ACCURACY OF THE KATO-KATZ METHOD IN DIAGNOSIS OF *S. mansoni* AND SOIL TRANSMITTED HELMINTHS INFECTION IN ZIMBABWE

UNIVERSITY OF ZIMBABWE

COLLEGE OF HEALTH SCIENCES

DEPARTMENT OF COMMUNITY MEDICINE

BY

GEORGE NYANDORO

(R056730C)

MASTERS IN BIOSTATISTICS (COMMUNITY MEDICINE)
DECLARATION FORM

STUDENT:

I do hereby declare that this dissertation is the original work of GEORGE NYANDORO and has not been submitted before to the University of Zimbabwe or any other institution for the fulfilment of any degree requirements.

Name …………………………………………………………………………………………..

Signed …………………………………………………………………………………………..Date ..............................

SUPERVISOR:

I certify that I have supervised the writing of this dissertation and declare that it is indeed the original work of the student in whose name it is being submitted.

Name …………………………………………………………………………………………..

Signed …………………………………………………………………………………………..Date ..............................

DEPARTMENTAL CHAIRPERSON:

I do hereby declare all of the above statements to be true.

Name …………………………………………………………………………………………..

Signature …………………………………………………………………………………………..Date ..............................
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors, Professor S. Rusakaniko, Professor N. Midzi, Prof T. Mduluza and Mr V. Chikwasha for their valuable academic guidance throughout this project and also to Mr W Mkululi Wami Vundla, Dr Francisca Mutapi and Prof Woolhouse from University of Edinburg for their support towards modelling in winBUGs through their PhD student Welcome Wami Mkululi Vundla.

Last but not least, a special thank you to my fellow classmates for helping me to remain focussed on the reason we started this academic journey and my sincerest gratitude goes to my wife for soldiering on for so long without me.
ABSTRACT

Introduction: Mapping of schistosomiasis and soil transmitted helminths infection in Zimbabwe was prioritized for evidence based treatment control strategy in year 2010. Two diagnostics tests: Kato-Katz (KK) and formol ether concentration (FEC) were used during the mapping process.

Objective: This study aimed to determine the sensitivity and specificity of Kato-Katz technique in the absence of a gold standard using Bayesian modelling for the determination of S. mansoni and soil transmitted helminths infection (STHs) in Zimbabwe.

Methods: This is a secondary data analysis based on primary school children (n=15 818) aged 10-15 years who were enrolled in the national mapping of S. mansoni and STHs study, since they were reported to be the most exposed age group. Both parasitic infection by S. mansoni and STHs were diagnosed using a combination of two diagnostic techniques: the Kato Katz technique and the formol-ether concentration technique. A Bayesian approach was used to evaluate the diagnostic performance of the evaluation tool.

Results: The formol ether diagnostic technique was generally more sensitive with S. mansoni detection operational characteristics as follows (Sensitivity: 0.995; 95% Bayesian Credible Interval (BCI): 0.989 - 0.999), STHs- Hookworm detection (Sensitivity: 0.991; 95% BCI:0.988 - 0.993 ), STHs– A. lumbricoides detection (Sensitivity:0.992; 95% BCI;0.989 - 0.995) and STHs– T. trichiuria (Sensitivity:0.988 95% BCI: 0.926 - 1.00); than the Kato- Katz diagnostic technique (for S. mansoni detection (Sensitivity: 0.981; 95% BCI:0.971 - 0.994), STHs - Hookworm detection (Sensitivity:0.966; 95% BCI:0.963 - 0.970) , STHs – A. lumbricoides detection (Sensitivity: 0.967; 95% BCI;0.974-0.981) and STHs – T. trichiuria (Sensitivity:0.988; 95% BCI:0.974 - 0.981). However, specificity is higher for the Kato- Katz technique compared to the formol ether concentration technique.

Conclusion: The formol ether concentration technique has better sensitivity compared to the Kato-Katz technique, but however it has less specificity compared to the formol ether concentration technique. been shown for the Kato – Katz technique for all the infections respectively.
# TABLE OF CONTENTS

List of Tables.......................................................................................................................... vii
List of figures ............................................................................................................................. vii

## CHAPTER ONE......................................................................................................................... 1

1.1 TOPICAL AREA ..................................................................................................................... 1

1.2 INTRODUCING PRIMARY DATA ....................................................................................... 1

1.3 BACKGROUND TO THE ORIGINAL RESEARCH .............................................................. 2

1.4 THE PROBLEM STATEMENT OF THE ORIGINAL STUDY ................................................ 3

1.5 OBJECTIVES OF THE PRIMARY (ORIGINAL) STUDY ...................................................... 3

1.6 STUDY DESIGN OF THE ORIGINAL STUDY .................................................................... 4

1.7 ORIGINAL STUDY SAMPLE SIZE ...................................................................................... 4

1.8 SAMPLING METHODS IN ORIGINAL STUDY ................................................................... 4

1.9 DATA COLLECTION METHODS ......................................................................................... 5

1.10 MANAGEMENT OF DATA AND STATISTICAL ANALYSIS ................................................ 6

1.11 CRITICAL APPRAISAL ..................................................................................................... 6

1.11.1 PROBLEM STATEMENT & RESEARCH QUESTION ..................................................... 7

1.11.2 OBJECTIVES ............................................................................................................... 7

1.11.3 STUDY DESIGN .......................................................................................................... 7

1.11.4 SAMPLE SIZE ............................................................................................................. 7

1.11.5 SAMPLING PLAN ........................................................................................................ 8

1.11.6 DATA COLLECTION .................................................................................................... 8

1.11.7 DATA ANALYSIS ........................................................................................................ 8

1.11.8 PROBLEM STATEMENT ............................................................................................. 9

## CHAPTER TWO......................................................................................................................... 10

2.1 LITERATURE REVIEW ..................................................................................................... 10

2.2 RESEARCH QUESTION .................................................................................................... 12

2.3 JUSTIFICATION .............................................................................................................. 12

2.4 RESEARCH OBJECTIVES ............................................................................................. 13

## CHAPTER THREE................................................................................................................... 14
List of Tables

Table 1: Data dictionary for the variables selected for the secondary data analysis ....................... 14
Table 2: Evaluation of Kato Katz (KK) and Formal Ether concentration technique (FEC) in the absence of a gold standard .................................................................................................................. 16
Table 3: Schistosomiasis and STHS prevalence in the study population ........................................... 20
Table 4: Sensitivity and Specificity of the Kato-Katz and Formal Ether concentration technique ........ 21
Table 5: Positive predictive values and negative predictive values of S. mansoni using the formal ether technique ......................................................................................................................... 34
Table 6: Positive predictive values and negative predictive values of S. mansoni using the Kato-Katz technique ............................................................................................................................. 36

List of figures

Figure 1: Comparison of the area under curve for detecting Schistosomiasis and STHS between the methods ............................................................................................................................................. 22
Figure 2: Simulation of SEKK (Sensitivity of Kato-Katz) parameter ......................................................... 31
CHAPTER ONE

INTRODUCTION

1.1 TOPICAL AREA

Schistosomiasis is endemic in 76 countries world-wide (Chitsulo et al., 2000), with one hundred and ninety-three million people are infected annually. Of these 164 million (85%) live in the African continent (Chitsulo et al., 2000). Soil transmitted helminths (STHs) are nematodes transmitted from the soil to human. They include hookworms (Ancylostomes), round worms (Ascaris lumbricoides) and whip worms (Trichuris trichiura). About 400 million school age children world-wide are infected with STHs (Chen and Mott, 1989) and Schistosomiasis remains one of the top ten causes of outpatient attendances in Zimbabwe (Midzi et al., 2014). The extent and distribution dynamics of Schistosomiasis and STHs effects highlights the importance of diagnostic tools evaluation in terms of their diagnostic capabilities (sensitivity, specificity, Positive Predictive Value (PPV), and Negative predictive Value (NPV). There are generally two statistical evaluation approaches to these parameters; the Bayesian and Classical approaches. The authors of the primary study sought the expertise of a Biostatistician to evaluate the diagnostics capabilities of the diagnostic tools used in their survey. We explored the capabilities of these tools using the Bayesian approach recommended by literature, Dendukuri et al., 2001 and Zhou et al., 2002., Bayesian latent class models appropriately evaluate accuracy of diagnostic tests when the accuracy an imperfect gold standard.

1.2 INTRODUCING PRIMARY DATA

They are several parasitological diagnostics tests that are used to detect infection in human hosts; these methods enable the mapping of schistosomiasis and STHs infection distribution among populations. In the National mapping described by Midzi et al 2014, data was collected by determining infection status of the individuals using the Kato- Katz (KK) and Fomol Ether technique (FEC): the Kato-Katz technique involves microscopic examination of a fixed amount of faecal material to detect and count S. mansoni and STHs eggs. Full diagnosis of S.mansoni is completed by examination of three stool specimens on three consecutive days (WHO, 2002). However, STHs egg counts give an indirect measure of the worm burden. Alternative to the Kato Katz method is the formol-ether concentration technique (FECT) which is more sensitive method especially for detecting light infections. Individuals
were classified positive or negative using the two tests Kato-Katz and Fomol Ether Concentration techniques. Data was entered into the SPSS data base, the information of the individuals’ infection status and the specific diagnostic were recorded into three categories: Positive for KK, Positive for FEC and Positive for either KK or FEC and even positive for both tests. This information was important in order to meet the objectives of the primary study see section 1.5 for the objectives.

1.3 BACKGROUND TO THE ORIGINAL RESEARCH

According to the two national schistosomiasis surveys carried out in Zimbabwe, 1985 and 1992, the prevalence and epidemiological distribution of the disease has continued to increase (Midzi et al., 2014). However, there has not been any national control programme that includes mass treatment to reduce morbidity due to schistosomiasis (Midzi et al., 2014). As a result of the public health concern on the significance of schistosomiasis in Zimbabwe recently, the Ministry of Health and Child Care (MOHCC) included schistosomiasis control programme as a matter of high priority on the national agenda. Inspite of the occurrence of STHs as reported in some parts of the country, the possible overlap in epidemiological distribution of schistosomiasis and STHs in some areas and the associated public health impact, there has not been any documentation on the distribution of STHs at national level in Zimbabwe.

The life-time inter-provincial or district migration of people may also have caused transmission of schistosomiasis in previously non-endemic areas. Infected people could have continued to suffer from the morbidity associated with schistosomiasis and STHs, due to the scarcity of anti-helminthic and schistosomiasis drugs. The recent persistent public health problems related to schistosomiasis and STHs and the expiry of the patent on praziquantel (WHO 2002), made it possible for MOHCC to reconsider the proposal for the national schistosomiasis control programme but integrated with STHs.

A national survey was conducted in 2010 as a baseline to set benchmarks to regularly treat primary school aged children and the adult population at risk of infection in order to reduce the prevalence and morbidity due to schistosomiasis and soil transmitted helminths by at least 75% (WHO 2002), this was meant to contribute towards the improvement of school health.

A total of 315 primary schools across the country were selected for the national survey and 15 818 children were screened for *Schistosomiasis mansoni*, *STHs* using both the Kato Katz and Formal Ether
Concentration techniques. The national survey was used to stratify districts into treatment strategies for schistosomiasis and STHs according to WHO guidelines (WHO 2002).

The Kato Katz slides prepared in the field were examined within 2 hours of preparation to ensure hookworm eggs are quantified whilst viable. Any delay would result in egg digestion and disappearance.

### 1.4 THE PROBLEM STATEMENT OF THE ORIGINAL STUDY

The burden of schistosomiasis and STHs has been shown locally and elsewhere with primary school aged children being the most affected age group (Kjetland et al., 1996; Leutscher et al., 1997; Madden, 1899; Poggensee et al., 1998; Poggensee et al., 2000; Renaud et al., 1989; Van der Werf, 2003; WHO, 2003; and Wright et al., 1982). STHs impose more or less the same problems in the age group mentioned above. In Zimbabwe schistosomiasis remains one of the top ten causes of outpatient attendances (Taylor and Makura, 1985). The impact of schistosomiasis and STHs on school health, cognitive development, growth, and school attendance, on pregnancy outcomes and on the national economic development cannot be ignored. Integrated deworming (parasite control) was to be initiated in schools and communities where the parasites are prevalent. A national schistosomiasis and STHs control programme was therefore proposed in order to generate data required for proper planning for the control of morbidity associated with schistosomiasis and STHs in school age children and other high-risk groups in endemic communities to insignificant levels.

### 1.5 OBJECTIVES OF THE PRIMARY (ORIGINAL) STUDY

Study focus was to conduct a baseline survey in order to determine the distribution of schistosomiasis and STHs among the general population living in Zimbabwe and implement deworming strategies for the national control of schistosomiasis and STHs.

**Original Study Specific Objectives Were:**

1. To determine the distribution of schistosomiasis and STHs in the general population living in Zimbabwe for control by May 2010.
2. To estimate population requiring anti-helminthic drugs and to provide regular treatment of at least 75% children at risk of morbidity due to schistosomiasis and STHs infection by 2013.
3. To create a data-base of schistosomiasis and STHs that will be used to monitor success of de-worming activities during phase 2 (programme implementation) by June 2010.

4. To determine sensitivity, specificity, predictive values of both the Kato Katz and Formol ether techniques in field diagnosis of *S. mansoni* and STH.

### 1.6 STUDY DESIGN OF THE ORIGINAL STUDY

A cross-sectional survey was conducted to collect baseline data on the epidemiological distribution of schistosomiasis and STHs. This involved examination of urine and stool samples for parasites in selected school pupils from each district and administration of rapid schistosomiasis and school health questionnaires.

### 1.7 ORIGINAL STUDY SAMPLE SIZE

The national sample size of 15,818 was calculated from the enrolled primary school age population (N=2,490,568), the calculation of the national sample size was done in EPInfo version 6. The expected prevalence of 37% for schistosomiasis (Kjetland *et al.*, 2005) and the error margin of 0.075 were used.

A random proportional sampling of the target population was done. This involves weighting the sample population according to the size of the target population by province, district and randomly selecting the schools for baseline survey. (See Annex 1& 2 for detailed calculations of sample sizes by province and by district). On average the number of schools to be selected per district agrees with recommendations made by Lwanga and Lemeshow (1991).

### 1.8 SAMPLING METHODS IN ORIGINAL STUDY

The national sample size was proportionally distributed by district (taking into account of the district population denominator and numerator). The district sample was then divided by the number of children expected to be sampled per school (n=50 per school) to determine the number of schools per district. Schools in each district were selected using the lottery technique to achieve the required simple random sampling (Taylor and Makura, 1985); all schools were listed on small slips of paper and then slips were randomly picked after thoroughly mixing the small slips paper. Fifty primary school children per school, aged 10-15 years, were enrolled; only in a few cases were the number would be less than 50, would the researchers consider those aged 6-9 years (n=598) and >15 years (n=6). The selection of
participants was balanced between males and females this was done randomly to minimize selection bias done randomly (Chitsulo et al., 2000 and Kjetland et al., 2005).

1.9 DATA COLLECTION METHODS

Kato-Katz technique: The Kato Katz technique is a field based quantitative method used for the detection of intestinal helminths (S. mansoni and STH) infection. The test was used to test infection presence (positive or negative). In this method, the participant is asked to give a stool (faecal matter) in a plastic bottle measuring 50-100 ml in volume. The stool sample is sieved by straining it on nylon mesh screen measuring 70 µm pore size. The fine material that collects on the other side of the sieve is collected and used to fill in the Kato Katz template whole measuring 41.7 mg on a microscope slide. As the template is removed, the measured 41.7 mg fine stool remains on the slide. A thick (Kato-Katz) smear is prepared by covering the stool on the slide with malachite green soaked cellophane (30mm x 25mm size). The prepared slide is then examined under the microscope with a 10 x objective within 60 minutes of smear preparation for hookworms. The slide is allowed to stand for 24 hours clearing the background after which it examined again for S. mansoni (intestinal schistosomiasis), Trichuris trichiura and the Ascaris lumbricoides eggs (STH).

Interpretation of data using the Kato Katz method: A person is infected (referred to as positive for S. mansoni) if any S. mansoni eggs are visualized in the Kato Katz thick smear. The egg intensity is estimated as the number of eggs observed /gram stool by multiplying the number of eggs examined in the Kato thick smear by a factor of 24. The same procedure is done for each STH (hookworm, A. lumbricoides and T. trichiura)

The formol ether concentration technique: This is a qualitative method which also classifies infection status, either positive or negative, in which about a gram stool is dissolved in 7ml of 10% volume methanol in water. 4 ml of ether is added and the mixture is shaken briefly after which the suspension is sieved through a tea strainer and the filtrate is centrifuged at 3000 revolutions per minute. The supernatant is decanted and the deposit that remains in the centrifuge tube is transferred onto the microscope slide. This is covered with a microscope cover slip and examined under the microscope with a 10 x objective. Observation of eggs of S. mansoni eggs will indicate infection (positivity). No quantitative estimation of infection intensity can be done using this method.
**Combined Kato Katz and formal ether concentration technique results:** The purpose of using both the Kato Katz and the formal ether concentration technique is to improve the sensitivity of diagnosis of the intestinal helminths. Thus the results were combined as follows: If a participant had eggs examined in stool using either the Kato Katz or the formal ether concentration techniques of both, the individual was considered positive for *S. mansoni* or any of the STHs species. If there were no eggs observed using the Kato Katz and the formal ether concentration technique, the participant was considered negative for *S. mansoni* and STHs.

It should be noted that the use of both the Kato Katz and formal ether concentration technique in the diagnosis of *S. mansoni* and STHs was deliberately done in order to later explain on the sensitivity of the Kato Katz in field diagnosis of intestinal helminths. Under normal circumstances the diagnosis of intestinal helminths infection has been traditionally done using the Kato Katz which is recommended by the World Health organization (WHO 2002). This is regarded a field applicable technique in the sense that only a microscope and a trained laboratory technician is required whereas the formal ether regarded as non-field applicable method requires a centrifuge which is expensive to buy and the following additional reagents: formol and ether in addition to the requirement for the trained laboratory technician. Ether is transported under strict cool conditions to avoid explosion. However, the researcher wanted to demonstrate that the Kato Katz being recommended by WHO for field diagnosis of *S. mansoni* or STHs is not as sensitive as is expected for mapping such helminthic infections.

**1.10 MANAGEMENT OF DATA AND STATISTICAL ANALYSIS**

SPSS version 8 was used for both data entry and statistical analysis. The data was cleaned before analysis and frequencies were run for summary statistics. Except of objective four, objectives 1 to 3 were achieved by basic cross tabulations in SPSS that is differences in prevalence of infection among different groups by their demography, geographical locations and so on. The chi-square test was used to test for associations and the student t-test was used to determine differences between groups.

**1.11 CRITICAL APPRAISAL**
1.11.1 PROBLEM STATEMENT & RESEARCH QUESTION

The authors gave a clear focus of their research question which required them to determine the baseline burden and distribution of schistosomiasis and STH. The authors clearly stated and presented the disease burden; and evidently quoted statistics and references. They highlighted the public importance of the disease with risk factors (s. mansoni and STHs infection).

1.11.2 OBJECTIVES

The objectives of the original study are clearly stated, Specific, and Measurable in terms of measuring prevalences and mapping the distribution (for quantitative outcomes). The objectives were Achievable, Time bound and Realistic, Schistosomiasis and STHs mapping and baseline setting was implemented. All objectives had time lines except of objective 4 which is the utility of the Kato- Katz tool in the diagnosis of intestinal helminths. However, in order for them to demonstrate the requirements of objective 4 they needed expertise and skills from a Biostatistician which are availed through this dissertation.

1.11.3 STUDY DESIGN

The authors used a cross sectional study design to determine prevalence’s, and literature recommends the same design for prevalence studies.

1.11.4 SAMPLE SIZE

The sample size is adequate, had enough power (sample size power analysis was done in STATA 11, and it was above 84%) to reflect the diseases burden in the population as identified in the problem statement. A sample of 15 818 children (which allows determination of prevalence) was calculated using EPI Info 6 statistical package, the expected frequency of 37% for schistosomiasis (Kjetland et al., 2005) and the error margin of 0.75%. A random proportional sampling of the target population was done. This involves weighting the sample population according to the size of the target population by province, and district after which they had randomly selected for the baseline survey. On average, the number of schools selected per district was adequate and agrees with recommendations made by Ross, 1991 and Lwanga and Lemeshow, 1991.
1.11.5 SAMPLING PLAN

Nagelkerke et al., 2000, recommended a simple random sampling (SRS) of schools at district level for national surveillance of tuberculosis in school children. This agrees with what the authors have done in this survey. Each school within the district had the same probability of being included in the sample. In support to the authors sampling plan; literature show that increasing the number of selected districts is more efficient for increasing the precision of the estimate than increasing the number of children per district beyond several hundreds (Lwanga and Lemeshow, 1991).

The authors calculated the numbers of schools selected in each district proportionally according to population and sample size. Guyatt et al., 1999 and Monteresor et al 2002 recommended the selection of 50 grade three children (10-15 years) per school to be included in community surveillance of schistosomiasis and STHs. A probability sampling method was suitable in this study for the selection of schools in each district and here the authors used simple random sampling method to select the primary schools in each district.

1.11.6 DATA COLLECTION

The authors conducted a cross-sectional survey nationwide to collect baseline data on the epidemiological distribution of schistosomiasis and STHs. The data collected involved STHs and Schistosomiasis diagnosis examination of urine and stool samples for parasites in selected schools from each district and administration of rapid schistosomiasis and school health questionnaires. Urine and stool sample examination or diagnosis. The authors collected relevant data to answer to their objectives, detailed key variables selected for the secondary analysis are shown in section 3.5 see( Key variables)

The detailed laboratory diagnosis methods are described in the research background section 1.9.

1.11.7 DATA ANALYSIS

The authors were able to answer objectives 1-3 as they were able to map the distribution of schistosomiasis and STH including production of maps on point prevalence distribution. Their results...
were used to plan mass drug administration in Zimbabwe as outline in objective 2. However, a gap still exist where by one of their objective (objective 4) supposed to demonstrate the diagnostic performance of the Kato Katz in mapping intestinal helminths has not been fulfilled. This objective requires Statistician skill to analyze the available data in order to determine the sensitivity, specificity and predictive values of the Kato Katz compared with the gold standard (combined Kato Katz and the formal ether concentration technique).

1.11.8 PROBLEM STATEMENT

The authors in their manuscript (Midzi et al 2014) mentioned that combining the results of the KK and the FECT improved the sensitivity in diagnosis of intestinal helminths. However they simply used the prevalence values without elaborating on the sensitivity, specificity and PV which is the scope of objective number 4. A Bayesian approach to statistics best explains the objective 4, which is statistical inference based on the posterior distribution which expresses all that is known about the parameters of a statistical model, given the data and existing knowledge (Kerry, 2010) of the diagnostic parameters.
CHAPTER TWO

2.1 LITERATURE REVIEW

Generally diagnostic tests have been reported imperfect in most published articles reporting on sensitivity and specificity. Some articles highlighted that there is no a gold standard test for sensitivity and specificity in parasitology (Booth et al., 2003 and Santos et al., 2005). Researchers on STHs detection have demonstrated that ‘gold standard’ tests (with 100% accuracy) do not exist (Booth et al., 2003; Goodman et al., 2007; Santos et al., 2005).

Other studies on the accuracy of the Kato-Katz and formol ether technique in determining STH infection have reported sensitivity and/or specificity of the Kato-Katz and the Fomol Ether techniques for detecting STH infection and S. mansoni (Assefa et al., 2014; Booth et al., 2003; Coulibaly et al., 2011; Glinz et al., 2010; Goodman et al., 2007; Knopp et al., 2008; Santos et al., 2005; Steinmann et al., 2008; Utzinger et al., 2011).

Sensitivity and/or specificity of the Kato-Katz and Fomol Ether techniques have been estimated against other tests or a combination of tests, none of which is 100% accurate, ‘pseudo gold standard’ (Booth et al., 2003; Goodman et al., 2007; Joseph et al., 1995; Knopp et al., 2008; Santos et al., 2005; Steinmann et al., 2008).

Evaluation of diagnostic capabilities (sensitivity, specificity and Predictive values) of diagnostic tools is done in two different approaches; the classical or frequentist approach and the Bayesian approach. These two approaches mentioned above both view data as the observed realizations of stochastic processes with one or more random processes (Kerry, 2010). In the classical approach, the parameters are not known and fixed unlike to the Bayesian approach where they are considered as random processes (Kerry, 2010). The classical approach evaluates uncertainty in terms of frequency and inferences drawn from a single data set whilst the Bayesian approach evaluates uncertainty in terms of posterior distribution of the parameter (unknown quantities), the conditional probabilities of the parameters, given the data, the model and what is known about the parameters before analysis (Kerry, 2010).
Knopp et al., 2008 and other authors used a Bayesian approach to calculate sensitivity of the Kato-Katz (KK) for STH infections under a major assumption of perfect specificity for the KK test (Booth et al., 2003; Dendukuri and Joseph, 2001; Goodman et al., 2007; Joseph et al., 1995; Knopp et al., 2008; Nikolay et al., 2014; Santos et al., 2005; Steinmann et al., 2008).

Among other several authors who determined the sensitivity of the Kato-Katz in detecting hookworm infection in human hosts, Booth et al., 2003 used a Bayesian latent class model to estimate sensitivity (Joseph et al., 1995).

Several articles were using similar assumption; to carry out this estimation, the constraint of 100% specificity was imposed in the model (Booth et al., 2003; Dendukuri and Joseph, 2001; Goodman et al., 2007; Joseph et al., 1995; Knopp et al., 2008; Nikolay et al., 2014; Santos et al., 2005; Steinmann et al., 2008).

Utzinger et al., 2010 reported the sensitivity and specificity for diagnosis of S. mansoni for both the fomol ether-concentration technique with stool samples preserved for 40 days (sensitivity:85.0%) and triplicate Kato–Katz using fresh stool samples (sensitivity:77.4%). Whilst Nikolay et al., 2014 demonstrated the diagnostics test sensitivity and specificity variations across different parasitological techniques including Kato-Katz and Fomol Ether and even by infection intensity on detection of STH; A. lumbricoides, the 1-slide Kato-Katz method had a sensitivity of 48.8% (95% BCI: 37.6–58.2%) in the low intensity group compared with 95.8% (95% BCI: 91.8–98.5%) in the high intensity group.

Nikolay et al., 2014 estimated sensitivity for different diagnostic tests using Bayesian latent class model as described by Dendukuri and Joseph, 2001 and Branscum et al., 2005. The Bayesian Latent Class Analysis approach allows estimation of the sensitivity and specificity of imperfect diagnostic tests by assuming a probabilistic model for the relationship between unobserved, or latent, parameters: true disease prevalence and the sensitivity and specificity of different diagnostic methods.

New Algorithms and Computer knowledge have widely become more important in solving high-dimensional integrals: the denominator in Bayes rule (important for Bayesian approach estimation of sensitivity) was analytically intractable (Kerry, 2010), historically they could not be solved at all. The burden lessened with the coming of simulation based approaches like Markov Chain Monte Carlo (MCMC) computing algorithms (Kerry, 2010). The simulations draw samples from the posterior
distribution and inferences are drawn from MCMC analysis to ensure an equilibrium distribution, convergence of the chains (Kerry, 2010).

Epidemiologists in Cambridge developed a program called the BUGs (Bayesian analysis Using Gibbs Sampling) in the 1990s which was furthered into WinBUGs described by Spiegel et al., 2003.

2.2 RESEARCH QUESTION

What are the diagnostics capabilities; sensitivity, specificity and positive predictive values of the Kato-Katz and formol ether stool examination techniques?

2.3 JUSTIFICATION

Among other screening program factors, a suitable valid screening test should be available. The validity of a screening test is determined by its ability to correctly classify persons with preclinical disease as positive and those without as negative (Hennekens, 1987). Sensitivity is defined as the probability of testing positive if the disease is truly present (Hennekens, 1987). Specificity is defined as the probability of screening negative if the disease is truly absent (Hennekens, 1987). A higher sensitivity and specificity will ensure effectiveness in a screening program, thus reducing morbidity and mortality from the diseases under the screening program (Hennekens, 1987). A Bayesian approach seems more appropriate for this sort of problem as we can determine the sensitivity of the Kato-Katz and Formol Ether Concentration technique in the presence of an imperfect gold standard given the data, the information we know about the parameters (expert knowledge or published data which you cannot in a frequentist framework. A Bayesian inference approach is best to answer to the research question above (Kerry, 2010).

A Bayesian technique has been recommended for this scenario where we do not have a gold standard, and unobserved data in case of misclassifications which is our latent data (Booth et al., 2003; Dendukuri and Joseph, 2001; Goodman et al., 2007; Joseph et al., 1995; Knopp et al., 2008; Nikolay et al., 2014; Santos et al., 2005; Steinmann et al., 2008). The Bayesian technique considers the unknown parameters as random processes thus have no any fixed and constrain on parameters (Booth et al., 2003 and Kerry, 2010). In our estimations of the sensitivity and specificity of the Kato-Katz technique and formol ether stool examination technique to detect infection with *S. mansoni* and STHs (hookworm, *Ascaris and Trichuris*), we adapted the methods described by Joseph et al., 1995 into our context.
2.4 RESEARCH OBJECTIVES

The aim of the study was to model the diagnostic capabilities of Parasitological methods on schistosomiasis diagnostics

The specific objective

1. To determine the sensitivity, specificity, positive and negative predictive values of the Kato-Katz and Fomol Ether techniques using the Bayesian Latent Class modeling.
3  CHAPTER THREE

3.1  DATA SOURCE

The data set was accessed from the Medical Microbiology Department, the National Schistosomiasis and STHs Mapping Survey (2010-2011). For variables available in the data set refer to annex 2. We selected a few variables for secondary analysis as described below.

Table 1: Data dictionary for the variables selected for the secondary data analysis

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Type</th>
<th>Variable</th>
<th>Definition</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infection status</td>
<td>numeric</td>
<td>Smfect</td>
<td>S. mansoni infection status using Formal Ether Concentration Technique (FEC)</td>
<td>0=negative 1=positive</td>
</tr>
<tr>
<td>Infection status</td>
<td>numeric</td>
<td>cumsmst0</td>
<td>Cumulative S. mansoni infection status based on KK and FECT</td>
<td>0=negative 1=positive</td>
</tr>
<tr>
<td>Infection status</td>
<td>numeric</td>
<td>smstkk0</td>
<td>S. mansoni infection status based on Kato Katz Technique (KK)</td>
<td>0=negative 1=positive</td>
</tr>
<tr>
<td>Infection status</td>
<td>numeric</td>
<td>hwfec0</td>
<td>Hookworm infection status based on Formal Ether Concentration Technique (FEC)</td>
<td>0=negative 1=positive</td>
</tr>
<tr>
<td>Infection status</td>
<td>numeric</td>
<td>cuhw0</td>
<td>Cumulative Hookworm infection status based on KK and FEC techniques</td>
<td>0=negative 1=positive</td>
</tr>
<tr>
<td>Infection status</td>
<td>numeric</td>
<td>hwinkk0</td>
<td>Hookworm infection status using FEC</td>
<td>0=negative 1=positive</td>
</tr>
<tr>
<td>Infection status</td>
<td>numeric</td>
<td>cumal0</td>
<td>Cumulative Ascaris lumbricoides (Al) Al infection status based on KK and FEC</td>
<td>0=negative 1=positive</td>
</tr>
<tr>
<td>Infection status</td>
<td>numeric</td>
<td>alinkk0</td>
<td>Al Infection status based on KK</td>
<td>0=negative 1=positive</td>
</tr>
<tr>
<td>Infection status</td>
<td>numeric</td>
<td>alfect0</td>
<td>Al Infection status based on FEC</td>
<td>0=negative 1=positive</td>
</tr>
<tr>
<td>Infection status</td>
<td>numeric</td>
<td>cumtt0</td>
<td>Cumulative T. trichiura infection status based on KK and FEC techniques</td>
<td>0=negative 1=positive</td>
</tr>
<tr>
<td>Infection status</td>
<td>numeric</td>
<td>ttfect0</td>
<td>T. trichiura infection status based on FEC</td>
<td>0=negative 1=positive</td>
</tr>
<tr>
<td>Infection status</td>
<td>numeric</td>
<td>ttstkk0</td>
<td>Tt Infection status at baseline based on KK</td>
<td>0=negative 1=positive</td>
</tr>
</tbody>
</table>
3.2 SAMPLE SIZE

Calculation of the sample size for Bayesian estimation of sensitivity is described differently from the frequentist or classical approaches which rely on inferences from single data set frequencies (Kerry, 2010; Elham Rahme, 1997). Bayesian sample size calculation requires solving complex integrals which is best done by running Markov Chains MC (Elham Rahme, 1997).

According to Elham and Rahme, 1997. A minimum sample size to detect a sensitivity \( s = 0.955 \), specificity \( c = 1 \) and prevalence \( p = 0.723 \) at 95% confidence level, precision of 10% (\( w = 0.1 \)). Using the formulas below

\[
p = \theta s + (1 - \theta)(1 - c).
\]

\[
n_{adj} = \left( \frac{2Z_{\alpha/2}}{w(s + c - 1)} \right)^2 p(1 - p).
\]

A minimum sample size of 338 children would need to be screened S.mansonii and STHs infection to detect a prevalence of 72.3%.

3.3 DATA MANAGEMENT

Table 1 show the key variables which were selected for this analysis. The selected variables were scanned for values out of range in SPSS software (version 16) since the data set was in SPSS format. And there were no extreme values in the key variables. The key variables in the data set were correctly labeled, and coded into two categories (0= negative and 1= positive). The data was then exported to STATA 12 for generation of diagnostic performance evaluation 2x2 tables, a preparatory for WinBUGS software analysis.

3.4 THEORETICAL FRAMEWORK TO THE DATA ANALYSIS

In any diagnostic test there are two possible outcomes; Individuals are infected and the diagnosis is either positive or negative and Individuals are not infected and the diagnosis is either positive or negative. So in the absence of a gold standard, we need to estimate the sensitivity and specificity of the
test without constraints on the parameters hence Bay’s theorem. Using the Bayes theorem, we determined the posterior distribution from the prior distributions.

The Baysian Latent Class Analysis approach allows estimation of the sensitivity and specificity of imperfect diagnostic tests by assuming a probabilistic model for the relationship between five unobserved, or latent, parameters: true disease prevalence $\pi_k$ and the sensitivities $S_iS_i$, $S_jS_j$ and specificities $C_Ci$, $C_jC_j$ of diagnostic methods $i$ and $j$ (Pepe and Janes, 2007). The model additionally incorporates the covariance terms $covD_i^+covD_j^+$, $covD_i^-covD_j^-$ to account for conditional dependency between compared diagnostic tests amongst infected and non-infected individuals, which is necessary as the included diagnostic tests are based on the same biological principle (detection of $S. mansonii$ and STHs) and therefore factors other than the true infection status are likely to influence both test outcomes simultaneously (Dendukuri and Joseph, 2001).

Nikolay et al., 2014 went on to demonstrate that, the joint distribution of the results of a $2 \times 2$ table (see figure 1) follows a multinomial distribution,

**Table 2**: Evaluation of Kato Katz (KK) and Formal Ether concentration technique (FEC) in the absence of a gold standard

<table>
<thead>
<tr>
<th>KK or FEC technique</th>
<th>Positive</th>
<th>negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic technique</td>
<td>$X_{k++}$</td>
<td>$X_{k-}$</td>
</tr>
<tr>
<td></td>
<td>$X_{k+}$</td>
<td>$X_{k-}$</td>
</tr>
</tbody>
</table>

$(X_{k++}, X_{k-}, X_{k+}, X_{k-}) \sim \text{Multi}(p_{k++}, p_{k+}, p_{k-}, p_{k-}, N_k)(X_{k+}, X_{k-}, X_{k+}, X_{k-}) \sim \text{Multi}(p_{k++}, p_{k+}, p_{k-}, p_{k-}, N_k)$

With the multinomial probabilities calculated as follows:

$$p_{k++} = P(T_i^+, T_j^+ | k^{th} \text{ population})$$

$$= [S_iS_j + covD_i^+covD_j^+] \pi_k + [(1 - C_i)(1 - C_j) + covD_i^-covD_j^-](1 - \pi_k)_{pk++} = P(T_i^+, T_j^+ | k^{th} \text{ population})$$

$$= [S_iS_j + covD_i^+covD_j^+] \pi_k + [(1 - C_i)(1 - C_j) + covD_i^-covD_j^-](1 - \pi_k)$$
\[ p_{k+} = P(T_i^+, T_j^- | k^{\text{th}} \text{ population}) \]
\[ = [S_i(S_j - 1) - covD_{ij}^-] \pi_k + [(1 - C_i) C_j - covD_{ij}^-](1 - \pi_k)_{pk+} \]
=\(P(Ti^+,Tj^-|k^{th}\text{-}\text{population})\)
=\([Si(Sj-1)-covDij+]\pi_k+[(1-Ci)Cj-covDij-](1-\pi_k)\)

\[ p_{k+} = P(T_i^-, T_j^+ | k^{\text{th}} \text{ population}) \]
\[ = [(S_i - 1)S_j - covD_{ij}^+] \pi_k + [C_i(1 - C_j) - covD_{ij}^+](1 - \pi_k)_{pk+} \]
=\(P(Ti^-Tj^+|k^{th}\text{-}\text{population})\)
=\([(Si-1)Sj-covDij+]\pi_k+[Ci(1-Cj)-covDij-](1-\pi_k)\)

\[ p_{k-} = P(T_i^- , T_j^- | k^{\text{th}} \text{ population}) \]
\[ = [(S_i - 1)(S_j - 1) + covD_{ij}^+] \pi_k + [C_iC_j + covD_{ij}^+](1 - \pi_k)_{pk-} \]
=\(P(Ti^-Tj^-|k^{th}\text{-}\text{population})\)
=\([(Si-1)(Sj-1)+covDij+]\pi_k+[CiCj+covDij-](1-\pi_k)\)

Source: Nikolay et al., 2014

The equation uses the Bayes rules, thus we can calculate the sensitivity and specificity (posterior) given what we know about the data (priors). The beta distribution parameters (priors based on published prevalences, sensitivities, specificity) were obtained by using a software called Beta Buster.

3.5 DATA ANALYSIS

We used the Bayesian technique to estimate the sensitivity and specificity of the Kato-Katz and formol ether techniques to detect infection with *s. mansoni* and *STHs* which is described elsewhere (Joseph et al., 1995)

3.6 BAYESIAN TECHNIQUE

The Bayesian modeling technique was used to assess the accuracy of the two infection diagnostic techniques in the absence of a perfect gold standard diagnostic method. The priors were calculated using the Beta Buster software, free software and the 2×2 tables were generated in STATA 12 (see annex 3). Program Models were written and run in WinBUGs (see annex 2).
A Bayesian technique was used to estimate sensitivity and specificity of a diagnostic test in the absence of a ‘gold standard’ test and it does not have any constraints on the test parameters (Assefa et al., 2014). The objective of this analysis was to estimate the sensitivity and specificity of the Kato-Katz and formol ether stool examination techniques to detect infection with *S. mansoni* and STHs as described by Joseph et al., 1995.

The diagnostics tests $2 \times 2$ tables (see figure 1 and Annex for Stata log file) were generated in Stata 12 software, and later input data in vector form in winBUGs for Bayesian estimation.

Calculations for uninformative priors: sensitivity and underlying true prevalence (alpha and beta constants for the beta distribution shape with) were done using Beta Buster software described and provided elsewhere (Nikolay, 2014). The written program for this Bayesian analysis is not a new work but readily free published programs codes for modeling in winBUGs software. Estimation of sensitivity and specificity was done without any constraint to any parameters with room for covariance terms, to which a uniform prior distribution was assumed with limits as described by Dendukuri and Joseph, 2001.

The models were built separately for *S. mansoni*, Hookworm, *A. lumbricoides* and *T. trichiuria* in the WinBUGS software version 1.3 for 2 dependent tests, 1 population, no gold standard: kato-katz technique vs. Fomol ether concentration technique (Nikolay, 2014) and the following assumptions were made and checked for:

Assumptions (Nikolay, 2014);

1. The covariance terms, follow a uniform prior distribution
2. Specificity was assumed as a fixed term

### 3.7 CHECKING FOR ASSUMPTIONS

We were basing on the most parsimonious (best fitting model with low Deviance Information Criterion (DIC) which is was given the DIC tool, on the inference menu in winBUGs software. (Annex 1; for an example of model convergence checks, where simulations were done to ensure that the three chains which were run where converging and following same patterns).
3.8 ETHICS

Permission to use the data for secondary analysis was sought from the Principal Investigator, Professor N. Midzi, and Medical Microbiology department. See annex for both the MRCZ National Survey approval and the granted permission for secondary analysis.
4 CHAPTER FOUR

4.1 RESULTS

The table 3 below show the prevalences stratified by different general characteristics of the population under analysis. The average age was 11.20 ±1.38 years; there were 6,496 (49.23%) males and 6,698 (50.77%) females. The study was balanced in terms of gender.

Table 3: Schistosomiasis and STHS prevalence in the study population

<table>
<thead>
<tr>
<th>Prevalence category</th>
<th>Prevalence of combined schistosomiasis (95%CI)n</th>
<th>Prevalence of combined STHs(95%CI)n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>22.7(21.95-23.38)13165</td>
<td>5.5(5.13-5.94)12252</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td><strong>Prevalence</strong></td>
<td><strong>Prevalence</strong></td>
</tr>
<tr>
<td>Males</td>
<td>25.4(24.37-26.50)6482</td>
<td>5.8(5.20-6.40)6042</td>
</tr>
<tr>
<td>Females</td>
<td>20.0(19.0-20.96)6683</td>
<td>5.3(4.74-5.78)6210</td>
</tr>
<tr>
<td><strong>Rural based Province</strong></td>
<td><strong>Prevalence</strong></td>
<td><strong>Prevalence</strong></td>
</tr>
<tr>
<td>Manicaland</td>
<td>23.8(22.01-25.75)2051</td>
<td>4.4(3.54-5.40)1978</td>
</tr>
<tr>
<td>Mashonaland East</td>
<td>31.2(28.76-33.71)1388</td>
<td>18.3(16.19-20.52)1269</td>
</tr>
<tr>
<td>Mashonaland Central</td>
<td>39.3(36.32-42.35)1038</td>
<td>1.8(1.05-2.78)1018</td>
</tr>
<tr>
<td>Mashonaland West</td>
<td>22.8(20.54-25.21)1280</td>
<td>24(1.64-3.44)1238</td>
</tr>
<tr>
<td>Masvingo</td>
<td>37.0(34.48-38.69)2054</td>
<td>6.0(5.01-7.15)1995</td>
</tr>
<tr>
<td>Matebeleland North</td>
<td>3.8(2.69-5.11)1037</td>
<td>14.1(11.93-16.42)967</td>
</tr>
<tr>
<td>Matebeleland South</td>
<td>8.8(7.09-10.80)953</td>
<td>0.0(0.0-0.0)881</td>
</tr>
<tr>
<td>Midlands</td>
<td>30.4(27.67-33.31)1058</td>
<td>2.8(1.81-4.09)896</td>
</tr>
<tr>
<td><strong>Urban based (Metropolitan) Province</strong></td>
<td><strong>Prevalence</strong></td>
<td><strong>Prevalence</strong></td>
</tr>
<tr>
<td>Harare</td>
<td>9.6(7.98-11.45)1165</td>
<td>1.9(1.17-3.01)980</td>
</tr>
<tr>
<td>Bulawayo</td>
<td>3.3(2.24-4.75)871</td>
<td>0.5(0.13-1.25)815</td>
</tr>
</tbody>
</table>
Sensitivity and Specificity of the Kato – Katz and Fomol Ether Concentration technique

In this study, Formol Ether technique has shown a high sensitivity than specificity for detecting schistosomiasis - s. mansoni. Some STHs (Hookworm and A. lumbricoides) infection showed high sensitivities for the Formol Ether test. Hookworm and A. lumbricoides detection specificities did not differ with either technique except of s. mansoni, where the specificity was higher with the Kato-Katz, (shown in table 4 below).

Table 4: Sensitivity and Specificity of the Kato-Katz and Fomol Ether concentration technique

<table>
<thead>
<tr>
<th>Infection</th>
<th>Node</th>
<th>mean</th>
<th>Sd</th>
<th>95% Bayesian Credible Interval (BCI) (lower – upper)</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. mansoni detection</td>
<td>Sefec</td>
<td>0.995</td>
<td>0.002614</td>
<td>0.989 – 1.000</td>
<td>1100</td>
</tr>
<tr>
<td></td>
<td>Sekk</td>
<td>0.981</td>
<td>0.005674</td>
<td>0.971 – 0.994</td>
<td>1100</td>
</tr>
<tr>
<td></td>
<td>Spfec</td>
<td>0.849</td>
<td>0.06575</td>
<td>0.709 – 0.962</td>
<td>1100</td>
</tr>
<tr>
<td></td>
<td>Spkk</td>
<td>0.955</td>
<td>0.01647</td>
<td>0.917 – 0.981</td>
<td>1100</td>
</tr>
<tr>
<td>STHs – Hookworm</td>
<td>Sefec</td>
<td>0.991</td>
<td>0.001083</td>
<td>0.989 – 0.993</td>
<td>2000</td>
</tr>
<tr>
<td></td>
<td>Sekk</td>
<td>0.967</td>
<td>0.001785</td>
<td>0.963 – 0.970</td>
<td>2000</td>
</tr>
<tr>
<td></td>
<td>Spfec</td>
<td>0.900</td>
<td>0.005565</td>
<td>0.889 – 0.910</td>
<td>3000</td>
</tr>
<tr>
<td></td>
<td>Spkk</td>
<td>0.900</td>
<td>0.005833</td>
<td>0.888 – 0.911</td>
<td>3000</td>
</tr>
<tr>
<td>STHs – A. lumbricoides</td>
<td>Sefec</td>
<td>0.992</td>
<td>0.001364</td>
<td>0.989 – 0.995</td>
<td>3000</td>
</tr>
<tr>
<td></td>
<td>Sekk</td>
<td>0.977</td>
<td>0.001753</td>
<td>0.974 – 0.981</td>
<td>3000</td>
</tr>
<tr>
<td></td>
<td>Spfec</td>
<td>0.900</td>
<td>0.005643</td>
<td>0.888 – 0.910</td>
<td>3000</td>
</tr>
<tr>
<td></td>
<td>Spkk</td>
<td>0.900</td>
<td>0.005618</td>
<td>0.889 – 0.911</td>
<td>3000</td>
</tr>
<tr>
<td>STHs – T. trichiuria</td>
<td>Sefec</td>
<td>0.988</td>
<td>0.02136</td>
<td>0.926 – 1.000</td>
<td>3000</td>
</tr>
<tr>
<td></td>
<td>Sekk</td>
<td>0.988</td>
<td>0.02131</td>
<td>0.926 – 1.000</td>
<td>3000</td>
</tr>
<tr>
<td></td>
<td>Spfec</td>
<td>0.900</td>
<td>0.005902</td>
<td>0.888 – 0.911</td>
<td>3000</td>
</tr>
<tr>
<td></td>
<td>Spkk</td>
<td>0.900</td>
<td>0.005699</td>
<td>0.888 – 0.911</td>
<td>3000</td>
</tr>
</tbody>
</table>

Sekk-Sensitivity of the Kato- Katz technique, Spkk-Specificity of the Kato- Katz technique, Sefec-Sensitivity of the fomol ether concentration technique, Spfec-Specificity of the fomol ether concentration technique.

4.2 RECEIVER OPERATING CHARACTERISTIC CURVES (ROC)
Figure 1 below; illustrate the findings above using the classical approach to compare with the Bayesian latent approach findings. The classical approach is less rigorous and less robust it has pointed the difference in diagnostic capabilities between the two techniques (Kato-Katz versus Fomol Ether).

![Figure 1: Comparison of the area under curve for detecting Schistosomiasis and STHS between the methods](image)

The Fomol Ether diagnostic technique was generally more sensitive with S. mansoni detection operational characteristics as follows (Sensitivity: 0.995; 95% Bayesian Credible Interval (BCI): 0.989 - 0.999), STHs - Hookworm detection (Sensitivity: 0.991; 95% BIC: 0.988 - 0.993), STHs – A. lumbricoides detection (Sensitivity: 0.992; 95% BCI: 0.989 - 0.995) and STHs – T. trichiura detections.
(Sensitivity: 0.988; 95% BCI: 0.926 - 1.00); than the Kato- Katz diagnostic technique (for *S. mansoni* detection (Sensitivity: 0.981; 95% BCI: 0.971 - 0.994), STHs - Hookworm detection (Sensitivity: 0.966; 95% BCI: 0.963 - 0.970), STHs - *A. lumbricoides* detection (Sensitivity: 0.967; 95% BCI: 0.974 - 0.981) and STHs - *T. trichiura* (Sensitivity: 0.988; 95% BCI: 0.974 - 0.981). However, higher specificity has been shown for the Kato – Katz technique for all the infections respectively.

The Fomol Ether Technique has shown better performance through other operational characteristics like Positive Predictive Value (PPV = 0.508 for *S. mansoni* using fomol ether concentration technique compared to PPV = 0.468 for the Kato- Katz technique. However the other parameter which is Negative Predictive Value (NPV) remained constant see annex 1.
5 CHAPTER FIVE

5.1 DISCUSSION

The Kato Katz and formal ether concentration technique in the diagnosis of *S. mansoni* and STHs were deliberately used in pair in order to later explain on the sensitivity of the Kato Katz basing on literature reviews findings described in Midzi *et al.*, 2014

Basing on the evidence from our findings, the diagnosis of *S.mansoni* and STHs using the Kato – Katz technique alone would have left thousands of cases untreated and not mapped leading to under estimation of the countries burden of schistosomiasis and STHs. Traditionally the mapping of *S. mansoni* and STHs is done using the Kato Katz which is recommended by the World Health organization (WHO 2002). Although the Kato – Katz technique is regarded as field applicable technique our findings demonstrate that the Kato Katz which is recommended by WHO for field diagnosis of *S. mansoni* or STH is not as sensitive as the fomol ether concentration technique.

Fomol Ether Concentration technique was generally more sensitive compared to the Kato-Katz technique with *S. mansoni* detection Sensitivity: 0.995; 95% Bayesian Credible Interval (BCI): 0.989 - 0.999), *S.mansoni* has a high sensitivity with both the Kato-Katz and Fomol Ether Concentration technique. Generally the ability of the fomol ether technique against the recommended Kato-Katz technique cannot be ignored for control and mapping of *S.mansoni* and STHs.

Hookworms are detected better the fomol ether concentration technique compared to the Kato- Katz evidenced by Sensitivity of 0.991; 95% BIC; 0.988 - 0.993) compared to hookworm sensitivity of 0.966; 95% BCI; 0.963 - 0.970) . The knowledge of the performance of the diagnostic tools play a major role in control programs if proper mapping is to be done. The Kato-Katz alone would have underestimate the disease burden, for example *A. lumbricoides* detection had a Sensitivity of 0.992; 95% BCI; 0.989 - 0.995 using the fomol ether concentration technique compared to 0.967; 95% BCI; 0.974- 0.981 detected by the Kato-Katz technique and STHs – *T. trichiura* (Sensitivity:0.988 95% BCI: 0.926 - 1.00);
In our study the Kato- Katz diagnostic technique, *S. mansoni* sensitivity was 0.981; 95% BCI: 0.971 - 0.994), Whilst Utzinger *et al.*, 2010, reported the sensitivity of 77.4%). The differences may be due to different prevalence levels.

Our findings among STHs *A. lumbricoides* sensitivity was 0.967; 95% BCI: 0.974- 0.981) compared to 95.8% (95% BCI: 91.8–98.5%) in the high intensity group reported by Whilst Nikolay *et al.*, 2014 and we have almost similar sensitivity of the Kato- Katz technique on detecting *A. lumbricoides* (Nikolay *et al.*, 2014).

The, *T. trichiura* sensitivity was 0.988; 95% BCI; 0.974 - 0.981 for both the Kato-Katz and fomol ether concentration technique. However, in this study higher specificity has been shown for the Kato – Katz technique for all the infections respectively.

The Positive Predictive Value (PPV) was 0.508 for *S, mansoni* using fomol ether concentration technique compared to PPV =0.468 for the Kato- Katz technique. Which imply that one is more likely to be correctly diagnosed positive with fomol ether technique as compared to the Kato-Katz. However the other parameter which is Negative Predictive Value (NPV) remained constant which implies that the chances with both techniques of one to be diagnosed negative when the infection is not present are the same.

### 5.2 STUDY LIMITATIONS

The study may not be generalizable to different country settings where the nature of infection is varying from the Zimbabwean set up. And the study was targeting the most risk primary school aged between 10- 15 years. This may not explain the situation infection levels among adults, this analysis should have been stratified by age groups and weighted for populations by geographical populations (provinces) and it would have been clearer and help in the evidence based control program initiatives, however we did not consider this as it is outside the objectives of the study.

In our study we did not do a stratified analysis on by intensity levels for a fair comparison with other papers with results stratified by infection intensity (low or high intensity).
CHAPTER SIX

6.1 CONCLUSION

The Kato-Katz technique has been widely used to map S.mansoni and STHs in developing countries and there has been under estimations of the burden and extent which may have led to increased chances of undetected illness and preclinical diseases in communities. Screening for S.mansoni and STHs is better achieved with fomol ether concentration technique although its regarded not field friendly (Midzi et al 2014).

6.2 RECOMMENDATIONS

Policy makers and World Health Organization should consider and plan for the under estimation of the S.mansoni and STHs mapping. There is need for National Mapping using both the Kato-Katz and fomol ether techniques to estimate the unknown proportions missed by the Kato-Katz technique who are not on treatment. Globally, where Kato-Katz technique was used alone, there are some missed undetected infections which impose a huge threat in terms of morbidity and mortality. Those countries should consider remapping with both the Kato-Katz and Fomol Ether techniques.
6.3 REFERENCES


7 ANNEXE

7.1 ANNEXE 1

We have shown a few oscillations run as criteria for selecting the best “burn in”, this is where there is convergence of the MCM chains, see below

![Simulations of SEKK (Sensitivity of Kato-Katz) parameter](image)

**Figure 2:** Simulation of SEKK (Sensitivity of Kato-Katz) parameter

Comment: see the model chains have the same patterns and converging this is the best burn in. However the priors for Sekk were not varying much compared to the Sefect priors.
7.2 MODEL PROGRAM

The following is a WinBUGs program used to estimate sensitivity and specificity for:

7.2.1 S. mansoni

#S. mansoni detection for 2 dependent tests, 1 population, no gold standard: kato-katz technique (KK) vs. Fomol ether concentration technique (FEC)

model{
  x[1:4] ~ dmulti(p[1:4], n)
  p[1] <- pi*(Sekk*Sefec+covDp) + (1-pi)*((1-Spkk)*(1-Spfec)+covDn)
  ls <- (Sekk-1)*(1-Sefec)
  us <- min(Sekk,Sefec) - Sekk*Sefec
  lc <- (Spkk-1)*(1-Spfec)
  uc <- (Spkk*Spfec) - min(Spkk,Spfec)
  pi ~ dbeta(3.6291,5.8827)
  Sekk ~ dbeta(10.7821,1.5149)
  Spkk ~ dbeta(99.6983,6.1946)
  Sefec ~ dbeta(21.2019,2.0633)
  Spfec ~ dbeta(15.0342, 2.5594)
  covDn ~ dunif(lc, uc)
  covDp ~ dunif(ls, us)
  rhoD <- covDp / sqrt(Sekk*(1-Sekk)*Sefec*(1-Sefec))
  rhoDc <- covDn / sqrt(Spkk*(1-Spkk)*Spfec*(1-Spfec))
}

list(n=12062, x=c(11170,15,251,626))

list(pi=0.727, Sekk=0.71, Spkk=0.90, Sefec=0.76, Spfec=0.95)

7.2.2 STHs – Hookworm

#STHs<Hookworm detection for 2 dependent tests, 1 population, no gold standard: kato-katz technique (KK) vs. Fomol ether concentration technique (FEC)

model{
  x[1:4] ~ dmulti(p[1:4], n)
  p[1] <- pi*(Sekk*Sefec+covDp) + (1-pi)*((1-Spkk)*(1-Spfec)+covDn)
  ls <- (Sekk-1)*(1-Sefec)
  us <- min(Sekk,Sefec) - Sekk*Sefec
  lc <- (Spkk-1)*(1-Spfec)
uc <- min(Spkk,Spfec) - Spkk*Spfec
pi ~ dbeta(1.0675,1.0075)
Sekk ~ dbeta(2.3764,1.1529)
Spkk ~ dbeta(2414.9512,269.2169)
Sefec ~ dbeta(130.7069,15.4119)
Spfec ~ dbeta(2414.9512,269.2169)
covDn ~ dunif(lc, uc)
covDp ~ dunif(ls, us)
  rhoD <- covDp / sqrt(Sekk*(1-Sekk)*Sefec*(1-Sefec))
  rhoDc <- covDn / sqrt(Spkk*(1-Spkk)*Spfec*(1-Spfec))
}
list(n=12177, x=c(11597,3,290,102))

list(pi=0.324, Sekk=0.2602, Spkk=0.95, Sefec=0.8528, Spfec=0.95)

7.2.3 STHs – Ascaris lumbricoides

#STHs<-Ascaris lumbricoides detection for 2 dependent tests, 1 population, no gold standard: kato-katz technique (KK) vs. Fomol ether concentration technique (FEC)
model{
  x[1:4] ~ dmulti(p[1:4], n)
  p[1] <- pi*(Sekk*Sefec+covDp) + (1-pi)*((1-Spkk)*(1-Spfec)+covDn)
  ls <- (Sekk-1)*(1-Sefec)
  us <- min(Sekk,Sefec) - Sekk*Sefec
  lc <- (Spkk-1)*(1-Spfec)
  uc <- min(Spkk,Spfec) - Spkk*Spfec
  pi ~ dbeta(1.0675,1.0075)
  Sekk ~ dbeta(3.7574,1.3064)
  Spkk ~ dbeta(2414.9512,269.2169)
  Sefec ~ dbeta(42.5732,5.6192)
  Spfec ~ dbeta(2414.9512,269.2169)
covDn ~ dunif(lc, uc)
covDp ~ dunif(ls, us)
  rhoD <- covDp / sqrt(Sekk*(1-Sekk)*Sefec*(1-Sefec))
  rhoDc <- covDn / sqrt(Spkk*(1-Spkk)*Spfec*(1-Spfec))
}
list(n=11990, x=c(11693,1,179,117))

list(pi=0.247, Sekk=0.3953, Spkk=0.95, Sefec=0.7515, Spfec=0.95)
7.2.4 STHs – T. trichiura

#STHs<T trichiura detection for 2 dependent tests, 1 population, no gold standard: kato-katz technique (KK) vs. Fomol ether concentration technique (FEC)

model{
  x[1:4] ~ dmulti(p[1:4], n)
  p[1] <- pi*(Sekk*Sefec+covDp) + (1-pi)*((1-Spkk)*(1-Spfec)+covDn)
  ls <- (Sekk-1)*(1-Sefec)
  us <- min(Sekk,Sefec) - Sekk*Sefec
  lc <- (Spkk-1)*(1-Spfec)
  uc <- min(Spkk,Spfec) - Spkk*Spfec
  pi ~ dbeta(1.0675,1.0075)
  Sekk ~ dbeta(8.3045,1.8116)
  Spkk ~ dbeta(2414.9512,269.2169)
  Sefec ~ dbeta(5.3842,1.4871)
  Spfec ~ dbeta(2414.9512, 269.2169)
  covDn ~ dunif(lc, uc)
  covDp ~ dunif(ls, us)
  rhoD <- covDp / sqrt(Sekk*(1-Sekk)*Sefec*(1-Sefec))
  rhoDc <- covDn / sqrt(Spkk*(1-Spkk)*Spfec*(1-Spfec))
}

list(n=11992, x=c(11976,0,6,10))
list(pi=0.13, Sekk=0.625, Spkk=0.95, Sefec=50, Spfec=95)

1. Positive Predictive Values for KK versus FEC

**TABLE 5: PPV AND NPV OF S. mansoni USING THE FOMOL ETHER TECHNIQUE**

<table>
<thead>
<tr>
<th>Infection</th>
<th>node</th>
<th>mean</th>
<th>sd</th>
<th>MC error</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>Start</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. mansoni</td>
<td>LRN</td>
<td>0.003243</td>
<td>0.003186</td>
<td>1.015E-4</td>
<td>8.763E-5</td>
<td>0.002254</td>
<td>0.01182</td>
<td>1001</td>
<td>3000</td>
</tr>
<tr>
<td></td>
<td>LRN</td>
<td>0.003243</td>
<td>0.003186</td>
<td>1.015E-4</td>
<td>8.763E-5</td>
<td>0.002254</td>
<td>0.01182</td>
<td>1001</td>
<td>3000</td>
</tr>
<tr>
<td></td>
<td>LRP</td>
<td>24250.0</td>
<td>86580.0</td>
<td>1488.0</td>
<td>2323.0</td>
<td>9006.0</td>
<td>125700.0</td>
<td>1001</td>
<td>3000</td>
</tr>
<tr>
<td>NPV</td>
<td>1.0</td>
<td>6.703E-7</td>
<td>2.167E-8</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1001</td>
<td>3000</td>
<td></td>
</tr>
<tr>
<td>PPV</td>
<td>0.4682</td>
<td>0.2771</td>
<td>0.008008</td>
<td>0.0287</td>
<td>0.455</td>
<td>0.9552</td>
<td>1001</td>
<td>3000</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>--------</td>
<td>----------</td>
<td>--------</td>
<td>-------</td>
<td>--------</td>
<td>------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>Pi</td>
<td>1.233E-4</td>
<td>1.077E-4</td>
<td>3.19E-6</td>
<td>4.6E-6</td>
<td>9.596E-5</td>
<td>4.154E-4</td>
<td>1001</td>
<td>3000</td>
<td></td>
</tr>
<tr>
<td>Se</td>
<td>0.9968</td>
<td>0.003186</td>
<td>1.015E-4</td>
<td>0.9882</td>
<td>0.9977</td>
<td>0.9999</td>
<td>1001</td>
<td>3000</td>
<td></td>
</tr>
<tr>
<td>Sp</td>
<td>0.9999</td>
<td>1.136E-4</td>
<td>3.393E-6</td>
<td>0.9996</td>
<td>0.9999</td>
<td>1.0</td>
<td>1001</td>
<td>3000</td>
<td></td>
</tr>
</tbody>
</table>

### 7.3 PROGRAM FOR CALCULATING PPV AND NPV AND OTHER PARAMETERS

```r
model{
  x ~ dbin(AP, n)
  AP <- pi*se + (1-pi)*(1-sp)
  se ~ dbeta(ase,bse)
  sp ~ dbeta(asp,bsp)
  pi ~ dbeta(api,bpi)
  PPV <- pi*se/AP
  NPV <- (1-pi)*sp/(1-AP)
  LRP <- se/(1-sp)
  LRN <- (1-se)/sp
}

DAT
list(n=12062,x=1,ase=298.0731175,bse=1,asp=20.636834805,bsp=1.198351676,api=1.0412,bpi=2.3315)
## Use BetaBuster to obtain the following results (ase,...)
## Mode=1, 95% sure se > 0.99
## Mode=0.99, 95% sure sp > 0.85
## Mode=0.05, 95% sure pi < 0.723
## Initial values
list(pi=0.05,se=1,sp=0.85)
```
# Positive predictive value (PPV) = The proportion of all individuals with a positive test who actually have the disease
# Negative predictive value (NPV) = The proportion of all individuals with a negative test who do not have the disease
# The likelihood ratio for a positive test result (LR+):
# LR+ = sensitivity / (1-specificity)
# The likelihood ratio for a negative test result (LR-):
# LR- = (1-sensitivity) / specificity

Table 6: Positive predictive values and negative predictive values of *S. mansoni* using the Kato-Katz technique

<table>
<thead>
<tr>
<th>node</th>
<th>Mean</th>
<th>sd</th>
<th>MC error</th>
<th>2.5%</th>
<th>97.5%</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>2.49E-4</td>
<td>1.427E-4</td>
<td>3.714E-6</td>
<td>2.226E-4</td>
<td>1001</td>
<td>3000</td>
</tr>
<tr>
<td>LRN</td>
<td>0.006425</td>
<td>0.006335</td>
<td>2.19E-4</td>
<td>0.004385</td>
<td>1001</td>
<td>3000</td>
</tr>
<tr>
<td>LRP</td>
<td>40780.0</td>
<td>244800.0</td>
<td>4648.0</td>
<td>10260.0</td>
<td>1001</td>
<td>3000</td>
</tr>
<tr>
<td>NPV</td>
<td>1.0</td>
<td>1.216E-6</td>
<td>3.362E-8</td>
<td>1.0</td>
<td>1001</td>
<td>3000</td>
</tr>
<tr>
<td>PPV</td>
<td>0.5084</td>
<td>0.283</td>
<td>0.007049</td>
<td>0.5105</td>
<td>1001</td>
<td>3000</td>
</tr>
<tr>
<td>Pi</td>
<td>1.263E-4</td>
<td>1.061E-4</td>
<td>2.769E-6</td>
<td>9.894E-5</td>
<td>1001</td>
<td>3000</td>
</tr>
<tr>
<td>Se</td>
<td>0.9936</td>
<td>0.006334</td>
<td>2.189E-4</td>
<td>0.9956</td>
<td>1001</td>
<td>3000</td>
</tr>
<tr>
<td>Sp</td>
<td>0.9999</td>
<td>1.094E-4</td>
<td>2.973E-6</td>
<td>0.9999</td>
<td>1001</td>
<td>3000</td>
</tr>
</tbody>
</table>

7.4 PROGRAM FOR CALCULATION OF PPV AND NPV AND OTHER PARAMETERS

model{
  x ~ dbin(AP, n)
  AP <- pi*se + (1-pi)*(1-sp)
  se ~ dbeta(ase,bse)
  sp ~ dbeta(asp,bsp)
  pi ~ dbeta(api,bpi)
  PPV <- pi*se/AP
  NPV <- (1-pi)*sp/(1-AP)
LRP <- se/(1-sp)
LRN <- (1-se)/sp

}

DAT

list(n=12062,x=1,ase=148.2838368,bse=1,asp=73.385204813,bsp=1.0,api=1.0412,bpi=2.3315)
## Use BetaBuster to obtain the following results (ase,...)
## Mode=1, 95% sure se > 0.98
## Mode=0.99, 95% sure sp > 0.96
## Mode=0.05, 95% sure pi < 0.723
## Initial values

list(pi=0.05,se=1,sp=0.99)

# Positive predictive value (PPV) = The proportion of all individuals with a
# positive test who actually have the disease
# Negative predictive value (NPV) = The proportion of all individuals with a
# negative test who do not have the disease
# The likelihood ratio for a positive test result (LR+):
# LR+ = sensitivity/(1-specificity)
# The likelihood ratio for a negative test result (LR-):
# LR- = (1-sensitivity)/specificity
8.1 STATA LOG FILE

name: <George>
log: E:\George-1\model basics\Evaluation of KK in the absence of a gold standard.log
log type: text
opened on: 30 Aug 2014, 12:56:48

.codebook cuhw0 smstkk0 smfect cuhw0 hwfect0 hwinkk0 cumtt0 alinkk0 alfect0 cumtt0 ttstkk0 ttfect0

---

**cuhw0**

Cumulative Hookworm infection status based on KK and FECT techniques at baseline

<table>
<thead>
<tr>
<th>type: numeric (double)</th>
</tr>
</thead>
<tbody>
<tr>
<td>label: cuhw0</td>
</tr>
<tr>
<td>range: [0,1]</td>
</tr>
<tr>
<td>units: 1</td>
</tr>
<tr>
<td>unique values: 2</td>
</tr>
<tr>
<td>missing .: 943/13195</td>
</tr>
</tbody>
</table>

**tabulation:**

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Numeric</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>11858</td>
<td>0</td>
<td>Negative</td>
</tr>
<tr>
<td>394</td>
<td>1</td>
<td>Positive (Either KK or FECT or both = 1)</td>
</tr>
<tr>
<td>943</td>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>

**smstkk0**

*S. mansoni* infection status based on KK

<table>
<thead>
<tr>
<th>type: numeric (double)</th>
</tr>
</thead>
<tbody>
<tr>
<td>label: smstkk0</td>
</tr>
<tr>
<td>range: [0,1]</td>
</tr>
<tr>
<td>units: 1</td>
</tr>
<tr>
<td>unique values: 2</td>
</tr>
<tr>
<td>missing .: 1133/13195</td>
</tr>
</tbody>
</table>

**tabulation:**

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Numeric</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>11421</td>
<td>0</td>
<td>negative</td>
</tr>
<tr>
<td>641</td>
<td>1</td>
<td>Positive</td>
</tr>
<tr>
<td>1133</td>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>
**smfect**  
*S. mansoni* infection status using formal ether concentration technique

- **type:** numeric (double)  
- **label:** smfect  
- **range:** [0,1]  
- **units:** 1  
- **unique values:** 2  
- **missing:** 1035/13195

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Numeric</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>11488</td>
<td>0</td>
<td>Negative</td>
</tr>
<tr>
<td>672</td>
<td>1</td>
<td>Positive</td>
</tr>
<tr>
<td>1035</td>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>

**cuhw0**  
Cumulative Hookworm infection status based on KK and FECT techniques at baseline

- **type:** numeric (double)  
- **label:** cuhw0  
- **range:** [0,1]  
- **units:** 1  
- **unique values:** 2  
- **missing:** 943/13195

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Numeric</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>11858</td>
<td>0</td>
<td>Negative</td>
</tr>
<tr>
<td>394</td>
<td>1</td>
<td>Positive (Either KK or FECT or both = 1)</td>
</tr>
<tr>
<td>943</td>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>

**hwfect0**  
Hookworm infection status based on formal ether concentration technique at baseline

- **type:** numeric (double)  
- **label:** hwfect0  
- **range:** [0,1]  
- **units:** 1  
- **unique values:** 2  
- **missing:** 1017/13195

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Numeric</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>11841</td>
<td>0</td>
<td>Negative (No hookworm ova detected)</td>
</tr>
<tr>
<td>337</td>
<td>1</td>
<td>Positive (Hookworm ova detected)</td>
</tr>
<tr>
<td>1017</td>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>
hwinkk0  
Hookworm infection status based on KK technique at baseline

type: numeric (double) 
label: hwinkk0

range: [0,1]  units: 1
unique values: 2  missing :: 1203/13195

tabulation: Freq  Numeric  Label
11887  0  Negative
105    1  Positive
1203   .

cumtt0  
Cumulative T. trichiura infection status based on KK and FECT techniques

type: numeric (double) 
label: cumtt0

range: [0,1]  units: 1
unique values: 2  missing :: 943/13195

tabulation: Freq  Numeric  Label
12236  0  Negative
16     1  Positive
943    .

alinkk0  
Al infection status based on KK at baseline

type: numeric (double) 
label: alinkk0

range: [0,1]  units: 1
unique values: 2  missing :: 1205/13195

tabulation: Freq  Numeric  Label
11872  0  Negative
118    1  Positive
1205   .
### alfct0
*Ascaris lumbricoides* infection status at baseline using FECT

<table>
<thead>
<tr>
<th>Type</th>
<th>Numeric (double)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>alfct0</td>
</tr>
<tr>
<td>Range</td>
<td>[0,1]</td>
</tr>
<tr>
<td>Units</td>
<td>1</td>
</tr>
<tr>
<td>Unique Values</td>
<td>2</td>
</tr>
<tr>
<td>Missing</td>
<td>1016/13195</td>
</tr>
</tbody>
</table>

**Tabulation**

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Numeric</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>11949</td>
<td>0</td>
<td>Negative</td>
</tr>
<tr>
<td>230</td>
<td>1</td>
<td>Positive</td>
</tr>
<tr>
<td>1016</td>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>

### cumtt0
Cumulative *T. trichiura* infection status based on KK and FECT techniques

<table>
<thead>
<tr>
<th>Type</th>
<th>Numeric (double)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>cumtt0</td>
</tr>
<tr>
<td>Range</td>
<td>[0,1]</td>
</tr>
<tr>
<td>Units</td>
<td>1</td>
</tr>
<tr>
<td>Unique Values</td>
<td>2</td>
</tr>
<tr>
<td>Missing</td>
<td>943/13195</td>
</tr>
</tbody>
</table>

**Tabulation**

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Numeric</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>12236</td>
<td>0</td>
<td>Negative</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>Positive</td>
</tr>
<tr>
<td>943</td>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>

### ttstkk0
*Tt* Infection status at baseline based on KK

<table>
<thead>
<tr>
<th>Type</th>
<th>Numeric (double)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>ttstkk0</td>
</tr>
<tr>
<td>Range</td>
<td>[0,1]</td>
</tr>
<tr>
<td>Units</td>
<td>1</td>
</tr>
<tr>
<td>Unique Values</td>
<td>2</td>
</tr>
<tr>
<td>Missing</td>
<td>1203/13195</td>
</tr>
</tbody>
</table>

**Tabulation**

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Numeric</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>11982</td>
<td>0</td>
<td>Negative</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>Positive</td>
</tr>
<tr>
<td>1203</td>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>
ttfect0  
*T. trichiura* infection status based on FECT at baseline

- type: numeric (double)
- label: ttfect0
- range: [0,1]  
- units: 1
- unique values: 2  
- missing :: 1016/13195

<table>
<thead>
<tr>
<th>Tabulation</th>
<th>Freq.</th>
<th>Numeric</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>12171</td>
<td>0</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>1016</td>
<td>.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- help diagtest
- codebook cumsmst0

---

cumsmst0  
Cumulative *S. mansoni* infection status based on KK and FECT at baseline

- type: numeric (double)
- label: cumsmst0
- range: [0,1]  
- units: 1
- unique values: 2  
- missing :: 946/13195

<table>
<thead>
<tr>
<th>Tabulation</th>
<th>Freq.</th>
<th>Numeric</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>11368</td>
<td>0</td>
<td>Negative (Both KK and FECT=0)</td>
<td></td>
</tr>
<tr>
<td>881</td>
<td>1</td>
<td>Either KK or FECT or both =1</td>
<td></td>
</tr>
<tr>
<td>946</td>
<td>.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- diagtest smstkk0 cumsmst0, all

<table>
<thead>
<tr>
<th align="left">S. mansoni</th>
<th align="left">Cumulative S.</th>
</tr>
</thead>
<tbody>
<tr>
<td align="left">infection</td>
<td align="left">mansoni infection</td>
</tr>
<tr>
<td align="left">status</td>
<td align="left">status based on KK</td>
</tr>
<tr>
<td align="left">based on</td>
<td align="left">and FECT at baseline</td>
</tr>
<tr>
<td align="left">KK</td>
<td align="left">Negative</td>
</tr>
<tr>
<td align="left">------------</td>
<td align="left">-----------------</td>
</tr>
<tr>
<td align="left">negative</td>
<td align="left">11,170</td>
</tr>
<tr>
<td align="left">Positive</td>
<td align="left">15</td>
</tr>
<tr>
<td align="left">Total</td>
<td align="left">11,185</td>
</tr>
</tbody>
</table>
Pearson chi2(1) = 8.2e+03  Pr = 0.000  
likelihood-ratio chi2(1) = 3.7e+03  Pr = 0.000  
Cramér's V = 0.8247  
gamma = 0.9989  ASE = 0.000  
Kendall's tau-b = 0.8247  ASE = 0.010

True D defined as cumsmst0 =~ 0  [95% Conf. Inter.]

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pr(+</td>
<td>D)</td>
<td>Pr(-</td>
<td>~D)</td>
</tr>
<tr>
<td></td>
<td>71.38%</td>
<td>99.87%</td>
<td>97.66%</td>
<td>97.80%</td>
</tr>
<tr>
<td></td>
<td>70.57%</td>
<td>99.80%</td>
<td>97.39%</td>
<td>97.54%</td>
</tr>
<tr>
<td></td>
<td>72.19%</td>
<td>99.93%</td>
<td>97.93%</td>
<td>98.06%</td>
</tr>
</tbody>
</table>

Prevalence  Pr(D)  7.27%  6.81%  7.73%

. diagtest smfect cumsmst0,all

S. mansoni  |
infection  |
status  |
using  |
   Formal  | Cumulative S. Ether  |
mansoni  infection  |
Concentrat  | status based on KK Concentrat  |
ion  | and FECTat baseline  |
Technique  | Negative  Either KK  | Total  |
---+----------------------|
Negative  | 11,273  215  11,488 |
Positive  | 8  664  672  |
---+----------------------|
Total  | 11,281  879  12,160 |

Pearson chi2(1) = 8.9e+03  Pr = 0.000  
likelihood-ratio chi2(1) = 4.1e+03  Pr = 0.000  
Cramér's V = 0.8553  
gamma = 0.9995  ASE = 0.000  
Kendall's tau-b = 0.8553  ASE = 0.009

True D defined as cumsmst0 =~ 0  [95% Conf. Inter.]

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pr(+</td>
<td>D)</td>
<td>Pr(-</td>
<td>~D)</td>
</tr>
<tr>
<td></td>
<td>75.54%</td>
<td>99.93%</td>
<td>98.81%</td>
<td>98.13%</td>
</tr>
<tr>
<td></td>
<td>74.78%</td>
<td>99.88%</td>
<td>98.62%</td>
<td>97.89%</td>
</tr>
<tr>
<td></td>
<td>76.30%</td>
<td>99.98%</td>
<td>99.00%</td>
<td>98.37%</td>
</tr>
</tbody>
</table>
Prevalence                  Pr(D)     7.23%     6.77%     7.69%
-----------------------------------------------

. diagtest hwink0 cuhw0,all

<table>
<thead>
<tr>
<th>Hookworm infection</th>
<th>Cumulative Hookworm</th>
</tr>
</thead>
<tbody>
<tr>
<td>status</td>
<td>infection status</td>
</tr>
<tr>
<td>based on KK</td>
<td>infection status</td>
</tr>
<tr>
<td>technique</td>
<td>based on KK and FECT</td>
</tr>
<tr>
<td>at baseline</td>
<td>techniques at baseline</td>
</tr>
<tr>
<td>Negative</td>
<td>Positive</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>11,597</td>
<td>290</td>
</tr>
<tr>
<td>Positive</td>
<td>3</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Total</td>
<td>11,600</td>
</tr>
</tbody>
</table>

Pearson chi2(1) = 3.0e+03   Pr = 0.000
likelihood-ratio chi2(1) = 699.0612   Pr = 0.000
Cramér's V = 0.4962
gamma = 0.9985   ASE = 0.001
Kendall's tau-b = 0.4962   ASE = 0.022

True D defined as cuhw0 =~ 0                     [95% Conf. Inter.]
-----------------------------------------------
Sensitivity                  Pr(+ | D)     26.02%     25.24%     26.81%
Specificity                  Pr(- | ~D)     99.97%     99.95%     100.00%
Positive predictive value    Pr( D | +)     97.14%     96.84%     97.44%
Negative predictive value    Pr(~D | -)     97.56%     97.28%     97.84%
-----------------------------------------------

Prevalence                  Pr(D)     3.27%     2.95%     3.59%
-----------------------------------------------

. diagtest hwfec0 cuhw0,all

<table>
<thead>
<tr>
<th>Hookworm infection</th>
<th>Cumulative Hookworm</th>
</tr>
</thead>
<tbody>
<tr>
<td>status</td>
<td>infection status</td>
</tr>
<tr>
<td>Formal Ether</td>
<td>based on KK and FECT</td>
</tr>
<tr>
<td>Concentration</td>
<td>techniques at baseline</td>
</tr>
<tr>
<td>Technique at baseline</td>
<td>Negative Positive</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Negative (No hookworm)</td>
<td>11,782</td>
</tr>
<tr>
<td>Positive(Hookworm ov)</td>
<td>1</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Total</td>
<td>11,783</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
</tr>
</tbody>
</table>

Pearson chi2(1) = 1.0e+04  Pr = 0.000  
likelihood-ratio chi2(1) = 2.7e+03  Pr = 0.000  
Cramér's V = 0.9198  
gamma = 1.0000  ASE = 0.000  
Kendall's tau-b = 0.9198  ASE = 0.010

True D defined as cuhw0 ~ 0  [95% Conf. Inter.]

| Sensitivity | Pr(+|D) | 85.28% | 84.65% | 85.91% |
|-------------|--------|--------|--------|--------|
| Specificity | Pr(-|~D) | 99.99% | 99.98% | 100.01% |

Positive predictive value  Pr(D|+)  99.70%  99.61%  99.80%
Negative predictive value  Pr(~D|-)  99.51%  99.39%  99.63%

Prevalence  Pr(D)  3.24%  2.92%  3.55%

```
. diagtest alinkk0 cumal0,all
```

<table>
<thead>
<tr>
<th>AI</th>
<th>Infection</th>
<th>Cumulative AI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>status</td>
<td>infection status</td>
</tr>
<tr>
<td></td>
<td>based on</td>
<td>based on KK and FECT</td>
</tr>
<tr>
<td></td>
<td>KK at</td>
<td>at baseline</td>
</tr>
<tr>
<td></td>
<td>baseline</td>
<td>Negative</td>
</tr>
<tr>
<td>---</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>Negative</td>
<td>11,693</td>
<td>179</td>
</tr>
<tr>
<td>Positive</td>
<td>1</td>
<td>117</td>
</tr>
<tr>
<td>Total</td>
<td>11,694</td>
<td>296</td>
</tr>
</tbody>
</table>

Pearson chi2(1) = 4.6e+03  Pr = 0.000  
likelihood-ratio chi2(1) = 907.4312  Pr = 0.000  
Cramér's V = 0.6212  
gamma = 0.9997  ASE = 0.000  
Kendall's tau-b = 0.6212  ASE = 0.023

True D defined as cumal0 ~= 0  [95% Conf. Inter.]

| Sensitivity | Pr(+|D) | 39.53% | 38.65% | 40.40% |
|-------------|--------|--------|--------|--------|
| Specificity | Pr(-|~D) | 99.99% | 99.97% | 100.01% |

Positive predictive value  Pr(D|+)  99.15%  98.99%  99.32%
Negative predictive value  Pr(~D|-)  98.49%  98.27%  98.71%

Prevalence  Pr(D)  2.47%  2.19%  2.75%

45
. diagtest  alfect0 cumal0,all

|         | Ascaris | lumbricoid |
|------------------|------------------|
| Cumulative Al     | infection status at baseline using FECT |
| Negative          | 11,874 75 | 11,949  |
| Positive          | 3 227 | 230  |
| Total             | 11,877 302 | 12,179  |

Pearson chi2(1) = 9.0e+03  Pr = 0.000
likelihood-ratio chi2(1) = 1.9e+03  Pr = 0.000
Cramér's V = 0.8584
gamma = 0.9998  ASE = 0.000
Kendall's tau-b = 0.8584  ASE = 0.015

True D defined as cumal0 ~ = 0  [95% Conf. Inter.]

| Sensitivity          | Pr( +| D) 75.17% 74.40% 75.93% |
|----------------------|------------------|
| Specificity          | Pr(-| ~D) 99.97% 99.95% 100.00% |
| Positive predictive value | Pr( D| +) 98.70% 98.49% 98.90% |
| Negative predictive value | Pr(~D| -) 99.37% 99.23% 99.51% |
| Prevalence           | Pr(D) 2.48% 2.20% 2.76% |

. diagtest  ttstkk0 cumtt0,all

|         | Tt | Infection | Cumulative T. |
|------------------|------------------|
| status at        | trichiura infection |
| baseline         | status based on KK |
| based on | and FECT techniques |
| KK | Negative | Positive | Total |
|------------------|------------------|
| Negative          | 11,976 6 | 11,982  |
| Positive          | 0 10 | 10  |
| Total             | 11,976 16 | 11,992  |
Pearson chi2(1) = 7.5e+03   Pr = 0.000
likelihood-ratio chi2(1) = 140.6098   Pr = 0.000
Cramér's V = 0.7904
gamma = 1.0000  ASE = 0.000
Kendall's tau-b = 0.7904  ASE = 0.077

True D defined as cumtt0 =~ 0   [95% Conf. Inter.]

Sensitivity         Pr(+| D) 62.50%    61.63%   63.37%
Specificity         Pr(-|~D) 100.00%   100.00%  100.00%
Positive predictive value  Pr(D+) 100.00%  100.00%  100.00%
Negative predictive value  Pr(~D-) 99.95%   99.91%   99.99%

Prevalence  Pr(D) 0.13%    0.07%    0.20%

. diagtest tfect0 cumtt0,all

T. |  trichiura |
infection | Cumulative T.  
status | trichiura infection 
based on | status based on KK 
FECT at  | and FECT techniques 
baseline | Negative  Positive | Total

|                | 12,163   8 | 12,171
Negative |
| 0 | 8 | 8
Positive |

|                | 12,163   16 | 12,179
Total |
| 12,163 | 16 | 12,179

Pearson chi2(1) = 6.1e+03   Pr = 0.000
likelihood-ratio chi2(1) = 111.0625   Pr = 0.000
Cramér's V = 0.7069
gamma = 1.0000  ASE = 0.000
Kendall's tau-b = 0.7069  ASE = 0.088

True D defined as cumtt0 =~ 0   [95% Conf. Inter.]

Sensitivity         Pr(+| D) 50.00%    49.11%  50.89%
Specificity         Pr(-|~D) 100.00%   100.00%  100.00%
Positive predictive value  Pr(D+) 100.00%  100.00%  100.00%
Negative predictive value  Pr(~D-) 99.93%   99.89%   99.98%

Prevalence  Pr(D) 0.13%    0.07%    0.20%
. log close
   name: <unnamed>
   log: E:\George-1\model basics\Evaluation of KK in the absence of a gold standard.log
   log type: text
   closed on: 1 Sep 2014, 07:49:45

---------------------------------------------------------------------------------------------------