Process-based Bayesian modelling of imperfect diagnostic tests in the absence of a 'gold' standard: applications to egg-count tests for helminth infections

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Oliver Bärenbold

aus Roggliswil, LU

Basel, 2020

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Dr. Jürg Utzinger, PD Dr. Penelope Vounatsou und Prof. Dr. Laura Rinaldi.

Basel, den 19.11.2019

Prof. Dr. Martin Spiess
Dekan

Summary

Schistosomiasis and soil-transmitted helminthiasis are two of the most prevalent infections among the human helminthiasis with more than 2 billion infected individuals worldwide. Despite their low direct mortality the two diseases cause a considerable burden with more than 20'000 deaths annually, and wide spread anaemia and growth stunting. Their burden is highest in sub-Saharan Africa where mass-drug administration of preventive chemotherapy is the main intervention strategy. The World Health Organization has set a goal to achieve elimination as a public health problem for schistosomiasis by 2025. Treatment needs are evaluated based on observed prevalence traditionally determined using parasitological methods, i.e. counting eggs in stool or urine. However, diagnostic sensitivity is low, especially for light infections. Furthermore, identifying and counting eggs requires trained technicians and a laboratory posing a considerable effort. Hence, there is a need for sensitive and field-applicable diagnostic methods. Alternative diagnostics have been developed for example based on detecting a specific antigen in urine, blood in stool or urine, or by directly detecting genetic material. However, no 'gold' standard diagnostic exists and all observations have to be understood in the context of their limitations, making incorporation of different data sources difficult.

Currently, the impact a diagnostic method has on the accuracy of disease prevalence estimation is largely ignored. Observed prevalence from differing diagnostic methods are often integrated without adjustment, leading to bias. Methods are often compared assuming an artificial 'gold' standard, thus, underestimating both specificity and sensitivity of alternative diagnostic tests. Furthermore, egg counts are dichotomised in positives/negatives before analysis removing

important information about the day-to-day and slide-to-slide variations in egg counts from the data and hence, there is a need for improved methodology. In this thesis, we develop a modelling framework which allows estimation of diagnostic error as a function of infection intensity. Our approach estimates the 'true' prevalence and can relate prevalence based on different diagnostic methods without assuming any 'gold' standard.

In Chapter 2, we develop an egg count model for the Kato-Katz technique which is typically used to diagnose *S. mansoni* and soil-transmitted helminthiasis. We apply the model to data from Côte D'Ivoire *S. mansoni* and hookworm and estimate the diagnostic sensitivity of Kato-Katz in relation to infection intensity using samples taken on multiple days. Moreover, we determine the change of prevalence with age.

In Chapter 3, we extend our egg count model to fit data based on the point-of-care circulating cathodic (POC-CCA) antigen test when Kato-Katz data on the same individual are available. The model allows us to quantify Kato-Katz day-to-day and slide-to-slide variations and is fitted on a suite of 30 datasets from Africa and the Americas. Results inform about the relation between the sensitivity of Kato-Katz and the infection intensity for different sampling designs as well as about the infection intensity-dependent sensitivity of POC-CCA for multiple interpretations of the test (POC-CCA trace reading treated as positive or negative). We use simulations to translate intervention prevalence thresholds proposed by the World Health Organization (WHO) from the Kato-Katz technique into POC-CCA.

In Chapter 4, we estimate the 'true' prevalence of *S. mansoni* from population summary measures of Kato-Katz such as the arithmetic, the geometric mean, and the observed prevalence, obtained from one to three sample slides, taken over one to three days. We use the egg count model derived in previous chapters to estimate the individual-level sensitivity of Kato-Katz and make common assumptions about the distribution of worms in the population combined with a worm-mating model to create summary measures of simulated populations. A parametric model is fitted to the simulated data, which enables us to predict the 'true' prevalence from population summary measures and validate it.

In Chapter 5, we simulate the worm and the worm-pair distribution of *S. mansoni* using data from Togo and generate observations at individual level from the egg count model described in

Chapter 3. The simulated data is used as a testbed for survey sampling designs based on either the Kato-Katz or the POC-CCA diagnostic, varying the numbers of individuals per location and the locations per district. The survey sampling designs are evaluated according to their accuracy in determining treatment needs at the district level.

In Chapter 6, we extend the egg count model developed in Chapter 2 to *S. haematobium* data derived from a single sample, urine filtration diagnostic and multiple additional haematuria based diagnostics, obtained from the 2010 schistosomiasis and soil-transmitted helminthiasis national survey in Zimbabwe. The specificity and sensitivity in relation to infection intensity was estimated for each diagnostic method at the individual level. Furthermore, we estimated the ability of each diagnostic to detect districts with prevalence above 10% by urine filtration.

In Chapter 7, we conduct an in-depth analysis of the urine filtration and the reagent strip diagnostic used for *S. haematobium* analysing data from Tanzania with measurements over five days. We estimate sensitivity in relation to infection intensity for both diagnostic tests and translate prevalence thresholds from urine filtration into the reagent strip.

The main contributions of the thesis in the analysis of helminthiasis data that are subject to diagnostic error are: i) flexible Bayesian models for data from tests counting eggs for both *S. mansoni* and *S. haematobium* under various sampling schemes; ii) parametric models for the diagnostic sensitivity and specificity of diagnostic tests alternative to egg count based ones; iii) simulation models to translate prevalence thresholds from one diagnostic method to another; iv) simulation methods to estimate the 'true' prevalence of *S. mansoni* from population disease summary measures; and v) simulation methods for survey sampling designs when taking into account diagnostic methods in terms of the number of locations per district and the individuals screened per location. Furthermore, we derive results directly relevant for policy such as: vi) sensitivity and specificity of various diagnostic methods for *S. mansoni* and *S. haematobium*; vii) translation of WHO prevalence thresholds for Kato-Katz into the POC-CCA test; viii) parameters to estimate 'true' prevalence from population summary measures; ix) prevalence thresholds for haematuria-based methods to detect *S. haematobium*; and x) recommendations regarding POC-CCA as an alternative to Kato-Katz for *S. mansoni* surveillance.

Zusammenfassung

Bilharziose und die Gruppe der über den Boden übertragenen Helminth Parasiten sind zwei der häufigsten Infektionen unter den menschlichen Helminthiasis Erkrankungen mit mehr als 2 Milliarden infizierten Personen weltweit. Trotz ihrer geringen direkten Sterblichkeit verursachen die beiden Krankheiten eine erhebliche Belastung mit mehr als 20'000 Todesfällen pro Jahr und weit verbreiteter Anämie und Wachstumsstörungen. Ihre Belastung ist in Afrika südlich der Sahara am höchsten, wo die medikamentöse Verabreichung einer vorbeugenden Chemotherapie die Hauptinterventionsstrategie ist. Die Weltgesundheitsorganisation hat sich zum Ziel gesetzt, bis 2025 die Beseitigung als öffentliches Gesundheitsproblem für Bilharziose zu erreichen. Der Behandlungsbedarf wird auf der Grundlage der beobachteten Prävalenz bewertet, die traditionell unter Verwendung parasitologischer Methoden bestimmt wird, d. H. Zählen der Eier im Stuhl oder Urin. Die diagnostische Sensitivität ist jedoch gering, insbesondere bei leichten Infektionen. Darüber hinaus erfordert das Identifizieren und Zählen von Eiern geschulte Techniker und ein Labor mit welches mit erheblichem Aufwand verbunden ist. Es besteht daher ein Bedarf an sensitiven und vor Ort anwendbaren Diagnosemethoden. Es wurden alternative Diagnosen entwickelt, die beispielsweise auf dem Nachweis eines bestimmten Antigens im Urin, Blut im Stuhl und Urin, oder auf dem direkten Nachweis von genetischem Material beruhen. Es gibt jedoch keine 'Gold'-Standarddiagnose, und alle Beobachtungen müssen im Kontext ihrer Limitierungen verstanden werden, was die Einbeziehung verschiedener Datenquellen erschwert.

Gegenwärtig wird der Einfluss einer diagnostischen Methode auf die Genauigkeit der Abschätzung der Prävalenz einer Krankheit weitgehend ignoriert. Die beobachtete Prävalenz von

unterschiedlichen Diagnosemethoden wird häufig ohne Anpassung integriert, was zu Verzerrungen führt. Die Methoden werden oft unter der Annahme eines künstlichen ‘Gold-Standards verglichen, wodurch sowohl die Spezifität als auch die Sensitivität alternativer diagnostischer Tests unterschätzt werden. Darüber hinaus werden die Ei-Zählungen vor der Analyse in Positiv/Negativ eingeteilt, wobei wichtige Informationen zu den täglichen und von Spezimen zu Spezimen schwankenden Ei-Zählungen aus den Daten entfernt werden. Daher besteht ein Bedarf an einer verbesserten Methodik. In dieser Arbeit entwickeln wir eine Modelstruktur, mit dem diagnostische Fehler in Abhängigkeit von der Infektionsintensität abgeschätzt werden können. Unser Ansatz schätzt die ‘wahre’ Prävalenz und kann die Prävalenz basierend auf verschiedenen diagnostischen Methoden in Beziehung setzen, ohne einen "Gold"-Standard anzunehmen.

In Kapitel 2 entwickeln wir ein Eizählmodell für die Kato-Katz-Technik, welche normalerweise zur Diagnose von *S. mansoni* und bodenübertragener Helminthiasis verwendet wird. Wir wenden das Modell auf Daten für *S. mansoni* und Hakenwurm aus Côte D’Ivoire an und schätzen die diagnostische Empfindlichkeit von Kato-Katz in Bezug auf die Infektionsintensität anhand von Proben, die an mehreren Tagen entnommen wurden. Darüber hinaus bestimmen wir die Abhängigkeit der Prävalenz vom Alter.

In Kapitel 3 erweitern wir unser Eizählmodell, um Daten basierend auf dem Point-of-Care-Test für zirkulierende kathodische Antigene (POC-CCA) anzupassen, wenn Kato-Katz-Daten für dieselbe Person verfügbar sind. Das Modell ermöglicht die Quantifizierung von Kato-Katz-Variationen von Tag zu Tag und von Spezimen zu Spezimen und ist auf eine Sammlung von 30 Datensätzen aus Afrika und Amerika abgestützt. Die Ergebnisse geben Auskunft über den Zusammenhang zwischen der Sensitivität von Kato-Katz und der Infektionsintensität bei verschiedenen Probenahmedesigns sowie über die infektionsintensitätsabhängige Sensitivität von POC-CCA bei Mehrfachinterpretationen des Tests (POC-CCA-Spurenwert als positiv oder negativ behandelt). Wir verwenden Simulationen, um die von der Weltgesundheitsorganisation (WHO) vorgeschlagenen Interventionsprävalenzschwellenwerte aus der Kato-Katz-Technik in POC-CCA zu übersetzen.

In Kapitel 4 schätzen wir die ‘wahre’ Prävalenz von *S. mansoni* anhand von Populationsdurchschnitten von Kato-Katz Resultaten wie dem arithmetischen Mittel, dem geometrischen

Mittel und der beobachteten Prävalenz, welche von ein bis drei Spezimen von über ein bis drei Tage entnommenen Kotproben bestimmt wurden. Wir verwenden das in den vorangegangenen Kapiteln hergeleitete Eizählmodell, um die Sensitivität von Kato-Katz auf individueller Ebene abzuschätzen, und treffen gemeinsame Annahmen über die Verteilung von Würmern in der Population in Kombination mit einem Modell welches den Paarungsprozess von Würmern beschreibt, um Populationsdurchschnitte simulierter Populationen zu simulieren. Anhand der simulierten Daten werden die Parameter eines Modells geschätzt, das es uns ermöglicht, die 'wahre' Prävalenz anhand von Populationsdurchschnitten vorherzusagen und zu validieren.

In Kapitel 5 simulieren wir die Wurm- und die Wurmpaarverteilung von *S. mansoni* unter Verwendung von Daten aus Togo und generieren Beobachtungen auf individueller Ebene aus dem in Kapitel 3 beschriebenen Eizählmodell. Die simulierten Daten werden als Testumgebung für Probenahmedesigns verwendet, die entweder auf der Kato-Katz- oder der POC-CCA-Diagnose basieren und die Anzahl der Personen pro Standort und die Standorte pro Distrikt variieren. Die Probenahmedesigns werden anhand ihrer Genauigkeit bei der Ermittlung des Behandlungsbedarfs auf Distriktebene bewertet.

In Kapitel 6 erweitern wir das in Kapitel 2 entwickelte Eizählmodell auf *S. haematobium*-Daten, aus einer einzelnen Urinprobe, einer Urinfiltrationsdiagnose und mehreren zusätzlichen auf dem Detektieren von Hämaturie basierenden Diagnostiken, erhalten aus der nationalen Umfrage zu Bilharziose und bodenübertragener Helminthiasis 2010 in Simbabwe. Die Spezifität und Sensitivität in Bezug auf die Infektionsintensität wurde für jede Diagnosemethode auf individueller Ebene geschätzt. Darüber hinaus schätzten wir die Fähigkeit jeder Diagnostischen Methode, Distrikte mit einer Prävalenz von über 10% bestimmt durch Urinfiltration zu erkennen.

In Kapitel 7 führen wir eine vertiefte Analyse der Urinfiltration und der Reagenzstreifen-Diagnose als *S. haematobium* Diagnostik, anhand von Daten aus Tansania mit Messungen über fünf Tage durch. Wir schätzen die Sensitivität in Abhängigkeit von der Infektionsintensität für beide diagnostischen Tests und übersetzen die Interventionsprävalenzschwellenwerte von der Urinfiltration zum Reagenzstreifen.

Die Hauptbeiträge der Dissertation in der Analyse von Helminthiasis-Daten, bei denen diagnostische Fehler auftreten, sind: i) Flexible Bayes'sche Modelle für Daten aus Tests, die sowohl

für *S. mansoni* als auch *S. haematobium* Eier nach verschiedenen Probenahmedesigns zählen; ii) parametrische Modelle für die diagnostische Sensitivität und Spezifität diagnostischer Tests als Alternative zu Ei-Zählungen-basierten Tests; iii) Simulationsmodelle zur Umrechnung von Prävalenzschwellenwerten von einer Diagnosemethode in eine andere; iv) Simulationsmethoden zur Abschätzung der 'wahren' Prävalenz von *S. mansoni* anhand von Populationsdurchschnitten von Ei-Zählungen-basierten Tests; und v) Simulationsmethoden für Probenahmedesigns unter Berücksichtigung von Diagnosemethoden in Bezug auf die Anzahl der Standorte pro Bezirk und die pro Standort gescreenten Personen. Darüber hinaus leiten wir Ergebnisse ab, die für die Strategie zur Unterstützung der öffentlichen Gesundheit direkt relevant sind, wie: vi) Sensitivität und Spezifität verschiedener Diagnosemethoden für *S. mansoni* und *S. haematobium*; vii) Übersetzung der WHO-Prävalenzschwellenwerte für Kato-Katz in den POC-CCA-Test; viii) Parameter zur Schätzung der 'wahren' Prävalenz anhand von Populationsdurchschnitten; ix) Prävalenzschwellenwerte für Hämaturie basierte Methoden zum Nachweis von *S. haematobium*; und x) Empfehlungen zu POC-CCA als Alternative zu Kato-Katz für die *S. mansoni* Überwachung.

Acknowledgement

It is said it takes a village to complete a thesis and this one is no different. I am deeply grateful to the Swiss Tropical and Public Health institute and its people for having had the chance to spend the past four years here.

In particular, I would like to thank PD Dr. Penelope Vounatsou for accepting me in the Bayesian Biostatistics group and guiding me throughout the PhD. Her continued trust gave me an amazing level of freedom to pursue whatever idea I felt like being worthwhile. Her hard work, mathematical rigorousness and attention to detail when editing manuscripts has been a most important contribution to this thesis. I am grateful for her openness in discussing new ideas and her willingness to challenge everything. I am also grateful to the European Research Council (ERC) for funding my thesis.

I would like to thank Prof. Dr. Jürg Utzinger who made sure that the work always stayed connected to what is important in applications to public health. His relaxed but dedicated approach to work, his motivation and deep understanding have always been inspiring. I am grateful for all the epidemiological input into my work, the collaborations he made possible, and the occasional travel photo attached to a mail. Many thanks also to Anja Schreier for always helping me find time with Jürg even when constrained by his many responsibilities.

I am thankful to Prof. Dr. Laura Rinaldi for agreeing to examine my thesis, which includes coming to grey and rainy Basel in November and to Prof. Dr. Pascal Mäser for agreeing to chair my defence.

I would like to thank the whole Bayesian Biostatistics group for having been an a very

welcoming group and the perfect company for countless breakfasts, lunchs, coffees: Elizaveta Semenova, Carla Grolimund, Alex Karagiannis-Voules, Frédérique Chammartin, Serena Scudella, Christos Kokaliaris, Isidoros Papaioannou, Anton Beloconi, Betty Nambuusi, Julius Ssempiira, Ourohiré Millogo, Sammy Khagayi, Ying-Si Lai, Eric Diboulo, Abbas Adigun, Sabelo Dlamini, Salomon Massoda Tonye, Guojing Yang, and Martin Matuska.

I am grateful for all the people that I got to collaborate with on one project or the other keeping my methodological work grounded in reality: Dan Colley, Fiona Fleming, Nicholas Midzi, Amadou Garba. Statistics is nothing without data. Therefore, I would like to thank all the collaborators who did the hard work of collecting data in the field without whom none of this work would be possible or even make sense: Giovanna Raso, Jean T. Coulibaly, Eliézer K. N'Goran, Ayat A. Haggag, Reda M. Ramzy, Rufin K. Assaré, Edridah M. Tukahebwa, Jean B. Mbonigaba, Victor Bucumi, Biruck Kebede, Makoy S. Yibi, Aboulaye Meité, Louis-Albert Tchuem Tchuenté, Pauline Mwinzi, Ameyo M. Dorkenoo, Portia Manangazira, Isaac Phiri, Masceline Mutsaka, Gibson Mhlanga, Christoph Hatz.

Swiss TPH has been a most stimulating environment and there have been countless people who crossed my path in one way or the other of whom I would just like to name few who happened to be around for the whole journey: Liliana Rutaihwa, Monica Ticlla, Rhastin Castro, Natalie Wiedemar, Anna Fesser, Josephine Malinga. A thanks also to the student who take care of the social side of Swiss TPH which is part of what makes this institute so special: Marta Palmeirim, Katrina Obas, Wendelin Moser.

I would also like to thank a number of friends that have supported me over the years: Wander Engbers, Marco Trombetta, Aurore Flaceliere, Jeremy Laniel, Lukas Muri, David Naef, Roman Käch, Pablo Estevez Romero, Angels Garcia Hervas, Gabriela Silveira Borges, Emilie Mathys, Mulako Jaeger Silumesii, Laura Orozco, Maxine Gachoka.

Finally a big thank you for always being there and never doubting: Fabian, Anina, Milo, Yann, Jana, Rebekka, Ruth, and Thomas.

Contents

Chapter 1

Introduction

1.1 Rationale

Helminthiasis, a group of infections caused by parasitic worms, are some of the most prevalent diseases in the world. Due to their low direct mortality and unspecific symptoms it is difficult to gather information the exact number of infected individuals in the world [Stoll, 1947]. Although direct mortality is low, due to the high prevalence a considerable burden is caused by the diseases with more than 20,000 deaths annually and 2 billion infected individuals [GBD 2016 DALYs and HALE Collaborators, 2017; GBD 2016 Causes of Death Collaborators, 2017; GBD 2016 Disease and Injury Incidence and Prevalence Collaborators, 2017]. Treatment for schistosomiasis and soil-transmitted helminthiasis (STH), the helminth infections discussed in this thesis, is cost effective and there has recently been considerable support by donors companies towards elimination [King et al., 2011; Rollinson et al., 2013].

Schistosomiasis and STH are known as neglected tropical diseases (NTD). The term NTDs has been established in the early 2000s in the context of the Millennium Development Goals (MDGs), a concentrated global commitment to ensure human dignity worldwide [Molyneux, 2004; Hotez et al., 2007]. Helminthiasis was addressed in the sixth of the MDGs that aimed to combat HIV-AIDS, malaria and 'other diseases'. The goal brought some large-scale funding to the fight against infectious disease. However, the 'other diseases' such as the NTDs did not yet sufficiently benefit from it [Hotez et al., 2007]. For schistosomiasis, strategies have been

defined in the 54th world health assembly in 2001 and then again in a 2006 manual by the World Health Organization (WHO) that rely primarily on mass-drug-administration (MDA) to combat the disease [World Health Assembly, 2001; WHO, 2006]. In 2013 a progress report has been formulated by the WHO shifting the policy landscape towards elimination at least as a public health problem [Rollinson et al., 2013; WHO, 2013]. The progress in the fight against infectious diseases is also reflected in the Sustainable Development Goals (SDGs), the successor of the MDGs after 2015, being less concentrated on health issues. Despite the NTDs are still included as a crucial part in the universal health coverage concept of 'leaving no one behind' and in the combat of open defecation [Bangert et al., 2017; Molyneux et al., 2017].

Helminthiasis control is primarily achieved using mass drug administration (MDA) but can be complemented by improvements to sanitary infrastructure or by reduction of the population of disease specific intermediate host snails [Hotez et al., 2007; King, 2007]. Targeted interventions become necessary for efficient resource use once prevalence reduces and the disease becomes more focal, requiring disease maps with high resolution [Tchuem Tchuente et al., 2018]. However, data on disease prevalence and on the associated burden is sparse since the people affected are disproportionately poor and live in remote areas with little access to health care [King, 2010]. Estimating disease risk at high geographical resolution for targeted interventions requires new statistical methods which combine existing data sets in terms of population included and diagnostic methods used [Chammartin et al., 2013; Hay et al., 2013].

The report for 'Research Priorities for Helminth Infections' by the WHO in 2012 [TDR Disease Reference Group on Helminth infections, 2012] mentions the need for improved understanding of diagnostic methods as well as investments in developing new ones. Currently, the topic of statistical modelling of diagnostic tests is largely ignored even though there exists no 'gold' standard diagnostic method and all measurements are to be understood as a proxy for the 'true' underlying disease burden. Surveys are currently performed using various diagnostic methods and sampling schemes without strong justification, hampering the comparability across studies. Observed prevalence from studies with different diagnostic methods are reported and integrated without adjustments leading to bias. New diagnostic methods are commonly compared with parasitological methods considered as 'gold' standard leading to under-estimation of

specificity because all additional positives are labelled as false-positives.

This thesis concentrates on the specific aspect of data integration described above, namely, the absence of affordable, reliable and field applicable 'gold' standard diagnostic tests with well-known accuracy for both schistosomiasis and the STHs [Stothard et al., 2014; Utzinger et al., 2015]. As worm infections are traditionally diagnosed using parasitological methods where the accuracy is strongly dependent on the number of worms harboured by the infected person, observations in each study give only partial and possibly biased information on what is actually happening [Colley et al., 2017]. We develop methods to increase the information extracted from observations based on imperfect diagnostic tests, primarily parasitological methods, using process-based models of the diagnostic method and knowledge of disease transmission. There is a multitude of possible applications on diagnostic tests for egg shedding parasites of which a subset was explored and will be presented in this thesis.

1.2 Disease characteristics

Parasitological methods are widespread and the most common methods to diagnose helminth infections. This thesis concentrates on applications to schistosomiasis with one example including STH infections, namely hookworms, the former is the most severe and the latter the most common helminth infections and both belonging to the NTDs.

1.2.1 Schistosomiasis

Schistosomiasis, bilharzia, or snail fever is a parasitic disease caused by flatworms first described in 1851 by the physician Theodor Bilharz. There are six species of the genus Schistosoma of which *Schistosomiasis mansoni* and *Schistosomiasis haematobium* are the two common ones in sub-Saharan Africa. All schistosomes rely on an intermediate host, a fresh-water snail, for asexual reproduction that release the infectious stage into the water where infection of humans or other mammals occurs by penetration of the skin (see Figure 1.1). After penetration, the schistosomes migrate through the blood vessels to the liver for maturation into separate-sex adults. Worms pair up in permanent couples and migrate to the veins in proximity to the intestines or bladder for

S. mansoni and *S. haematobium*, respectively. About five weeks after infection egg production starts which migrate from the blood stream to the bladder or intestines where they are excreted completing the cycle. A female worm can survive for many years in adult stage and produce several hundred eggs per day for both *S. mansoni* and *S. haematobium* leading to a large number of eggs over their life-cycle if no treatment occurs that either leave the body continuing the cycle or are lodged in various organs and tissue.

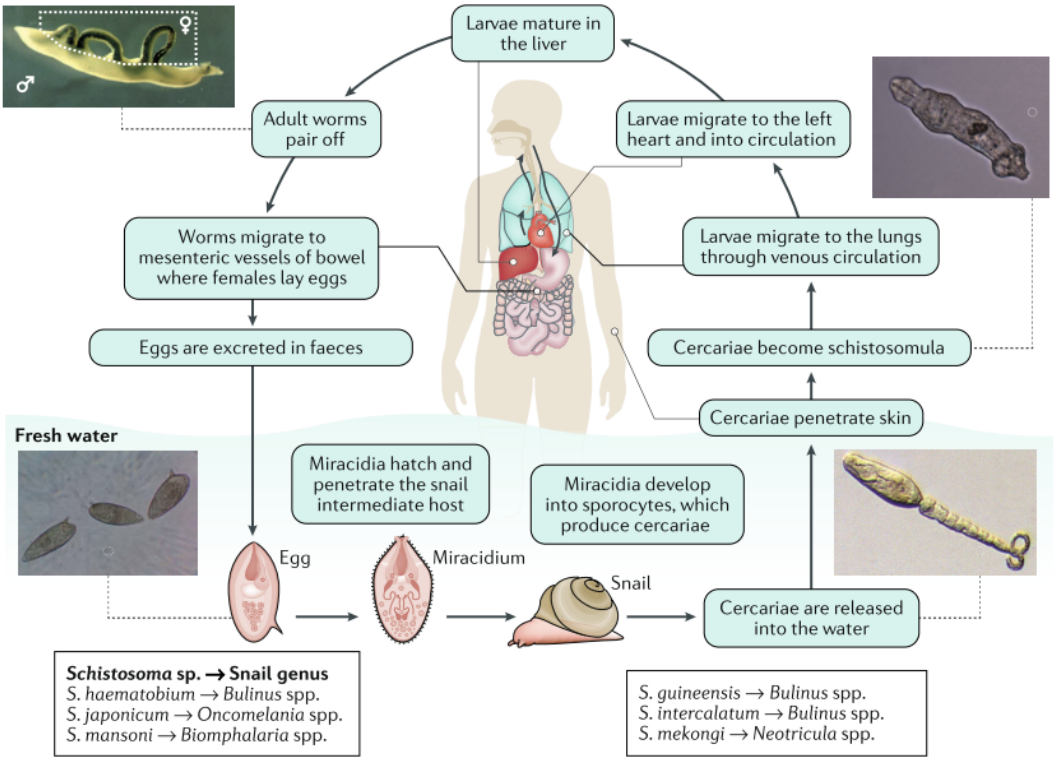
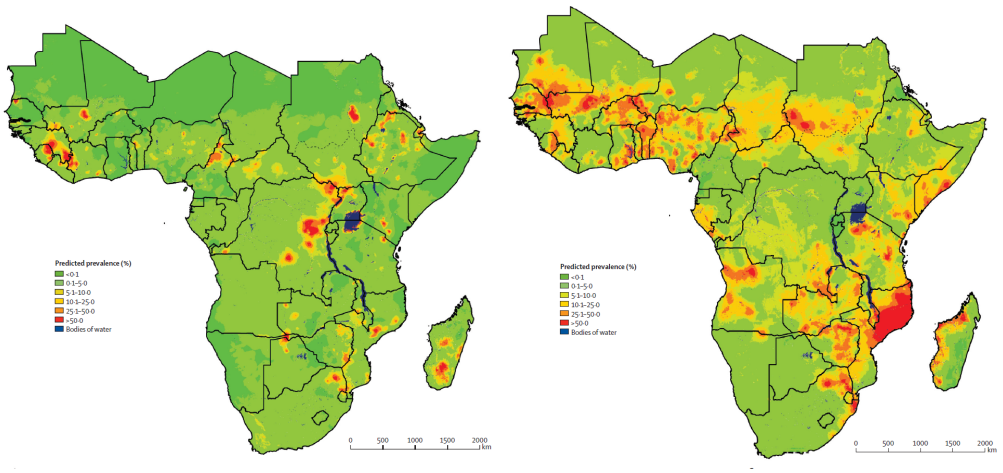


Figure 1.1: Life-cycle of *S. mansoni*. Figure taken from McManus et al. [2018].

The disease burden of untreated schistosomiasis is primarily caused by eggs getting trapped in tissue leading to substantial morbidity and mortality in the shape of growth stunting, impaired cognition, anaemia, higher risk for HIV infections, and even increased risk for cancer after years of exposure [Abdulamir et al., 2009; Kjetland et al., 2014]. The complicated life cycle involving multiple actors interconnected with environmental factors leads to highly variable disease prevalence and mean worm burden even from village to village. Burden of the highly focal

disease is distributed according to available fresh-water bodies with suitable snail populations, access to clean water, common professions and uncountable possible further influences. It is estimated that globally there are more than 700 million people at risk for infection and more than 200 million people infected causing a combined 2.1 million disability-adjusted life years lost (DALYs) in 2016 [GBD 2016 Disease and Injury Incidence and Prevalence Collaborators, 2017; GBD 2016 Causes of Death Collaborators, 2017; GBD 2016 DALYs and HALE Collaborators, 2017].

S. mansoni and *S. haematobium* are most prevalent in sub-Saharan Africa while there are also endemic regions in the Americas for the former and in the Middle East for the latter. *S. haematobium* is the more common species but both are wide-spread in Africa (see Figure 1.2). There are large uncertainties in the estimated disease prevalence because of a lack of data from many regions [Lai et al., 2015].



(a) (b)

Figure 1.2: Estimated prevalence of a) *S. mansoni* and b) *S.haematobium* in Africa. Taken from Lai et al. [2015]

The highest burden is often found among impoverished people in remote locations because of poor hygiene and water infrastructure with school-aged children and pregnant women being most affected by the associated morbidity. The distribution of worms in a population is highly over-dispersed meaning the individuals with the highest infection-intensity harbour a large proportion of the total worm number. Although burden is highest in severe infections there is no such

thing as asymptomatic or harmless infection but rather that the effects are difficult to separate from co-morbidities in studies with imperfect diagnostics for weak infections. Therefore, it is important to also find and treat light infections if elimination as a public health problem is the goal [King, 2015].

1.2.2 Soil-transmitted helminthiasis

The STHs consist of three species of parasitic worms that are very common in the Global South with more than 1.5 billion individuals infected and 7.5 million DALYs lost [GBD 2016 Disease and Injury Incidence and Prevalence Collaborators, 2017; GBD 2016 DALYs and HALE Collaborators, 2017]. Co-infections are very common and evaluation and control for these diseases is usually integrated in a common framework due to shared diagnostic methods and treatment options. The STHs are one of the major causes of growth stunting with a large impact on educational and economic development [Bethony et al., 2006]. *Ascaris lumbricoides* roundsworms and *Trichuris trichiura* whipworms are transmitted by ingesting viable eggs from where the former develops into egg-laying worms in the small intestines after extra-intestinal migration through the liver and the lungs, and the latter in the large intestine after direct travel through the colon. *Necator americanus* and *Ancylostoma duodenale* hookworms are transmitted through contact of the larvae with human skin from where they enter the host's circulation to finally arrive in the upper small intestine where they develop to egg-laying adults. Eggs are excreted through stool for all three species and no reproduction happens in the host.

Soil-transmitted helminths are common in the tropics and sub-tropics due to climate conditions being important determinants combined with a lack of sanitation infrastructure. Figure 1.3 depicts the hookworm prevalence in Africa before the year 2000 and after indicating that the disease is still wide-spread in sub-Saharan Africa [Karagiannis-Voules et al., 2015]. The WHO formulates a goal to eliminate STHs as a public health problem in children by 2020. However, in 2010 only about a third of the children at risk received preventive treatment supporting the need for continued effort to deliver MDA to all eligible children [WHO, 2012].

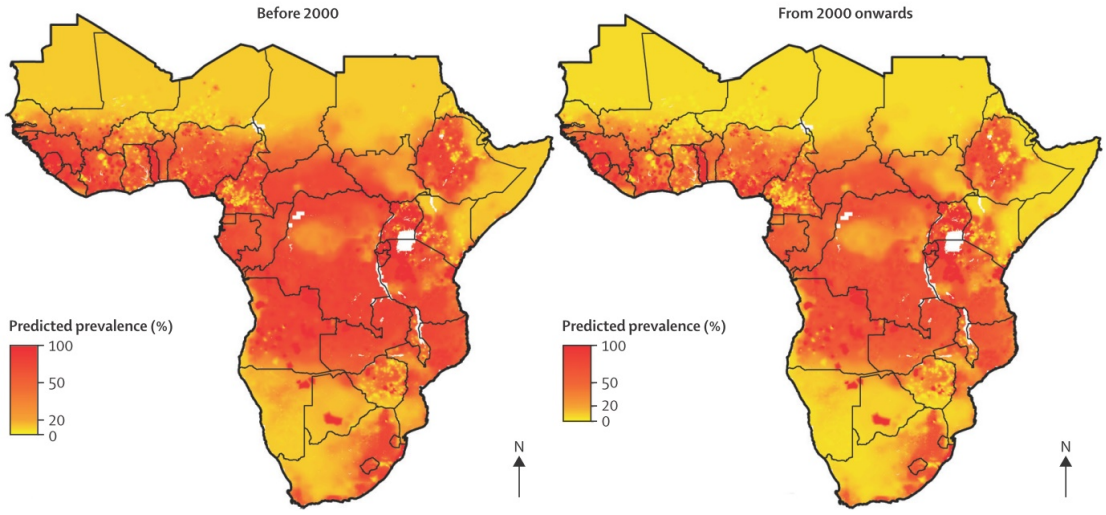


Figure 1.3: Estimated prevalence of Hookworm in Africa. Taken from Karagiannis-Voules et al. [2015]

1.3 Diagnostics

Diagnostics for STHs and schistosomiasis is an active field of research with development of high-tech methods based on the detection of genetic material, low-tech methods like simple questionnaires, as well as continued evaluation of traditional techniques in newly occurring settings currently happening [Stothard et al., 2014; Utzinger et al., 2015].

Diagnosing established infections of STH or schistosomiasis is most simply done by detecting eggs in stool or urine depending on the location of the parasite species in the host body. The STHs and *S. mansoni* shed eggs in stool while *S. haematobium* releases egg through urine. The standard approach for stool based diagnostics is the so-called Kato-Katz method [Kato and Miura, 1954; Katz et al., 1972]. The technique is performed by pressing stool through a sieve and preparing a specified amount of 41.7 mg as a stool-smear on a slide where glycerine is used to make the eggs visible. Subsequently, the eggs are classified and counted using light-microscopy allowing for quantification of infection intensity [WHO, 1991]. A common classification for *S. mansoni* is to define 1-100 eggs per gram (EPG) as light infection, 101-400 EPG as moderate infection, and >400 EPG as severe infection [WHO, 2013]. However, classification does not reproduce well due to large variability from day to day [Kongs et al., 2001; Krauth et al., 2012]. To improve

diagnostic sensitivity multiple slides of the same stool sample are taken or additional samples are collected on consecutive days to limit the influence of highly variable egg excretion [Utzinger et al., 2001; Enk et al., 2008]. Specificity is close to 100% in absence of mishandling of samples while the sensitivity is as low as 50% for double slide Kato-Katz on a single stool sample. In low transmission settings where most samples are negative it has been suggested to pool stool samples to increase cost-effectiveness [Kure et al., 2015]. Alternative egg concentration methods have been proposed to enhance sensitivity and compared to Kato-Katz but have not yet seen wide-spread adoption [Glinz et al., 2010; Becker et al., 2011; Moser et al., 2018].

Parasitological diagnosis of *S. haematobium* is done by filtering 10 ml of urine collected between 10 am and 2 pm where filtrates are examined for eggs using light-microscopy [WHO, 1991]. Urine filtration is favoured over techniques like centrifugation due to its independence from electricity. The sensitivity can be increased by increasing the amount of urine or by collecting samples on multiple days as egg output varies considerably across time [Kosinski et al., 2011; Stete et al., 2012; Knopp et al., 2018]. Infections are classified according to infection intensity where $<$50 eggs/10 ml is considered a light infection and $>$50 eggs/10 ml a moderate or severe infection [WHO, 2013]. A common, easily detectable symptom for *S. haematobium* is blood in urine which is why there are multiple alternative for diagnosis based on haematuria [Mafe, 1997]. The most easily applied ones are a questionnaire regarding the recent history of visible blood in urine and the examination of a urine sample for blood [Mafe, 1997; Lengeler et al., 2000]. In order to detect blood that is not visible by eye(i.e. microhaematuria), a reagent strip is used that results in semi-quantitative readings indicating the severity of infection [Kosinski et al., 2011]. However, there are other conditions causing haematuria and therefore in contrast to urine filtration the specificity, that is the probability of a negative test in the absence of infection, is reduced [Krauth et al., 2015].

Additional methods for the detection of schistosomiasis have recently been developed based on the detection of specific antigens in urine and have been compared to parasitological methods for both *S. mansoni* and *S. haematobium* [Obeng et al., 2008; Midzi et al., 2009; Knopp et al., 2015; Ochodo et al., 2015; Danso-Appiah et al., 2016]. The antigens are produced by the worms directly requiring no egg production for a positive test and therefore, diagnosis based on antigens

detects individuals that harbour worms even if there are no viable worm-pairs and hence, no current egg-shedding [van Lieshout et al., 2000]. Antigen based tests are more sensitive for light infections than parasitological ones [Kittur et al., 2016; Danso-Appiah et al., 2016]. However, the specificity is difficult to evaluate because it is not always clear if additional positives are infections not detectable by parasitological methods or false-positives [Greter et al., 2016; Colley et al., 2017; Peralta and Cavalcanti, 2018]. For *S. haematobium* there is currently no field applicable rapid antigen test while the point-of-care circulating cathodic antigen (POC-CCA) test is currently in use for *S. mansoni* at comparable costs to Kato-Katz when operational costs are included [Worrell et al., 2015; Knowles et al., 2017]. Schistosomiasis can also be diagnosed through detection of genetic material of the parasite species by polymerase chain reaction (PCR) which has high sensitivity and specificity [Obeng et al., 2008; Enk et al., 2012; Vinkeles Melchers et al., 2014; Shiff, 2015]. However, this is usually not feasible in settings of interest [Shiff, 2015].

Although the costs of developing new diagnostic tools is relatively high compared to performing parasitological diagnostic techniques, reducing the reliance on stool samples would considerably increase the flexibility in data collection [Turner et al., 2017]. When parasitological techniques become replaced more and more in the future by alternative ones the need for statistical methods that integrate data from different diagnostic techniques without introducing bias increases in order to maintain the ability to track progress over time.

1.4 Disease control guidelines

S. haematobium and *S. mansoni* have first been described in 1852 and 1907, respectively and already in the early 20th century first attempts at disease control have been made in Egypt and Japan with the first MDA being carried out in 1920 in Egypt. Additional control attempts happened in multiple national programs targeting the intermediate host snail as well as health education to prevent exposure. Control strategies evolved in relation to available chemotherapy where the development of single-dose oral treatments was a major milestone with the development of praziquantel in the mid-1970s by Bayer and Merck. MDA with the newly developed treatments proved to be the most cost-effective control strategy in the following years to prevent heavy-

intensity infections that cause a large proportion of the burden. Provision of MDA was still not possible in many countries without external funding and only with more recent significant reductions in the price and donations of praziquantel in conjecture with advocacy for NTD control the efforts have increased [WHO, 2013].

Although WHO provided guidelines for disease evaluation at community level in 1998, recommending a cluster-randomized design with double slide Kato-Katz diagnostics for *S. mansoni* and single urine filtration for *S. haematobium*, MDA for schistosomiasis has only been endorsed as a key public health strategy in 2001 at the 54th World Health Assembly [Montresor et al., 1998; World Health Assembly, 2001]. A goal was set to deliver chemotherapy to more than 75% of school-aged children at risk for infection by 2010. In 2006 the strategy was revised to take into account that also light and moderate infection intensities lead to morbidity including adults in the target population in high risk areas [WHO, 2006]. Treatment guidelines and corresponding intervention thresholds were defined with $< 10\%$ observed prevalence indicating a low-risk community where MDA twice during schooling is advised, 10% to 50% a moderate-risk community where MDA every second year is recommended, and $> 50\%$ where yearly treatment is needed including adults (see Figure 1.4) [WHO, 2013].

Category	Baseline prevalence among school-age children	Action to be taken[b]	
High-risk community	50% by parasitological methods[c] (intestinal and urogenital schistosomiasis) or 30% by questionnaire for history of haematuria	Treat all school-age children (enrolled and not enrolled) once a year	Also treat adults considered to be at risk (from special groupsd to entire communities living in endemic areas)
Moderate-risk community	10% but <50% by parasitological methods (intestinal and urogenital schistosomiasis) or <30% by questionnaire for history of haematuria	Treat all school-age children (enrolled and not enrolled) once every 2 years	Also treat adults considered to be at risk (special groupsd only)
Low-risk community	<10% by parasitological methods (intestinal and urogenital schistosomiasis)	Treat all school-age children (enrolled and not enrolled) twice during their primary schooling age (e.g. once on entry and once on exit)	Praziquantel should be available in dispensaries and clinics for treatment of suspected cases

Figure 1.4: Treatment guidelines and intervention thresholds for both *S. mansoni* and *S. haematobium* taken from the schistosomiasis progress report by WHO in 2013 [WHO, 2013]

While treatment guidelines are formulated in terms of observed prevalence the targets include requirements on the prevalence of heavy-intensity infections creating a mismatch between strategy and goals (see Figure 1.5). The first stage in control efforts towards elimination is the control of

morbidity, which is defined as a prevalence of heavy-intensity infections below 5%. The next stage is the 'elimination as a public health problem' (EPHP) which is defined as prevalence of heavy-infections below 1%, continuing to the elimination stage with a zero incidence rate of infection. The final stage consists of continued surveillance while no new infections occur. The overarching goal is for all countries to control morbidity by 2020, eliminate schistosomiasis as a public health problem and to interrupt transmission and reach elimination stage in suitable settings outside of Africa and in selected countries within the African Region by 2025 [WHO, 2013; Rollinson et al., 2013].

GROUP	1. Countries eligible for control of morbidity	2. Countries eligible for elimination as a public-health problem	3. Countries eligible for elimination (interruption of transmission)	VERIFICATION	4. Countries that have achieved elimination
Goal	Control of morbidity	Elimination as a public-health problem	Elimination (interruption of transmission)		Post-elimination surveillance
Recommended intervention	Preventive chemotherapy Complementary public-health interventions, where possible	Adjusted preventive chemotherapy Complementary public-health interventions strongly recommended	Intensified preventive chemotherapy in residual areas of transmission Complementary public-health interventions essential		Surveillance to detect and respond to resurgence of transmission and to prevent reintroduction (schistosomiasis should be made notifiable)
Target	100% geographical coverage and at least 75% national coverage Prevalence of heavy-intensity infection <5% across sentinel sites[a]	Prevalence of heavy-intensity infection <1% in all sentinel sites	Reduction of incidence of infection to zero		Incidence of infection remains zero (no autochthonous cases)
Group progression (1 to 4)	Up to 5–10 years from joining the group	Up to 3–6 years from joining the group	Up to 5 years from joining the group		Until all countries have interrupted transmission

Figure 1.5: WHO strategy to achieve disease elimination for both *S. mansoni* and *S. haematobium* taken from the schistosomiasis progress report by WHO in 2013 [WHO, 2013]

In 2015 coverage of school-aged children at risk was 42.2% while 57.5% of intervention units reached the targeted coverage of 75% [WHO, 2016]. Research focuses on what is needed to achieve elimination, such as surveillance methods for targeted treatment to maximise MDA efficacy, best timing to stop MDA and still achieve elimination, and control measures needed to complement MDA and new outbreaks [Truscott et al., 2014; Anderson et al., 2015; Colley et al., 2017; Secor, 2019; Toor et al., 2019].

The control of STHs was discussed in the 54th World Health Assembly as well resulting in

endorsement for preventive chemotherapy [World Health Assembly, 2001]. Annual coverage rates with MDA based on mebendazole and albendazole have increased considerably in recent years but have not reached the goal of 75%. Treatment guidelines were defined in a progress report by WHO with re-treatment twice a year above 50% prevalence, once a year between 20% and 50%, and no treatment below 20%. The goals were set modestly aiming at starting control programs in all endemic countries by 2015 and reaching 75% national as well as 100% geographical coverage [WHO, 2012].

1.5 Modelling diagnostic error

In disease diagnostics usage of multiple diagnostic tests with imperfect diagnostic sensitivity (i.e. the probability of a positive test in a positive individual), and specificity (i.e. the probability of a negative test in a negative individual), is common. Most diagnostic test do not only record binary data but continuous or discrete results that are then dichotomised, i.e. classified in positives and negatives, for further analysis. Results of diagnostic tests are also often not independent of each other but depend on disease characteristics like the number of worms in an infected individual.

When only binary data of the diagnostic tests is available, the common approach is the so-called latent class modelling. The approach is based on taking the disease status D_i of individual i as latent that is not directly measurable variable. The probability of a positive test is then depending on the disease status $P(T = 1 \mid D = 1) = S$ or $P(T = 1 \mid D = 0) = 1 - C$ where S is the sensitivity and C the specificity. For multiple tests Dendukuri et al. (2001) present two classes of models where the correlation between diagnostic tests is either modelled as a fixed effects model where for example $P(T_1 = 1, T_2 = 1 \mid D = 1) = S_1 S_2 + COV_{12}$ with an additional parameter COV_{12} or as an individual random effect [Dendukuri and Joseph, 2001; Wang et al., 2016]. The random effect can be interpreted as a latent disease characteristic increasing for example the probability of a positive result for both diagnostics and thereby introducing correlation. The model can readily be extended to larger number of diagnostic tests as long as the latent variable influencing the diagnostic test is common [Qu et al., 1996]. Latent class models are common in the analysis of diagnostic tests, see for example Menten et al. (2008)

for leishmaniasis, and Koukounari et al. (2009, 2013) for schistosomiasis and are often performed in a Bayesian framework [Menten et al., 2008; Koukounari et al., 2009, 2013].

When non-binary results of diagnostic tests are available it is beneficial to minimise lost information by modelling the continuous or discrete results directly. This thesis is concerned with diagnostic tests for helminth infections where the primary diagnostic techniques are parasitological, i.e. rely on the detection of eggs, and therefore the results of the diagnostic technique is in form of count data. The mean number of eggs recorded is related to the number of worms harboured by the patient and there are variations in egg excretion from day-to-day and from sample to sample that can be modelled [Levecke et al., 2014, 2015]. Non-parasitological diagnostic tests that are for example based on the detection of antigens or genetic material do likely have a sensitivity that depends on the number of worms harboured by an individual as well. Bayesian hierarchical statistical models provide flexible formulations to capture the uncertainties in the observed and latent processes. They also allow efficient computation via Markov chain Monte Carlo simulation algorithms [Carpenter et al., 2016]. Figure 1.6 shows a schematic of a hierarchical egg count model. De Vlas et al. (1992) recognized the potential of analysing variations in egg counts to infer on latent parameters like the 'true' disease prevalence in a population but the technology for implementation was not yet wide-spread [de Vlas et al., 1992, 1993].

In this thesis we developed a general modelling framework assuming that the population consists of a proportion of infected individuals p with a mean infection intensity μ in units of eggs per parasitological test. The infections are assumed to follow a continuous distribution in the population with α defining the aggregation of worms. In this thesis usually a shifted gamma distribution is used to take into account the fact that there has to be at least one worm-pair in any infection giving a non-zero infection intensity. Egg-excretion varies log-normally from day-to-day around a common mean λ_i related to the number of worm-pairs in individual i. Variation from one parasitological test to the next within the same sample are modelled as a negative binomial distribution taking into account the non-random distribution of eggs in the sample and variations due to technicians. A second binary diagnostic test can be understood in terms of the previously presented latent class model with random-effects. λ_i is the latent

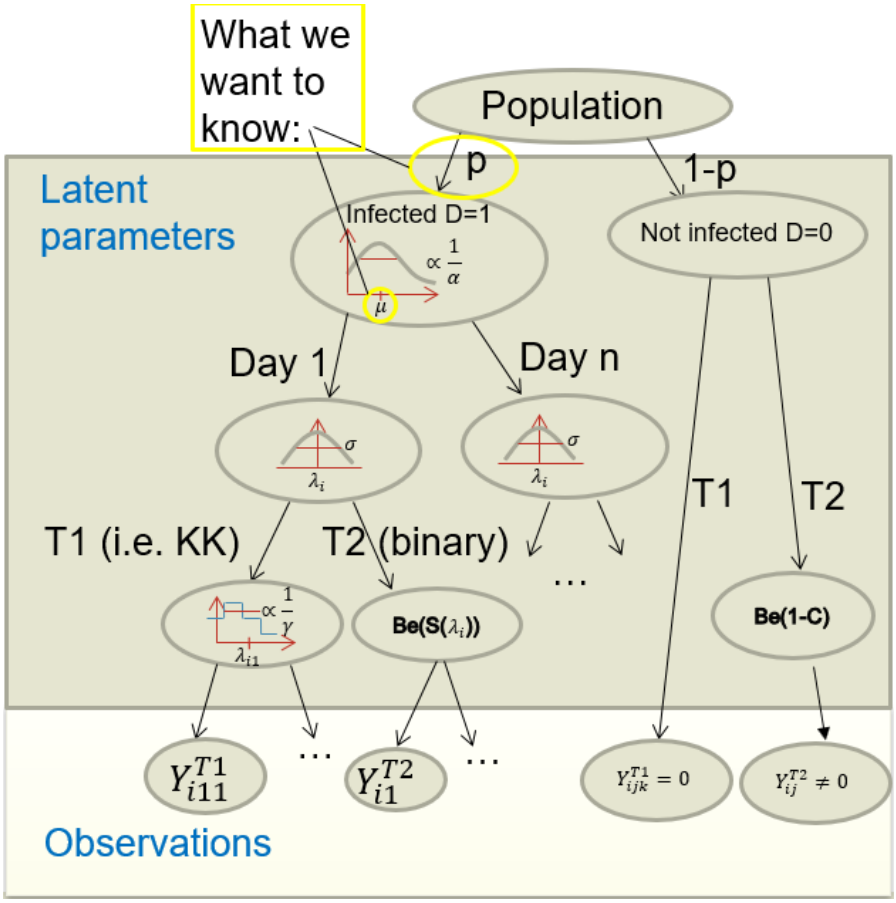


Figure 1.6: Schematic of a hierarchical model for Kato-Katz diagnostic on multiple days in parallel with an additional diagnostic test. p is the proportion of infected individuals in the population, μ the mean infection intensity of an infected individual in eggs per slide of the count test, e.g. Kato-Katz, α describes the aggregation of infection intensities in the population, σ the variation in egg excretion from day-to-day, γ defines the variation from egg count test slide to egg count test slide, and Y_{i11}^{T1} is the result of the first egg count test on the first sample from individual i. It is assumed that the alternative diagnostic test T2 is a Bernoulli process the sensitivity depends on the individual infection-intensity λ_i or the egg excretion on day j, λ_{ij}. Observations are on the lowest level while latent parameters of interest are linked to observations through multiple latent variables.

variable that modifies the sensitivity that could be interpreted as a random-effect if not identified by the egg count model. A parametric model for the sensitivity can be chosen to for example introduce tests with monotonically increasing sensitivity with infection intensity. The distribution of infection-intensities λ_i is in principle a zero-inflated shifted gamma distribution which is

related to the idea of zero-inflated models fitted directly to the egg count data [Wang et al., 2017].

1.6 Input from transmission models

The above framework enabled us to infer on latent variables describing the disease, 'true' prevalence of infections and mean infection intensity of an infected individual. However, it does not include any information about the relation of these parameters or for example dependence on the age of an individual. However, from mathematical models of disease transmission there is information available that can be incorporated in statistical models or into simulation studies [Gregory and Woolhouse, 1993; Gurarie and King, 2014].

Generally, it is assumed that schistosomiasis worms are distributed according to a negative binomial distribution in the population with parameters that are related to the number of individuals with at least one worm and the non-infected ones [Anderson and May, 1991; Grafen and Woolhouse, 1993]. Schistosomiasis worms pair up in monogamous couples when they reach their adult stage creating a distribution of worm-pairs in the population that cannot be described by a known count distribution [May, 1977; May and Woolhouse, 1993].

The prevalence of the disease increases from age zero till reaching an equilibrium stage at a certain age depending on transmission intensity where worm death is balanced with new infections [Woolhouse, 1998]. In older age groups prevalence can decrease again due to acquired immunity and changing behaviour [Chan et al., 1996; Chan, 1997]. There are many models describing potential relations between age and prevalence [Woolhouse, 1991, 1992]. However, a relatively simple model is sufficient to describe the age-prevalence relation in school-aged children before peak prevalence is achieved.

1.7 Objectives

The goal of this PhD thesis is to develop methods that allow integration of data from multiple imperfect diagnostic tests into a coherent model allowing estimation of the diagnostic accuracy of each of the tests as well as translation of observed prevalence from one diagnostic technique to another. We address the different parts of the goal in the following specific objectives:

1. estimate the diagnostic sensitivity in relation to infection intensity of the Kato-Katz diagnostic technique for *S. mansoni* and for different sampling schemes (Chapters 2, 3, 4);

2. estimate the relation between infection intensity and sensitivity of POC-CCA for *S. mansoni* determine equivalent prevalence thresholds to intervention thresholds based on the Kato-Katz diagnostic technique (Chapter 3);

3. estimate the 'true' prevalence of *S. mansoni* when only population summary measures of observed prevalence and mean infection intensity are available (Chapter 4);

4. evaluate the accuracy of survey designs based on POC-CCA or Kato-Katz for determining *S. mansoni* treatment needs (Chapter 5);

5. estimate sensitivity of a reagent strip for microhaematuria, a questionnaire, and visual inspection of urine to determine treatment needs for *S. haematobium* (Chapter 6);

6. estimate sensitivity of urine filtration and a reagent strip on multiple days to detect microhaematuria for *S. haematobium* (Chapter 7).

Chapter 2

Estimating sensitivity of the Kato-Katz technique for the diagnosis of *Schistosoma mansoni* and hookworm in relation to infection intensity

Authors

Oliver Bärenbold[1,2], Giovanna Raso[1,2], Jean T. Coulibaly[1,2,3,4], Eliézer K. N'Goran[3,4], Jürg Utzinger[1,2], Penelope Vounatsou[1,2]

1 Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland

2 Faculty of Science, University of Basel, Basel, Switzerland

3 Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, Abidjan, Côte d'Ivoire

4 Unité de Formation et de Recherche Biosciences, Université Félix Houphouët-Boigny, Abidjan,

Côte d'Ivoire

Published in *PLoS neglected tropical diseases*, 11(10), (2017): e0005953.

Abstract

The Kato-Katz technique is the most widely used diagnostic method in epidemiologic surveys and drug efficacy trials pertaining to intestinal schistosomiasis and soil-transmitted helminthiasis. However, the sensitivity of the technique is low, particularly for the detection of light-intensity helminth infections. Examination of multiple stool samples reduces the diagnostic error; yet, most studies rely on a single Kato-Katz thick smear, thus underestimating infection prevalence. We present a model, which estimates the sensitivity of the Kato-Katz technique in *Schistosoma mansoni* and hookworm, as a function of infection intensity for repeated stool sampling and provide estimates of the age-dependent 'true' prevalence. We find that the sensitivity for *S. mansoni* diagnosis is dominated by missed light infections, which have a low probability to be diagnosed correctly even through repeated sampling. The overall sensitivity strongly depends on the mean infection intensity. In particular, at an intensity of 100 eggs per gram of stool (EPG), we estimate a sensitivity of 50% and 80% for one and two samples, respectively. At an infection intensity of 300 EPG, we estimate a sensitivity of 62% for one sample and 90% for two samples. The sensitivity for hookworm diagnosis is dominated by day-to-day variation with typical values for one, two, three, and four samples equal to 50%, 75%, 85%, and 95%, respectively, while it is only weakly dependent on the mean infection intensity in the population. We recommend taking at least two samples and estimate the 'true' prevalence of *S. mansoni* considering the dependence of the sensitivity on the mean infection intensity and the 'true' hookworm prevalence by taking into account the sensitivity given in the current study.

Author summary

The World Health Organization (WHO) has defined a roadmap for schistosomiasis and soil-transmitted helminthiasis morbidity control and interruption of transmission with targets to be reached by 2025. Control efforts require reliable estimates of at-risk populations, number of

infections, and disease burden estimates in population subgroups in terms of age and location. Intervention guidelines are based on insensitive diagnostic techniques, such as the Kato-Katz method and do not take into account the effect of sampling effort and infection intensity. Our proposed methodology estimates the infection intensity-dependent sensitivity and the 'true' age-prevalence of the blood fluke *Schistosoma mansoni* and hookworm. We also provide recommendations on the number of stool samples required and the methodology to be used to reliably estimate the 'true' prevalence of parasitic worm infections.

Acknowledgements

This study received financial support from the European Research Council (ERC-2012-AdG-323180). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

2.1 Introduction

Soil-transmitted helminthiasis (STH) and schistosomiasis are two of the most prevalent neglected tropical diseases with more than 1 billion and over 250 million people affected worldwide, respectively [Hotez et al., 2014; Pullan et al., 2014]. Their collective global burden is 6 million disability-adjusted life years [GBD 2016 DALYs and HALE Collaborators, 2017], with school-aged children at the highest risk of associated morbidity. Control efforts have intensified over the past 15 years, with preventive chemotherapy serving as the main pillar [WHO, 2012, 2013, 2016].

Information about the spatial and temporal distribution of STH and schistosomiasis are important to guide interventions. Moreover, it is necessary to know which age groups contribute most to transmission, in terms of helminth egg output, in order to effectively use the available resources. There are two approaches to obtain this information. First, large-scale, standardized studies including all age groups with intense sampling which, however, is difficult to pursue due to the high cost. Second, reanalysis of data from previously published studies. Inference is hampered by the paucity of quality data [Hürlimann et al., 2011; Karagiannis-Voules et al., 2015; Lai et al., 2015]. Individual-level data are usually not reported; instead, only the number of participants tested positive for a specific helminth infection is considered. In addition, diagnosis relies on the Kato-Katz technique [Katz et al., 1972], i.e., counting of helminth eggs in a small amount of stool. This approach, however, has a low, setting-dependent sensitivity, which is governed by variation in day-to-day production of eggs per worm, non-random distribution of eggs within a stool sample, decay of eggs in the sample due to methods and duration of the experimental procedure, transportation, and storage [Hall, 1981; Bundy et al., 1992; de Vlas and Gryseels, 1992; Booth et al., 2003; Krauth et al., 2012]. Collecting multiple stool samples over consecutive days increases the accuracy but there are no guidelines on the optimal number of samples [Utzinger et al., 2001; Knopp et al., 2010]. Consequently, the comparison of studies that employed different sampling efforts, which is necessary for monitoring progress of control programs, is hampered.

Statistical modelling can help studying the age-prevalence and its dependence on the diagnos-

tic error but has been restricted by the aforementioned limitations that compromise the quality of the data [Brooker et al., 2009; Hürlimann et al., 2011]. Although the qualitative shape of helminthiasis age-prevalence curves is known, there has been little progress in the application of quantitative transmission models, especially for STH infections [Holford and Hardy, 1975; Anderson, 1986; Chan et al., 1994; Yang and Yang, 1998]. Furthermore, the dependency of the intensity of the infection on the diagnostic sensitivity has been largely neglected. The negative binomial distribution has commonly been used to fit helminth egg count data. For example, De Vlas and Gryseels [de Vlas and Gryseels, 1992] and Levecke et al. [Levecke et al., 2015] proposed models that separate the measurement process from the true underlying infection intensity distribution. However, none of the preceding models is able to infer on the dependence of the sensitivity of the Kato-Katz method for repeated stool sampling on infection intensity.

We developed a model for faecal egg count (FEC) data, which quantifies the relation between sampling effort, infection intensity, and diagnostic sensitivity. The model separates the infected from the non-infected individuals and the measurement process from the infection status. Variability due to egg output, experimental conditions, and aggregation within a population are taken into account. We calculate the 'true' prevalence and other biological and transmission-related parameters based on the probability of false-negatives. Our model improves estimation of the age-related disease burden and provides inputs for mathematical transmission models.

2.2 Materials and methods

2.2.1 Ethics statement

This study consists of a secondary analysis of published data. Ethics approval, written informed consent procedures, and treatment of infected individuals have been described elsewhere [Utzinger et al., 2000a,b; Raso et al., 2004; Coulibaly et al., 2013a].

2.2.2 Data

We tested our model performing a secondary analysis of individual level FEC data from three separate studies in medium and high transmission settings in Côte d'Ivoire conducted in 1998,

2002, and 2011, respectively. All data used were from baseline surveys with no previous mass drug administration in the area. The studies took place in Fagnampleu [Utzinger et al., 2000a,b], Zouatta [Raso et al., 2004], and Azaguié [Coulibaly et al., 2013a]. Based on the Kato-Katz assay, hookworm prevalence varied from 11.4% to 59.0%, and mean or infected population based infection intensity from 280 eggs per gram of stool (EPG) to 396 EPG. For *S. mansoni*, prevalence varied from 35.6% to 76.3%, and infection intensity from 152 EPG to 307 EPG. Between two and four stool samples were collected and analysed on consecutive days of a total of 1423 participants. Azaguié and Zouatta were surveys performed in the full age range from 0 to 90 years, while Fagnampleu only included school-aged children. Prevalence of *Ascaris lumbricoides* and *Trichuris trichiura* were too low to be analysed. Summary measures are included in Table 2.1, a more detailed description can be found in the appendix A1, and the individual level data used is included in the published version of the paper.

2.2.3 Model

We utilized a hierarchical Bayesian model to address the objectives given in the introduction. Let Y_{ij} be the FEC, i.e., the number of helminth eggs found in sample j in individual i, k_i the number

Table 2.1: Resulting parameters of the simulations (mean and 95% BCI[1]) parameters of the mean infection and of the prevalence of *S. mansoni* and hookworm in Azaguié, Zouatta, and Fagnampleu in Côte d'Ivoire. Observed prevalence is the ratio of positively tested individuals in the original study, observed mean infection the arithmetic mean egg count of the individuals with a positive test.

	S. mansoni			**Hookworm**		
	Azaguié	**Zouatta**	**Fagnampleu**	**Azaguié**	**Zouatta**	**Fagnampleu**
Parameters	(N=500, k=2)[2]	(N=559, k=3)	(N=354, k=4)	(N=500, k=2)	(N=559, k=3)	(N=354, k=4)
Observed prevalence (%)	35.6 (31.4, 39.8)	40.8 (36.7, 44.9)	76.3 (71.8, 80.5)	11.4 (8.6, 14.2)	35.4 (31.5, 39.4)	59.0 (53.9, 64.1)
Estimated 'true' prevalence (%)	49.3 (40.4, 61.2)	59.6 (50.7, 69.3)	83.8 (78.3, 89.3)	14.3 (10.9, 18.5)	43.7 (38.6, 49.2)	62.2 (56.6, 67.6)
Observed mean infection (EPG)[3]	179 (171, 188)	152 (141, 163)	307 (289, 325)	396 (326, 466)	331 (301, 361)	283 (260, 306)
Estimated mean infection (EPG)[4]	132 (101, 167)	104 (84, 128)	282 (249, 321)	220 (150, 316)	261 (208, 325)	262 (221, 312)
Sensitivity[5]						
1 sample (%)	59.4 (47.6, 70.2)	48.0 (40.8, 55.8)	70.2 (66.1, 74.1)	57.1 (44.5, 68.8)	47.1 (41.7, 52.5)	53.8 (50.1, 57.7)
2 samples (%)	72.9 (59.5, 84.6)	62.3 (53.5, 71.3)	83.5 (79.3, 87.3)	81.0 (69.1, 90.1)	71.8 (65.9, 77.3)	78.5 (74.59, 81.9)
3 samples (%)	-	69.0 (59.8, 78.2)	88.2 (84.2, 91.8)	-	84.9 (80.0, 89.1)	89.9 (87.3, 92.2)
4 samples (%)	-	-	90.7 (86.8, 94.0)	-	-	95.2 (93.6, 96.6)
Day-to-day variation (r)	1.10 (0.80, 1.46)	0.83 (0.67, 1.02)	0.87 (0.77, 0.99)	0.25 (0.15, 0.37)	0.15 (0.13, 0.19)	0.20 (0.17, 0.23)
Aggregation (α)	0.09 (0.05, 0.13)	0.08 (0.05, 0.11)	0.05 (0.04, 0.07)	0.22 (0.04, 0.90)	0.32 (0.06, 1.23)	0.19 (0.05, 0.68)

[1] Parameter posterior mean estimates and 95% Bayesian credible interval
[2] k is the number of samples taken in a study; N the number of participants with at least one sample
[3] The observed mean infection is the mean egg count of all individuals tested positive in at least one Kato-Katz thick smear
[4] The estimated mean infection is $\mu_f + \mu_m$
[5] The sensitivity was calculated using equation 2.4

of stool samples from individual i, and x_i the age of individual i.

We assumed that a population consists of a proportion of infected individuals p, i.e., people that carry at least one pair of worms, and that of uninfected individuals. Thus p is interpreted as the prevalence. Each infected individual has a characteristic infection intensity λ_i, measured in units of mean eggs per sample and assumed to be distributed within the population according to a shifted gamma distribution, given by

$$\begin{aligned} \lambda_i &= v_i + \mu_m \\ v_i &\sim \text{Gamma}(\mu_f \cdot \alpha, \alpha) \\ &= \frac{\alpha^{\mu_f \cdot \alpha}}{\Gamma(\mu_f \cdot \alpha)} \, v_i^{\mu_f \cdot \alpha - 1} \, \exp(-\alpha \cdot v_i) \end{aligned} \tag{2.1}$$

with a mean number of eggs $\mu_f + \mu_m$ in an infected individual, variance $\frac{\mu_f}{\alpha}$ that corresponds to the aggregation of infection intensities, and hence, the aggregation of worms within the population. The shift parameter μ_m is the mean number of eggs per sample that can be expected from an individual carrying exactly one female worm and thus the minimal possible infection intensity. Direct inference on the worm load is not possible in this frame as the dependence on mean egg output is non-linear and not well known [Anderson and Schad, 1985].

The process of taking k_i samples from an infected individual i with infection intensity λ_i is modelled by a negative binomial distribution with mean λ_i and a variance given by $\lambda_i + \lambda_i^2/r$. r reflects the additional variation, due to changes in the day-to-day helminth eggs output, the aggregation of eggs in stool, and the precise experimental procedure but not the within-population variation which is given by α. If a perfectly random distribution of the eggs and perfect measurement is assumed $r \to \infty$, the measurement process becomes a Poisson process. By including the uninfected, the model is written as

$$\mathbf{Y_i} \sim \begin{cases} (1-p) + p \cdot \text{NB}(0, \lambda_i, r)^{k_i} & \text{, if } I_i = 0 \\ p \cdot \prod_{j=1}^{k_i} \text{NB}(Y_{ij}, \lambda_i, r) & \text{, if } I_i = 1 \end{cases} \tag{2.2}$$

which corresponds to a zero-inflated negative binomial model with p corresponding to the mixing proportion. NB is the negative binomial distribution and I_i the result of the Kato-Katz test over

all samples from an individual. They are defined as follows:

$$I_i = \begin{cases} 0 & \text{, if } \max(\mathbf{Y_i}) = 0 \\ 1 & \text{, if } \max(\mathbf{Y_i}) \neq 0 \end{cases} \tag{2.3}$$

$$\mathrm{NB}(y, \lambda, r) = \binom{y + r - 1}{y} \left(\frac{\lambda}{\lambda + r}\right)^y \left(\frac{r}{r + \lambda}\right)^r$$

False-negatives are included in the model as repeated zero measurements for an infected individual. Thus, the sensitivity depending on the number of repeated measurements becomes

$$s_i[k_i] = 1 - \mathrm{NB}(0, \lambda_i, r)^{k_i} = 1 - \left(\frac{r}{\lambda_i + r}\right)^{k_i \cdot r} \tag{2.4}$$

where s is the sensitivity, and k and λ vary for each individual.

Low-rank thin-plate splines are used to study the age dependence of p, μ_f, r, and α. For a detailed derivation of the spline model, see Crainiceanu et al. [Crainiceanu et al., 2005]. The representation of p is

$$\mathrm{logit}(p_i) = \beta_0 + \beta_1 \, x_i + \sum_{m=1}^{M} u_m \, (x_i - \kappa_m)^3 \tag{2.5}$$

where $\mathbf{\Theta}_1 = (\beta_0, \beta_1, u_1, \ldots, u_M)^T$ is the vector of regression coefficients and $\kappa_1 < \cdots < \kappa_M$ are the fixed knots. The other parameters can be represented analogously using logarithmic or linear spline models. The spline regression makes only a few very general assumptions about the shape of the curve, e.g., continuity and differentiability, and is therefore able to infer without requiring prior knowledge about the biology of the process, i.e., the transmission model.

The minimum eggs per sample μ_m is fixed to the average egg output of a worm divided by an average amount of faeces per day, multiplied by the weight of a sample. For hookworm, μ_m is 5 eggs in a sample which corresponds to 120 EPG and for *S. mansoni*, μ_m is 0.03 eggs which corresponds to 0.72 EPG [Cheever et al., 1994; Despommier et al., 2012]. We choose the following semi-informative priors for the model: gamma for r with mean 1 and variance 1; normal for $\log(\alpha)$ with mean 0 and variance 1; gamma for μ_f with mean 2 and variance 4; normal for $\beta_{0,1}$ with mean 0 and variance 1; normal for $u_{1,\ldots,M}$ with mean 0 and variance τ, where τ is distributed as a gamma with mean 2 and variance 4. The results were not sensitive to the specific shape of the priors.

Bayesian inference was performed using Markov chain Monte Carlo (MCMC) simulations implemented in Stan [Carpenter et al., 2016]. Validity of the model was checked using simulated data. Models with splines on p, μ_f, r, and α were run to check for age dependence. μ_f, r, and α showed no significant age dependence and were set as independent of age for the simulations presented in the results section. μ_m was varied from 1 egg to 6 eggs for hookworm and from 0.01 eggs to 0.1 eggs for *S. mansoni*, which also showed no significant influence. The lower limit of 0.01 eggs per slide for *S. mansoni* corresponds to roughly 100 eggs in 500 g of stool. The upper limit of 0.1 eggs per slide corresponds to 1000 eggs per 500 g of stool therefore any value larger than 0.1 is most likely unrealistic for a single worm pair. The model was run with 25 chains, with 20,000 iterations each, of which 2,000 were used as warm up and adaption, for each study and for each of the two infections separately. Convergence was achieved, and assessed using Gelman + Rubin diagnostics and visual inspection of the chains [Gelman and Rubin, 1992].

2.3 Results

We applied a Bayesian hierarchical model to FEC data from three studies carried out in Zouatta, Azaguié and Fagnampleu in Côte d'Ivoire, as described in the "Materials and methods" section. Parameter posterior mean estimates and 95% Bayesian credible interval (BCI) are summarized in Table 2.1.

2.3.1 Estimated 'true' prevalence and its relation to age

The three studies are from different hookworm transmission settings with observed prevalence of 11.4% and mean infection intensity of an infected individual of 396 EPG for Azaguié, 35.4% and 331 EPG for Zouatta, and 59.0% and 283 EPG for Fagnampleu. Based on our model, we estimated the 'true' hookworm prevalence at 14.3% (95% BCI 10.9–18.5%), 43.7% (95% BCI 38.6–49.2%), and 62.2% (95% BCI 56.6–67.6%), for Azaguieé, Zouatta, and Fagnampleu, respectively. The estimated mean infection intensity does not significantly differ from one study to another and mean estimates ranged from 220 EPG to 262 EPG (see Table 2.1). Age-prevalence curves in Fig. 2.1 from the three studies show similar features such as a steep increase from

birth until an equilibrium is reached at ages of around 20 years for Zouatta, and 45 years for Azaguié. The prevalence stays constant till an age of about 60 years from where the rate of infection declines. For Fagnampleu only the initial steep increase is visible due to the fact that no individuals older than 15 years were included.

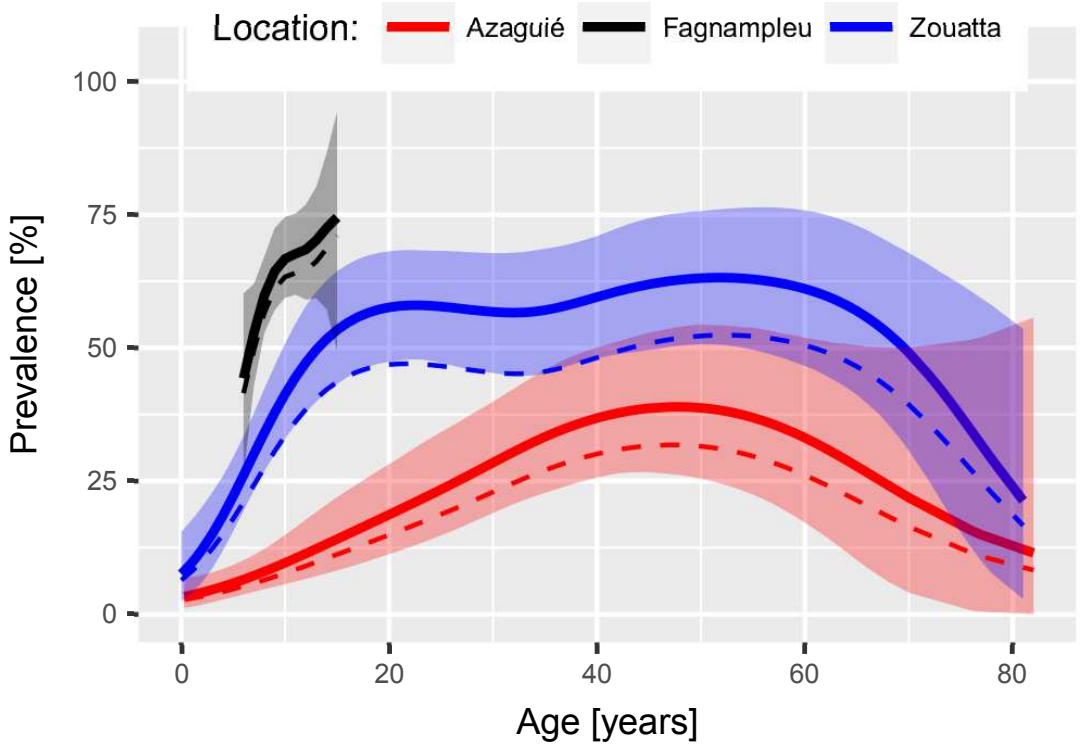


Figure 2.1: Prevalence of hookworm infection in relation to age for the three studies in Côte d'Ivoire, including 95% BCI indicated by shaded areas, and observed prevalence as dashed line.

For *S. mansoni*, the Azaguié and Zouatta studies show similar transmission levels with an observed prevalence of 35.6% and 40.8% and observed mean infection intensity of 179 EPG and 152 EPG, respectively. In contrast, the study in Fagnampleu had a prevalence of 76.3% and a mean infection intensity of 307 EPG. We estimated a 'true' prevalence of 49.3% (95% BCI 40.4–61.2%) and a mean infection intensity of 132 EPG (95% BCI 101–167 EPG) for Azaguié, 59.6% (95% BCI 50.7–69.3%) and 104 EPG (95% BCI 84–128 EPG) for Zouatta, and 83.8% (95% BCI 78.3–89.3%) and 282 EPG (249–321 EPG) for Fagnampleu. The estimated age-prevalence curves displayed in Fig. 2.2 show similar qualitative features. The prevalence

increases up to a peak between the ages of 15 and 20 years, and subsequently declines slowly up to an age of 60 years, followed by a stronger decrease. The lower prevalence after the peak is not significant but it appears both in the Azaguié and Zouatta data.

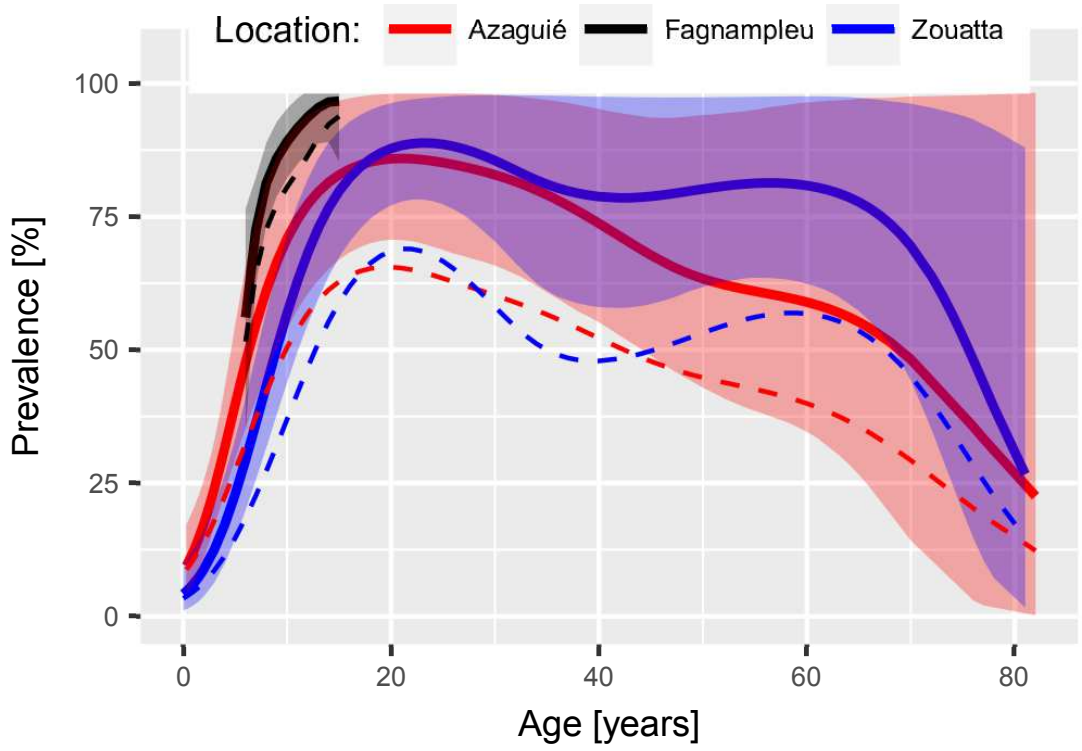


Figure 2.2: Prevalence of *S. mansoni* infection in relation to age for three studies in Côte d'Ivoire, including 95% BCI indicated by shaded areas, and observed prevalence as dashed line.

2.3.2 Variations in egg output

For hookworm, the day-to-day variation given by r is consistent across study sites, ranging from 0.15 (95% BCI 0.13–0.17) to 0.25 (95% BCI 0.15–0.37) (see Table 2.1), indicating strong overdispersion. The aggregation of egg output within the population is also consistent across the studies with α estimates ranging from 0.19 (95% BCI 0.05–0.68) to 0.32 (95% BCI 0.06–1.23).

For *S. mansoni* the day-to-day variation is consistent across studies and significantly different from hookworm with values ranging from 0.83 (95% BCI 0.67–1.02) to 1.10 (95% BCI 0.80–1.46). The aggregation of infections within the population shows no significant differences

between studies with α ranging from 0.05 (95% BCI 0.04–0.07) to 0.09 (95% BCI 0.05–1.13), which indicates a significantly higher variance than for hookworm.

2.3.3 Estimated sensitivity and its relation to infection intensity

For hookworm, the estimates of the diagnostic sensitivity of Kato-Katz did not vary between locations. Based on a single Kato-Katz thick smear, sensitivity estimates were in the range of 47% to 57%, for two samples obtained from different days from 72% to 81%, for three samples estimates were within the range of 85% to 90%, and for four samples around 95%. For *S. mansoni*, data from Azaguié and Zouatta revealed similar sensitivity estimates within the range of 48% to 59% for one Kato-Katz thick smear, 62% to 73% for two samples, and 69% for three samples. Fagnampleu has a higher sensitivity of 70%, 84%, 88%, and 91% for one, two, three, and four samples, respectively (see Table 2.1).

The sensitivity of the Kato-Katz technique for different infection intensities was calculated using equation 2.4 and is plotted in Fig. 2.3. For hookworm the dependence on infection intensity is weak, e.g., only increasing from 40% to 55% from a very light infection of 120 EPG to a still light infection of 500 EPG. For moderate and heavy infections (>2000 EPG) the sensitivity did not significantly improve with infection intensity. However, the sensitivity can be greatly increased by examining several stool samples, e.g., for an infection intensity of 360 EPG the sensitivity can raise from 50% based on a single sample to 75% for two samples, and 92.5% for three samples.

The sensitivity was strongly associated with *S. mansoni* infection intensity. In particular, for very light infections (<5 EPG), it was below 50% even after three samples. For light infections (<100 EPG), it was still heavily dependent on infection intensity. For moderate infections (100–399 EPG), two samples gave a high sensitivity above 90%. Heavy infections (>400 EPG) were reliably detected (i.e. >99%) by testing two samples.

Fig. 2.4 shows the overall sensitivity in a population with a day-to-day variation of $r = 1.0$ and a population aggregation of $\alpha = 0.07$ as a function of the mean infection intensity in the population. For lower transmission settings with 100 EPG comparable to Zouatta, the sensitivity after four samples is still below 75%. However, sensitivity rose to more than 95% for a setting

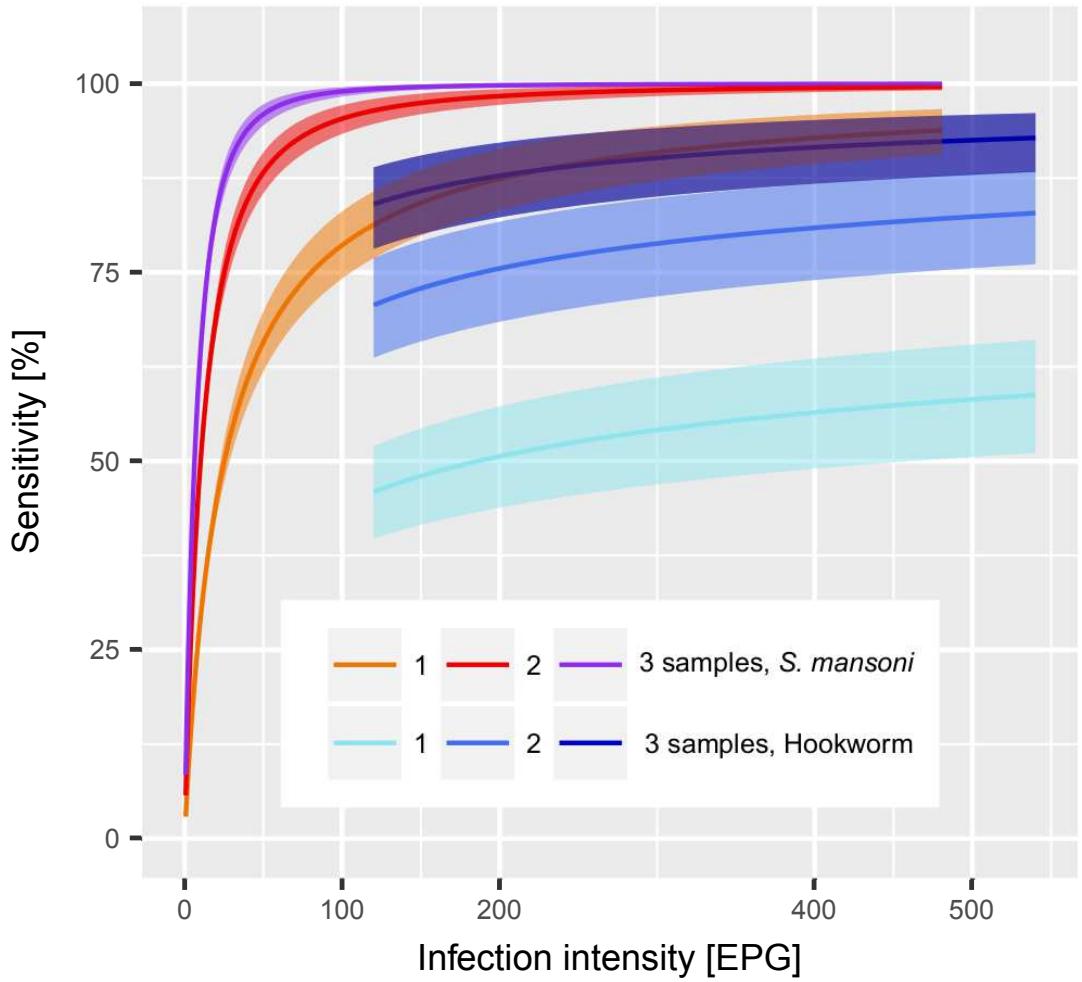


Figure 2.3: Relation between *S. mansoni* and hookworm infection intensity with Kato-Katz diagnostic sensitivity, including 95% BCI for k equal to 1, 2, and 3 samples per individual.

with a mean infection intensity of over 300 EPG.

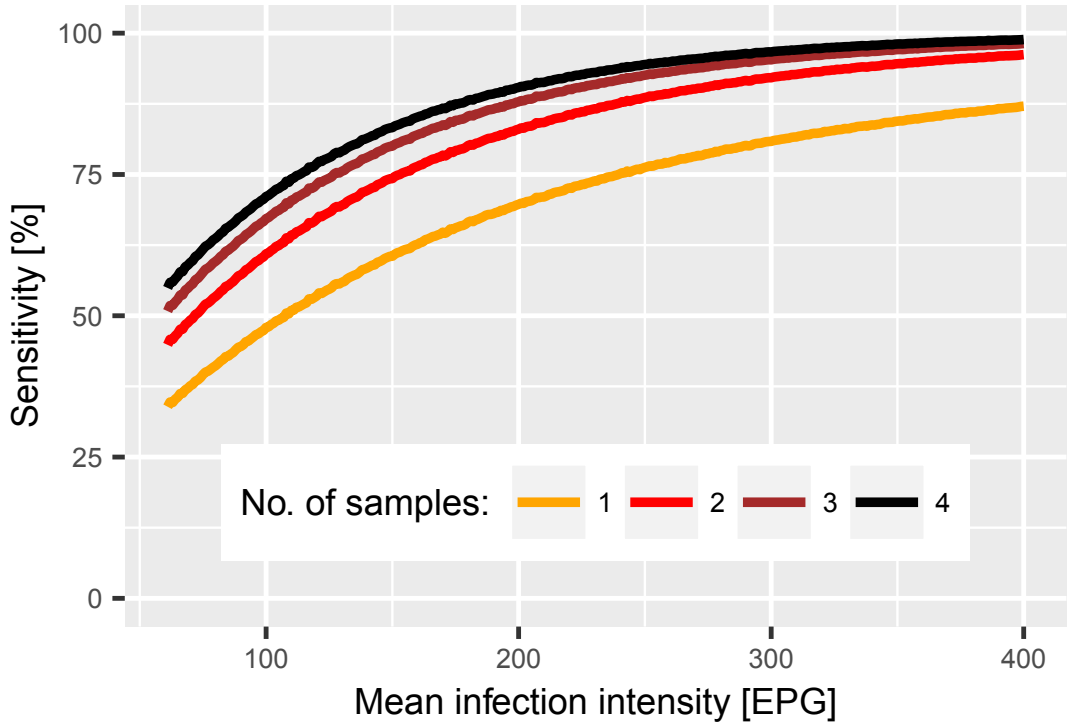


Figure 2.4: Estimated relation between sensitivity and arithmetic mean *S. mansoni* infection in a population with $\alpha = 0.07$ and $r = 1.0$ for one to four samples.

2.4 Discussion

We present a model which determines the relation between the sensitivity of the Kato-Katz technique and intensity of *S. mansoni* and hookworm infections. The model takes into account day-to-day variations in helminth egg output and within population aggregation of worms. Additionally, we were able to test various parameters for age-dependence, especially the age structure of the prevalence.

The overall sensitivity point estimates for hookworm corroborate with estimates derived from latent class modelling approaches. Tarafder et al. [Tarafder et al., 2010] give an estimate of 65% sensitivity in a setting with 35% prevalence where no infection intensity is given, and Nikolay et al. [Nikolay et al., 2014] estimated the sensitivity for one sample at 59.5%, for two at 74.2%, and for three samples at 74.3%. For *S. mansoni* in a setting with an observed prevalence of

33%, Lamberton et al. [Lamberton et al., 2014] predict a sensitivity of 29.6% for one sample, 51.9% for two samples, 70.4% for three samples, and 77.8% for four samples in agreement with our estimates for Zouatta and Azaguié. The same authors predicted a much higher sensitivity of 83.5%, 97.8%, and 100% for one, two, or three samples, respectively in a setting with 95% prevalence in agreement with our estimates for Fagnampleu. Glinz et al. [Glinz et al., 2010] also predicted a sensitivity of 70% for a single Kato-Katz thick smear in a high transmission setting with prevalence close to 90%, which is again consistent with our estimates for Fagnampleu.

Our estimates are in agreement with those given in the literature. However, the added value of our modelling approach is that we can predict the sensitivity depending on infection intensity and the 'true' prevalence in various settings and we can understand the factors influencing the sensitivity for hookworm and *S. mansoni*.

There are three main parameters that directly influence the sensitivity, i.e., the minimum infection intensity, the day-to-day variation, and population aggregation. These parameters are specific for *S. mansoni* and hookworm. The minimum infection μ_m, which is fixed to 0.03 eggs per sample for *S. mansoni* and 5 eggs per sample for hookworm, is based on estimates of egg output for a single worm. The estimates carry a large uncertainty, which were shown to be non-critical for the results. The low output of eggs per worm in *S. mansoni* makes it almost impossible to detect light infections with only a few female worms. The low sensitivity for light infections can also be seen in Fig. 2.3 where the three *S. mansoni* curves show very low sensitivity when approaching 0 EPG. For hookworm, the output of a single female worm is about 5 eggs per sample, which already leads to a reasonable probability for detection.

The parameter r is determined by the day-to-day variation. The three studies agree on a value between 0.17 and 0.19 for hookworm and between 0.8 and 0.99 for *S. mansoni*. Both values conclusively contradict a Poisson process which has $r \to \infty$ and could be interpreted as worms producing and excreting eggs randomly. Thus, our model clearly indicates that a worm produces eggs in a clustered fashion. A single pair of *S. mansoni* produces in the order of 100 eggs per day [Cheever et al., 1994], while a female hookworm sheds around 10,000 eggs [Despommier et al., 2012]. Thus, an infection that manifests itself with a similar egg count for both diseases indicates a much higher number of *S. mansoni* compared to hookworm. This difference in worm

numbers has important implications, because the variation in egg output of a single worm is partly averaged out over the population of worms that produce eggs independently in an individual. The much lower variation in day-to-day output of *S. mansoni* can therefore be explained by the, on average, larger number of worms in an infected individual. The larger variation for hookworm leads to a lower sensitivity of the Kato-Katz technique for this helminth species for a similar infection intensity. Thus, the increase in sensitivity due to repeated sampling is much larger for hookworm (see Fig. 2.3).

The population aggregation parameter α describes the distribution of worm output within the population. A small value indicates a relatively wider distribution, while a larger value indicates a more uniform distribution. The estimate of α has larger uncertainty for hookworm compared to *S. mansoni*, most likely due to the large day-to-day variation for the former species. Furthermore, α is overall larger for hookworm compared to *S. mansoni* indicating that egg output is more evenly distributed among individuals infected with hookworm than for *S. mansoni*. These results are in agreement with findings by Krauth et al. [Krauth et al., 2012].

The severity of disease burden at a location is generally given by two parameters, prevalence and intensity of infections. If negative individuals are included in the calculation of the mean infection intensity, the high number of zeros will skew the mean downwards, especially in lower prevalence settings. Thus, a strong correlation to the prevalence will be induced, making the mean infection intensity to be a mixed measure of prevalence and infection intensity. We decided to only include positive individuals in the mean infection intensity to separate the effects of the sensitivity on observed prevalence from that on infection intensity. Therefore, we decided to estimate the mean infection intensity from only positive individuals.

It is evident from our estimates that the sensitivity of the Kato-Katz technique in hookworm is dominated by the large day-to-day variation in egg output and has only a weak dependence on infection intensity. Thus, increasing the number of samples is an effective strategy to increase sensitivity even in low transmission settings for hookworm. In contrast, for *S. mansoni* the infections that give false-negative results are largely those with light intensity. Increasing the number of samples to more than two does only marginally improve the sensitivity because the low density of eggs and not the variation in excretion and production limit the sensitivity. However,

taking two thick smears from the same samples at 100 EPG mean infection intensity would increase the sensitivity from 50% to 70% for a comparably low additional effort. For mean infection intensities below 100 EPG no directly representative data was included in this model study. Nevertheless, extrapolations to low prevalence and infection intensity settings are valid because the model considers individual-level data. In a high intensity setting there are still many individuals with low intensity infections, and therefore our model includes information on the full range of infection intensities. Thus, inference about a low intensity and prevalence population consisting primarily of light infections is possible without making unreasonable extrapolations. For infections with an intensity above 240 EPG, the strong relation between infection intensity and sensitivity suggests that the sensitivity for two samples is close to 100%, ensuring that the most heavily infected individuals are detected and can be treated. Still, the results indicate a possibility of bias when comparing different study sites. The infection intensity-dependent sensitivity will become increasingly important in the new era with the goal to eliminate STH and schistosomiasis [Rollinson et al., 2013; Utzinger et al., 2015].

In the study sites of Zouatta and Azaguié, the observed mean infection of *S. mansoni* was moderate compared to Fagnampleu. Hence, it is conceivable that there was a comparably larger share of light infections that were missed due to the low sensitivity of the Kato-Katz technique, and therefore, the 'true' mean infection intensity will be even lower. Moreover, the difference between estimated and observed prevalence is also larger due to the higher number of missed cases. For Fagnampleu, the observed mean infection intensity and therefore the overall sensitivity are high. Thus, the estimates agree with the observations and the 'true' prevalence is only slightly higher than the one observed. For hookworm, the likelihood of correctly diagnosing a heavy infection is still larger than for a light infection. Accordingly, the difference between the observed and the estimated mean infection intensity is larger the fewer the number of samples taken.

Our estimates of the underlying 'true' age-prevalence for *S. mansoni* and for hookworm are in agreement with those obtained from transmission models [Holford and Hardy, 1975; Anderson and May, 1991] and from latent class statistical models [Raso et al., 2007]. For example, the *S. mansoni* age-prevalence is comparable to those obtained by the Yang et al. [Yang, 2003] models, which differentiate between the influence of water contact patterns and the acquired

immunity. However, due to large uncertainty in older age groups, our results do not allow choosing the transmission model, which resembles best. For hookworm, a peak shift, as proposed by Woolhouse [Woolhouse, 1998], is clearly visible in our age-prevalence curve. The decreasing prevalence at older age is apparent in all locations but it has yet to be discussed in greater detail in the literature. It could indicate a significantly lower life expectancy for infected individuals although other explanations and confounding factors are conceivable but cannot be tested with the available data.

2.4.1 Conclusion

The proposed model succeeds in predicting the intensity-dependent sensitivity of the Kato-Katz technique directly from the day-to-day variation in helminth egg output. Hence, the model is able to explain the differences between the sensitivity of hookworm and *S. mansoni*. The sensitivity of Kato-Katz for hookworm is dominated by a high day-to-day variation. We recommend collecting at least two stool samples over subsequent days combined with the given sensitivity values to estimate 'true' prevalence. For *S. mansoni* infection the sensitivity is largely driven by light infections that are hard to detect by a single Kato-Katz thick smear. We also recommend collecting two samples due to almost perfect sensitivity for moderate and heavy infections and low benefit of additional samples for light infections. We predict that improving the sensitivity for *S. mansoni* can be achieved more cost effectively by increasing the number of Kato-Katz thick smears from the same stool sample instead of increasing the number of samples taken. Additionally, it is necessary to take into account the infection intensity-dependent sensitivity of Kato-Katz for *S. mansoni* when comparing data from several studies. Including the infection dependence becomes more important when close to elimination due to the larger changes in sensitivity of Kato-Katz with infection intensity.

A further consequence of the results is due to the fact that the guidelines of WHO are defined in terms of observed prevalence. An observed prevalence of e.g. 10% for *S. mansoni*, which is the lower limit for yearly MDA, is indicative of a 'true' prevalence of roughly 14%, 20%, and 29% for 200 EPG, 100 EPG, and 50 EPG, respectively. Hence, the observed prevalence is a measure of both, the 'true' prevalence and the infection intensity. We advise the disentanglement

of these two components by defining thresholds separately for 'true' prevalence and infection intensity. The results also suggest that the current disease burden estimates underestimate the true prevalence.

The spline model for age-dependence used in this study can be replaced by appropriate transmission models to determine which age groups should be treated and how frequently that has to happen to increase the intervention effectiveness. The model can be further extended to analyse studies with multiple Kato-Katz thick smears performed per stool sample and thus separate day-to-day from within-sample variation. This would enable us to address the question of how repeated testing of the same sample compares to taking several samples in order to reduce cost and increase compliance.

Appendix

A1 data description

Zouatta. This dataset was obtained from a study carried out in Zouatta II in the district of Man, western Côte d'Ivoire in 2002, a region where *S. mansoni* is known to be endemic. Stool and urine samples were obtained from 561 individuals over three consecutive days. 545 individuals provided at least 2 stool samples with similar numbers of male and female participants. The age range was from 5 days to 91 years with a median of 17 years and a mean of 23 years. Single 41.7 mg Kato-Katz thick smears were prepared from each stool sample and examined under a light microscope. Eggs of *S. mansoni*, hookworm, *Ascaris lumbricoides*, and *Trichuris trichiura* were counted and recorded separately. 40.9% of the participants delivered at least one positive sample for *S. mansoni* and 35.3% for hookworm.

Azaguié. The study was carried out in the two villages, Azaguié Makouguié and Azaguié M'Bromé in the district of Azaguié in south Côte d'Ivoire between June and September 2011. *S. mansoni* and *S. haematobium* are known to co-occur in this area. Stool samples were collected over two consecutive days and two Kato-Katz thick smears were prepared from each sample. 501 individuals completed the study (241 males and 260 females). The age ranged from 3 months to 82 years, with a mean age of 16.3 years, and a median age of 6 years. The overall prevalence for

S. mansoni was 42.7%, and that of hookworm was 14.2%.

Fagnampleu. Data were obtained from the initial diagnosis from a clinical trial done between November 1998 and July 1999 in the region of Man in western Côte d'Ivoire. *S. mansoni* is known to be endemic in the region and there has been no control campaign before the study. Stool samples were taken and Kato-Katz thick smears prepared and analysed for soil-transmitted helminths and *S. mansoni* eggs over four consecutive days from 354 schoolchildren. The age range was from 6 to 15 years, the mean 9.4 years, and the median 9 years. The overall prevalence was 76.3% for *S. mansoni* and 59.0% for hookworm.

Chapter 3

Translating preventive chemotherapy prevalence thresholds for *Schistosoma mansoni* from the Kato-Katz technique into the point-of-care circulating cathodic antigen diagnostic test

Authors

Oliver Bärenbold[1,2], Amadou Garba[3], Daniel G. Colley[4], Fiona M. Fleming[5], Ayat A. Haggag[6], Reda M. R. Ramzy[7], Rufin K. Assaré[1,2,8,9], Edridah M. Tukahebwa[10], Jean B. Mbonigaba[11], Victor Bucumi[12], Biruck Kebede[13], Makoy S. Yibi[14], Aboulaye Meité[15], Jean T. Coulibaly[1,2,8,9],

Eliézer K. N'Goran[8,9], Louis-Albert Tchuem Tchuenté[16,17], Pauline Mwinzi[18], Jürg Utzinger[1,2], Penelope Vounatsou[1,2]

1 Swiss Tropical and Public Health Institute, Basel, Switzerland

2 University of Basel, Basel, Switzerland

3 Department of Control of Neglected Tropical Diseases, World Health Organization, Geneva, Switzerland

4 Center for Tropical and Emerging Global Diseases and Department of Microbiology, University of Georgia, Athens, GA, United States of America

5 Schistosomiasis Control Initiative, Imperial College, London, United Kingdom

6 Ministry of Health and Population, Cairo, Egypt

7 National Nutrition Institute, General Organisation for Teaching Hospitals and Institutes, Cairo, Egypt

8 Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, Abidjan, Côte d'Ivoire

9 Unité de Formation et de Recherche Biosciences, Université Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire

10 Vector Control Division, Ministry of Health, Kampala, Uganda

11 Ministry of Health, Kigali, Rwanda

12 Programme National Intégré de Lutte contre les Maladies Tropicales Négligées et la Cécité au Burundi, Bujumbura, Burundi

13 Ministry of Health, Addis Ababa, Ethiopia

14 Neglected Tropical Disease Department, Ministry of Health, Juba, South Sudan

15 Programme National de Lutte contre les Maladies Tropicales Négligées à Chimiothérapie Préventive, Ministère de la Santé et de l'Hygiène Publique, Abidjan, Côte d'Ivoire

16 Laboratory of Parasitology and Ecology, University of Yaoundé I, Yaoundé, Cameroon

17 Centre for Schistosomiasis and Parasitology, Yaoundé, Cameroon

18 Centre for Global Health Research, Kenya Medical Research Institute, Nairobi, Kenya

Published in *PLoS neglected tropical diseases*, 12(6), (2018): e0006562.

Abstract

Background

Intervention guidelines against *Schistosoma mansoni* are based on the Kato-Katz technique. However, Kato-Katz thick smears show low sensitivity, especially for light-intensity infections. The point-of-care circulating cathodic antigen (POC-CCA) is a promising rapid diagnostic test detecting antigen output of living worms in urine and results are reported as trace, 1+, 2+, and 3+. The use of POC-CCA for schistosomiasis mapping, control, and surveillance requires translation of the Kato-Katz prevalence thresholds into POC-CCA relative treatment cut-offs. Furthermore, the infection status of egg-negative but antigen-positive individuals and the intensity-dependent sensitivity of POC-CCA should be estimated to determine its suitability for verification of disease elimination efforts.

Methodology

We used data from settings in Africa and the Americas characterized by a wide range of *S. mansoni* endemicity. We estimated infection intensity-dependent sensitivity and specificity of each test at the unit of the individual, using a hierarchical Bayesian egg count model that removes the need to define a 'gold' standard applied to data with multiple Kato-Katz thick smears and POC-CCA urine cassette tests. A simulation study was carried out based on the model estimates to assess the relation of the two diagnostic tests for different endemicity scenarios.

Principal findings

POC-CCA showed high specificity ($> 95\%$), and high sensitivity ($> 95\%$) for moderate and heavy infection intensities, and moderate sensitivity ($> 75\%$) for light infection intensities, and even for egg-negative but antigen-positive infections. A 10% duplicate slide Kato-Katz thick smear prevalence corresponded to a 15–40% prevalence of $\geq$ trace-positive POC-CCA, and

10–20% prevalence of $\geq$ 1+ POC-CCA. The prevalence of $\geq$ 2+ POC-CCA corresponded directly to single slide Kato-Katz prevalence for all prevalence levels.

Conclusions/significance

The moderate sensitivity of POC-CCA, even for very light *S. mansoni* infections where the sensitivity of Kato-Katz is very low, and the identified relationship between Kato-Katz and POC-CCA prevalence thresholds render the latter diagnostic tool useful for surveillance and initial estimation of elimination of *S. mansoni*. For prevalence below 10% based on a duplicate slide Kato-Katz thick smear, we suggest using POC-CCA including trace results to evaluate treatment needs and propose new intervention thresholds that need to be validated in different settings.

Author summary

The World Health Organization (WHO) has defined goals for schistosomiasis morbidity control to be reached by 2025 that are based on preventive chemotherapy. Intervention thresholds for *Schistosoma mansoni* are currently defined for prevalence measured by stool microscopy using the Kato-Katz technique. However, the Kato-Katz technique shows low sensitivity, particularly for the detection of light-intensity infections. Replacing it with the semi-quantitative point-of-care circulating cathodic antigen (POC-CCA) urine cassette test requires translation of the thresholds and precise characterization of the diagnostic sensitivity and specificity. In this study, we applied a novel egg count model to a suite of data obtained from different settings in Africa and the Americas with diverse endemicity levels. We used a simulation study to infer on the relation between Kato-Katz and POC-CCA prevalence. Based on our study, we were able to provide recommendations for POC-CCA thresholds taking into account semi-quantitative results of the test. We found that a *S. mansoni* prevalence of 10% based on duplicate slide Kato-Katz thick smear is equivalent to 15–40% POC-CCA prevalence when trace results are considered positive and to 10–20% POC-CCA prevalence when trace results are considered negative. Our results have important bearings for mapping, control, surveillance, and verification of elimination of intestinal schistosomiasis.

Acknowledgements

This study received financial support from the European Research Council (ERC-2012-AdG-323180). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We acknowledge the SCORE secretariat and their advisory board for their input and encouragement. The research team of the Centre for Schistosomiasis and Parasitology in Yaoundé, Cameroon and the Ministry of Health survey teams in Burundi, Rwanda and Uganda are acknowledged for their contributions to data collection and the respective donors of the Bill & Melinda Gates Foundation, The END Fund, and UKAID.

3.1 Introduction

Schistosomiasis is a neglected tropical disease (NTD) caused by parasitic flatworms of the genus *Schistosoma* that require freshwater snails as intermediate hosts to complete their life cycle [Colley et al., 2014]. Currently, around 250 million people are infected, primarily school-aged children in rural areas of sub-Saharan Africa, and, to a lesser extent in South-east Asia, and the Americas, causing an estimated 2.1 million disability-adjusted life years (DALYs) in 2016 [Hotez et al., 2014; GBD 2016 DALYs and HALE Collaborators, 2017]. There are six species infecting humans, of which *Schistosoma haematobium*, *S. japonicum*, and *S. mansoni* are the most common ones [Colley et al., 2014].

This study discusses the diagnosis of *S. mansoni* whose adult stage live in pairs within the mesenteric venules. Experiments in mice showed that female worms shed in the order of several hundred eggs daily with a large variability. Eggs are either trapped in the intestines and liver causing inflammatory reactions, or are excreted through faeces continuing the life cycle [Cheever et al., 1994]. Morbidity, among others, includes anaemia, growth stunting, impaired cognition, increased susceptibility to other infections (e.g., HIV), and severe pathologies of the liver and spleen [van der Werf et al., 2003; King et al., 2005; Colley et al., 2014; Ezeamama et al., 2018].

The World Health Organization (WHO) has articulated a road-map for morbidity control, and elimination of the disease in suitable settings, by 2025 based on preventive chemotherapy using praziquantel administered to school-aged children and other high-risk communities, control of intermediate host snails, and behavioural and environmental changes [WHO, 2013; Lo et al., 2017]. To achieve these goals, intervention guidelines were set that define the communities requiring preventive chemotherapy and treatment frequency. Enacting these guidelines requires estimates of the disease prevalence, which are obtained from survey data. Transmission levels may vary over short geographic scales (2–5 km), and thus, having the ability to sample large numbers of communities with rapid, low cost, and sensitive tests are critical for efficient and effective disease control [Stothard et al., 2014; Utzinger et al., 2015; Knowles et al., 2017].

Infections with *S. mansoni* can be diagnosed using various techniques to detect eggs in stool. The most widely used method is the Kato-Katz technique based on a thick smear, usually with a

volume equivalent to 41.7 mg [Katz et al., 1972]. Of note, the Kato-Katz technique has been used to define intervention thresholds for preventive chemotherapy [Montresor et al., 1998]. While faecal egg detection techniques have generally high specificity, they suffer from low sensitivity, especially for light-intensity infections, which lead to underestimation of community prevalence and burden [King, 2015; Bärenbold et al., 2017]. To reduce the diagnostic error, repeated sampling of stool and preparation of multiple Kato-Katz thick smears from a single stool specimen is recommended, which is impractical and expensive for disease control purposes [Booth et al., 2003]. Furthermore, Kato-Katz detects only infections with mature, egg-shedding worms. To overcome these shortcomings, novel techniques that are not based on egg detection are required.

A promising candidate is a test that detects a specific antigen in urine; the point-of-care circulating cathodic antigen (POC-CCA) [van Lieshout et al., 2000]. Diagnosis takes about 20 min and the outcome is reported semi-quantitatively; namely, trace, 1+, 2+, and 3+, although interpretation may depend on the laboratory technician. Systematic reviews estimate the sensitivity to be around 90%. POC-CCA classifies individuals as positive that were classified as negative by Kato-Katz due to its low sensitivity for light infection intensities. Hence, the specificity of POC-CCA is underestimated when Kato-Katz is considered as diagnostic 'gold' standard [Ochodo et al., 2015; Danso-Appiah et al., 2016]. Colley et al. performed regression analysis to estimate the relation between POC-CCA and single Kato-Katz thick smears and indicated the need for further investigation to determine how it depends on different sampling schemes and levels of endemicity [Colley et al., 2013]. Converting Kato-Katz prevalence thresholds from existing treatment guidelines into POC-CCA analogues requires information on the variation of the sensitivity of POC-CCA with infection intensity. Moreover, the prevalence of egg-negative but antigen- or worm-positive infections and its influence on the prevalence measured by POC-CCA in comparison with Kato-Katz testing has to be evaluated.

In this study, we addressed the aforementioned issues using pairs of data with simultaneous testing of faecal samples by Kato-Katz and urine samples by POC-CCA from various settings in Africa and the Americas, characterized by a wide range of endemicity levels. We developed a model, which infers on the infection intensity-dependent sensitivity profile of POC-CCA for semi-quantitative test results without using an artificial 'gold' standard. Additionally, we estimated the

specificity of POC-CCA and the prevalence of egg-negative/antigen-positive infections. Model outputs were employed in a simulation study to obtain insights on the relation between measured Kato-Katz and POC-CCA prevalence, assuming a range of scenarios with various infection intensities and prevalence of egg-negative infections. Our findings translate Kato-Katz based *S. mansoni* prevalence thresholds put forth in the current WHO intervention guidelines into POC-CCA prevalence and enable the latter diagnostic approach for mapping, disease control, and surveillance.

3.2 Materials and methods

3.2.1 Ethics statement

The data included in this study were published elsewhere or are currently in the process of being published [Coulibaly et al., 2011; Shane et al., 2011; Tchuem Tchuenté et al., 2012; Colley et al., 2013; Coulibaly et al., 2013b; Erko et al., 2013; Adriko et al., 2014; Mwinzi et al., 2015; Haggag et al., 2017; Ortu et al., 2017; Assare et al., 2018]. Hence, ethics approval, written informed consent procedures, and treatment of infected individuals are given in the respective studies where the original data were or are being published.

3.2.2 Data

We analysed a suite of 30 datasets with simultaneous Kato-Katz and POC-CCA results available at individual level. A description of the data related to Kato-Katz and POC-CCA results is given in Tables 3.1 and 3.2, respectively. The data were grouped in four categories depending on their characteristics. In particular, a number of datasets (group 1) from Cameroon, Côte d'Ivoire, Ethiopia, Kenya, and Uganda have at least duplicate Kato-Katz readings on two different days and at least one POC-CCA urine cassette test result. These datasets were used in the most complex egg count model to estimate infection intensity-dependent sensitivity of both diagnostic methods and can be found in in the published version of the paper. Data from Ecuador and Ethiopia with binary Kato-Katz results (i.e., egg-positive or egg-negative; group 2) were collected at locations known to be non-endemic and therefore include information about the specificity of POC-CCA.

Datasets from Burundi, Côte d'Ivoire, Rwanda, and South Sudan, with only duplicate Kato-Katz slides from one stool sample (group 3) did not provide information on the day-to-day variation. Datasets from Egypt (group 4) with single Kato-Katz readings of binary outcome were collected from locations with very low transmission. Data from groups 3 and 4 were used in the simulation study.

3.2.3 Statistical model

The Kato-Katz results either were available as binary or egg count measurements. Hence, we developed separate but interlinked Bayesian hierarchical models for each type of data to include all available information in the evaluation of POC-CCA.

Table 3.1: Summary of the Kato-Katz results of the databases

Country	**Location**	**Date** (year)	**Age range** (years)	$\mathbf{z_d}$ [1]	$\mathbf{z_s}$ [2]	$\mathbf{N_{KK}}$	**P. KK** (%)	μ **KK** (EPG)	**geom.** μ (EPG)	**P 1 KK** (%)
Group 1										
Cameroon	Makenene [Tchuem Tchuenté et al., 2012]	2010	6-16	3	3	251	71.7	161	43.3	41.7 (34.8, 48.7)
Cameroon	Njombe[Tchuem Tchuenté et al., 2012]	2010	8-16	3	3	245	63.3	173	27.5	30.6 (26.9, 34.3)
Cameroon	Yaounde[Tchuem Tchuenté et al., 2012]	2010	7-14	3	3	233	27.9	235	40.9	16.5 (12.3, 20.7)
Côte d'Ivoire	Man[Assare et al., 2018]	2016	9-13	2	3	695	6.5	72	22.0	3.8 (2.6, 4.9)
Côte d'Ivoire	1 [Coulibaly et al., 2013b]	2011	0.2-5.5	2	2	109	25.7	90	37.0	16.5 (12.3, 20.8)
Côte d'Ivoire	2 [Coulibaly et al., 2013b]	2011	0.2-5.5	2	2	133	21.1	122	30.8	11.7 (9.1, 14.3)
Côte d'Ivoire	1 [Coulibaly et al., 2011]	2010	8-12	3	3	170	91.7	525	248.0	70.2 (62.4, 78.1)
Côte d'Ivoire	2 [Coulibaly et al., 2011]	2010	8-12	3	3	130	53.1	116	36.8	24.5 (14.8, 34.3)
Côte d'Ivoire	3 [Coulibaly et al., 2011]	2010	8-12	3	3	146	32.9	50	8.5	8.3 (3.1, 13.5)
Ethiopia	Harbu[Erko et al., 2013]	2010	8-12	3	2	300	57.0	69	31.0	33.1 (24.8, 41.4)
Ethiopia	Jiga[Erko et al., 2013]	2010	8-12	3	2	320	49.4	153	70.9	35.8 (32.1, 39.5)
Kenya	[Shane et al., 2011]	2007	1-15	3	2	1,845	22.1	106	32.1	11.4 (7.7, 15.2)
Uganda	1 [Adriko et al., 2014]	2010	7-13	3	2	100	55.0	240	34.2	29.3 (24.0, 34.6)
Uganda	2 [Adriko et al., 2014]	2010	7-13	3	2	100	54.0	122	33.3	29.8 (23.6, 36.1)
Uganda	3 [Adriko et al., 2014]	2010	7-13	3	2	100	31.0	37	19.8	14.9 (9.7, 20.1)
Uganda	4 [Adriko et al., 2014]	2010	7-13	3	2	100	35.0	247	58.0	21.1 (16.8, 25.4)
Uganda	5 [Adriko et al., 2014]	2010	7-13	3	2	100	12.0	58	28.4	6.8 (3.8, 9.8)
Uganda	Baseline	2013	6-16	3	2	775	6.3	48	22.0	3.1 (1.5, 4.7)
Uganda	Follow-up	2015	6-16	3	2	659	4.2	68	33.5	2.7 (1.5, 3.9)
Uganda	Mapping	2013	9-14	3	2	711	3.8	182	26.9	1.8 (1.0, 2.6)
Group 2										
Ecuador	[Mwinzi et al., 2015]	2014	6-16	1	1	144	0	-	-	-
Ethiopia	[Colley et al., 2013]	2010	8-12	1	1	100	0	-	-	-
Group 3										
Burundi	[Ortu et al., 2017]	2014	12-16	1	2	8,482	1.5	56	34.4	1.2 (1.1, 1.3)
Côte d'Ivoire	All		6-15	1	2	11,449	8.0	267	80.3	6.1 (5.6, 6.6)
Rwanda	All	2014		1	2	8,695	2.0	84	52.0	1.7 (1.5, 2.0)
South Sudan	All		10-14	1	2	5,649	7.1	128	54.1	5.7 (5.1, 6.3)
Group 4										
Egypt	Gov 1 [Haggag et al., 2017]	2016	6-15	1	1	3,000	3.5	-	-	-
Egypt	Gov 2 [Haggag et al., 2017]	2016	6-15	1	1	5,000	1.7	-	-	-
Egypt	Gov 3 [Haggag et al., 2017]	2016	6-15	1	1	2,946	0.1	-	-	-
Egypt	Gov 4 [Haggag et al., 2017]	2016	6-15	1	1	974	0.4	-	-	-
Egypt	Gov 5 [Haggag et al., 2017]	2016	6-15	1	1	2,997	0.1	-	-	-

[1] z_d is the number of stool specimens taken on different days
[2] z_s is the number of Kato-Katz thick smears prepared by a single stool specimens

3.2.3.1 Egg count model

We extended our previous work modelling Kato-Katz egg count data without the need of an artificial 'gold' standard to include POC-CCA results and estimated the POC-CCA infection intensity-dependent sensitivity [Bärenbold et al., 2017]. Similar models have recently been applied by Bottomley et al. (2016)[Bottomley et al., 2016] to model the diagnostic sensitivity for *Onchocerca volvulus*, and by Prada et al. (2018) [Prada et al., 2018] to compare POC-CCA to Kato-Katz diagnostics for *S. mansoni*.

Let Y_{jids}^{KK} be the Kato-Katz egg count of individual i in study j on day d when reading slide s, and Y_{jirk}^{CCA} the binary result of the POC-CCA reading based on proxy k from individual i in

Table 3.2: Summary of POC-CCA results of databases employed for the current modeling study to translate Kato-Katz to POC-CCA prevalence intervention thresholds

				P. CCA	(%)				
Country	**Location**	$\mathbf{z_r}$ [1]	$\mathbf{N_{CCA}}$	**Tr+** [2]		**Tr−** [3]		$\mathbf{2-3+}$ [4]	
Group 1									
Cameroon	Makunene	3	270	85.2	75.9 (68.2, 83.7)	60.4	52.1 (48.7, 55.5)	-	-
Cameroon	Njombe	3	270	87.8	75.2 (73.0, 77.4)	55.9	43.1 (34.4, 51.7)	-	-
Cameroon	Yaounde	3	237	72.1	50.8 (45.7, 55.9)	24.1	17.2 (13.6, 20.7)	-	-
Côte d'Ivoire	Man	1	700	32.7	-	20.4	-	-	-
Côte d'Ivoire	1	2	109	81.7	67.0 (61.8, 72.2)	44.0	35.8 (22.8, 48.8)	27.5	22.0 (66.7, 76.4)
Côte d'Ivoire	2	2	109	72.2	58.3 (50.8, 65.7)	45.9	32.0 (24.5, 39.4)	24.1	17.7 (8.1, 27.2)
Côte d'Ivoire	1	3	170	-	-	86.5	83.3 (81.5, 85.1)	77.6	69.4 (66.3, 72.5)
Côte d'Ivoire	2	3	130	-	-	51.5	40.5 (27.4, 53.6)	23.1	15.9 (13.5, 18.2)
Côte d'Ivoire	3	3	146	-	-	34.2	23.1 (22.3, 23.9)	6.8	5.0 (3.4, 6.6)
Ethiopia	Harbu	3	300	80.0	71.2 (65.0, 77.4)	-	-	-	-
Ethiopia	Jiga	3	320	62.5	59.6 (58.6, 60.5)	-	-	-	-
Kenya		3	1,845	74.4	53.3 (51.0, 55.5)	-	-	11.6	7.7 (6.8, 8.6)
Uganda	1	1	100	70.0	-	52.0	-	28.0	-
Uganda	2	1	100	74.0	-	56.0	-	22.0	-
Uganda	3	1	100	65.0	-	52.0	-	20.0	-
Uganda	4	1	100	56.0	-	46.0	-	20.0	-
Uganda	5	1	100	48.0	-	35.0	-	7.0	-
Uganda	Base	3	775	33.7	21.0 (19.0, 22.9)	13.4	8.5 (6.5, 10.5)	5.2	3.1 (1.0, 5.3)
Uganda	F1	3	659	37.0	21.2 (13.9, 28.6)	19.0	11.4 (8.7, 14.2)	2.4	1.6 (0.7, 2.4)
Uganda	Mapping	3	711	19.0	11.2 (10.3, 12.2)	7.3	4.6 (3.8, 5.5)	2.0	1.6 (1.3, 1.8)
Group 2									
Ecuador		1	144	0	-	-	-	-	-
Ethiopia		1	100	1	-	-	-	-	-
Group 3									
Burundi		1	8,482	41.3	-	10.9	-	-	-
Côte d'Ivoire	All	1	11,453	20.9	-	-	-	-	-
Rwanda	All	1	8,695	37.5	-	8.6	-	-	-
South Sudan	All	1	5,649	41.5	-	-	-	-	-
Group 4									
Egypt	Gov 1	1	3,000	17.6	-	-	-	-	-
Egypt	Gov 2	1	5,000	9.4	-	-	-	-	-
Egypt	Gov 3	1	2,946	4.6	-	-	-	-	-
Egypt	Gov 4	1	974	9.7	-	-	-	-	-
Egypt	Gov 5	1	2,997	12.3	-	-	-	-	-

[1] z_r is the number of POC-CCA tests performed on different days
[2] $\mathbf{Tr+}$ is the prevalence when trace, 1+, 2+, and 3+ results are considered positive
[3] $\mathbf{Tr-}$ is the prevalence when 1+, 2+, and 3+ results are considered positive
[4] $\mathbf{2-3+}$ is the prevalence when 2+ and 3+ results are considered positive

study j on repeat r. Three proxies have been considered to convert the semi-quantitative reading into a binary outcome and estimate the equivalent Kato-Katz prevalence; i.e., $k = 1$ treats all positive results, including trace, as positive; $k = 2$ takes 1+, 2+, and 3+ results as positive; and $k = 3$ groups only 2+ and 3+ results as positive. We assumed that the population in study j consists of a proportion of infected individuals that are egg-positive p_j with individual disease status $D_{ji} = 2$, the proportion of antigen-positive individuals that is egg-negative l_j with disease status $D_{ji} = 1$, and the uninfected individuals $D_{ji} = 0$ with proportion $1 - p_j - l_j$. We assumed that the diagnostic results conditional on the disease status and infection intensity are independent. This means that we assume that the sensitivity of a diagnostic test for an individual is fully determined by the egg-density in stool and it is independent of other individual level factors. Let $\mathbf{Y}_{ji} = (Y_{ji11}, Y_{ji12}, \ldots, Y_{jids})^T$.

$$P(\mathbf{Y}_{jik}^{KK}, \mathbf{Y}_{ji}^{CCA}) = \sum_{Z=0}^{2} P(\mathbf{Y}_{jik}^{KK} \mid D_{ji} = Z) \cdot P(\mathbf{Y}_{ji}^{CCA} \mid D_{ji} = Z) \cdot P(D_{ji} = Z) \tag{3.1}$$

For the non-infected individuals ($D_{ji} = 0$), with infection intensity $\lambda_{ji} = 0$, the egg count Y_{jids}^{KK} is modelled by a negative binomial distribution, and the POC-CCA binary results Y_{jirk}^{CCA} by a binomial distribution, as follows:

$$\begin{aligned} P(\mathbf{Y}_{ji}^{KK} \mid D_{ji} = 0) &\equiv \prod_{d,s} NB(\mu = \gamma_1 \cdot ((c^{KK})^{-1/\gamma_1} - 1), \gamma_1) \\ P(\mathbf{Y}_{jik}^{CCA} \mid D_{ji} = 0) &\equiv \prod_{r} Be(1 - c_k^{CCA}) \end{aligned} \tag{3.2}$$

where c^{KK} is the specificity of the Kato-Katz technique, which is assumed to be the same for each study, γ_1 is the dispersion parameter of the false positives, and c_k^{CCA} is the specificity of the POC-CCA test for proxy k.

For egg-positive individuals ($D_{ji} = 2$), we assumed that each individual has a mean infection intensity λ_{ji} that is related to the number of worm pairs. Daily variations in egg-output are described using a mean egg output λ_{jid} for day d. We assumed that the Kato-Katz result for the infected individuals on day d and slide s are independent, conditional on the infection intensity λ_{ji} for individual i and on the day-to-day variation ϵ_{jid}, and modelled by a negative binomial

distribution.

$$
\begin{aligned}
P(\mathbf{Y}_{ji}^{KK} \mid D_{ji} = 2) &\equiv \prod_{d,s} NB(\lambda_{jid} + \mu_{min}, \gamma_{2j}) \\
\log(\lambda_{jid}) \mid \lambda_{ji}, \epsilon_{jid} &= \log(\lambda_{ji}) + \epsilon_{jid} \\
\epsilon_{jid} &\sim N(0, \sigma_j^2) \\
\lambda_{ji} &\sim \text{Gamma}(\mu_{f,j} \cdot \alpha_j, \alpha_j)
\end{aligned}
\tag{3.3}
$$

α_j models the variation within the population of infected individuals in study j, σ_j^2 captures the extent of the day-to-day variation ϵ_{jid}, γ_{2j} takes into account the non-random distribution of eggs within a sample, μ_{min} is the minimum possible infection intensity corresponding to one pair of worms, and $\mu_{f,j}$ corresponds to the mean infection intensity of infected individuals in study j. False-negatives are included in the model as repeated zero measurements, e.g., the sensitivity for a single Kato-Katz reading becomes

$$
s_{jid}^{KK} = 1 - \text{NB}(Y_{jids}^{KK} = 0; \lambda_{jid}, \gamma_{2j}) = 1 - \left(\frac{\gamma_{2j}}{\lambda_{jid} + \gamma_{2j}}\right)^{\gamma_{2j}} \tag{3.4}
$$

Y_{jirk}^{CCA} is modeled by a binomial distribution with the sensitivity of POC-CCA, s_{jik}^{CCA}, dependent on the infection intensity of the individual, given by

$$
\begin{aligned}
P(Y_{jirk}^{CCA} \mid D_{ji} = 2) &\equiv Bn(s_{jik}^{CCA}) \\
s_{jik}^{CCA}(\lambda_{ji}) &= \text{logit}^{-1}(a_{0kj} + a_{1kj} \cdot \sqrt{\lambda_{ji}}) \cdot a_{2kj}
\end{aligned}
\tag{3.5}
$$

a_{0kj} determines the sensitivity of a very light infection intensity in study j, a_{1kj} describes the dependence of the sensitivity on the infection intensity, and a_{2kj} determines the limit of the sensitivity for severe infections.

The infected but non egg-shedding individuals $D_{ji} = 1$ are modelled similar to the uninfected ones but with differing parameters for POC-CCA.

$$
\begin{aligned}
P(Y_{jids}^{KK} \mid D_{ji} = 1) &\equiv NB(\mu = \gamma_1 \cdot (c_{KK}^{-1/\gamma_1} - 1), \gamma_1) \\
P(Y_{jirk}^{CCA} \mid D_{ji} = 1) &\equiv Be(s_{jik}^{CCA}(\lambda_{ji} = 0))
\end{aligned}
\tag{3.6}
$$

where $s_{jik}^{CCA}(\lambda_{ji} = 0)$ is the sensitivity of POC-CCA for egg-negative infections, which is considered to be the same as for egg shedding infections with an infection intensity equal to zero

eggs per gram of stool (EPG). Kato-Katz readings are assumed to have the same distribution as for non-infected individuals.

The parameters that are related to the biology of the worms, to transmission behavior or to the diagnostic technique and expected to be related between studies, i.e., the within-population variation α_j, the day-to-day variation σ_j^2, the slide-to-slide variation γ_j, the parameters determining the sensitivity of POC-CCA, namely a_{0kj}, a_{1kj}, and a_{2kj}, were partially pooled, using a common mean and a normally distributed random effect on the log scale.

3.2.3.2 Binary latent class model

Binary POC-CCA data from locations with zero Kato-Katz positive individuals were included to extract information about the specificity of POC-CCA. The model was formulated as follows:

$$Y_{ikr}^{CCA} \sim B(s_{jik}^{CCA}(\lambda_{ji} = 0)\pi + (1 - c_k^{CCA})(1 - \pi)) \tag{3.7}$$

where $s_{jik}^{CCA}(\lambda_{ji} = 0)$ is the sensitivity of POC-CCA for egg-negative infections, used as a lower boundary, and π the prevalence in the given setting.

3.2.3.3 Implementation details

The aforementioned model was fitted using Markov chain Monte Carlo (MCMC) simulations in Stan version 2.16.2 (Stan Development Team; http://mc-stan.org) with 25 chains consisting of 500 warm-up and 500 sampling steps [Carpenter et al., 2016]. Subsequent analyses and simulations were performed using R version 3.4.1 (The R Foundation for Statistical Computing; Vienna, Austria) and RStudio version 1.0.143 (RStudio, Inc.; Boston, United States of America).

An informative Beta prior was considered for the specificity of the Kato-Katz technique c^{KK} with mean 0.98 and standard deviation (SD) 0.01. We assumed a rather informative truncated normal prior with mean 0.03 and SD 0.01 for the infection intensity of one pair of worms μ_{min}. The mean was chosen by assuming an average egg output of a pair of worms in the order of 100 eggs per day, multiplied with the ratio between the weight of a Kato-Katz sample of 41.7 mg and a daily production of faeces of 150 g which corresponds to about 0.03 eggs per sample or 0.72 EPG for *S. mansoni* [Cheever et al., 1994]. For the dispersion of eggs detected in a non-infected

individual γ_1, we assumed a truncated normal prior with mean and variance of 1 to ensure that the typical false positive has a low egg count. The mean infection intensity of an infected individual in study j was calculated as follows:

$$\mu_j = \mu_{f,j} \cdot e^{\sigma_j^2/2} + \mu_{min} \tag{3.8}$$

Semi-informative or non-informative priors were adopted for the rest of the parameters: a uniform distribution between 0 and 1 for c_k^{CCA}, gamma distributions for σ_j, α_j, and γ_{2j} with mean 1 and SD 1, gamma distribution with mean 10 and SD 10 for $\mu_{f,j}$, normal distribution for a_0 with mean 0 and SD 2, gamma distribution with mean 5 and SD 5 for a_1, normal distribution with mean 2 and SD 2 on the logit of a_2, and uniform priors between 0 and 1 for all prevalence parameters.

3.2.4 Simulation study

Using the parameter estimates described in our statistical model, we simulated single and duplicate slide Kato-Katz thick smears and POC-CCA prevalences under 84 different scenarios. The simulations were carried out with a large number of individuals per scenario to avoid variations due to sampling. We assumed a wide range of egg-positive infections (i.e., 5%, 10%, 20%, 30%, 50%, and 70%) to mimic different endemicity scenarios. Prevalence of egg-negative infections was varied independently from the egg-positive prevalence and assumed to be 5%, 10%, 20%, and 30%. Mean infection intensity of an egg-positive infected individual was set to 50 EPG, 100 EPG, 200 EPG, and 400 EPG.

3.3 Results

3.3.1 Characterization of POC-CCA

Our model estimates the sensitivity and specificity in relation to the 'true' prevalence of infection in the population. This avoids the use of a 'gold' standard, which does not exist. Posterior estimates of parameters that influence the sensitivity and specificity can be found in Table A1 in the appendix.

Based on these parameters, Fig 3.1 shows the infection-dependent sensitivity of POC-CCA. All three binary proxies of POC-CCA indicate a very high sensitivity for moderate and heavy infections ($\geq$ 100 EPG). The sensitivity for light infections differ between proxies: while including trace results as positives enables detection of very light infections, considering $2+$ and $3+$ results as positives leads to missing of many infected individuals. The uncertainty in the estimated sensitivity is largest in the proxy considering trace results as negative (1+/2+/3+).

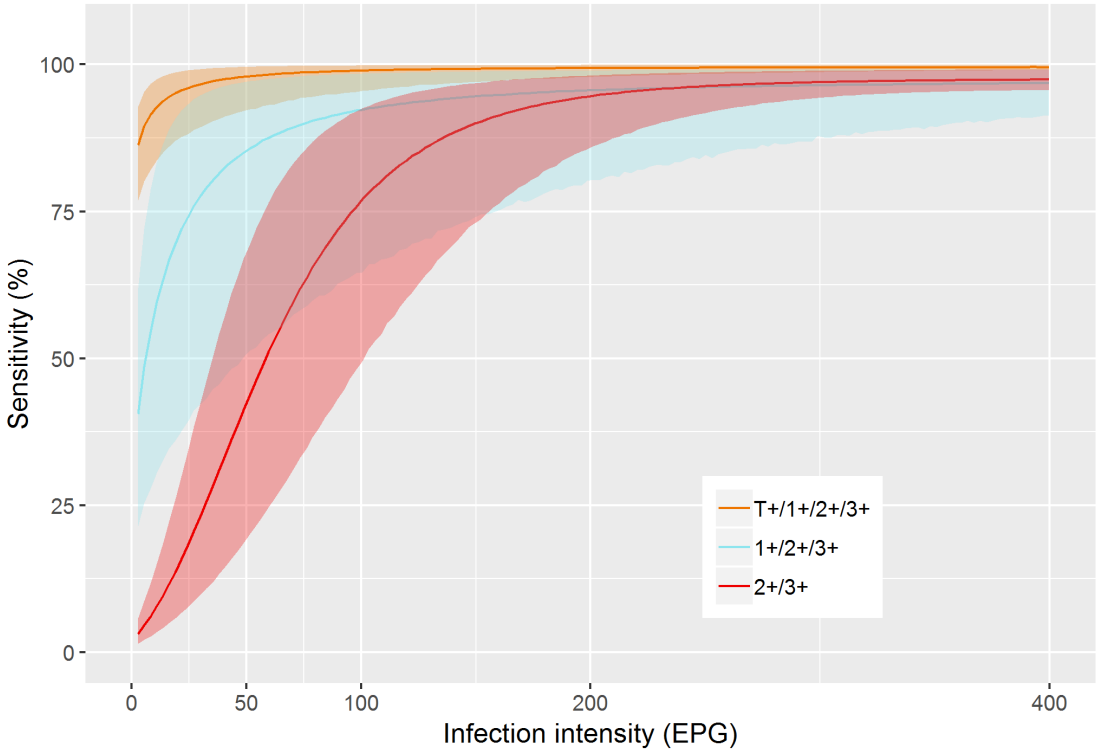


Figure 3.1: Model-based estimate of the infection intensity-dependent sensitivity of POC-CCA

Table 3.3 shows the specificity and sensitivity estimates for egg-negative infections. POC-CCA revealed very high sensitivity even for light infection intensities when trace results were interpreted as positive. Furthermore, sensitivity was above 60% for non-egg-shedding but antigen-positive individuals that apparently harbor worms. The proxy $2+$ and $3+$ revealed low sensitivity for egg-negative/antigen-positive individuals. Specificity is over 95% irrespective of the chosen proxy.

In Fig 3.2 the infection intensity-dependent sensitivity of Kato-Katz for one to two slides

Table 3.3: Model-based estimates of the specificity and sensitivity for egg-negative infections of POC-CCA

	Specificity	Sensitivity–egg-negative/antigen-positive
T/1+/2+/3+	0.96 (0.95, 0.97)	0.75 (0.65, 0.84)
1+/2+/3+	1.00 (0.99, 1.00)	0.23 (0.12, 0.39)
2+/3+	1.00 (0.99, 1.00)	0.01 (0.01, 0.02)

from one to three samples are shown. Single slide Kato-Katz testing is equivalent to the 2+/3+ POC-CCA proxy with a sensitivity of 60% at 50 EPG, and 70% at 100 EPG. A single slide on two different days clearly outperforms two slides from the same sample. Taking two slides from three different samples leads essentially to perfect sensitivity for infection intensities above 50 EPG but even this extensive sampling scheme does not allow reliable detection of very light infections below 10 EPG.

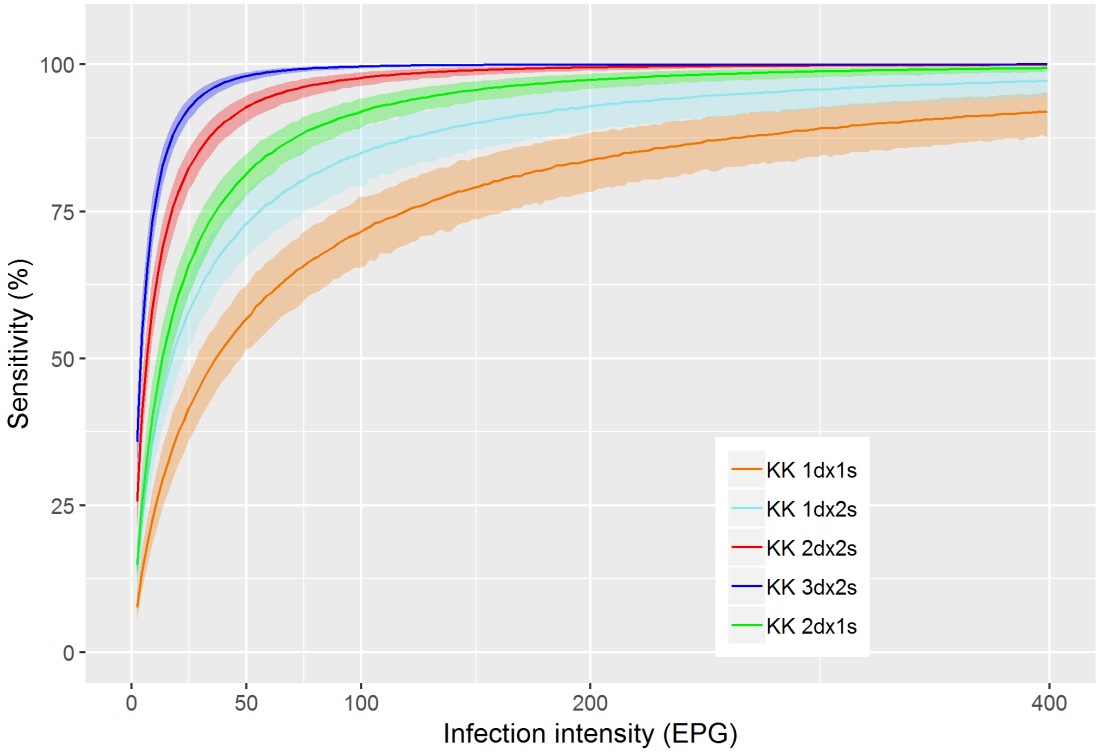


Figure 3.2: Model-based estimates of the infection intensity-dependent sensitivity of Kato-Katz for various numbers of days (d) with one or two samples (s).

3.3.2 Prevalence and mean infection intensity model estimates

Table 3.4 shows estimates of egg-positive and egg-negative infection prevalence, mean infection intensity of an egg-positive infected individual, and mean egg count in the population. We found a wide variation in the estimates between datasets with egg-positive infection prevalence ranging from 6% to 91%, and egg-negative prevalence varying from 3% to 45%. The mean intensity of an infected individual ranged from 80 EPG to more than 500 EPG and the mean egg count in the population ranged from around 4 EPG to more than 300 EPG.

There was no simple relation between the prevalence of egg-negative and egg-positive infections, the infection intensity of the infected individuals, and the total infection intensity in the population (Fig 3.3). However, there was evidence of a positive relation between the prevalence of egg-positive and egg-negative infections at low prevalence.

3.3.3 Relation between Kato-Katz and POC-CCA

Fig 3.4 displays the relation between the observed prevalence measured with a single Kato-Katz thick smear and POC-CCA. We observed a clear relation for each of the three proxies of

Table 3.4: Prevalence of egg-positive and egg-negative/antigen-positive cases and infection intensities for each dataset.

Country	Location	Prev. egg-pos. (%) [1]	Prev. egg-neg. (%) [2]	Mean infect. (EPG) [3]	Mean pop. egg count [4] (EPG)
Cameroon	Makenene	91 (85, 96)	3 (0, 8)	221.8 (149.8, 325.8)	75.0 (53.3, 106.2)
Cameroon	Njombe	84 (75, 91)	5 (0, 13)	189.0 (113.4, 308.9)	48.2 (30.9, 72.2)
Cameroon	Yaounde	38 (29, 48)	38 (27, 49)	476.5 (225.2, 986.7)	46.5 (25.1, 81.5)
Côte d'Ivoire	Man	8 (6, 12)	24 (20, 28)	132.2 (65.5, 250.0)	7.6 (3.5, 14.3)
Côte d'Ivoire	1	42 (29, 58)	42 (26, 58)	222.1 (84.9, 515.4)	41.8 (19.4, 80.2)
Côte d'Ivoire	2	50 (31, 70)	23 (4, 41)	245.8 (94.9, 572.1)	65.4 (27.8, 132.6)
Côte d'Ivoire	1	93 (88, 97)	5 (1, 10)	683.5 (535.9, 891.6)	313.9 (250.6, 393.5)
Côte d'Ivoire	2	62 (52, 74)	25 (7, 42)	243.0 (127.5, 442.9)	59.7 (34.7, 97.2)
Côte d'Ivoire	3	51 (39, 63)	8 (0, 28)	868.5 (134.1, 3656.4)	48.4 (17.7, 100.6)
Ethiopia	Harbu	67 (58, 76)	14 (5, 22)	80.5 (59.8, 107.8)	28.2 (21.4, 37.6)
Ethiopia	Jiga	52 (47, 58)	10 (6, 15)	156.7 (125.9, 197.7)	65.0 (50.9, 82.8)
Kenya		30 (27, 34)	45 (41, 49)	119.0 (90.6, 158.8)	11.1 (8.8, 13.9)
Uganda	1	72 (58, 83)	5 (0, 17)	274.3 (143.4, 508.3)	86.5 (46.6, 154.3)
Uganda	2	70 (56, 81)	7 (0, 20)	217.2 (111.5, 398.1)	74.4 (40.0, 126.9)
Uganda	3	38 (25, 53)	27 (14, 39)	167.1 (55.9, 426.8)	27.6 (9.9, 60.9)
Uganda	4	43 (29, 62)	21 (6, 34)	586.6 (247.7, 1,350.2)	70.2 (35.0, 128.9)
Uganda	5	16 (8, 28)	35 (23, 48)	366.3 (136.4, 896.9)	26.4 (9.3, 55.4)
Uganda	Base	12 (9, 16)	13 (9, 17)	303.9 (56.9, 1,347.0)	3.8 (1.6, 8.2)
Uganda	F1	9 (6, 13)	20 (15, 25)	137.5 (53.3, 328.0)	7.5 (3.0, 16.3)
Uganda	Mapping	6 (3, 8)	6 (3, 9)	693.4 (215.0, 2,117.2)	9.0 (3.8, 17.1)

Posterior mean and 95% BCI for all model estimates.
[1] Estimated prevalence of egg-shedding individuals
[2] Estimated prevalence of non egg-shedding individuals that harbor worms
[3] Estimated arithmetic mean infection intensity of an egg-positive infected individual
[4] Estimated mean egg count in the population

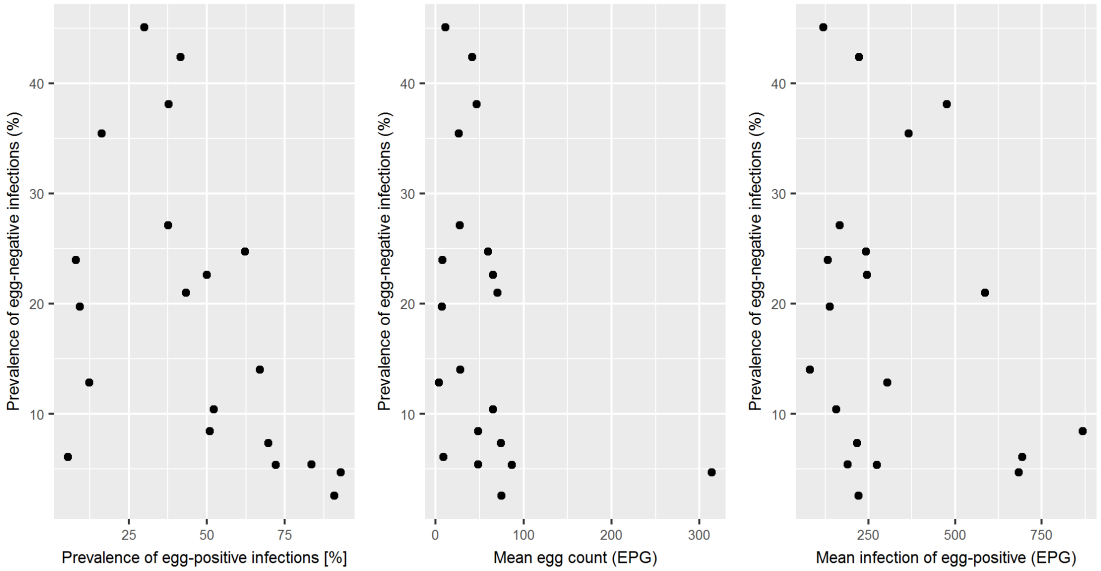


Figure 3.3: Relation between egg-negative prevalence, egg-positive prevalence, mean egg count in the population, and mean infection intensity of an infected individual.

POC-CCA, but there is still insufficient data to give a clear indication of the variability.

The results of the simulation study (Fig 3.5) showed the estimated relation between single and duplicate slide Kato-Katz prevalence and POC-CCA prevalence for varying egg-positive prevalence for the 16 combinations with the four infection intensity scenarios on the x-axis and the four scenarios for the prevalence of egg-negative infections on the y-axis.

For trace considered positive, the relation between POC-CCA and Kato-Katz prevalence were dependent on infection intensity and prevalence of egg-negative infections. When Kato-Katz prevalence is zero, the POC-CCA prevalence is determined by the prevalence of egg-negative infections. For low mean infection intensities, there are primarily light infections for which POC-CCA has a higher sensitivity, and hence, the slope of the relation is small. When only $2+$ and $3+$ readings are considered positive, the relation neither depends on the infection intensity nor on the prevalence of egg-negative infections. When trace is considered negative, there is lower variation in the Kato-Katz/POC-CCA relation. These patterns hold for both single and duplicate slide Kato-Katz with a change in the slope of the relation. Fig 3.6 and Fig 3.7 show a scatter plot of prevalence measured by single and duplicate slide Kato-Katz, respectively, and POC-CCA for

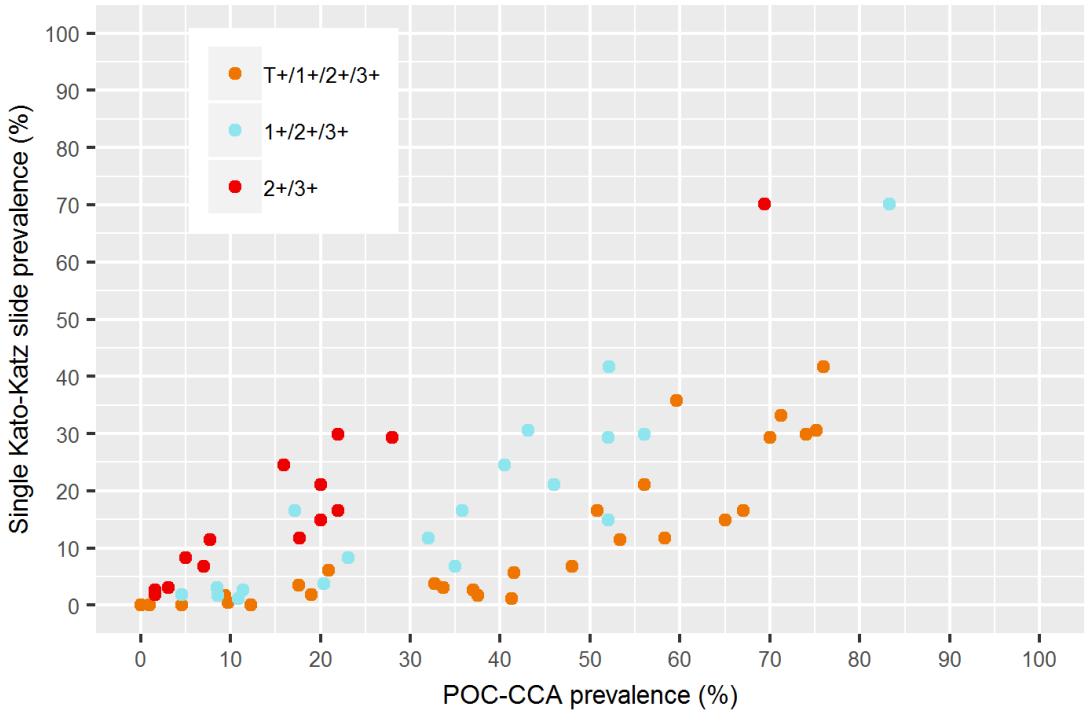


Figure 3.4: Relation between observed Kato-Katz and POC-CCA prevalence based on the datasets in Tables 3.1 and 3.2.

all simulation scenarios to highlight the variability present in the Kato-Katz/POC-CCA relation for each proxy.

Both the prevalence of egg-negative infections and mean infection intensity are, in practice, unknown when only POC-CCA diagnostics is applied. There is large uncertainty in the estimates for the trace-positive proxy due to the dependence on unknown parameters. For the trace-negative proxy there is weaker but still considerable uncertainty. The 2+ and 3+ proxy shows close to perfect accordance with Kato-Katz and only low uncertainty. Based on the simulation scenarios, Table 3.5 translates single and duplicate slide Kato-Katz prevalence thresholds into POC-CCA equivalents. For trace considered positive, the variability is large because of the previously mentioned unknown number of egg-negative/antigen-positive individuals. Single slide Kato-Katz and POC-CCA prevalence are identical when 2+ and 3+ proxy is considered. Due to the uncertainty in the Kato-Katz/POC-CCA relation when trace-positive or trace-negative proxies are

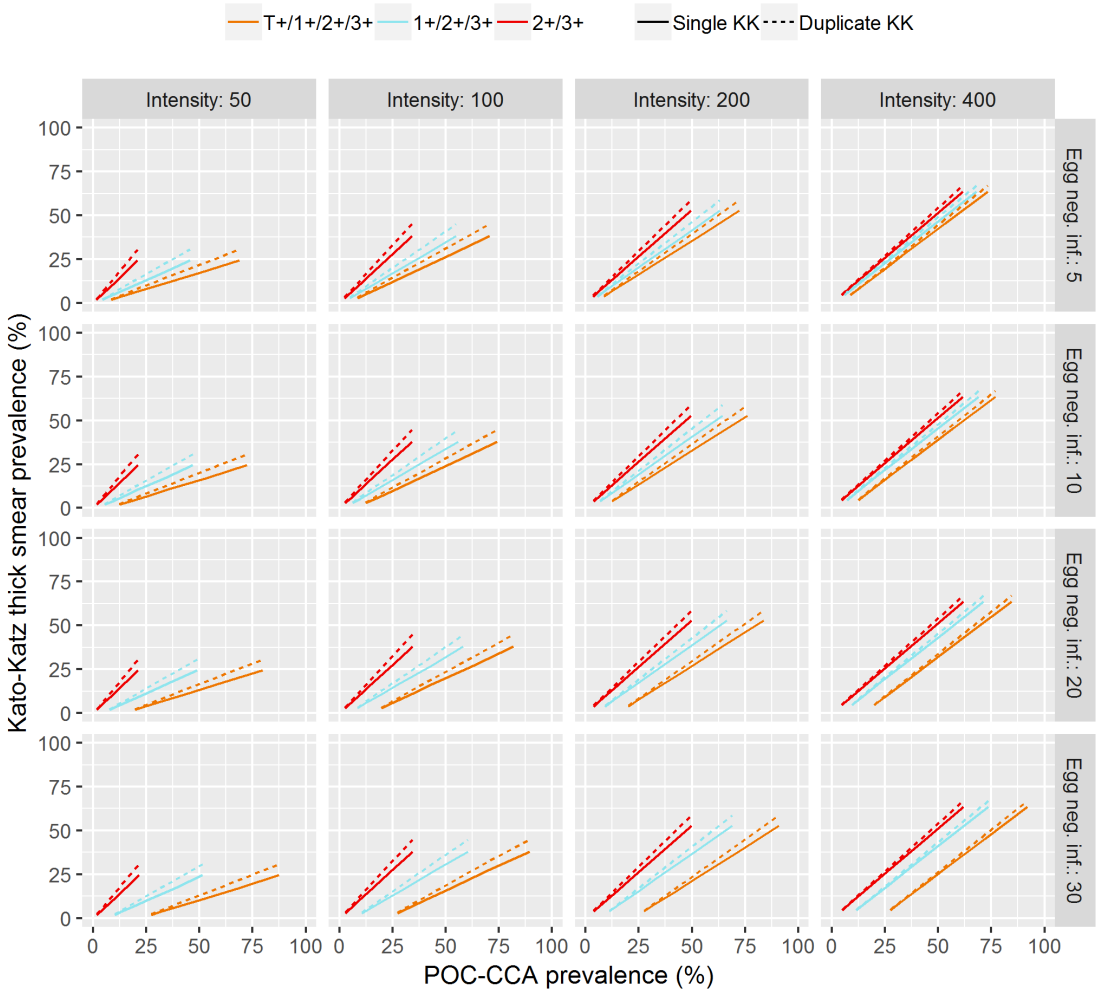


Figure 3.5: Predictions of the dependence between Kato-Katz and POC-CCA prevalence for various infection intensities and prevalence of egg-negative cases. Infection intensity is given in EPG and egg-negative prevalence in %.

used, the Kato-Katz prevalence corresponds to a range of POC-CCA prevalence values, which are generally higher than the Kato-Katz one. A single slide Kato-Katz prevalence of 10% corresponds to 20–40% POC-CCA prevalence when traces are considered positive and to 15–25% when traces are considered negative. For a duplicate slide Kato-Katz prevalence of 10% the equivalent ranges are 15–40% when traces are considered positive and 10–20% when traces are considered negative.

When avoiding under-treatment is the priority, conservative thresholds could be defined at

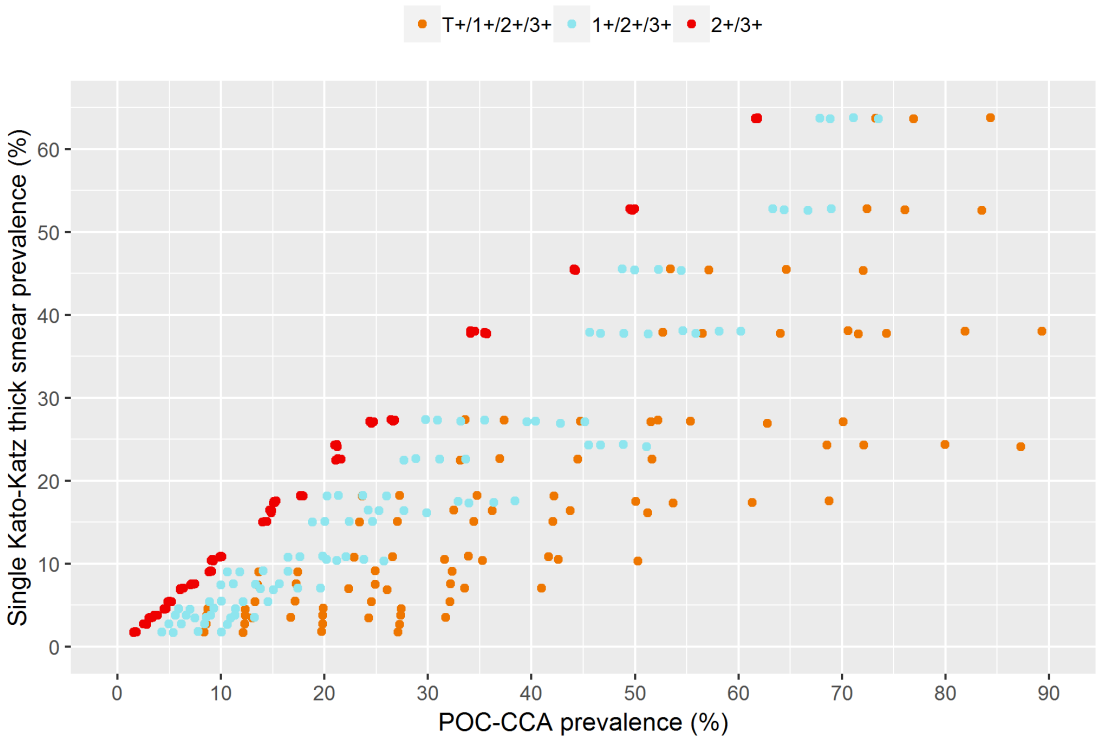


Figure 3.6: Scatter plot of all results for single slide Kato-Katz from Fig 3.5.

the lower end of each range. However, the lower end of the range corresponds to a high mean infection-intensity of an infected individual, which is only a realistic scenario in a high prevalence setting. For lower prevalence, a lower mean infection intensity is a more realistic threshold. Hence, we suggest a threshold of 60% POC-CCA to be equivalent to 50% duplicate Kato-Katz, and 30% POC-CCA to be equivalent to 10% Kato-Katz to define treatment categories. Table 3.6 shows the WHO guidelines for *S. mansoni* from 2013 [WHO, 2013]. In addition to the threshold given for parasitological methods we added the suggested thresholds from Table 3.5.

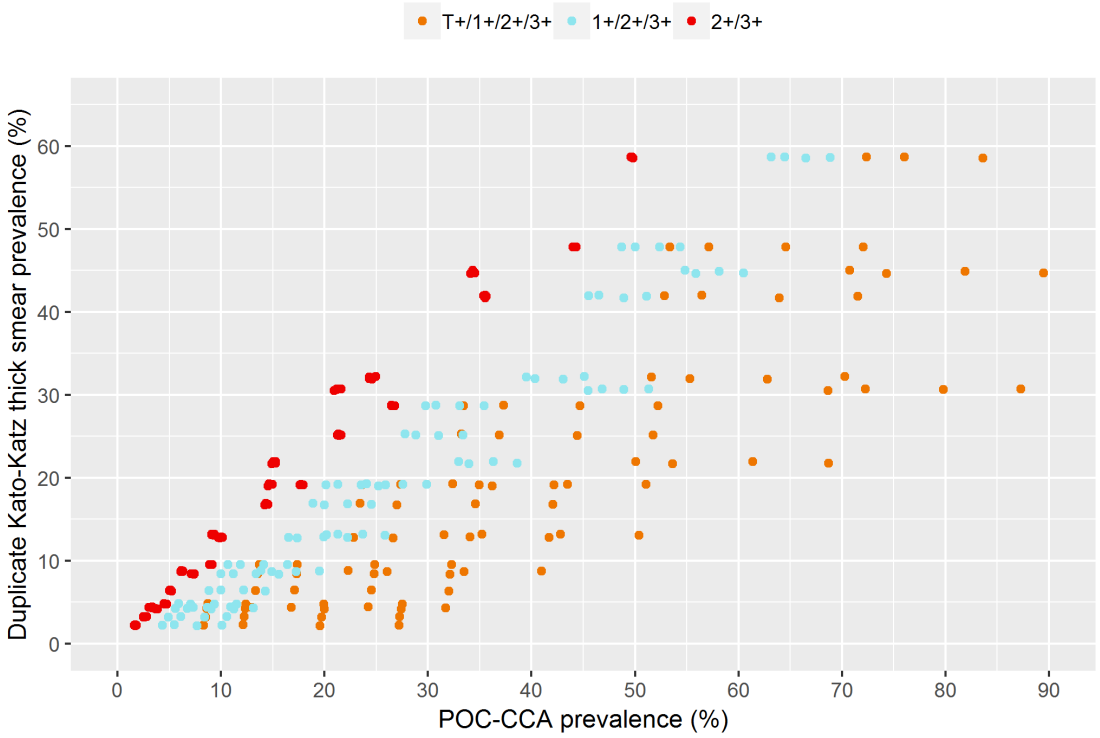


Figure 3.7: Scatter plot of all results for duplicate slide Kato-Katz from Fig 3.5.

Kato-Katz	**POC-CCA** T/1+/2+/3+	Suggested threshold	1+/2+/3+	2+/3+
Single				
1%	5-30	10%	3-10%	1%
5%	10-30%	20%	5-15%	5%
10%	20-40%	30%	15-25%	10%
25%	35-70%	50%	30-50%	25%
50%	>75%	75%	>60%	50%
Duplicate				
1%	5-25	10%	3-10%	1%
5%	10-35%	20%	5-15%	5%
10%	15-40%	30%	10-20%	5-10%
25%	30-70%	45%	25-40%	15-25%
50%	>60%	60%	>50%	>40%

Table 3.5: Estimated equivalent prevalence of POC-CCA to single and duplicate slide Kato-Katz and suggested equivalent prevalence threshold

Category	Prevalence among school-aged children	Action to be taken
High-risk community	$\geq 50\%$ by parasitological methods or $\geq 60\%$ by POC-CCA	Treat all school-aged children (enrolled and not enrolled) once a year. Also treat adults considered to be at risk (from special groups or once a year to entire communities living in endemic areas).
Moderate-risk community	$\geq 10\%$ by parasitological methods or $\geq 30\%$ by POC-CCA	Treat all school-aged children (enrolled and not enrolled) once every 2 years. Also treat adults considered to be at risk (special groups only).
Low-risk community	$\leq 10\%$ by parasitological methods or $\leq 30\%$ by POC-CCA	Treat all school-aged children (enrolled and not enrolled) twice during their schooling age (e.g., once on entry and once on exit). Praziquantel should be available in dispensaries and clinics for treatment of suspected cases.

Table 3.6: WHO recommended treatment strategy for schistosomiasis mansoni [WHO, 2013] extended with the suggested thresholds for POC-CCA from Table 3.5.

3.4 Discussion

We established the relation between observed Kato-Katz and POC-CCA prevalence of *S. mansoni* infections using rigorous Bayesian modelling and extensive simulation studies. Moreover, our analysis provided estimates of the infection intensity-dependent sensitivity, the sensitivity for egg-negative/antigen-positive infections, and the specificity of POC-CCA without making use of an artificial 'gold' standard.

We found that a duplicate slide Kato-Katz prevalence of 10% corresponds to 15–40% POC-CCA when trace are considered positive and 10–20% POC-CCA when trace are considered negative. The uncertainty in the estimates cannot be eliminated due to the exact value depending on quantities not known by POC-CCA, namely the mean infection intensity in the population and the prevalence of egg-negative but antigen-positive individuals. More accurate estimates can be given using the POC-CCA $\geq 2+$ proxy. It shows similar change of the sensitivity with infection intensity and insensitivity to light-intensity infections as single slide Kato-Katz leading to a one-to-one correspondence. However, this stringent proxy misses light intensity infections, as does the Kato-Katz assay.

Previous analyses were unable to give clear recommendations due to various limitations in the models and data used. Kittur et al. (2016) [Kittur et al., 2016] performed a regression analysis on the datasets included in the reviews by Danso-Appiah et al. (2016) [Danso-Appiah et al., 2016] and Ochodo et al. (2015) [Ochodo et al., 2015] to establish the relation between Kato-Katz and POC-CCA prevalence. The researchers found a clear correlation between the semi-quantitative results of POC-CCA and the egg output measured by Kato-Katz, but the variation in the relation between Kato-Katz prevalence and POC-CCA was too large for predictive use. Prada et al. (2018) [Prada et al., 2018] applied a Bayesian model on pre- and post-treatment data from a study in Uganda taking into account the infection intensity-dependent sensitivity of POC-CCA. They estimated a specificity of only 85% and determined a relation between Kato-Katz and POC-CCA prevalence for true prevalence above 35%. Caveats of their analysis are as follows. First, the semi-quantitative readings were modelled by a binomial distribution, which is not supported by the process generating the data. Second, POC-CCA positive but egg-negative individuals

were defined as negative, thus underestimating specificity. Third, model-based estimates do not represent low endemic settings.

The results presented here revealed that POC-CCA has close to perfect sensitivity for moderate and heavy *S. mansoni* infections (≥ 100 EPG) and a specificity of over 95% regardless of the endemicity. Generally, the uncertainty in the sensitivity estimates is larger than for Kato-Katz due to considerable variation introduced by the lack of a standardized reader. For light infections (1–99 EPG) the sensitivity varies with the proxy used to categorize semi-quantitative POC-CCA readings into a binary result (see Fig 3.1). The proxy that classifies traces as positives has a very high sensitivity even for infections below 50 EPG and it detects egg-negative/antigen-positive infections with a moderate sensitivity of 75%, while the Kato-Katz technique is unreliable in detecting infections with intensities below 10 EPG even for repeated sampling. The proxy that considers POC-CCA $\geq 1+$ as positive has lower sensitivity, but larger variability in the estimates of the infection dependent sensitivity. It follows that POC-CCA has a very high diagnostic sensitivity and specificity at the individual level when traces are included in the positive results.

In two recent systematic reviews and meta-analyses of diagnostic accuracy of POC-CCA, a sensitivity of 90% of POC-CCA was found when trace was included in the positives, which is in agreement with our results [Ochodo et al., 2015; Danso-Appiah et al., 2016]. However, the prior analyses assumed Kato-Katz as the diagnostic 'gold' standard and did not take into account the dependence of the sensitivity on infection intensity. The former assumption implies that the additional positives of POC-CCA are false-positives. They report a specificity of 55% against single Kato-Katz and a specificity of 66% against duplicate or triplicate Kato-Katz. The increase in specificity when compared to a more accurate 'gold' standard indicates that not all additional positives are false-positives and that they underestimate the true specificity. The lack of positive POC-CCA tests in non-endemic regions, the correlation of egg-negative/antigen-positive prevalence with egg-positive prevalence, and our model-based estimates suggest that most of the additional positives detected by POC-CCA are true infections. This interpretation is in agreement with Mwinzi et al. (2015) [Mwinzi et al., 2015] who showed that the number of POC-CCA-positives but egg-negative decreased after treatment with praziquantel.

3.4.1 Conclusion

Kato-Katz prevalence can be translated to a range of POC-CCA prevalence. Choosing a single equivalent threshold can be justified for simplicity and applicability but leads to misclassification. A conservative threshold could be chosen at the lower end of the range, which ensures that there is no under-treatment but doing so would also underestimate the true prevalence of disease. Instead, we recommend a more balanced approach suggesting a threshold at the lower end of the range for high prevalence, and a more central value for lower prevalence to reflect the accompanying decrease in mean infection-intensity of the infected population. Therefore, the 10% and 50% duplicate slide Kato-Katz thresholds are to be translated to 30% and 60% POC-CCA, respectively, when traces are considered positive. Additionally, new treatment categories for scenarios close to elimination can be defined for POC-CCA at 10% and 20% roughly corresponding to 1% and 5% Kato-Katz, respectively. For a more accurate translation, which is especially useful when integrating studies based on Kato-Katz and POC-CCA diagnostics into a single analysis, we recommend recording fully semi-quantitative results and using the POC-CCA $\geq 2+$ prevalence as equivalent to single slide Kato-Katz.

In suspected low-endemicity settings, we recommend replacing Kato-Katz irrespective of sampling effort with the trace positive proxy of POC-CCA. This is solely based on the diagnostic accuracy determined in this study, while cost-effectiveness shall be evaluated taking into account the specific situation. The test's sensitivity to infections with none or only erratic egg shedding, which are difficult to detect even by Kato-Katz repeated on multiple days, make it useful for surveillance in settings approaching elimination and for diagnostics on the individual level. The presented evidence suggests that the egg-negative but antigen-positive infections are real infections, and hence, it is conceivable that the current Kato-Katz based estimates for prevalence and morbidity underestimate the reality. Better tools and further studies are needed to determine worm burdens and morbidity associated with egg-negative infections.

Appendix

	Posterior mean	95% BCI
Kato-Katz		
α	0.07	0.04–0.11
γ_2	5.34	3.07–10.10
σ	1.28	1.08–1.49
T/1+/2+/3+		
a_0	1.10	0.62–1.67
a_1	3.64	1.97–6.25
a_2	5.58	4.55–6.87
1+/2+/3+		
a_0	-1.20	-1.95– -0.42
a_1	3.75	1.90–6.59
a_2	3.69	2.78–4.68
2+/3+		
a_0	-4.46	-5.16– -3.74
a_1	4.34	3.47–5.32
a_2	3.72	2.71–4.53

A1: Posterior means and 95% BCI of model parameters that determine sensitivity of POC-CCA and Kato-Katz

Chapter 4

Estimating true prevalence of *S. mansoni* from population summary measures based on the Kato-Katz diagnostic technique

Authors

Oliver Bärenbold[1,2], Daniel G. Colley[3], Fiona M. Fleming[4], Rufin K. Assaré[1,2,5,6], Edridah M. Tukahebwa[7], Biruck Kebede[8], Jean T. Coulibaly[1,2,5,6], Eliézer K. N'Goran[5,6], Louis-Albert Tchuem Tchuenté[9,10], Pauline Mwinzi[11], Jürg Utzinger[1,2], Penelope Vounatsou[1,2]

1 Swiss Tropical and Public Health Institute, Basel, Switzerland

2 University of Basel, Basel, Switzerland

3 Center for Tropical and Emerging Global Diseases and Department of Microbiology, University of Georgia, Athens, GA, United States of America

4 Schistosomiasis Control Initiative, Imperial College, London, United Kingdom

5 Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, Abidjan, Côte d'Ivoire

6 Unité de Formation et de Recherche Biosciences, Université Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire

7 Vector Control Division, Ministry of Health, Kampala, Uganda

8 Ministry of Health, Addis Ababa, Ethiopia

9 Laboratory of Parasitology and Ecology, University of Yaoundé I, Yaoundé, Cameroon

10 Centre for Schistosomiasis and Parasitology, Yaoundé, Cameroon

11 Centre for Global Health Research, Kenya Medical Research Institute, Nairobi, Kenya

To be submitted to *PLoS neglected tropical diseases*.

Abstract

Background

Prevalence of *Schistosomiasis mansoni* is usually evaluated by the Kato-Katz diagnostic technique, which has low sensitivity especially for light infections and depends on the replicate samples taken. Egg count models fitted on individual level data can adjust for the infection intensity-dependent sensitivity and estimate the 'true' prevalence in a population. Most frequently however, individual data are not available. This study provides estimates of the 'true' prevalence from population summary measures of infection intensity and observed prevalence using extensive simulations parametrized with data from different settings in sub-Saharan Africa.

Methodology

An individual level egg count model was applied to Kato-Katz data to determine the infection intensity-dependent sensitivity for various sampling schemes varying from one sample taken for a single day to three slides per day for three days. Observations in populations with varying transmission strengths were simulated using standard assumptions about the distribution of worms, their mating behaviour and the between days variations in Kato-Katz egg counts. Summary measures such as the geometric mean infection, arithmetic mean infection, and the observed prevalence of the simulations were calculated and parametric statistical models were fitted

to the summary measures for each sampling scheme. Validation was performed comparing simulation-based estimates to real data.

Principal findings

Overall, the sensitivity of Kato-Katz in a population varies strongly with the mean infection intensity. Both, the geometric and arithmetic mean infection intensity improve estimation of sensitivity using a parametric model which takes into account different sampling schemes varying from one day, one slide to three days with three slides per day. The relation between observed and 'true' prevalence is remarkably linear and three slides per day on three subsequent days ensure close to perfect sensitivity. The validation indicates an underestimation of sensitivity for high infection intensities due to over-estimation of the number of light infections in the simulation.

Conclusions/significance

Estimation of 'true' prevalence is improved when taking into account geometric or arithmetic mean infection intensity in a population. We supply parametric functions and corresponding estimates of their parameters to calculate the 'true' prevalence for sampling schemes up to three days with three samples per day that allow estimation of the 'true' prevalence.

Acknowledgements

This study received financial support from the European Research Council (ERC-2012-AdG-323180). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

4.1 Introduction

Schistosomiasis mansoni is a neglected tropical disease caused by a parasitic flatworm with a complex life cycle that includes freshwater snails as an intermediate host [McManus et al., 2018]. The most recent Global Burden of Disease study gives an estimate of 250 million infected individuals and a total of 2.1 million disability-adjusted life years primarily among the school-aged population of sub-Saharan Africa [GBD 2016 DALYs and HALE Collaborators, 2017; GBD 2016 Causes of Death Collaborators, 2017; GBD 2016 Disease and Injury Incidence and Prevalence Collaborators, 2017]. The World Health Organization (WHO) has set the goal of eliminating morbidity due to *S. mansoni* by 2025 using preventive chemotherapy in regions that surpass given thresholds for the prevalence of the disease [WHO, 2013].

S. mansoni is most commonly diagnosed using the Kato-Katz diagnostic technique, a parasitological method that relies on the detection of eggs in a stool-smear of 41.7 mg using a light-microscope [Katz et al., 1972]. However, the diagnostic sensitivity of the technique is low, in the order of 50%, for one or two stool samples on a single day, and varies substantively with the sampling scheme, i.e. the number of days on which sampling is done and the number of slides per sample [Kongs et al., 2001; Krauth et al., 2012]. The sensitivity increases strongly with the density of eggs in the stool of an individual, however, for very light infections even sampling on many days does not lead to a sensitivity close to 100% and therefore prevalence remains underestimated [Bärenbold et al., 2017; Colley et al., 2017].

The WHO defines intervention thresholds for mass drug administration with praziquantel based on observed prevalence while not taking into account infection intensity and sensitivity of the diagnostic method [WHO, 2006]. Using observed prevalence from widely different settings unadjusted is at risk for introducing bias. To administer preventive chemotherapy effectively and to track improvements, historical data generated under various sampling schemes needs to be integrated in a coherent model to reduce the risk for bias.

Recent studies have shown that the underlying ‘true’ prevalence can be estimated effectively when individual level data are given [Bärenbold et al., 2017, 2018]. However, very often individual level data is not available but only the overall prevalence and in some cases the arithmetic or

geometric mean infection intensity, and therefore, individual level models cannot be applied. Therefore, there is a need to estimate the 'true' prevalence from population summary measures. The most basic approach is to use the overall diagnostic sensitivity as for example determined in studies using artificial 'gold' standards to compare diagnostic techniques [Glinz et al., 2010]. This study aims to improve on the basic approach of constant sensitivity by taking into account the influence of the mean infection intensity and observed prevalence on the sensitivity of Kato-Katz.

We carried out a simulation study using a transmission model that takes into account the diagnostic process to determine the relation between the 'true' prevalence and the observed prevalence as well as summary measures for infection intensity. We evaluate the results using national survey data from Uganda. Thus, we propose a simple way for practitioners to calculate 'true' prevalence from observed prevalence while taking into account infection intensity. Additionally, we supply informative priors to be used in *S. mansoni* modelling studies, such as latent class analysis using the Bayesian formulation.

4.2 Material and methods

4.2.1 Ethics statement

All data included in this study were published elsewhere [Coulibaly et al., 2011; Shane et al., 2011; Tchuem Tchuenté et al., 2012; Colley et al., 2013; Coulibaly et al., 2013b; Erko et al., 2013; Adriko et al., 2014; Mwinzi et al., 2015; Assare et al., 2018]. Ethics approval, written informed consent procedures and treatment of infected individuals are given in the aforementioned studies where the data were published.

4.2.2 Data

In this study we first use a suite of 20 datasets with Kato-Katz results available for between two and three slides on two or three days. A description of the data is given in Table 4.1. The data originates from five countries in sub-Saharan Africa, namely Cameroon, Côte d'Ivoire, Ethiopia, Kenya, and Uganda. Observed prevalence ranged from 3.8% to 91.7%, mean infection intensity from 37 EPG to 525 EPG, and the geometric mean from 8.5 to 248 EPG. Two or three Kato-Katz

slides were taken from stool samples on two or three days from between 100 and 1845 individuals. This data is used in the egg count model in Section 4.2.3.1 to infer on the infection intensity dependent sensitivity of the Kato-Katz technique.

Table 4.1: Summary of the survey data used in the study.

Country	Location	age range	$\mathbf{n_d}$ [1]	$\mathbf{n_s}$ [2]	$\mathbf{N_{KK}}$	P. KK (%)	μ KK (EPG)	geom. $\mu\mu$ (EPG)	P 1 KK (%)	Source
Cameroon	Makanene	6-16	3	3	251	71.7	161	43.3	41.7 (34.8, 48.7)	Tchuem Tchuenté et al. [2012]
Cameroon	Njobe	8-16	3	3	245	63.3	173	27.5	30.6 (26.9, 34.3)	Tchuem Tchuenté et al. [2012]
Cameroon	Yaonde	7-14	3	3	233	27.9	235	40.9	16.5 (12.3, 20.7)	Tchuem Tchuenté et al. [2012]
Côte d'Ivoire	?		2	3	695	6.5	72	22.0	3.8 (2.6, 4.9)	Assare et al. [2018]
Côte d'Ivoire	1	0.2-5.5	2	2	109	25.7	90.2	37.0	16.5 (12.3, 20.8)	Coulibaly et al. [2013b]
Côte d'Ivoire	2	0.2-5.5	2	2	133	21.1	122	30.8	11.7 (9.1, 14.3)	Coulibaly et al. [2013b]
Côte d'Ivoire	1	8-12	3	3	170	91.7	525	248	70.2 (62.4, 78.1)	Coulibaly et al. [2011]
Côte d'Ivoire	2	8-12	3	3	130	53.1	116	36.8	24.5 (14.8, 34.3)	Coulibaly et al. [2011]
Côte d'Ivoire	3	8-12	3	3	146	32.9	50	8.5	8.3 (3.1, 13.5)	Coulibaly et al. [2011]
Ethiopia	Harbu	8-12	3	2	300	57	69	31.0	33.1 (24.8, 41.4)	Erko et al. [2013]
Ethiopia	Jiga	8-12	3	2	320	49.4	153	70.9	35.8 (32.1, 39.5)	Erko et al. [2013]
Kenya			3	2	1845	22.1	106	32.1	11.4 (7.7, 15.2)	Shane et al. [2011]
Uganda	1	7-13	3	2	100	55.0	240	34.2	29.3 (24.0, 34.6)	Adriko et al. [2014]
Uganda	2	7-13	3	2	100	54.0	122	33.3	29.8 (23.6, 36.1)	Adriko et al. [2014]
Uganda	3	7-13	3	2	100	31.0	37	19.8	14.9 (9.7, 20.1)	Adriko et al. [2014]
Uganda	4	7-13	3	2	100	35.0	247	58.0	21.1 (16.8, 25.4)	Adriko et al. [2014]
Uganda	5	7-13	3	2	100	12.0	58	28.4	6.8 (3.8, 9.8)	Adriko et al. [2014]
Uganda	Base		3	2	775	6.3	48	22.0	3.1 (1.5, 4.7)	
Uganda	F1		3	2	659	4.2	68	33.5	2.7 (1.5, 3.9)	
Uganda	Mapping		3	2	711	3.8	182	26.9	1.8 (1.0, 2.6)	

[1] n_d is the number of stool specimens taken on different days;
[2] n_s is the number of Kato-Katz slides taken from each stool specimens;

Secondly, data from a national survey in Uganda in 2016, primarily along the coast of lake Victoria is used for validation purposes. There are 146 locations of which we used the 34 schools for the validation that had observed prevalence above 10% after two slides Kato-Katz from one stool-smear. There were 33 schools with 48 to 56 students and one school with 104 students. Observed prevalence ranges from 12% to 87.5% with 26 schools below 50%, arithmetic mean ranges from 9.8 EPG to 820 EPG, and geometric mean from 3.8 EPG to 233 EPG.

4.2.3 Methods

We fitted the real data across Africa on an egg count model estimating Kato-Katz diagnostic sensitivity as a function of infection intensity. We developed the egg count model in our earlier work by Bärenbold et al. (2017) and extended it in this study as described in Section 4.2.3.1 [Bärenbold et al., 2017]. We then simulated worm burden and observed data at individual level in a population using the model explained in 4.2.3.2 and taking into account the Kato-Katz

sensitivity estimated in 4.2.3.1. Statistical curves were fitted on the simulated data relating diagnostic sensitivity with 'observed' population mean egg intensity and prevalence. The curves were used to translate observed prevalence data by Kato-Katz into 'true' prevalence when the 'observed' arithmetic or geometric mean is available taking into account diagnostic sensitivity across sampling schemes varying from one to two slides per stool sample from one to three days. Details on the modelling, simulation approach and validation methods are given below.

4.2.3.1 Modelling diagnostic sensitivity of Kato-Katz

To model the data generating process we extended our Kato-Katz egg count model described in Bärenbold et al. (2017) [Bärenbold et al., 2017]. In this model the infected population j with *S. mansoni* prevalence p_j is characterized by an arithmetic mean egg intensity μ_j, while the infection intensity λ_{ji} of individual i is assumed to follow a shifted gamma distribution with a rate parameter α_j and a shift μ_{min} corresponding to the lowest possible infection with one worm pair.

$$\lambda_{ji} \sim \text{Gamma}((\mu_j - \mu_{min}) \cdot \alpha_j, \alpha_j) \tag{4.1}$$

To determine the sensitivity of Kato-Katz we modified the Bärenbold model to take into account i) the day-to-day variation of egg output of an individual and ii) the heterogeneous distribution of eggs within a single slide. In particular, we modelled the variation by a log-normal distribution, that is the egg output of individual i in population j on day d, $\log(\lambda_{jid}) = \log(\lambda_{ji} + \mu_{min}) + \epsilon_{jid}$ where $\epsilon_{jid} \sim N(-\sigma_j^2/2, \sigma_j)$. We assumed that the observed Kato-Katz egg count data Y_{jids}^{KK} are negative binomially distributed $Y_{jids}^{KK} \sim NB(\lambda_{jid}, \gamma_j)$. The parameter γ_j captures the over-dispersion in the egg count data from a single individual in population j [Krauth et al., 2012]. γ_j, and σ_j are allowed to vary between studies around a common mean, $\gamma_j \sim LN(\log(\gamma) - \sigma_\gamma^2/2, \sigma_\gamma)$ and $\sigma_j \sim LN(\log(\sigma) - \sigma_\sigma^2/2, \sigma_\sigma)$.

False-negatives are included in the model as repeated zero measurements, thus, the sensitivity for a single Kato-Katz reading of an individual with egg-density λ_{jid} becomes

$$s_{jid}^{KK} = 1 - \text{NB}(0, \lambda_{jid}, \gamma_j) = 1 - \left(\frac{\gamma_j}{\lambda_{jid} + \gamma_j}\right)^{\gamma_j} \tag{4.2}$$

For the non-infected individuals the mean egg-density λ_{ji}, and hence, the counts Y_{jids}^{KK} are set to zero.

The infection intensity of one pair of worms μ_{min} is fixed to the expected average egg output of a pair of worms, which is in the order of 100 eggs, multiplied with by ratio between the weight of a Kato-Katz sample and the average daily production of faeces which corresponds to about 0.05 eggs per sample (EPS) or 1.2 eggs per gram (EPG) for *S. mansoni* [Anderson, 1986]. All models were formulated using the Bayesian framework of inference and fitted by Markov chain Monte Carlo (MCMC) simulation in Stan version 2.16.2 (Stan Development Team; mc-stan.org) [Carpenter et al., 2016]. Priors were chosen as a normal distribution with mean 0.5 and standard deviation (SD) 0.5 for α_j, a gamma distribution with mean 25 and SD 125 for μ_j, truncated normal distribution with mean parameter 0 and SD parameter 3 for σ, truncated normal distribution with mean parameter 0 and SD parameter 1 for σ_σ, truncated normal distribution with mean parameter 1 and SD parameter 3 for γ, and a truncated normal distribution with mean parameter 0 and SD parameter 1 for σ_γ.

4.2.3.2 Simulating worm burden and observed disease data

We assume that the distribution of worms in the population can be simulated using a negative Binomial distribution which is a commonly employed assumption in transmission models for Schistosomiasis [Anderson, 1986]. Using the model for the worm mating process by May et al. (1977, 1993) the distribution of worm-pairs is determined [May, 1977; May and Woolhouse, 1993]. The distribution is defined by the mean number of worms μ_j^w in the population j, the aggregation parameter of the negative binomial distribution of worms k_j in the population j, q_w the proportion of female worms, and the mean number of eggs per worm-pair n_w set to 0.2 eggs per slide [Chan et al., 1995]. The aggregation parameters k_j are assumed to be related between different populations j and distributed log-normally around a common mean $k_j \sim LN(\log(k_0) - \delta^2/2, \delta)$, while the μ_j^w are independent and related to the strength of transmission. Under the assumption of a negative Binomial distribution the prevalence, defined as the individuals with zero worm

pairs, can be calculated as follows with p_j being the prevalence.

$$p_j = 1 - \left(1 + \frac{q_w \mu_j^w}{k_j}\right)^{-k_j} - \left(1 + \frac{(1-q_w)\mu_j^w}{k_j}\right)^{-k_j} + \left(1 + \frac{\mu_j^w}{k_j}\right)^{-k_j} \tag{4.3}$$

We simulated individual level data for 9,000 populations according to the egg count model presented above using the posterior of the fit to the datasets above as priors for σ and γ. Population size N was chosen once as 30 and as 50 which are two commonly used size for mapping studies, and as 5,000 to exclude influence of sampling error. The mean infection intensity of infected individuals was varied from an average of 10 to 400 worms per individual in 15 steps on a logarithmic scale, which covers a wide range of possible scenarios. We simulated 200 populations for each combination using a new draw from the joint posterior distribution for each population. Sampling schemes considered for Kato-Katz were $d1s1$ (1 day and 1 slide per sample), $d1s2$ (1 day and 2 slides), $d2s1$ (2 days and 1 slide), $d2s2$, $d3s2$, $d3s3$, where the number after s denotes the number of slides per day and the number after d the number of days where sampling was done.

For each simulated population and for each sampling scheme we calculated four different summary measures for the mean (i.e. arithmetic and geometric mean for all individuals as well as for only the positive ones). The full arithmetic mean μ_{ja} of the population j is simply the mean egg count of all slides over all days and individuals. To calculate the full geometric mean μ_{jg} the 1 has to be added to the counts to eliminate the zeros, then the geometric mean can be calculated using the standard formula, and finally, the 1 has to be subtracted again. The mean arithmetic, μ_{ja}^+ and geometric μ_{jg}^+ infection intensity in the positive population was calculated by

$$\begin{aligned} \mu_{ja}^+ &= \frac{\mu_{ja}}{p_j^{obs}} \\ \mu_{jg}^+ &= (\mu_{jg} + 1)^{1/p_j^{obs}} - 1 \end{aligned} \tag{4.4}$$

where p_j^{obs} is the observed prevalence. Because the full mean and the mean of positives are related directly via the prevalence, having one of the two means is sufficient to calculate the other.

The individual level sensitivity of the Kato-Katz diagnostics was obtained using Equation 4.2 based on the true infection intensity of individual i, in the simulated population j, on day d. The

mean of all sensitivities in a population was calculated to determine the sample sensitivity, i.e. $s_{ji}^{KK,ds} = 1 - \prod_{D=1}^{d} \left(\frac{\gamma_j}{\lambda_{jiD}+\gamma_j}\right)^{s\cdot\gamma_j}$.

4.2.4 Relating 'true' prevalence to observed prevalence and mean infection intensity

We fitted a parametric statistical model on simulated data to estimate population level sensitivity under the assumptions that i) the mean sensitivity can be expressed as a function of the infection intensity and ii) the sensitivity values are described by a beta distribution. z represents either a linear transformation of the geometric mean $\mu_g/8$ or the arithmetic mean $\mu_a/25$ of the full population. Division by 8 and 25 for geometric and arithmetic mean, respectively, ensures that the range of values for z is mostly within 0 and 1 optimising computation using MCMC for this model.

$$\begin{gathered} s_n^{KK,ds} \sim Beta(\alpha_n, \beta_n),\ \alpha_n = \frac{m_n}{\nu_n},\ \beta_n = \frac{1-m_n}{\nu_n} \\ \operatorname{logit} m_n = a_0 + a_1 \cdot z_n^{1/a_2},\ \log \nu_n = b_0 + b_1 \cdot \log(z_n) \end{gathered} \tag{4.5}$$

m_n is the estimated sensitivity for a given value z_n, where a_0 determines the sensitivity at low values for z, a_1 the increase with z, and a_2 the shape of the curve. ν_n determines the variance and is modelled using a linear model in $\log(z_n)$ with parameters b_0 and b_1.

We also determined the relation between observed p_n^{obs} and 'true' p_n prevalence using the following model ensuring a linear relation between p_n^{obs} and p_n.

$$\begin{gathered} p_n^{KK,ds} \sim Beta(\alpha_n, \beta_n),\ \alpha_n = \frac{m_n}{\nu_n},\ \beta_n = \frac{1-m_n}{\nu_n} \\ m_n = (2\operatorname{logit}^{-1}(a_1 p_n^{obs}) - 1)(1-a_0) + a_0),\ \nu_n = b_0 + b_1 \cdot p_n^{obs} \end{gathered} \tag{4.6}$$

m_n is the estimated 'true' prevalence p_n for a given observed prevalence p_n^{obs}, where a_0 determines the true prevalence at very low observed prevalence, a_1 describes the increase of the 'true' prevalence with increasing observed prevalence. ν_n determines the variance and is modelled using the two parameters b_0, b_1.

4.2.5 Validation

We validated the estimated relation between the population sensitivity and the mean infection intensity using survey data from Uganda described in Section 4.2.2. We estimated sensitivity and 'true' prevalence using the egg count model presented in Section 4.2.3.1 and calculated summary measures according to the definitions in Section 4.2.4. Validation was done by visual comparison between simulated populations and the estimates from the schools in the validation dataset.

4.3 Results

4.3.1 Kato-Katz day-to-day and slide-to-slide variations

Our model estimates the day-to-day and slide-to-slide variation in egg counts of the Kato-Katz diagnostic technique from 20 different datasets collected in five countries across sub-Saharan Africa while allowing for variations between different study locations. The parameter that catches the day-to-day variation, σ, was estimated to 1.16 (1.03 - 1.23), and the parameter of the negative binomial count distribution that captures the slide-to-slide variation, δ, to 6.14 (4.36 - 8.50) (Posterior mean and 95% BCI in brackets).

The two parameters γ and σ define the infection-intensity dependent sensitivity of Kato-Katz for every possible sampling scheme. The sensitivity is calculated using the probability to measure zero under the negative binomial model for the slide-to-slide variation while simulating λ_{jid} from the day-to-day variation. Figure 4.1 shows the infection-intensity dependent sensitivity for six different sampling schemes including one or two slides per stool sample from one to three days. Above 200 EPG, even a single slide achieves a sensitivity of 90% while at 10 EPG not even three samples with two slides each reach a sensitivity of 75%. Two slides from a different stool sample each show consistently higher sensitivity than two slides on the same day due to the stronger variations between days than slides.

Figure 4.2 depicts the sensitivity at low infection intensities below 25 EPG for the same six sampling schemes. The vertical line indicates a common estimate of the infection intensity of a single pair of worms at 0.2 EPS or 4.8 EPG. However, this estimate is uncertain and the true value

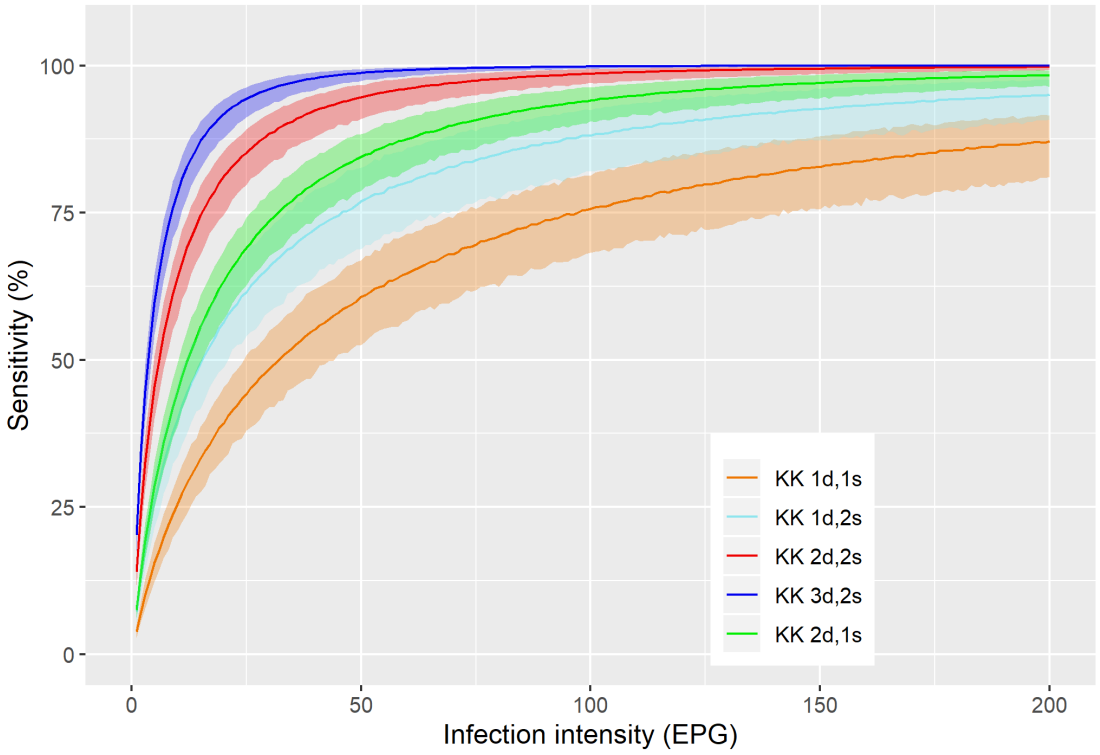


Figure 4.1: Infection intensity-dependent sensitivity of Kato-Katz for intensities up to 200 EPG and five different sampling schemes. d denotes the number of days where stool samples are collected, and s the number of slides per stool sample.

might very well be 1 EPG. Thus, the minimum sensitivity of Kato-Katz for three stool samples on consecutive days with two slides each might be as low as 25%.

4.3.2 Simulation of populations

The distributional assumptions about the transmission of worms in a population and the model for mating of worms to worm-pairs were combined with the diagnostic model using the parameter values determined through fit to many datasets as described in the methods section. Summary measures, arithmetic and geometric mean of all individuals, arithmetic and geometric mean of positives only, and observed prevalence were then calculated for the 9,000 simulated populations for six different sampling schemes with up to three slides from one to three stool samples on consecutive days.

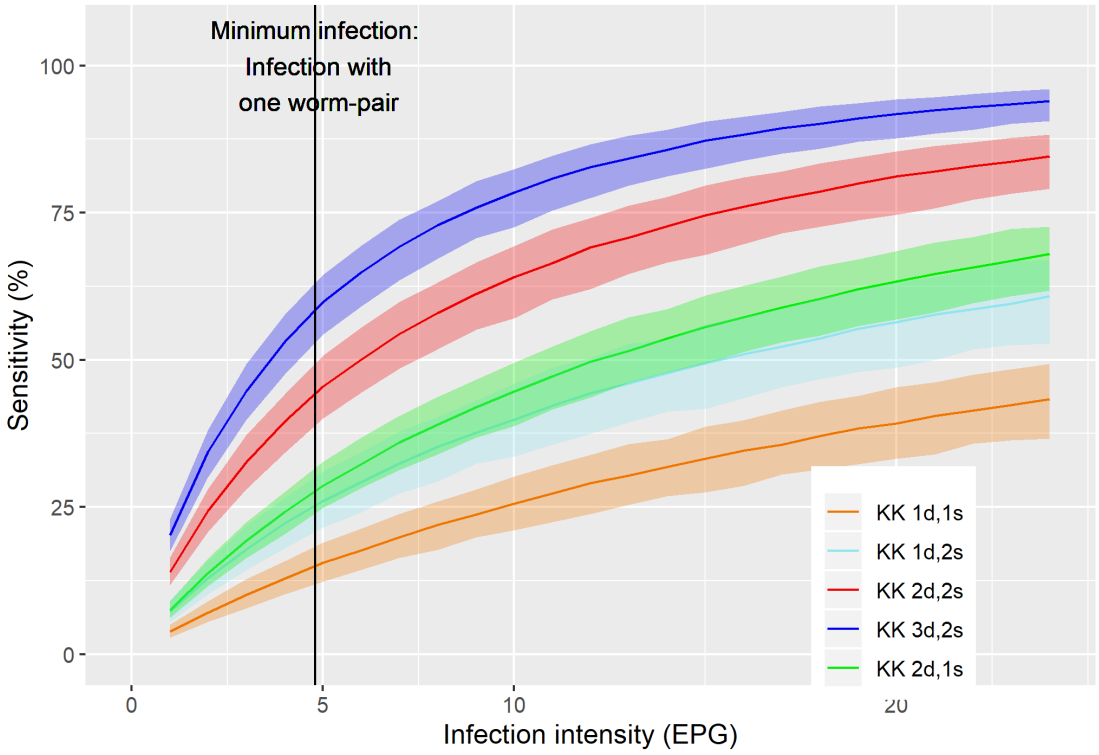


Figure 4.2: Infection intensity dependent sensitivity of Kato-katz for intensities up to 25 EPG for five different sampling schemes. The number before the d denotes the number of days where stool samples are collected, and the one before s the number of slides per stool sample. The vertical line denotes an estimate of the infection intensity of an infection with a single worm-pair.

Figure 4.3 shows total sensitivity in a population of 50 individuals, that is the mean sensitivity across all positive individuals, in relation to arithmetic and geometric mean infection intensity of the total population for six different sampling schemes. There is a clear relation between infection intensity and sensitivity for both geometric and arithmetic mean while variations for the former are smaller. Sampling on three days with three slides leads to very high overall sensitivity over 80% for even very low mean infection intensity. Two slides on a single day, the sampling scheme recommended by WHO, shows a sensitivity between 40 and 80% depending on infection intensity.

The relation between the mean of only the positive individuals and the total sensitivity is shown in Figure A1 in the appendix. The relation between infection intensity and sensitivity is still visible in both arithmetic and geometric mean of the positives but variations are much larger.

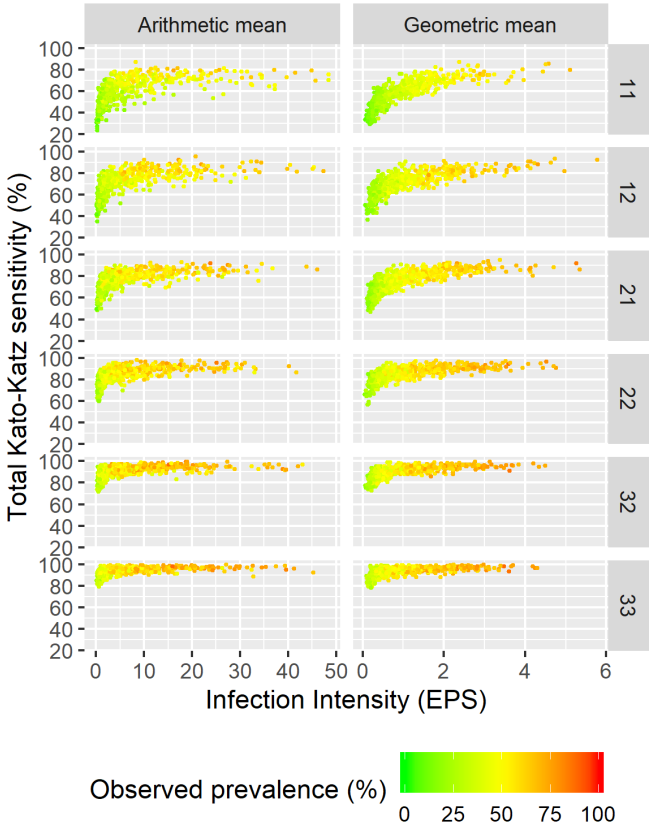


Figure 4.3: Overall sensitivity of Kato-Katz in relation to geometric and arithmetic mean of total population for six different sampling schemes in simulations with 50 individuals per location. The sampling scheme is denoted on the right side with the first number referring to the number of stool samples and the second to the number of slides per stool sample. Observed prevalence is shown in colour.

Estimation of mean of the positives only is based on a much lower sample size than mean of the total population and hence, carries a larger uncertainty. Therefore, we fit the statistical model to the means of the total population only to profit from the lower uncertainty.

The relation between observed prevalence and 'true' prevalence is presented in Figure 4.4 for sample sizes of 30, 50, and 5,000 and each of the six sampling schemes. The relation appears fairly linear. As expected the larger the sample size, the smaller the uncertainty. The observed prevalence is almost equivalent to 'true' prevalence for samples based on three days with three slides each, confirming that sensitivity is very high for that sampling scheme. Uncertainty is

much larger for lower sample sizes of 30 or 50 making the 'true' prevalence more difficult to estimate.

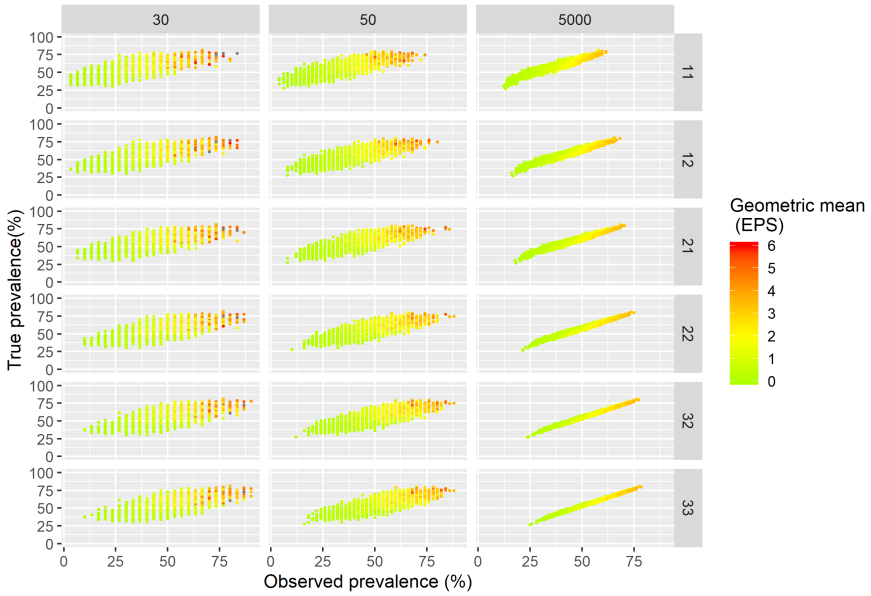


Figure 4.4: Relation between observed prevalence and 'true' prevalence for sample sizes of 30, 50, and 5,000 with geometric mean of the total population in colour. The sampling scheme applied is denoted on the right hand side where the first number is the number of days and the second the number of slides per day.

4.3.3 Relation between observed summary measures and 'true' prevalence

The statistical model described in Section 4.2.4 was fitted to the simulated data to determine the relations between the arithmetic and geometric mean of a population with the diagnostic sensitivity as well as the relation between the observed and 'true' prevalence for all sampling schemes.

Table 4.2 shows the posterior means for the parameters defining the relation between diagnostic sensitivity of different Kato-Katz sampling schemes and the geometric mean intensity of the total population. The mean estimate for the sensitivity can be calculated from the parameters using Equations 4.5 given the geometric mean infection intensity. a_0 the parameter that describes the sensitivity for very low mean infection intensity increases steadily with larger sampling effort

indicating a higher sensitivity. The change is less pronounced in the fit for sample size 30. There is a slight reduction in b_1 the parameter describing the change of the uncertainty in the sensitivity with infection intensity.

Scheme	a_0	a_1	a_2	b_0	b_1
N=50					
11	-7.19 (-9.55 - -5.36)	8.98 (7.19 - 11.31)	13.54 (10.28 - 17.73)	-4.54 (-4.69 - -4.39)	-0.06 (-0.12 - 0)
12	-6.24 (-8.55 - -4.39)	8.45 (6.63 - 10.73)	13.51 (10.01 - 17.92)	-4.45 (-4.6 - -4.3)	-0.08 (-0.14 - -0.02)
21	-6.38 (-8.89 - -4.36)	8.86 (6.88 - 11.34)	13.92 (10.16 - 18.63)	-4.54 (-4.7 - -4.4)	-0.09 (-0.15 - -0.02)
22	-5.09 (-7.62 - -3.09)	8.06 (6.1 - 10.57)	13.52 (9.57 - 18.59)	-4.62 (-4.77 - -4.47)	-0.14 (-0.2 - -0.08)
32	-4.12 (-6.61 - -2.26)	7.57 (5.73 - 10.03)	13.29 (9.37 - 18.41)	-4.97 (-5.13 - -4.82)	-0.25 (-0.31 - -0.19)
33	-3.3 (-5.54 - -1.4)	7.08 (5.22 - 9.28)	13.32 (9.06 - 18.3)	-5.3 (-5.45 - -5.14)	-0.33 (-0.39 - -0.27)
N=30					
11	-6.03 (-8.03 - -4.54)	7.86 (6.41 - 9.83)	11.41 (8.73 - 15)	-4.11 (-4.24 - -3.97)	-0.07 (-0.12 - -0.01)
12	-5.16 (-7.37 - -3.5)	7.44 (5.82 - 9.61)	11.25 (8.14 - 15.36)	-3.99 (-4.13 - -3.85)	-0.09 (-0.15 - -0.04)
21	-5.41 (-7.68 - -3.6)	7.93 (6.16 - 10.17)	12.03 (8.66 - 16.25)	-4.1 (-4.24 - -3.96)	-0.1 (-0.16 - -0.05)
22	-4.12 (-6.43 - -2.29)	7.12 (5.34 - 9.39)	11.55 (7.93 - 16.11)	-4.15 (-4.29 - -4)	-0.17 (-0.22 - -0.11)
32	-3.62 (-6.04 - -1.67)	7.05 (5.14 - 9.46)	12.42 (8.26 - 17.6)	-4.42 (-4.56 - -4.27)	-0.22 (-0.28 - -0.17)
33	-2.8 (-5.12 - -0.96)	6.56 (4.75 - 8.83)	12.51 (8.25 - 17.72)	-4.76 (-4.9 - -4.61)	-0.31 (-0.37 - -0.25)

Table 4.2: Posterior estimates (mean and 95% BCI) of the parameters describing the relation between geometric mean of the total population and sensitivity for six sampling schemes and two sample.

There is a good fit in the relation between the geometric mean infection intensity and the diagnostic sensitivity in a population (see Figure 4.5). Tables A2 and A4 that present the parameter values of the fit between arithmetic mean infection intensity and sensitivity as well as observed prevalence and 'true' prevalence can be found in the appendix.

The relation between the arithmetic mean infection intensity and the sensitivity (Figure A3, appendix) is very similar to the one determined for the geometric mean infection but the uncertainty is slightly larger possibly because the arithmetic mean is influenced more by outliers. Taking three Kato-Katz slides from three stool samples achieves observed prevalence basically equivalent to the 'true' prevalence while for two Kato-Katz slides from one stool sample the observed prevalence is only half of the 'true' prevalence. For example, an observed prevalence of 25% corresponds to a 'true' prevalence of 50%. The relation between observed and 'true' prevalence appears linear for each sampling scheme (Figure A5, appendix), however, uncertainty is large for a sample size of only 50.

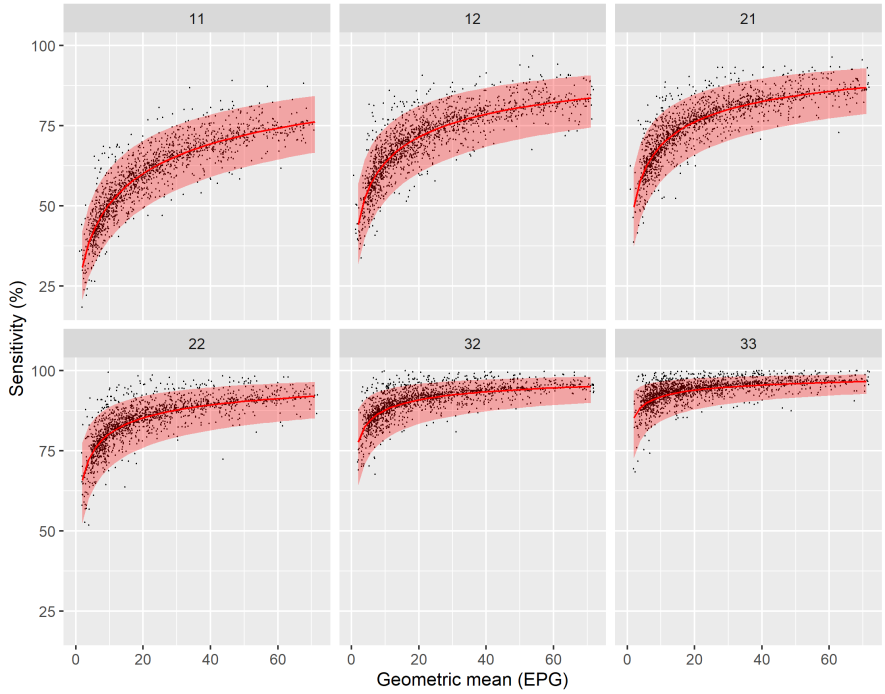
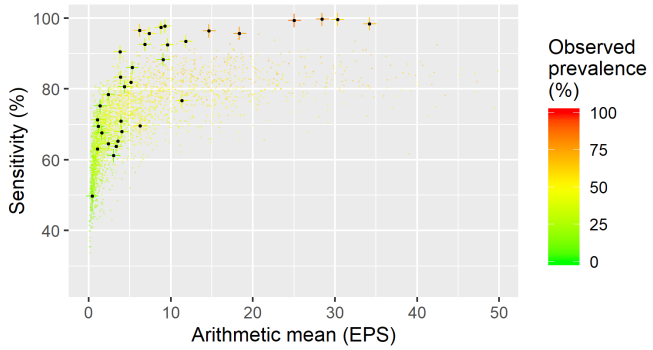


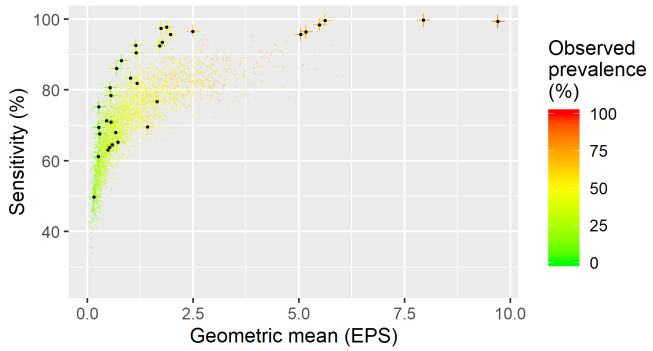
Figure 4.5: Estimated relation between geometric mean infection intensity of the total population and sensitivity (posterior mean and 95% BCI as red line and shading) for each sampling scheme and a sample size of 50. The black dots indicate the simulated data.

4.3.4 Validation

To see if the simulated populations adequately match real populations we fit an individual level model to data from 34 schools of about 50 individuals with observed prevalence above 10% from two Kato-Katz slides on one day. Comparison with simulation results is done in Figure 4.6 where we plotted the estimated sensitivities for the validation data in relation to arithmetic and geometric mean infections as well as all the simulated populations for two slides of Kato-Katz on a single sample. For low infection-intensities, the relations match reasonably well, starting off around 50% and increasing to 80% for 5 EPS arithmetic mean and 1.25 EPS geometric mean, respectively. The simulations underestimate the sensitivity for larger infection intensity by never surpassing 90% even for high infection intensity. This indicates that in the simulation there is a larger number of light infections compared to the validation datasets.



(a)



(b)

Figure 4.6: Estimated sensitivity in relation to infection intensity for a) geometric and b) arithmetic mean and for two Kato-Katz slides of a single stool sample. Black dots on crosses are the validation results from 34 locations in Uganda and the dots are the simulated results.

4.4 Discussion

We estimated the 'true' prevalence of *S. mansoni* from population summary measures of Kato-Katz diagnostics for sampling schemes varying from a single slide on one day to three slides on three days. We considered two population summary measures of infection intensity, the arithmetic and geometric mean infection intensity for both the full population or only the positive individuals and observed prevalence. The aim is to improve estimation of the 'true' prevalence from the basic approach of using a constant sensitivity when there is no access to individual level data.

We determined the infection intensity-dependent sensitivity for individuals by fitting a Bayesian egg count model for Kato-Katz to data from different transmission settings in sub-Saharan Africa. We have previously shown that it is possible to infer on the 'true' prevalence

from individual level data[Bärenbold et al., 2017, 2018]. Severe infections above 400 EPG are reliably detected even from a single stool sample while for infection intensities between 100 and 400 EPG Kato-Katz diagnostics, two days are necessary to achieve more than 90% sensitivity. Very light infections below 10 EPG are difficult to detect even for Kato-Katz performed on three consecutive days. Thus, diagnostic sensitivity mostly affects light infections. An infection with a single worm-pair has been estimated to be between 1 and 5 EPG a range where the sensitivity of Kato-Katz is strongly dependent on the infection intensity [Cheever et al., 1994]. For example, for three stool samples with two slides each, sensitivity of Kato-Katz is 20% at 1 EPG but 50% at 4 EPG. Therefore, the number of samples needed to detect even infection with a single worm-pair cannot easily be inferred.

Simulating populations was done using the assumptions of a negative binomial distribution of worms and monogamous mating [May, 1977; Anderson, 1986; May and Woolhouse, 1993]. Geometric and arithmetic means were calculated for both, the total population and the infected individuals alone as it was expected that the infection intensity of infected people could contain valuable information about the sensitivity. However, estimating means from only the positives in sample sizes of 30 or 50 individuals leads to large uncertainty making this measure impractical. Our results showed the strong dependence of the sensitivity on the infection intensity confirming once more that a constant sensitivity is not sufficient to estimate 'true' prevalence. The assumption of negative binomial distribution of worms in a population links the prevalence of the disease to the mean worm burden. Furthermore, the relation between the observed and 'true' prevalence is rather linear.

The statistical models used to describe the relation between sensitivity and infection intensity are based on the assumption that the sensitivity is monotonically increasing with infection-intensity and reaches 100% for very severe infections. We also assumed that the relation between the observed and the 'true' prevalence is monotonically increasing and at 100% observed prevalence the 'true' prevalence is 100% too. Estimating not just the mean but a whole distribution about the sensitivity enabled us to quantify the uncertainty of the sensitivity and therefore obtain prior distributions for Bayesian models including this parameter. Translation of the observed prevalence to the 'true' one can be done by hand using the parameters given in the results

section and provided in the appendix or an excel table can be created to facilitate application by practitioners. De Vlas et al. (1997) developed and validated a chart to translate Kato-Katz data from a single stool sample to 'true' prevalence [de Vlas et al., 1997]. Here we present parametric functions for six different sampling schemes to estimate both sensitivity and 'true' prevalence from arithmetic and geometric mean infection intensities.

The validation data agrees with the simulated one for lower infection intensities. However, sensitivity saturates around 90% in the simulated data while fitting a model to individual level data estimates a sensitivity of close to 100% being reached. This indicates that either the negative binomial distribution does not accurately capture the distribution of worms at high infection intensities but over-estimates the number of light infections or the gamma distribution under-estimates the number of light infections. The former could imply that at high transmission intensity, exposed individuals harbour a large number of worms while a certain number of people is still not exposed at all leading to fewer light infections than expected under the negative binomial assumption. An improvement of our model could be to vary the worm aggregation parameter with infection intensity.

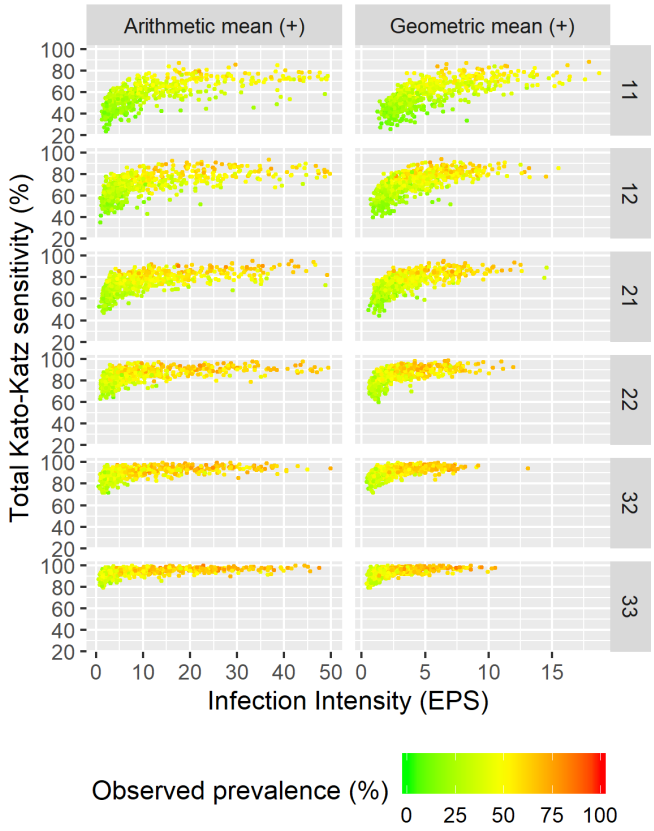
The uncertainty about the number of light infections of *S. mansoni* is important in the context of disease elimination. Discussions about diagnostic specificity of the point-of-care circulating cathodic antigen diagnostic technique (POC-CCA) include a similar argument when considering whether the additional positives by POC-CCA are true infections or false-positives [Colley et al., 2017]. Haggag et al. (2019) find that a large number of POC-CCA positives but Kato-Katz negatives shows no egg excretion even when samples are taken over a complete month which supports that the number of very light infections might be over-estimated [Haggag et al., 2019b].

4.4.1 Conclusion

In conclusion, we showed that there is important variation in the population level sensitivity of Kato-Katz, i.e. the ratio between observed and 'true' prevalence, with mean infection intensity. We confirmed that the relation between the sensitivity and the infection intensity can be simulated using our egg count model for Kato-Katz observations combined with standard assumptions for the worm distribution in a population. Our parametric model fitted on simulated data can be

used to translate observed prevalence into ‘true’ prevalence when either the arithmetic or the geometric mean infection intensity are available. Moreover, our results provide Bayesian priors when modelling historical survey data aggregated at the population level.

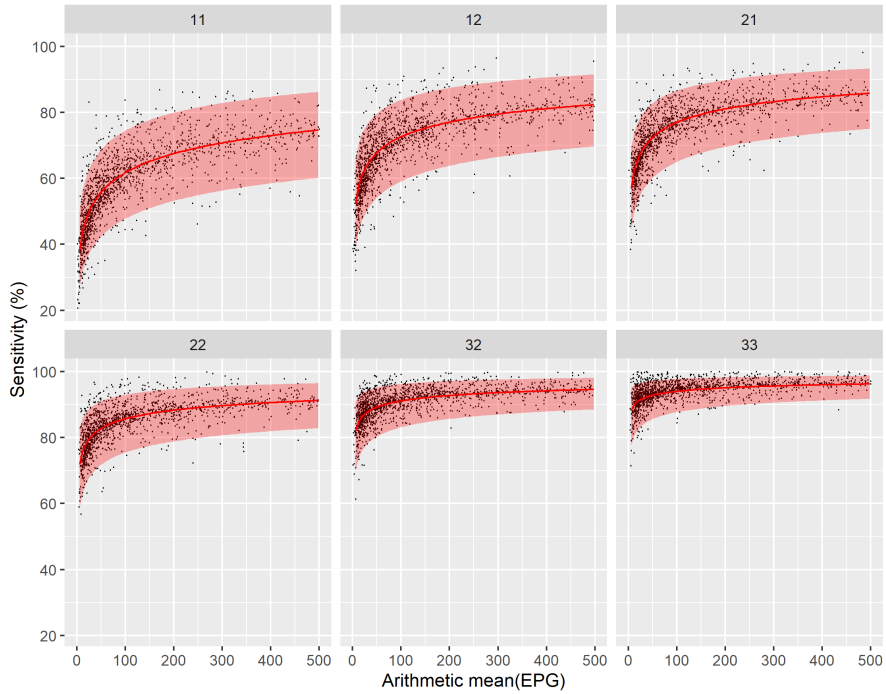
Appendix



A1: Overall sensitivity of Kato-Katz in relation to geometric and arithmetic mean of only positive individuals for six different sampling schemes in simulations with 50 individuals per location. The sampling scheme is denoted on the right side with the first number referring to the number of stool samples and the second to the number of slides per stool sample. Observed prevalence is shown in colour.

Scheme	a_0	a_1	a_2	b_0	b_1
N=50					
11	-7.81 (-9.89 - -5.98)	8.94 (7.12 - 11.02)	23.39 (18.22 - 29.35)	-3.71 (-3.8 - -3.61)	0.11 (0.07 - 0.15)
12	-6.23 (-8.32 - -4.46)	7.81 (6.03 - 9.9)	21.84 (16.45 - 28.1)	-3.82 (-3.92 - -3.73)	0.04 (0 - 0.08)
21	-6.41 (-8.76 - -4.56)	8.24 (6.39 - 10.58)	22.05 (16.69 - 28.99)	-4.01 (-4.11 - -3.91)	0.02 (-0.02 - 0.06)
22	-4.51 (-6.46 - -2.9)	6.87 (5.26 - 8.81)	19.79 (14.68 - 25.97)	-4.2 (-4.3 - -4.1)	-0.05 (-0.1 - -0.01)
32	-3.28 (-5.15 - -1.73)	6.14 (4.59 - 8)	18.31 (13.13 - 24.5)	-4.5 (-4.6 - -4.41)	-0.12 (-0.17 - -0.08)
33	-2.28 (-4.14 - -0.76)	5.51 (3.99 - 7.37)	17.7 (12.3 - 24.29)	-4.78 (-4.88 - -4.68)	-0.18 (-0.22 - -0.14)
N=30					
11	-7.96 (-10.14 - -6.12)	9.12 (7.29 - 11.3)	23.92 (18.65 - 30.15)	-3.32 (-3.42 - -3.23)	0.11 (0.07 - 0.14)
12	-6.33 (-8.32 - -4.65)	7.94 (6.26 - 9.92)	22.15 (16.99 - 28.25)	-3.37 (-3.47 - -3.28)	0.04 (0 - 0.07)
21	-5.88 (-7.89 - -4.24)	7.74 (6.11 - 9.76)	20.46 (15.67 - 26.34)	-3.57 (-3.66 - -3.47)	0.02 (-0.02 - 0.06)
22	-4.08 (-5.98 - -2.48)	6.46 (4.86 - 8.35)	18.5 (13.37 - 24.61)	-3.7 (-3.8 - -3.61)	-0.05 (-0.09 - -0.01)
32	-3.11 (-5.03 - -1.53)	5.95 (4.38 - 7.89)	18.28 (12.93 - 24.77)	-4 (-4.1 - -3.91)	-0.11 (-0.15 - -0.07)
33	-2.1 (-3.83 - -0.67)	5.31 (3.89 - 7.04)	17.71 (12.32 - 24.15)	-4.29 (-4.38 - -4.19)	-0.17 (-0.21 - -0.13)

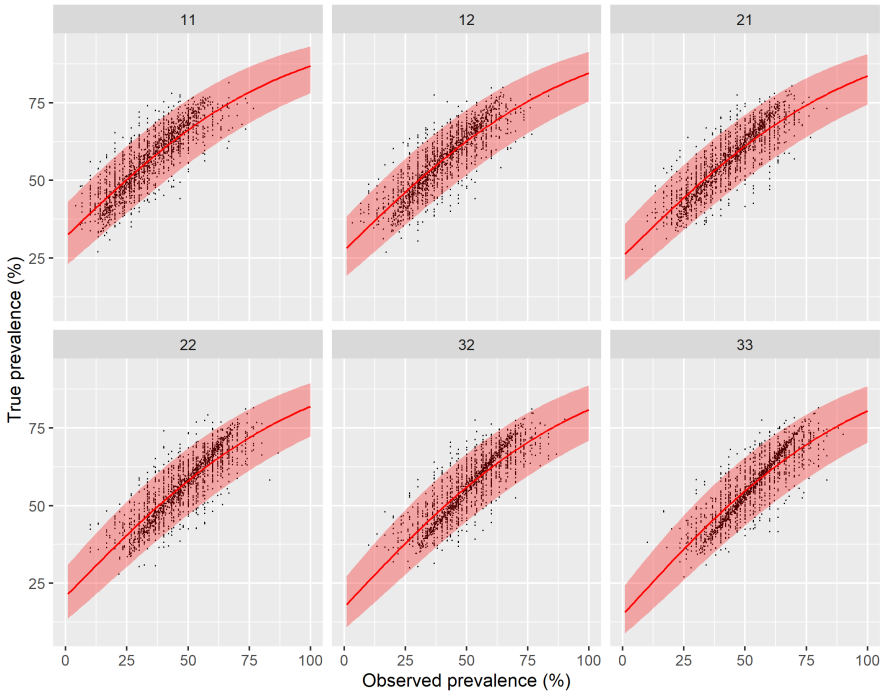
A2: Posterior estimates (mean and 95% BCI) of the parameters describing the relation between arithmetic mean of the total population and sensitivity for six sampling schemes and two sample.



A3: Estimated relation between arithmetic mean infection intensity of the total population and sensitivity (posterior mean and 95% BCI as red line and shading) for each sampling scheme and a sample size of 50. The black dots indicate the simulated data.

Scheme	a_0	a_1	b_0	b_1
N=50				
11	0.32 (0.31 - 0.33)	2.21 (2.17 - 2.25)	0.01 (0.01 - 0.01)	0 (0 - 0)
12	0.28 (0.27 - 0.28)	2.11 (2.07 - 2.14)	0.01 (0.01 - 0.01)	0 (0 - 0)
21	0.26 (0.25 - 0.26)	2.07 (2.03 - 2.11)	0.01 (0.01 - 0.01)	0 (0 - 0)
22	0.21 (0.2 - 0.22)	2.03 (1.99 - 2.07)	0.01 (0.01 - 0.01)	0 (0 - 0)
32	0.18 (0.16 - 0.19)	2.02 (1.98 - 2.05)	0.01 (0.01 - 0.01)	0 (0 - 0.01)
33	0.15 (0.14 - 0.16)	2.02 (1.98 - 2.06)	0.01 (0.01 - 0.01)	0 (0 - 0.01)
N=30				
11	0.36 (0.35 - 0.36)	1.98 (1.93 - 2.02)	0.02 (0.02 - 0.02)	0 (0 - 0)
12	0.32 (0.31 - 0.33)	1.88 (1.84 - 1.92)	0.02 (0.01 - 0.02)	0 (0 - 0)
21	0.31 (0.3 - 0.31)	1.84 (1.8 - 1.88)	0.02 (0.01 - 0.02)	0 (0 - 0)
22	0.27 (0.26 - 0.28)	1.8 (1.75 - 1.84)	0.02 (0.02 - 0.02)	0 (0 - 0)
32	0.24 (0.23 - 0.25)	1.79 (1.75 - 1.83)	0.02 (0.02 - 0.02)	0 (0 - 0.01)
33	0.23 (0.21 - 0.24)	1.79 (1.74 - 1.83)	0.02 (0.02 - 0.02)	0 (0 - 0.01)

A4: Posterior estimates (mean and 95% BCI) of the parameters describing the relation between observed prevalence and 'true' prevalence for six sampling schemes and two sample.



A5: Estimated relation between observed prevalence and 'true' prevalence (posterior mean and 95% BCI as red line and shading) for each sampling scheme and a sample size of 50. The black dots indicate the simulated data.

Chapter 5

The influence of Kato-Katz and POC-CCA diagnostics on survey sampling designs for assessing *S. mansoni* treatment needs

Authors

Oliver Bärenbold[1,2], Christos Kokaliaris[1,2], Amadou Garba[3], Ameyo M. Dorkenoo[4,5], Jürg Utzinger[1,2], Penelope Vounatsou[1,2]

1 Swiss Tropical and Public Health Institute, Basel, Switzerland

2 University of Basel, Basel, Switzerland

3 Department of Control of Neglected Tropical Diseases, World Health Organization, Geneva, Switzerland

4 National Lymphatic Filariasis Elimination Program, Ministry of Health and Social Protection, Lomé, Togo

5 Department of Fundamental Science, University of Lomé, Lomé, Togo

Abstract

Background

Preventive treatment with Praziquantel is the main intervention for *S. mansoni*. Treatment schedules depend on infection prevalence thresholds based on Kato-Katz diagnostic and estimated at district level from sampling surveys. Recently, efforts have been made to translate the thresholds into an antigen-based diagnostic, namely the POC-CCA that promises higher sensitivity especially for light infection intensities. However, the influence of the diagnostic technique on the survey sampling design requirements is not known.

Methodology

A number of survey sampling designs were evaluated based on simulated data on the mean number of worms, prevalence of worms, and worm-pairs following standard assumptions for the distribution of worms in a population. Subsequently, the measurement process for Kato-Katz and POC-CCA was simulated at the individual level based on an egg count model that incorporates the infection-intensity dependent sensitivity of the diagnostic tests. Simulations were carried out on real data from the 2009 national survey in Togo, sampling from the actual locations of the communities in the country.

Principal findings

We find that the relative error in observed district prevalence is about 50% lower for POC-CCA compared to Kato-Katz for the same survey design. Taking more locations per districts with fewer individuals is clearly superior compared to increasing the number of individuals per location at a constant total sample size per district. Improvements in relative error become marginal above 500 tested individuals per district. Testing only 5 locations in a district with Kato-Katz leads to under-treatment in up to 40% of the districts while the corresponding ratio of districts

under-treated with POC-CCA is 20%. When the number of locations per district increases to 20 the above difference between diagnostic methods becomes small.

Conclusions/significance

Increasing the number of locations per district while decreasing the number of individuals tested per location is preferable for *S. mansoni* which is a focal disease. Surveillance based on the POC-CCA diagnostic method requires fewer locations per district being tested than Kato-Katz to achieve a low number of under-treated districts.

Acknowledgements

This study received financial support from the European Research Council (ERC-2012-AdG-323180). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

5.1 Introduction

Schistosomiasis is a neglected tropical disease causing a considerable burden, primarily in sub-Saharan Africa. Currently, there are about 800 million people at risk of which 190 million are infected inflicting more than 2 million DALYs lost [Lai et al., 2015; GBD 2016 DALYs and HALE Collaborators, 2017; GBD 2016 Disease and Injury Incidence and Prevalence Collaborators, 2017; GBD 2016 Causes of Death Collaborators, 2017]. The World Health Organization (WHO) set goals to eliminate schistosomiasis as a public health problem by 2025 [WHO, 2013]. The preferred method of disease control is periodic mass drug administration with Praziquantel while improved water and sanitation infrastructure (WASH) and snail control are supplementary options for morbidity control [McManus et al., 2018].

The disease is caused by a parasitic flatworm of the genus *Schistosoma* with a complex life cycle that includes fresh water snails. *S. mansoni*, the species considered in this study, resides in blood vessels next to the intestines in pairs, producing eggs that are excreted through faeces. Infection occurs through contact with water, contaminated with free-swimming cercariae after expulsion from their intermediate host-snail. In chronic disease, eggs are trapped in host tissue causing an inflammatory and obstructive reaction. Schistosomiasis causes abdominal pain, diarrhoea, anaemia and growth stunting and plays a role in the development of cancer and transmission of HIV [McManus et al., 2018]. Due to the unspecific nature of symptoms, chronic schistosomiasis represents an under-recognized disease burden [King and Dangerfield-Cha, 2008].

For estimating treatment needs within an ecologically homogeneous region, the World Health Organization (WHO) recommends a two stage-cluster, school-based survey sampling design which includes 5 schools with 35-50 children per school using two slides of Kato-Katz per person, a parasitological method to detect eggs in a stool smear of 41.7 mg [Katz et al., 1972; Montresor et al., 1998]. Decisions about treatment needs are then made at district level based on the average observed prevalence in the district. Tchuem Tchuente et al. [2018] suggest that this design is not suitable anymore due to high disease focality and advocate targeted treatment in order to reach the global target of elimination of schistosomiasis as a public health problem by 2025. Knowles et al. [2017] suggest that a survey design including 15 schools with 20-30 children per school

is more reliable and cost-effective in detecting endemic schistosomiasis. Giardina et al. [2019] reached similar conclusions in a modelling study looking at sampling designs for evaluating a prevalence of soil-transmitted helminths below 1%.

Apart the arguments about the appropriate number of survey locations per district and of individuals per location, the role of the diagnostic method in the survey design has not been investigated. Imperfect and varying diagnostic sensitivity affects mapping efforts across Africa and the lack of sensitivity of Kato-Katz has been noted repeatedly in the past years [Kongs et al., 2001; Chammartin et al., 2013]. Furthermore, Turner et al. [2017] provides justifications for considering alternatives to Kato-Katz. Recently, the point-of-care circulating cathodic antigen (POC-CCA), a diagnostic test detecting a specific antigen in urine, was shown to have lower error, especially in settings with low infection intensities [Colley et al., 2013; Ochodo et al., 2015; Danso-Appiah et al., 2016]. Moreover, the POC-CCA is sensitive to infections without egg-producing worm-pairs. This leads to a discrepancy in measured prevalence by POC-CCA compared to Kato-Katz that has to be better understood in order to differentiate between false-positives and infections without worm-pairs [Colley et al., 2017]. Recently, two studies by Haggag et al. show that at least a proportion of the trace positives but Kato-Katz negatives are false-positives and raise questions regarding the usability of the test in settings approaching elimination [Haggag et al., 2019b,a]. Bärenbold et al. (2018) translated the WHO recommended intervention thresholds based on Kato-Katz into POC-CCA analogues supporting further the use of the latter for disease surveillance [Bärenbold et al., 2018]. To our knowledge, no work has been done regarding the influence of the diagnostic technique on the accuracy of survey designs.

We assessed survey sampling designs for *S. mansoni* based on simulation data on the mean worm number, prevalence of worms and worm-pairs following standard assumptions for the distribution of worms in the population [May, 1977; Anderson, 1986; May and Woolhouse, 1993]. The measurement process generating the observed data was simulated for each individual assuming Kato-Katz and POC-CCA diagnostics using an egg count model which takes into account the infection intensity and the sensitivity of each diagnostic [Bärenbold et al., 2018]. Simulations were carried out based on real data from the 2009 national survey in Togo, sampling from the actual location of the communities in the country [Dorkenoo et al., 2012].

5.2 Material and methods

5.2.1 Simulation scenario

We simulated the spatial distribution of worms and worm-pairs of *S. mansoni* in Togo based on a grid of 'observed' prevalences by Kato-Katz, obtained by the predictions of a geostatistical model fitted on the 2009 Kato-Katz Togo survey data [Dorkenoo et al., 2012].

In particular, for each community j, (in Togo, $j = 1, \ldots, 3631$), we simulate the average number of worms μ_j assuming that $\log(\mu_j) = \mu_0 + \beta p_j + \phi_j$, where p_j is the prevalence at j obtained by a geostatistical model, fitted on the Togo survey data, ϕ_j is a Gaussian spatial process across locations j,

$$\phi \sim N(0, \tau^2 I + \sigma_1^2 R) \text{ where } R_{nm} = \exp(-\rho d_{nm}) \tag{5.1}$$

where d_{nm} is the distance between locations n and m, ρ defines the range of the spatial process, σ_1^2 the strength of the spatial correlation, and τ^2 the local random effect. For each individual i at location j we simulate the worm burden from a negative binomial distribution, $w_{ji} \sim NB(\mu_j, k_1)$ where k_1 is the variation of the worm counts in the population. We assumed that each worm has a 30% probability to be female and calculated the number of worm pairs $w_{ji}^{(p)}$ accordingly as the minimum of the numbers of males and females in an individual considering perfect mating. The worm and worm-pair burden determine the infection status D_{ji} of individual i, where $D_{ji} = 0$ if $w_{ji} = 0$, $D_{ji} = 2$ if $w_{ji}^{(p)} > 0$, and $D_{ji} = 1$ otherwise.

Observations for POC-CCA and double slide Kato-Katz are simulated following the model developed by Bärenbold et al (2018). For individuals with active worm-pairs ($D_{ji} = 2$) the infection intensity λji is defined as $0.2 \cdot w_{ji}^{(p)}$ assuming that a single worm-pair has an egg output of 0.2 eggs per slide [Chan et al., 1995; Gurarie and King, 2014]. POC-CCA observations are simulated from a Bernoulli distribution, that is, $Y_{ji}^{CCA} \mid D_{ji} = 2 \sim Be(s_{ji}^{CCA})$ with $\text{logit}(s_{ji}^{CCA} \cdot a_2) = a_0 + a_1\sqrt{\lambda_{ji}}$ where s_{ji} is the sensitivity parameter. a_0 determines the sensitivity of POC-CCA for very light infections, a_1 describes the dependence of the sensitivity on the infection intensity, and a_2 determines the limit of the sensitivity for severe infections.

Kato-Katz observations for the same individual are simulated taking into account the day-to-day, and the slide-to-slide variation, modelling $Y_{jids} \mid D_{ji} = 2 \sim NB(\lambda_{jid}, \gamma_2)$, where

$\log(\lambda_{jid}) = \log(\lambda_{ji}) + \epsilon_{jid}$ and $\epsilon_{jid} \sim N(-\sigma^2/2, \sigma)$ to ensure that the mean over d of λjid is λji. σ^2 captures the day-to-day variation, and γ_2 takes into account the non-random distribution of eggs within a sample. In the simulation, we draw egg counts for two slides on one day.

For individuals with no worm-pairs, i.e., $D_{ji} = 0, 1$, we assumed that the observed Kato-Katz counts are 0, corresponding to a specificity of 100%, that is $c^{KK} = 1$. Antigen output of people with living worms but no worm-pairs, $D_{ji} = 1$, was assumed to be similar to those with $D_{ji} = 2$ when λ_{ji} approaches zero. Therefore, POC-CCA results were simulated as from a Bernoulli distribution that is $Y_{ji}^{CCA} \mid D_{ji} = 1 \sim Be(s_{ji}^{CCA}(\lambda_{ji} = 0))$. For individuals that do not harbor worms, i.e. $D_{ji} = 0$, we simulated POC-CCA assuming t hat $Y_{ji}^{CCA} \mid D_{ji} = 0 \sim Be(1 - c^{CCA})$.

Parameters were chosen based on prior information. The specifications of the infection intensity dependence of Kato-Katz and POC-CCA were chosen based on Bärenbold et al. (2018), that is $a_0 = 1.1(0.62 - 1.67)$, $a_1 = 3.64(1.97 - 6.25)$, $a_2 = 5.58(4.55 - 6.87)$, $c^{CCA} = 0.98$, $\gamma_2 = 5.34(3.07 - 10.10)$, and $\sigma = 1.28(1.08 - 1.49)$ [Bärenbold et al., 2018]. β and μ_0 were chosen to replicate the magnitude of the true prevalence distribution with $\beta = 7$ and $\mu_0 = -1$. The spatial process was defined by $\tau^2 = 0.05$, $\sigma_1^2 = 2$, and the range $1/\rho = 40$ km. The worm aggregation parameter k_1 was simulated from a normal distribution with mean 0.5 and standard deviation 0.1 [Truscott et al., 2019]. Estimated prevalence based on the Togo survey data is shown for the grid and at the 3631 locations in Figure A1 in the appendix.

5.2.2 Survey designs

Survey designs were simulated by randomly drawing N individuals from M locations per district and subsequently calculating the observed prevalence by summarising over the district. We varied M to take the values 5, 10, 15, and 20, and consider N equals to 20, 40, and 60. Each combination was repeated 100 times to reduce sampling error. Performance of survey designs was evaluated first in comparison to the true district prevalence based on sampling 1000 individuals at each location in a district, and secondly, by comparing the accuracy of the treatment decisions made based on a given survey design. Accuracy of the district prevalence was calculated using the relative error $RE = |\pi_{obs} - \pi_{true}|/\pi_{true}$.

Classification of districts using Kato-Katz was done using WHO treatment guidelines [WHO,

2006]. For POC-CCA, two sets of thresholds were considered; one proposed by Bärenbold et al (2018) and a second one more conservative [Bärenbold et al., 2018].

5.3 Results

5.3.1 Simulation

We simulated the mean number of worms per individual from the negative binomial model presented above with a spatially varying μ_j that depends on the prevalence p_j. Using the mean number of worms, and the worm-aggregation parameter, the prevalence of worms in a population is calculated and from the distribution of worms the prevalence of worm-pairs in the population. Simulating a large number of individuals at each location, we determined the true observed prevalence by POC-CCA and Kato-Katz based on a large sample size of $N = 1000$ per location. In Table 5.1 we present mean and ranges of the simulated parameters across all communities and in Figure 5.1 maps of communities and corresponding simulation parameter. Similar Figures of grid estimates can be found in Figure A2 in the appendix.

To calculate the true classification of a district we took the mean over all communities in a district of the true observed prevalence by Kato-Katz and POC-CCA, respectively. Each district consists of between 10 and 542 communities. Using 1%, 5%, and 10% thresholds for the Kato-Katz based prevalence 6, 22, 9, and 0 districts were classified in the corresponding group, respectively. The first set of cut-offs for POC-CCA of 10%, 20%, and 30% implies a higher treatment need with 0, 12, 13, and 12 districts in each group. Higher cut-offs for CCA at 15%, 30%, and 50% classify the 4, 21, 12, and 0 districts in each group. Figure A3 in the appendix

Table 5.1: Mean and range of values of simulation parameters across all communities

	Mean	**Range**
Mean number of worms	0.84	0.005 - 76
Mean number of worm-pairs	0.18	0.0 - 22
Prevalence of worms (%)	29.0	0.5 – 94.4
Observed prevalence by POC-CCA(%)	25.1	1.7 – 91.9
Prevalence of worm-pairs (%)	9.3	0.0 – 87
Observed prevalence by Kato-Katz (%)	3.2	0.0 – 65

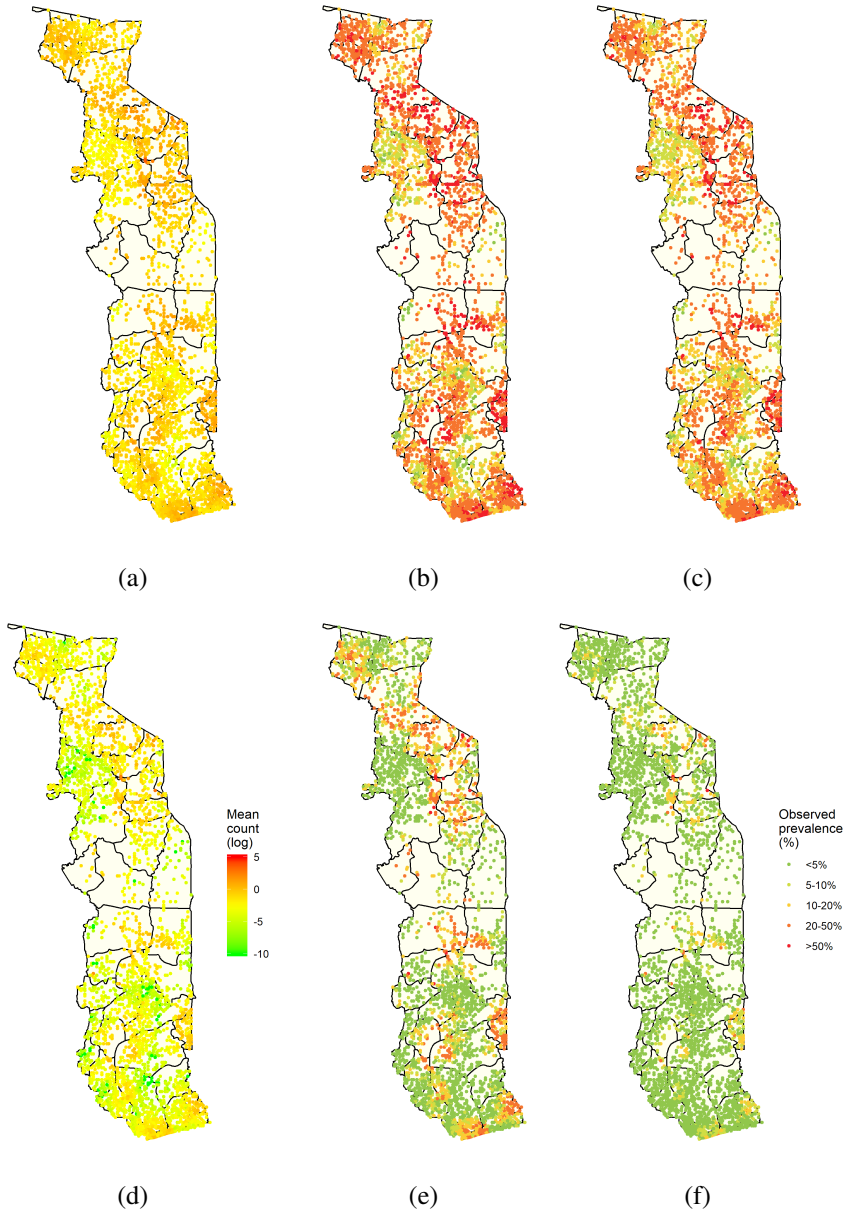


Figure 5.1: Parameters at community level of logarithmic mean number of worms (a), prevalence of worms (b), observed prevalence by POC-CCA (c), logarithmic mean number of worm-pairs (d) prevalence of worm-pairs (e), and observed prevalence by Kato-Katz (f). Colour scale for logarithmic mean worm counts is given in the lower left and for prevalence in the lower right.

show the spatial distribution of district classifications for each diagnostic and set of thresholds.

5.3.2 Survey Design

We used the simulated data above as a testbed to look at the impact of varying number of locations per district and of the number of people per location on the accuracy of prevalence estimation and classification of districts to treatment categories. The number of locations per district was chosen to be 5, 10, 15, and 20, and the number of individuals per location was set to 20, 40, and 60. Figure 5.2 shows the relative error of Kato-Katz and POC-CCA, respectively for various sampling designs. Relative errors for POC-CCA are generally much lower than Kato-Katz, varying from 20% error for 3 locations with 20 individuals each to almost 100% for the same survey design using the Kato-Katz diagnostic. The diagnostic error is more efficiently reduced when increasing the number of locations compared to the number of individuals per location.

For a fixed total number of individuals screened the diagnostic error is consistently lower when the number of locations is larger. POC-CCA performs better with lower diagnostic error, which happens at much higher observed prevalence, and thus, a higher absolute error. Diagnostic error seems to saturate in the order of 500 individuals per district.

Figure 5.3 compares the accuracy of classification based on Kato-Katz and the two sets of POC-CCA thresholds for each of the 12 survey designs, showing the proportion of districts over- or under-treated in each treatment group. Of most interest are the proportions of under-treated districts as over-treatment has fewer consequences. At 5 locations with 60 individuals each, which is close to strategy currently recommended by WHO, Kato-Katz leads to about 25% of the districts in the second and third treatment group being wrongly classified in a lower category than appropriate. Using POC-CCA in the same sampling design would reduce the proportion of under-treated districts by up to 50%. When 20 locations with only 20 individuals are chosen, all three classification methods perform equally well under-treating districts in the order of 10%. Generally, POC-CCA performs well already at sample sizes as low as 5 locations with 20 individuals screened while Kato-Katz performs consistently worse when only 5 locations per district are sampled.

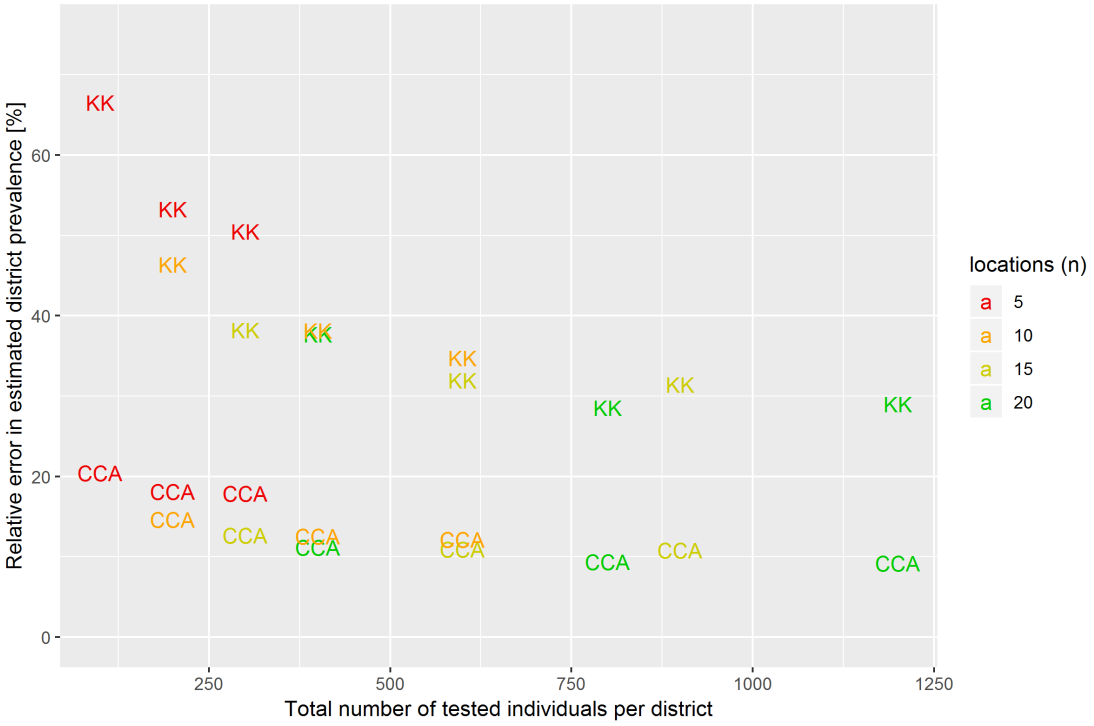


Figure 5.2: Dependence between the total number of individuals screened per district, the number of locations visited, and the estimated prevalence for both Kato-Katz and POC-CCA.

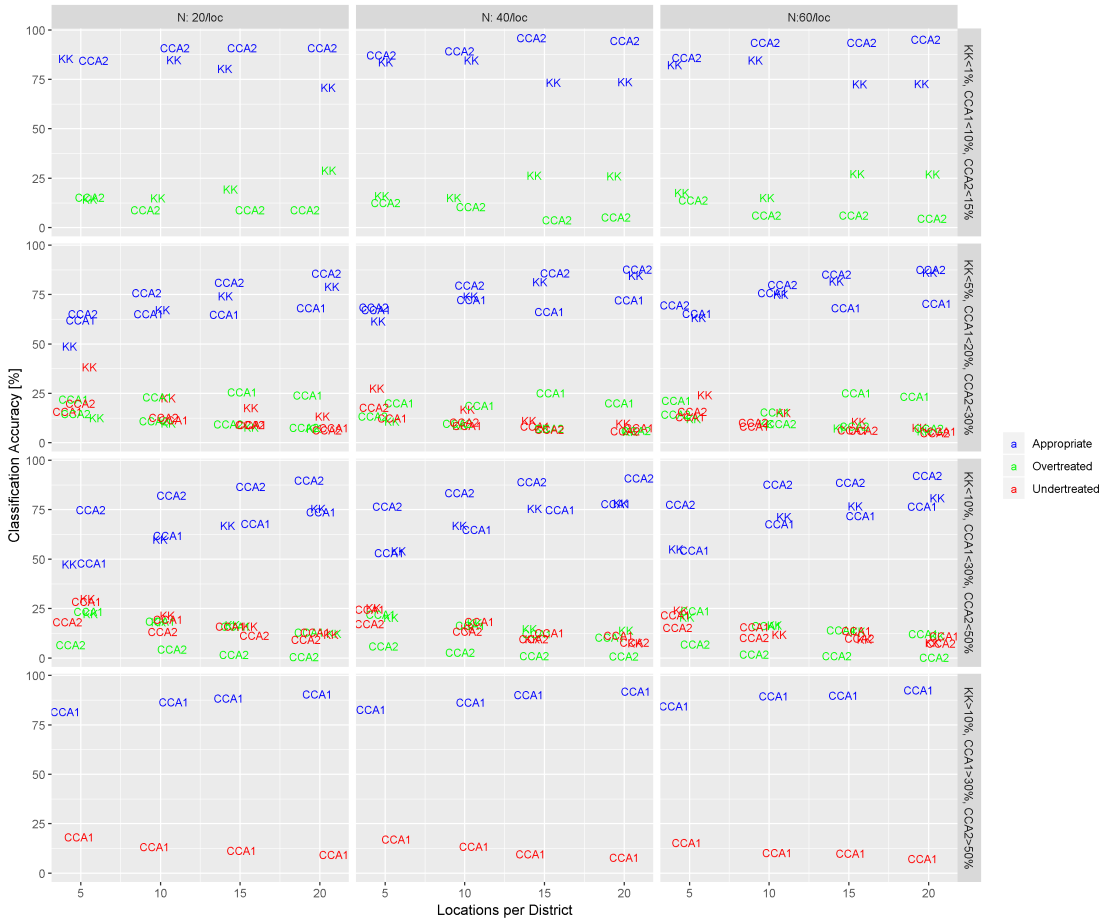


Figure 5.3: Accuracy of district classification for Kato-Katz and two sets of CCA thresholds for 20 to 60 individuals per location, and 5 to 20 locations per district.

5.4 Discussion

In this study, we compare the influence of diagnostic technique, specifically Kato-Katz and POC-CCA, on the accuracy of estimation of prevalence at district level using simulated data from Togo. We show that generally increasing the number of locations is much more effective than increasing the number of individuals per location. This is consistent with findings by Giardina et al. (2019) that study this question for soil-transmitted helminthiasis [Giardina et al., 2019] due to the focal distribution of the studied diseases. Knowles et al. (2017) find that 15 to 20 schools with 20 to 30 children per school lead to optimised cost-efficiency [Knowles et al., 2017]. Sturrock et al. (2010) also find that the numbers of individuals screened per school has little influence on accuracy of the survey in a study regarding optimal survey designs for soil-transmitted helminthiasis [Sturrock et al., 2010]. However, they deem 4 to 5 locations per district as the most cost-efficient strategy in their analysis of data from eastern Africa. The differences to findings for *S. mansoni* can occur due to different spatial heterogeneity of the diseases as well as changing cost estimates.

Alternatives to random sampling have been explored with lot quality assurance sampling (LQAS) and a variation of lattice plus close neighbour designs where locations are chosen to maximise spatial information being prominent ones [Sturrock et al., 2011]. Compared to random sampling both designs have the disadvantage of being more complicated to implement locally. Because decisions about the sample size in LQAS are made based on the results of the first 15 screened individuals the prevalence estimate is not unbiased. Sturrock et al. (2011) do not compare their spatial design to random sampling but note that LQAS performs better in classification tasks but at a higher cost.

Simulating worm distributions directly and calculating observed Kato-Katz and POC-CCA prevalence enabled us to evaluate the expected discrepancy due to the occurrence of single worm infections. The simulated grids show that widespread prevalence of up to 50% of the population harbouring worms is compatible with observed Kato-Katz results from Togo. Results from Haggag et al. [Haggag et al., 2019b,a] put into question whether trace-positive but egg-negative individuals actually harbour worms. Our simulation looks at the discrepancies between observed Kato-Katz and POC-CCA prevalence, explained by commonly used assumptions for

worm distributions in populations and mating behaviour but cannot answer the question of false-positives as the specificity is assumed to be $> 95\%$ (determined by our earlier study [Bärenbold et al., 2018]).

Intervention thresholds that were based on Kato-Katz have been translated to POC-CCA [Bärenbold et al., 2018]. We showed that the suggested translation of $< 1\%$,$< 5\%$,$< 10\%$,$> 10\%$ Kato-Katz into $< 10\%$,$< 20\%$,$< 30\%$,$> 30\%$ POC-CCA might lead to an overestimation of treatment needs. A more conservative set of thresholds, namely $< 15\%$,$< 30\%$,$< 50\%$,$> 50\%$, leads to s classification that is similar to the one from Kato-Katz thresholds. The discrepancy comes from the fact that the earlier publication did not use a model to link the mean infection-intensity, prevalence of worm-pairs and prevalence of worms together leading to a less accurate estimation of appropriate thresholds.

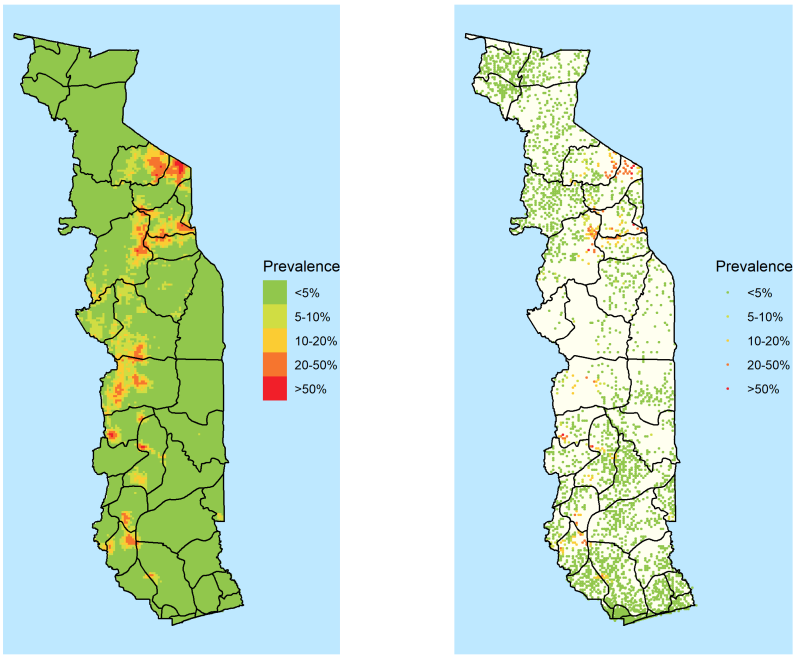
To our knowledge there is no earlier study looking at the influence of diagnostic technique on survey designs for POC-CCA and Kato-Katz. However, it has been acknowledged that taking into account diagnostic method when determining the spatial distribution of a disease is important [Chammartin et al., 2013; Giorgi et al., 2015; Amoah et al., 2018]. We show that the relative error in estimating district prevalence for POC-CCA is about half that of the Kato-Katz technique and decreases to only about 10% for sampling more than 300 individuals per district. This is partly due to the higher prevalence of worms in a population requiring a smaller sample size for the same relative error.

More important is however, how well the diagnostic methods perform in estimating treatment needs and especially how often a district is under-treated for a given survey design. We show that POC-CCA performs much better here as well. Kato-Katz leads to up to 40% of districts being under-treated for certain categories and lower effort survey designs with for example only 5 locations with 20 individuals. For larger sample sizes, the errors become comparable. Thus, rapid evaluation of treatment need with POC-CCA requires lower sample sizes for a reasonable accuracy.

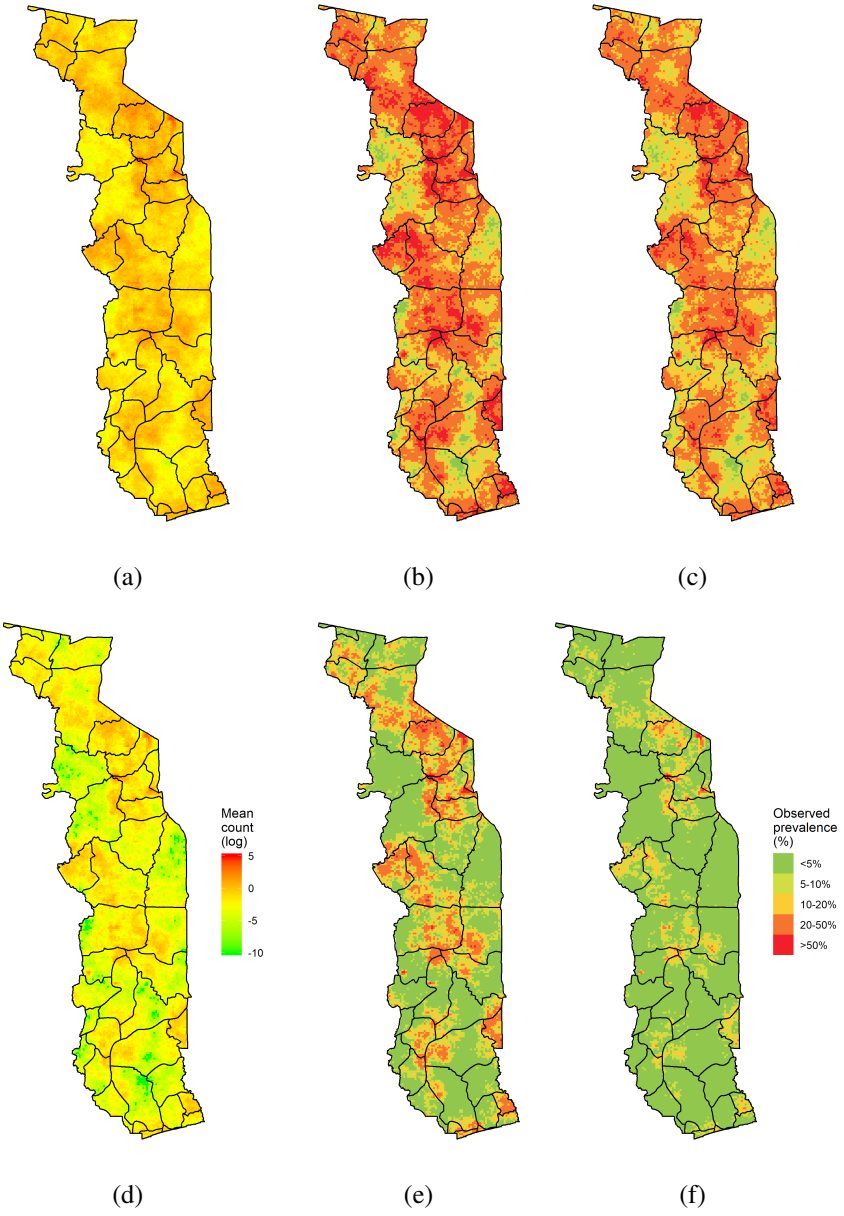
5.4.1 Conclusion

Simulating worm and worm-pair distributions directly can create realistic maps of observed Kato-Katz prevalence and explain large discrepancies between observed POC-CCA and Kato-Katz prevalence. Further studies are needed to find out the nature of false-positive POC-CCA trace results and their prevalence to confidently use POC-CCA in disease surveillance especially when approaching elimination. Due to the less focal distribution of worms compared to worm-pairs and a generally higher prevalence of worms in a population, POC-CCA estimates district prevalence of *S. mansoni* with higher accuracy and leads to fewer under-treated districts. Therefore, POC-CCA is well suited for performing a rapid evaluation of treatment needs especially when only urine samples need to be collected.

Appendix



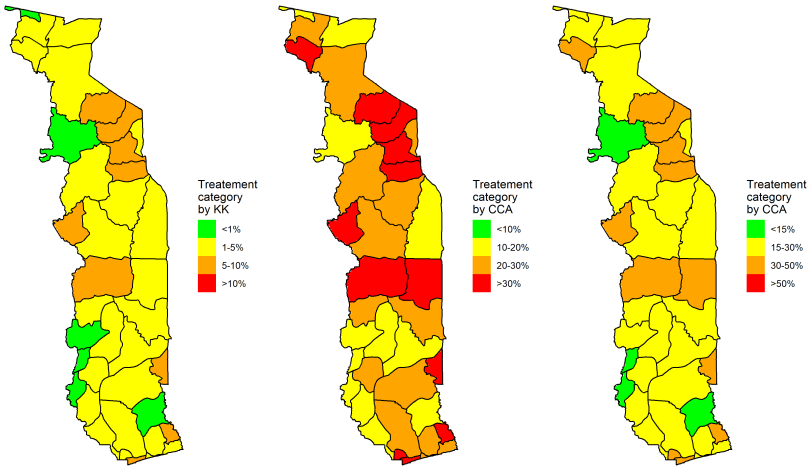
A1: Predictions of *S. mansoni* prevalence obtained from a Bayesian geostatistical model fitted on the 2009 survey data (using Kato-Katz) at a) pixels of a grid overlayed Togo with a resolution of 2 km and b) locations of actual communities (i.e. 3631) in the country.



(a) (b) (c)

(d) (e) (f)

A2: Grid of logarithmic mean number of worms (a), prevalence of worms (b), observed prevalence by POC-CCA (c), logarithmic mean number of worm-pairs (d) prevalence of worm-pairs (e), and observed prevalence by Kato-Katz (f). Colour scale for logarithmic mean counts are given in the lower left and for prevalence in the lower right.



(a) (b) (c)

A3: True classification of a district based on Kato-Katz and two different sets of thresholds for POC-CCA. a) Double slide Kato-Katz classification with 1%, 5%, and 10% as intervention thresholds. b) POC-CCA classification with 10%, 20%, and 30% as intervention thresholds. c) POC-CCA classification with 15%, 30%, and 50% as intervention thresholds.

Chapter 6

Accuracy of different diagnostic techniques for *Schistosoma haematobium* to estimate treatment needs in Zimbabwe: application of a hierarchical Bayesian egg count model

Authors

Nicholas Midzi[1], Oliver Bärenbold[2,3], Portia Manangazira[4], Isaac Phiri[4], Masceline Mutsaka[1,4], Gibson Mhlanga[4], Jürg Utzinger[2,3], Penelope Vounatsou[2,3*]

1 Department of Medical Microbiology, University of Zimbabwe, Harare, Zimbabwe

2 Swiss Tropical and Public Health Institute, Basel, Switzerland

3 Faculty of Science, University of Basel, Basel, Switzerland

4 Ministry of Health and Child Care, Harare, Zimbabwe

To be submitted to *PLoS neglected tropical diseases*.

Abstract

Background

Treatment needs for *Schistosoma haematobium* are commonly evaluated using urine filtration and detection of parasite eggs under a microscope. A widely seen symptom of *S. haematobium* is haematuria, the passing of blood in urine. Hence, the use of haematuria-based diagnostic techniques as a proxy for the assessment of treatment needs has been considered. This study evaluates data from a national survey in Zimbabwe, where three haematuria-based diagnostic techniques, that is microhaematuria, macrohaematuria, and an anamnestic questionnaire pertaining to self-reported blood in urine, have been included in addition to urine filtration in 280 schools across 70 districts.

Methodology

We developed an egg count model, which evaluates the infection intensity-dependent sensitivity and the specificity of each diagnostic technique without relying on a 'gold' standard. Subsequently, we determined prevalence thresholds for each diagnostic technique, equivalent to a 10% urine filtration-based prevalence and compared classification of districts according to treatment strategy based on the different diagnostic methods.

Principal findings

A 10% urine filtration prevalence threshold corresponded to a 17.9% and 13.3% prevalence based on questionnaire and microhaematuria, respectively. Both the questionnaire and the microhaematuria showed a sensitivity and specificity of more than 85% for estimating treatment needs at the above thresholds. For diagnosis at individual level, the questionnaire showed the highest sensitivity at (70.0%) followed by urine filtration (53.8%) and microhaematuria (52.2%).

Conclusions/significance

The high sensitivity and specificity of a simple questionnaire to estimate treatment needs of *S. haematobium* suggests that it can be used as a rapid, low-cost method to estimate district prevalence. Our modelling approach can be expanded to include setting-dependent specificity of the technique and should be assessed in relation to other diagnostic methods due to potential cross-reaction with other diseases.

Acknowledgements

This study received financial support from the European Research Council (ERC-2012-AdG-323180). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

6.1 Introduction

Schistosomiasis is one of the world's neglected tropical diseases primarily affecting school-aged children in rural areas of low- and middle-income countries where there is little or no sanitation and safe water [King, 2010]. Urogenital schistosomiasis, caused by bloodfluke *Schistosoma haematobium* is highly endemic in many countries of Africa and the Middle East [Bichler et al., 2006; Abdulamir et al., 2009; Botelho et al., 2010]. Haematuria, dysuria, cystitis and strictures associated with the bladder and ureter, as well as hydronephrosis and renal failure are sequelae of advanced cases of urogenital schistosomiasis. Carcinoma of the bladder is also a late complication.

At the 54th World Health Assembly in 2001, the World Health Organization (WHO) encouraged member states to regularly provide mass drug administration (MDA) using praziquantel to treat at least 75% of the population at risk of morbidity due to schistosomiasis [World Health Assembly, 2001]. Subsequently, a manual for control programme managers was published specifying MDA once every year if the prevalence of schistosomiasis is $\geq 50\%$ by a standard parasitological approach such as urine filtration or if $\geq 30\%$ by morbidity questionnaires (i.e. anamnestic self-reported haematuria). MDA should be given once every 2 years in moderate risk areas (parasitological prevalence $\geq 10\%$ but $< 50\%$, or prevalence of haematuria $< 30\%$ [WHO, 2006]).

To target MDA effectively, rapid assessment procedures are needed for identifying high-risk communities [Lengeler et al., 2002]. However, there has not been any provision of alternative cost-effective field applicable tools for the rapid screening of urogenital schistosomiasis with superior sensitivity than the urine filtration method. At a time when the global target has been shifted from morbidity control to elimination, rigorous strategies aiming at rapid transitioning from the state of schistosomiasis infectiousness, and transmission interruption to elimination phase are needed [WHO, 2012]. These phases would naturally require specific and sensitive tools with proven diagnostic efficiency to cater for the diagnostic error demanded at each stage [Knopp et al., 2015]. Whilst urine filtration is the current 'gold' standard for the detection of *S. haematobium* with high specificity, this tool is more technical requiring trained microscopist

and costly consumables, including special filter holders and filter papers that are not locally manufactured in sub-Saharan Africa. The use of haematuria to determine treatment strategies at national level has only been assessed at small scale, as one among the tools that can be applied for screening communities with high risk of schistosomiasis for morbidity control [Mott et al., 1985; Lengeler et al., 2002; Morenikeji et al., 2014]. It is not well known whether, indeed, haematuria would correctly identify and categorize the different endemic areas in a country into specific risk categories as parasitological tests do. As part of the Plan of Action for the control of schistosomiasis and soil-transmitted helminthiasis in Zimbabwe a baseline survey that involved mapping of both urogenital and intestinal schistosomiasis was conducted collecting data on history of haematuria as described by Mott et al. (1985) and Lengeler et al. (2002), macrohaematuria, microhaematuria and the urine filtration method for the diagnosis of urogenital schistosomiasis [Mott et al., 1985; Lengeler et al., 2002; Midzi et al., 2014].

The aim of this study was to assess the accuracy of alternative diagnostic methods for *S. haematobium* based on the presence of blood in urine for evaluating treatment needs. We extended our earlier egg count model, developed for estimating the diagnostic error of the Kato-Katz method for *S. mansoni* and estimated the sensitivity of *S. haematobium* diagnostic methods taking into account the dependence of the sensitivity on infection intensity, and the correlation of false positives among the blood-based diagnostic methods [Bärenbold et al., 2017]. Previous analyses were either based on latent class methods, neglecting infection intensity and reducing the urine filtration reading to a binary outcome, or summarising repeated urine filtration results by a composite value which is treated as a 'gold' standard, thus over-estimating sensitivity and under-estimating specificity [Kosinski et al., 2011; Koukounari et al., 2013].

6.2 Materials and methods

6.2.1 Study design

A school-based cross-sectional survey was conducted in Zimbabwe to estimate the geographical distribution of schistosomiasis and soil-transmitted helminthiasis in the country; first in rural provinces (September and October 2010), then in the metropolitan provinces of Harare, Bulawayo

and Chitungwiza town (July and August 2011). The survey was done in collaboration between the Ministry of Health and Child Care (MOHCC) and the Ministry of Education, Sport Arts and Culture (MOESC). There are a total of 73 districts in the country, of which 58 are rural ones, seven are districts in Harare including peri-urban areas, six districts in Bulawayo, and two districts in Chitungwiza. The survey was conducted in 70 (95.8%) districts. Three districts (Gweru, Kwekwe, Bindura) were left out due to limited resources.

Ten teams, each consisting of two laboratory technicians, one technical assistant, one district community nurse, one district education officer, the district environmental health officer and a driver, conducted data collection. Half of the laboratory technicians were drawn from the National Institute of Health Research to lead the teams with the overall responsibilities of organizing and managing field data collection. These technicians were also responsible for performing the urine filtration and the Kato-Katz techniques [Katz et al., 1972]. The other technicians were drawn from the province and were responsible for performing the formol-ether concentration technique as well as for assisting the team leader in executing other duties.

The technical assistants helped in processing specimens and cleaning filters, templates and sieves for re-use. The community nurses were responsible for visual assessment of blood in urine and treatment of study participants. The district environmental officers ensured clean environment in schools and at every stage during sample collection, as well as provided food to children during treatment. The district education officers located the primary schools randomly selected for the national survey, introduced the research team to the school authorities and enrolled a random sample of schoolchildren into the study.

6.2.2 Ethics statement

The proposal to conduct the national schistosomiasis and soil-transmitted helminthiasis survey was approved by the national ethical review board and the Medical Research Council of Zimbabwe. The ethical approval registration number for the study is MRCZ/A/1207. The Secretary for Education, Sport, Arts, and Culture also approved the study. Written informed consent was sought from the parents/guardians of study participants. UNICEF delivered parental/guardian informed consent forms addressed to each school by the Secretary for Education, Sport, Arts, and

Culture throughout the country in advance to allow school heads sufficient time to liaise with parents/guardians for their consent. On the day of sample collection, only the assenting children whose consent forms were signed by their parents/guardians participated. Enrolment into the study was voluntary; hence, participants were free to withdraw from the study at any time.

6.2.3 Study population

Schoolchildren aged 10–15 years were targeted for the study, as they constitute the high-risk age group for schistosomiasis and soil-transmitted helminth infections in the community [Nagelkerke et al., 2000; WHO Expert Committee, 2002; WHO, 2006].

6.2.4 Sample size calculation and selection of participants

The sample size at national level was based on the total enrolment of primary school children, i.e., n = 2,490,568 (MOESC 2005). A sample of 15,818 children was calculated using Epi Info (Epi Info version 6, Centers for Disease Control and Prevention, Atlanta, GA, USA) assuming an average national prevalence of schistosomiasis equal to 37% and an error margin of 0.75% [Ndhlovu et al., 1992]. The number of children per district was considered proportional to these enrolled nationally. Simple random sampling was used to select schools per district following the lottery method [Nagelkerke et al., 2000]. At each school, 50 children equally distributed by gender were randomly selected using the lottery approach [WHO Expert Committee, 2002; Montresor et al., 2002]. While schoolchildren aged 10–15 years constituted the desired sampling frame, children aged 6–9 years (n = 598) and some aged above 15 years (n = 6) were included in some schools where the number of children aged 10–15 years was less than 50 [Midzi et al., 2014].

6.2.5 Diagnosis and treatment

A single mid-day urine sample was collected in 100 ml screw cap plastic specimen bottles from each participant [Doehering et al., 1983; Engels et al., 1996]. For the quantitative diagnosis of S. haematobium infection, the standard urine filtration method was applied [WHO, 1991]. The

technique involves filtration of 10 ml of a thoroughly mixed urine specimen, through a Nytrile filter (12–14 mm diameter).

In addition, each individual was screened using the following indirect tests: (i) a questionnaire regarding recent history of haematuria; (ii) inspection of the urine specimen for visible blood; and (iii) use of reagent strips to detect haematuria. To determine the presence and severity of microhaematuria, all urine specimens were tested for presence of detectable blood using urine reagent strips (Bayer Hemastix; Leverkusen, Germany). The results were recorded semi-quantitatively: negative, trace hemolyzed, weakly positive (+), moderately positive (++), and highly positive (+++). Using this test, an individual was diagnosed negative if there was no positive reaction of the reagent strip from trace to highly positive and positive if trace, weakly, moderately, and highly positive was detected. The recent history of blood in urine was assessed asking the following question: "'Have you seen blood in your urine in the past month?"' [Mott et al., 1985; Lengeler et al., 2002]. The individual was regarded as positive if the answer was yes, and negative if the answer was no to the question. For macrohaematuria, the community nurse assessed each urine specimen for visible blood in urine as they normally do at health facility. If there was no blood detected in urine, the individual was diagnosed as negative and otherwise as positive.

Following submission of stool and urine samples, all participants received 100 ml of orange juice and a piece of bread to eat after which they simultaneously received a single dose of praziquantel (40 mg/kg) and albendazole (400 mg) in tablet form regardless of their infection status since both drugs are considered safe [WHO, 2006].

6.2.6 Statistical Methods

Data collected from the field were coded into binary variables representing positive or negative for the reagent strip, recent history of blood in urine and macrohaematuria. The urine filtration technique records the number of eggs per 10 ml of urine. Mapping of the disease risk and comparisons between diagnostic methods were done at the district level with data from the 70 districts described previously.

We developed a Bayesian hierarchical model to estimate the diagnostic sensitivity and

specificity of the four tests. In particular, we extended our earlier work on *S. mansoni* and took into account an infection intensity-dependent sensitivity and specificity of macrohaematuria (M), microheamaturia (m), and the questionnaire (Q) [Bärenbold et al., 2017, 2018]. The results of urine filtration and the alternative diagnostic methods were assumed to be only dependent on infection intensity when infection was present, while the alternative, blood detecting methods were considered to be correlated via a latent variable describing blood in urine for uninfected individuals, that is

$$\begin{aligned} P(Y_i^F, Y_i^m, Y_i^M, Y_i^Q) = P(Y_i^F \mid D_i = 1)P(Y_i^m \mid D_i = 1)P(Y_i^M \mid D_i = 1)P(Y_i^Q \mid D_i = 1)\pi_d \\ + P(Y_i^F \mid D_i = 0)P(Y_i^m, Y_i^M, Y_i^Q \mid D_i = 0)(1 - \pi_d) \end{aligned} \tag{6.1}$$

where Y_i^F is the urine filtration egg count for individual i, Y_i^m is the binary result of the microhaematuria test including traces, Y_i^M is the binary result of the macrohaematuria test and Y_i^Q indicates the binary result of the questionnaire. D_i is the disease status, taking 1 for infected and 0 for uninfected individuals and π_d is the true infection prevalence of district $d = 1, 2, \ldots, 70$. Urine filtration measurements were modelled by a negative binomial count distribution, that is $P(Y_i^F \mid D_i = 1) \equiv NB(\lambda_i, k)$ and infection intensities, λ_i were considered to be gamma distributed in the population of infected individuals, $\lambda_i \sim Gamma(\mu_d\alpha, \alpha)$, with μ_d describing the mean infection intensity among infected individual in district d and α specifying the aggregation of infection intensities in the population.

Microhaematuria, macrohaematuria and the questionnaire data were modelled by a Bernoulli distribution with the sensitivity depending on the infection intensity of an infected individual, that is,

$$P(Y_i^z \mid D_i) = \begin{cases} Be(s_i^z) & \text{, if } D_i = 1 \\ Be(1 - c_i^z) & \text{, if } D_i = 0 \end{cases} \tag{6.2}$$

The specificity parameter c_i^z for the z diagnostic method was defined as $c_i^z = c^z b_i$ where b_i reflects the correlation between false-positives from blood detecting methods. The mean of the c_i^z was calculated separately for girls aged 12 and older from the rest of the subjects to investigate

influence of menstruation on the probability of false positives. The egg intensity-dependent sensitivity s_i^z of z diagnostic method was described by $\text{logit}(s_i^z a_3^z) = a_0^z + a_1^z \lambda_i^{\log(a_2^z+1)}$ where the infection intensity λ_i corresponds to the number of eggs in 10 ml of urine for individual i. a_0^z determines the sensitivity in the limit of very low infection intensity; a_1^z quantifies the dependence of sensitivity on infection intensity; a_2^z determines the shape of the curve; and a_3^z the sensitivity in the limit of very high infections. We assumed that the sensitivity of the alternative diagnostic methods for an individual depends on infection intensity only and not on other individual variables like age and sex.

To complete Bayesian model formulation, we chose a uniform prior $\pi_d \sim U(0,1)$, a gamma distribution with mean 40 and standard deviation (SD) 40 for the mean infection intensity μ_d, a normal distribution with mean 0.1 and SD 0.1 for the population variation α_d, a normal distribution with mean -1 and SD 1.5 for a_0, a gamma distribution with shape and scale parameters 5 and 30 respectively, for a_1, a normal distribution with mean 0.5 and SD 0.5 for a_2, and a beta distribution with parameters 10 and 1 for a_3 to ensure a non-informative distribution of sensitivity curves in the relevant range of infection intensities. Markov chain Monte Carlo (MCMC) simulations were run for 1,000 iterations with 50 chains in Stan version 2.16.2 (Stan Development Team; mc-stan.org) [Carpenter et al., 2016]. Convergence was assessed using the Gelman+Rubin diagnostics [Gelman and Rubin, 1992].

6.3 Results

6.3.1 Demographics

A total of 13,195 primary school children drawn from 280 schools were included in the study. Of these, 12,656 (95.9%) had all diagnostic methods performed and results recorded. The mean age was 11.2 years and 95.6% of the children were between the age of 10 and 15 years.

Country-level prevalence and diagnostic accuracy

Table 6.1 shows observed results of the diagnostic tests in the study population. The prevalence measured by single urine filtration was 17.6%, by microhaematuria 20.8%, by macrohaematuria

4.3% and by questionnaire 28.5%, while the cumulative prevalence was 37.6%. The arithmetic mean egg count among those positive by urine filtration was 79.6 eggs/10 ml of urine, and in the full population was 13.5 eggs/10 ml of urine.

Table 6.2 shows model-based estimates of the diagnostic errors and the 'true' prevalence, and mean egg count of all individuals and of the positively tested ones. The overall 'true' prevalence was estimated to be 35.3% which leads to an overall sample sensitivity of 53.8% for urine filtration, 52.2% for microhaematuria, 11.5% for macrohaematuria and 70.0% for the questionnaire. The specificity was estimated to be 95.6% for microhaematuria, 99.3% for macrohaematuria, and 93.7% for the questionnaire. Specificity of the three diagnostic methods did not differ in the 12- to 14- years-old female population. The mean infection intensity of an infected individual was estimated to 45.4 eggs/ 10ml of urine and the mean infection intensity of all individuals to 15.4 eggs/10 ml of urine.

6.3.2 District-level prevalence

The urban districts of Bulawayo have much lower prevalence than Harare and Chitungwiza but there is considerable variability within a city, e.g. from 10.4% to 25.3% in Harare, and from 2.8% to 14.8% in Bulawayo (Table A1, appendix). The estimated 'true' prevalence and the observed

Table 6.1: Summary of prevalence and infection intensity of observed data

	Estimate	**95%CI**[1]
Urine filtration prevalence (%)	17.6	16.9 – 18.2
Microhaematuria prevalence (%)	20.8	20.1 – 21.5
Macrohaematuria prevalence (%)	4.3	3.9 – 4.6
Questionnaire prevalence (%)	28.5	27.7 – 29.3
Cumulative prevalence (%)	37.6	36.7– 38.4
Mean infection intensity among those positive by filtration (eggs/10 ml of urine)	76.6	69.3 – 84.4
Population mean infection intensity (eggs/10 ml of urine)	13.5	12.2 – 15.0

[1] Confidence Intervals (CI) were determined by bootstrap

one by district for the whole country are shown in Fig 6.1. All maps depict low prevalence in the western districts while there is some disagreement between haematuria-based techniques showing higher prevalence in the northern districts compared to urine filtration.

6.3.3 Infection-intensity dependent sensitivity

All four diagnostic methods show a clear dependence of sensitivity on infection intensity, as determined by the urine filtration (Fig 6.2). Urine filtration, microhaematuria and the questionnaire reached considerable sensitivity ($> 60\%$) at an intensity of 25 eggs per 10 ml of urine. While the sensitivity of the questionnaire did not surpass 90%, microhaematuria detected almost all cases with an intensity of more than 50 eggs per 10 ml of urine. The sensitivity of macrohaematuria stayed very low up to moderate infection intensities with only 25% at 100 eggs and less than 10% at 50 eggs per 10 ml of urine. The questionnaire shows the weakest dependence on infection intensity with sensitivity above 50% even at very low infection intensities.

6.3.4 Evaluation of treatment needs

The WHO defines treatment needs according to prevalence thresholds at 10% and 50% based on parasitological methods. Table 6.3 presents estimates of the corresponding thresholds of haematuria-based diagnostics. A 10% prevalence observed by urine filtration corresponds to a

Table 6.2: Model-based estimates of the prevalence, diagnostic errors and mean infection intensity (posterior mean and 95% Bayesian credible intervals, BCI).

	Estimate	**95% BCI**
'True' prevalence (π) (%)	35.3	33.7– 36.9
Urine filter sensitivity (%)	53.8	51.3– 56.2
Microhaematuria sensitivity (%)	52.2	49.7 – 54.7
Macrohaematuria sensitivity (%)	11.5	10.4 – 12.6
Questionnaire sensitivity (%)	70.0	67.7 – 72.4
Microhaematuria specificity (%)	95.6	95.0 – 96.2
Macrohaematuria specificity (%)	99.3	99.1– 99.5
Questionnaire specificity (%)	93.7	92.6 – 94.7
'True' infected mean (μ) (%)	45.4	42.2 – 48.9
'True' full mean ($\mu\pi$) (%)	15.4	14.3 – 16.4

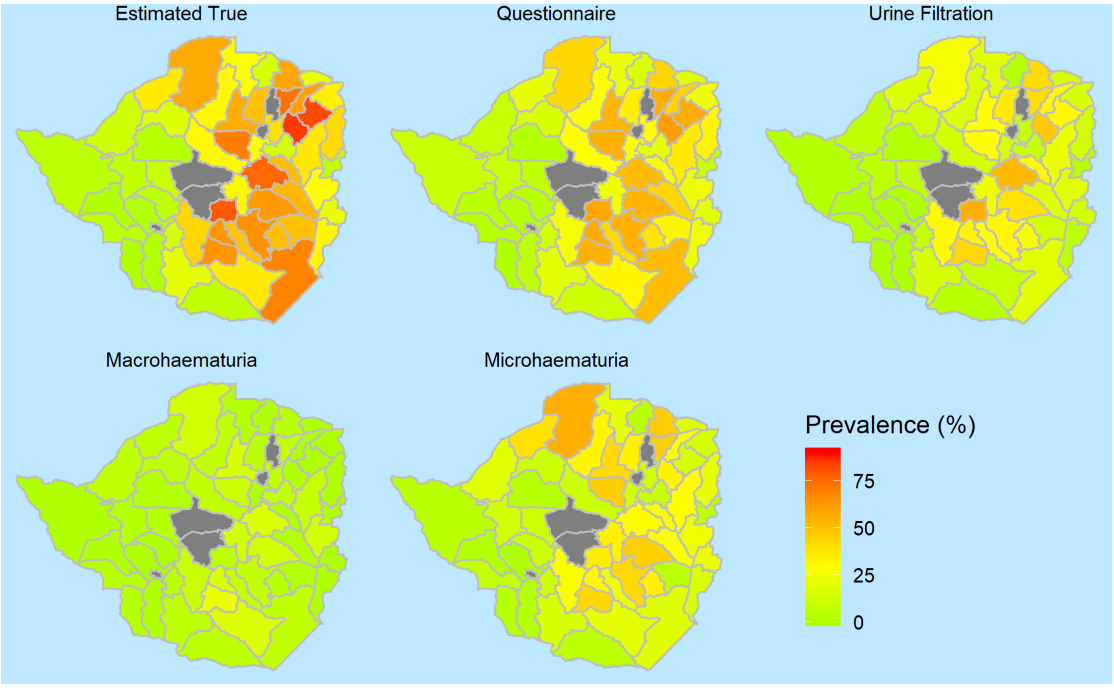


Figure 6.1: Estimated 'true' and observed prevalence by district and diagnostic method. Grey areas indicate the urban districts of Harare, Bulawayo and the three districts where no data was collected (Zimbabwe, 2010/2011).

'true' prevalence of 18.5% and an observed prevalence of 17.9%, 13.3%, and 2.7% based on the questionnaire, microhaematuria, and macrohaematuria methods, respectively.

Only two districts had urine filtration-based prevalence of *S. haematobium* above 50%. Therefore, we show the district classification results for the treatment category of 10-49%. There were 40 districts with *S. haematobium* prevalence above 10% and 30 districts with prevalence below this threshold giving a decent sample size for both groups. Table 6.4 compares the ability of each diagnostic method to classify a district in the correct treatment category. All three methods showed few false positives, i.e. 4, 4, and 7 for the questionnaire, microhaematuria and

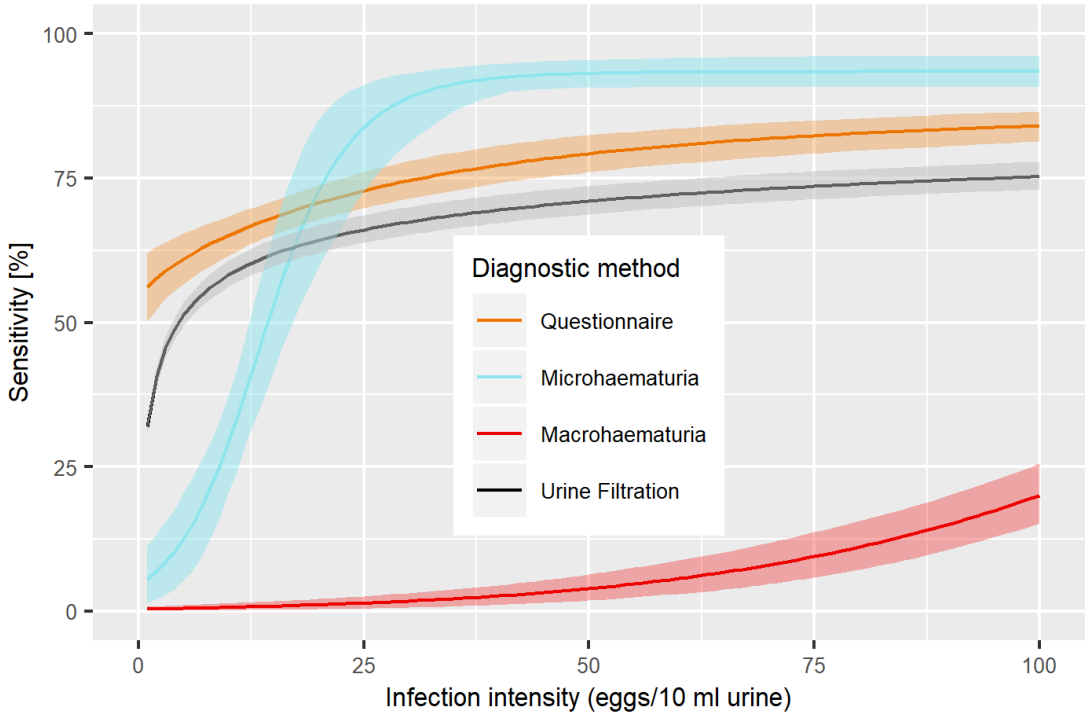


Figure 6.2: Infection-intensity dependent sensitivity estimates for urine filtration, microhaematuria, macrohaematuria and a questionnaire for diagnosis of *S. haematobium* (Zimbabwe, 2010/2011).

macrohaematuria, respectively. The questionnaire and microhaematuria classified 35 and 34, or 87.5% and 85.0%, respectively, of the 40 districts in the correct treatment category while macrohaematuria does so only for 62.5% of the districts (Table A2, appendix).

Sensitivity and specificity can be represented in a receiver operating characteristics (ROC) curve when the threshold is varied from 0% to 100% (see Fig A3, appendix). Both micro-

Table 6.3: Prevalence thresholds of haematuria-based diagnostic methods corresponding to urine filtration ones

	Threshold 1 (%)	Threshold 2 (%)
Urine filtration	10	50
True prevalence	18.5	92.9
Questionnaire	17.9	62.7
Microhaematuria	13.3	48.8
Macrohaematuria	2.7	10.7

haematuria and the questionnaire show very similar curve with sensitivity of 95% reached at specificities between 70% and 80%, and specificity of 95% at sensitivities between 55% and 65%. Macrohaematuria shows lower accuracy with 95% specificity at 35% sensitivity and 70% sensitivity at 60% specificity. The classification of districts to a treatment category (Fig 6.3) indicates that there was no spatial pattern in the districts that were wrongly excluded from treatment for microhaematuria and the questionnaire while macrohaematuria missed districts that needed treatment mostly in the north-western part of Zimbabwe.

Fig 6.4 depicts the district prevalence measured by each diagnostic method, ordered by urine filtration. False positives by questionnaire were all in districts with urine filtration above 5%. The same is true for microhaematuria with one outlier recording 40% observed prevalence in a district with $< 5\%$ positives according to urine filtration. Macrohaematuria determined treatment needs in districts with $< 5\%$ urine filtration and simultaneously missed districts approaching 50%. The districts that were falsely identified as negative by microhaematuria and the questionnaire were

Table 6.4: Ability of each diagnostic method of detecting districts with treatment needs according to WHO (10-49% parasitological) using the adjusted thresholds.

	Urine filtration$> 10\%$, n=40	Urine filtration$< 10\%$, n=30
Questionnaire $> 17.9\%$	35	4
Questionnaire $< 17.9\%$	5	26
Microhaematuria $> 13.3\%$	34	4
Microhaematuria $< 13.3\%$	6	26
Macrohaematuria $> 2.7\%$	25	7
Macrohaematuria $< 2.7\%$	15	23
Questionnaire $> 10\%$	38	11
Questionnaire $< 10\%$	2	19
Microhaematuria $> 10\%$	38	9
Microhaematuria $< 10\%$	2	21
Macrohaematuria $> 10\%$	8	1
Macrohaematuria $< 10\%$	32	29

[1] The sensitivity is calculated by dividing the number of districts correctly identified in the $> 10\%$ urine filtration group by the total number; e.g. 35/40 for the questionnaire at threshold 17.9%.

[2] The specificity is calculated by dividing the number of districts correctly identified in the $< 10\%$ urine filtration group by the total number; e.g. 26/30 for the questionnaire at threshold 17.9%.

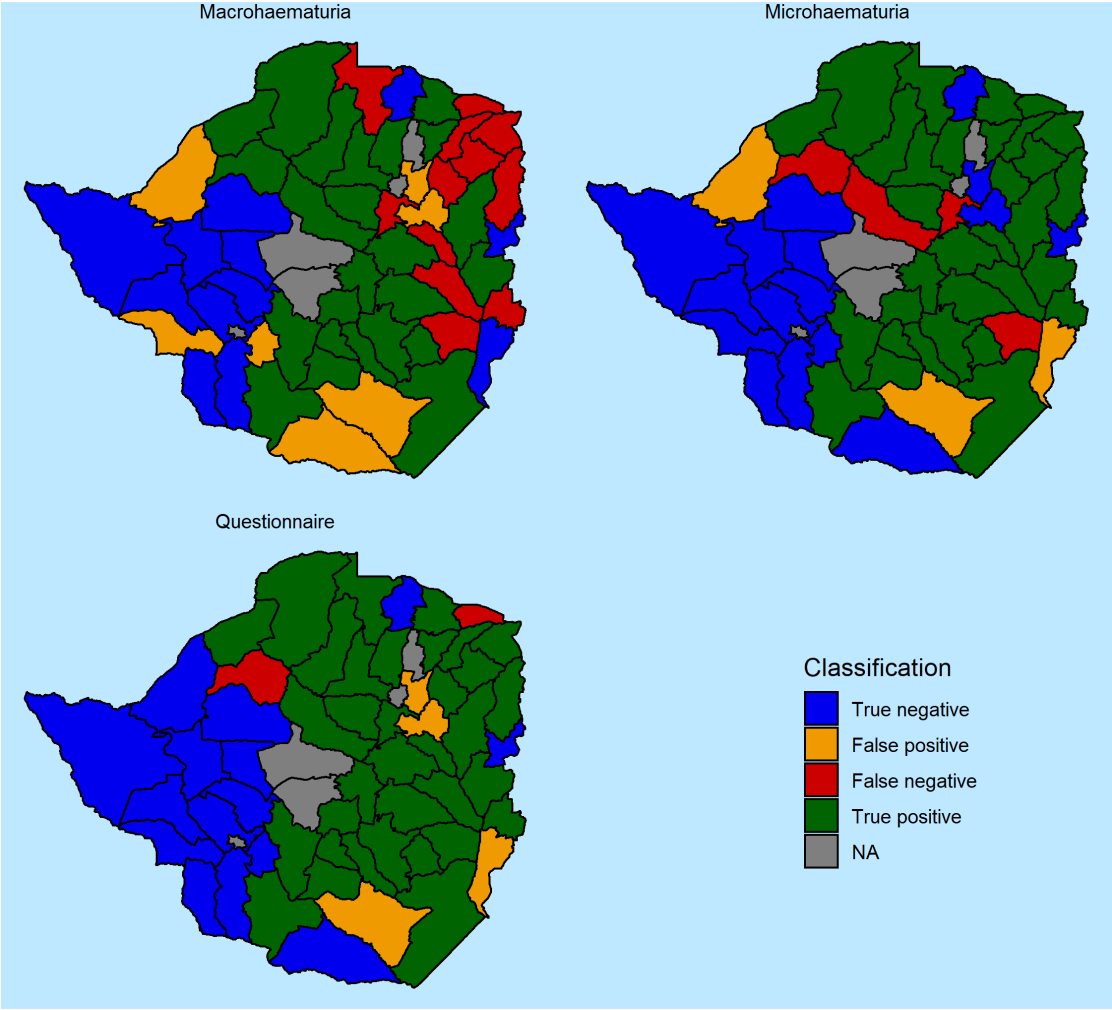


Figure 6.3: Classification of districts according to haematuria diagnostic. True negative indicates classification below threshold for both urine filtration and the given alternative; false positive indicates exceeding threshold in alternative but not in urine filtration; false negative indicates urine filtration above threshold but alternative below; and true positive indicates both urine filtration and alternative above threshold.

always at urine filtration prevalence below 15%.

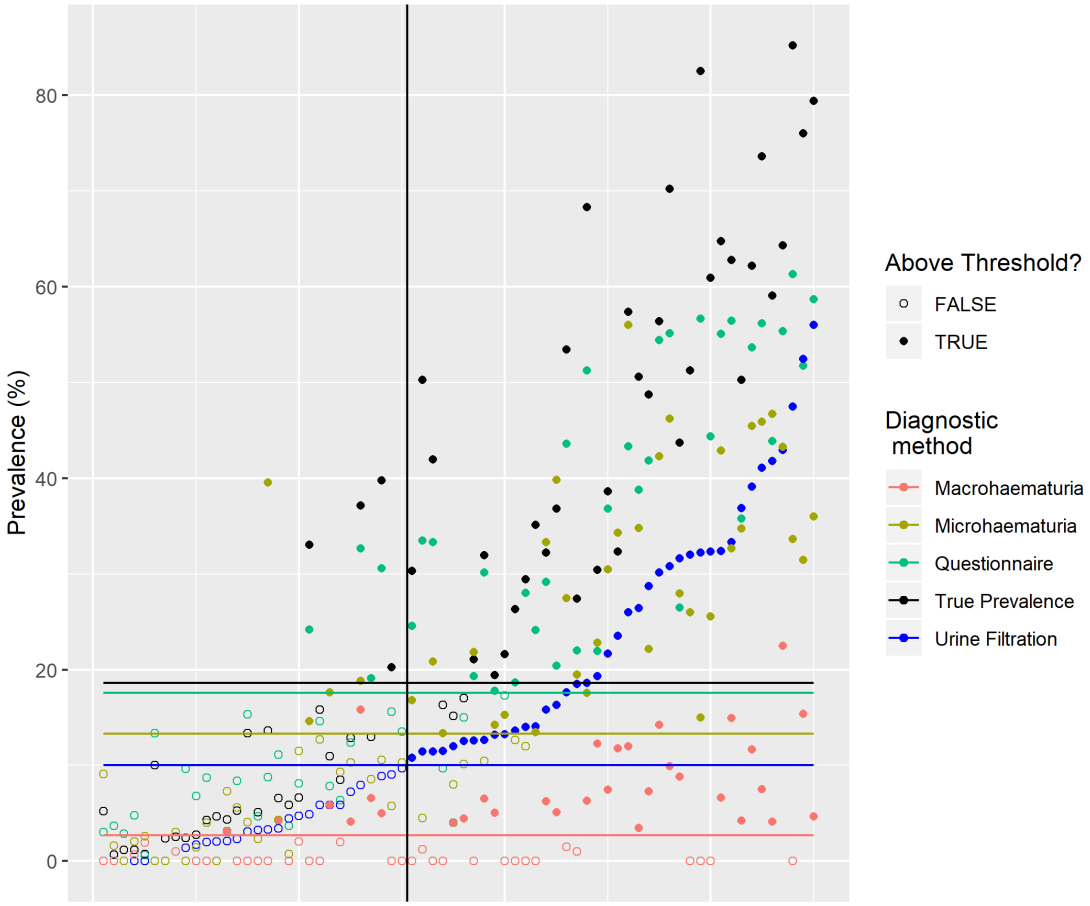


Figure 6.4: District prevalence for all 70 districts and the four diagnostic methods including the estimated 'true' prevalence ordered by urine filtration prevalence on the x-axis. The vertical line indicates the 10% urine filtration prevalence border. Filled dots to the left of this line are false positive, while empty dots to the right are false negatives. The horizontal lines show the determined threshold for each diagnostic method.

6.4 Discussion

This is the first study to estimate the diagnostic error of *S. haematobium* diagnostics, i.e. urine filtration, a questionnaire regarding recent history of blood in urine, a microhaematuria reagent strip, and visual inspection of urine to detect macrohaematuria, taking into account the dependence of sensitivity on the egg intensity. We further evaluated the ability of the three haematuria-based diagnostic techniques to determine treatment needs. Analysis was done using an adaptation of the egg count model previously used to determine the dependence of the Kato-Katz diagnostic

technique for *S. mansoni* on the infection intensity and to translate intervention thresholds from Kato-Katz to the point-of-care circulating cathodic antigen (POC-CCA) methods [Bärenbold et al., 2017, 2018]. The model was fitted on *S. haematobium* data from the Zimbabwe national survey of 2010.

Generally, the strong dependence on infection-intensity of all diagnostic methods indicates that an assumption of constant sensitivity is not valid. An advantage of urine filtration is that its sensitivity increases by repeated sampling which is not possible for the alternative diagnostics. However, this improvement cannot be measured due to lack of repeats allowing the estimation of within person variability.

The questionnaire showed the highest overall sensitivity (70.0%) but a rather low specificity of 93.7% for individual diagnosis. Low specificity is less of a problem when adverse events of possible over-treatment are small. Macrohaematuria shows very high specificity ($> 99\%$), but a low sensitivity of 11.5% on the individual level. Microhaematuria, with a sensitivity of 52.2% and specificity of 95.6% does not improve the diagnostic error of urine filtration. Furthermore, the effort required to perform the diagnostics is comparable.

Mafe et al. (1997) evaluated macrohaematuria, microhaematuria, and a different form of the questionnaire asking for the current passing of blood in urine in a study of 1056 individuals in the Kainji Lake area of Nigeria [Mafe, 1997]. Urine filtration was used as a reference standard, leading to overestimation of the sensitivity in a highly endemic area with no previous treatment. Hence, macrohaematuria and microhaematuria have both higher sensitivity at 69% and 38%, respectively, compared to our estimates. Sensitivity of the questionnaire was low (44%) which is not surprising given that the question was about the current blood passing instead of the recent history like in our study. Fatiregun et al. (2005) compared different variants of questionnaires and a reagent strip to detect microhaematuria with urine filtration as reference standard [Fatiregun et al., 2005]. 'Unqualified haematuria' the most similar version of the questionnaire to this study showed a specificity of 93.1% similar to our estimate, while the sensitivity was much lower at 41.7%. Lower sensitivity can be explained by lower overall prevalence and infection-intensity in the study setting. Koukounari et al. (2009) used latent class modelling to compensate for the lack of reference test, therefore comparison of the results is difficult [Koukounari et al., 2009].

Krauth et al. show that in low prevalence settings microhaematuria is an unstable proxy for *S. haematobium* infections [Krauth et al., 2015]. A number of other conditions like for example sickle cell disease that have important spatial variations can cause Haematuria. We estimated specificity of the tests in girls in the age between 12 years and 14 years separately but no difference was found. Our analysis could not control for additional conditions that cause haematuria because this data was not collected. We assumed that the specificity of the diagnostic tests is constant which may not necessarily be true when other conditions, possibly spatially varying, can cause false-positives. The absence of a well-performing diagnostic test for *S. haematobium* motivates the development of alternative diagnostics based on either specific antigens, or PCR but none of them is currently accurate and cost-effective enough to use in large scale surveys [Ibironke et al., 2012; Shiff, 2015; Knopp et al., 2015; Danso-Appiah et al., 2016; de Dood et al., 2018].

Our results do not mean that haematuria-based diagnostic techniques cannot be used to evaluate treatment needs. Lengeler et al. showed that questionnaires perform well in identifying high risk *S. haematobium* schools first in a study in Kongo and then in a review of studies from 10 countries involving a total of 133,880 children [Lengeler et al., 2000, 2002]. They showed a striking relation between the prevalence observed by parasitological methods and the number of positive responses to the questionnaire. Sensitivity at the community level was in the order of 90% for most countries while specificity was mostly above 80% with the exception of Cameroon, Malawi and Zimbabwe.

In the presented national survey of Zimbabwe using an intervention prevalence threshold of 17.9% for the questionnaire (equivalent to 10% urine filtration) almost 90% of districts with treatment needs were correctly identified while the specificity was similarly high above 85%. Misclassified districts showed urine filtration prevalence between 5% and 15% representing a range of prevalence where differences in treatment decisions can be justified. Thus, the data supports that a simple questionnaire can be used as a low effort diagnostic alternative to collecting urine for filtration for evaluating treatment needs. The suggested threshold of 17.9% has to be estimated prospectively in different settings at different stages of control. Microhaematuria performs equally well to the questionnaire for evaluation of treatment needs but requires the collection of urine samples. Therefore, the advantage of the method compared to urine filtration

is reduced. Macrohaematuria shows a 62.5% sensitivity for district-level treatment needs and a lower specificity of 76.7%. Misclassified districts by macrohaematuria cover almost the whole range of urine filtration prevalence from $< 5\%$ to almost 50%.

6.4.1 Conclusion

We showed that the questionnaire has high diagnostic accuracy in evaluating treatment needs for *S. haematobium* at a level of 10% urine filtration when a threshold of 17.9% is used. The threshold suggested here has to be evaluated and confirmed in different settings. The corresponding specificity was high in Zimbabwe but this is not necessarily the case for other locations due to varying prevalence of haematuria causing conditions. Still, the low effort of performing questionnaire surveys compared to urine filtration can make the questionnaire a promising tool for low-cost rapid assessment. For optimal individual diagnosis, urine filtration is still the best method among the ones tested in this study due to its high specificity, the possibility of increasing its sensitivity through repeated testing, and the additional information that is provided regarding the infection intensity.

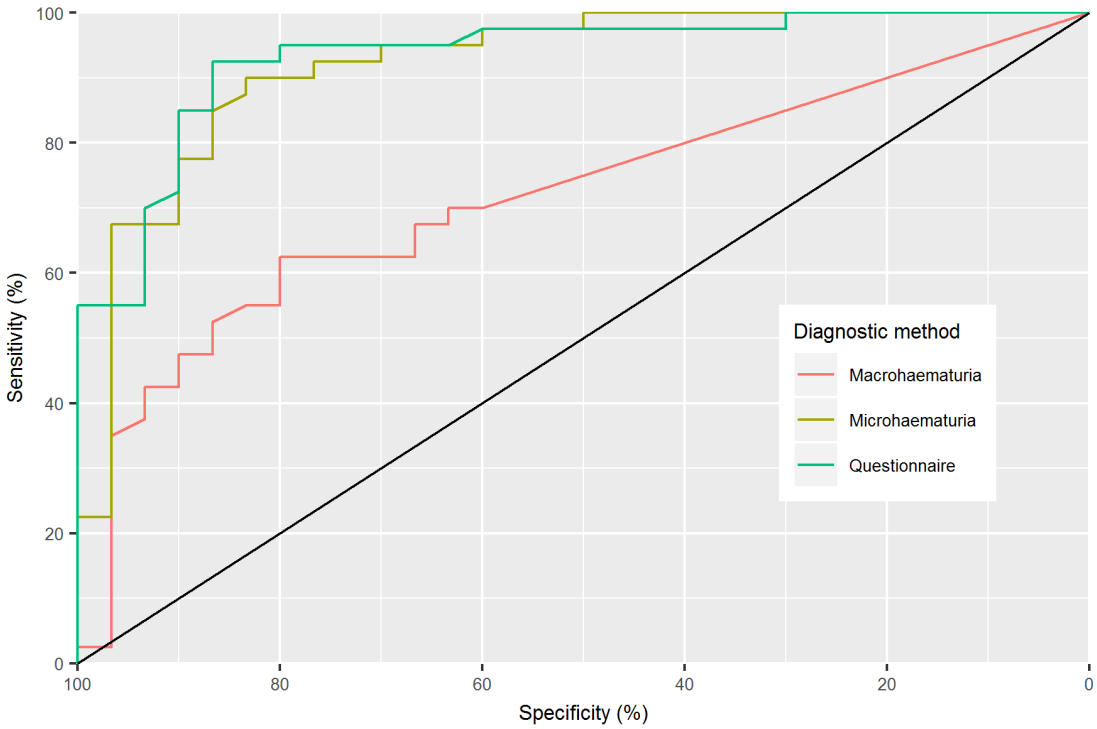
Appendix

A1: Observed and 'true' estimated prevalence of urban districts in Harare, Bulawayo, and Chitungwiza

(%)	True prevalence	Questionnaire	Urine filtration	Macrohaematuria	Microhaematuria
Harare					
Peri-urban	25.3 (15.8 - 36.3)	24.6	10.8	0.0	16.8
Glen View / Mufakose	26.3 (18.9 - 36.2)	18.7	13.6	0.0	12.6
High Fields / Glen Norah	13.6 (8.8 - 19.8)	13.5	9.7	0.0	10.3
Maborein / W-Park	16.3 (10.6 - 22.7)	9.7	11.5	0.0	13.4
Mabvuku / Tafara	13.2 (1.7 - 26.0)	8.8	3.3	0.0	39.6
Mbare / Hatfield	15.9 (7.4 - 21.7)	14.6	5.9	0.0	12.7
North Central	10.4 (0.3 - 35.4)	13.0	0.0	0.0	0.0
Bulawayo					
Peri-urban	2.8 (1.1 - 6.1)	6.8	1.7	0.0	1.4
Reigate	5.1 (2.5 - 12.1)	4.7	3.3	0.0	2.3
Imbizo	2.2 (0.0 - 8.6)	0.0	0.0	0.0	0.0
Khami	5.9 (2.4 - 11.5)	3.7	4.4	0.0	0.7
Mzilikazi	4.4 (0.6 - 12.1)	0.0	2.0	0.0	0.0
Sizinda	14.8 (4.8 - 25.8)	4.0	12.0	4.0	8.0
Chitungwiza					
Seke-Chitungwiza	20.4 (10.8 - 31.9)	15.6	9.0	0.0	5.7
Zengeza	2.6 (0.0 - 9.4)	9.6	1.4	0.0	0.0

A2: Sensitivity and specificity of determining the treatment need

	District sensitivity	District specificity
Questionnaire 17.9%	87.5%	86.7%
Microhaematuria 12.4%	85.0%	86.7%
Macrohaematuria 2.6%	62.5%	76.7%
Questionnaire 10%	95.0%	63.3%
Microhaematuria 10%	95.0%	70.0%
Macrohaematuria 10%	20.0%	96.7%



A3: Diagnostic error of district classification in treatment categories (ROC-curve) based on prevalence thresholds of the questionnaire, microhaematuria and macrohaematuria corresponding to a 10% urine filtration prevalence.

Chapter 7

Infection intensity dependent accuracy of a reagent strip to diagnose *S. haematobium* infection and equivalent treatment prevalence thresholds

Authors

Oliver Bärenbold[1,2], Christoph F. Hatz[1,2], Brigitte J. Vennervald[3], Jürg Utzinger[1,2], Penelope Vounatsou[1,2]

1 Swiss Tropical and Public Health Institute, Basel, Switzerland

2 University of Basel, Basel, Switzerland

3 Danish Bilharziasis Laboratory, Copenhagen, Denmark

To be submitted to *PLoS neglected tropical diseases*.

Abstract

Background

Reagent strips to detect microhaematuria as a proxy for *S. haematobium* infections have long been considered an alternative to urine filtration for both individual diagnosis and for estimation of treatment needs. However, the diagnostic accuracy of reagent strips is not known especially at low infection intensity levels. Furthermore, community treatment guidelines estimates based on reagent strip (Boehringer Mannheim) data are not available. We evaluate data from a study conducted in Tanzania with parallel diagnosis by urine filtration and a semi-quantitative reagent strip in two villages including a baseline and six follow-up surveys representing a wide range of transmission settings.

Methodology

We develop a Bayesian model linking individual egg count data based on urine filtration to reagent strip binary test results available on multiple days. The model is used to estimate the relation between infection intensity and sensitivity of the reagent strip for observations on one to five days and when trace reading is considered either positive or negative. Furthermore, we simulated data from 3,000 populations with varying mean infection intensity to infer on the relation between prevalence observed by urine filtration and the interpretation of the reagent strip readings.

Principal findings

The reagent strip including has a very high sensitivity even for single measurement reaching 100% at around 7 eggs/10 ml when traces are included in the positives. The specificity is still high equal to 96%. When traces are excluded, the diagnostic accuracy of the reagent strip in this study is equivalent to the urine filtration on data from single day. A 10% and 50% urine filtration prevalence based on a single day sampling design corresponds to 11.9% and 49.4% prevalence by reagent strip, respectively when traces are treated as positives and 19.3% and 61.1%, respectively when traces are excluded from the positives.

Conclusions/significance

Traces can be included in the positives when high sensitivity is required, and can be excluded when almost perfect specificity is needed. The observed prevalence of the reagent strip used in this study when traces are treated as negative is a very good proxy for prevalence observed by urine filtration on a single day. There is need to evaluate the diagnostic error of reagent strips from alternative producers applied in different settings with varying levels of endemicity.

Acknowledgements

This study received financial support from the European Research Council (ERC-2012-AdG-323180) and the Schistosomiasis Consortium for Operational Research and Evaluation (SCORE). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

7.1 Introduction

Schistosomiasis haematobium is a parasitic infection caused by the most common species of flatworms called schistosomes. The disease is also known as urinary schistosomiasis because the parasitic worms live in veins next to the bladder and eggs are excreted through urine creating symptoms distinct from intestinal schistosomiasis [Colley et al., 2014; McManus et al., 2018]. There are currently about 190 million infected individuals for all species of schistosomiasis combined causing an estimated 2.1 million disability-adjusted life years (DALYs) lost [GBD 2016 DALYs and HALE Collaborators, 2017; GBD 2016 Disease and Injury Incidence and Prevalence Collaborators, 2017]. While direct mortality is low there is a considerable burden in the form of anemia, growth stunting, impaired cognition, increased risk for cancer of the bladder and HIV infections [Abdulamir et al., 2009; Kjetland et al., 2014].

There has been a surge in investment in the control of schistosomiasis as part of the Millennium Development Goals (MDGs) after 2000 and the Sustainable Development Goals (SDGs) from 2015 onwards [Molyneux, 2004; Molyneux et al., 2017]. Reduction of disease burden is primarily achieved using mass drug administration (MDA) while treatment needs are evaluated using parasitological methods [Montresor et al., 1998; World Health Assembly, 2001]. The World Health Organization (WHO) has formulated a roadmap for elimination of schistosomiasis as a public health problem (i.e. a prevalence of heavy infections under a threshold of 1%) and even elimination in suitable settings by 2025 [WHO, 2013].

The WHO recommended diagnostic method for *S. haematobium* is urine filtration when the aim is decision making regarding interventions. Urine is collected between 10 am and 2 pm and after vigorous shaking 10 ml of urine it is filtered and examined using light microscopy [Montresor et al., 1998; Knopp et al., 2018]. Alternative diagnostic techniques exist which are often used in parallel on the same individual to increase diagnostic accuracy. Blood in urine is a very common symptom of urinary schistosomiasis although it is not fully specific to the disease [Krauth et al., 2015]. As blood in urine is relatively easier to detect than eggs, three blood-based diagnostic tests are available to test for *S. haematobium*. A questionnaire regarding recent history of visible blood in urine, inspection of a urine sample for visible blood, and a

reagent strip for the detection of microhaematuria [Mafe, 1997; Lengeler et al., 2000]. Reagent strips for microhaematuria generally give a semi-quantitative assessment of infection-intensity with four positive and one negative levels. Other diagnostic approaches are the detection of a specific antigen in urine and polymerase chain reaction (PCR) based methods to detect genetic material in urine [Knopp et al., 2015; Shiff, 2015].

In this study we evaluate the infection intensity-dependent diagnostic accuracy of a reagent strip to detect microhaematuria and urine filtration incorporating data from measurements on multiple days. Previous studies have evaluated the sensitivity and specificity of the above mentioned methods but there is no study that models directly the egg counts directly and takes into account the day-to-day variation of infection intensity [Mafe, 1997; Obeng et al., 2008; Midzi et al., 2009; Kosinski et al., 2011; Stete et al., 2012; Knopp et al., 2018]. We extend our egg count model for individual level data developed for the analysis of Kato-Katz and POC-CCA diagnostics for *S. mansoni* and predict 'true' prevalence and the infection intensity-dependent sensitivity for urine filtration and reagent strip diagnostic tests as well as the diagnostic specificity of the latter [Bärenbold et al., 2017, 2018]. Furthermore, we use a simulation study to translate urine filtration intervention thresholds to microhaematuria analogues.

7.2 Materials and methods

7.2.1 Data

The analysis was carried out using a data set from a study conducted in two villages in Tanzania in 1993. The study involved school-aged children and included a baseline survey and six follow-ups over 24 months after an initial treatment with praziquantel. At each survey, measurements are were collected by urine filtration and a reagent strip (Boehringer Mannheim, Mannheim, Germany) on up to five consecutive days. Readings of the reagent strip are semi-quantitative with values corresponding to 0 (negatives), trace (< 5 red blood cells (RBC)/μl), 1 ($5 - 10$ RBC/μl), 2 (~ 50 RBC/μl), and 3 (~ 250 RBC/μl) [Hatz et al., 1998]. The study aimed to look at the evolution of *S. haematobium* pathology after MDA. Summary measures of the data by village at the seven time points are presented in Table 7.1.

Table 7.1: Observed prevalence and mean infection intensities by urine filtration as well as prevalence by microhaematuria reagent strips on one and five cumulative days by village (Tanzania survey, 1993 [Hatz et al., 1998]).

		Urine filtration						**Reagent strip**					
		1 day		**5 days**				**1 day**			**5 days**		
Village	Survey	N	Positive (%)	N	Positive (%)	μ	μ^+	N	Trace (%)	1(%)	N	Trace (%)	1 (%)
Mikumi	Baseline	310	56.5	357	70.0	52.4	74.8	311	59.1	53.2	357	77.6	63.6
	2 months	288	15.3	328	22.3	0.8	3.5	287	24.0	19.5	328	38.4	27.7
	4 months	205	12.2	249	20.9	1.44	6.9	206	22.8	19.9	248	39.5	31.4
	6 months	241	40.7	244	59.0	5.5	9.4	172	32.0	23.3	244	50.4	39.3
	12 months	221	38.5	288	59.4	13.5	22.8	220	48.2	22.2	288	60.7	48.6
	18 months	202	56.9	237	69.2	22.9	33.1	141	100	77.3	187	100	89.3
	24 months	175	41.7	210	60.0	11.7	19.8	76	100	84.2	139	100	88.5
Msimba	Baseline	178	82.6	207	91.8	107.0	116.6	179	86.6	58.7	207	93.2	81.6
	2 months	163	15.3	197	32.5	1.1	3.5	163	42.3	23.3	196	54.6	40.3
	4 months	129	10.9	171	24.6	0.4	1.8	129	17.1	9.3	171	33.9	22.8
	6 months	173	23.7	177	42.9	2.5	5.9	150	23.3	15.3	176	51.1	26.1
	12 months	76	32.9	77	98.7	5.8	5.8	0	-	-	32	96.9	65.6
	18 months	144	59.7	163	72.4	66.7	92.1	89	100	83.1	117	100	88.9
	24 months	113	44.2	151	50.3	36.8	73.0	68	98.5	70.6	85	98.8	81.2

UF is short for urine filtration of 10 ml urine; DST is short for dipstick to detect microhaematuria trace considered positive; DS1 is dipstick with trace considered negative; ending with 1 means a single test; ending with 5 is the maximum of 5 tests; N-xx is the number of individuals tested; μ is the mean number of eggs in 10 ml of urine in the population after 5 tests; μ^+ is the mean number of eggs in 10 ml of urine in the positive individuals after 5 tests.

Prevalence by a single urine filtration ranges from 10% to 60% and cumulatively over five days from 20% to 98.7%. The mean infection intensity in the population goes from 0.4 to 107 eggs/10 ml of urine and in the positive individuals only from 1.8 to 116.6 eggs/10 ml. Reagent strip prevalence ranges from 17.1% to 100% when traces are considered positive and from 9.3% to 83.1% when traces are treated as negative based on observations from a single day. Application of reagent strip diagnostics on five days increased the observed prevalence from 33.3% to 100% when traces were considered positive and from 22.8% to 89.3% when traces are grouped in the negatives. It is common to have individuals where only one diagnostic test was performed and therefore, observed prevalence by reagent strip and urine filtration at the same time point and village are not directly comparable.

7.2.2 Egg count model

We assumed that the infection status and intensity of each individual was independent across follow-ups. Therefore, with two villages and seven surveys (baseline and all follow-ups) we have for analysis 14 independent populations with varying numbers of individuals and different prevalence levels.

For each individual i in population j, $j = 1, \ldots, 14$ we observed egg counts, Y_{jid}^{UF} according

to the urine filtration test day d, $d = 1, 2, \ldots, 5$ and results from a semi-quantitative reagent strip for microhaematuria that is coded in 15 binary variables, $Y_{ji}^{DS,T,z}$, $Y_{ji}^{DS,1,z}$, $Y_{ji}^{DS,2,z}$ are 1 if the highest result of the first z days is at least a T for trace, 1, or 2, respectively. The model was fitted separately for each of the 15 binary results $Y_{ji}^{DS,x,z}$, $x = T, 1, 2$, combined with all Y_{jid}^{UF}. Thereby, we can infer on the sensitivity of repeated microhaematuria measurements without having to model the correlation structure explicitly.

We assumed that each population consists of a proportion of infected individuals π_j where each individual had a disease status D_{ji} and infection intensity $\lambda_{ji} + \mu_{min}$. For infected individuals with $D_{ji} = 1$ measurements on consecutive days were modelled as a negative binomial distribution with dispersion parameter depending on the infection-intensity $\lambda_{ji} + \mu_{min}$ which furthermore were assumed to be distributed as a shifted gamma distribution as follows.

$$\begin{aligned} Y_{jid}^{UF} &\sim NB(\lambda_{ji} + \mu_{min}, k_{ji}) \\ \log(k_{ji}) &= k_0 + (\lambda_{ji} + \mu_{min})k_1 \\ \lambda_{ji} &\sim Gamma((\mu_j^+ - \mu_{min})\alpha_j, \alpha_j) \end{aligned} \tag{7.1}$$

μ_{min} was the minimum infection intensity of a single pair of worms, k_{ji} was the dispersion parameter of individual i in population j that depended on parameters k_0 and k_1, μ_j^+ was the mean infection intensity of a positive individual in population j, and α_j was the aggregation parameter that described the worm-pair aggregation in population j. Individuals with infection status $D_{ji} = 0$ have observations $Y_{jid}^{UF} = 0$ which is equivalent to 100% specificity. The sensitivity of urine filtration s_{jin} for individual i in population j after n days was calculated using the probability of repeated zero measurements under the negative binomial model, that is $s_n(\lambda) = 1 - \left(\frac{k(\lambda)}{\lambda + k(\lambda)}\right)^{n \cdot k(\lambda)}$, where $k(\lambda) = e^{k_0 + \log(\lambda + \mu_{min})k_1}$.

Microhaematuria measurements are modelled by a Bernoulli distribution,

$$P(Y_{ji}^{DS,x,z} \mid D_{ji}) = \begin{cases} Be(s_{ji}^{DS,x,z}) & \text{, if } D_{ji} = 1 \\ Be(1 - c^{DS,x,z}) & \text{, if } D_{ji} = 0 \end{cases} \tag{7.2}$$

$$\text{logit}(s_{ji}^{DS,x,z} \cdot a_3^{x,z}) = a_0^{x,z} + a_1^{x,z} \cdot \log(\lambda_{ji})^{\log(a_2^{x,z}+1)}$$

A parametric model was used to relate infection intensity with diagnostic sensitivity, where s^{DSx-z} is the infection intensity dependent sensitivity defined by four parameters. a_0 sets the

sensitivity for infections approaching to zero intensity, a_1 describes the rate of increase with infection intensity, a_2 changes the shape of the curve, and a_3 limits the maximum sensitivity for very severe infections. $c^{DS,x,z}$ is the specificity of the microhaematuria reagent strip, that is the probability to have no result of at least x after z observations of an uninfected individual.

Models were also run with sensitivity and specificity parameters of the reagent strip stratified by sex. The model was formulated using the Bayesian paradigm. We have chosen a uniform prior for π_j, a gamma distribution with mean 100 and standard deviation (SD) 100 for the mean infection intensity μ_j^+, a gamma distribution with mean 1 and SD 1 for the population variation α_j, a beta distribution with parameters 10 and 1 for c^{DS}, normal distributions with mean 0 and SD 2 for k_0 and k_1, a normal distribution with mean -1 and SD 1.5 for a_0, a gamma distribution with shape and scale parameters 5 and 30 for a_1, a normal distribution with mean 10 and SD 10 for a_2, and a beta distribution with parameters 10 and 1 for a_3 to ensure a non-informative distribution of sensitivity curves in the relevant range of infection intensities, μ_{min} was set to 0.5 eggs per 10 ml [Chan et al., 1996]. Inference was done using Markov chain Monte Carlo (MCMC) simulation. Model fit was carried out in Stan version 2.16.2 (Stan Development Team; mc-stan.org) running 20 chains for 2000 iterations of which the first 500 were thrown away as warm-up [Carpenter et al., 2016]. There were no divergent transitions and convergence was assessed using the Gelman-Rubin diagnostics as well as visual inspection of chains [Gelman and Rubin, 1992].

7.2.3 Simulation

To assess the relation between prevalence observed by microhaematuria and urine filtration for one to five days, we run extensive simulations of populations in diverse transmission settings. We assume that worms are negative binomially distributed in the population with worm aggregation parameter w^{agg} and that a proportion of 30% of the worms are female [Anderson, 1986; May and Woolhouse, 1993]. The mean number of eggs/10 ml per worm-pair was taken as 1 [Chan et al., 1996]. For the parameters, a_0, a_1, a_2, a_3, k_0, k_1 100 draws were taken directly from the posterior distribution and thereby correlation between parameters incorporated in the simulation. The mean number of worms per individual in a population was varied in 30 equal steps on the

log-scale from 1 to 400. w^{agg} according to a normal distribution with a mean of 0.2 and a SD of 0.03. For each population worm load for 5000 individuals and corresponding urine filtration and reagent strip results are simulated for all five days.

7.3 Results

7.3.1 Sensitivity of urine filtration

Using the egg count model, we estimate the sensitivity of the urine filtration from a single sample to a total of five samples collected across the five days. The posterior medians for parameters k_0 and k_1 which define the relation between the sensitivity and mean infection intensity and were estimated to around -2.3 and 0.4, respectively. The influence of the choice of DS, x, z on estimates of k_0 and k_1 was small (see Tables A1-3 in the appendix). The aggregation parameter of the negative binomial distribution increases from 0.1 to 0.45 for infection intensity increasing from 0 to 50 eggs/10 ml of urine which corresponds to a reduction in overdispersion when the infection intensity increases. This is equivalent to a standard deviation of 21 at a density of 10 eggs/10 ml, 76 at a density of 50 eggs/10 ml, and 3.4 at a density of 1 egg/10 ml. The posterior distributions of k_0 and k_1 combined with the probability for repeated zeros under the negative binomial model enabled us to estimate the sensitivity of urine filtration. In Figure 7.1 we present the estimated sensitivity of urine filtration cumulatively for one to five days (i.e. single vs five tests) as a function of infection intensity.

A single urine filtration reaches sensitivity of more than 85% for heavy infections above 50 eggs/10 ml while the sensitivity drops below 50% at around 7 eggs/10 ml. The sensitivity increases substantially after repeated urine filtration. For example, the sensitivity increases from 50% to 75% from a single to a double test at 7 eggs/10 ml. After five days of urine filtration, an average infection with a single worm-pair at 1 egg/10 ml has a probability of around 60% to be detected.

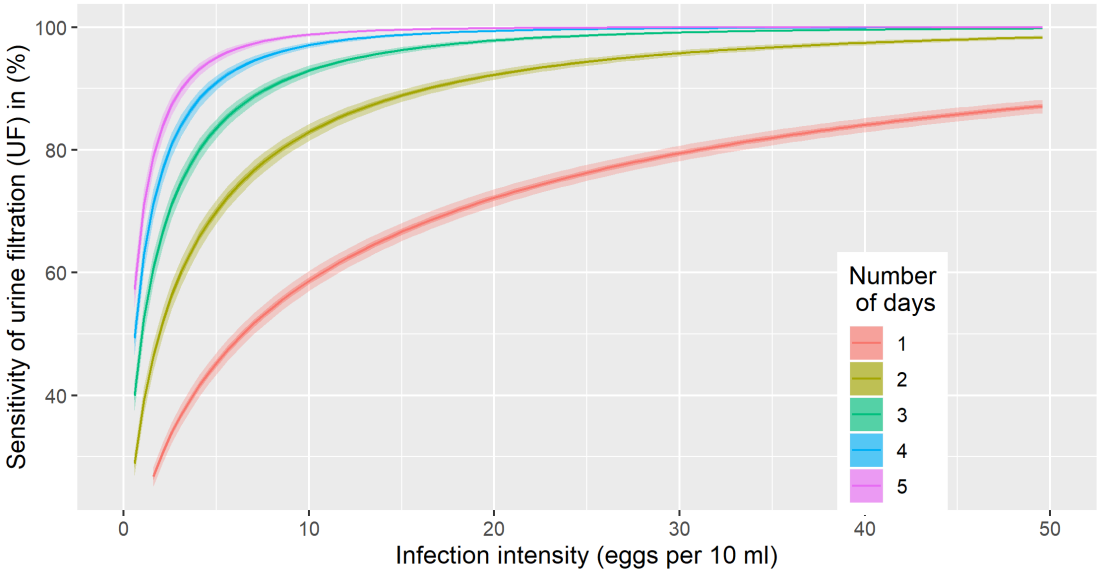


Figure 7.1: Sensitivity estimates of urine filtration (10 ml of urine) from a single to a total of five samples over five days on one to five days. Curves indicate posterior medians and shaded areas provide a 50% and 95% Bayesian credible interval.

7.3.2 Sensitivity of reagent strip

The model is run for each interpretation of the reagent strip DS, x, z separately, where x is either T for trace, 1 or 2 and z the total number of samples goes from 1 to 5 (i.e. the test is positive when at least on one of the first z days a value of at least x in the semi-quantitative interpretation of the reagent strip is reported). The parameters a_0, a_1, a_2, a_3, and c fully define the diagnostic accuracy of each DS, x, z. We present the specificity c for each interpretation in Table 7.2. The specificity is estimated to basically 100% even after multiple days when only the 2 and 3 test readings are considered positive while when trace result is treated as positive the specificity of even a single sample is only 96% and reduces to 85% after five days.

Table 7.2: Specificity of each interpretation of the reagent strip DS, x, z defined as a result of at least x (x being T for trace, 1, or 2), on the first z days. Values are given between 0 and 1 where 1 corresponds to zero probability of a false-positive.

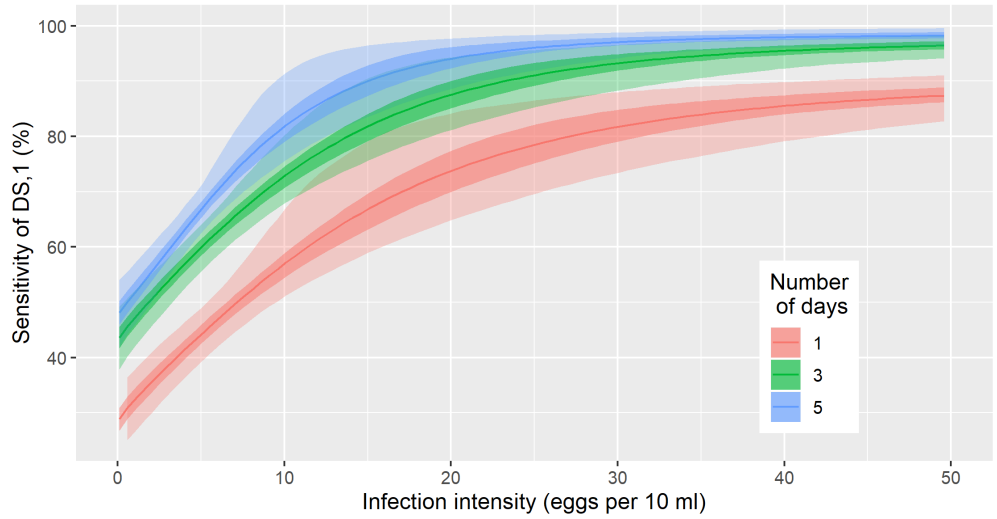
Reading	**DS,x,1**	**DS,x,2**	**DS,x,3**	**DS,x,4**	**DS,x,5**
$x = T$	0.96 (0.91 - 1.00)	0.90 (0.84 - 0.97)	0.89 (0.83 - 0.97)	0.87 (0.81 - 0.94)	0.85 (0.79 - 0.92)
$x = 1$	0.99 (0.96 - 1.00)	0.99 (0.96 - 1.00)	0.99 (0.95 - 1.00)	0.97 (0.93 - 1.00)	0.95 (0.91 - 0.99)
$x = 2$	0.99 (0.98 - 1.00)	0.99 (0.97 - 1.00)	0.99 (0.97 - 1.00)	0.99 (0.96 - 1.00)	0.98 (0.95 - 1.00)

The parameters a_0, a_1, a_2, a_3 are reported for all 15 interpretations in Tables A1-A3 in the appendix. Combined with Equation 7.2, these parameters estimate the infection intensity-dependent sensitivity of the reagent strip for microhaematuria to diagnose *S. haematobium*. Figures 7.2a and 7.2b show the sensitivity of the reagent strip when trace is considered positive or negative, for cumulative results from one, three, and five days. When traces are considered positive a single reagent strip detects basically all infections with an intensity above 10 eggs/10 ml, and still 40% of infections with a single worm-pair. Repeating the test on consecutive days increases the sensitivity, for example from 40% to 70% for very weak infections after five days, but this increase is modest when compared to urine filtration.

Considering traces as negative leads to a higher specificity of 99% for a single reagent strip and still 95% after five samples. When traces are positive, the specificity after 5 samples is reduced to 85%, however, the sensitivity is higher compared to traces as negatives. A single reagent strip has only a 60% chance to detect an infection of 10 eggs/10 ml and still less than 90% at 50 eggs/10 ml of urine. Even repeated sampling over five days increases the sensitivity only up to 80% at 10 eggs/10 ml compared to the virtually 100% sensitivity at the same intensity when traces are considered positive. The sensitivity of a single reagent strip when traces are considered negative is very similar to a single urine filtration at all levels of infection intensity. Stratification by sex did not show any differences in parameter estimates that would indicate an important interaction for example with menstruation.

The semi-quantitative results of the reagent strip are associated with the infection intensity of an individual. We show the proportion of trace, 1, and > 1 results for infection intensities up to 50 eggs/10 ml in Figure 7.3. The non-monotonic behavior close to an infection intensity of zero is due to increased uncertainty in the sensitivity estimates (not shown in the plot, but visible in Figures 7.2a and 7.2b). At very low infection intensities, there is still a considerable probability of about 40% for readings of 2 or 3, while at 50 eggs/10 ml almost 80% of tests show a 2 or 3. Trace results on the other hand decrease from a proportion of about 30% to less than 10% at 50 eggs/10 ml while the proportion of samples with reading 1 remains relatively constant.

(a)



(b)

Figure 7.2: Sensitivity of a reagent strip in relation to infection intensity for measurements from a single to a total of 3 and 5 cumulative tests a) trace results are included in the positives and b) trace results are included in the negatives. Shaded areas are the 50% and 95% BCI.

7.3.3 Relation between urine filtration and reagent strip observed prevalence

We conducted extensive simulations of 3,000 populations with 5,000 individuals each and generated observations of the two diagnostic tests over five consecutive days based on the

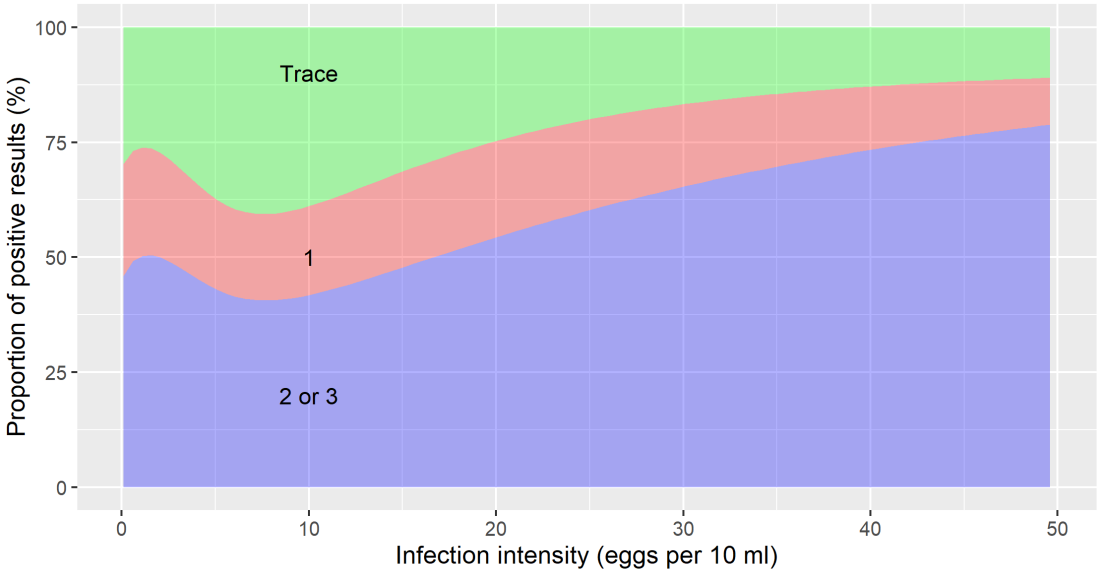


Figure 7.3: Proportion of semi-quantitative results of reagent strips for microhaematuria in relation to *S.haematobium* infection intensity. The 2 and 3 readings are grouped together.

individual level results from the egg count model above. Figure 7.4 depicts the relation between the observed prevalence by reagent strip and urine filtration for up to five samples collected over five days. Traces are treated as either positives or negatives. Measurements based on a single sample of reagent strip shows equal or higher observed prevalence than urine filtration. When all five samples are considered, the prevalence is higher by urine filtration compared to reagent strip when traces are treated as negatives. For the sampling scheme recommended by WHO, i.e. single day sample, and reagent strip without trace results the observed prevalence by both tests are very similar, independent of the setting.

We translated the WHO prevalence thresholds from urine filtration into microhaematuria by taking all populations with observed prevalence by urine filtration within a narrow interval of $\pm 0.5\%$ around the threshold and calculating the mean of the observed prevalence by reagent strip and the corresponding Bayesian credible intervals. Table 7.3 shows results for three thresholds of urine filtration, 10%, 30%, and 50%, and the reagent strip when traces are considered positive or negative, for sampling schemes based on a single up to five urine samples over five days. Most relevant for evaluating treatment needs are results for sampling on a single day. When traces are

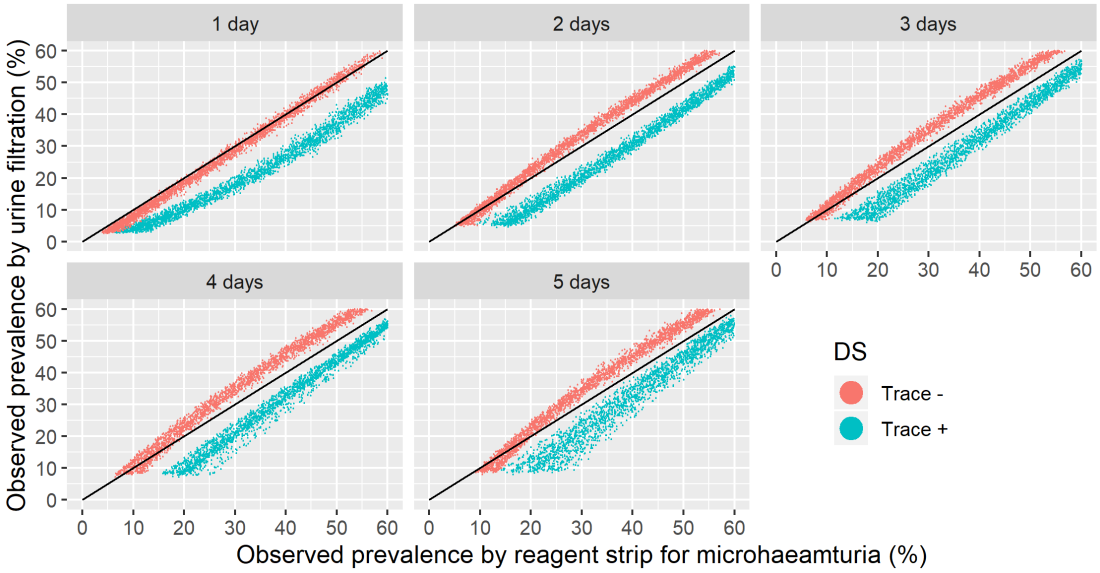


Figure 7.4: Simulated relation between observed prevalence of *S. haematobium* by urine filtration and by reagent strip based on one to five samples collected over one to five days and when traces are considered positive or negative. The black line indicates equivalence between urine filtration and microhaematuria.

considered negative, the prevalence thresholds for microhaematuria are very close to the urine filtration ones. Including traces in the positive requires upwards adjustment of the thresholds, for example, a 10% urine filtration corresponds to about 20% prevalence by reagent strip.

Table 7.3: Translation of prevalence thresholds from urine filtration (UF) into reagent strip for sampling schemes varying from on one to five urine samples over 5 days when traces are considered as positive or negative. All numerical values are percentages and the brackets contain 95% percentiles from the simulation.

DS	UF (%)	1 day	2 days	3 days	4 days	5 days
Traces +	10	19.3 (16.0 - 22.6)	19.4 (16.1 - 22.7)	19.3 (15.0 - 23.5)	21.0 (16.9 - 25.1)	21.0 (14.7 - 27.3)
	30	43.4 (40.1 - 46.7)	39.4 (37.3 - 41.5)	37.2 (33.1 - 41.4)	37.8 (34.6 - 41.0)	37.1 (33.4 - 40.8)
	50	61.1 (58.8 - 63.4)	57.0 (55.0 - 59.0)	55.6 (53.2 - 58.0)	55.5 (53.2 - 57.9)	54.4 (50.7 - 58.1)
Traces -	10	11.9 (10.1 - 13.6)	10.0 (8.2 - 11.9)	9.8 (7.4 - 12.1)	10.1 (7.3 - 13.0)	11.9 (9.6 - 14.3)
	30	30.8 (28.4 - 33.1)	26.4 (24.6 - 28.3)	25.2 (23.1 - 27.3)	25.4 (23.4 - 27.4)	26.2 (24.4 - 28.0)
	50	49.4 (47.2 - 51.5)	45.8 (44.0 - 47.6)	43.8 (41.6 - 45.9)	44.1 (41.3 - 46.8)	44.6 (42.1 - 47.2)

7.4 Discussion

We determined the diagnostic accuracy in relation to infection intensity of a reagent strip used for detection of *S. haematobium* in urine, from survey data with parallel diagnosis by urine filtration at individual level, using a Bayesian egg count model. The semi-quantitative results of the reagent strip were incorporated in the modelling by determining the sensitivity and specificity profiles, when trace results are treated as positive or negative, and when only the 2 and 3 readings are considered as positive. Furthermore, we related the observed prevalences by the two tests and translated the urine filtration treatment prevalence thresholds recommended by WHO into reagent ones.

For individual level diagnosis, the reagent strip shows almost perfect sensitivity above 7 eggs/10 ml of urine while specificity is still relatively high at 96% for a single urine sample when traces are included in the positives. When traces are excluded, a much more similar profile to urine filtration is created across all infection intensities. A key difference between microhaematuria and urine filtration is that results are more correlated across samples for microhaematuria leading to less improvement in sensitivity by repeating measurements. It is reasonable to believe that the presence of eggs in urine is less variable than the egg excretion by parasitic worms in the case where few worm-pairs are present. The reduction in the overdispersion of the negative binomial distribution of egg outputs with increasing infection intensity supports arguments that in the presence of a larger number of worms, the day-to-day variations reduce due to averaging.

Point estimates of urine filtration and reagent strip sensitivity as well as specificity from other studies are difficult to compare to our results because previous studies do not take into account the relation to infection intensity and they are averaged over the population, therefore, they are much more setting-dependent. Most analysis defined urine filtration as the 'gold' standard. This approach makes the results of each study additionally dependent on the sampling scheme. Nonetheless, for example the results by Kosinski et al. (2011) show around 50% sensitivity reagent strip using for a single urine sample, 60% for a double, and 70% for a triple sample. For urine filtration these estimates were 50%, 80% and 100%, respectively [Kosinski et al., 2011]. Specificity of the dipstick decreases from 93% for single, to 88% for double and 83% for triple

sample. These results are generally in agreement with our estimates, especially considering that those values over-estimate the diagnostic error because of the definition of urine filtration as a 'gold' standard. Day-to-day variation in egg output is large making a stable classification of infection intensity into severity classes difficult. We show that the semi-quantitative results of the reagent strip have a clear relation to infection intensity. Variation in microhaematuria seems to be lower indicating that classification based on a reagent strip might be more stable and more insightful.

Results were derived from the analysis of a study conducted in Tanzania in 1993 using a reagent strip produced by Boehringer Mannheim, a company that was bought and incorporated by Roche in 1997, it is unclear how representative the results are for other similar tests [Hatz et al., 1998]. For example a study published in 1999 conducted in Kenya utilised a reagent strips called Hemastix (Ames, Bie and Bernsten, Copenhagen, Denmark) instead as well as a study by Mafe et al. (1997) in Nigeria published in 1997 [Mafe, 1997; Kahama et al., 1999]. More recently Kosinski et al. (2011) evaluated a semi-quantitative dipstick called U-11 Urinalysis Reagent Strips (Mindray Co. Ltd., China), however, this test seems to be older as well and has been evaluated already in 1985 by Mott et al. [Mott et al., 1985; Kosinski et al., 2011]. Thus, more data has to be incorporated to study possible differences between reagent strips from different companies or from the same company over time. A part of the data is from just after MDA and depending on how fast microhaematuria clears after treatment, this is a possible confounder that was not included into the modelling.

On the population level, we see a clear relation between observed prevalence by urine filtration and by reagent strip. The size of the 3,000 simulated populations was fixed at 5,000 to be able to observe the influence of uncertainty in diagnostic accuracy while limiting the influence of sampling error. Important assumptions in the simulation model are primarily the negative binomial distribution of worms in the population and the aggregation parameter that was considered independent of infection intensity. The latter may not hold because it has been assessed for hookworm infections, showing that the aggregation parameter increases with infection intensity [Truscott et al., 2019]. In accordance with WHO recommendations for evaluation of *S. haematobium* prevalence from a single day, we recommend translating urine

filtration thresholds of 10% and 50% into 12% and 50% ones by reagent strip, classifying traces as negative. Trace positive individuals should however still be treated, as they are likely to be infected with *S. haematobium*. Nevertheless, the reagent strip with traces considered negative serves as a convenient proxy to estimate prevalence almost equivalent to single day urine filtration of 10 ml.

7.4.1 Conclusion

In conclusion, the reagent strip for microhaematuria shows a high sensitivity for *S. haematobium* above 7 eggs/10 ml while still maintaining a high specificity above 95%. When higher specificity is required, traces can be considered as negative to exclude the majority of false-positives and lead to a sensitivity almost equal to a single sample urine filtration, therefore enabling direct translation of observed prevalence in the population. Caveats of the study are that the data is from a single type of reagent strip from a survey carried out in two locations albeit at multiple time points after treatment. It is imperative to validate the results with additional data for example using the Hemastix and U-11 reagent strips mentioned above in settings close to elimination. However, our results indicate that reagent strips are potentially very sensitive and specific diagnostic tools for *S. haematobium*.

Appendix

A1: Parameters describing the sensitivity and specificity of a reagent strip for microhaematuria including trace results (T) from one to 5 days.

Number of tests	1	2	3	4	5
Reagent strip					
c	0.96 (0.91 - 1.00)	0.90 (0.84 - 0.97)	0.89 (0.83 - 0.97)	0.87 (0.81 - 0.94)	0.85 (0.79 - 0.92)
a_0	-0.33 (-0.67 - -0.03)	0.14 (-0.16 - 0.40)	0.46 (0.17 - 0.73)	0.68 (0.40 - 0.93)	0.77 (0.50 - 1.05)
a_1	0.11 (0.03 - 0.25)	0.09 (0.03 - 0.23)	0.10 (0.03 - 0.24)	0.09 (0.03 - 0.21)	0.09 (0.03 - 0.21)
a_2	3.48 (1.83 - 8.49)	3.84 (1.93 - 8.88)	4.07 (1.92 - 10.00)	4.23 (2.19 - 9.76)	4.53 (2.26 - 11.29)
a_3	0.98 (0.97 - 0.99)	0.99 (0.99 - 1.00)	0.99 (0.98 - 1.00)	1.00 (0.99 - 1.00)	1.00 (0.99 - 1.00)
Urine filtration					
k_0	-2.34 (-2.46 - -2.22)	-2.36 (-2.49 - -2.25)	-2.37 (-2.49 - -2.24)	-2.36 (-2.48 - -2.25)	-2.36 (-2.48 - -2.24)
k_1	0.38 (0.35 - 0.42)	0.39 (0.36 - 0.42)	0.39 (0.36 - 0.42)	0.39 (0.36 - 0.42)	0.39 (0.36 - 0.42)

A2: Parameters describing the sensitivity and specificity of a reagent strip for microhaematuria excluding trace results (1) from one to 5 days.

Number of tests	1	2	3	4	5
Reagent strip					
c	0.99 (0.96 - 1.00)	0.99 (0.96 - 1.00)	0.99 (0.95 - 1.00)	0.97 (0.93 - 1.00)	0.95 (0.91 - 0.99)
a_0	-0.80 (-1.13 - -0.50)	-0.47 (-0.73 - -0.22)	-0.23 (-0.49 - 0.00)	-0.09 (-0.35 - 0.15)	-0.06 (-0.33 - 0.19)
a_1	0.20 (0.07 - 0.41)	0.18 (0.07 - 0.36)	0.16 (0.06 - 0.33)	0.16 (0.06 - 0.32)	0.15 (0.05 - 0.33)
a_2	1.25 (0.72 - 2.53)	1.21 (0.78 - 2.15)	1.46 (0.90 - 2.91)	1.62 (0.96 - 3.45)	1.79 (1.00 - 3.90)
a_3	0.90 (0.85 - 0.96)	0.98 (0.94 - 1.00)	0.97 (0.95 - 1.00)	0.98 (0.95 - 1.00)	0.99 (0.96 - 1.00)
Urine filtration					
k_0	-2.31 (-2.43 - -2.19)	-2.31 (-2.42 - -2.19)	-2.30 (-2.42 - -2.19)	-2.31 (-2.42 - -2.19)	-2.30 (-2.42 - -2.18)
k_1	0.38 (0.35 - 0.41)	0.38 (0.35 - 0.41)	0.38 (0.34 - 0.41)	0.38 (0.35 - 0.41)	0.37 (0.34 - 0.41)

A3: Parameters describing the sensitivity and specificity of a reagent strip for microhaematuria excluding trace and 1+ results (2) from one to 5 days

Number of tests	1	2	3	4	5
Reagent strip					
c	0.99 (0.98 - 1.00)	0.99 (0.97 - 1.00)	0.99 (0.97 - 1.00)	0.99 (0.96 - 1.00)	0.98 (0.95 - 1.00)
a_0	-1.43 (-1.78 - -1.12)	-1.30 (-1.60 - -1.04)	-1.09 (-1.37 - -0.83)	-0.98 (-1.26 - -0.72)	-0.93 (-1.21 - -0.66)
a_1	0.25 (0.11 - 0.46)	0.21 (0.09 - 0.39)	0.19 (0.08 - 0.36)	0.19 (0.07 - 0.36)	0.17 (0.07 - 0.35)
a_2	0.89 (0.60 - 1.48)	1.08 (0.76 - 1.72)	1.25 (0.85 - 2.02)	1.32 (0.91 - 2.13)	1.49 (1.00 - 2.49)
a_3	0.93 (0.86 - 0.99)	0.98 (0.94 - 1.00)	0.98 (0.95 - 1.00)	0.99 (0.96 - 1.00)	0.99 (0.97 - 1.00)
Urine filtration					
k_0	-2.29 (-2.41 - -2.18)	-2.30 (-2.42 - -2.18)	-2.29 (-2.40 - -2.17)	-2.29 (-2.41 - -2.18)	-2.29 (-2.40 - -2.17)
k_1	0.37 (0.34 - 0.40)	0.37 (0.34 - 0.41)	0.37 (0.34 - 0.40)	0.37 (0.34 - 0.40)	0.37 (0.34 - 0.40)

Chapter 8

Discussion

This thesis contributes with methodology to analyse imperfect diagnostic tests for helminthiasis in the absence of a 'gold' standard, presents their applications to schistosomiasis and soil-transmitted helminths to inform decision making. Models have been developed for egg count tests, for binary, and for semi-quantitative tests measured on the same individual using random-effect models to incorporate their dependence on infection intensity. Specifically, we model infection intensity-dependent sensitivity of urine filtration and of the Kato-Katz techniques based on samples collected from multiple days to diagnose *S. haematobium* and *S. mansoni* infection, respectively. We evaluated alternative diagnostic tests such as the POC-CCA test for *S. mansoni*, a reagent strip to detect microhaematuria, a questionnaire regarding recent history of blood in urine, and visible blood in urine for *S. haematuria*. Furthermore, we evaluated the 'true' prevalence and treatment decisions made for different diagnostic methods. Furthermore, we showed that it is possible to estimate 'true' prevalence of *S. mansoni* from summary measures of observed prevalence and geometric or arithmetic mean infection intensity from Kato-Katz alone and compared survey designs based on Kato-Katz and POC-CCA diagnostics.

Chapters 2 to 7 consist of manuscripts either published or in preparation for publication and contain their own detailed discussion and conclusion. The current chapter summarises the key findings and highlights the main contributions, and discusses limitations and potential extensions.

8.1 Significance of the work

8.1.1 Modelling egg count diagnostics

Previous modelling work of helminthiasis data that took into account diagnostic error, categorised the test results (i.e. egg counts) into a binary outcome (positive/negative) or into few egg intensity-based categories. However, variations in counts from sample to sample and from measurement to measurement carry important information that is not preserved by categorization in a few groups [de Vlas et al., 1993; Utzinger et al., 2001]. In our work, we developed models, which directly fitted to the egg count data rather than to their transformation into few categories. Chapters 2, 3, and 4 include variations of the same egg count model for *S. mansoni* based on the Kato-Katz diagnostic. In Chapter 2 the model considers sampling designs with a single measurement per day over few consecutive days, therefore, day-to-day and slide-to-slide variations are inseparable. Chapter 3 and 4 analyse basically the same suit of datasets which have up three Kato-Katz slides on two or three days. Thereby, it becomes possible to separate the two sources of variations where the day-to-day variation captures the variation in egg excretion of the worm-pairs and the slide-to-slide variation can be understood as the non-random distribution of eggs in a sample and also incorporates variations due to technicians and experimental conditions. Thus, in Chapters 3 and 4 the relation between infection intensity and sensitivity is defined by two parameters while the model in Chapter 2 requires only one.

In Chapters 6 and 7 two extensions of the ideas in new domains are presented in analysis of data from urine filtration to detect *S. haematobium*. Chapter 6 models measurements of urine filtration from a single day; however, the statistical model remains the same. This is made possible by multiple diagnostic tests on the same individual, which inform about the diagnostic sensitivity even when repeated egg counts are not done. Chapter 7 considers data from five urine filtration tests and assumes that the aggregation parameter of the negative binomial distribution modelling the egg count depends on the infection intensity. We find that the aggregation parameter of the negative binomial distribution increases with increasing infection intensity, which is not surprising when the infection intensity of an individual is understood as a collection of worm-pairs independently excreting eggs and therefore averaging out each other's variations.

An important aspect of the model is the minimum infection possible caused by a single worm-pair. In *S. mansoni* estimates are in the range of 1 to 5 EPG while for *S. haematobium* we expect about 1 egg/10 ml of urine [Cheever et al., 1994; Chan et al., 1996]. The sensitivity of count based diagnostic tests as determined by our model varies strongly within the range of likely infection intensities for a single worm-pair. Therefore, it is difficult to estimate the sensitivity for single worm-pair infections or to know what would be needed to detect every single infected individual with a viable worm-pair in an elimination setting [Colley et al., 2017]. The most likely values for the sensitivity of a double slide Kato-Katz in case of a single worm-pair of *S. mansoni* are between 10% and 30%. Similarly, the sensitivity is below 30% for a single urine filtration to detect a pair of *S. haematobium* (see Chapters 4 and 7).

In every application of the model it becomes clear that sensitivity varies strongly across infections classified as light which is below 100 EPG for *S. mansoni* and below 50 eggs/10 ml for *S. haematobium*. The difference between a single and repeated observation of the diagnostic test is large and depends on whether the test is performed on the same sample or on samples collected from separate days. Furthermore, the composition of infection intensities in a population is different from setting to setting making replacement of individual level measures with an averaged sensitivity difficult (see Chapter 4). Therefore, application on an individual-level egg count model taking into account the sampling design and the infection intensity clearly adds value to the information extracted from a dataset compared to modelling approaches that are based on binary data and a constant value for the sensitivity. The application of our model to hookworm data in Chapter 2 shows that our developed approach is useful for helminth infections more generally, rather than only schistosomiasis data.

8.1.2 Evaluation of alternative diagnostic tests

In addition to determining the sensitivity of the egg count test for different sampling schemes our modelling approach produces estimates of the individual infection intensity and the 'true' prevalence in the population. We show that the infection intensity can be used as an individual-level random-effect in a latent class model for diagnostic tests other than egg counts given that the sensitivity of those tests is also primarily determined by the infection intensity. Thereby, we

can estimate the relation between sensitivity and infection intensity for any binary test measured in addition to an egg count test. In Chapters 3, 6, and 7 we present applications of our model to the point-of-care circulating cathodic antigen (POC-CCA) test results for *S. mansoni* and for tests based on the presence of blood in urine for *S. haematobium* such as inspection of urine samples for visible blood, a reagent strip test to detect microhaematuria, and a questionnaire regarding recent history of blood in urine.

We utilise parametric models to capture the relation between infection intensity and sensitivity with parameters describing sensitivity for very light infections, the rate of increase of sensitivity with infection intensity, the shape of the relationship, and the sensitivity for very heavy infections wherever appropriate. The main assumption behind the parametric model is that the sensitivity increases monotonically with infection intensity. The specificity can be estimated given that we have an estimate of the 'true' prevalence together with the sensitivity and the distribution of infection intensity in the population.

Challenges in applying the model to real test arose in two distinct ways. The first becomes clear when examining the application of multiple POC-CCA and Kato-Katz tests per individual for *S. mansoni* in Chapter 3. In the best case, our egg count model for the Kato-Katz technique is able to estimate the prevalence of infections with viable worm-pairs that excrete eggs. However, infections with juvenile worms as well as worms of a single sex cannot be detected. In itself, this might not seem like an important drawback because both transmission and burden is primarily caused by egg excretion. However, the POC-CCA test is sensitive for antigens unrelated to egg excretion and therefore measures the prevalence of individuals with worms which is different from the prevalence of viable worm-pairs especially in settings approaching elimination [van Lieshout et al., 2000]. In Chapter 3, we approach the problem by separating the population in three categories instead of two, that is infected with viable worm-pairs, infected with worms but no worm-pairs, and non-infected. Information on the infected with worms but no worm-pairs is derived by the model, which is identifiable due to data from non-endemic settings with few positives that allow estimation of the specificity of the POC-CCA. Haggag et al. (2019) have recently called into question the estimates of high specificity of trace positive POC-CCA by demonstrating that patients stay trace positive after repeated treatment with praziquantel and

when no eggs are detected over a month [Haggag et al., 2019b,a]. Our simulations in Chapters 4 and 5 generated the distribution of worms directly taking into account the worm mating process to model the distribution of worm-pairs. Thereby, we can simulate observed data by Kato-Katz and POC-CCA realistically and study what range of discrepancy between the observed prevalence we could expect. It is an open question however whether the additional positives by POC-CCA can be attributed to egg-negative infections or they are false-positives due to a so far not understood mechanism. Greter et al. (2016) for example reported cross-reactions with pregnancy as a possible source of false-positives [Greter et al., 2016].

A second challenge becomes clear in Chapters 6 and 7. In Chapter 6 we analyse data where for each individual in addition to urine filtration for *S. haematobium*, multiple diagnostic tests based on haematuria were performed while in Chapter 7 we incorporate data from reagent strip measurements on multiple days. Although different situations, the problem arises in both cases from the fact that haematuria-based measurements are not independent when conditioned on infection intensity. On one hand when an individual suffers from an infection intensity λ_i the individual that was tested positive for haematuria has a higher likelihood to be positive on the next day as the negatively tested one. On the other hand, an individual that showed haematuria for other reasons than *S. haematobium* has a higher probability to be a false-positive repeatedly. In Chapter 6, we take this into account by assuming an additional latent 'blood-in-urine' variable while in Chapter 7 we run the model for each alternative test to urine filtration separately to avoid the above mentioned complication. The approach chosen in Chapter 6 makes assumptions about the correlation between the diagnostic tests but this was necessary, as we only had data from a single urine filtration which is not sufficient to estimate variations in egg counts and therefore the 'true' prevalence. In Chapter 7, there were urine filtration measurements on five consecutive days and therefore no assumptions about the multiple diagnostic tests are needed. In fact, we were able to determine not only the sensitivity of the reagent strip when it was performed on a single day but also for repeats without making additional assumptions. Unsurprisingly, the sensitivity increased less from day to day for the reagent strip compared to the urine filtration due to the strong correlation across days.

Finally, all alternative tests to egg count diagnostic were considered as binary in this thesis,

converting semi-quantitative test results into two categories according to different interpretations of what is considered as positive. In Chapters 3 and 7 we repeated the analysis treating semi-quantitative tests as binary; once with trace considered positive, once with trace considered negative and once with even 1 considered negative. In Chapter 6, we considered trace results as positive for the complete analysis. Decisions made regarding the ideal interpretation of an alternative test depend on the goal of the test. If the goal is to create a proxy for the egg count methods, the interpretation that matches the sensitivity of the egg count test best will be the ideal interpretation. If no positive individuals should be missed then traces should be considered positive while if the specificity is needed to be high then even a result of 1 may be considered negative. Chapter 3 shows an example where the best proxy for double slide Kato-Katz diagnostic using POC-CCA is achieved when only 2+, and 3+ results are considered positive while in Chapter 7 a reagent strip with traces considered negative shows very similar diagnostic accuracy to urine filtration on a single day.

8.1.3 Estimating 'true' prevalence for communities and districts

In the previous two subsections, we explained how we modelled the relation between sensitivity and infection intensity on the individual level. However, decisions about treatment within a mass-drug administration (MDA) framework are made at the community or district level based on the observed prevalence. Therefore, it is of interest to translate observed prevalence from one diagnostic method to another. Once sensitivity at individual-level is determined, the link to population prevalence is created by the distribution of infection intensities in the population. Whenever we fit the egg count model, we do not make assumption regarding the distribution of worm-pairs in the population, which creates a link between prevalence and mean infection intensity of an infected individual. We only assume that the positive infected individuals are distributed according to a shifted gamma distribution in the population, which enables us to make an estimate regarding how many weakly infected individuals, for whom sensitivity of the egg count test is low, were missed. The shift of the gamma distribution captures a feature of the ultimately discrete distribution of worm-pairs. Namely, that there can be no infection intensities below the intensity related to a single worm-pair while above a few worm-pairs the distribution

becomes reasonably continuous due to the variation in egg excretion of female worms.

In Chapters 4, 5, and 7 we simulated data assuming a negative binomial distribution of worms in the population and creating a distribution of worm-pairs distinct from the negative binomial ones, via a monogamous mating procedure [May and Woolhouse, 1993]. Thereby a link is created between the mean infection intensity and prevalence of worm-pair positive individuals, the prevalence of worm positive but worm-pair negative individuals, and truly negative individuals. To translate population level observed prevalence from one diagnostic test to another, assumptions about the distribution of infection intensities in the population have to be made. This distribution might be sensitive to local phenomena like common professions, culture, and history of MDA treatments. For example, it is likely that the worm aggregation parameter of the negative binomial distribution increases with mean infection intensity but the relation is not well known. Indeed, in Chapter 4 we indicate that the negative binomial distribution of worms with aggregation parameter independent of the mean worm burden over-estimates the number of very weak infections at higher mean infection intensities.

In Chapter 3 prevalence thresholds we translated without using the assumption of negative binomial distribution of worm in the population. This leads to larger uncertainty due to an increased probability of scenarios with large number of worm positives but worm-pair negatives even when there are only a small number of worm-pair positives. The simulations in Chapters 4, 5, and 7 decrease the likelihood of these scenarios and led to more accurate estimation of the relation between different diagnostic methods (Chapters 5 and 7) and different sampling schemes for Kato-Katz (Chapter 4).

When decisions are to be made not at the community level but for example at district level it is of interest to know how many communities should be tested and how many individuals should be screened per school to achieve good accuracy in the estimation of treatments needs. Simple sample size formulas cannot capture the additional uncertainty induced by the uneven distribution of schools in a district and the focality of the disease. In Chapter 5, we run simulations for *S. mansoni* based on the egg count model for different Kato-Katz sampling schemes with parallel POC-CCA for each individual. We assumed a negative binomial distribution for worm counts in the population and used it as a testbed for survey sampling designs. The diagnostic method is

shown to have a large influence on the precision of treatment decisions made and when POC-CCA is used a much smaller number of locations and children screened per location is sufficient.

8.2 Limitations and possible extensions

The main limitation of the egg count model is that the mean infection intensity μ of infected individuals is not directly linked to the prevalence of infections π. This leads to large uncertainty in the estimation of the mean infection intensity when prevalence becomes low and only few individuals are infected. The same is valid for the aggregation parameter α of the infection intensities which might be linked to the prevalence and the mean infection intensity. The simulations in Chapters 4, 5, and 7 showed how we linked the above three parameters by assuming a negative binomial distribution of worms in the population combined with a worm mating process. This produces a distribution of worm-pairs in the population that is described by the two parameters mean worm burden and worm aggregation but cannot be written as any known discrete distribution function. An important possible extension of the current work is to find appropriate approximations of the distribution for the parameter ranges occurring in the analysis of helminth data.

The choice of a negative binomial distribution for the distribution of worms in a population has been debated regarding what is to be considered a population in this case. It is likely to be a good approximation when either variations within one rather homogeneous population are continuous or when the population is large and from diverse settings as for example a complete district. Difficulty arises when the population consists of a few very distinct clusters where it is not recommended to apply the model to the full population but rather the clusters separately [Grafen and Woolhouse, 1993]. One of the main unknowns of the negative binomial model for the worm distribution is how the worm aggregation parameter varies with the mean worm burden in the population. It is largely clear that worms become more equally distributed in the population the higher the burden is because almost everybody is exposed [Truscott et al., 2019]. At low prevalence, only few people have been exposed but those might still harbour a considerable number of worms. We suggest including a term in the worm aggregation to estimate variation

with mean worm burden in the population. The ability to do this depends on the capacity to fit the above model assuming a negative binomial distribution for worms.

An assumption that can be further explored is related to the change in day-to-day and slide-to-slide variation with infection intensity. A lognormal day-to-day variation with mean 1 and variance parameter σ carries an implicit assumption about the changes in the variance with the infection intensity. We explored some alternatives based on the idea that the total infection intensity consists of a larger number of worms excreting eggs independently, which reduces the variance parameter with increasing infection intensity. The same is valid for the slide-to-slide variation. We suggest exploring more datasets with many repeats, in order to find the optimal functional form and improve on the assumption of constant variance. The egg count model assumes very high specificity with non-infected individuals test results modelled as a negative binomial with very low mean but high variance. In case where there is no mixing up and mislabelling of samples this is most likely a good assumption. However, it is not clear if this is always true and those kind of errors are difficult to model because they strongly depend on the organisation and laboratory procedures.

The alternative tests have consistently been analysed as binary in this thesis. Semi-quantitative data has been used by comparing different thresholds to translate into a binary result. As with all dichotomization to binary data, this process neglects some information present in the repeated measurements of the semi-quantitative tests. We suggest developing models for the semi-quantitative data directly including the latent infection intensity and the correlation between repeated measurements of the test for both positive and negative individuals. Additional information could then be extracted from the data and estimation of mean infection intensity in egg counts from the semi-quantitative tests could be feasible.

The age and sex of the individuals have been largely ignored. Sex has been included in the analysis of haematuria based diagnostic techniques but not for example for differences in disease prevalence, which could be due to varying exposure. We have modelled the relation between age and prevalence via a spline model in Chapter 2. However, there exist many transmission models, predicting age-specific prevalence and age-worm burden relations that could be incorporated in the statistical model. For school-aged children, an age group where prevalence and burden are

monotonically increasing, we suggest applying a simple immigration-death model to adjust for differences in age. When data of all age-groups is available more complicated, models including behavioural and immunity can be used [Chan et al., 1994; Yang, 2003].

For spatially referenced data available across many locations, the model can incorporate spatial correlation using a location-specific Gaussian spatial process on the mean worm burden. Smooth maps at high spatial resolution of the 'true' prevalence could then be created. As the model with the multitude of variance parameters is difficult to fit in itself, we suggest using approximations for the model fit like the nearest-neighbour Gaussian process model to exploit the fact that correlation decays fast for schistosomiasis [Datta et al., 2016].

Bibliography

Abdulamir, A. S., Hafidh, R. R., Kadhim, H. S., and Abubakar, F. (2009). Tumor markers of bladder cancer: The schistosomiasis bladder tumors. *J. Exp. Clin. Cancer Research*, 28:27.

Adriko, M., Standley, C. J., Tinkitina, B., Tukahebwa, E., Fenwick, A., Fleming, F. M., Sousa-Figueiredo, J. C., Stothard, J. R., and Kabatereine, N. B. (2014). Evaluation of circulating cathodic antigen (CCA) urine-cassette assay as a survey tool for schistosomiasis mansoni in different transmission settings within Bugiri District, Uganda. *Acta Trop.*, 136:50–57.

Amoah, B., Giorgi, E., and Diggle, P. J. (2018). A geostatistical framework for combining spatially referenced disease prevalence data from multiple diagnostics. *arXiv preprint*, page 1808.03141v1.

Anderson, R. M. (1986). The population dynamics and epidemiology of intestinal nematode infections. *Trans. R. Soc. Trop. Med. Hyg.*, 80:686–696.

Anderson, R. M. and May, R. M. (1991). *Infectious Diseases of Humans*. Oxford University Press.

Anderson, R. M. and Schad, G. A. (1985). Hookworm burdens and faecal egg counts: an analysis of the biological basis of variation. *Trans. R. Soc. Trop. Med. Hyg.*, 79:812–825.

Anderson, R. M., Turner, H. C., Farrell, S. H., Yang, J., and Truscott, J. E. (2015). What is required in terms of mass drug administration to interrupt the transmission of schistosome parasites in regions of endemic infection? *Parasit. Vectors*, 8:553.

Assare, R. K., Tra, M. B. I., Ouattara, M., Hürlimann, E., Coulibaly, J. T., N'Goran, E. K., and Utzinger, J. (2018). Sensitivity of the point-of-care circulating cathodic antigen urine cassette

test for diagnosis of *Schistosoma mansoni* in low-endemicity settings in Côte d'Ivoire. *Am. J. Trop. Med. Hyg.*, 99(6):1567–1572.

Bangert, M., Molyneux, D. H., Lindsay, S. W., Fitzpatrick, C., and Engels, D. (2017). The cross-cutting contribution of the end of neglected tropical diseases to the sustainable development goals. *Infectious Diseases of Poverty*, 6:73.

Becker, S. L., Lohourignon, L. K., Speich, B., Rinaldi, L., Knopp, S., N'Goran, E. K., Cringoli, G., and Utzinger, J. (2011). Comparison of the Flotac-400 dual technique and the Formalin-Ether concentration technique for diagnosis of human intestinal protozoon infection. *Journal of Clinical Microbiology*, 49:2183–2190.

Bethony, J., Brooker, S., Albonico, M., Geiger, S. M., Loukas, A., Diemert, D., and Hotez, P. J. (2006). Soil-transmitted helminth infections; ascariasis, trichuriasis, and hookworm. *The Lancet*, 367:1521–32.

Bichler, K. H., Savatovsky, I., Members of the Urinary Tract Infection (UTI) Working Group of the Guidelines Office of the European Association of Urology (EAU), Naber, K. G., Bischop, M. C., and Bjerklund-Johansen, T. E. (2006). EAU guidelines for the management of urogenital schistosomiasis. *Eur. Urol.*, 86:998–1003.

Booth, M., Vounatsou, P., N'Goran, E. K., Tanner, M., and Utzinger, J. (2003). The influence of sampling effort and the performance of the Kato-Katz technique in diagnosing *Schistosoma mansoni* and hookworm co-infections in rural Côte d'Ivoire. *Parasitology*, 127:525–531.

Botelho, M. C., Machado, J. C., and da Costa, J. M. (2010). *Schistosoma haematobium* and bladder cancer: what lies beneath? *Virulence*, 1:84–87.

Bottomley, C., Isham, V., Vivas-Martinez, S., Kuesel, A. C., Attah, S. K., Opoku, N. O., Lustigman, S., Walker, M., and Basañez, M.-G. (2016). Modelling neglected tropical diseases diagnostics: the sensitivity of skin snips for *Onchocerca volvulus* in near elimination and surveillance settings. *Parasit. Vectors*, 9:343.

Bärenbold, O., Garba, A., Colley, D. G., Fleming, F. M., Haggag, A. A., Ramzy, R. M. R., Assaré, R. K., Tukahebwa, E., Mbonigaba, B., Bucumi, V., Kebede, B., Yibi, M. S., Meite, A., Coulibaly, J. T., N'Goran, E. K., Tchuem Tchuenté, L.-A., Mwinzi, P., Utzinger, J., and Vounatsou, P. (2018). Translating preventive chemotherapy prevalence thresholds for *Schistosoma mansoni* from the Kato-Katz technique into the point-of-care circulating cathodic antigen diagnostic test. *PLoS Negl. Trop. Dis.*, 12:e0006941.

Bärenbold, O., Raso, G., Coulibaly, J. T., N'Goran, E. K., Utzinger, J., and Vounatsou, P. (2017). Estimating sensitivity of the Kato-Katz technique for the diagnosis of *Schistosoma mansoni* and hookworm in relation to infection intensity. *PLoS Negl. Trop. Dis.*, 11:e0005953.

Brooker, S., Kabatereine, N. B., Gyapong, J. O., Stothard, J. R., and Utzinger, J. (2009). Rapid mapping of schistosomiasis and other neglected tropical diseases in the context of integrated control programmes in Africa. *Parasitology*, 136:1707–1718.

Bundy, D. A. P., Hall, A., Medley, G. F., and Savioli, L. (1992). Evaluating measures to control intestinal parasitic infections. *World Health Stat.Q.*, 45:168–179.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (2016). Stan: a probabilistic programming language. *J. Stat. Softw.*

Chammartin, F., Hürlimann, E., Raso, G., N'Goran, K., E., Utzinger, J., and Vounatsou, P. (2013). Statistical methodological issues in mapping historical schistosomiasis survey data. *Acta Trop.*, 128:345–352.

Chan, M. S. (1997). The global burden of intestinal nematode infections - fifty years on. *Parasitol. Today*, 13(11):438 – 443.

Chan, M. S., Guyatt, H. L., Bundy, D. A. P., Booth, M., Fulford, A. J. C., and Medley, G. F. (1995). The development of an age structured model for schistosomiasis transmission dynamics and control and its validation for *Schistosoma mansoni*. *Epidemiol Infect*, 115:325–344.

Chan, M. S., Guyatt, H. L., Bundy, D. A. P., and Medley, G. F. (1994). The development and validation of an age-structured model for the evaluation of disease control strategies for intestinal helminths. *Parasitology*, 109:389–396.

Chan, M. S., Guyatt, H. L., Bundy, D. A. P., and Medley, G. F. (1996). Dynamic models of schistosomiasis morbidity. *Am. J. Trop. Med. Hyg.*, 55(1):52–62.

Cheever, E. A., W., C. A., Macedonia, J. G., and Mosimann, J. E. (1994). Kinetics of egg production and egg excretion by *Schistosoma mansoni* and *S. japonicum* in mice infected with a single pair of worms. *American Journal of Tropical Medicine and Hygiene*, 50:281–295.

Colley, D. G., Andros, T. S., and Campbell Jr., C. H. (2017). Schistosomiasis is more prevalent than previously thought: what does it mean for public health goals, policies, strategies, guidelines and intervention programs? *Infect Dis Poverty*, 6:63.

Colley, D. G., Binder, S., Campbell, C., King, C. H., Tchuem Tchuenté, L.-A., N'Goran, E. K., Erko, B., Karanja, M. S., Kabatereine, N. B., van Lieshout, L., and Rathbun, S. (2013). A five-country evaluation of a point-of-care circulating cathodic antigen urine assay for the prevalence of *Schistosoma mansoni*. *Am. J. Trop. Med. Hyg.*, 88(3):426–432.

Colley, D. G., Bustinduy, A. L., Secor, W. E., and King, C. H. (2014). Human schistosomiasis. *Lancet*, 383:2253–2264.

Coulibaly, J. T., Knopp, S., N'Guessan, N. A., Silue, K. D., Fürst, T., Lohourignon, L. K., Brou, J. K., Y.-K., N., Vounatsou, P., N'Goran, E. K., and Utzinger, J. (2011). Accuracy of urine circulating cathodic antigen (CCA) test for *Schistosoma mansoni* diagnosis in different settings of Côte d'Ivoire. *PLoS Negl. Trop. Dis.*, 5:e1384.

Coulibaly, J. T., N'Gbesso, Y.-K., N'Guessan, N. A., Winkler, M. S., Utzinger, J., and N'Goran, E. K. (2013a). Epidemiology of schistosomiasis in two high-risk communities of south Côte d'Ivoire with particular emphasis on pre-school – aged children. *Am. J. Trop. Med. Hyg.*, 89:32–41.

Coulibaly, J. T., Y.-K., N., Knopp, S., N'Guessan, N. A., Silue, K. D., van Dam, G., N'Goran, E. K., and Utzinger, J. (2013b). Accuracy of urine circulating cathodic antigen test for the diagnosis of *Schistosoma mansoni* in preschool-aged children before and after treatment. *PLoS Negl. Trop. Dis.*, 7(3):e2109.

Crainiceanu, C. M., Ruppert, D., and Wand, M. P. (2005). Bayesian analysis for penalized spline regression using WinBugs. *J. Stat. Softw.*, 14:1–24.

Danso-Appiah, A., Minton, J., Boamah, D., Otchere, J., Asmah, R. H., Rodgers, M., Bosompem, K. M., Eusebi, P., and de Vlas, S. J. (2016). Accuracy of point-of-care testing for circulatory cathodic antigen in the detection of schistosome infection: systematic review and meta-analysis. *Bull. World Health Organ.*, 94:522–533.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111:514:800–812.

de Dood, C. J., Hoekstra, P. T., Mngara, J., Kalluvya, S. E., van Dam, G., Downs, J. A., and Corstjens, P. L. A. M. (2018). Refining diagnosis of *Schistosoma haematobium* infections: antigen and antibody detection in urine. *Front Immunol.*, 9:2635.

de Vlas, S. J., Engels, D., Rabello, A., Oostburg, B. F. J., van Lieshout, L., Polderman, A. M., van Oortmarssen, G. J., Habbema, J. D. F., and Gryseels, B. (1997). Validation of a chart to estimate true *Schistosoma mansoni* prevalences from simple egg counts. *Parasitology*, 114:113–121.

de Vlas, S. J. and Gryseels, B. (1992). Underestimation of *Schistosoma mansoni* prevalences. *Parasitol. Today*, 8:274–277.

de Vlas, S. J., Gryseels, B., van Oortmarssen, G. J., Polderman, A. M., and Habbema, J. D. F. (1992). A model for variations in single and repeated egg counts in *Schistosoma mansoni* infections. *Parasitology*, 104:451–460.

de Vlas, S. J., Nagelkerke, N. J. D., Habbema, J. D. F., and van Oortmarssen, G. J. (1993). Review papers: Statistical models for estimating prevalence and incidence of parasitic diseases. *Statistical Methods in Medical Research*, 2:3–21.

Dendukuri, N. and Joseph, L. (2001). Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics*, 57:158–167.

Despommier, D. D., Gawadz, R. W., and Hotez, P. J. (2012). *Parasitic diseases*. Springer Science & Business Media.

Doehering, E., Feldmeier, H., and Daffalla, A. A. (1983). Day to day variation and circardian rhythm of egg excetion in urinary schistosomiasis in the Sudan. *Ann. Trop. Med. Parasitol.*, 77:587–594.

Dorkenoo, A. M., Bronzan, R. N., Ayena, K. D., Anthony, G., Agbo, Y. M., Sognikin, K. S. E., Dogbe, K. S., Amza, A., Sodahlon, Y., and Mathieu, E. (2012). Nationwide integrated mapping of three neglected tropical diseases in Togo: countrywide implementation of a novel approach. *Tropical Medicine and International Health*, 17:896–903.

Engels, D., Sinzinkayo, E., and Griyseels, B. (1996). Day-to-day egg count fluctuation in *Schistosoma mansoni* infection and its operational implications. *Am. J. Trop. Med. Hyg.*, 54:319–324.

Enk, M. J., Lima, A. C. L., Drummond, S. C., Schall, V. T., and Coelho, P. M. Z. (2008). The effect of the number of stool samples on the observed prevalence and the infection intensity with *Schistosoma mansoni* among a population in an area of low transmission. *Acta Trop.*, 108:222–228.

Enk, M. J., Oliveira e Silva, G., and Rodrigues, N. B. (2012). Diagnostic accuracy and applicability of a PCR system for the detection of *Schistosoma mansoni* DNA in human urine samples from an endemic area. *PLoS ONE*, 7:e38947.

Erko, B., Medhin, G., Teklehaymanot, T., Degarege, A., and Legesse, M. (2013). Evaluation of urine-circulating cathodic antigen (urine-CCA) cassette test for the detection of *Schistosoma mansoni* infection in areas of moderate prevalence in Ethiopia. *Trop. Med. Int. Health*, 18(8):1029 – 1035.

Ezeamama, A. E., Bustinduy, A. L., Nkwata, A. K., Martinez, L., Pabalan, N., Boivin, M. J., and King, C. H. (2018). Cognitive deficits and educational loss in children with schistosome infection – a systematic review and meta-analysis. *PLoS Negl. Trop. Dis.*, 12:e0005524.

Fatiregun, A. A., Osungbade, K. O., and Olumide, E. A. (2005). Diagnostic performance of screening methods for urinary schistosomiasis in a school-based control programme, in Ibadan, Nigeria. *J Community Med Prim Health Care.*, 17:24–27.

GBD 2016 Causes of Death Collaborators (2017). Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, 390:1151–1210.

GBD 2016 DALYs and HALE Collaborators (2017). Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, 390:1260–1344.

GBD 2016 Disease and Injury Incidence and Prevalence Collaborators (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, 390:1211–1259.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Stat Sci*, 7(4):457–511.

Giardina, F., Coffeng, L. E., Farrell, S. H., Vegvari, C., Werkman, M., Truscott, J., Anderson, R., and de Vlas, S. J. (2019). Sampling strategies for monitoring and evaluation of morbidity targets for soil-transmitted helminths. *PLoS Negl. Trop. Dis.*, 13(6):e0007514.

Giorgi, E., Sesay, S. S. S., Terlouw, D. J., and Diggle, P. J. (2015). Combining data from mulitple spatially referenced prevalence surveys using generalized linear geostatistical models. *J. R. Statist. Soc. A.*, 178:445–464.

Glinz, D., Silué, K. D., Knopp, S., Lohourignon, L. K., Yao, P. K., Steinmann, P., Rinaldi, L., Cringoli, G., N'Goran, and Utzinger, J. (2010). Comparing diagnostic accuracy of Kato-Katz, Koga agar plate, ether-concentration, and FLOTAC for *Schistosoma mansoni* and soil-transmitted helminths. *PLoS Negl. Trop. Dis.*, 4(7):e754.

Grafen, A. and Woolhouse, M. E. J. (1993). Does the negative binomial distribution add up? *Parasitol. Today*, 9:475–477.

Gregory, R. D. and Woolhouse, M. E. J. (1993). Quantification of parasite aggregation: a simulation study. *Acta Trop.*, 54:131–139.

Greter, H., Krauth, S. J., Ngandolo, B. N. R., Alfaroukh, I. O., Zinsstag, J., and Utzinger, J. (2016). Validation of a point-of-care circulating cathodic antigen urine cassette test for *Schistosoma mansoni* diagnosis in the Sahel, and potential cross-reaction in pregnancy. *Am. J. Trop. Med. Hyg.*, 94(2):361–364.

Gurarie, D. and King, C. H. (2014). Population biology of *Schistosoma* mating, aggregation and transmission breakpoints: More reliable model analysis for the end-game in communities at risk. *PLoS ONE*, 9(12):e115875.

Haggag, A. A., Partal, M. C., Rabiee, A., Elaziz, K. M. A., Campbell Jr., C. H., Colley, D. G., and Ramzy, R. M. R. (2019a). Multiple praziquantel treatments of *Schistosoma mansoni* egg-negative, CCA-positive schoolchildren in a very low endemic setting in Egypt do not consistently alter CCA results. *Am. J. Trop. Med. Hyg.*, 100(6):1507–1511.

Haggag, A. A., Rabiee, A., Abd Elaziz, K. M., Gabrielli, A. F., Abdel Hay, R., and Ramzy, R. M. R. (2017). Mapping of *Schistosoma mansoni* in the Nile Delta, Egypt: Assessment of the prevalence by the circulating cathodic antigen urine assay. *Acta Trop.*, 167:9–17.

Haggag, A. A., Rabiee, A., Elaziz, K. M. A., Campbell Jr., C. H., Colley, D. G., and Ramzy, R. M. R. (2019b). Thirty-day daily comparisons of Kato-Katz and CCA assays of 45 Egyptian children in areas with very low prevalence of *Schistosoma mansoni*. *Am. J. Trop. Med. Hyg.*, 100(3):578–583.

Hall, A. (1981). Quantitative variability of nematode egg counts in faeces: a study among rural Kenyans. *Trans. R. Soc. Trop. Med. Hyg.*, 75(5):682–687.

Hatz, C., Vennervald, B. J., Nkulila, T., Vounatsou, P., Kombe, Y., Mayombana, C., Mshinda, H., and Tanner, M. (1998). Evolution of *Schistosoma haematobium*-related pathology over 24 months after treatment with praziquantel among school children in southeastern Tanzania. *Am. J. Trop. Med. Hyg.*, 59:775–781.

Hay, S. I., Battle, K. E., Pigott, D. M., Smith, D. L., Moyes, C. L., Bhatt, S., Brownstein, J. S., Collier, N., Myers, M. F., George, D. B., and Gething, P. W. (2013). Global mapping of infectious disease. *Philosophical Transactions of the Royal Society, B*, 368:20120250.

Holford, T. R. and Hardy, R. J. (1975). A stochastic model for the analysis of age-specific prevalence curves in schistosomiasis. *J. Chronic Dis.*, 29:445–458.

Hotez, P. J., Alvarado, M., Basañez, M.-G., Bolliger, I., Bourne, R., Boussinesq, M., Brooker, S., Brown, A. S., Buckle, G., Budke, C. M., Carabin, H., Coffeng, L. E., Fevre, E. M., Fürst, T., Halasa, Y. A., Jasrasaria, R., Johns, N. E., Keiser, J., King, C. H., Lozano, R., Murdoch, M. E., O'Hanlon, S., Pion, S. D. S., Pullan, R. L., Ramaiah, K. D., Roberts, T., Shepard, D., Smith, J. L., Stolk, W. A., Undurraga, E. A., Utzinger, J., Wang, M., Murray, C. J. L., and Naghavi, M. (2014). The Global Burden of Disease Study 2010: Interpretation and implications for the Neglected Tropical Diseases. *PLoS Negl. Trop. Dis.*, 8(7):e2865.

Hotez, P. J., Molyneux, D., Fenwick, A., Kumaresan, J., Ehrlich Sachs, S., Sachs, J. D., and Savioli, L. (2007). Control of neglected tropical diseases. *N. Engl. J. Med.*, 357:10:1018–1027.

Hürlimann, E., Schur, N., Boutsika, K., Stensgaard, A.-S., de Himpsl, M. L., Ziegelbauer, K., Laizer, N., Camenzind, L., Di Pasquale, A., Ekpo, U. F., Simoonga, C., Mushinge, G., Saarnak, C. F. L., Utzinger, J., Kristensen, T. K., and Vounatsou, P. (2011). Toward an open-access global database for mapping, control, and surveillance of Neglected Tropical Diseases. *PLoS Negl. Trop. Dis.*, 5(12):e1404.

Ibironke, O., Koukounari, A., Asaolu, S., Moustaki, I., and Shiff, C. (2012). Validation of a new test for *Schistosoma haematobium* based on detection of Dra1 DNA fragments in urine: evaluation through latent class analysis. *PLoS Negl. Trop. Dis.*, 6:e1464.

Kahama, A. I., Odek, A. E., Kihara, R. W., Vennervald, B. J., Kombe, Y., Nkulila, T., Hatz, C., Ouma, J. H., and Deelder, A. M. (1999). Urine circulating soluble egg antigen in relation to egg counts, hematuria, and urinary tract pathology before and after treatment in children infected with *Schistosoma haematobium* in Kenya. *Am. J. Trop. Med. Hyg.*, 61:215–219.

Karagiannis-Voules, D.-A., Biedermann, P., Ekpo, U. F., Garba, A., Langer, E., Mathieu, E., Midzi, N., Mwinzi, P., Polderman, A. M., Raso, G., Sacko, M., Talla, I., Tchuem Tchuente, L.-A., Touré, S., Winkler, M. S., Utzinger, J., and Vounatsou, P. (2015). Spatial and temporal distribution of soil-transmitted helminth infection in sub-Saharan Africa: a systematic review and geostatistical meta-analysis. *Lancet Infect. Dis.*, 15:74–84.

Kato, K. and Miura, M. (1954). Comparative examinations. *Japanese Journal of Parasitology*, 3(5).

Katz, N., Chaves, A., and Pellegrino, J. (1972). A simple device for quantitative stool thick-smear technique in schistosomiasis mansoni. *Rev. Inst. Med. Trop. São Paulo*, 14:397–400.

King, C. H. (2007). Lifting the burden of schistosomiasis–defining elements of infection-associated disease and the benefits of antiparasite treatment. *The Journal of Infectious Diseases*, 196:653–655.

King, C. H. (2010). Parasites and poverty: the case of schistosomiasis. *Acta Trop.*, 113:95–104.

King, C. H. (2015). It's time to dispel the myth of "asymptomatic" schistosomiasis. *PLoS Negl. Trop. Dis.*, 9:e0003504.

King, C. H. and Dangerfield-Cha, M. (2008). The unacknowledged impact of chronic schistosomiasis. *Chronic Illness*, 4:65–79.

King, C. H., Dickman, K., and Tisch, D. J. (2005). Reassessment of the cost of chronic helmintic infection: a meta-analysis of disability-related outcomes in endemic schistosomiasis. *The Lancet*, 365:1561–1569.

King, C. H., Olbrych, S. K., M., S., Singer, M. E., Carter, J., and Colley, D. G. (2011). Utility of repeated praziquantel dosing in the treatment of schistosomiasis in high-risk communities in Africa: A systematic review. *PLoS Negl. Trop. Dis.*, 5(9):e1321.

Kittur, N., Castleman, J. D., Campbell Jr., C. H., King, C. H., and Colley, D. G. (2016). Comparison of *Schistosoma mansoni* prevalence and intensity of infection, as determined by the circulating cathodic antigen urine assay or by the Kato-Katz fecal assay: A systematic review. *Am. J. Trop. Med. Hyg.*, 94(3):605–610.

Kjetland, E. F., Hegertun, I. E., Baay, M. F., Onsrud, M., Ndhlovu, P. D., and Taylor, M. (2014). Genital schistosomiasis and its unacknowledged role on HIV transmission in the STD intervention studies. *Int. J. STD AIDS*, 25(10):1060–1064.

Knopp, S., Ame, S. M., Hattendorf, J., Ali, S. M., Khamis, I. S., Bakar, F., Khamis, I. S., Person, B., Kabole, F., and Rollinson, D. (2018). Urogenital schistosomiasis elimination in Zanzibar: accuracy of urine filtration and haematuria reagent strips for diagnosing light intensity *Schistosoma haematobium* infections. *Parasit. Vectors*, 11:552.

Knopp, S., Corstjens, P. L. A. M., Koukounari, A., Cercamondi, C. I., Ame, S. M., Ali, S. M., de Dood, C. J., Mohammed, K. A., Utzinger, J., Rollinson, D., and van Dam, G. (2015). Sensitivity and specificity of a urine circulating anodic antigen test for the diagnosis of *Schistosoma haematobium* in low endemic settings. *PLoS Negl. Trop. Dis.*, 9(5):e0003752.

Knopp, S., Mohammed, A. S., Stothard, J. R., Khamis, I. S., Rollinson, D., Marti, H., and Utzinger, J. (2010). Patterns and risk factors of helminthiasis and anemia in a rural an a peri-urban community in Zanzibar, in the context of helminth control programs. *PLoS Negl. Trop. Dis.*, 4(5):e681.

Knowles, S. C. L., Sturrock, H. J. W., Turner, H. C., Whitton, J. M., Gower, C. M., Jemu, S., Phillips, A. E., Meite, A., Thomas, B., Kollie, K., Thomas, C., Rebollo, M. P., STyles, B., Clements, M., Fenwick, A., Harrison, W. E., and Fleming, F. M. (2017). Optimising cluster survey design for planning schistosomiasis preventive chemotherapy. *PLoS Negl. Trop. Dis.*, 11:e0005599.

Kongs, A., Marks, G., Verlé, P., and Van der Stuyft, P. (2001). The unreliability of the Kato-Katz technique limits its usefulness for evaluating *S. mansoni* infections. *Trop. Med. Int. Health*, 6(3):163–169.

Kosinski, K. C., Bosompem, K., Stadecker, M. J., Wagner, A. D., Plummer, J., Durant, J. L., and Gute, D. M. (2011). Diagnostic accuracy of urine filtration and dipstick tests for *Schistosoma haematobium* infection in a lightly infected population of Ghanaian schoolchildren. *Acta Trop.*, 118:123–127.

Koukounari, A., Donnelly, C. A., Moustaki, I., Tukahebwa, E., Kabatereine, N. B., Wilson, S., Webster, J. P., Deelder, A. M., Vennervald, B. J., and van Dam, G. (2013). A latent Markov modelling approach to the evaluation of circulating cathodic antigen strips for schistosomiasis diagnosis pre- and post-praziquantel treatment in Uganda. *PLoS Comput Biol.*, 9:e1003402.

Koukounari, A., Webster, J. P., Donnelly, C. A., Bray, B. C., Naples, J., Bosompem, K., and Shiff, C. (2009). Sensitivities and specificities of diagnostic tests and infection prevalence of *Schistosoma haematobium* estimated from data on adults in villages northwest of Accra, Ghana. *Am. J. Trop. Med. Hyg.*, 80(3):435–441.

Krauth, S. J., Coulibaly, J. T., Knopp, S., Traoré, M., N'Goran, E. K., and Utzinger, J. (2012). An in-depth analysis of a piece of shit: distribution of *Schistosoma mansoni* and hookworm eggs in human stool. *PLoS Negl. Trop. Dis.*, 6(12):e1969.

Krauth, S. J., Greter, H., Stete, K., Coulibaly, J. T., Traoré, S. I., Ngandolo, B. N. R., Achi, L. Y., Zinsstag, J., N'Goran, E. K., and Utzinger, J. (2015). All that is blood is not schistosomiasis: experiences with reagent strip testing for urogenital schistosomiasis with special consideration to very-low prevalence settings. *Parasit. Vectors*, 8:584.

Kure, A., Mekonnen, Z., Dana, D., Bajiro, M., Ayana, M., Vercruysse, J., and Levecke, B. (2015). Comparison of individual and pooled stool samples for the assessment of intensity of *Schistosoma mansoni* and soil-transmitted helminth infections using the Kato-Katz technique. *Parasit. Vectors*, 8:1–9.

Lai, Y.-S., Biedermann, P., Ekpo, U. F., Garba, A., Mathieu, E., Midzi, N., Mwinzi, P., N'Goran, E. K., Raso, G., Assare, R. K., Sacko, M., Schur, N., Talla, I., Tchuem Tchuente, L.-A., Touré, S., Winkler, M. S., Utzinger, J., and Vounatsou, P. (2015). Spatial distribution of schistosomiasis and treatment needs in sub-Saharan Africa: a systematic review and geostatistical analysis. *Lancet Infect. Dis.*, 15:927–940.

Lamberton, P. H. L., Kabatereine, N. B., Oguttu, D. W., Fenwick, A., and Webster, J. P. (2014). Sensitivity and specificity of multiple Kato-Katz thick smears and a circulating cathodic antigen test for *Schistosoma mansoni* diagnosis pre- and post-repeated-Praziquantel treatment. *PLoS Negl. Trop. Dis.*, 8(9):e3139.

Lengeler, C., Makwala, J., Ngimbi, D., and Utzinger, J. (2000). Simple school questionnaire can map both *Schistosoma mansoni* and *Schistosoma haematobium* in the Democratic Republic of Congo. *Acta Trop.*, 74:77–87.

Lengeler, C., Utzinger, J., and Tanner, M. (2002). Questionnaires for rapid screening of schistosomiasis in sub-Saharan Africa. *Bull. World Health Organ.*, 80:235–242.

Levecke, B., Anderson, R. M., Berkvens, D., Charlier, J., Devleesschauwer, B., Speybroeck, N., Vercruysse, J., and Van Aelst, S. (2015). Mathematical inference on helminth egg counts in stool and its applications in mass drug administration programmes to control soil-transmitted helminthiasis in public health. *Adv. Parasitology*, 87:193–247.

Levecke, B., Brooker, S. J., Knopp, S., Steinmann, P., Sousa-Figueiredo, J. C., Stothard, J. R., Utzinger, J., and Vercruysse, J. (2014). Effect of sampling and diagnostic effort on the assessment of schistosomiasis and soil-transmitted helminthiasis and drug efficacy: a meta-analysis of six drug efficacy trials and one epidemiological survey. *Parasitology*, 141(14):1826–1840.

Lo, N. C., Addiss, D. G., Hotez, P. J., King, C. H., Stothard, J. R., Evans, D. S., Colley, D. G., Lin, W., Coulibaly, J. T., Bustinduy, A. L., Raso, G., Bendavid, E., Bogoch, I. I., Fenwick, A., Savioli, L., Molyneux, D., Utzinger, J., and Andrews, J. R. (2017). A call to strengthen the global strategy against schistosomiasis and soil-transmitted helminthiasis: the time is now. *Lancet Infect. Dis.*, 17:e64–69.

Mafe, M. A. (1997). The diagnostic potential of three indirect tests for urinary schistosomiasis in Nigeria. *Acta Trop.*, 68:277–284.

May, R. M. (1977). Togetherness among schistosomes: its effects on the dynamics of the infection. *Math Biosci*, 35:301–434.

May, R. M. and Woolhouse, M. E. J. (1993). Biased sex ratios and parasite mating probabilities. *Parasitology*, 107:287–295.

McManus, D. P., Dunne, D. W., Sacko, M., Utzinger, J., Vennervald, B. J., and Zhou, X.-N. (2018). Schistosomiasis. *Nat Rev Dis Primers*, 4:13.

Menten, J., Boelaert, M., and Lesaffre, E. (2008). Bayesian latent class models with conditionally dependent diagnostic tests: A case study. *Statistics in Medicine*, 27:4469–4488.

Midzi, N., Butterworth, A. E., Mduluza, T., Munyati, S. M., Deelder, A. M., and van Dam, G. (2009). Use of circulating cathodic antigen strips for the diagnosis of urinary schistosomiasis. *Trans. R. Soc. Trop. Med. Hyg.*, 103:45–51.

Midzi, N., Mduluza, T., Chimbari, M. J., Tshuma, C., Charimari, L., Mhlanga, G., Manangazira, P., Munyati, S. M., Phiri, I., Mutambu, S. L., Midzi, S. S., Ncube, A., Muranzi, L. P., Rusakaniko, S., and Mutapi, F. (2014). Distribution of schistosomiasis and soil transmitted helminthiasis in Zimbabwe: towards a national plan of action for control and elimination. *PLoS Negl. Trop. Dis.*, 8:e3014.

Molyneux, D. (2004). "Neglected" diseases but unrecognised successes - challenges and opportunities for infectious disease control. *Lancet*, 364:380–383.

Molyneux, D., Savioli, L., and Engels, D. (2017). Neglected tropical diseases: progress towards addressing the chronic pandemic. *Lancet*, 389:312–325.

Montresor, A., Crompton, D. W. T., Gyorkos, T. W., and Savioli, L. (2002). Helminths control in schoolchildren. a guide for managers of control. Technical report, World Health Organization.

Montresor, A., Crompton, D. W. T., Hall, A., Bundy, D. A. P., and Savioli, L. (1998). Guidelines for the evaluation of soil-transmitted helminthiasis and schistosomiasis at community level. Technical report, World Health Organization.

Morenikeji, L., Quazim, J., Omoregie, C., Hassan, A., Nwuba, R., Anumudu, C., Adejuwon, S., Salawu, O., Jegede, A., and Odaibo, A. (2014). A cross-sectional study on urogenital schistosomiasis in children; haematuria and proteinuria as diagnostic indicators in an endemic rural area of Nigeria. *Afr Health Sci*, 14:390–396.

Moser, W., Bärenbold, O., Mirams, G. J., Cools, P., Vlaminck, J., Ali, S. M., Ame, S. M., Hattendorf, J., Vounatsou, P., Levecke, B., and Keiser, J. (2018). Diagnostic comparison between FECPAK-G2 and the Kato-Katz method for analyzing soil-transmitted helminth eggs in stool. *PLoS Negl. Trop. Dis.*, 12(6):e0006562.

Mott, K. E., Dixon, H., Osei-Tutu, E., England, E. C., Ekue, K., and Tekle, A. (1985). Evaluation of reagent strips in urine tests for detection of *Schistosoma haematobium* infection: a comparative study in Ghana and Zambia. *Bull. World Health Organ.*, 63(1):125–133.

Mwinzi, P. N. M., Kittur, N., Ochola, E., Cooper, P. J., Campbell Jr., C. H., King, C. H., and Colley, D. G. (2015). Additional evaluation of the point-of-contact circulating cathodic antigen assay for *Schistosoma mansoni* infection. *Front. Public Health*, 3:48.

Nagelkerke, N. J. D., Borgdorff, M. W., Kalisvaart, N. A., and Broekmans, J. F. (2000). The design of multistage tuberculin surveys: some suggestions for sampling. *Int. J. Tuberc. Lung Dis.*, 4(4):314–320.

Ndhlovu, P., Chimbari, M., Ndamba, J., and Chandiwana, S. K. (1992). National schistosomiasis survey. Technical report, Ministry of Health and Child Welfare, Harare, Zimbabwe.

Nikolay, B., Brooker, S. J., and Pullan, R. L. (2014). Sensitivity of diagnostic tests for human soil-transmitted helminth infections: a meta-analysis in the absence of a true gold standard. *Int. J. Parasitol.*, 44:765–774.

Obeng, B. B., Aryeetey, Y. A., de Dood, C. J., Amoah, A. S., Larbi, I. A., Deelder, A. M., Yazdanbakhsh, M., Hargers, F. C., Boakye, D. A., Verweij, J. J., van Dam, G., and van Lieshout, L. (2008). Application of a circulating-cathodic-antigen (CCA) strip test and real-time PCR, in comparison with microscopy, for the detection of *Schistosoma haematobium* in urine samples from Ghana. *Ann. Trop. Med. Parasitol.*, 102(7):625–633.

Ochodo, E. A., Gopalakrishna, G., Spek, B., Reitsma, J. B., van Lieshout, L., Polman, K., Lamberton, P., Bossuyt, P., and Leeflang, M. (2015). Circulating antigen tests and urine reagent strips for diagnosis of active schistosomiasis in endemic areas. *Cochrane Database of Systematic Reviews*, Issue 3:CD009579.

Ortu, G., Ndayishimiye, O., Clements, M., Kayugi, D., Campbell Jr., C. H., Lamine, M. S., Zivieri, A., Magalhaes, F. S., Binder, S., King, C. H., Fenwick, A., Colley, D. G., and Jourdan, P. M. (2017). Countrywide reassessment of *Schistosoma mansoni* infection in Burundi using a urine-circulating cathodic antigen rapid test: Informing the national control program. *Am. J. Trop. Med. Hyg.*, 96:664–673.

Peralta, J. M. and Cavalcanti, M. G. (2018). Is POC-CCA a truly reliable test for schistosomiasis diagnosis in low endemic areas? the trace results controversy. *PLoS Negl. Trop. Dis.*, 12:e0006813.

Prada, J. M., Touloupou, P., Adriko, M., Tukahebwa, E., Lamberton, P., and Hollingsworth, T. D. (2018). Understanding the relationship between egg- and antigen-based diagnostics of *Schistosoma mansoni* infection pre- and post-treatment in Uganda. *Parasit. Vectors*, 11:21.

Pullan, R. L., Smith, J. L., Jasrasaria, R., and Brooker, S. J. (2014). Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. *Parasit. Vectors*, 7:37.

Qu, Y., Tan, M., and Kutner, M. H. (1996). Effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, 52:797–810.

Raso, G., N'Goran, E. K., Toty, A., Luginbühl, A., Adjoua, C. A., Tian-Bi, Y.-N. T., Bogoch, I. I., Vounatsou, P., Tanner, M., and Utzinger, J. (2004). Efficacy and side effects of Praziquantel against *Schistosoma mansoni* in a community of western Côte d'Ivoire. *Trans. R. Soc. Trop. Med. Hyg.*, 98:18–27.

Raso, G., Vounatsou, P., McManus, D. P., K., N. E., and Utzinger, J. (2007). A Bayesian approach to estimate the age-specific prevalence of *Schistosoma mansoni* and implications for schistosomiasis control. *Int. J. Parasitol.*, 37:1491–1500.

Rollinson, D., Knopp, S., Levitz, S., Stothard, J. R., Tchuem Tchuente, L.-A., Garba, A., Mohammed, A. S., Schur, N., Person, B., Colley, D. G., and Utzinger, J. (2013). Time to set the agenda for schistosomiasis elimination. *Acta Trop.*, 128:423–440.

Secor, W. E. (2019). Toward mass drug adminstration stopping criteria for *Schistosoma mansoni* control programs. *Am. J. Trop. Med. Hyg.*, 100:485–486.

Shane, H. L., Verani, J. R., Abudho, B., Montgomery, S. P., Blackstock, A. J., Mwinzi, P., Butler, S. E., Karanja, D. M. S., and Secor, W. E. (2011). Evaluation of urine CCA assays for detection of *Schistosoma mansoni* infection in western Kenya. *PLoS Negl. Trop. Dis.*, 5(1):e951.

Shiff, C. (2015). Accurate diagnostics for schistosomiasis: a new role for PCR. *Rep. Parasitol.*, 4:23–29.

Stete, K., Krauth, S. J., Coulibaly, J. T., Knopp, S., Hattendorf, J., Müller, I., Lohourignon, L. K., Kern, W. V., N'Goran, E. K., and Utzinger, J. (2012). Dynamics of *Schistosoma haematobium* egg output and associated infection parameters following treatment with praziquantel in school-aged children. *Parasit. Vectors*, 5:298.

Stoll, N. R. (1947). This wormy world. *J. Parasitol.*, 33:1–18.

Stothard, J. R., Stanton, M. C., Bustinduy, A. L., Sousa-Figueiredo, J. C., van Dam, G. J., Betson, M., Waterhouse, D., Ward, S., Allan, F., Hassan, A. A., Al-Helal, M. A., Memish, Z. A., and Rollinson, D. (2014). Diagnostics for schistosomiasis in Africa and Arabia: a review of present options in control and future needs for elimination. *Parasitology*, 141:1947–1961.

Sturrock, H. J. W., Gething, P. W., Ashton, R. A., Kolaczinski, J. H., Kabatereine, N. B., and Brooker, S. (2011). Planning schistosomiasis control: investigation of alternative sampling strategies for *Schistosoma mansoni* to target mass drug administration of praziquantel in East Africa. *International Health*, 3:165–175.

Sturrock, H. J. W., Gething, P. W., Clements, A. C. A., and Brooker, S. (2010). Optimal survey designs for targeting chemotherapy against soil-transmitted helminths: Effect of spatial heterogeneity and cost-efficiency of sampling. *Am. J. Trop. Med. Hyg.*, 82(6):1079–1087.

Tarafder, M. R., Carabin, H., Joseph, L., Balolong, E. J., Olveda, R., and McGarvey, S. T. (2010). Estimating the sensitivity and specificity of Kato-Katz stool examination technique for detection of hookworms, *Ascaris lumbricoides* and *Trichuris trichiura* infections in humans in the absence of a 'gold standard'. *Int. J. Parasitol.*, 40:399–404.

Tchuem Tchuenté, L.-A., Fouodo, C. J. K., Ngassam, R. I. K., Sumo, L., Noumedem, C. D., Kenfack, C. M., Gipwe, N. F., Nana, E. D., Stothard, J. R., and Rollinson, D. (2012). Evaluation of circulating cathodic antigen (CCA) urine-tests for diagnosis of *Schistosoma mansoni* infection in Cameroon. *PLoS Negl Trop Dis*, 6(7):e1758.

Tchuem Tchuente, L.-A., Stothard, J. R., Rollinson, D., and Reinhard-Rupp, J. (2018). Precision mapping: An innovative tool and way forward to shrink the map, better target interventions, and accelerate toward the elimination of schistosomiasis. *PLoS Negl. Trop. Dis.*, 12:e0006563.

TDR Disease Reference Group on Helminth infections (2012). Research priorities for helminth infections. Technical report, World Health Organization.

Toor, J., Truscott, J. E., Werkman, M., Turner, H. C., Phillips, A. E., King, C. H., Medley, G. F., and Anderson, R. (2019). Determining post-treatment surveillance criteria for predicting the elimination of *Schistosoma mansoni* transmission. *Parasit. Vectors*, 12:437.

Truscott, J., Hollingsworth, T. D., Brooker, S., and Anderson, R. (2014). Can chemotherapy alone eliminate the transmission of soil transmitted helminths? *Parasit. Vectors*, 7:266.

Truscott, J., Ower, A., Werkman, M., Halliday, K., Oswald, W. E., Gichuki, P. M., Mcharo, C., Brooker, S., Njenga, S. M., Mwandariwo, C., Walson, J., Pullan, R. L., and Anderson, R. (2019). Heterogeneity in transmission parameters of hookworm infection within the baseline data from the TUMKIA study in Kenya. *Parasit. Vectors*, 12:442.

Turner, H. C., Bettis, A. A., Dunn, J. C., Whitton, J. M., Hollingsworth, T. D., Fleming, F. M., and Anderson, R. (2017). Economic consierations for moving beyond the Kato-Katz technique for diagnosing intestinal parasites as we move towards elimination. *Trends in Parasitology*, 33:435–443.

Utzinger, J., Becker, S. L., van Lieshout, L., van Dam, G., and Knopp, S. (2015). New diagnostic tools in schistosomiasis. *Clin. Microbiol. Infect.*, 21:529–542.

Utzinger, J., Booth, M., N'Goran, E. K., Müller, I., Tanner, M., and Lengeler, C. (2001). Relative contribution of day-to-day and intra-specimen variation in faecal egg counts of *Schistosoma mansoni* before and after treatment with praziquantel. *Parasitology*, 122:537–544.

Utzinger, J., N'Goran, E. K., N'Dri, A., Lengeler, C., and Tanner, M. (2000a). Efficacy of praziquantel against *Schistosoma mansoni* with particular consideration for intensity of infection. *Trop. Med. Int. Health*, 5:771–778.

Utzinger, J., N'Goran, E. K., N'Dri, A., Lengeler, C., Xiao, S., and Tanner, M. (2000b). Oral artemether for prevention of *Schistosoma mansoni* infection: randomised controlled trial. *The Lancet*, 355:1320–1325.

van der Werf, M. J., de Vlas, S. J., Brooker, S., Looman, C. W. N., Nagelkerke, N. J. D., Habbema, J. D. F., and Engels, D. (2003). Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa. *Acta Trop.*, 86:125–139.

van Lieshout, L., Polderman, A. M., and Deelder, A. M. (2000). Immunodiagnosis of schistosomiasis by determination of the circulating antigens CAA and CCA, in particular in indiciduals with recent or light infections. *Acta Trop.*, 77:69–80.

Vinkeles Melchers, N. V. S., van Dam, G., Shaproski, D., Kahama, A. I., Brienen, E. A. T., Vennervald, B. J., and van Lieshout, L. (2014). Diagnostic performance of *Schistosoma* real-time PCR in urine samples from Kenyan children infected with *Schistosoma haematobium*: Day-to-day variation and follow-up after praziquantel treatment. *PLoS Negl. Trop. Dis.*, 8:e2807.

Wang, C., Torgerson, P. R., and Höglund, J. (2017). Zero-inflated hierarchical models for faecal egg counts to assess antihelmintic efficacy. *Veterinary Parasitology*, 235:20 – 28.

Wang, Z., Dendukuri, N., and Joseph, L. (2016). Understanding the effects of conditional dependence in research studies involving imperfect diagnostic tests. *Statistics in Medicine*.

WHO (1991). Basic laboratory methods in medical parasitology. Technical report, World Health Organization.

WHO (2006). *Preventive chemotherapy in human helminthiasis: coordinated use of anthelminthic drugs in control interventions: a manual for health professionals and programme managers*. World Health Organization, Geneva.

WHO (2012). Soil-transmitted helminthiasis: eliminating soil-transmitted helminthiasis as a public healthproblem in children, progress report 2001–2010 and strategic plan 2011–2020. Technical report, World Health Organization.

WHO (2013). Schistosomiasis: Progress report 2001–2011 and strategic plan 2012–2020. Technical report, World Health Organization.

WHO (2016). Schistosomiasis and soil-transmitted helminthiases: number of people treated in 2015. *Wkly Epidemiol. Rec.*, 91:585–600.

WHO Expert Committee (2002). Prevention and control of schistosomiasis and soil-transmitted helminthiasis. *World Health Organization Technical Report*, 912.

Woolhouse, M. E. J. (1991). On the application of mathematical models of schistosome transmission dynamics. i. natural transmission. *Acta Trop.*, 49:241–270.

Woolhouse, M. E. J. (1992). On the application of mathematical models of schistosome transmission dynamics. ii. control. *Acta Trop.*, 50:189–204.

Woolhouse, M. E. J. (1998). Patterns in parasite epidemiology: the peak shift. *Parasitol. Today*, 14:428–434.

World Health Assembly (2001). Schistosomiasis: A54/vr/9. Technical report, World Health Organization, Geneva.

Worrell, C. M., Bartoces, M., Karanja, D. M. S., Ochola, E., Matete, D., Mwinzi, P., Montgomery, S. P., and Secor, W. E. (2015). Cost analysis of tests for the detection of *Schistosoma mansoni* infection in children in western Kenya. *Am. J. Trop. Med. Hyg.*, 92(6):1233–1239.

Yang, H. M. (2003). Comparison between schistosomiasis transmission modelings considering ascquired immunity and age-structured contact pattern with infested water. *Math. Biosci.*, 184:1–26.

Yang, H. M. and Yang, A. C. (1998). The stabilizing effects of the acquired immunity on the schistosomiasis transmission modeling - the sensitivity analysis. *Mem. Inst. Oswaldo Cruz*, 93:Suppl. I: 63–73.