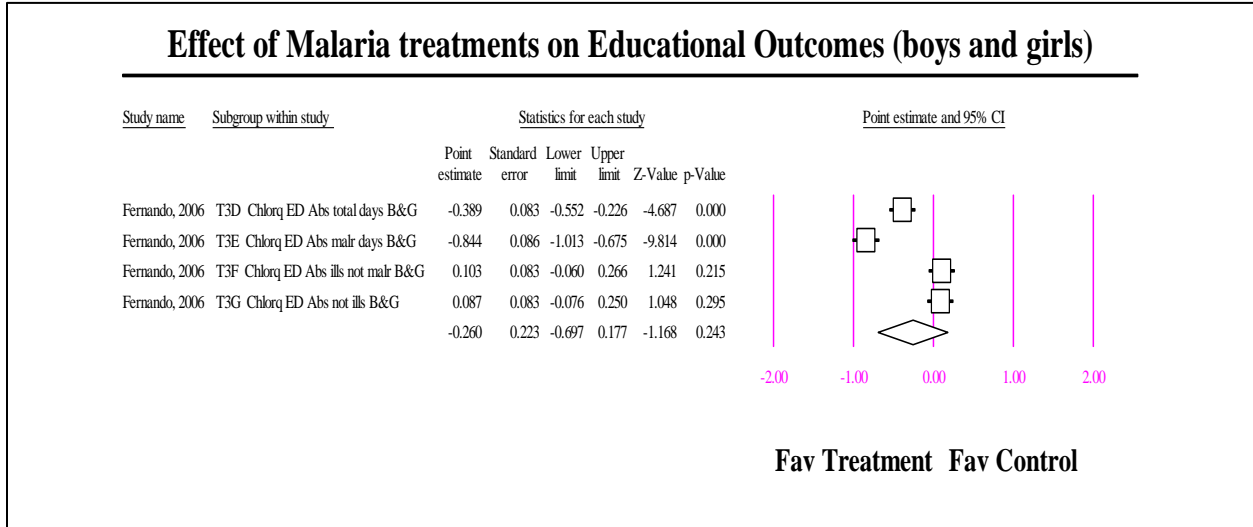
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
85.084	3.000	-	96.474	0.192	0.162	0.026	0.438



### Egger's regression intercept

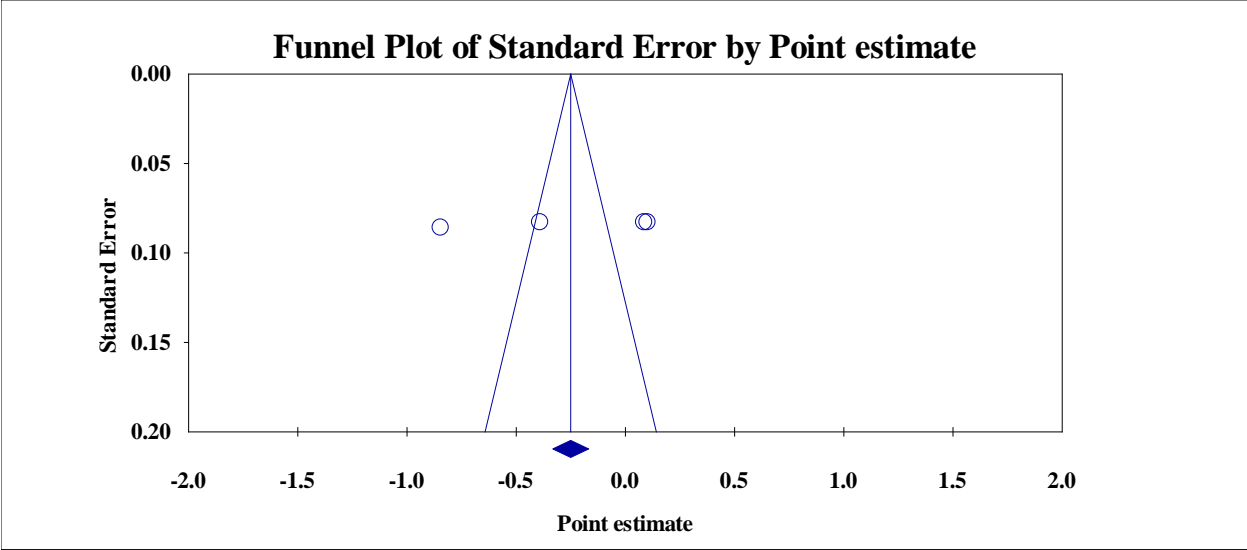
Intercept	-259.22222
Standard error	110.52931
95% lower limit (2-tailed)	-734.79146
95% upper limit (2-tailed)	216.34702
t-value	2.34528
df	2.00000
P-value (1-tailed)	0.07182
P-value (2-tailed)	0.14364

## Pooled Effect Sizes of Chloroquine Interventions on Absence All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	(0.260)	-	0.223

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
85.084	3.000	-	96.474	0.192	0.162	0.026	0.438

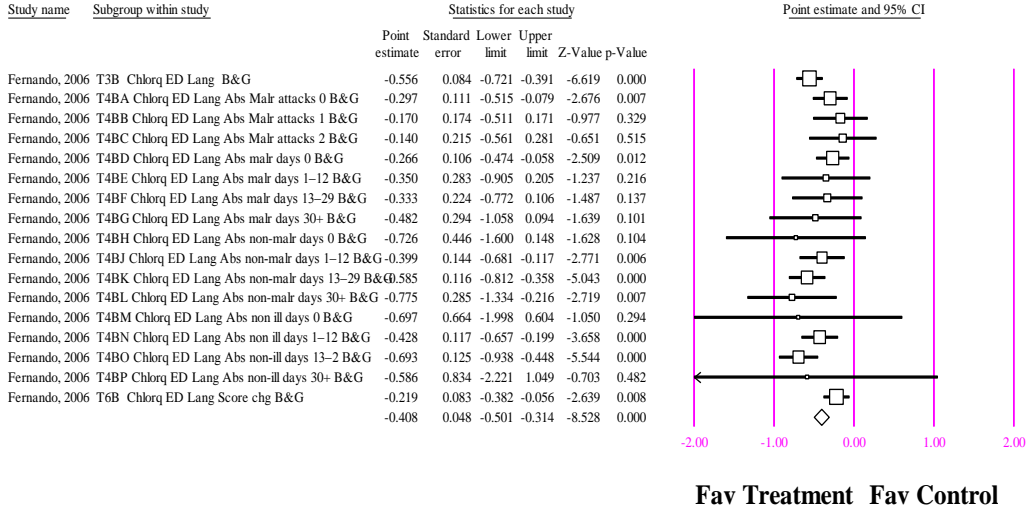


### Egger's regression intercept

Intercept	-259.22222
Standard error	110.52931
95% lower limit (2-tailed)	-734.79146
95% upper limit (2-tailed)	216.34702
t-value	2.34528
df	2.00000
P-value (1-tailed)	0.07182
P-value (2-tailed)	0.14364

## Pooled Effect Sizes of Chloroquine Interventions on Language All Children

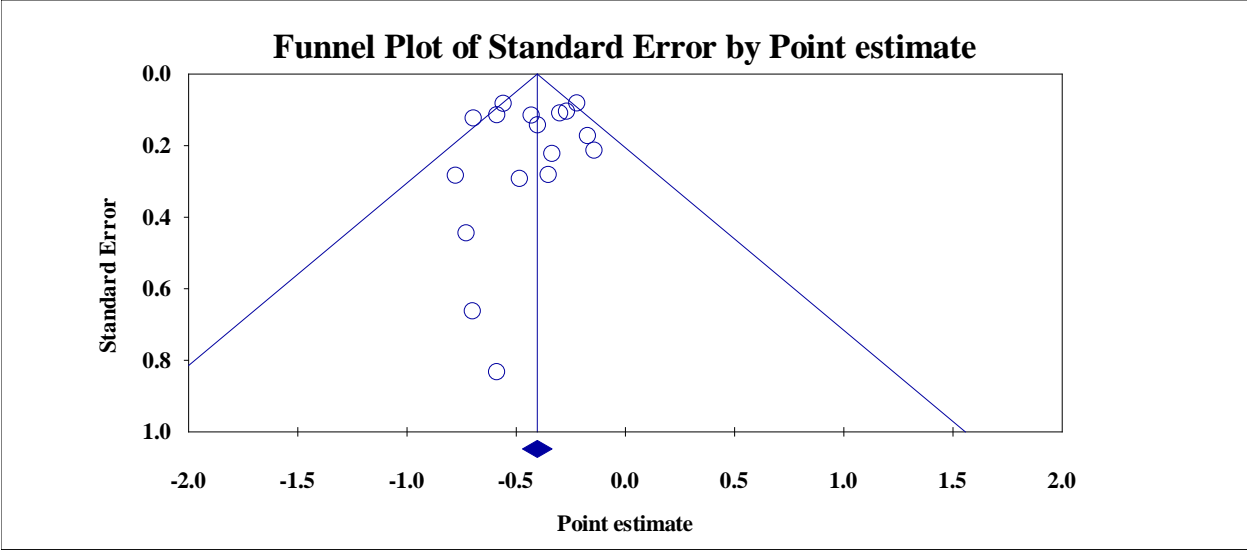
### Effect of Malaria treatments on Educational Outcomes (boys and girls)



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	17.000	(0.408)	***	0.048

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
24.671	16.000	0.076	35.148	0.012	0.012	0.000	0.108



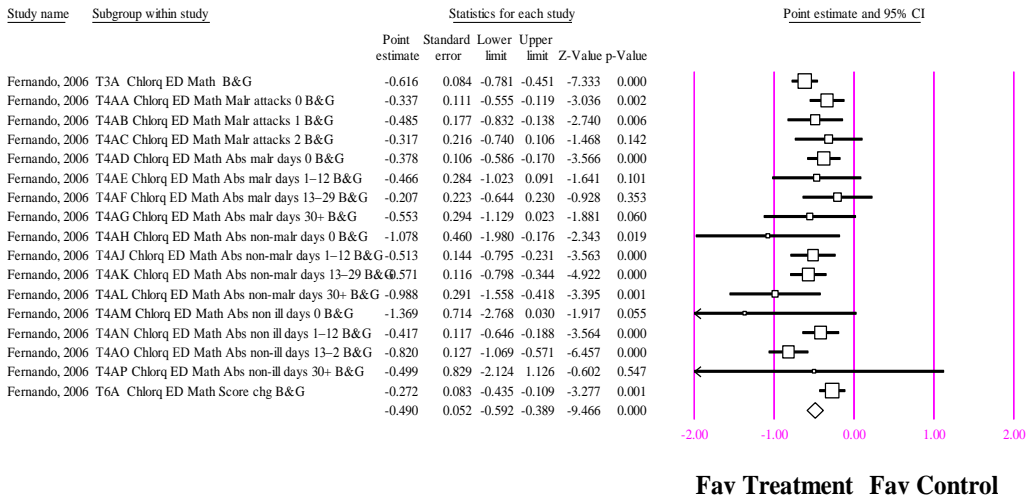


**Egger's regression intercept**

Intercept	-0.32687
Standard error	0.64583
95% lower limit (2-tailed)	-1.70341
95% upper limit (2-tailed)	1.04968
t-value	0.50612
df	15.00000
P-value (1-tailed)	0.31006
P-value (2-tailed)	0.62013

**Pooled Effect Sizes of Chloroquine Interventions on Math  
All Children**

## Effect of Malaria treatments on Educational Outcomes (boys and girls)



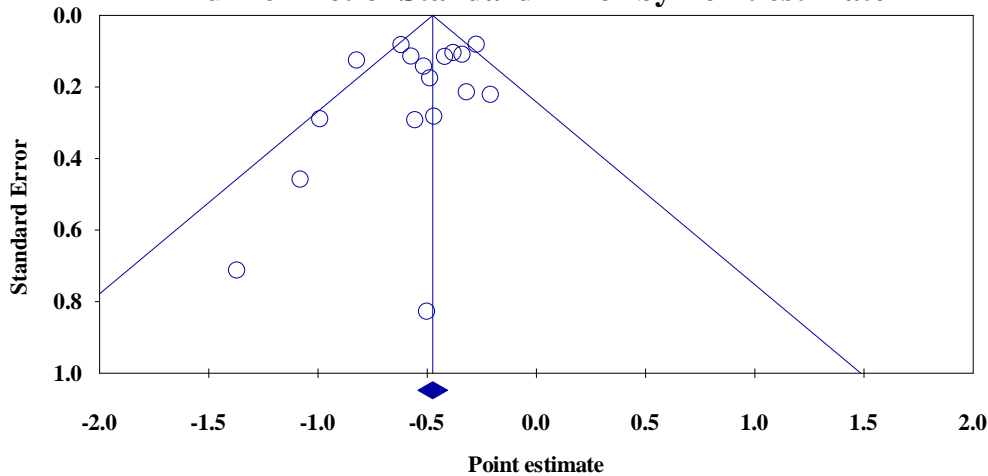
### Effect size and significance

Model	Number Studies	Point estimate	Significance	Standard error
Random effects	17.000	(0.490)	***	0.052

### Heterogeneity

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
28.011	16.000	0.032	42.880	0.016	0.014	0.000	0.127

### Funnel Plot of Standard Error by Point estimate

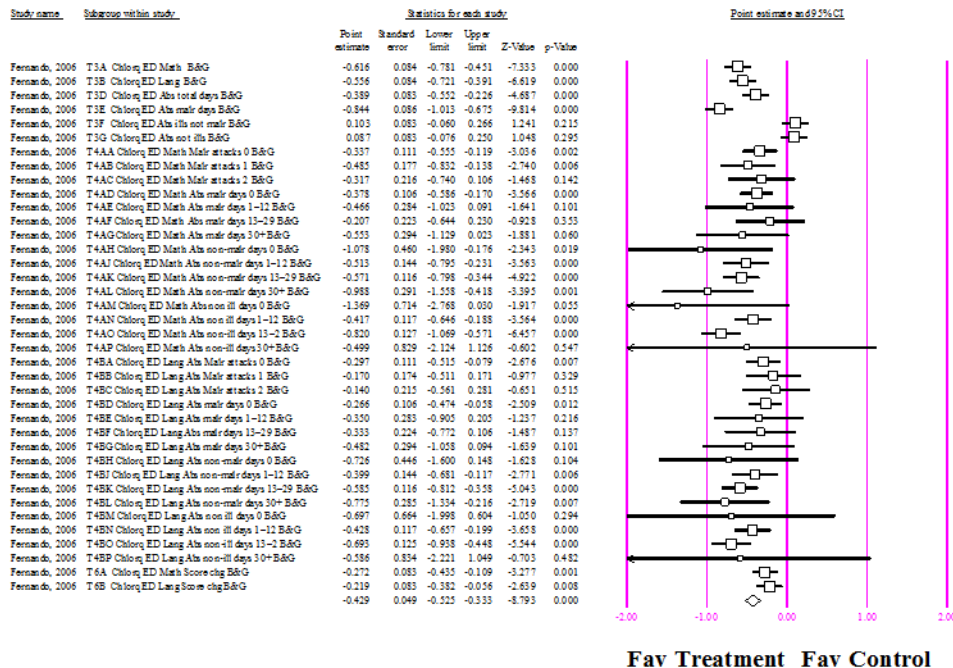


## Egger's regression intercept

Intercept	-0.79395
Standard error	0.65824
95% lower limit (2-tailed)	-2.19696
95% upper limit (2-tailed)	0.60906
t-value	1.20617
df	15.00000
P-value (1-tailed)	0.12322
P-value (2-tailed)	0.24643

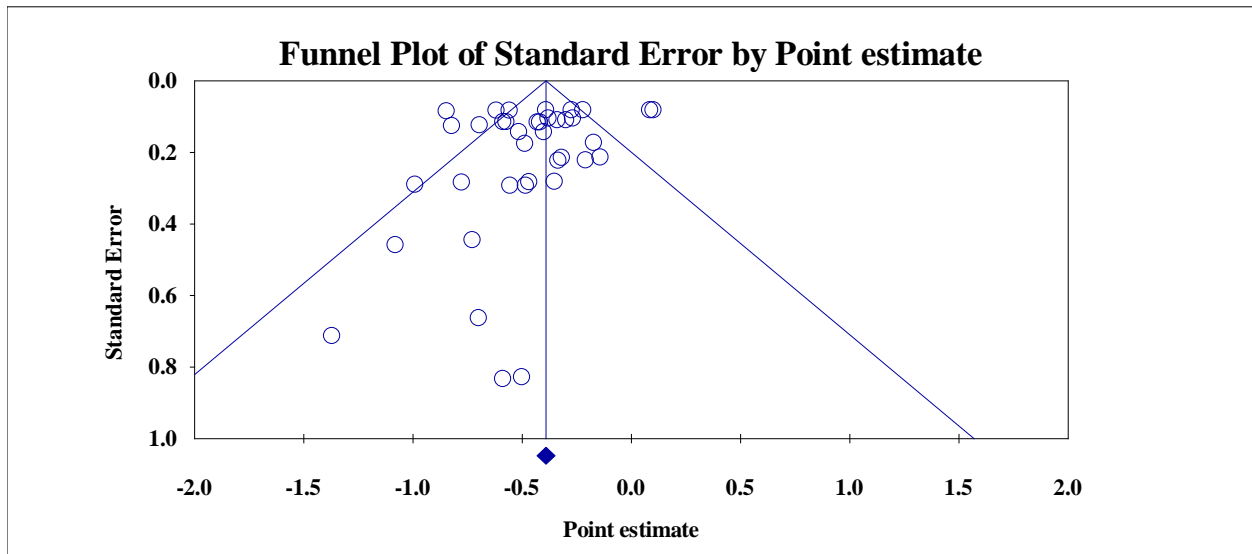
## Pooled Effect Sizes of Chloroquine Interventions on All Educational Outcomes All Children

### Effect of Malaria treatments on Educational Outcomes (boys and girls)



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	38.000	(0.429)	***	0.049

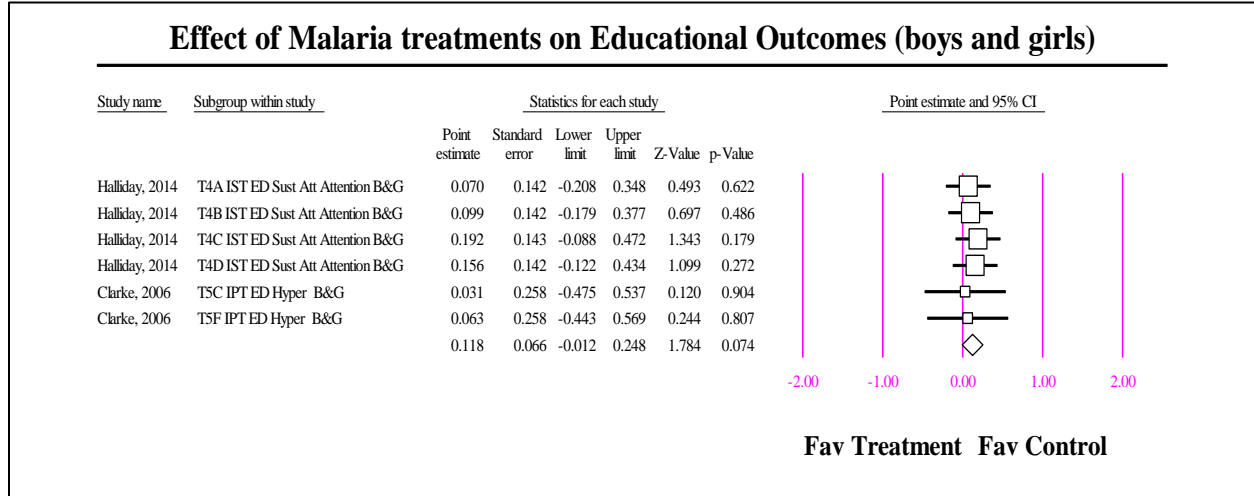
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
154.995	37.000	0.000	76.128	0.055	0.020	0.000	0.235



**Egger's regression intercept**

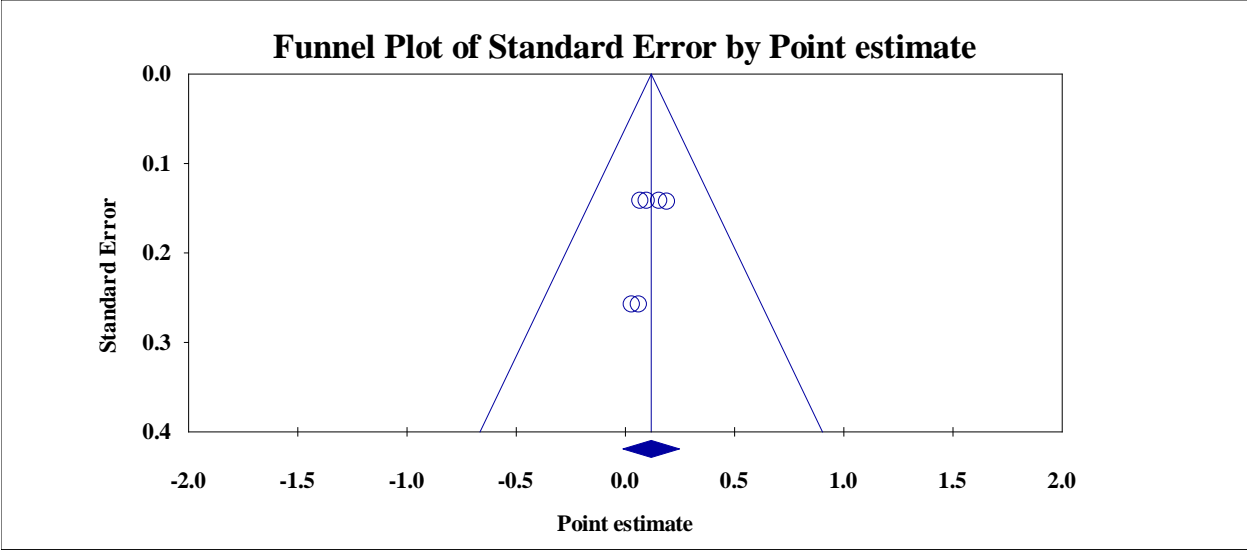
Intercept	-0.93226
Standard error	0.69272
95% lower limit (2-tailed)	-2.33715
95% upper limit (2-tailed)	0.47263
t-value	1.34581
df	36.00000
P-value (1-tailed)	0.09339
P-value (2-tailed)	0.18678

## Pooled Effect Sizes of IPT/IST Interventions on Attention All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	6.000	0.118	*	0.066

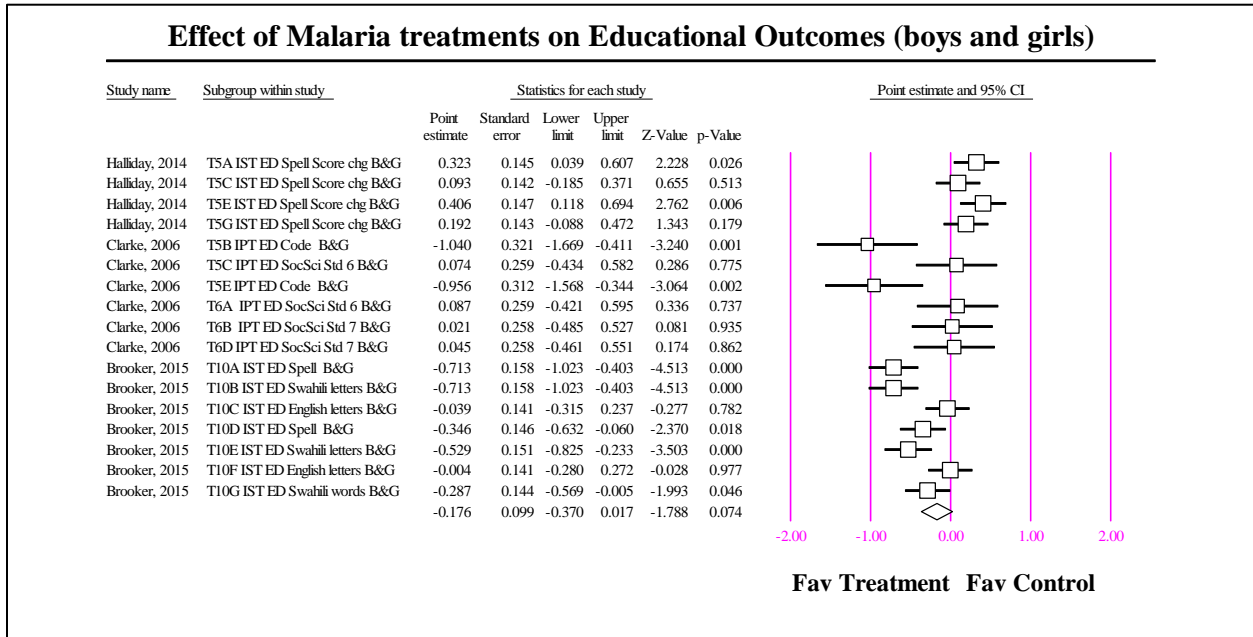
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
0.631	5.000	0.987	-	-	0.017	0.000	-



### Egger's regression intercept

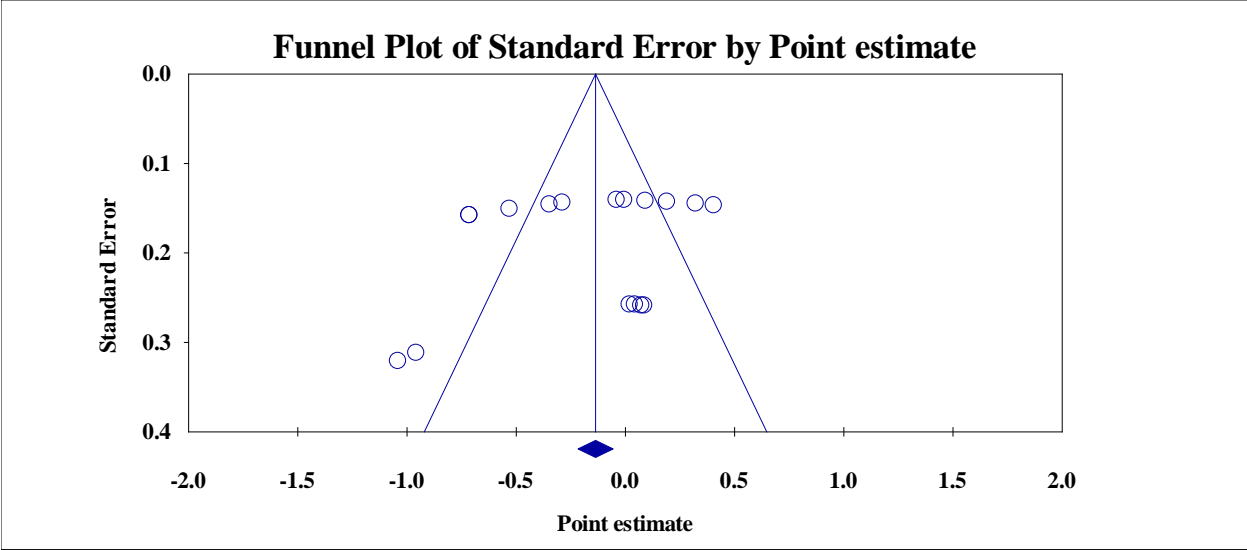
Intercept	-0.69979
Standard error	0.57335
95% lower limit (2-tailed)	-2.29165
95% upper limit (2-tailed)	0.89208
t-value	1.22053
df	4.00000
P-value (1-tailed)	0.14465
P-value (2-tailed)	0.28929

## Pooled Effect Sizes of IPT/IST Interventions on Language All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	17.000	(0.176)	*	0.099

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
86.541	16.000	0.000	81.512	0.128	0.060	0.004	0.358

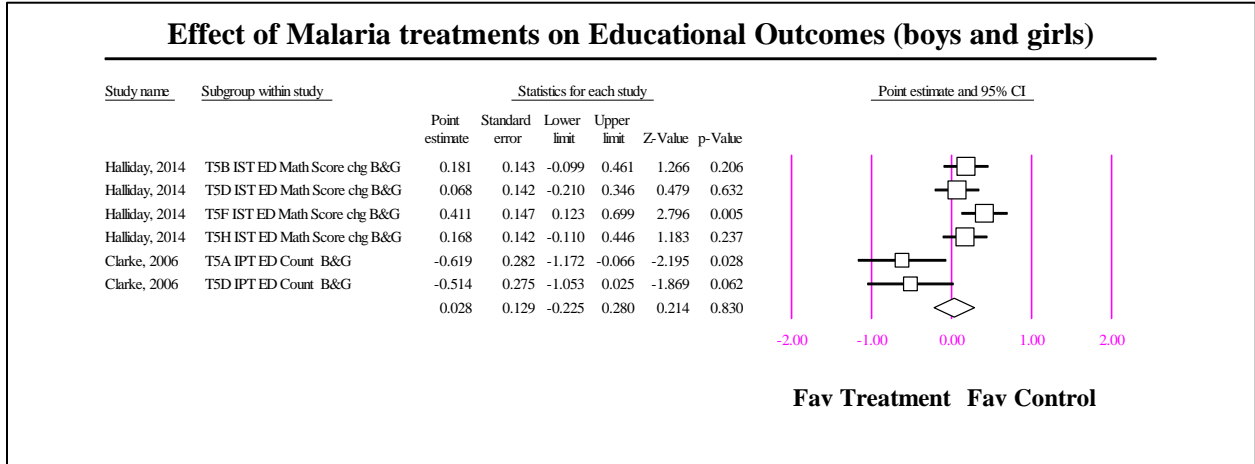


### Egger's regression intercept

Intercept	-2.04828
Standard error	2.15400
95% lower limit (2-tailed)	-6.63942
95% upper limit (2-tailed)	2.54286
t-value	0.95092
df	15.00000
P-value (1-tailed)	0.17836
P-value (2-tailed)	0.35672

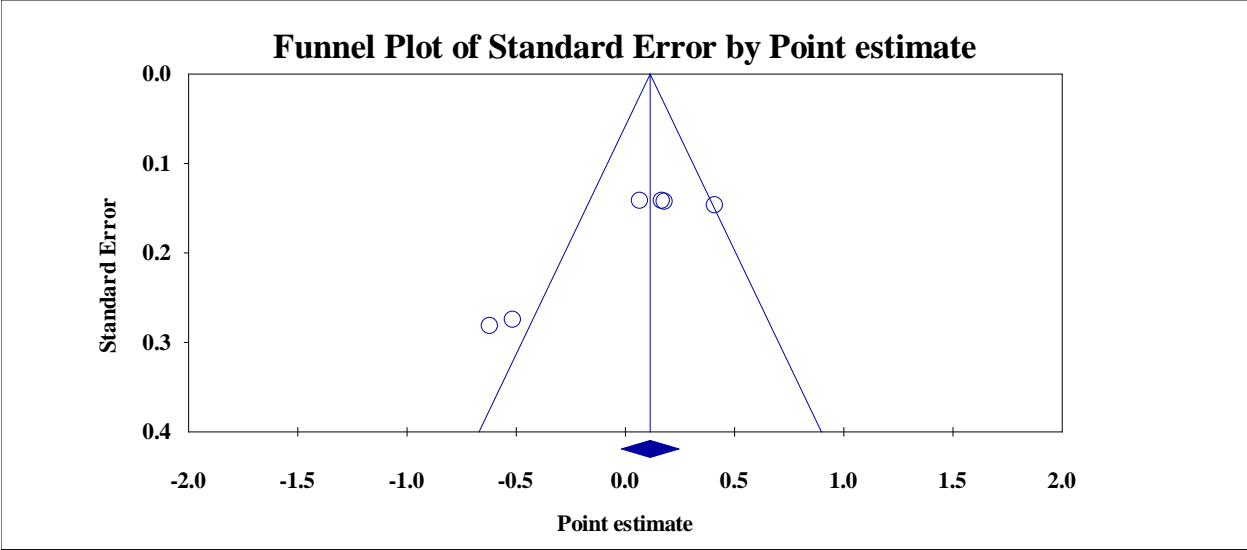


## Pooled Effect Sizes of IPT/IST Interventions on Math All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	6.000	0.028	-	0.129

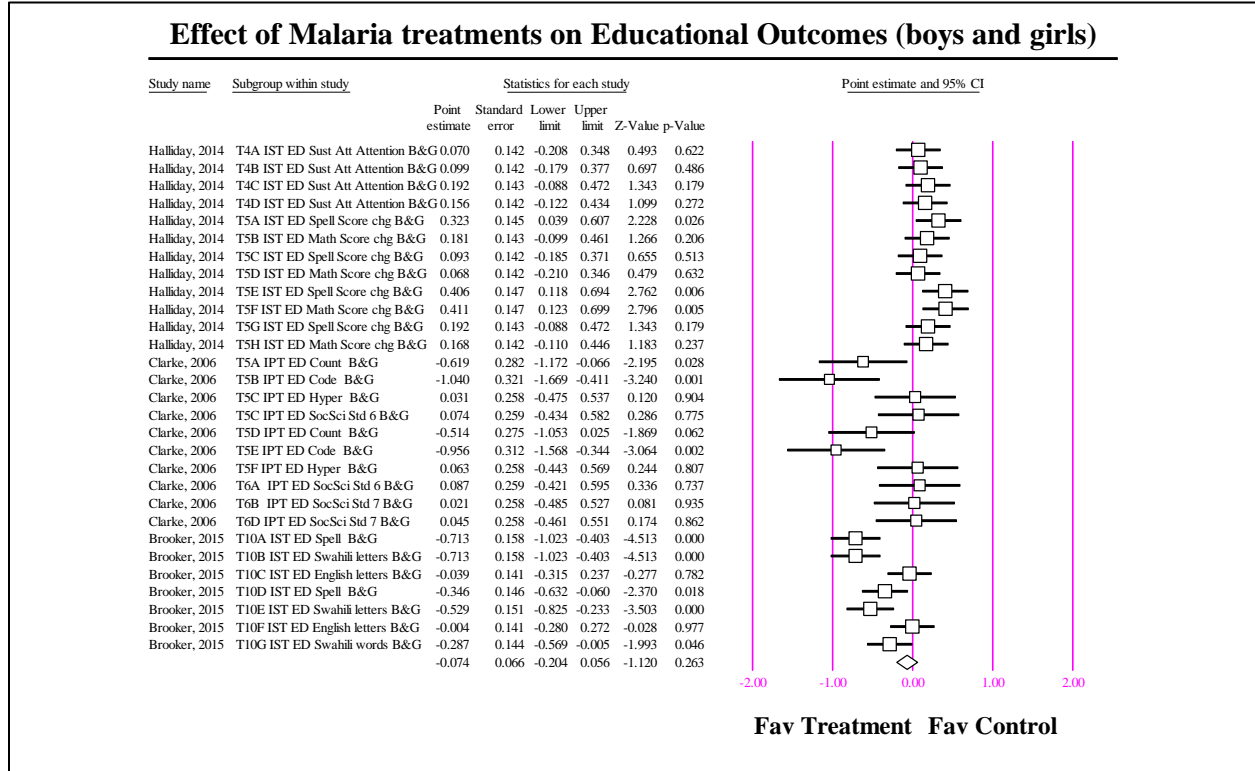
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
16.522	5.000	0.006	69.738	0.066	0.063	0.004	0.256



### Egger's regression intercept

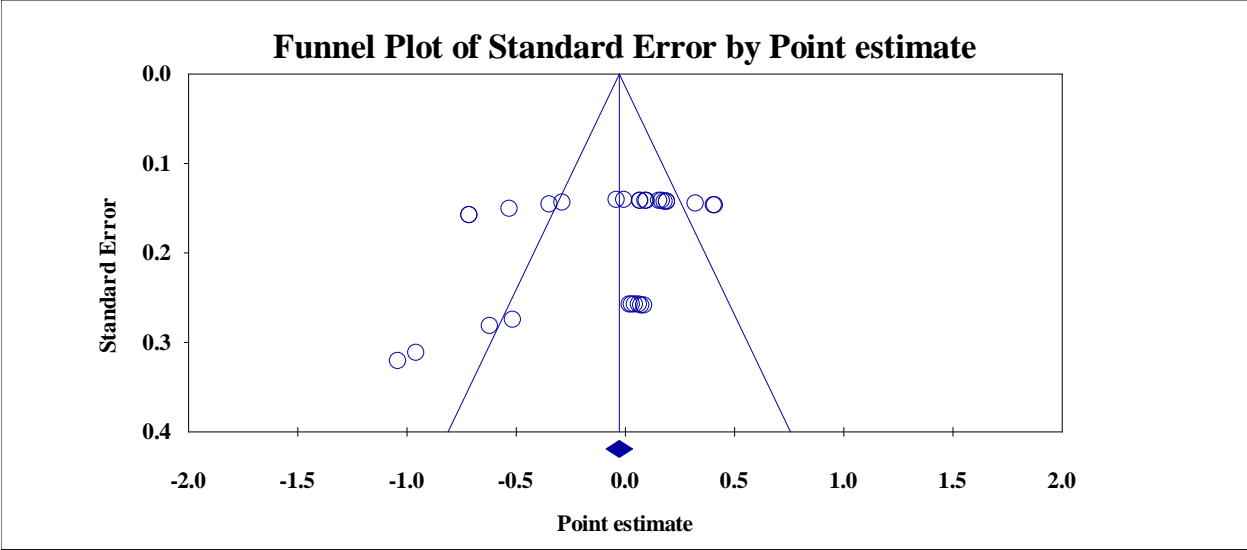
Intercept	-5.57942
Standard error	1.46514
95% lower limit (2-tailed)	-9.64730
95% upper limit (2-tailed)	-1.51154
t-value	3.80811
df	4.00000
P-value (1-tailed)	0.00949
P-value (2-tailed)	0.01897

## Pooled Effect Sizes of IPT/IST Interventions on All Educational Outcomes All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	29.000	(0.074)	-	0.066

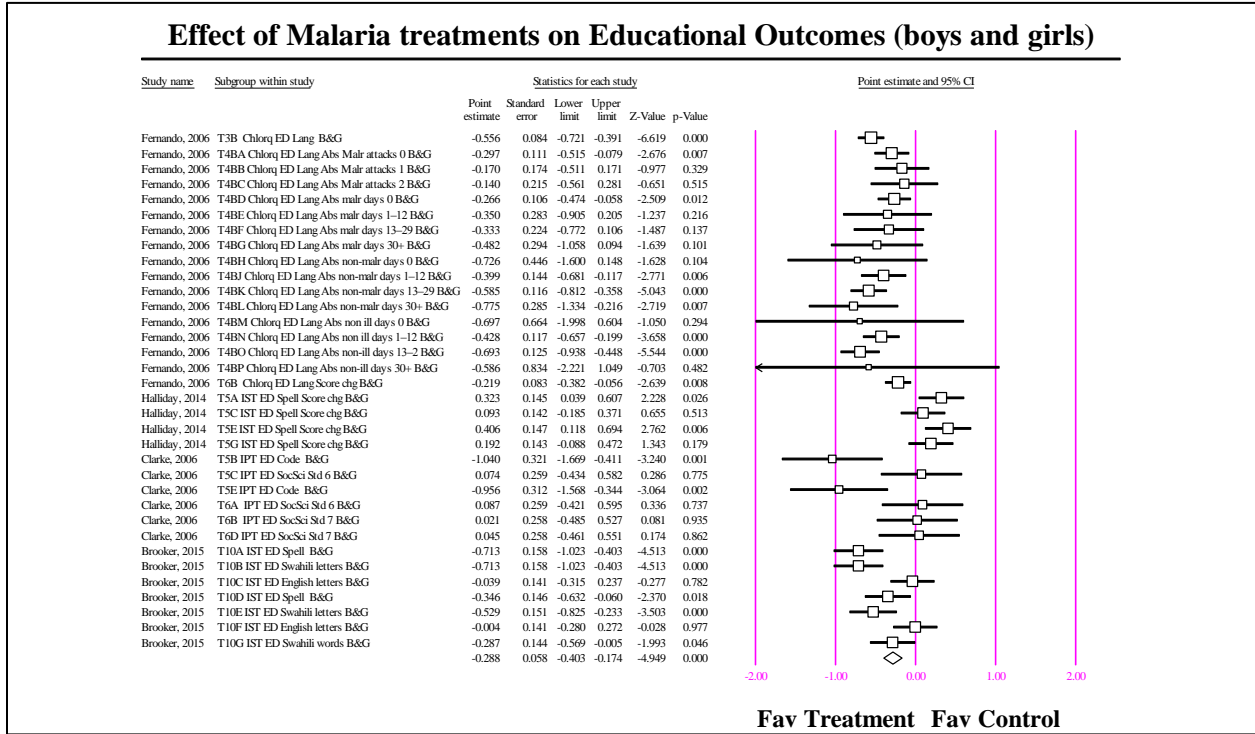
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
119.969	28.000	0.000	76.661	0.092	0.034	0.001	0.304



### Egger's regression intercept

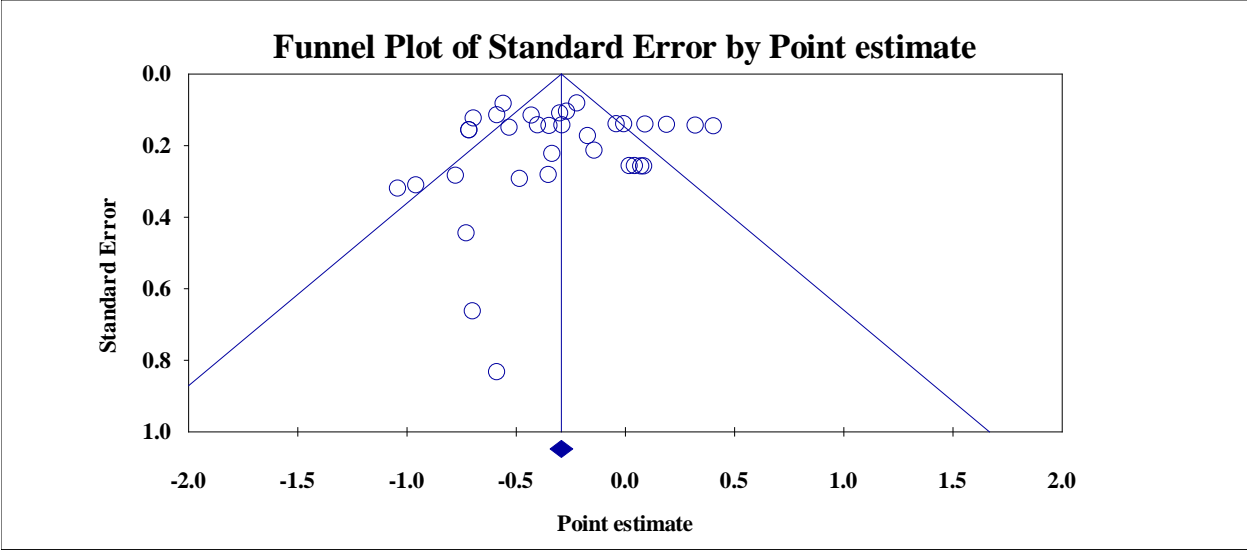
Intercept	-2.72467
Standard error	1.41741
95% lower limit (2-tailed)	-5.63296
95% upper limit (2-tailed)	0.18362
t-value	1.92228
df	27.00000
P-value (1-tailed)	0.03259
P-value (2-tailed)	0.06518

## Pooled Effect Sizes of All Malaria Interventions on Language All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	34.000	(0.288)	***	0.058

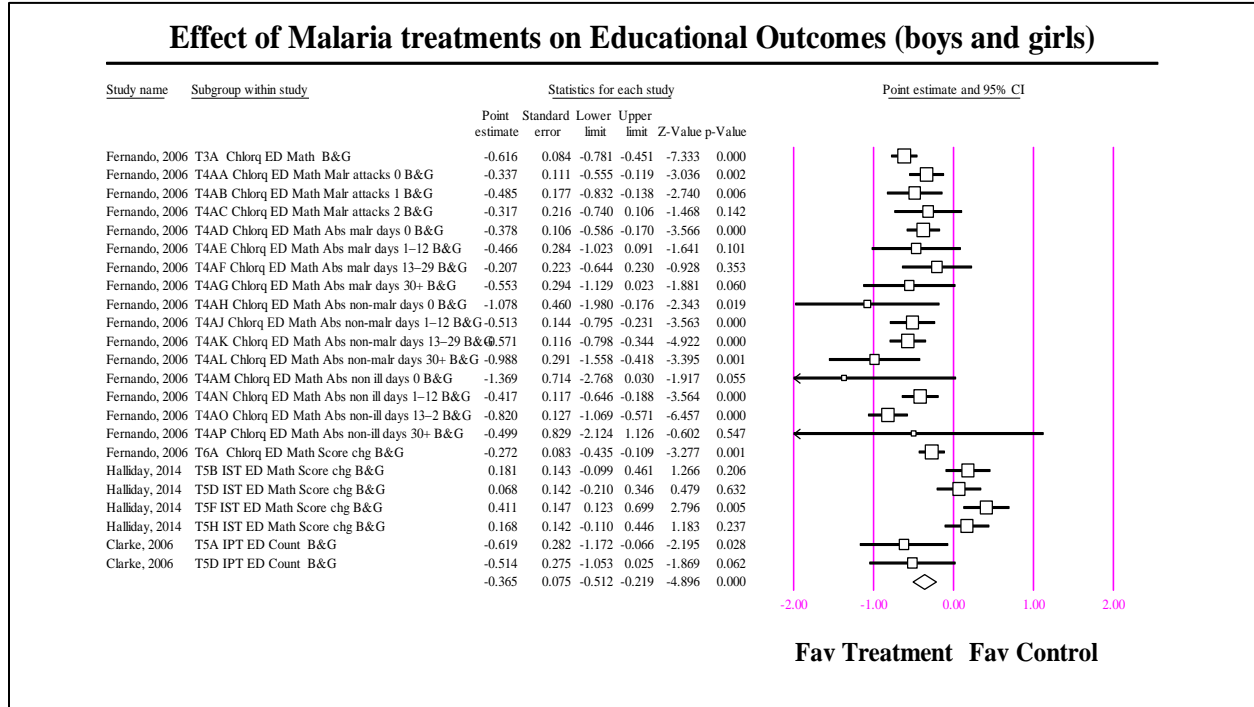
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
135.897	33.000	0.000	75.717	0.076	0.029	0.001	0.275



### Egger's regression intercept

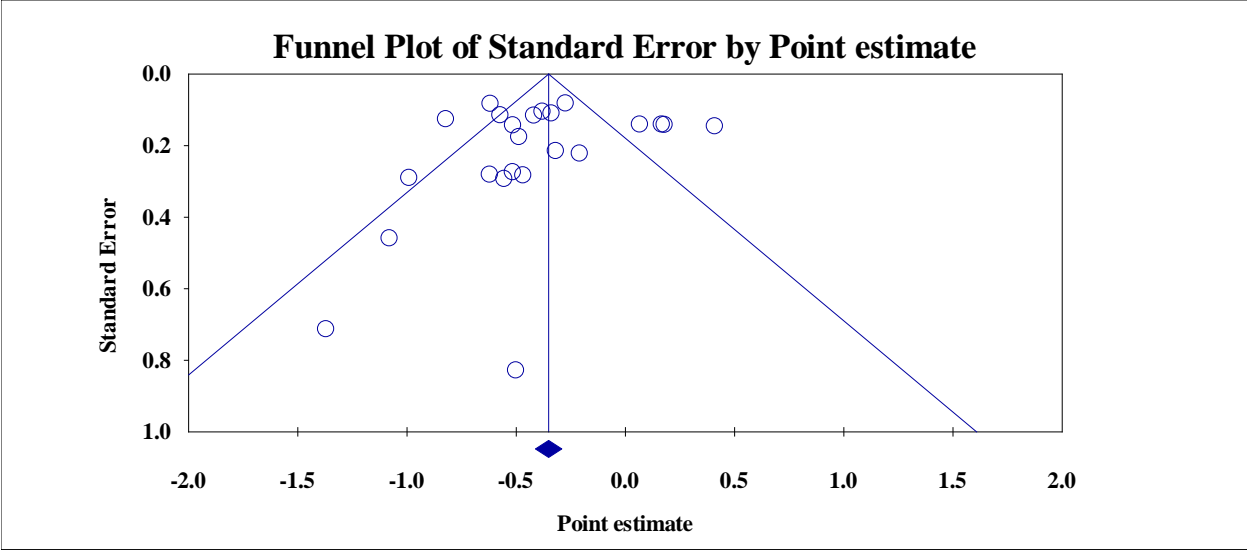
Intercept	-0.02348
Standard error	0.87416
95% lower limit (2-tailed)	-1.80408
95% upper limit (2-tailed)	1.75712
t-value	0.02686
df	32.00000
P-value (1-tailed)	0.48937
P-value (2-tailed)	0.97874

## Pooled Effect Sizes of All Malaria Interventions on Math All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	23.000	(0.365)	***	0.075

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
104.793	22.000	0.000	79.006	0.085	0.039	0.002	0.292



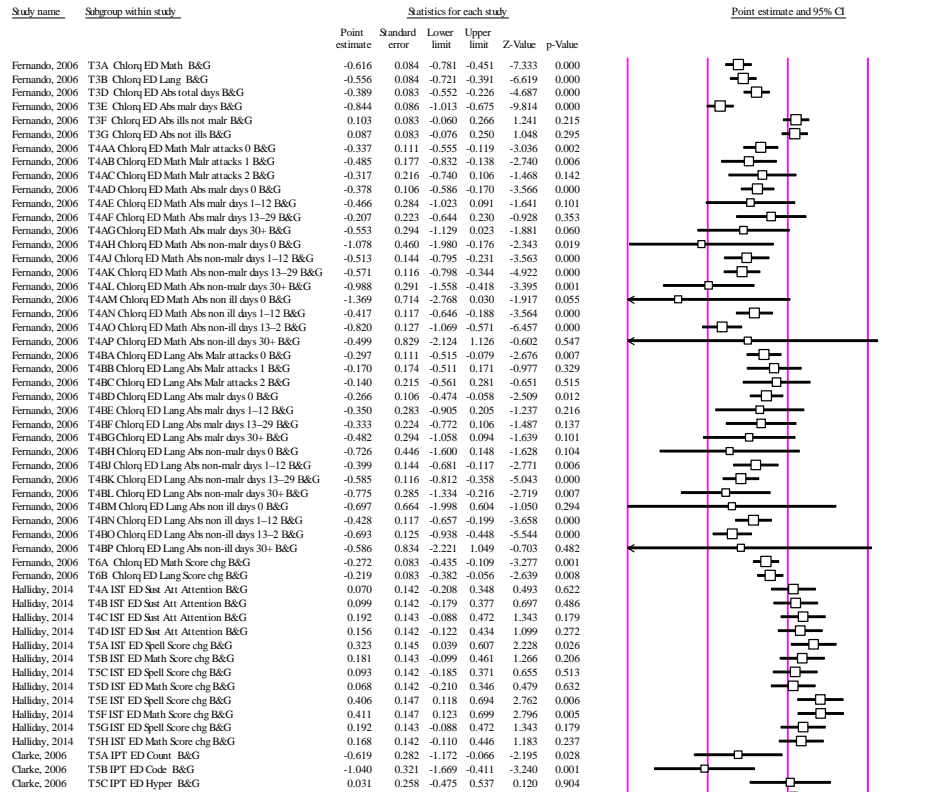
### Egger's regression intercept

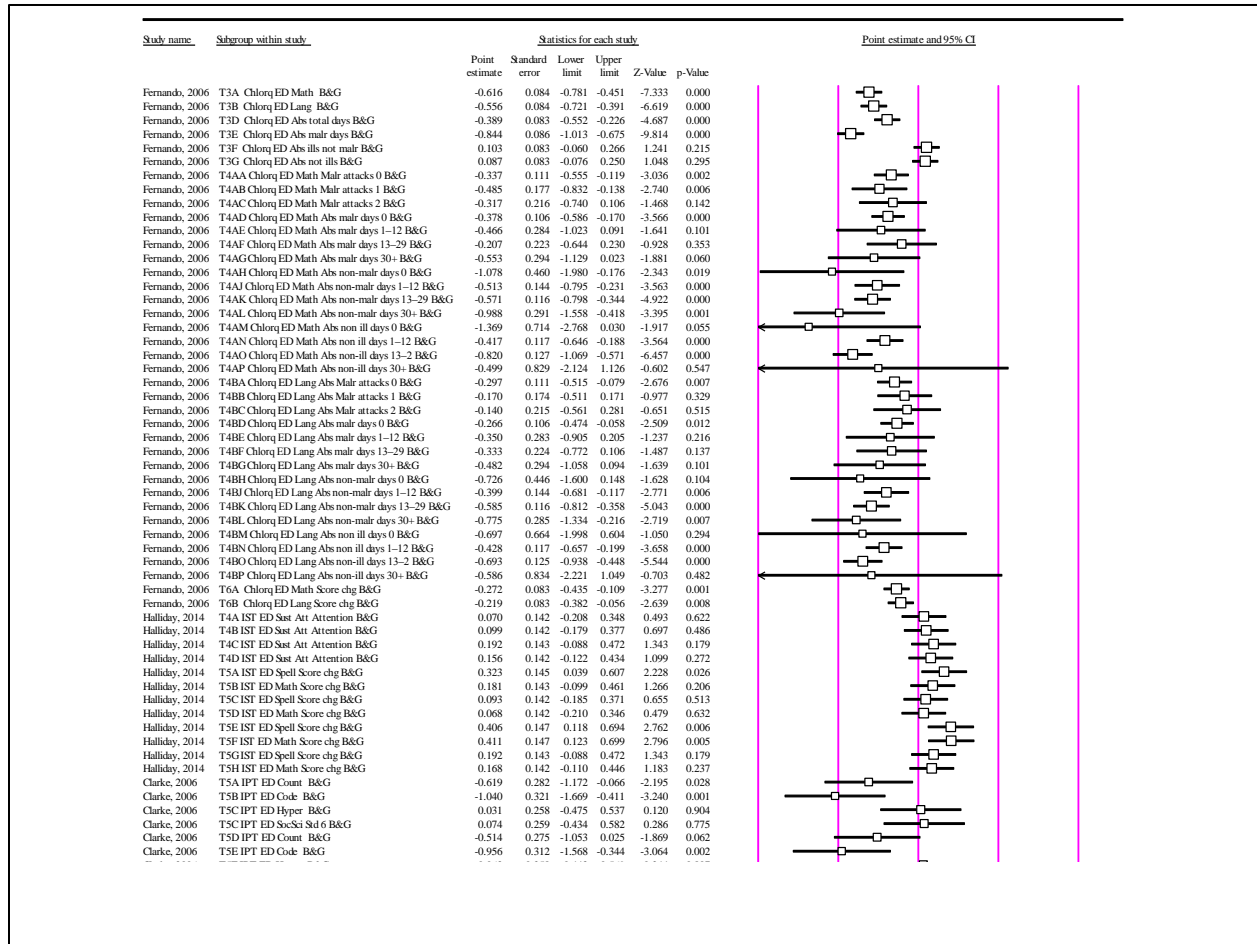
Intercept	-0.43265
Standard error	1.04355
95% lower limit (2-tailed)	-2.60283
95% upper limit (2-tailed)	1.73753
t-value	0.41459
df	21.00000
P-value (1-tailed)	0.34132
P-value (2-tailed)	0.68264



# Pooled Effect Sizes of All Malaria Interventions on All Educational Outcomes All Children

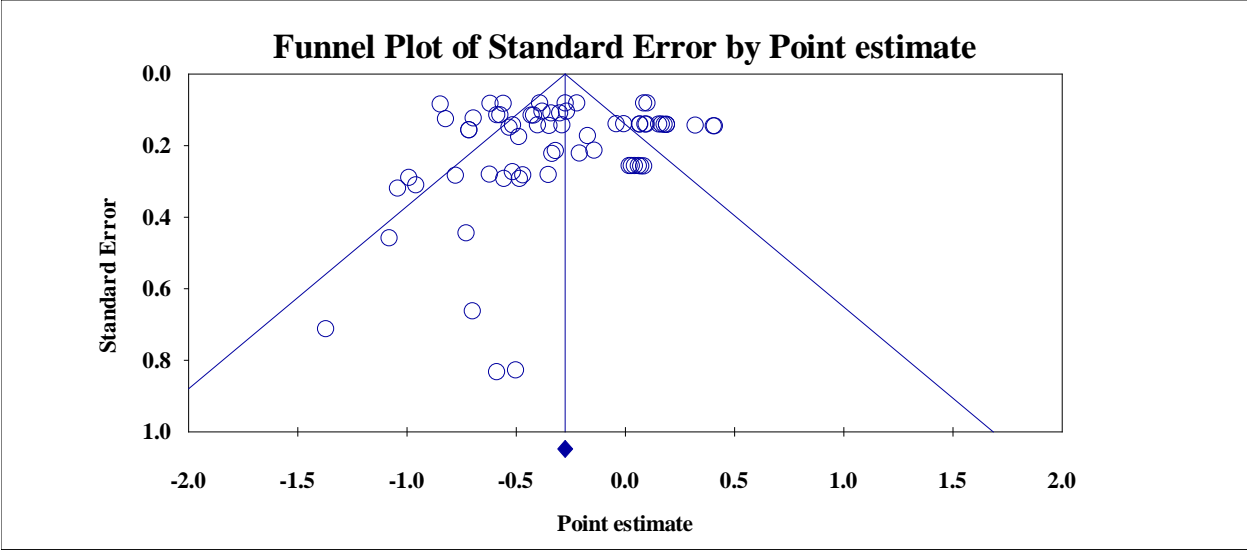
## Effect of Malaria treatments on Educational Outcomes (boys and girls)





Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	67.000	(0.276)	***	0.045

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
368.604	66.000	-	82.095	0.095	0.024	0.001	0.308

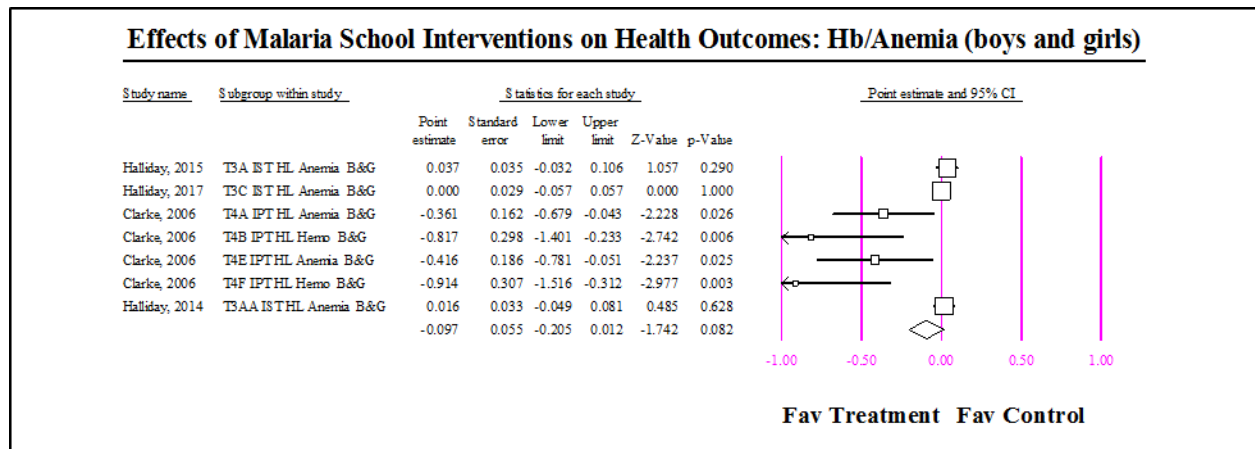


### Egger's regression intercept

Intercept	-0.10367
Standard error	0.68645
95% lower limit (2-tailed)	-1.47462
95% upper limit (2-tailed)	1.26727
t-value	0.15103
df	65.00000
P-value (1-tailed)	0.44021
P-value (2-tailed)	0.88042

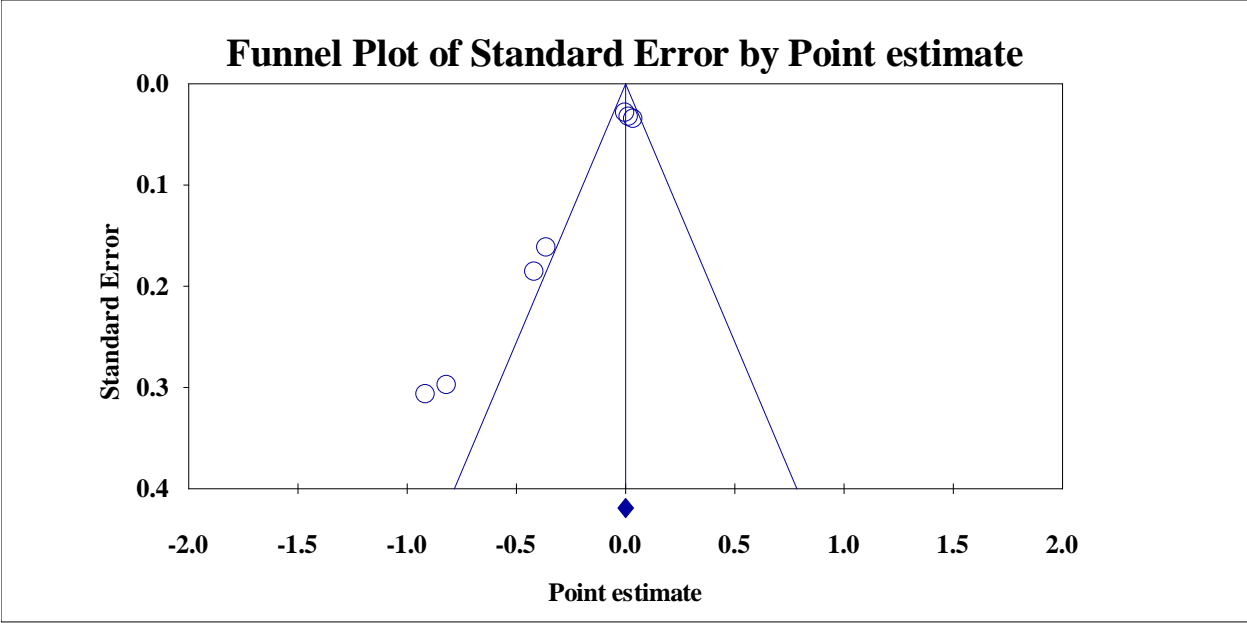
## Annex 4.2: Malaria Interventions, Health Outcomes

### Pooled Effect Sizes of IPT/IST Interventions on Anemia/Hemoglobin Status All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	7.000	(0.097)	*	0.055

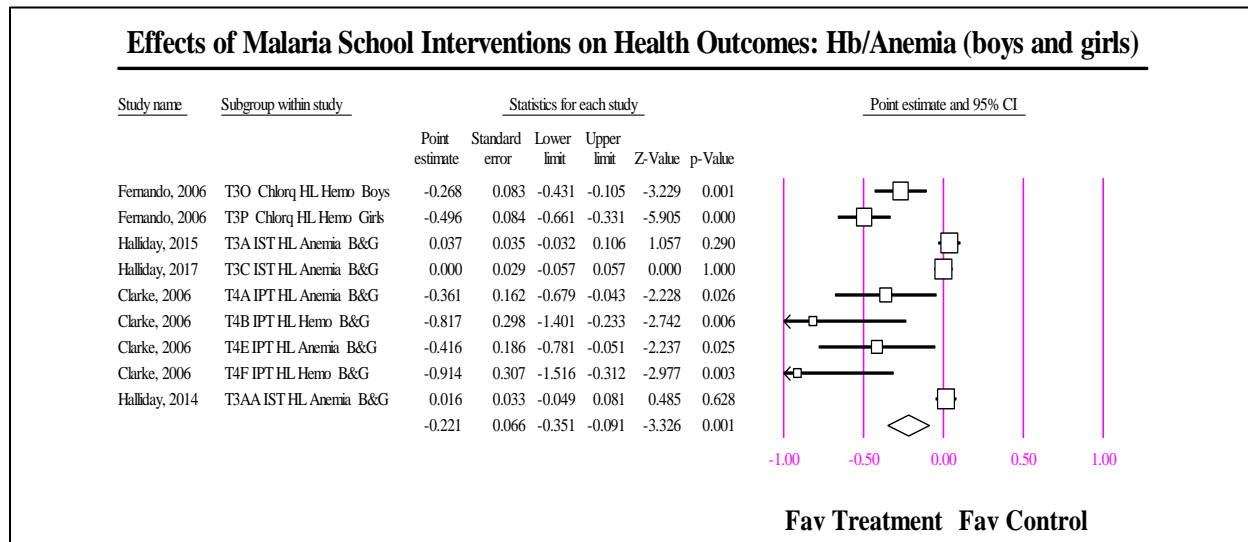
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
27.701	6.000	0.000	78.340	0.011	0.011	0.000	0.103



#### Egger's regression intercept

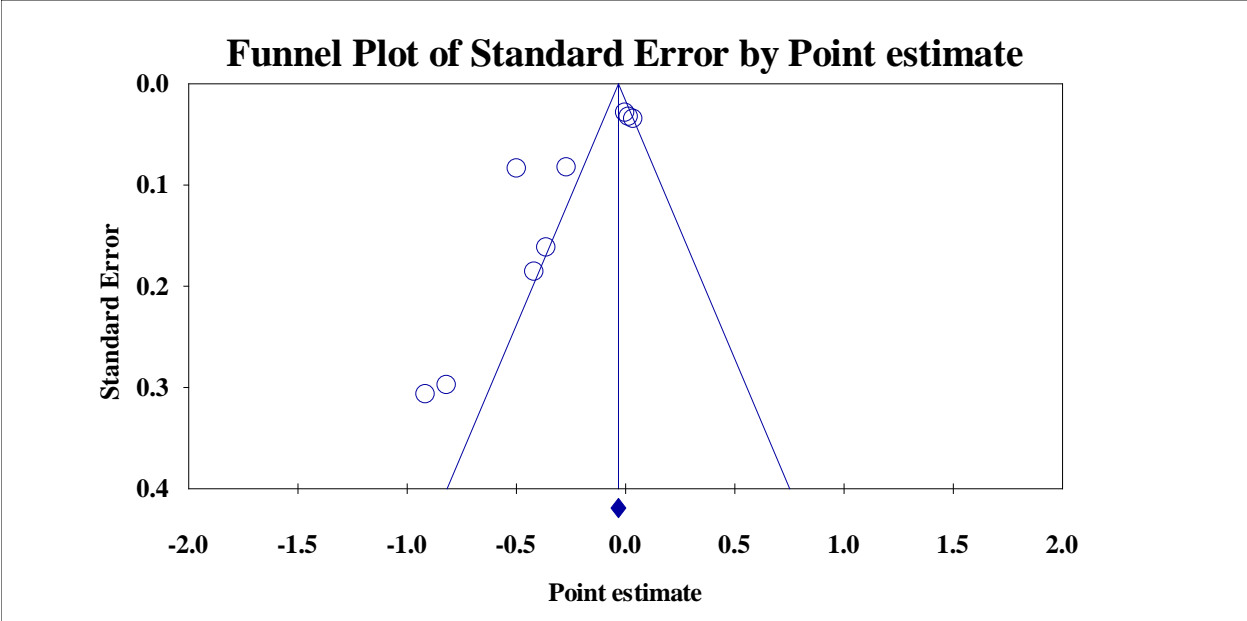
Intercept	-3.01208
Standard error	0.33864
95% lower limit (2-tailed)	-3.88259
95% upper limit (2-tailed)	-2.14157
t-value	8.89455
df	5.00000
P-value (1-tailed)	0.00015
P-value (2-tailed)	0.00030

## Pooled Effect Sizes of Chloroquine Interventions on Anemia/Hemoglobin Status All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	9.000	(0.221)	***	0.066

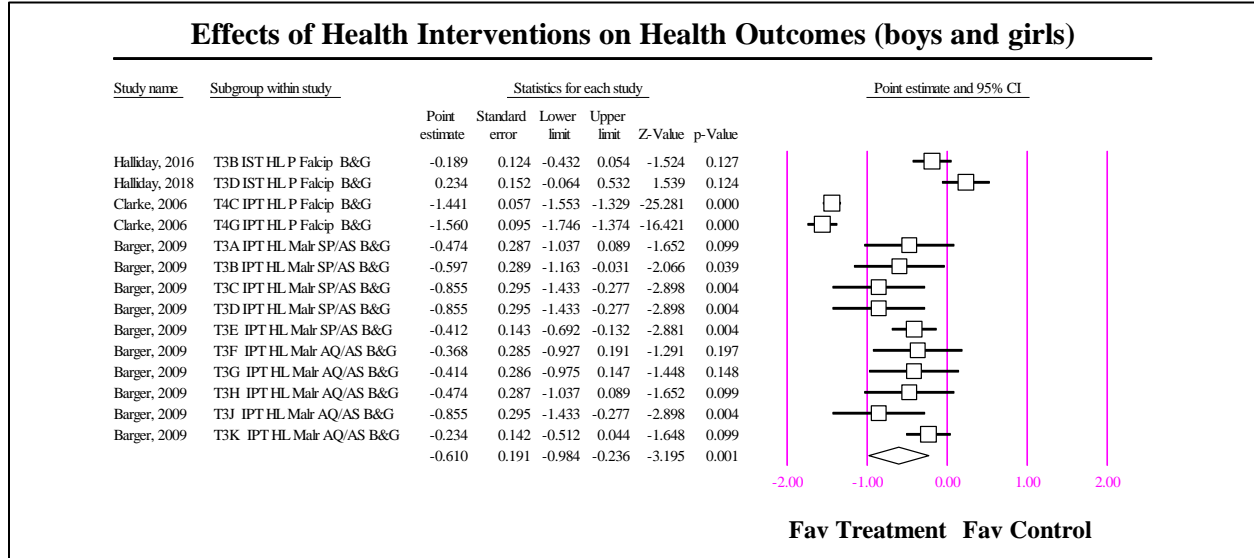
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
69.393	8.000	0.000	88.472	0.026	0.022	0.000	0.160



### Egger's regression intercept

Intercept	-3.86924
Standard error	0.86318
95% lower limit (2-tailed)	-5.91035
95% upper limit (2-tailed)	-1.82814
t-value	4.48253
df	7.00000
P-value (1-tailed)	0.00143
P-value (2-tailed)	0.00286

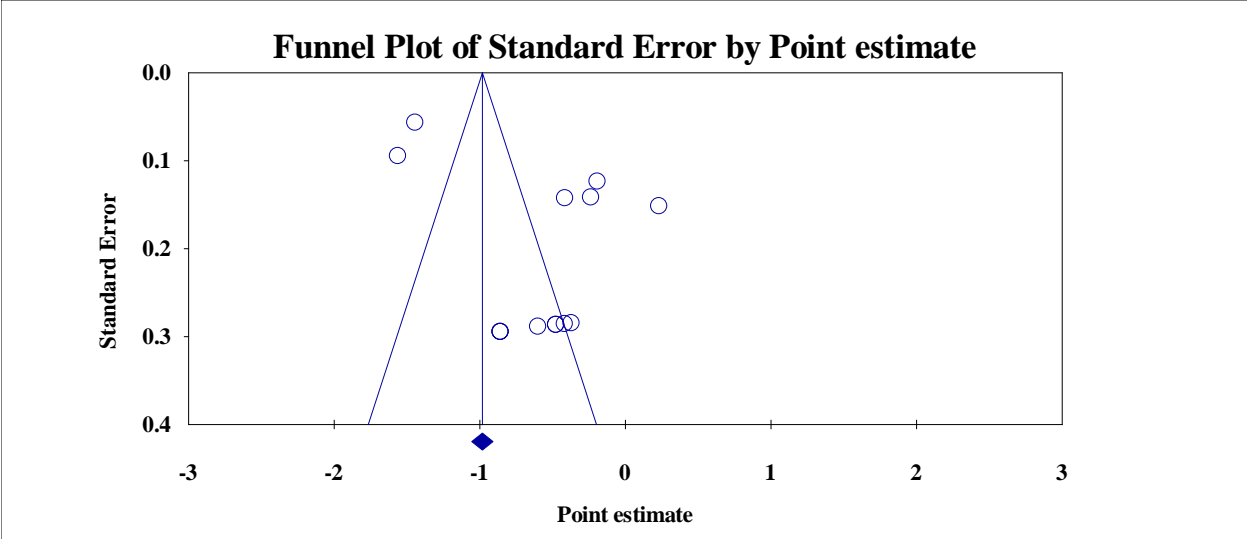
## Pooled Effect Sizes of IPT/ISP Interventions on Malaria All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	14.000	(0.610)	***	0.191

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
267.579	13.000	-	95.142	0.458	0.292	0.085	0.677

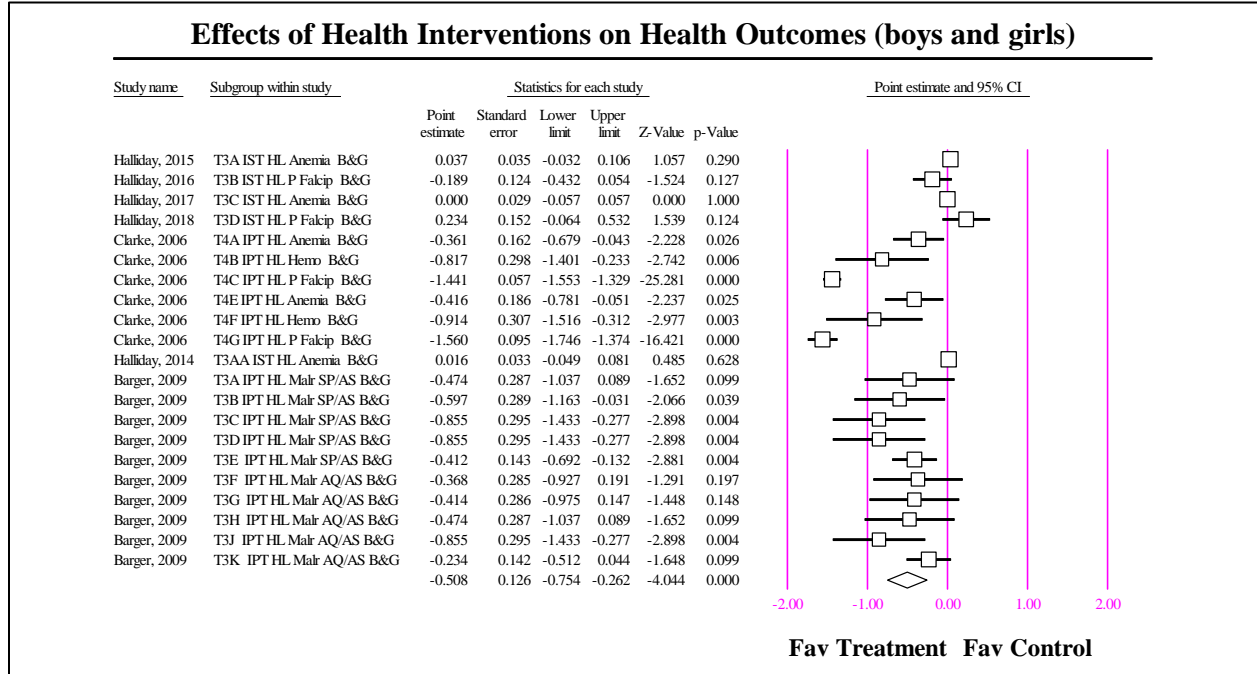




### Egger's regression intercept

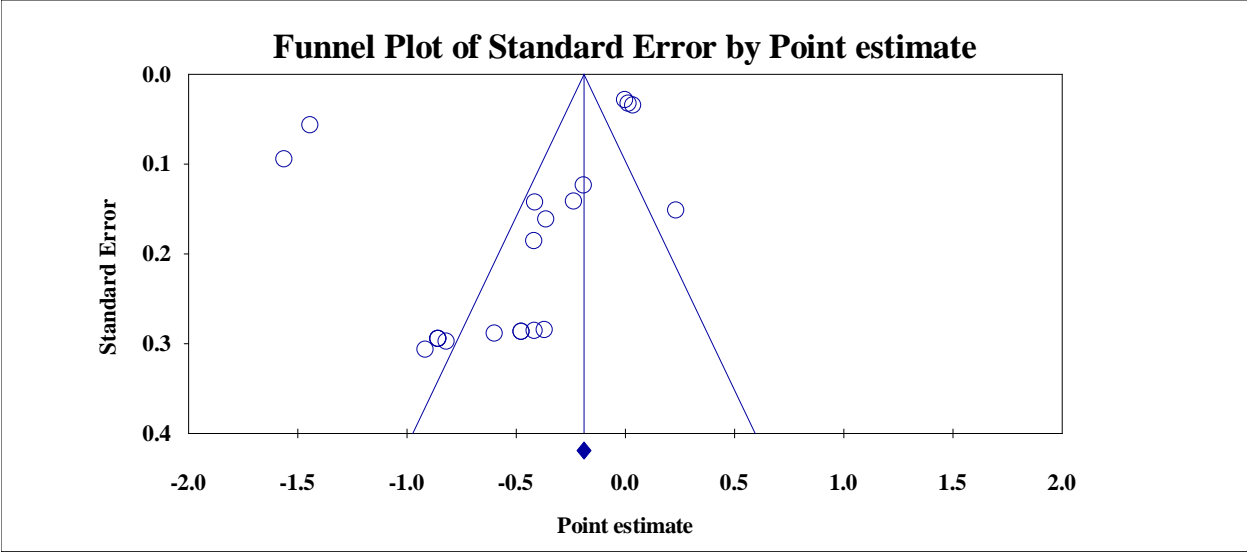
Intercept	4.81088
Standard error	1.86404
95% lower limit (2-tailed)	0.74949
95% upper limit (2-tailed)	8.87227
t-value	2.58089
df	12.00000
P-value (1-tailed)	0.01203
P-value (2-tailed)	0.02406

## Pooled Effect Sizes of IPT/ISP Interventions on Anemia/Hemoglobin Status All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	21.000	(0.508)	***	0.126

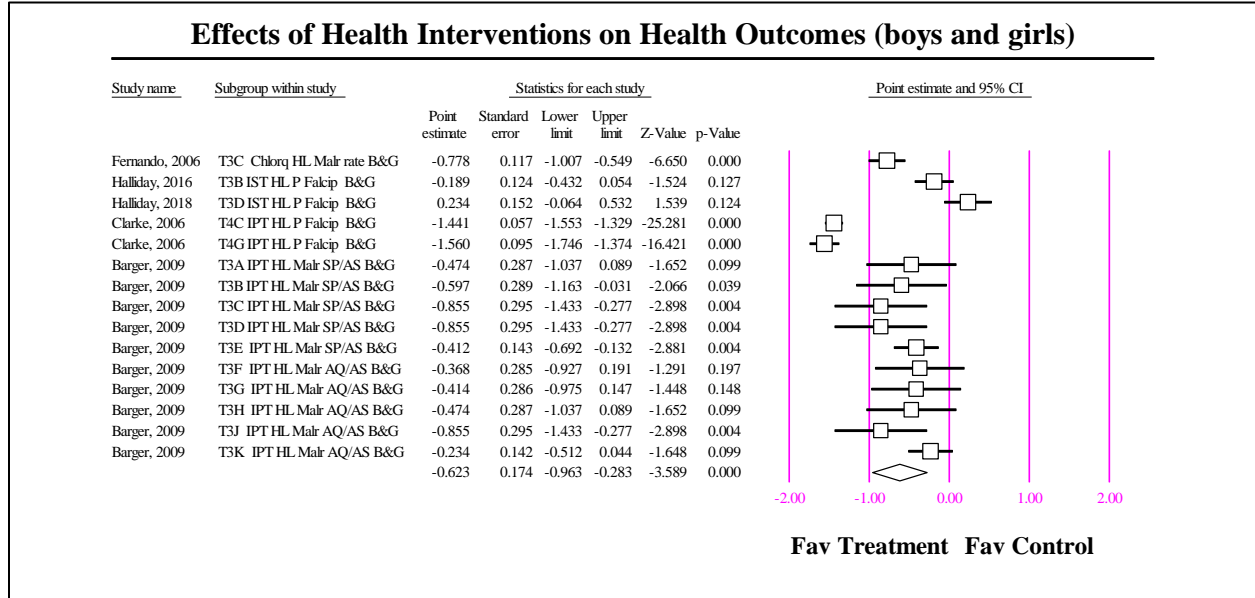
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
856.667	20.000	-	97.665	0.287	0.202	0.041	0.536



### Egger's regression intercept

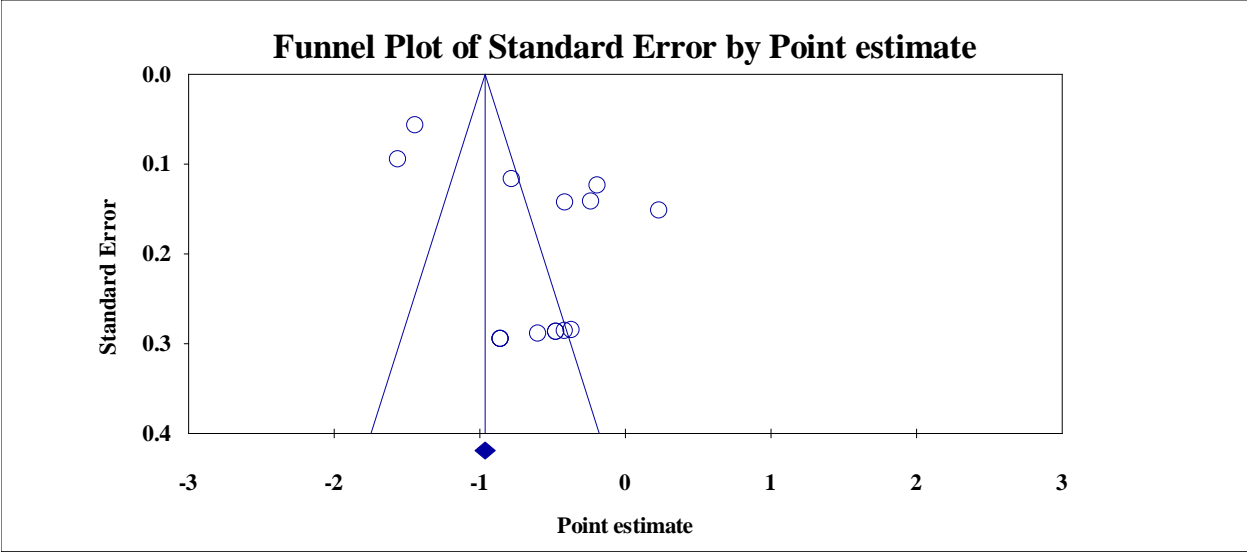
Intercept	-3.21956
Standard error	1.91290
95% lower limit (2-tailed)	-7.22330
95% upper limit (2-tailed)	0.78418
t-value	1.68308
df	19.00000
P-value (1-tailed)	0.05436
P-value (2-tailed)	0.10872

## Pooled Effect Sizes of All Health Interventions on Malaria All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	15.000	(0.623)	***	0.174

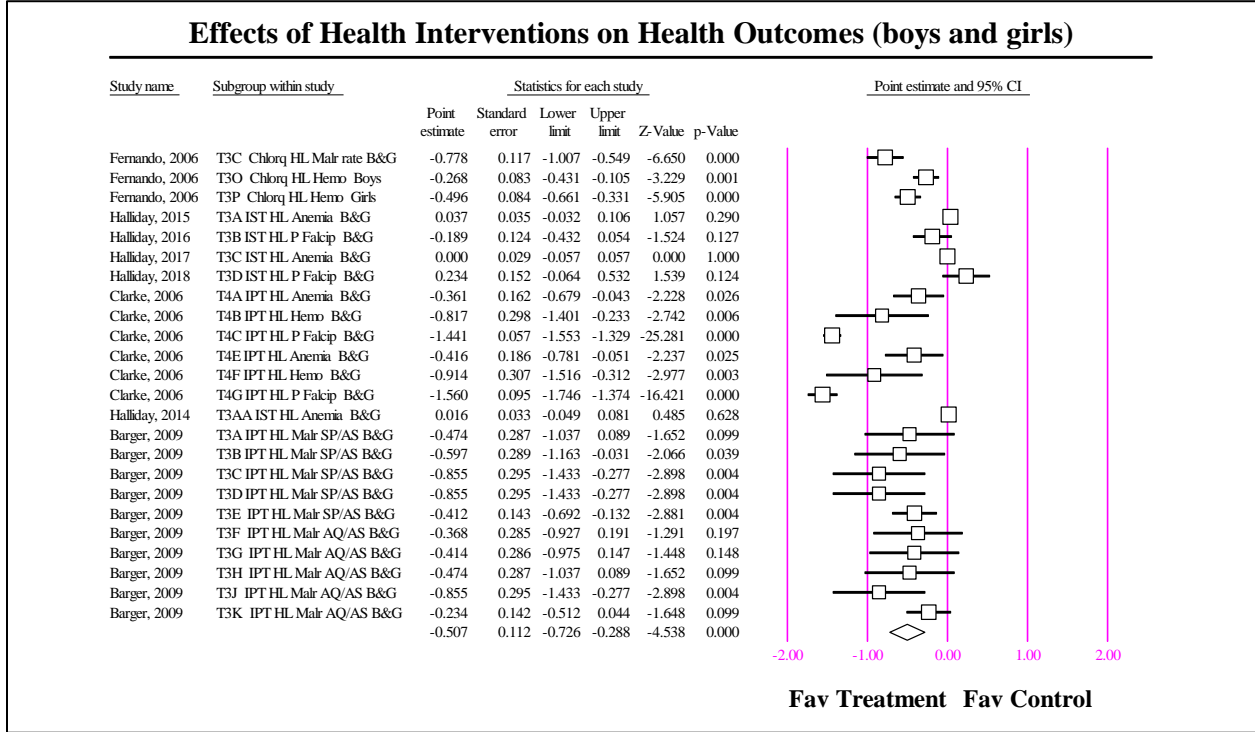
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
270.351	14.000	-	94.822	0.402	0.241	0.058	0.634



### Egger's regression intercept

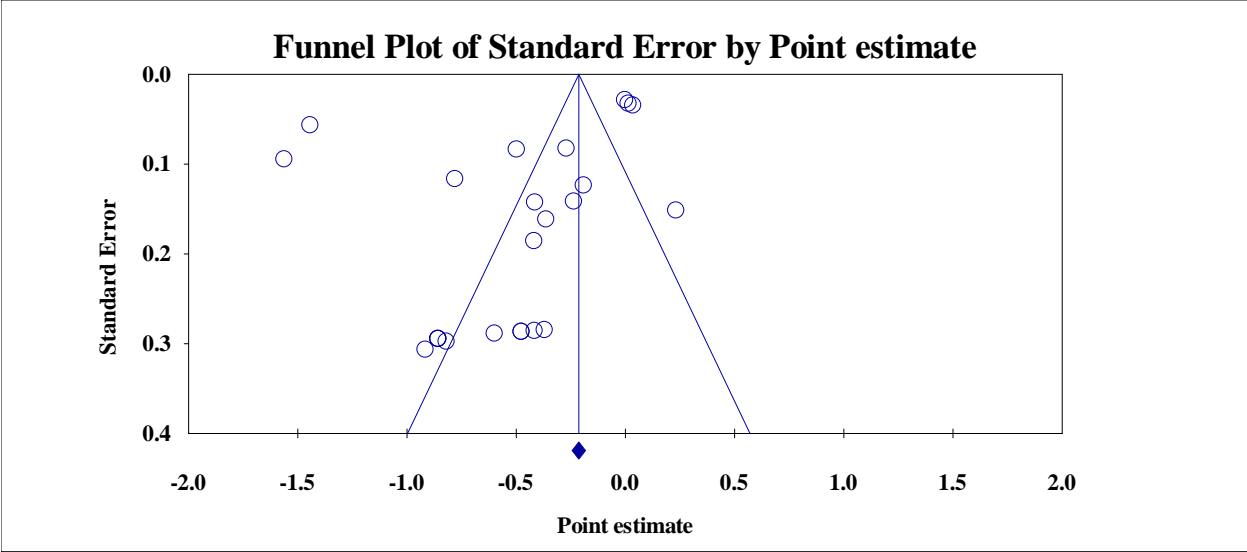
Intercept	4.81072
Standard error	1.80531
95% lower limit (2-tailed)	0.91058
95% upper limit (2-tailed)	8.71086
t-value	2.66476
df	13.00000
P-value (1-tailed)	0.00973
P-value (2-tailed)	0.01946

# Pooled Effect Sizes of All Health Interventions on All Health Outcomes All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	24.000	(0.507)	***	0.112

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
893.821	23.000	-	97.427	0.261	0.164	0.027	0.511

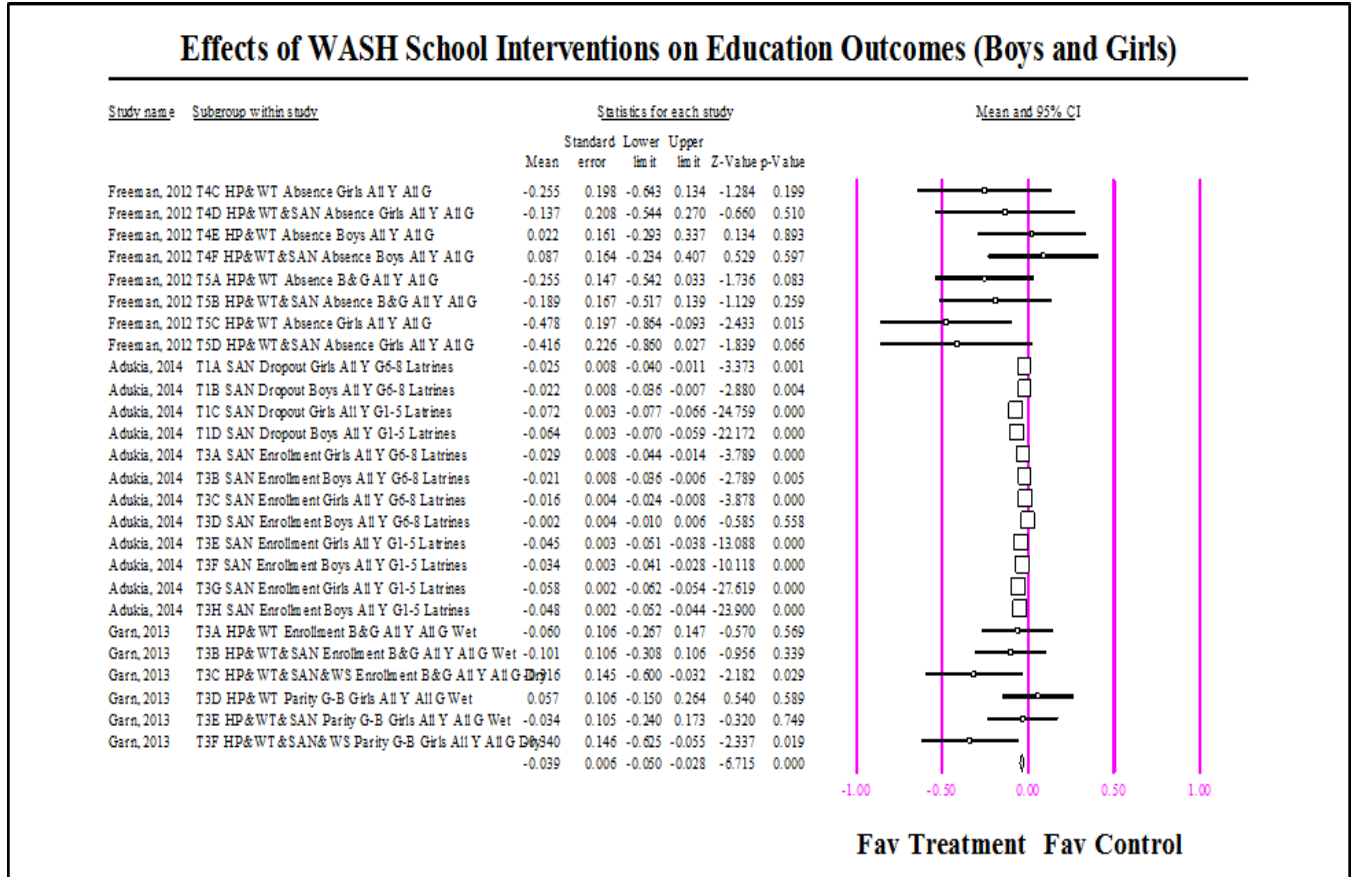


### Egger's regression intercept

Intercept	-3.43650
Standard error	1.75270
95% lower limit (2-tailed)	-7.07137
95% upper limit (2-tailed)	0.19838
t-value	1.96069
df	22.00000
P-value (1-tailed)	0.03135
P-value (2-tailed)	0.06270

## Annex 2.3: WASH Interventions, Educational Outcomes

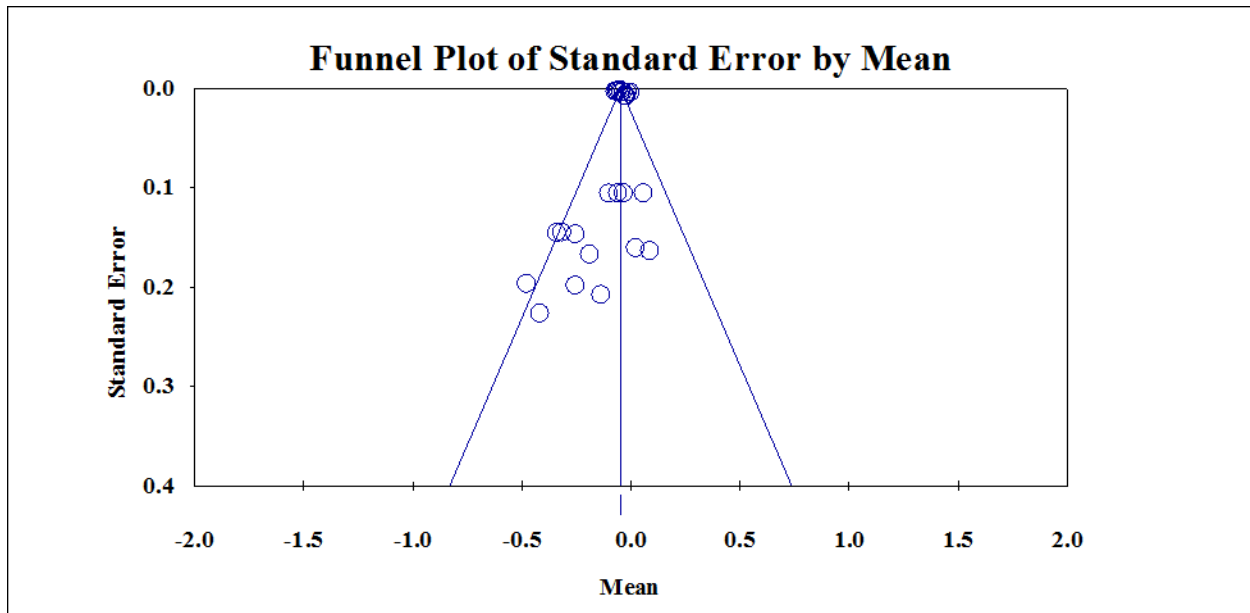
### Pooled Effect Sizes of all WASH Interventions for all Educational Outcomes All children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	26.000	(0.039)	***	0.006

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
384.553	25.000	-	93.499	0.000	0.000	0.000	0.020

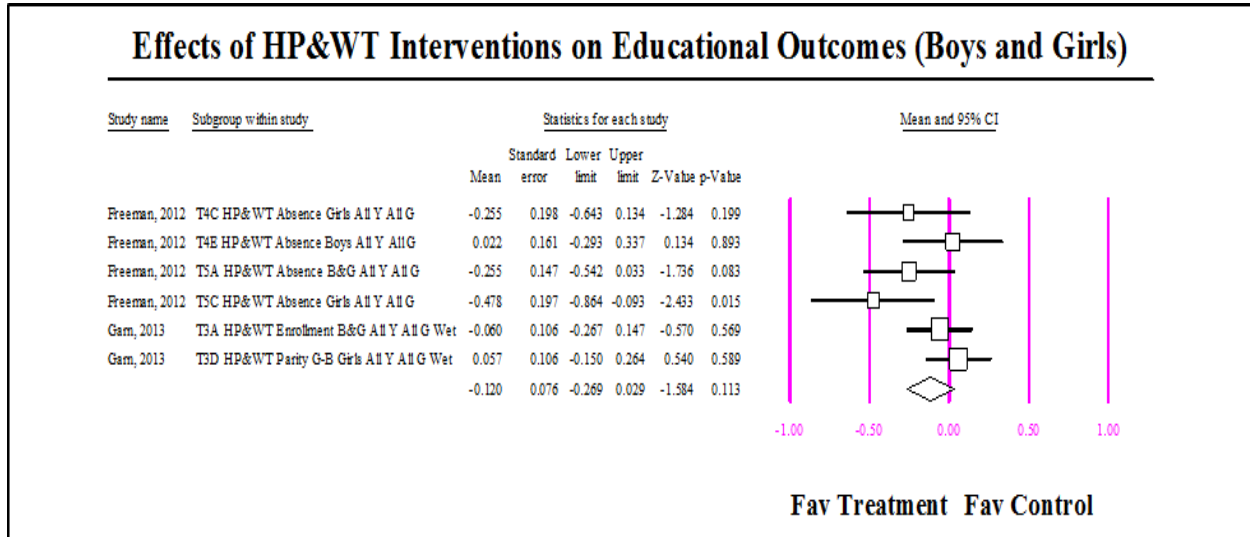




#### Egger's regression intercept

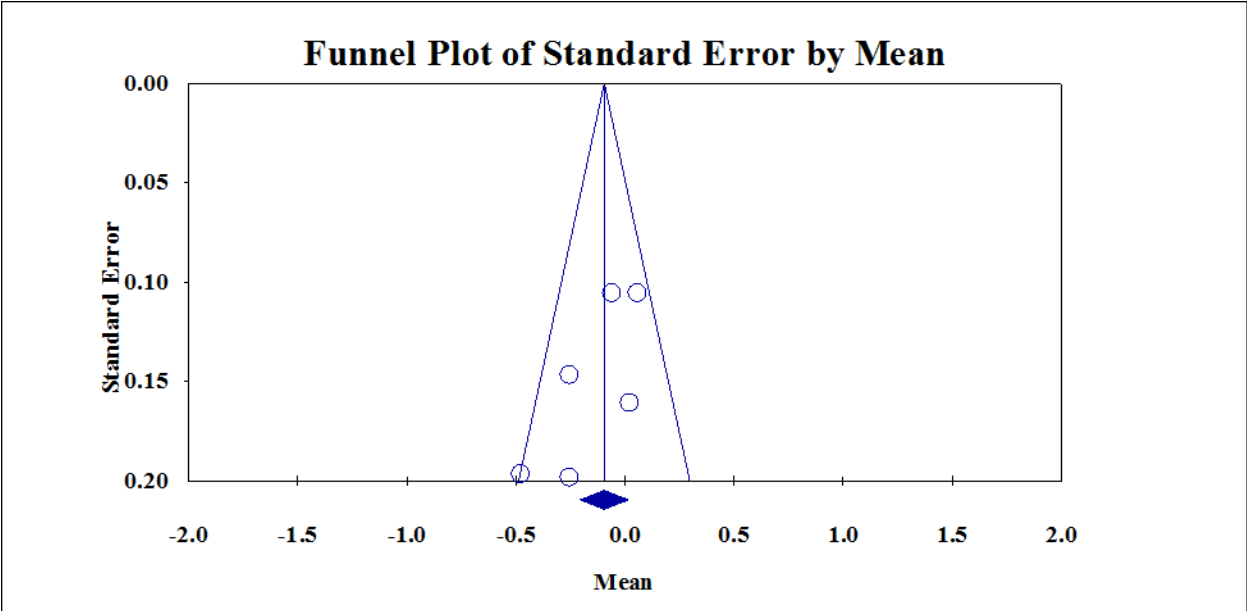
Intercept	0.39945
Standard error	1.01436
95% lower limit (2-tailed)	-1.69409
95% upper limit (2-tailed)	2.49298
t-value	0.39379
df	24.00000
P-value (1-tailed)	0.34861
P-value (2-tailed)	0.69721

**Pooled Effect Sizes of HP and WT Interventions for all Educational Outcomes  
All children**



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	6.000	(0.120)	-	0.076

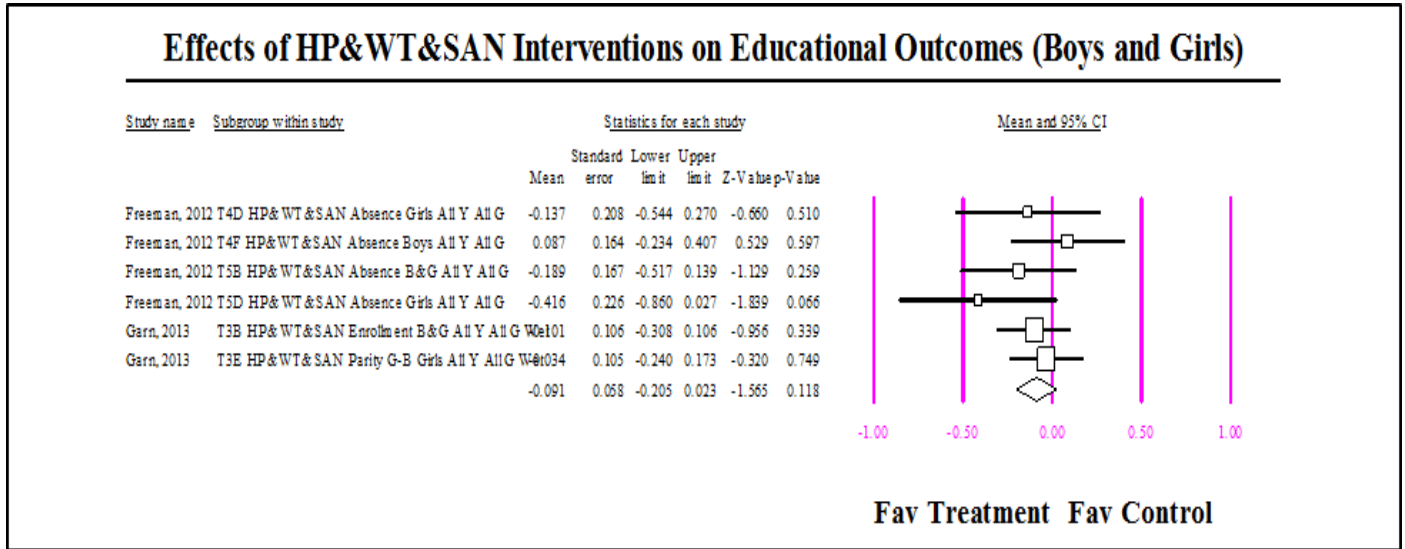
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
8.346	5.000	0.138	40.094	0.013	0.022	0.000	0.116



#### Egger's regression intercept

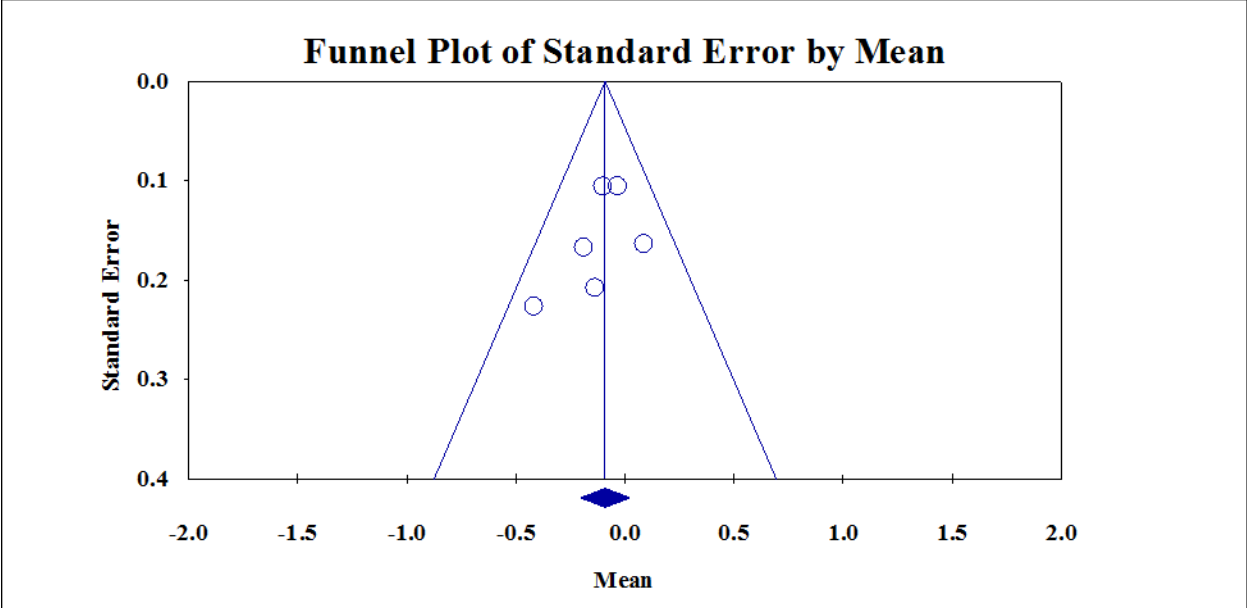
Intercept	-3.44418
Standard error	1.55584
95% lower limit (2-tailed)	-7.76388
95% upper limit (2-tailed)	0.87552
t-value	2.21372
df	4.00000
P-value (1-tailed)	0.04562
P-value (2-tailed)	0.09125

**Pooled Effect Sizes of HP, WT, and SAN Interventions for all Educational Outcomes  
All children**



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	6.000	(0.091)	-	0.058

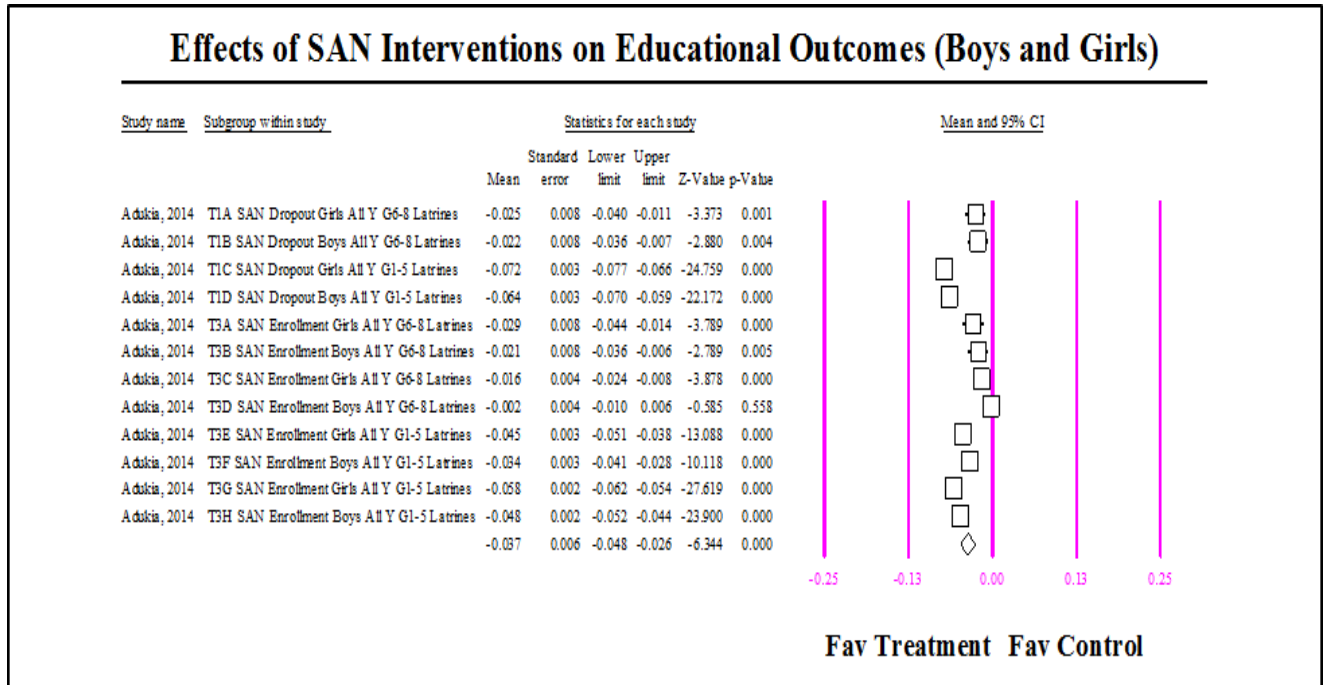
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
3.939	5.000	0.558	-	-	0.014	0.000	-



#### Egger's regression intercept

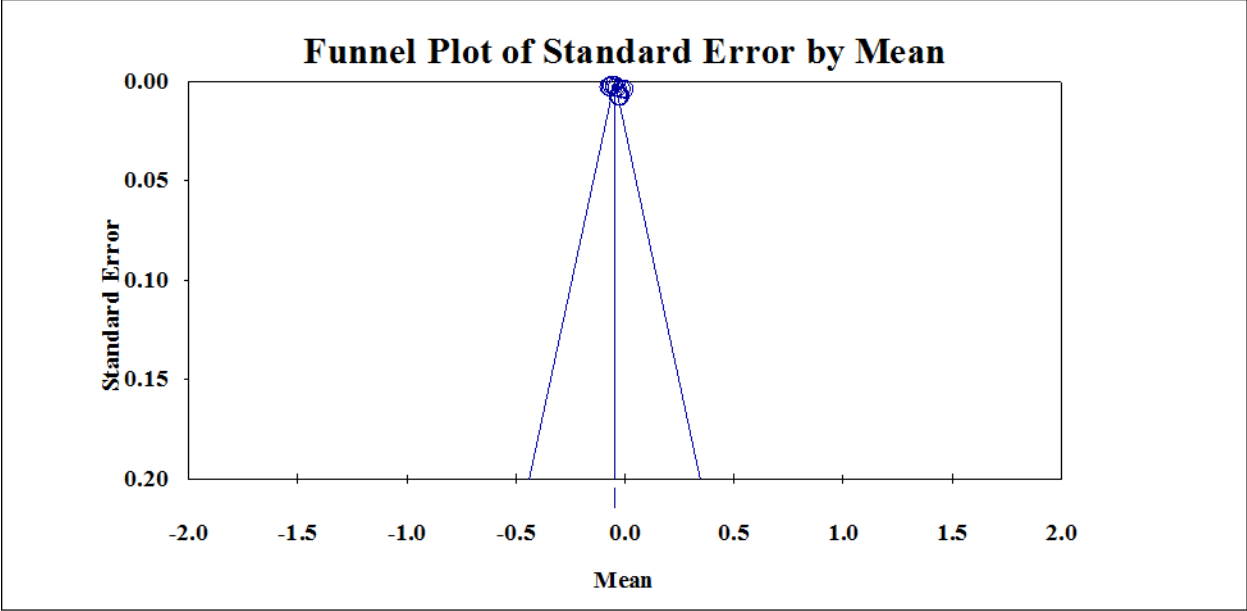
Intercept	-1.39013
Standard error	1.20587
95% lower limit (2-tailed)	-4.73815
95% upper limit (2-tailed)	1.95789
t-value	1.15281
df	4.00000
P-value (1-tailed)	0.15660
P-value (2-tailed)	0.31319

**Pooled Effect Sizes of SAN Interventions for all Educational Outcomes  
All children**



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	12.000	(0.037)	***	0.006

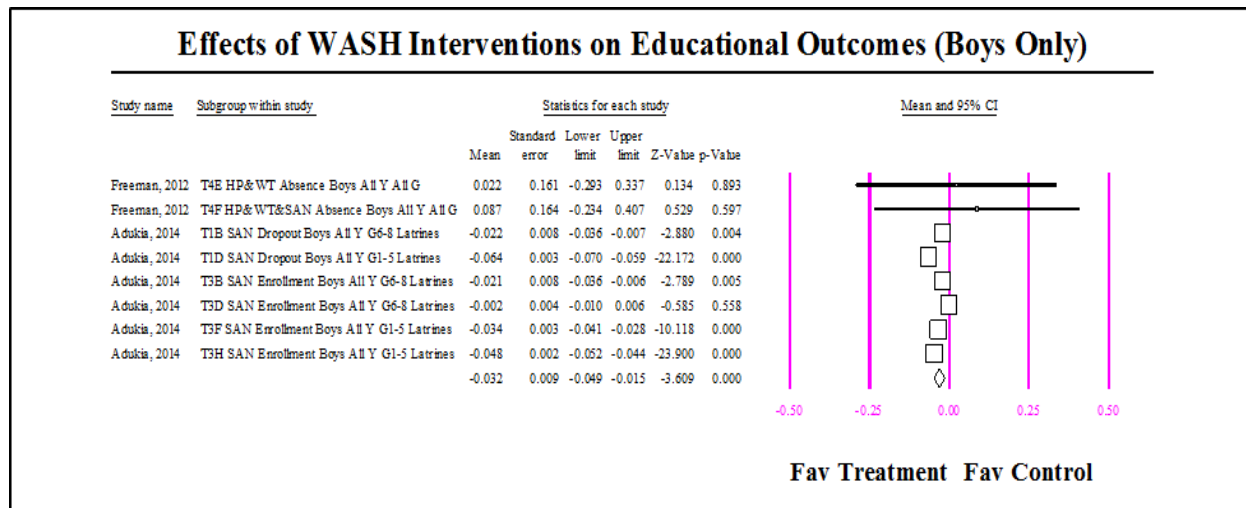
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
363.482	11.000	-	96.974	0.000	0.000	0.000	0.020



**Egger's regression intercept**

Intercept	7.62828
Standard error	3.44036
95% lower limit (2-tailed)	-0.03733
95% upper limit (2-tailed)	15.29389
t-value	2.21729
df	10.00000
P-value (1-tailed)	0.02546
P-value (2-tailed)	0.05093

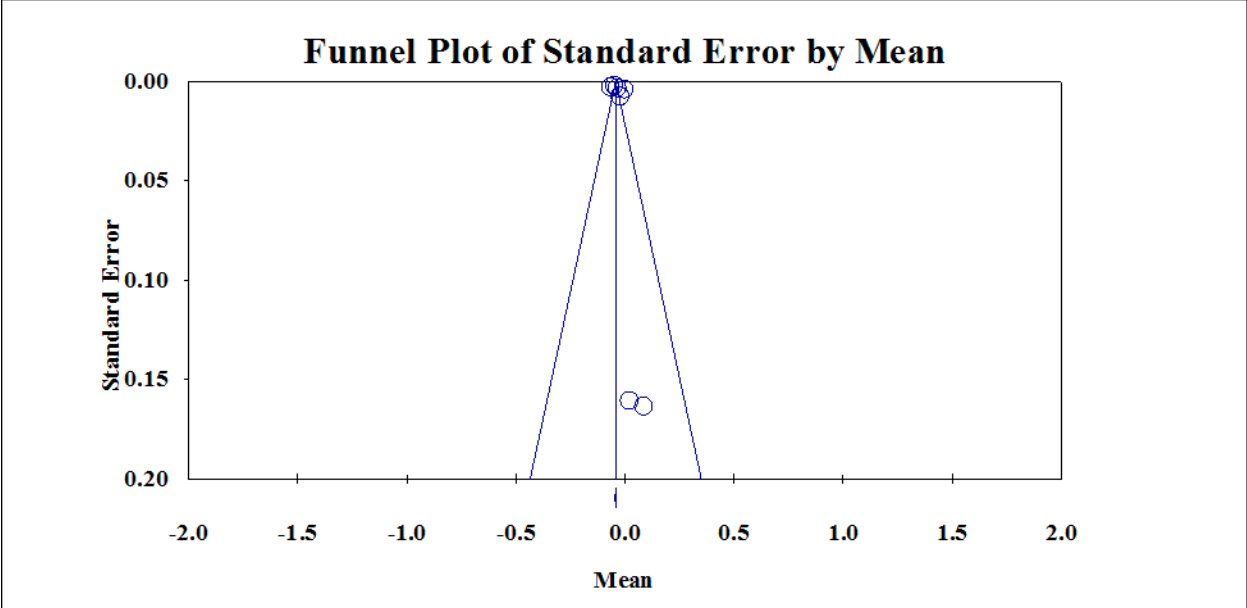
## Pooled Effect Sizes of All WASH Interventions for all Educational Outcomes Boys Only



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	8.000	(0.032)	***	0.009

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
181.254	7.000	-	96.138	0.000	0.000	0.000	0.021

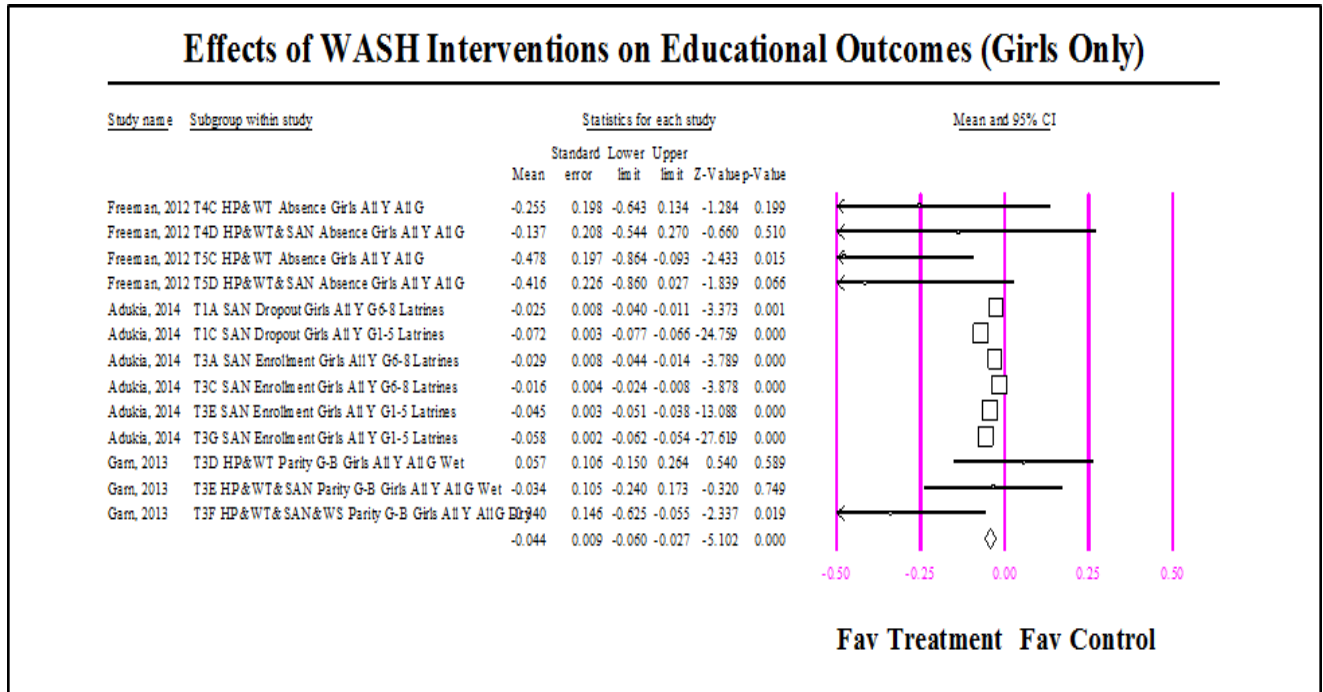




#### Egger's regression intercept

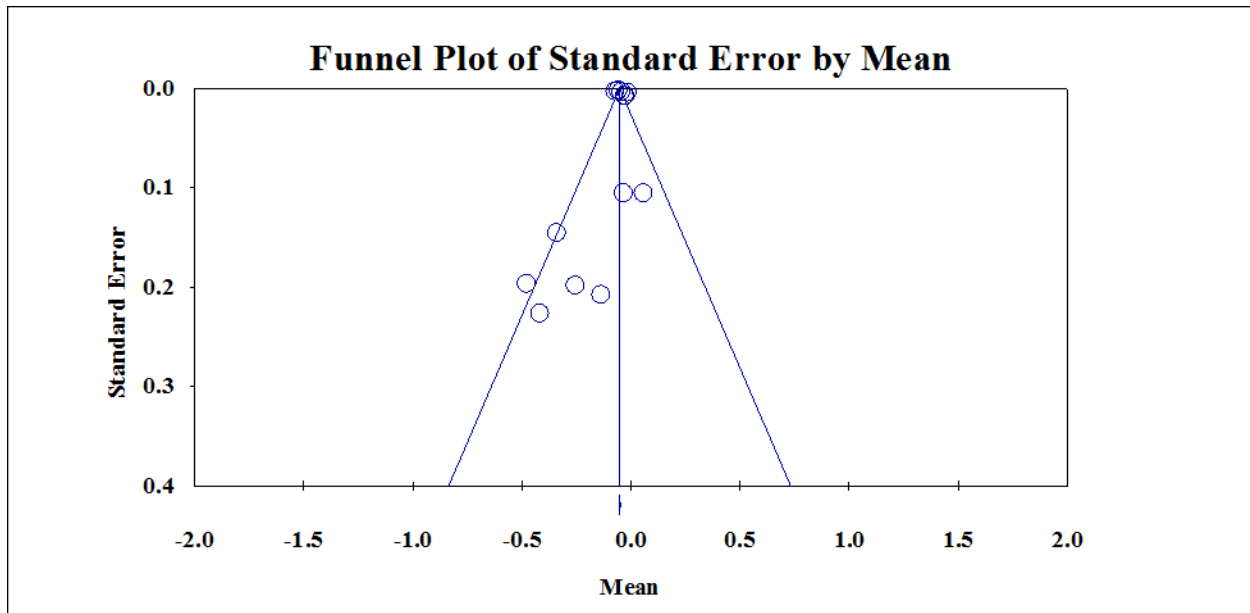
Intercept	3.01836
Standard error	2.93010
95% lower limit (2-tailed)	-4.15133
95% upper limit (2-tailed)	10.18804
t-value	1.03012
df	6.00000
P-value (1-tailed)	0.17134
P-value (2-tailed)	0.34268

## Pooled Effect Sizes of All WASH Interventions for all Educational Outcomes Girls Only



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	13.000	(0.044)	***	0.009

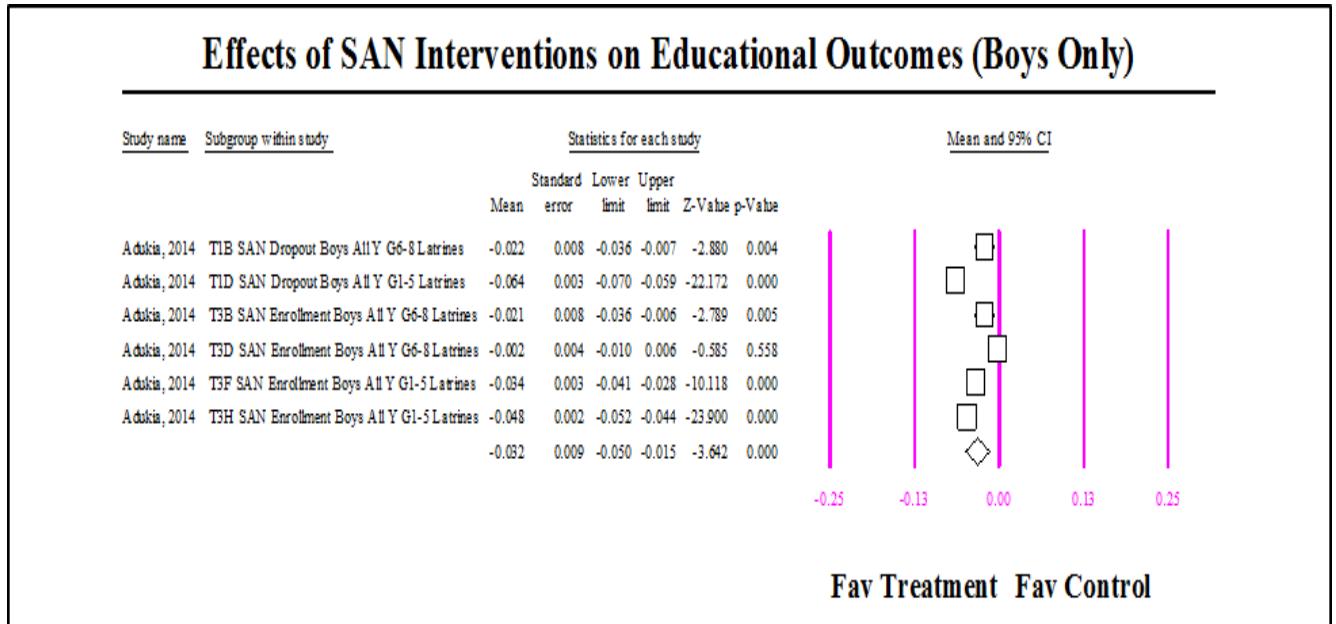
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
172.676	12.000	-	93.051	0.000	0.000	0.000	0.021



### Egger's regression intercept

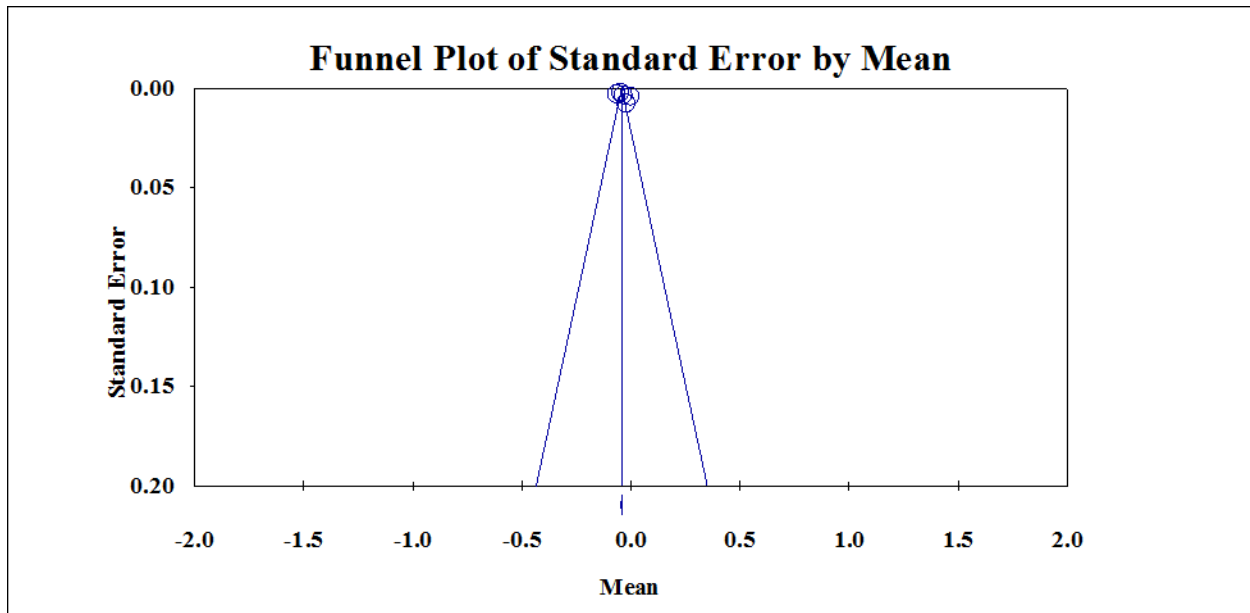
Intercept	0.28223
Standard error	1.42398
95% lower limit (2-tailed)	-2.85193
95% upper limit (2-tailed)	3.41639
t-value	0.19820
df	11.00000
P-value (1-tailed)	0.42325
P-value (2-tailed)	0.84651

**Pooled Effect Sizes of SAN Interventions for all Educational Outcomes  
Boys Only**



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	6.000	(0.032)	***	0.009

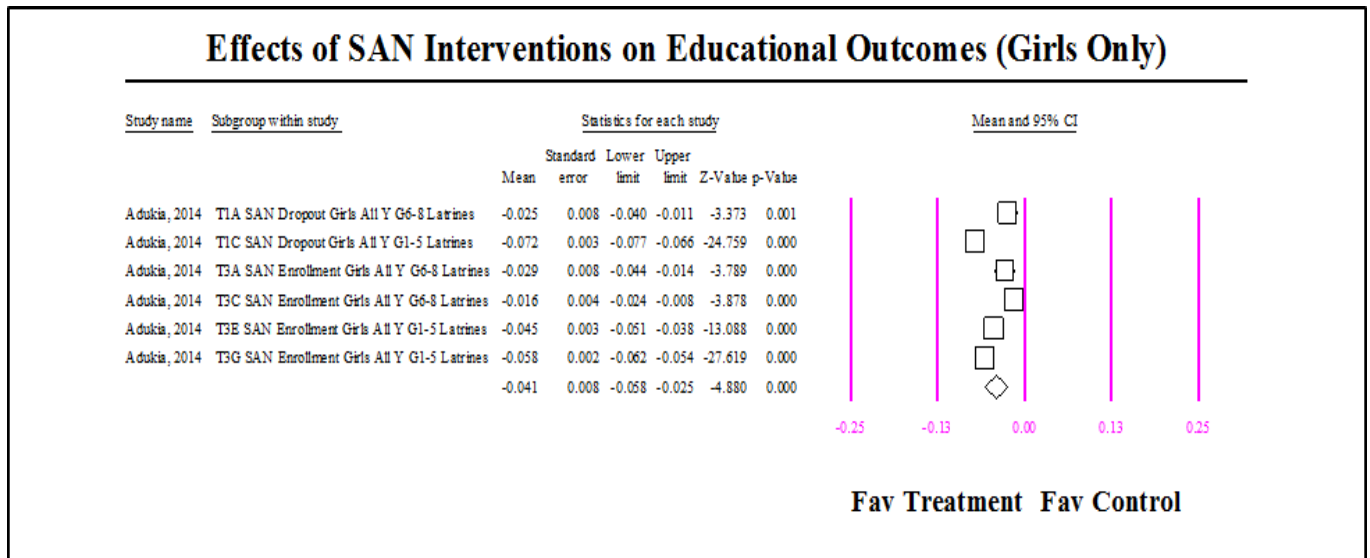
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
180.470	5.000	-	97.229	0.000	0.000	0.000	0.021



#### Egger's regression intercept

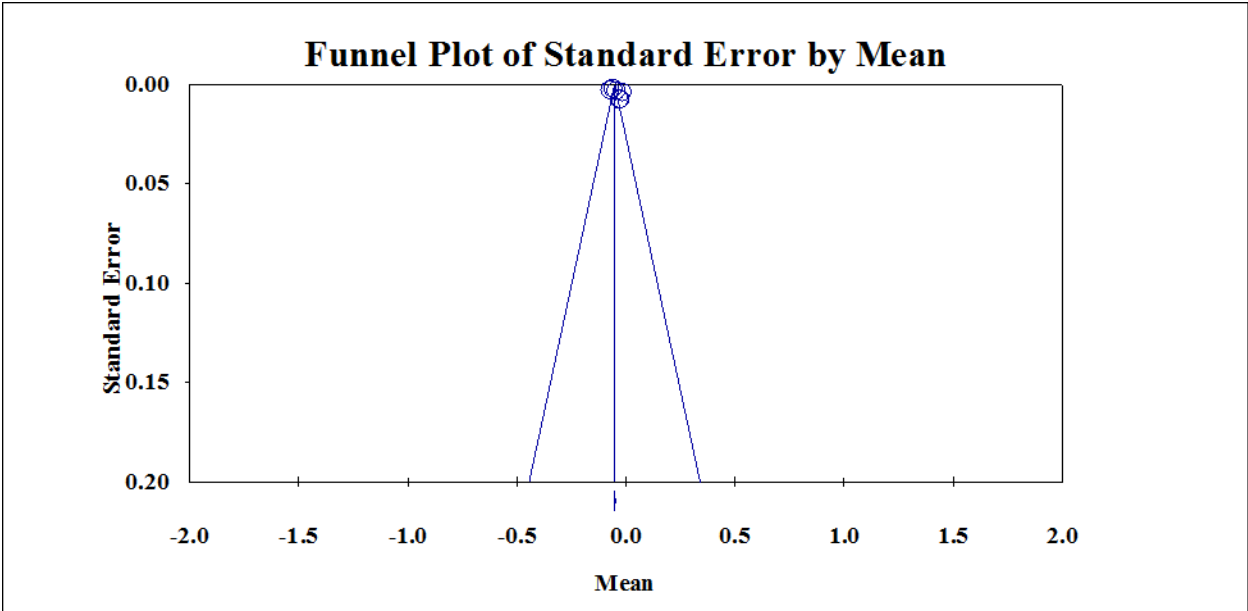
Intercept	7.38846
Standard error	5.36493
95% lower limit (2-tailed)	-7.50697
95% upper limit (2-tailed)	22.28390
t-value	1.37718
df	4.00000
P-value (1-tailed)	0.12025
P-value (2-tailed)	0.24050

## Pooled Effect Sizes of SAN Interventions for all Educational Outcomes Girls Only



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	6.000	(0.041)	***	0.008

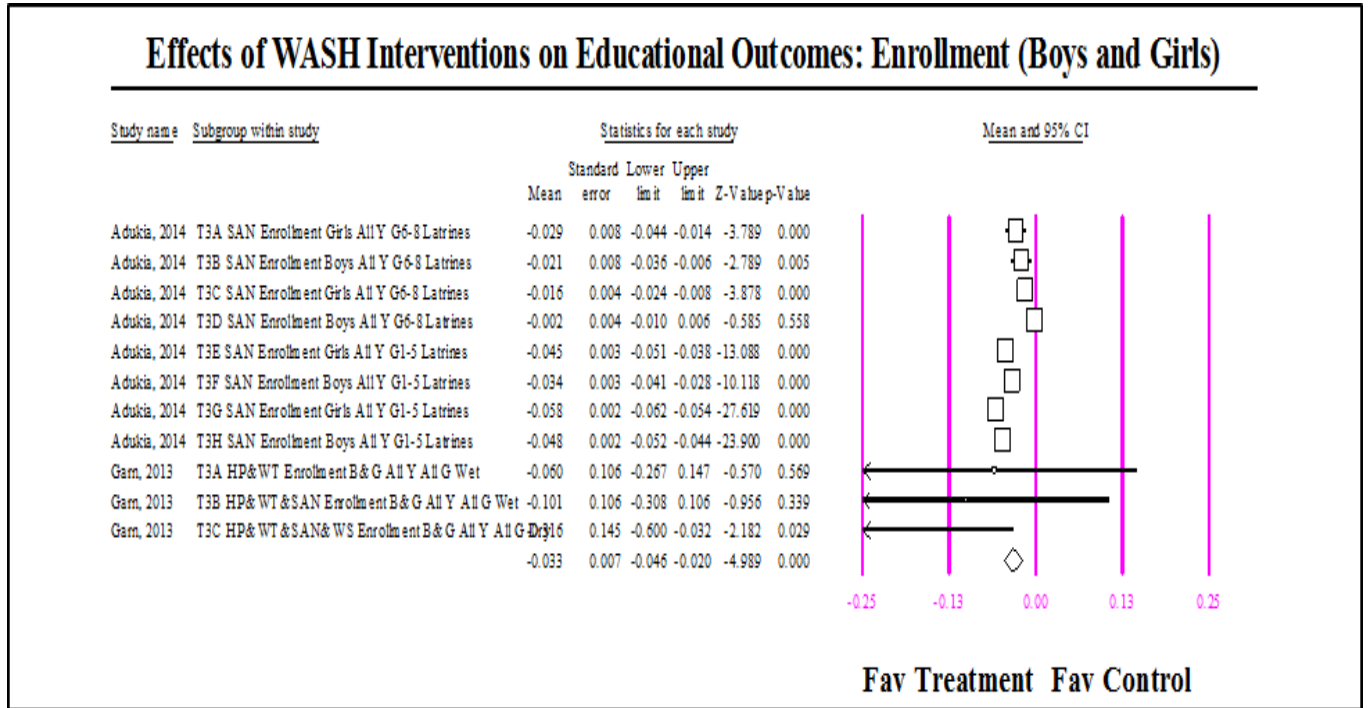
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
159.160	5.000	-	96.859	0.000	0.000	0.000	0.020



#### Egger's regression intercept

Intercept	8.12247
Standard error	4.85393
95% lower limit (2-tailed)	-5.35419
95% upper limit (2-tailed)	21.59914
t-value	1.67338
df	4.00000
P-value (1-tailed)	0.08478
P-value (2-tailed)	0.16957

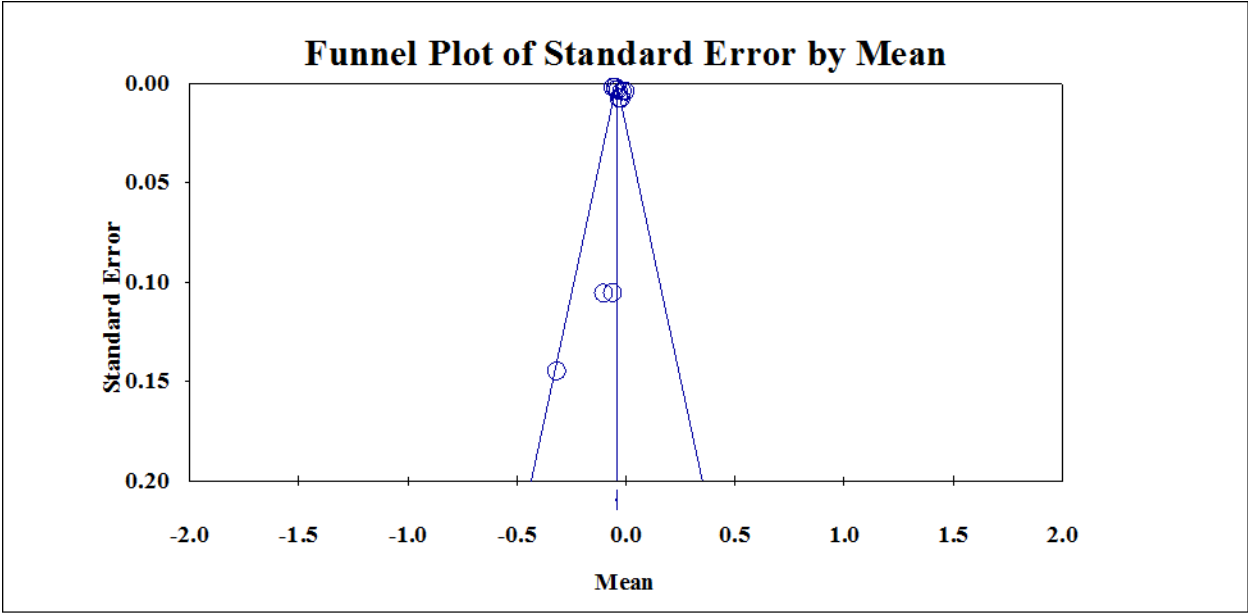
**Pooled Effect Sizes of All WASH Interventions for Enrollment  
All Children**



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	11.000	(0.033)	***	0.007

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
220.204	10.000	-	95.459	0.000	0.000	0.000	0.018

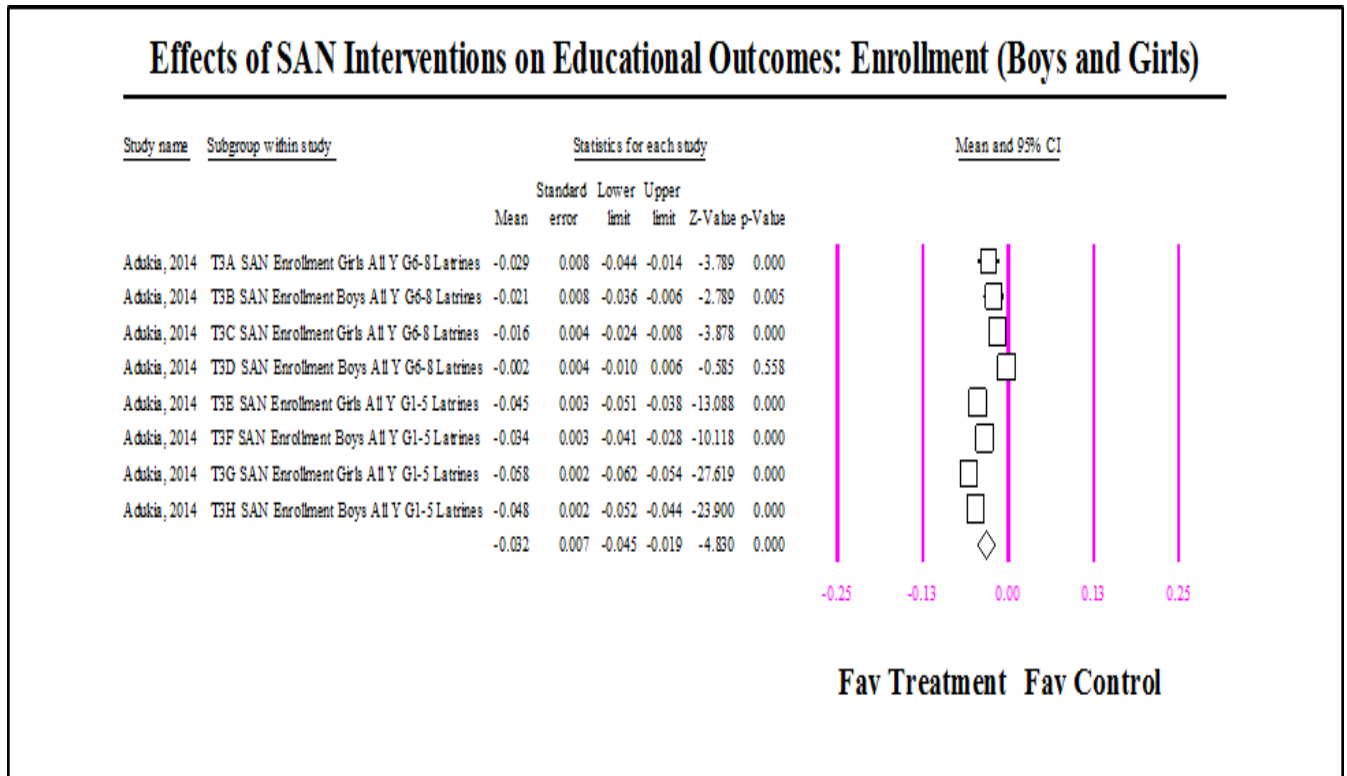




#### Egger's regression intercept

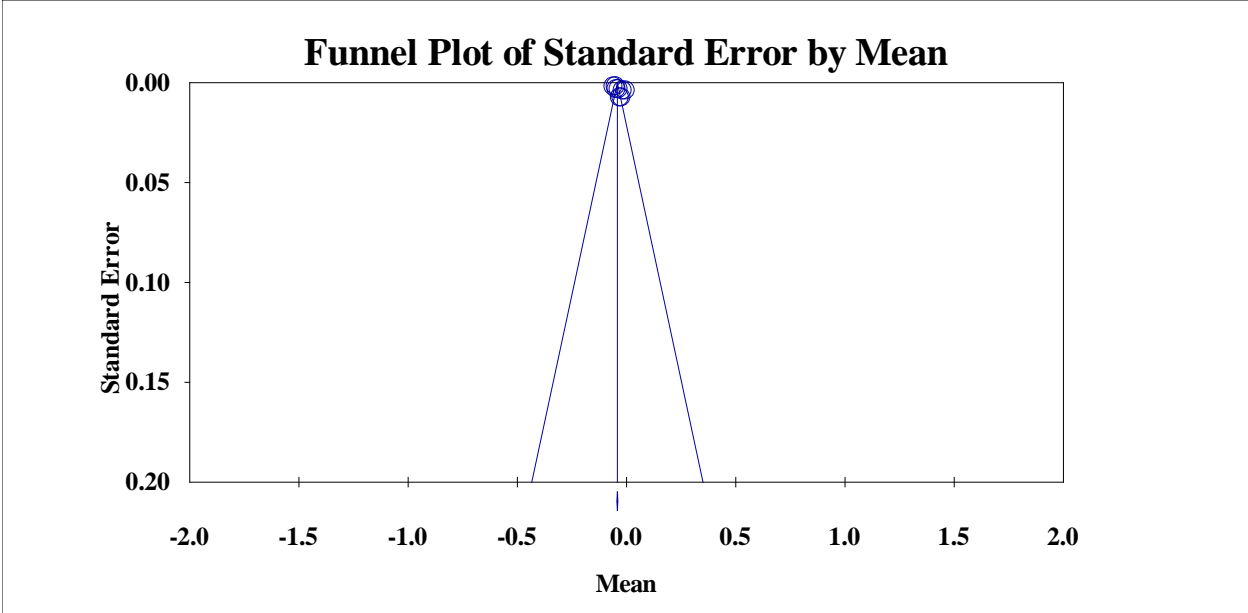
Intercept	2.23080
Standard error	2.30258
95% lower limit (2-tailed)	-2.97799
95% upper limit (2-tailed)	7.43960
t-value	0.96883
df	9.00000
P-value (1-tailed)	0.17897
P-value (2-tailed)	0.35794

## Pooled Effect Sizes of SAN Interventions for Enrollment All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	8.000	(0.032)	***	0.007

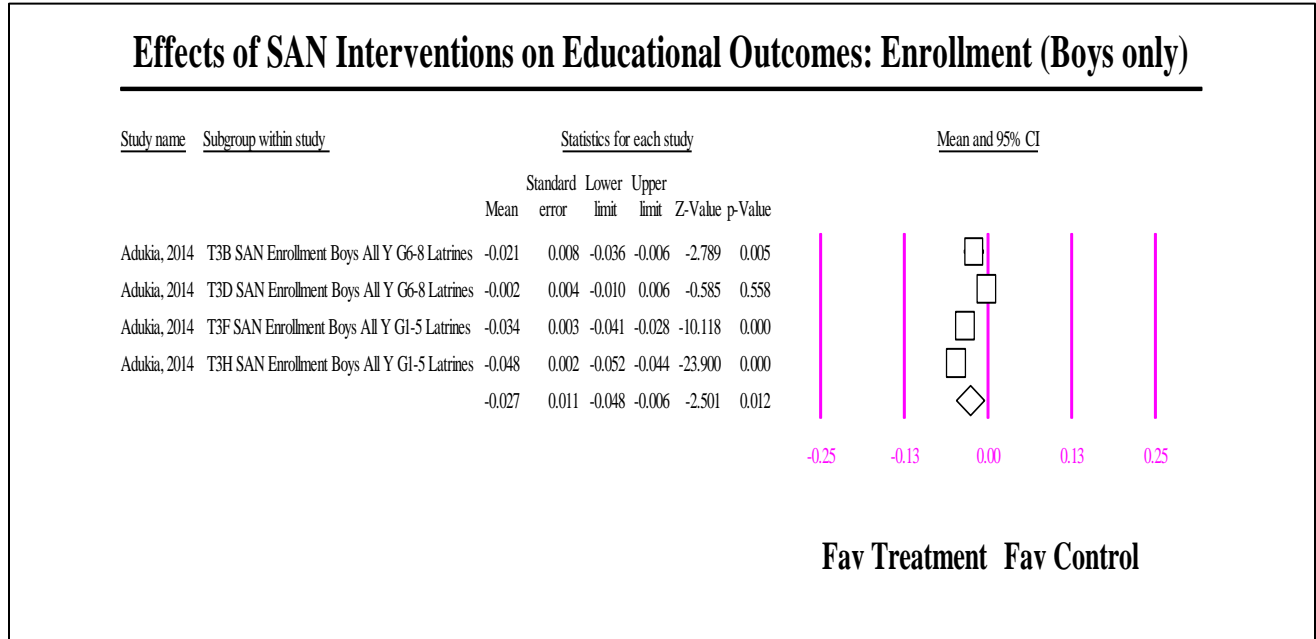
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
216.291	7.000	-	96.764	0.000	0.000	0.000	0.018



**Egger's regression intercept**

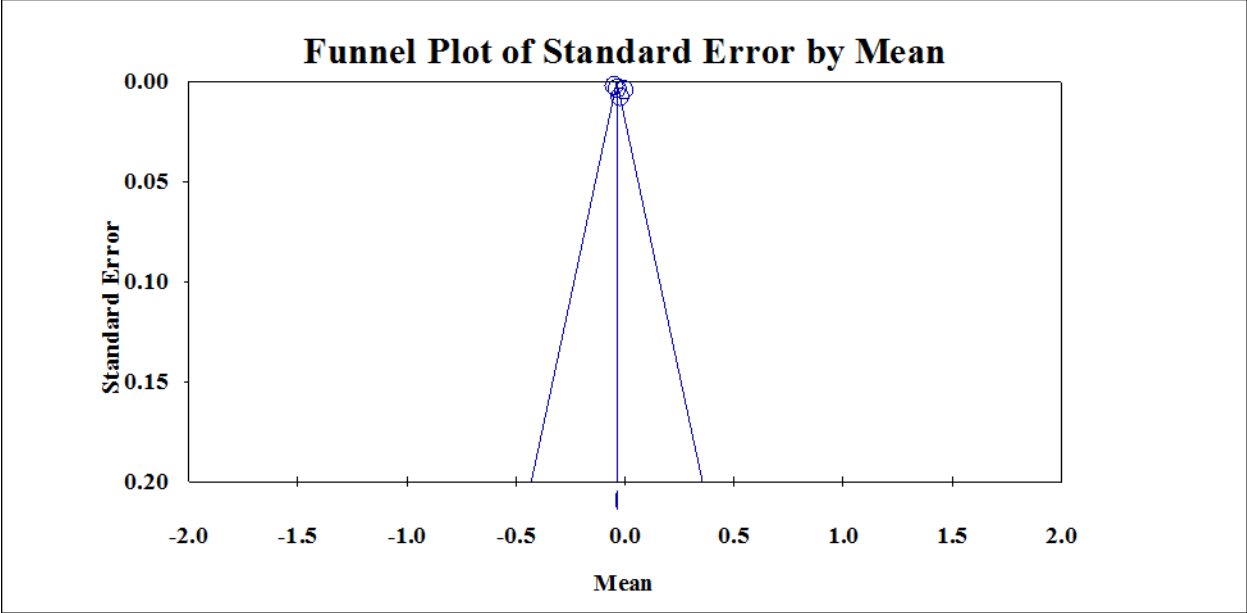
Intercept	8.95534
Standard error	3.71516
95% lower limit (2-tailed)	-0.13534
95% upper limit (2-tailed)	18.04602
t-value	2.41048
df	6.00000
P-value (1-tailed)	0.02627
P-value (2-tailed)	0.05254

## Pooled Effect Sizes of SAN Interventions for Enrollment Boys Only



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	(0.027)	**	0.011

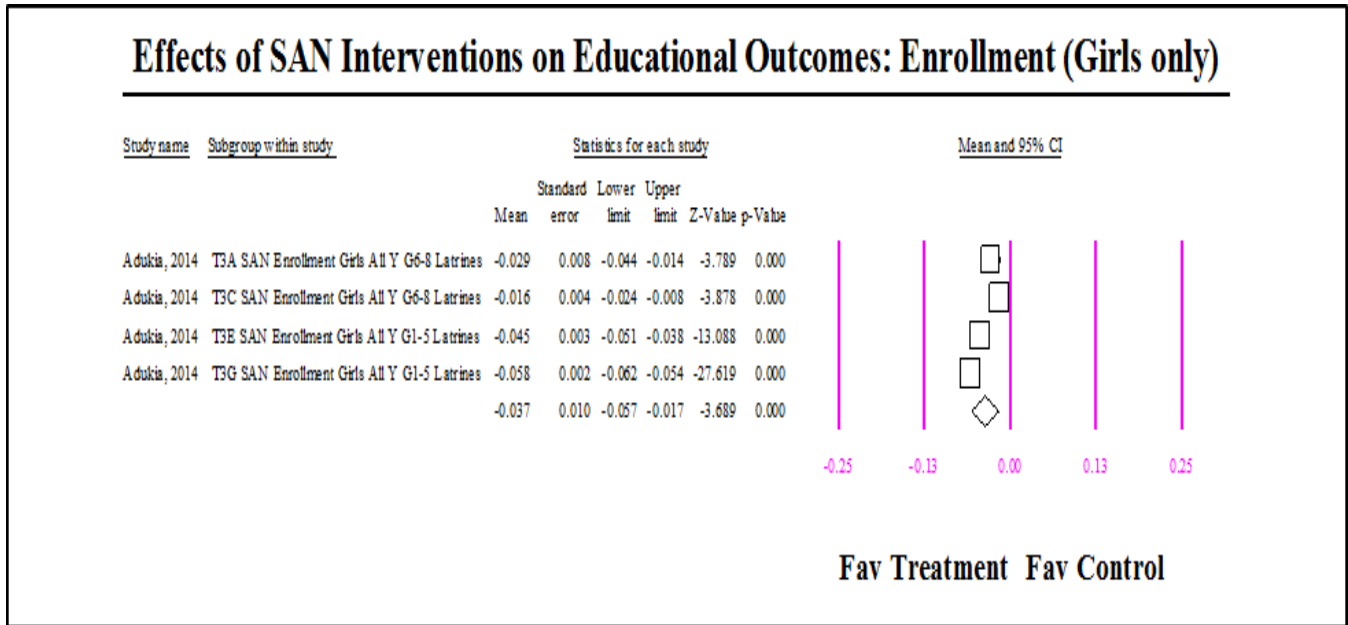
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
105.234	3.000	-	97.149	0.000	0.000	0.000	0.021



#### Egger's regression intercept

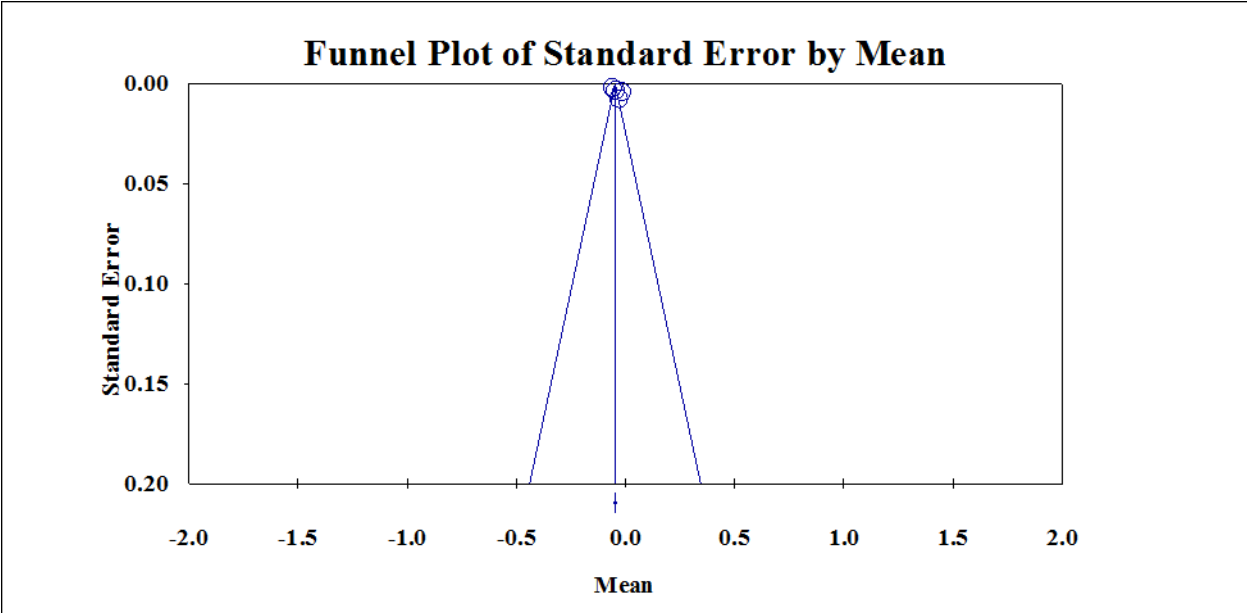
Intercept	9.11044
Standard error	5.89360
95% lower limit (2-tailed)	-16.24766
95% upper limit (2-tailed)	34.46854
t-value	1.54582
df	2.00000
P-value (1-tailed)	0.13109
P-value (2-tailed)	0.26218

**Pooled Effect Sizes of SAN Interventions for Enrollment  
Girls Only**



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	(0.037)	***	0.010

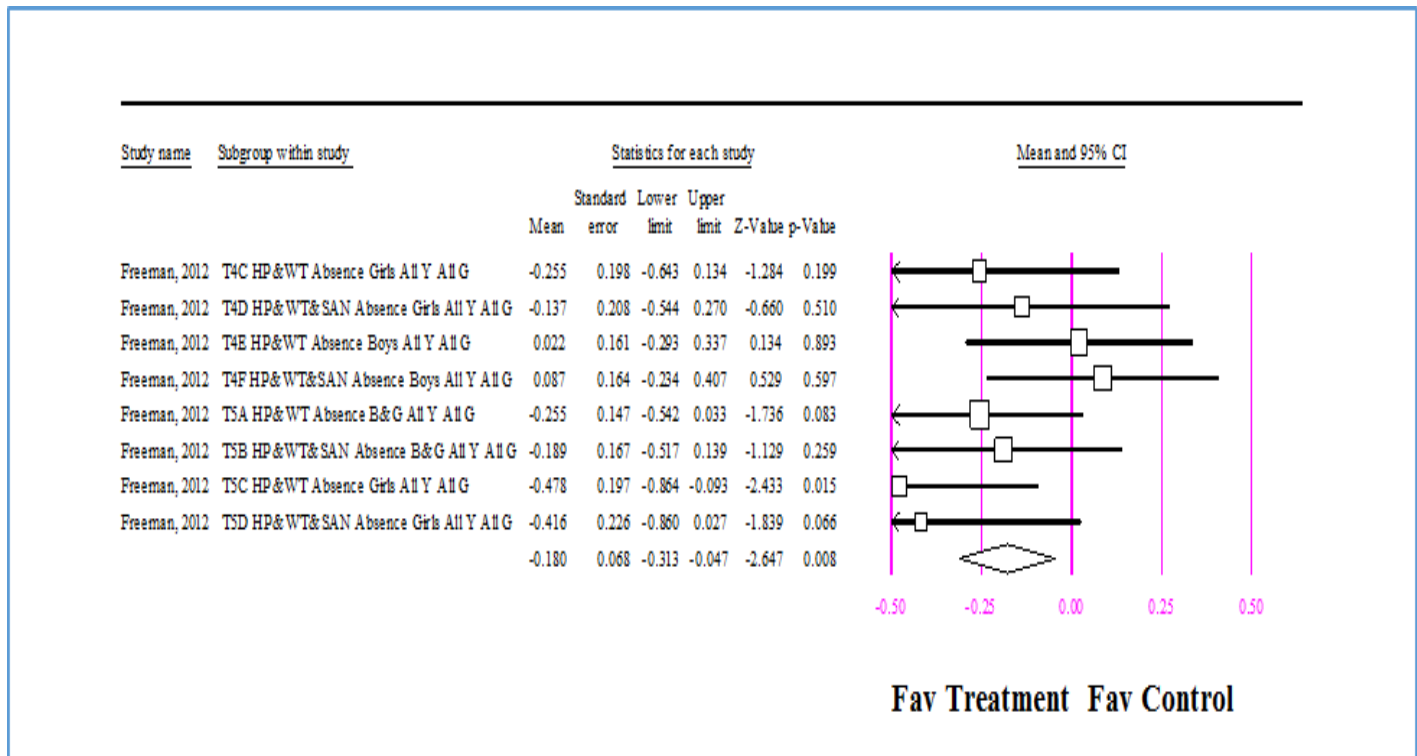
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
91.218	3.000	-	96.711	0.000	0.000	0.000	0.020



**Egger's regression intercept**

Intercept	9.15777
Standard error	5.47456
95% lower limit (2-tailed)	-14.39737
95% upper limit (2-tailed)	32.71291
t-value	1.67279
df	2.00000
P-value (1-tailed)	0.11817
P-value (2-tailed)	0.23634

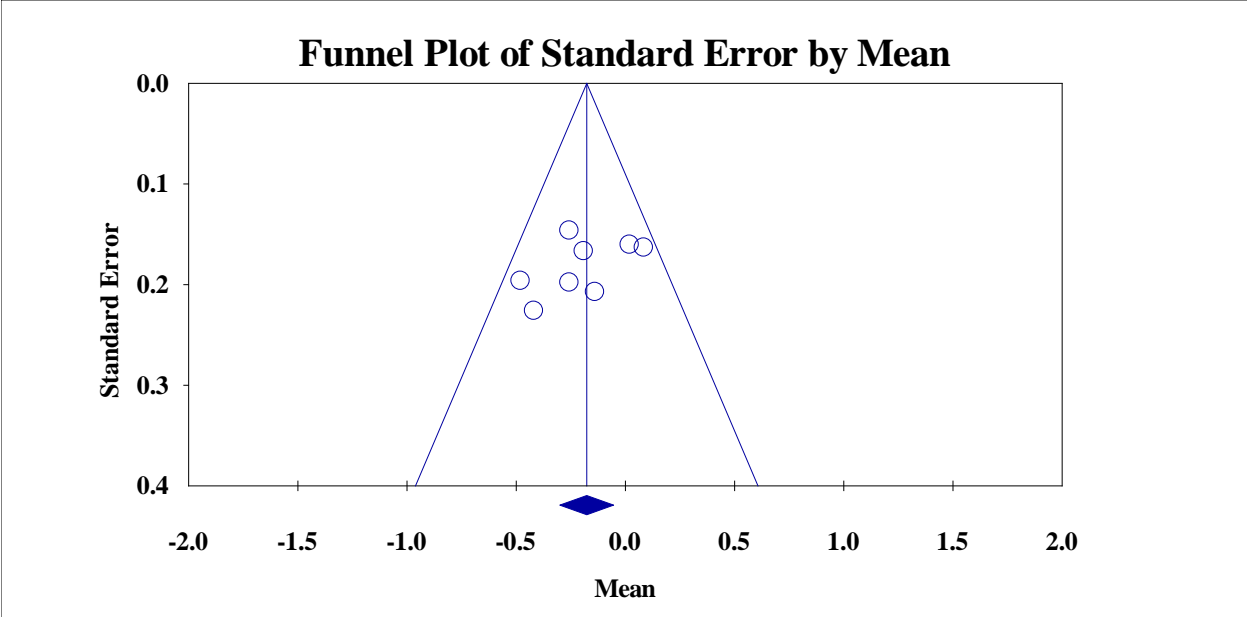
## Pooled Effect Sizes of All WASH Interventions for School Absence All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	8.000	(0.180)	***	0.068

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
8.065	7.000	0.327	13.210	0.005	0.020	0.000	0.070

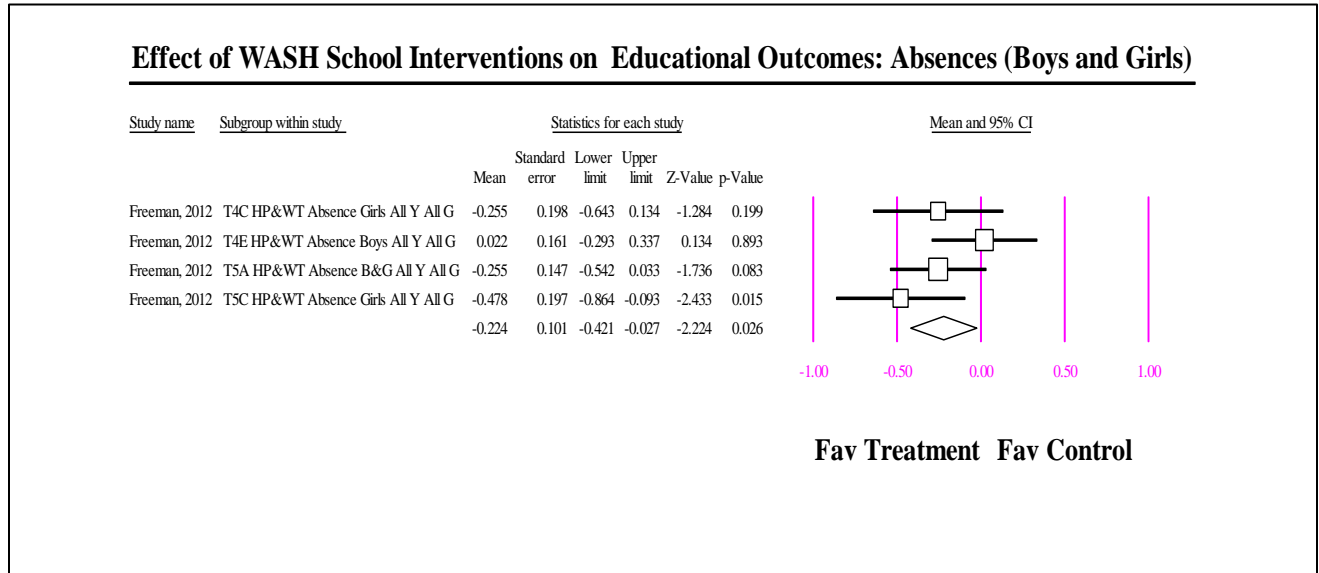




#### Egger's regression intercept

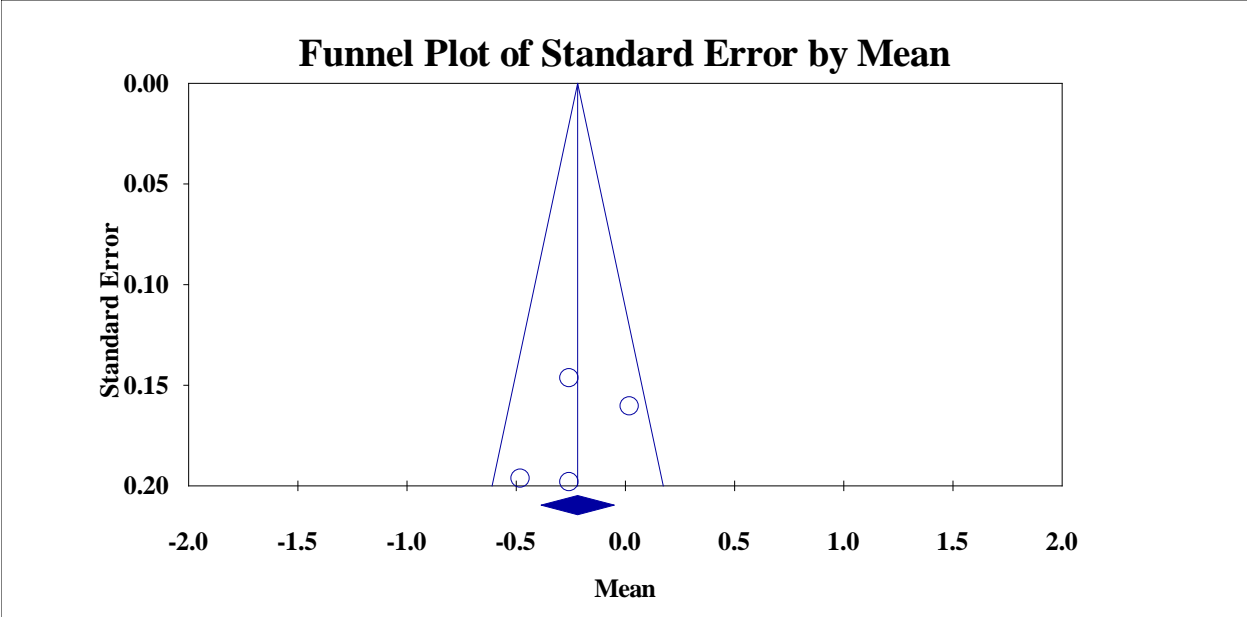
Intercept	-3.51380
Standard error	2.58374
95% lower limit (2-tailed)	-9.83599
95% upper limit (2-tailed)	2.80840
t-value	1.35996
df	6.00000
P-value (1-tailed)	0.11136
P-value (2-tailed)	0.22272

## Pooled Effect Sizes of HP and WT Interventions for School Absence All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	(0.224)	**	0.101

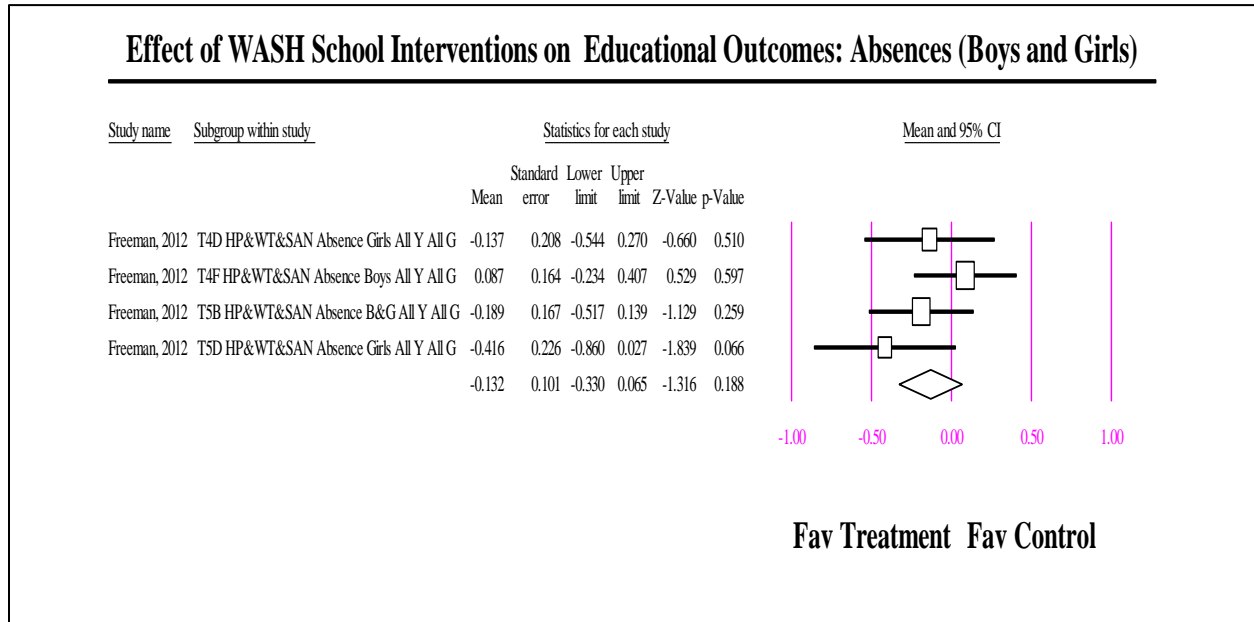
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
4.073	3.000	0.254	26.340	0.011	0.033	0.001	0.104



#### Egger's regression intercept

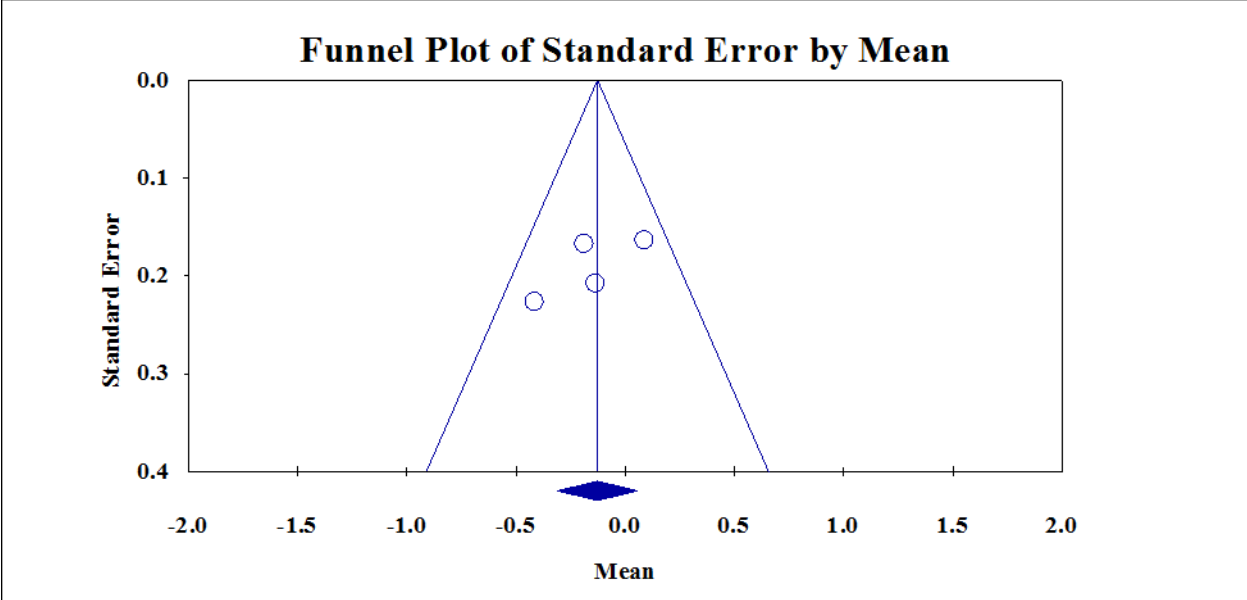
Intercept	-3.74968
Standard error	4.82240
95% lower limit (2-tailed)	-24.49880
95% upper limit (2-tailed)	16.99944
t-value	0.77755
df	2.00000
P-value (1-tailed)	0.25910
P-value (2-tailed)	0.51821

## Pooled Effect Sizes of HP, WT and SAN Interventions for School Absence All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	(0.132)	-	0.101

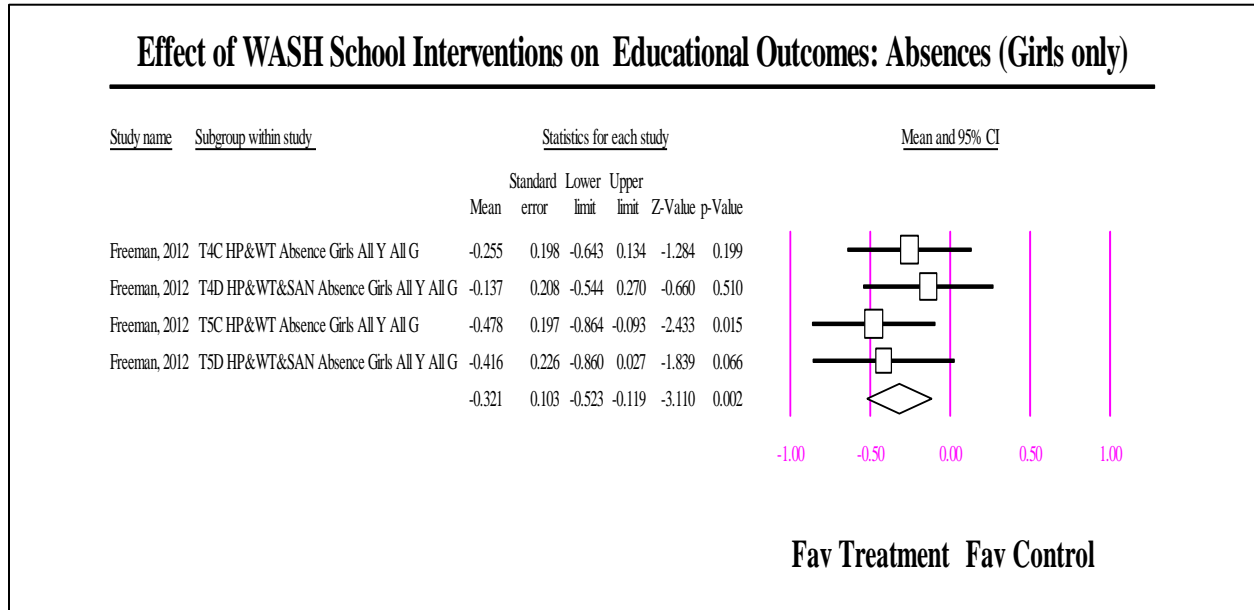
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
3.476	3.000	0.324	13.702	0.006	0.033	0.001	0.075



#### Egger's regression intercept

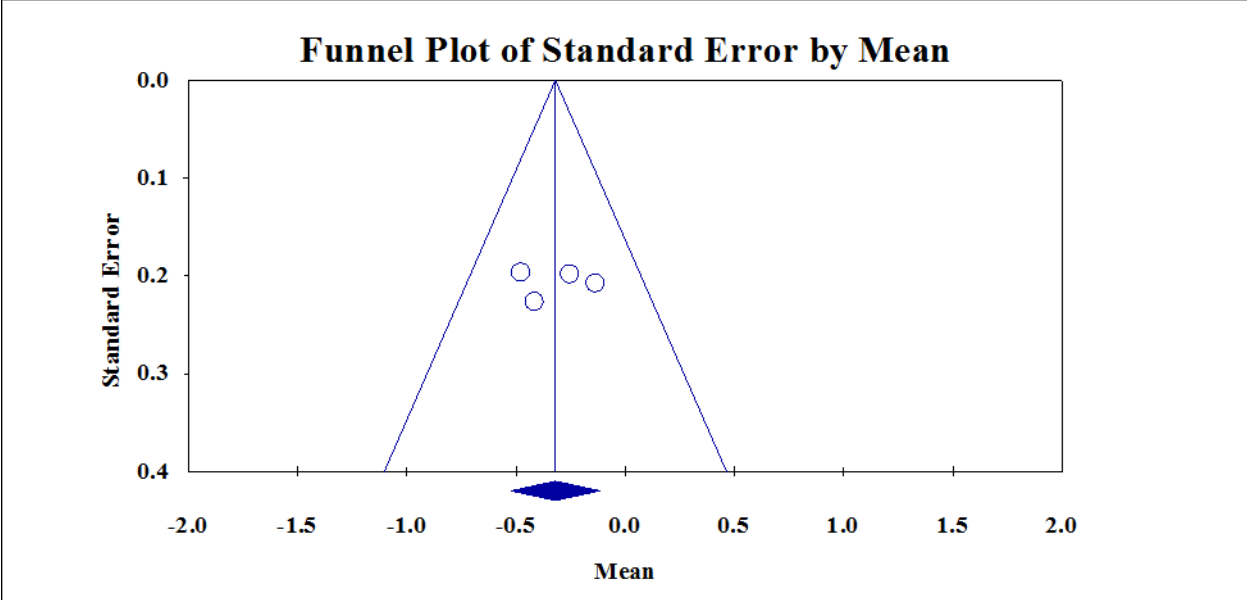
Intercept	-5.02404
Standard error	3.30939
95% lower limit (2-tailed)	-19.26319
95% upper limit (2-tailed)	9.21511
t-value	1.51812
df	2.00000
P-value (1-tailed)	0.13415
P-value (2-tailed)	0.26830

## Pooled Effect Sizes of All WASH Interventions for School Absence Girls Only



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	(0.321)	***	0.103

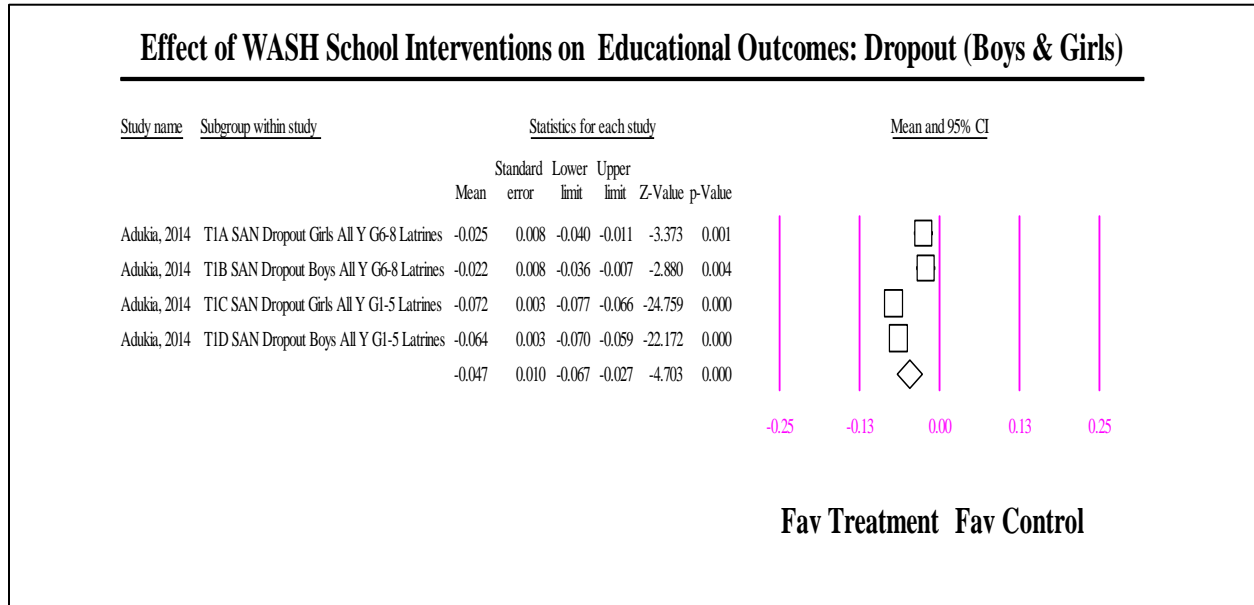
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
1.714	3.000	0.634	-	-	0.035	0.001	-



#### Egger's regression intercept

Intercept	-0.45117
Standard error	8.45175
95% lower limit (2-tailed)	-36.81611
95% upper limit (2-tailed)	35.91377
t-value	0.05338
df	2.00000
P-value (1-tailed)	0.48114
P-value (2-tailed)	0.96228

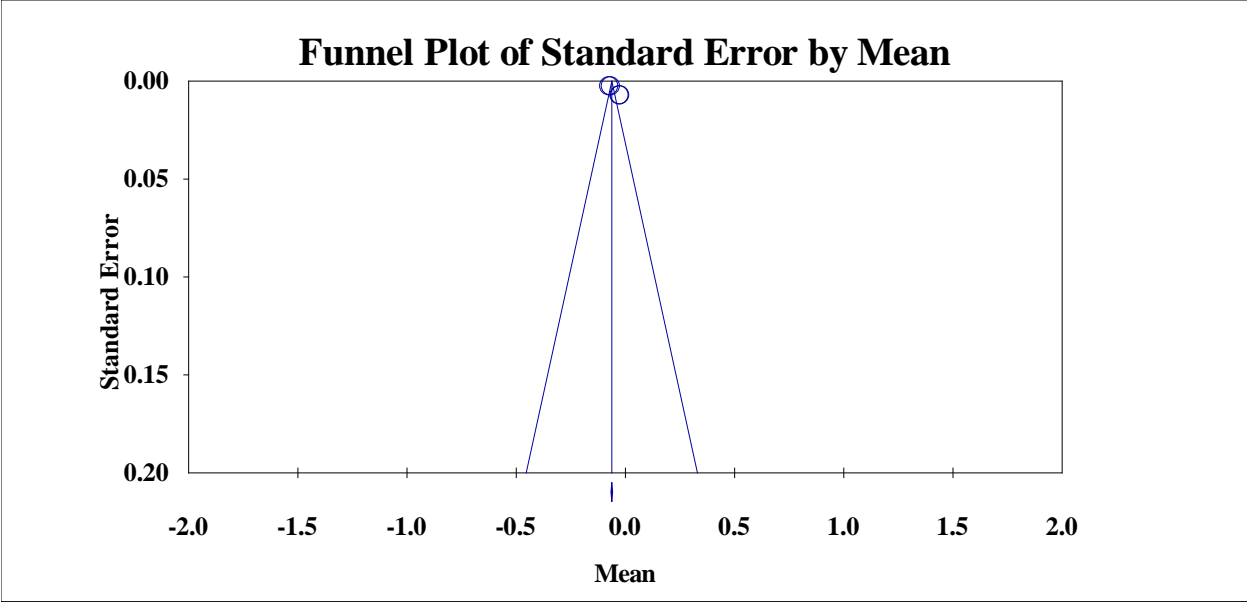
## Pooled Effect Sizes of All WASH Interventions for School Dropout All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	(0.047)	***	0.010

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
64.993	3.000	0.000	95.384	0.000	0.000	0.000	0.019



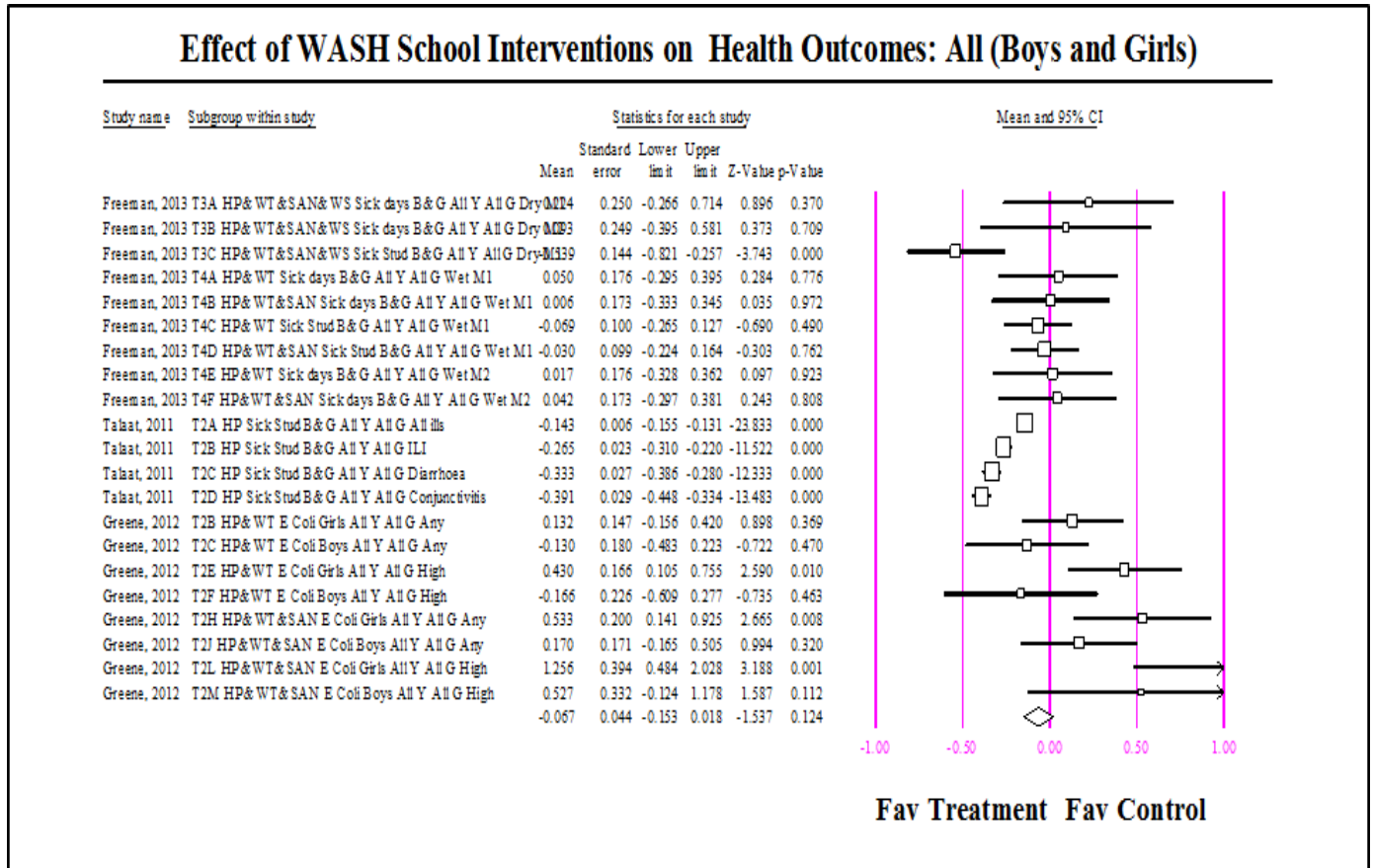


**Egger's regression intercept**

Intercept	9.69565
Standard error	1.62719
95% lower limit (2-tailed)	2.69440
95% upper limit (2-tailed)	16.69690
t-value	5.95851
df	2.00000
P-value (1-tailed)	0.01351
P-value (2-tailed)	0.02703

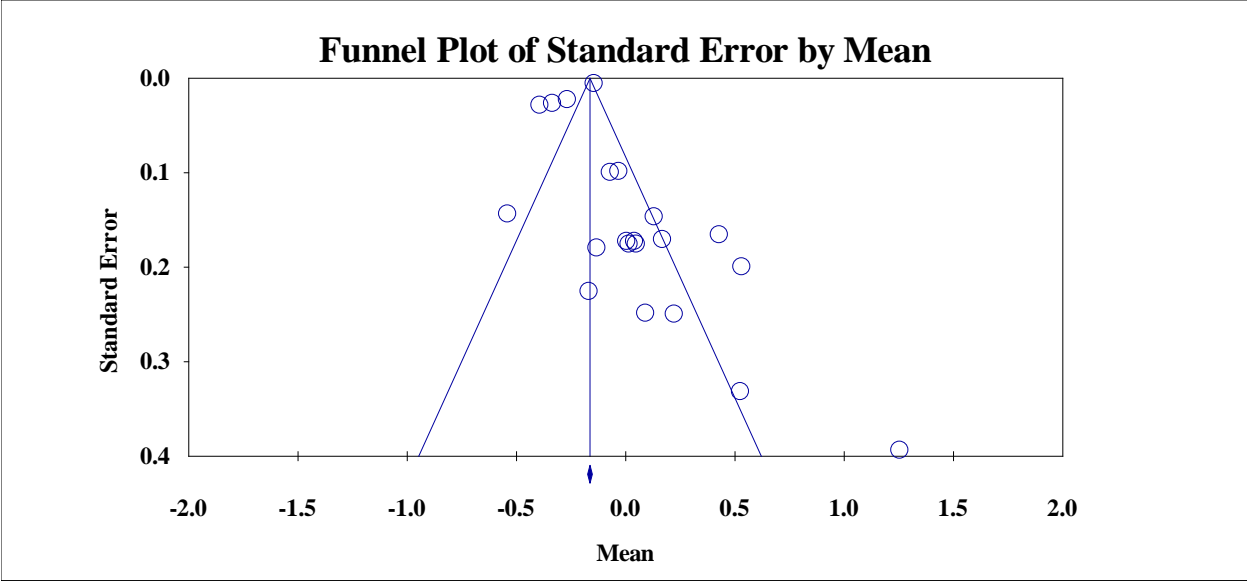
**Annex 2.4: WASH Interventions, Health Outcomes**

**Pooled Effect Sizes of All WASH Interventions for Health Outcomes  
All Children**



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	21.000	(0.067)	-	0.044

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
200.065	20.000	-	90.003	0.020	0.018	0.000	0.140

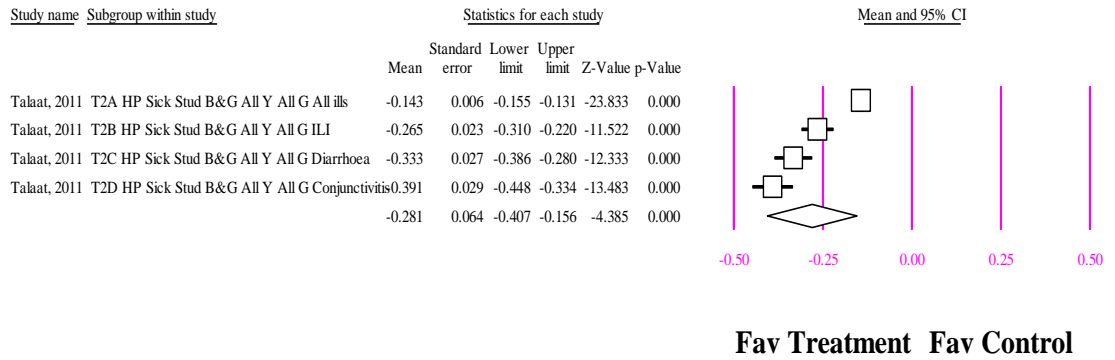


**Egger's regression intercept**

Intercept	0.50995
Standard error	0.78706
95% lower limit (2-tailed)	-1.13738
95% upper limit (2-tailed)	2.15728
t-value	0.64792
df	19.00000
P-value (1-tailed)	0.26240
P-value (2-tailed)	0.52479

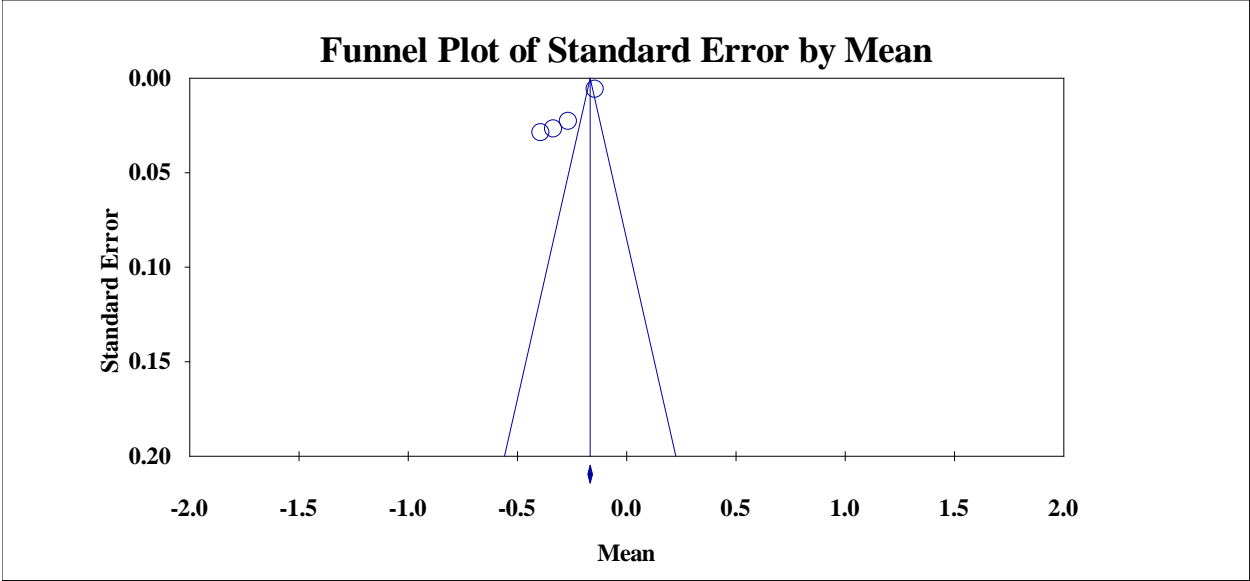
**Pooled Effect Sizes of HP Interventions for Health Outcomes  
All Children**

**Effect of HP School Interventions on Health Outcomes: All (Boys and Girls)**



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	(0.281)	***	0.064

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
131.612	3.000	-	97.721	0.016	0.016	0.000	0.126

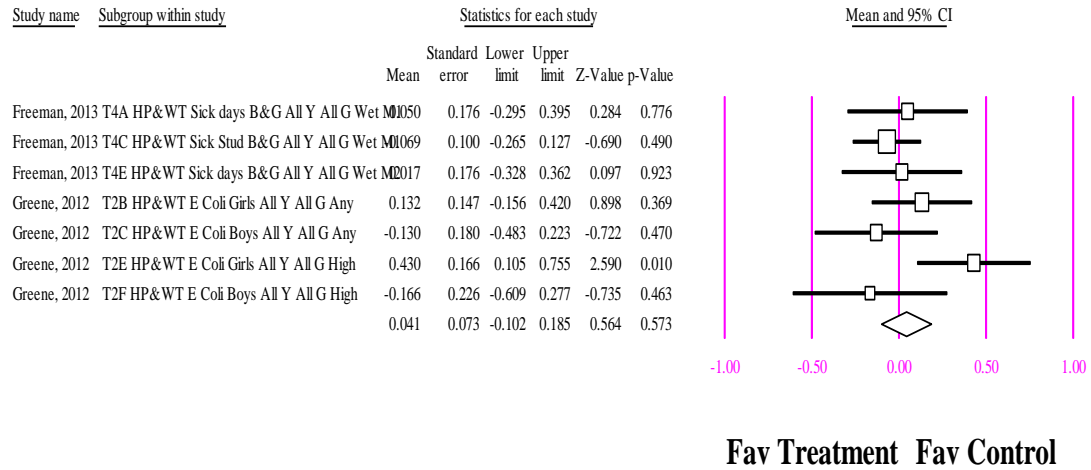


### Egger's regression intercept

Intercept	-9.11512
Standard error	1.11048
95% lower limit (2-tailed)	-13.89314
95% upper limit (2-tailed)	-4.33710
t-value	8.20825
df	2.00000
P-value (1-tailed)	0.00726
P-value (2-tailed)	0.01452

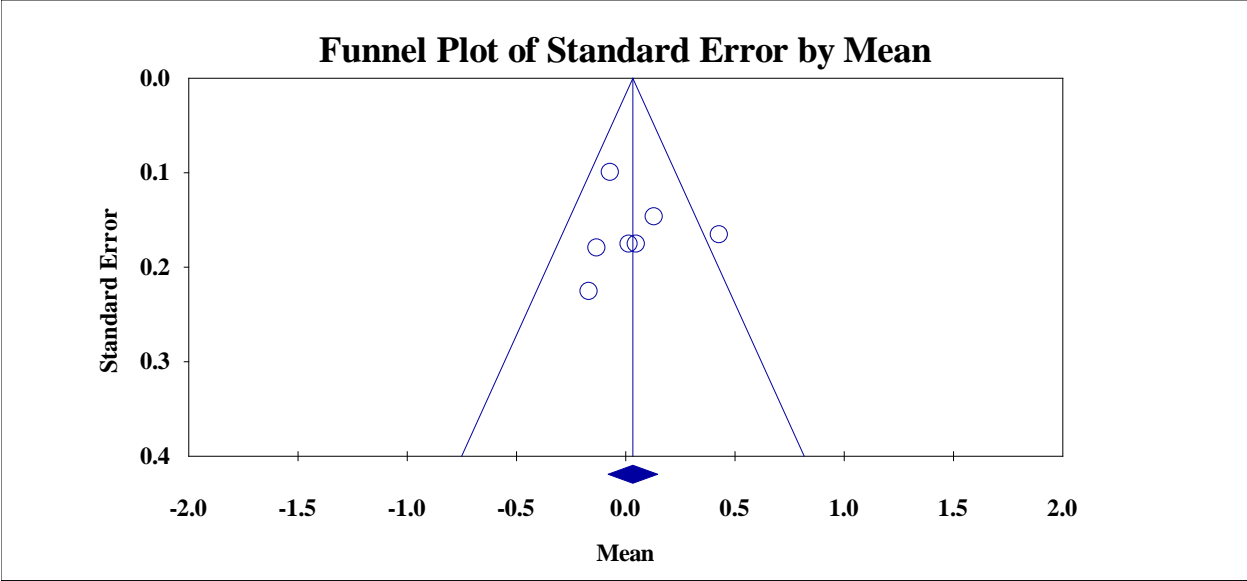
## Pooled Effect Sizes of HP and WT Interventions for Health Outcomes All Children

### Effect of HP&WT School Interventions on Health Outcomes: All (Boys and Girls)



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	7.000	0.041	-	0.073

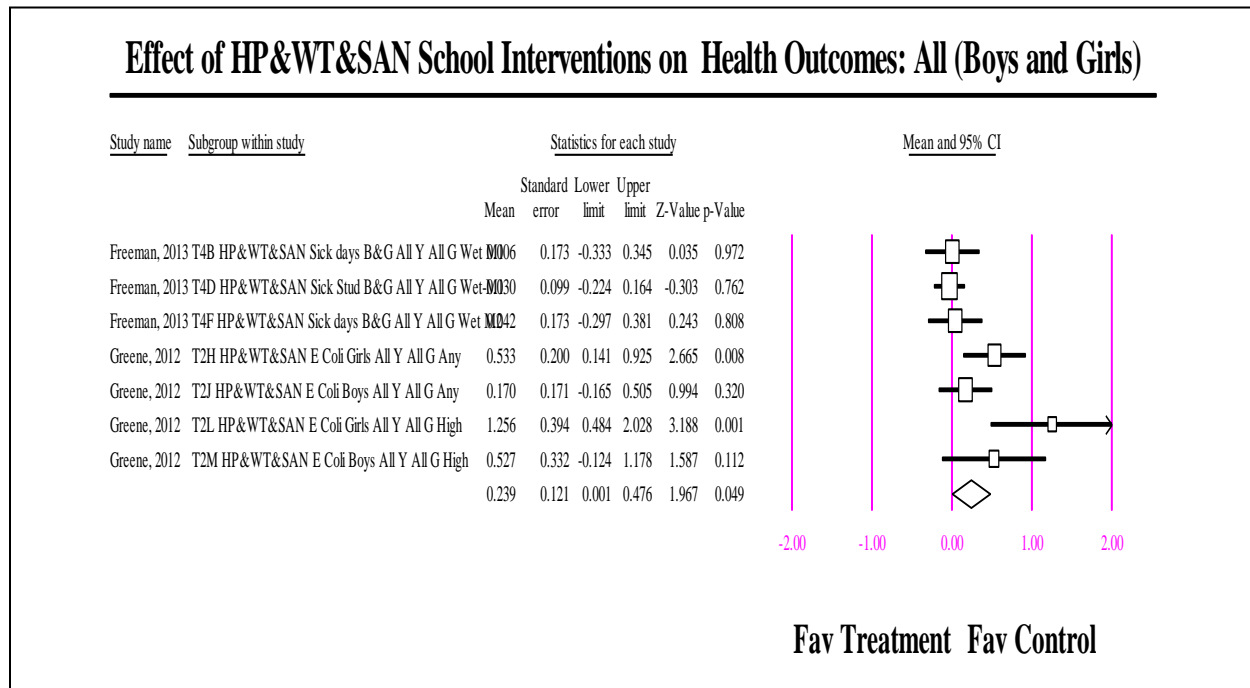
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
8.826	6.000	0.184	32.023	0.012	0.021	0.000	0.108



#### Egger's regression intercept

Intercept	0.63228
Standard error	1.97497
95% lower limit (2-tailed)	-4.44455
95% upper limit (2-tailed)	5.70910
t-value	0.32014
df	5.00000
P-value (1-tailed)	0.38090
P-value (2-tailed)	0.76180

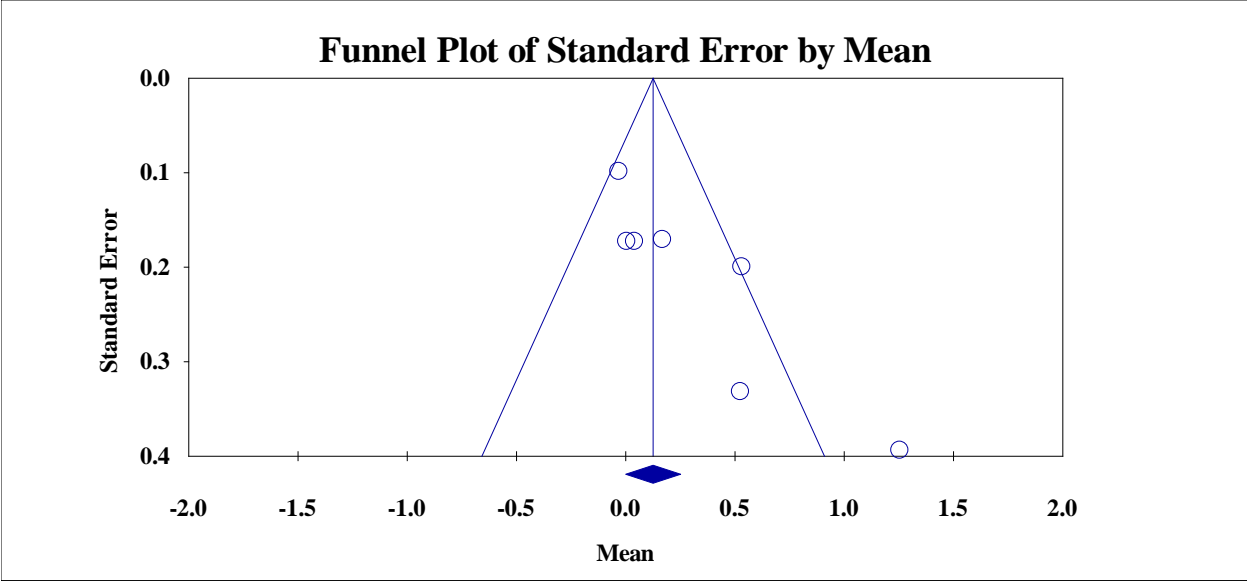
## Pooled Effect Sizes of HP, WT and SAN Interventions for Health Outcomes All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	7.000	0.239	**	0.121

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
17.092	6.000	0.009	64.895	0.060	0.059	0.003	0.246

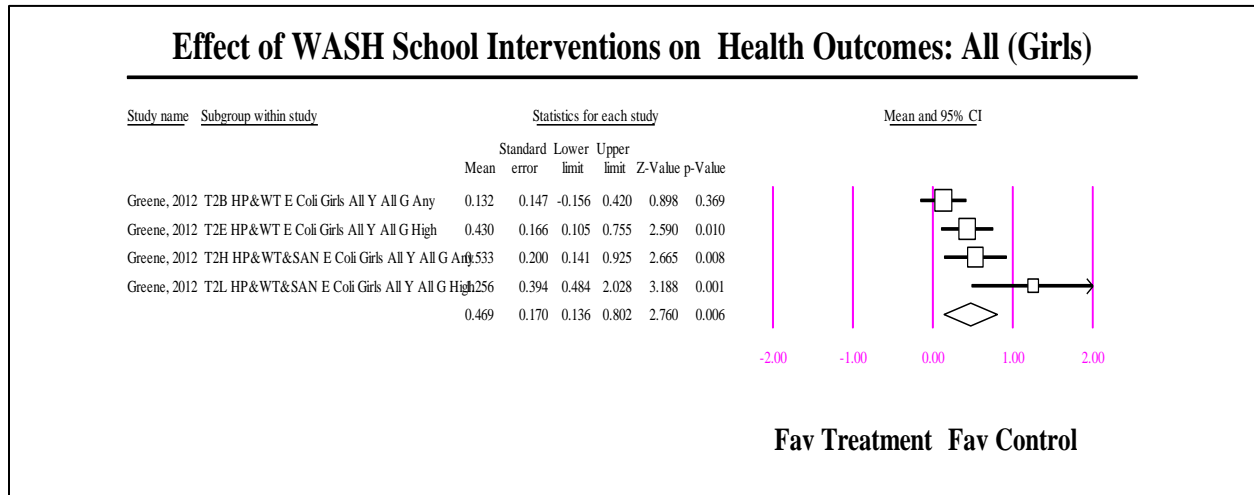




#### Egger's regression intercept

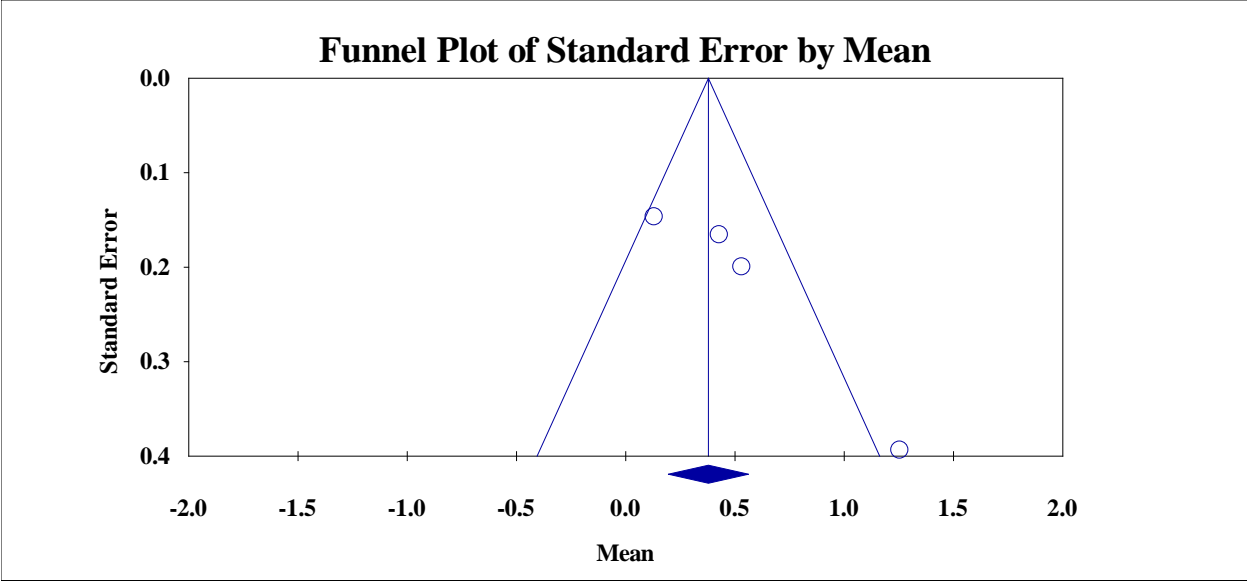
Intercept	3.45215
Standard error	0.92920
95% lower limit (2-tailed)	1.06357
95% upper limit (2-tailed)	5.84073
t-value	3.71519
df	5.00000
P-value (1-tailed)	0.00689
P-value (2-tailed)	0.01378

**Pooled Effect Sizes of All Interventions for Health Outcomes  
Girls only**



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	0.469	***	0.170

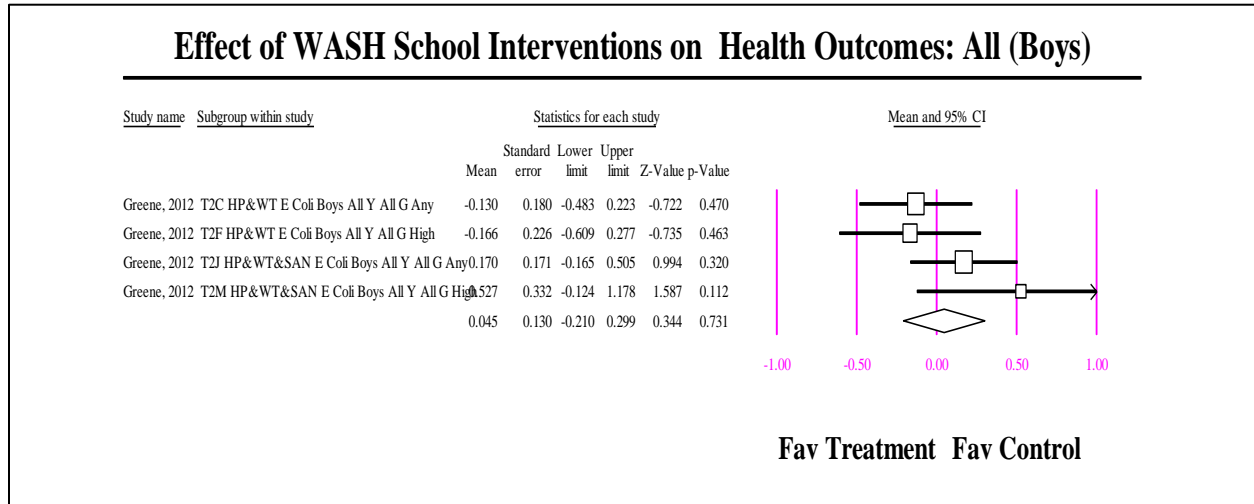
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
8.465	3.000	0.037	64.560	0.070	0.094	0.009	0.265



**Egger's regression intercept**

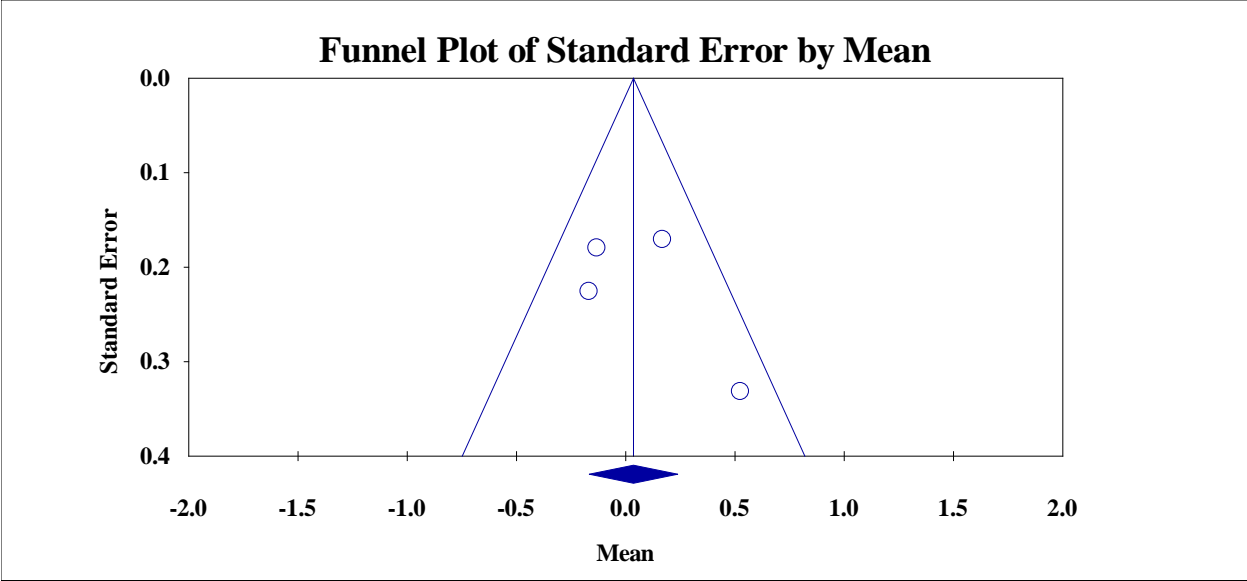
Intercept	4.51879
Standard error	1.21354
95% lower limit (2-tailed)	-0.70264
95% upper limit (2-tailed)	9.74022
t-value	3.72365
df	2.00000
P-value (1-tailed)	0.03258
P-value (2-tailed)	0.06515

## Pooled Effect Sizes of All WASH Interventions for Health Outcomes Boys Only



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	0.045	-	0.130

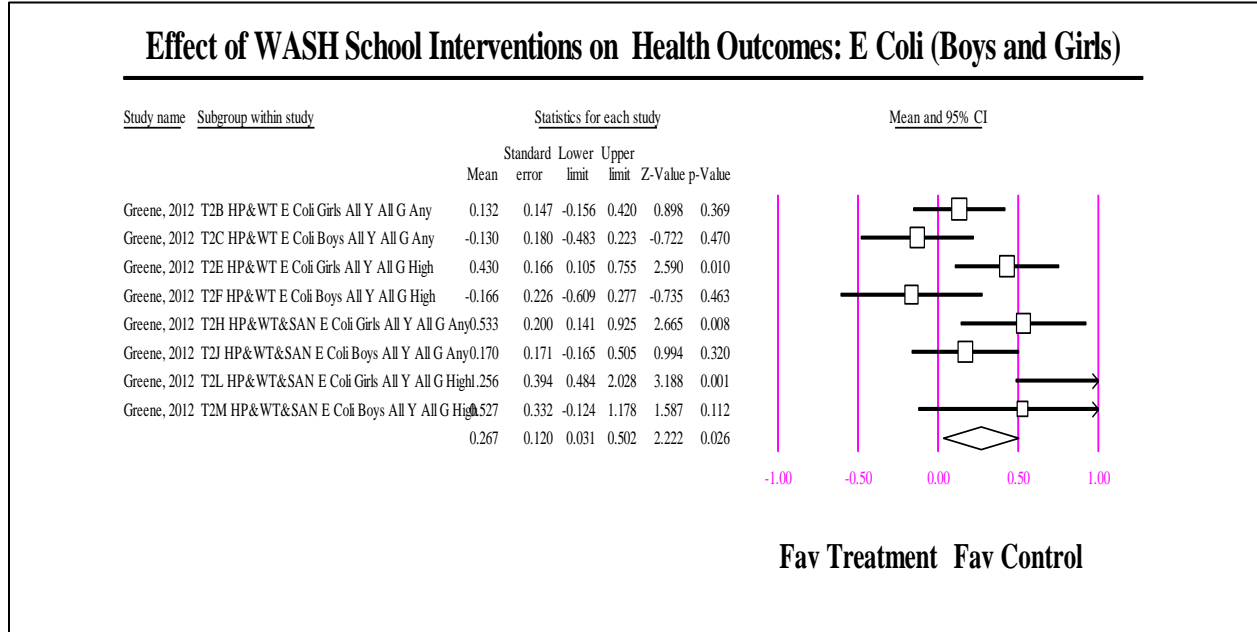
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
4.451	3.000	0.217	32.594	0.022	0.055	0.003	0.148



#### Egger's regression intercept

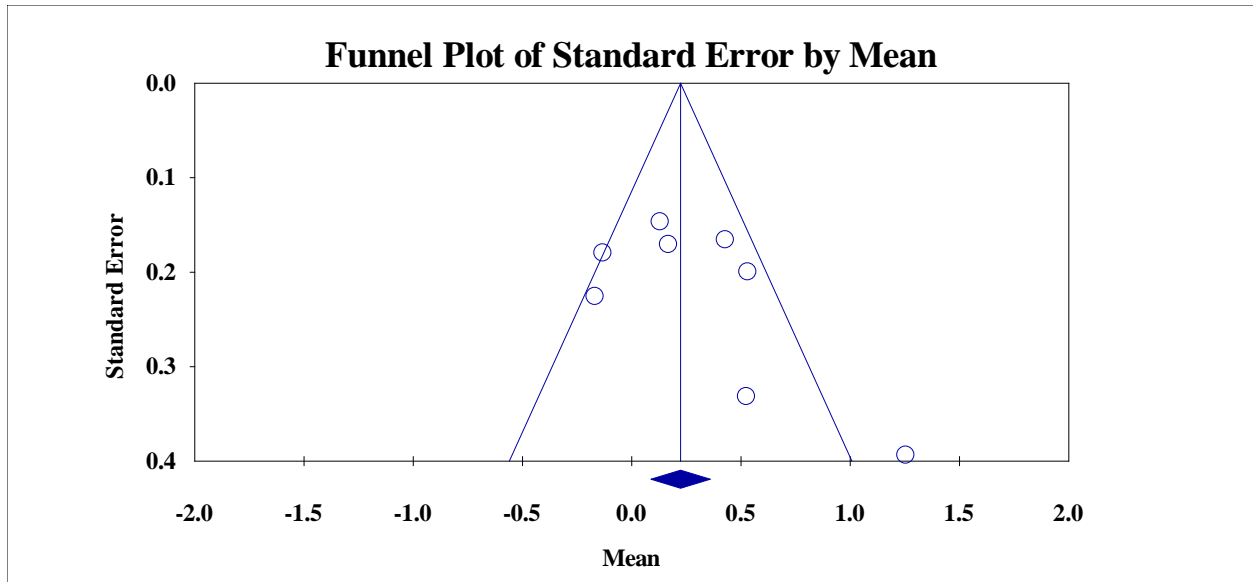
Intercept	2.14203
Standard error	2.86127
95% lower limit (2-tailed)	-10.16903
95% upper limit (2-tailed)	14.45309
t-value	0.74863
df	2.00000
P-value (1-tailed)	0.26607
P-value (2-tailed)	0.53215

## Pooled Effect Sizes of All WASH Interventions for E. coli All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	8.000	0.267	**	0.120

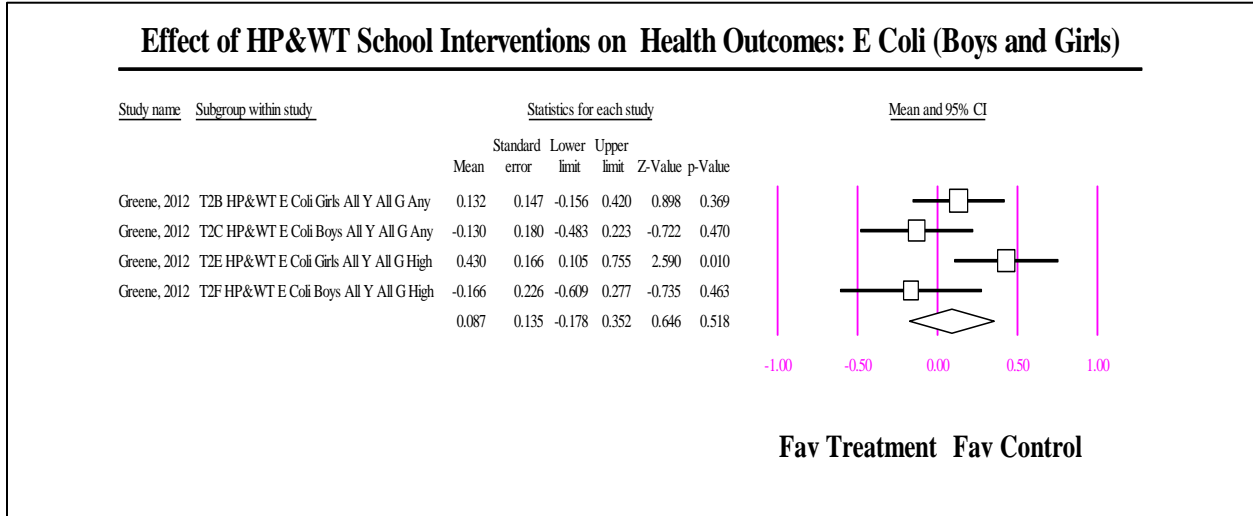
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
18.958	7.000	0.008	63.076	0.068	0.060	0.004	0.261



#### Egger's regression intercept

Intercept	2.82584
Standard error	1.98269
95% lower limit (2-tailed)	-2.02564
95% upper limit (2-tailed)	7.67731
t-value	1.42525
df	6.00000
P-value (1-tailed)	0.10198
P-value (2-tailed)	0.20396

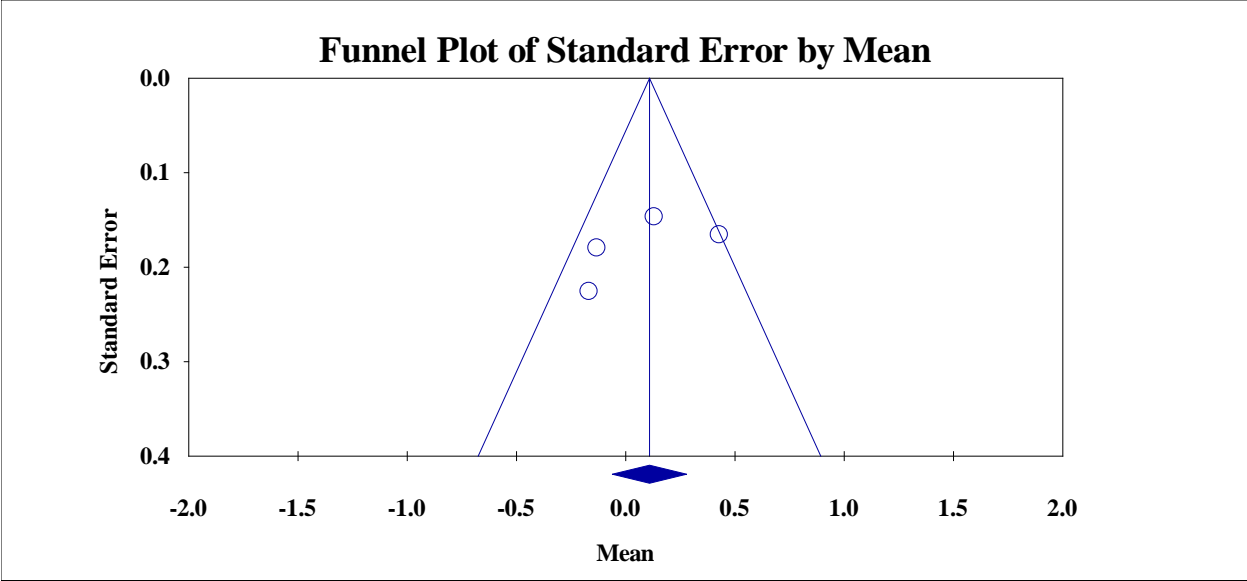
**Pooled Effect Sizes of HP and WT Interventions for E. coli  
All Children**



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	0.087	-	0.135

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
7.007	3.000	0.072	57.188	0.041	0.060	0.004	0.203

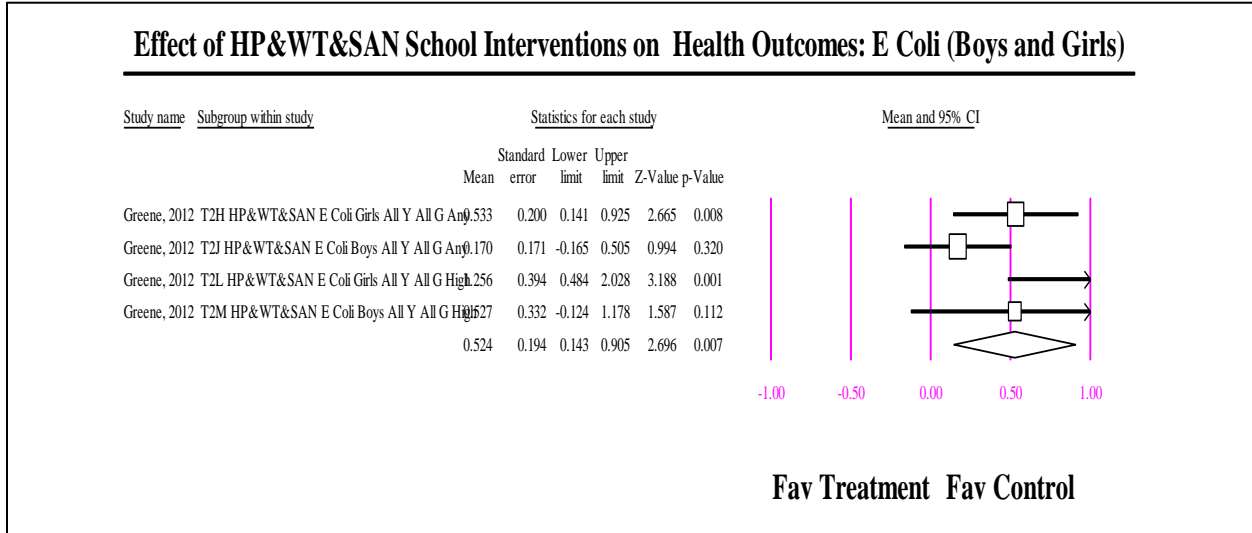




**Egger's regression intercept**

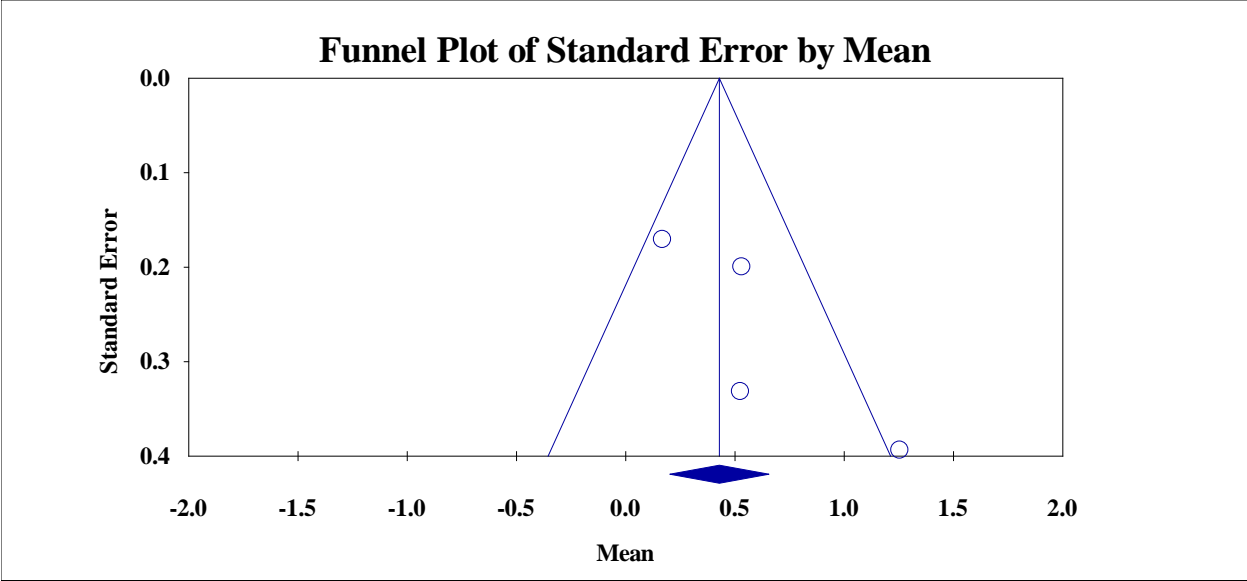
Intercept	-4.99988
Standard error	5.17242
95% lower limit (2-tailed)	-27.25502
95% upper limit (2-tailed)	17.25526
t-value	0.96664
df	2.00000
P-value (1-tailed)	0.21785
P-value (2-tailed)	0.43570

## Pooled Effect Sizes of HP, WT and SAN Interventions for E. coli All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	0.524	***	0.194

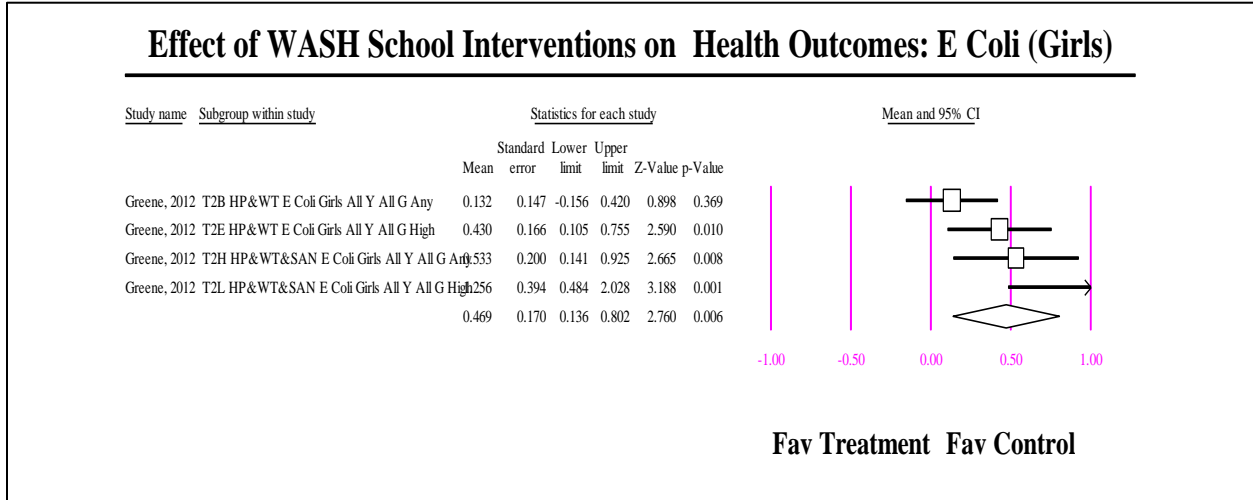
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
7.057	3.000	0.070	57.491	0.083	0.124	0.015	0.288



**Egger's regression intercept**

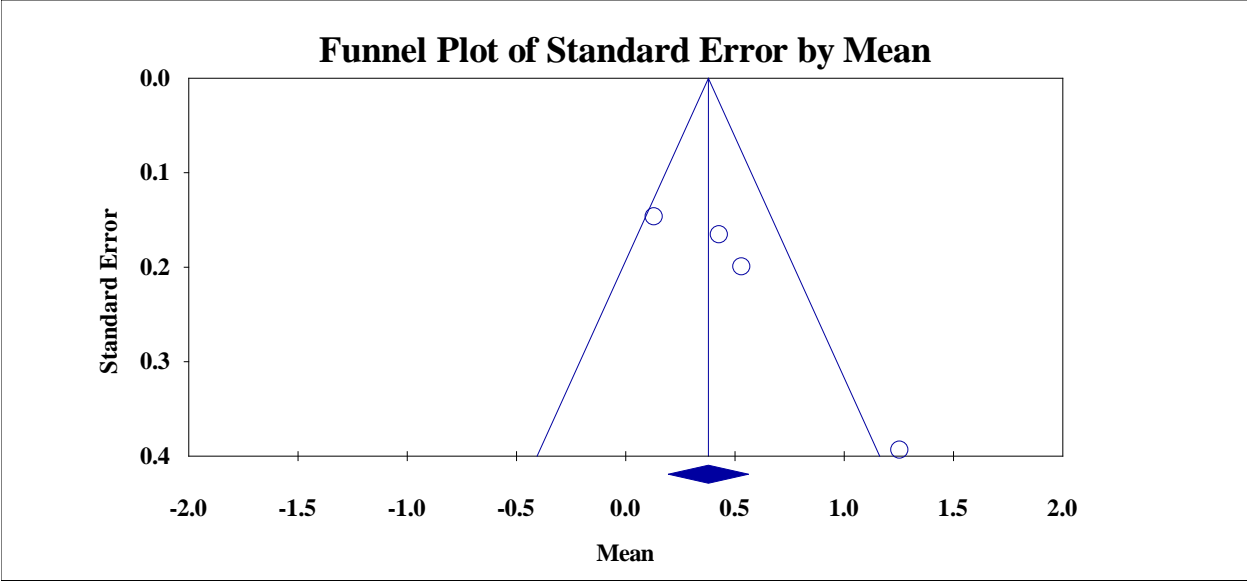
Intercept	3.51319
Standard error	1.62172
95% lower limit (2-tailed)	-3.46452
95% upper limit (2-tailed)	10.49091
t-value	2.16633
df	2.00000
P-value (1-tailed)	0.08132
P-value (2-tailed)	0.16264

## Pooled Effect Sizes of All WASH Interventions for E. coli Girls Only



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	0.469	***	0.170

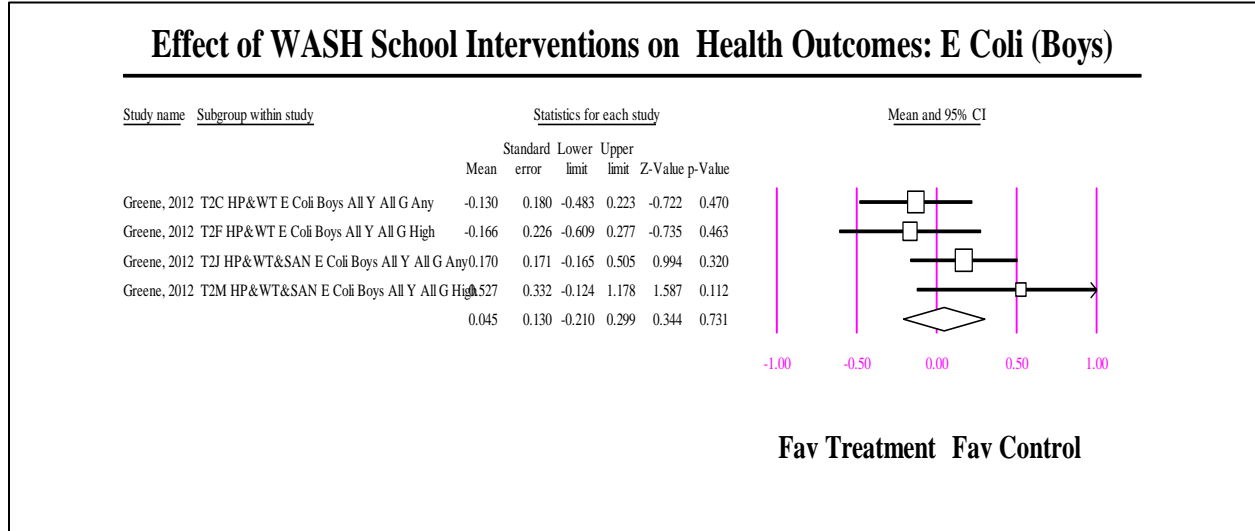
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
8.465	3.000	0.037	64.560	0.070	0.094	0.009	0.265



#### Egger's regression intercept

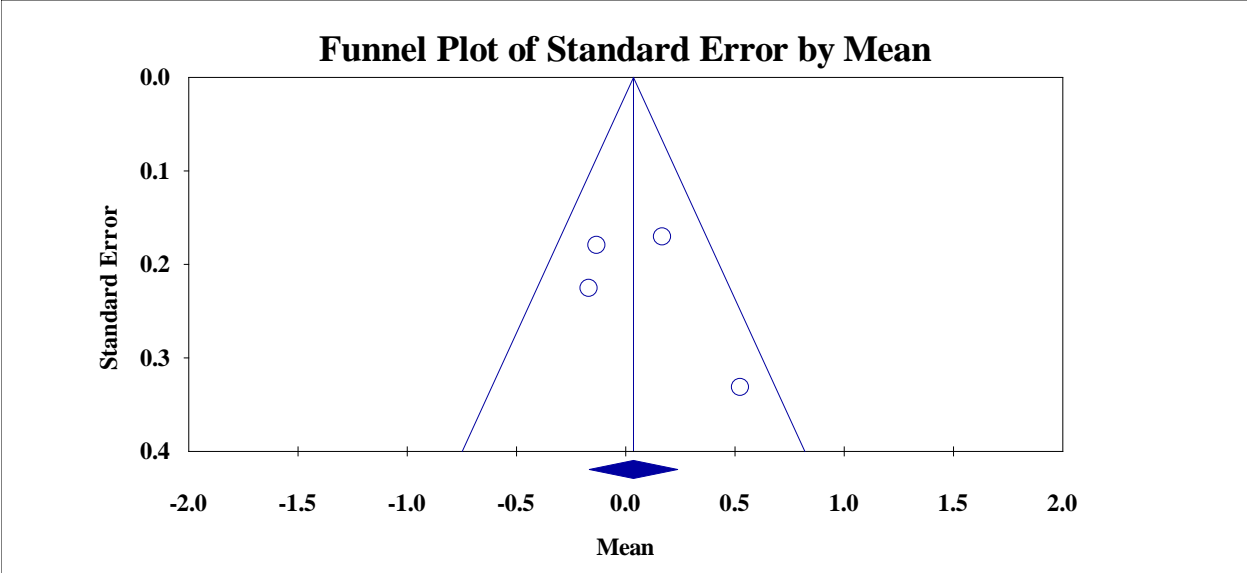
Intercept	4.51879
Standard error	1.21354
95% lower limit (2-tailed)	-0.70264
95% upper limit (2-tailed)	9.74022
t-value	3.72365
df	2.00000
P-value (1-tailed)	0.03258
P-value (2-tailed)	0.06515

## Pooled Effect Sizes of All WASH Interventions for E. coli Boy Only



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	0.045	-	0.130

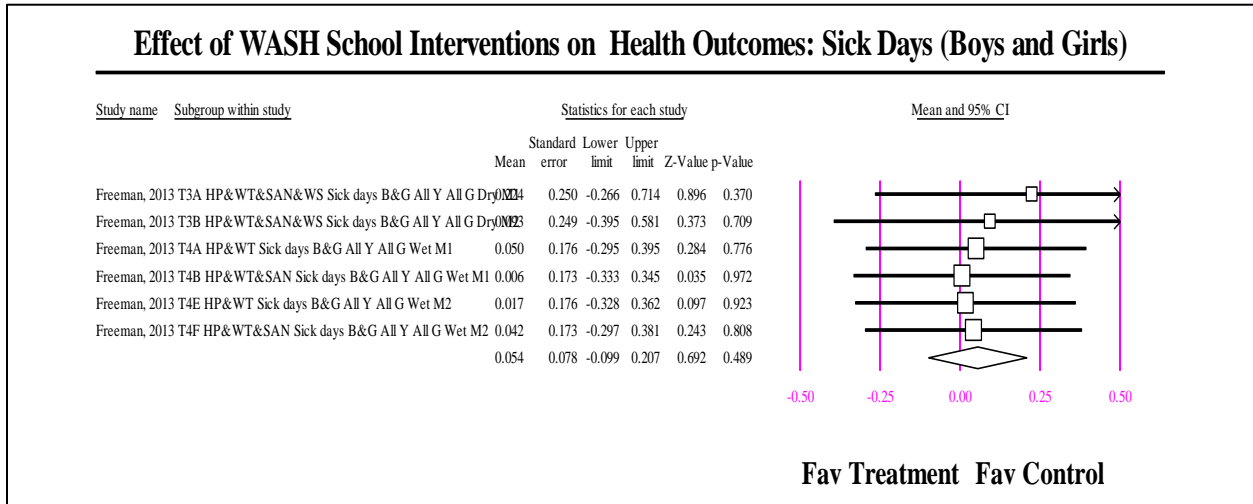
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
4.451	3.000	0.217	32.594	0.022	0.055	0.003	0.148



#### Egger's regression intercept

Intercept	2.14203
Standard error	2.86127
95% lower limit (2-tailed)	-10.16903
95% upper limit (2-tailed)	14.45309
t-value	0.74863
df	2.00000
P-value (1-tailed)	0.26607
P-value (2-tailed)	0.53215

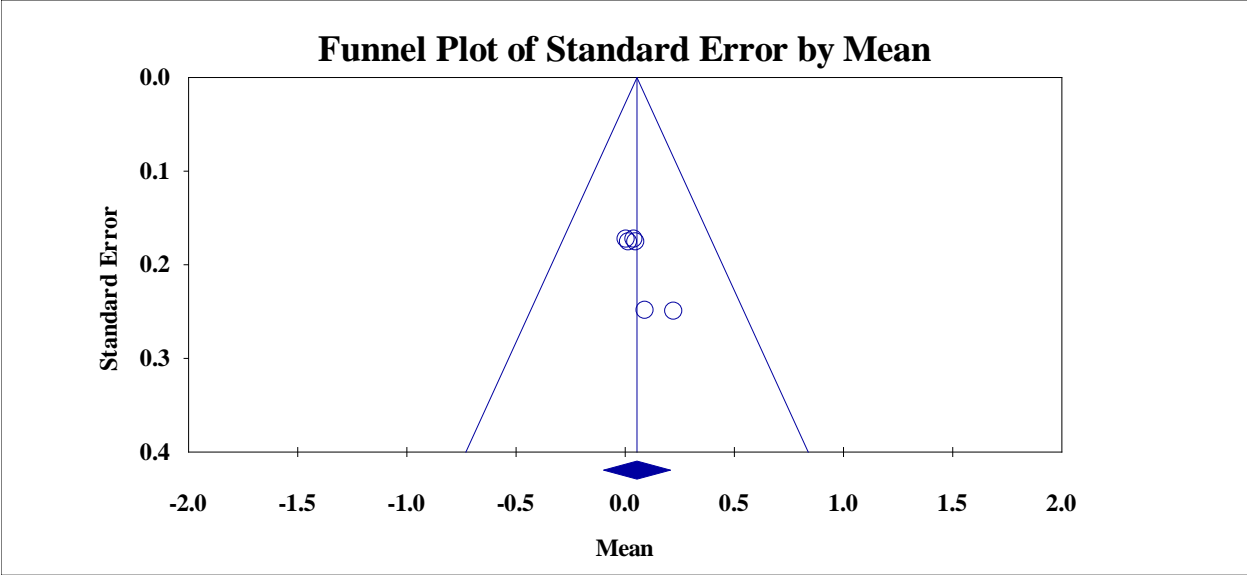
**Pooled Effect Sizes of All WASH Interventions for Sick Days  
All Children**



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	6.000	0.054	-	0.078

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
0.613	5.000	0.987	-	-	0.024	0.001	-

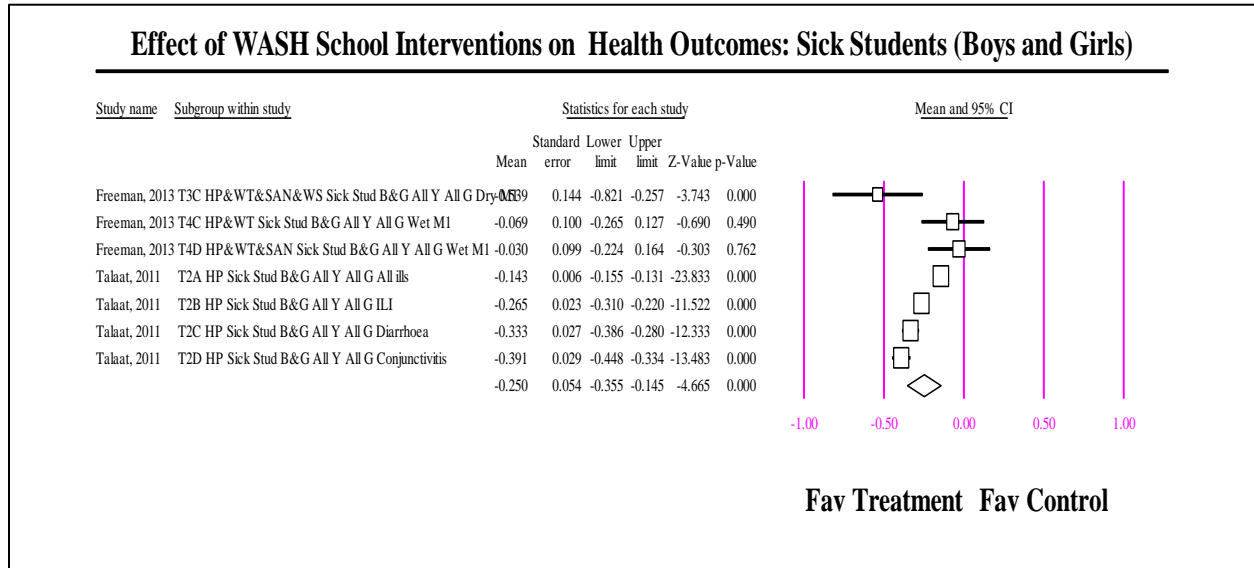




**Egger's regression intercept**

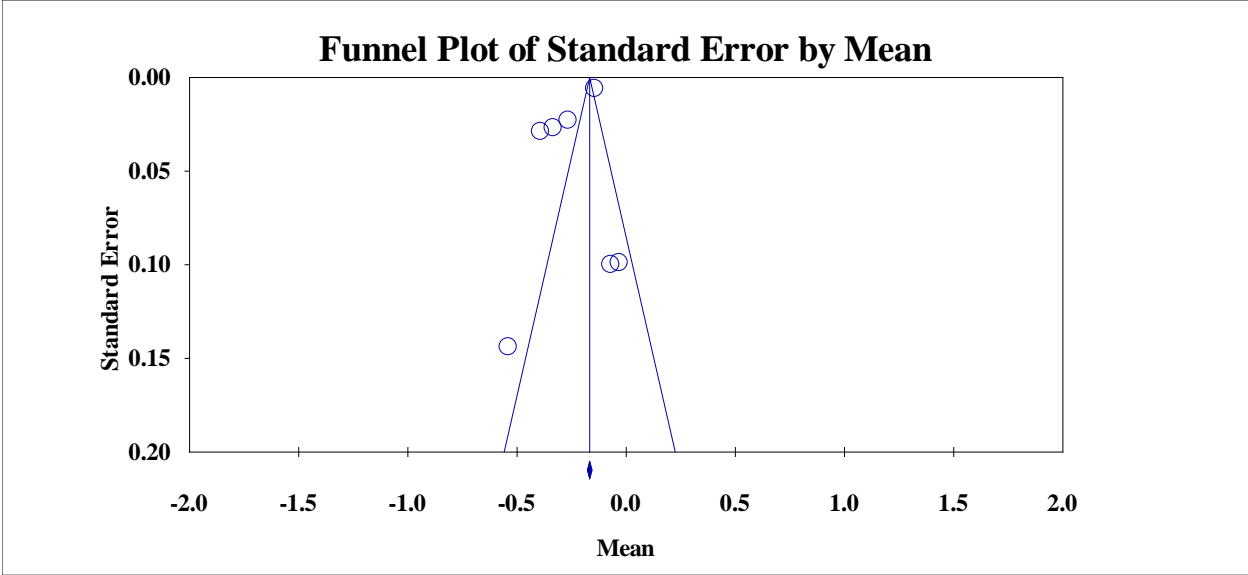
Intercept	1.73709
Standard error	0.54666
95% lower limit (2-tailed)	0.21933
95% upper limit (2-tailed)	3.25486
t-value	3.17766
df	4.00000
P-value (1-tailed)	0.01681
P-value (2-tailed)	0.03361

## Pooled Effect Sizes of All WASH Interventions for Sick Students All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	7.000	(0.250)	***	0.054

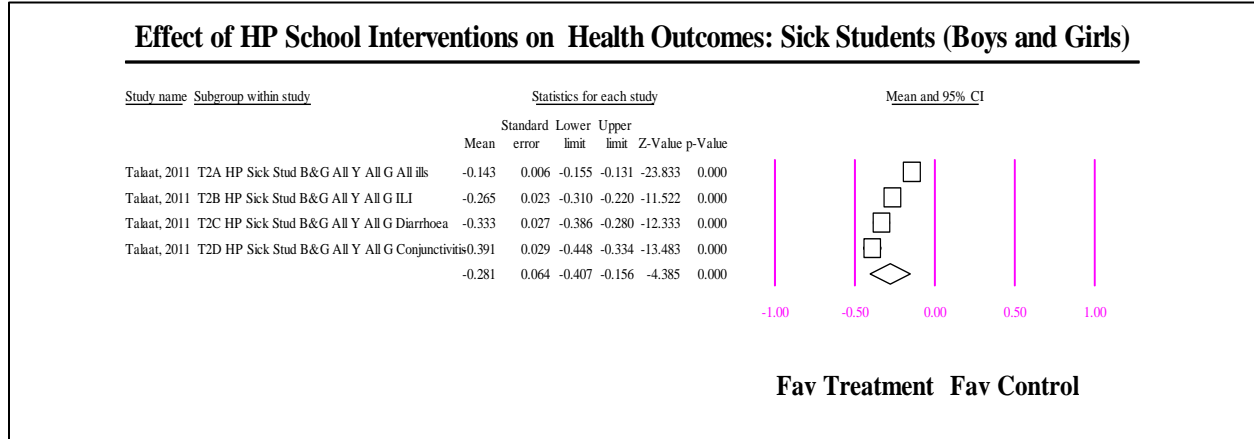
Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
141.165	6.000	-	95.750	0.016	0.015	0.000	0.126



#### Egger's regression intercept

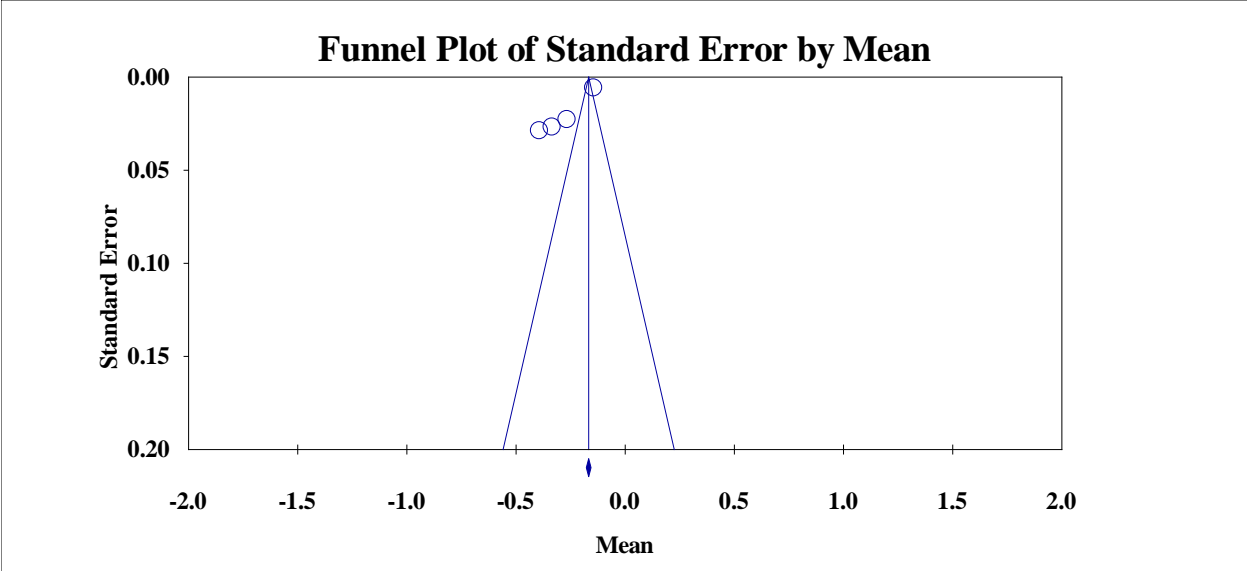
Intercept	-3.51422
Standard error	2.11522
95% lower limit (2-tailed)	-8.95155
95% upper limit (2-tailed)	1.92312
t-value	1.66140
df	5.00000
P-value (1-tailed)	0.07876
P-value (2-tailed)	0.15752

## Pooled Effect Sizes of HP Interventions for Sick Students All Children



Effect size and significance				
Model	Number Studies	Point estimate	Significance	Standard error
Random effects	4.000	(0.281)	***	0.064

Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
131.612	3.000	-	97.721	0.016	0.016	0.000	0.126



### Egger's regression intercept

Intercept	-9.11512
Standard error	1.11048
95% lower limit (2-tailed)	-13.89314
95% upper limit (2-tailed)	-4.33710
t-value	8.20825
df	2.00000
P-value (1-tailed)	0.00726
P-value (2-tailed)	0.01452