Markov model to forecast the change in prevalence of soil-transmitted helminths during a control programme: a case study in Vietnam

Antonio Montresor[a],*, Albis Francesco Gabrielli[a], Aya Yajima[a], Nam Lethanh[b], Beverley-Ann Biggs[c], Gerald J. Casey[c], Ta Thi Tinh[d], Dirk Engels[a] and Lorenzo Savioli[a]

[a]Department of Control of Neglected Tropical Diseases (NTD), World Heath Organization, Avenue Appia 20, 1211 Geneva, Switzerland; [b]Swiss Federal Institute of Technology (ETHZ), Zurich, Switzerland; [c]Department of Medicine (RMH/WH), University of Melbourne, Royal Melbourne Hospital, Parkville, Australia; [d]National Institute of Malariology, Parasitology and Entomology (NIMPE), Hanoi, Vietnam

*Corresponding author: Tel: + 41 22 791 21 11; Fax: + 41 22 791 31 11; E-mail: montresora@who.int

Background: A mathematical model based on the Markov methodology to predict the change in prevalence of soil-transmitted helminth (STH) infections during public health control activities is not available, but would be an extremely efficient planning tool.

Method: We used the parasitological data collected during a deworming and iron supplementation programme for women of child-bearing age conducted in Vietnam between 2006 and 2011 to develop a Markov transition probability model. The transition probabilities were calculated from the observed changes in prevalence in the different classes of intensity for each STH species during the first year of intervention. The model was then developed and used to estimate the prevalence in year 2, 3, 4 and 5 for each STH species and for 'any STH infection'. The prevalence predicted by the model was then compared with the prevalence observed at different times during programme implementation.

Results: The comparison between the model-predicted prevalence and the observed prevalence proved a good fit of the model.

Conclusions: We consider the Markov transition probability model to be a promising method of predicting changes in STH prevalence during control efforts. Further research to validate the model with observed data in different geographical and epidemiological settings is suggested to refine the prediction model.

Keywords: Soil-transmitted helminths, Prevalence-prediction-model

Introduction

Four species of nematode are collectively referred to as soil-transmitted helminths (STH): *Ascaris lumbricoides* (roundworms), *Trichuris trichiura* (whipworms), *Necator americanus* and *Ancylostoma duodenale*. The latter two species are indistinguishable by microscopic examination of the eggs and are therefore frequently reported collectively as hookworms. STH infections cause morbidity by adversely affecting nutritional status and impairing cognitive processes.[1] WHO recommends periodic administration of the anthelmintic medicines albendazole or mebendazole as a public health intervention for the control of STH.[2-3] Such intervention is referred to as preventive chemotherapy.

Preventive chemotherapy aims to control overall morbidity caused by STH by reducing the number of individuals with moderate-intensity and heavy-intensity infections, and in the longer term to reduce the number of infective eggs contaminating the environment, thereby reducing transmission rates in the community.[3]

Selection of the optimal intervention programme to control STH requires a reliable prediction of the change in infection prevalence during the intervention period. The prediction helps STH control programme managers to forecast the drug need and the required duration of the programme to achieve the control and/or elimination targets. It also helps to identify possible problems in a programme, such as low coverage and poor drug quality and efficacy, in the event that the anticipated epidemiological changes do not occur.

Since the 1980s, mathematical modelling has been extensively used to predict changes in the prevalence of communicable disease transmission in the absence of control and the expected impact of interventions, including those targeting STH.[4–7] Most of the models were developed using a probabilistic approach with underlying parametric distributions. For instance, Anderson and May[4] proposed a time-dependent dynamic mathematical model to represent the growth of worms and their transmission in a community. More recently,

ter Kuile at al.[8] applied a Cox-proportional hazard model to predict the rate of malaria transmission among infants in a region of Kenya.

A dynamic transmission model can also be used to predict changes in STH epidemiology during control interventions; however, application of such a model requires a large amount of empirical or experimental data to ensure accuracy of the prediction. The reality is that most STH control programme managers have to base their decisions on programme implementation on limited observed data.

To overcome this limitation we propose in the present paper an application of a Markov methodology that describes the expected changes in prevalence of any STH infection (i.e. infection with any among A. lumbricoides, T. trichiura and hookworms) as a result of transitions between different states, the states in this case being the different classes of intensity of infection defined by the WHO.[9] The advantage of using such a model is that a Markov transition probability can be estimated using only two most recent data sets, as the model is stable over time; a Markov model can therefore provide predictions for programme management with minimal data input. Furthermore, its properties can be refined, achieving greater accuracy, by using a larger set of data collected over time during programme implementation. The model has been used extensively in the field of public health.[8,10]

The present study aimed to examine the capacity of the proposed method to predict the change in prevalence of any STH infection over a 5-year period during a control programme based on biannual deworming.

Methods

The data used to develop and test the model are from an STH control intervention based on periodic administration of anthelmintic drugs and iron supplements that started in 2006 in Yen Bai province, northern Vietnam. Details on the control activities, data collection and the nutritional benefits of the intervention are reported elsewhere.[11,12] The parasitological component of the intervention was the distribution of albendazole every 6 months to women of child-bearing age; the programme reached approximately 250 000 women aged between 15 and 45 years in two rural districts for more than 5 years. Ten rounds of distribution of albendazole 400 mg to the target population were conducted between May 2006 and November 2010, achieving coverage constantly in excess of 80%.[12]

Parasitological investigations

Five parasitological surveys were conducted during the programme implementation: at baseline, i.e. before the intervention began, and 3, 12, 30 and 54 months after its start. The baseline cohort was followed up at each subsequent survey, and at 3 months and 12 months an additional group of randomly selected women was also surveyed. Therefore, the sample size varied through the five consecutive surveys (366 individuals at the baseline survey as follow: 435 individuals at 3 months, 503 individuals at 12 months, 187 individuals at 30 months and 253 individuals at 54 months).

A single faecal specimen was collected from each individual surveyed and a single sample was analysed with the Kato–Katz

Table 1. Classes of intensity of infection with soil-transmitted helminths (STH) recommended by WHO, and their classification as condition states (CS) according to Markov's terminology

Soil-transmitted helminth	CS1[a]	CS2[b]	CS3[c]	CS4[d]
A. lumbricoides	0	1–4999 epg	5000–49 999 epg	>50 000 epg
T. trichiura	0	1–999 epg	1000–9999 epg	>10 000 epg
Hookworms	0	1–1999 epg	2000–3999 epg	>4000 epg

A. lumbricoides: Ascaris lumbricoides; T. tricuria: Trichuris trichiura; epg: eggs per gram.
[a]Non-infected.
[b]Infected at light intensity.
[c]Infected at moderate intensity.
[d]Infected at heavy intensity.

technique;[13] each individual was classified for each of the three STHs as 'not infected', 'infected at light intensity', 'infected at moderate intensity' or 'infected at heavy intensity' according to the WHO thresholds[2] presented in Table 1.

In each survey, the following parameters were calculated: 1. Cumulative prevalence of infection with any of the three STH; 2. Prevalence of each STH infection; 3. Prevalence of different classes of intensity for each STH infection.

The model

The Markov model has been widely used in engineering, biotechnology and public health, to describe changes over time among discrete condition states (for example, 'healthy', 'ill', 'severely ill' and 'dead'), and analyse the probabilities of transition between such states.[13]

In mathematical terms, changes in the prevalence of infection of a disease can be represented by transition shifts through discrete condition states $i(i = 1,..,I)$, where $i = 1$ is the initial condition state (e.g. healthy state) and $i = I$ is the final condition state (e.g. death).

In one time period (in our case, 1 year), the Markov transition probability (MTP) matrix (P) can be described in cardinal form as:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1I} \\ p_{21} & p_{22} & \cdots & p_{2I} \\ \vdots & \vdots & \ddots & \vdots \\ p_{I1} & p_{I2} & \cdots & p_{II} \end{bmatrix} \qquad \text{Equation A}$$

where p_{ij} is the properties of the matrix, indicating the MTP from state i to state j, and can be estimated using observed data. As each row of the matrix P shows the possibility of

moving from one condition state to the other, the sum of each row in the matrix must be equal to 1.[14]

$$
\begin{cases}
\sum_{j=1}^{I} p_{ij=1} & \forall i, j \\
p_{ij} \geq 0 & \forall i, j
\end{cases}
\qquad \text{Equation B}
$$

Given a state probability $\pi_i(t)$ at time t and the MTP matrix P, it is possible to estimate the state probability $\pi_i(t+1)$ at time $t+1$ by using the following equation:

$$
\pi_i(t+1) = \pi_i(t) \cdot P \qquad \text{Equation C}
$$

The state probability $\pi_i(t)$ at time t is estimated by using observed data at one particular investigated time, while the estimation of the MTP p_{ij} of the matrix P can be estimated by using data of at least two observed times (e.g. one at time t and the other at time t-1). One of the popular estimation methods for P is using a brute force regression analysis with assumption that the MTP p_{ij} follows a parametric distribution.[15] This method is recommended to be used so as to give the best likelihood of the value of p_{ij}. It is believed that the more data are used in the regression analysis, the more likely it is that the value of p_{ij} is obtained. However, with a limited amount of data, the MTP can be approximated by a simplified method as described in the following equation:

$$
p_{ij} = \frac{\neq \left\{ CS(\tau_A^k) = i, CS(\tau_B^k) = j, \tau_B^k - \tau_A^k = l, k \in [1, K] \right\}}{\neq \left\{ CS(\tau_A^k) = i, k \in [1, K] \right\}}
$$

$$
\text{Equation D}
$$

Where $\neq$ {the formula} is the number of samples satisfying the formulae. In Equation D, τ_A^k and τ_B^k are referred to as observed time (e.g. time t-1 and t) of sample k (K is total sample size). CS = condition state.

In practical terms, Equation D allows us to estimate the probabilities of an individual in one class of intensity remaining in the same class of intensity or changing class during the investigated period (e.g. 1 year of intervention with two drug distributions). The transition probabilities obtained using Equation D allow us to estimate, with Equation C, the changes occurring for each parasite in each category of intensity of infection as a consequence of the intervention over a given time period (e.g. at year 3, 4 and 5). Transition probabilities were estimated for each STH species (*A. lumbricoides*, *T. trichiura* and hookworms); the predicted prevalence of each species was then calculated at annual intervals; the predicted prevalence of 'any STH infection' was then calculated from the STH-specific ones by applying the equation suggested by de Silva & Hall.[16]

Comparison between the predicted and the observed values

Transition probabilities were estimated on the basis of two sets of prevalence data: at baseline and at 1-year follow up, and were applied to the subsequent years of the programme to generate predicted values of prevalence.

The real changes observed in the 5 years of implementation of the programme were then used to validate the predictive capacity of the model. In the absence of other predictive models for STHs with which to compare the performance of our model, its predictive capacity was assessed in terms of statistical significance (χ^2) and programmatic significance. Programmatic significance was defined as a difference between predicted and observed levels of prevalence that would have resulted in any change of approach by a disease-control manager; this was on the basis that the WHO-recommended treatment strategy for preventive chemotherapy uses thresholds of prevalence of infection as the key decision-making indicator.[2]

Results

Observed reduction of prevalence over 5 years

The sample evaluated at baseline was 367 individuals, of which 308 (83.9%) were positive for any of the three infections. Of the total sample, 280/367 individuals (76.3%) were positive for hookworms, of whom 221 (60.2%) were positive with low intensity, 35 (9.5%) with moderate intensity and 24 (6.6%) with high intensity. For *T. trichiura*, 107/367 (29.2%) were positive, 99 (27%) with low intensity and 8 (2.2%) with moderate intensity. For *A. lumbricoides* 118/367 (32.1%) were positive, 45 (12.3%) with light intensity, 26 (7.1%) with moderate intensity and two (0.5%) with high intensity.

In total, 62/367 (16.8%) individuals had a helminth infection of moderate intensity and 25/367 (6.8%) an infection of heavy intensity.

During the implementation period of the control activity the prevalence of any STH infection was reduced to 65% after 3 months (sample = 435 individuals), to 33% after 1 year (sample = 503 individuals), 28.3% after 30 months (sample = 187 individuals) and to 17% after 54 months (sample = 253 individuals).

Estimation of the transition probabilities

A total of 178 individuals provided data both at baseline and at the 12-month survey, and the data collected in this group were used to assess the transition probabilities. Infection status at baseline was as follows: for hookworms, 139/178 (78%) were infected, 108 (60.7%) with light intensity, 21 (11.8%) with moderate intensity and 10 (5.6%) with heavy intensity; for *T. trichiura*, 59/178 (33%) were infected, 55 (30.8%) with light intensity and four (2.2%) with moderate intensity; for *A. lumbricoides*, 35/178 (19.6%) were infected, 21 (11.7%) with light intensity, 13 (7.3%) with moderate intensity and one (0.6%) with heavy intensity.

We recorded the changes in condition states after 12 months by analysing the data for the 178 individuals who provided data both at baseline and at the 12-month survey. The Markov transition probabilities were estimated according to Equation D, and are shown in Table 2.

Table 2 shows species-specific, 1-year transition probabilities among the different condition states (classes of intensity for STH); for example (row 1), if there are 100 people not infected by hookworm at baseline, i.e. in condition state one (CS = 1) 1 year later, approximately 87.18% of them will remain uninfected by hookworm, and 10.26% and 2.56% of them will be infected with light intensity (CS = 2) and moderate intensity (CS = 3), respectively.

The transition probabilities shown in Table 2 were then applied to observed baseline data to predict prevalence of infection at year 2, 3, 4 and 5, by using Equation C. Table 3 shows the results.

The estimated prevalence of any STH was 39% at 12 months (CI 31–46), 27.4% at 24 months (CI 21–34), 25.8% at 36 months (CI 19–33), 25.5% at 48 months (CI 19–33) and 25.5% at 60 months (CI 19–32); the changes in the prevalence of any STH and the changes in prevalence of each STH parasite are shown as curves in Figure 1.

Table 2. One-year transition probabilities among the different classes of intensity of infection (condition states [CS] in Markov terminology) for soil-transmitted helminths

	CS1	CS2	CS3	CS4
Hookworm				
CS1	0.8718	0.1026	0.0256	0.0000
CS2	0.6944	0.2963	0.0000	0.0093
CS3	0.5714	0.3810	0.0000	0.0476
CS4	0.7000	0.3000	0.0000	0.0000
Trichuris trichiura				
CS1	0.9244	0.0756	0.0000	0.0000
CS2	0.8364	0.1636	0.0000	0.0000
CS3	0.5000	0.5000	0.0000	0.0000
CS4	0.0000	0.0000	0.0000	1.0000
Ascaris lumbricoides				
CS1	0.9650	0.0140	0.0000	0.0210
CS2	0.9524	0.0476	0.0000	0.0000
CS3	0.8462	0.0769	0.0000	0.0769
CS4	1.0000	0.0000	0.0000	0.0000

Table 3. Observed prevalence at baseline and predicted prevalence of the different classes of intensity (condition states [CS] in Markov terminology) at year 2, 3, 4, and 5. The predicted prevalence for each year is calculated by applying the transition probability presented in Table 2 to the prevalence for the previous year

|	CS1	CS2	CS3	CS4
Hookworm				
Baseline	0.2191	0.6067	0.1180	0.0562
At 1 year	0.7191	0.2640	0.0056	0.0112
At 2 years	0.8213	0.1575	0.0184	0.0027
At 3 years	0.8379	0.1387	0.0211	0.0023
At 4 years	0.8405	0.1358	0.0215	0.0023
At 5 years	0.8409	0.1353	0.0216	0.0023
Trichuris trichiura				
Baseline	0.6685	0.3090	0.0225	0.0000
At 1 year	0.8876	0.1124	0.0000	0.0000
At 2 years	0.9145	0.0855	0.0000	0.0000
At 3 years	0.9168	0.0832	0.0000	0.0000
At 4 years	0.9171	0.0829	0.0000	0.0000
At 5 years	0.9171	0.0829	0.0000	0.0000
Ascaris lumbricoides				
Baseline	0.8034	0.1180	0.0730	0.0056
At 1 year	0.9551	0.0225	0.0000	0.0225
At 2 years	0.9655	0.0144	0.0000	0.0200
At 3 years	0.9656	0.0142	0.0000	0.0203
At 4 years	0.9656	0.0142	0.0000	0.0203
At 5 years	0.9656	0.0142	0.0000	0.0203
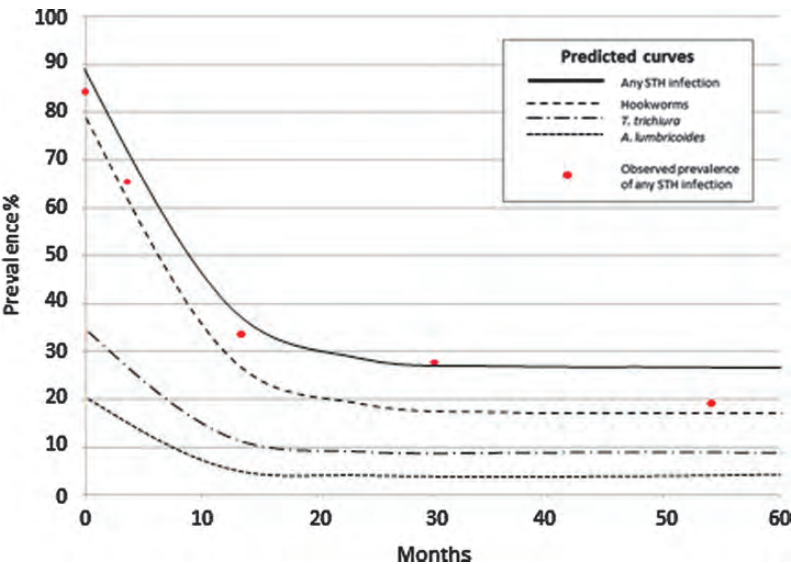
Figure 1. Predicted reduction of the prevalence of each soil-transmitted helminth (STH) and of the prevalence of any STH infection, for the 5-year control programme in Yen Bai (Vietnam) by the Markov model. The dots represent the values observed during programme implementation for any infection with STH.

The comparison between the data estimated by the model and those observed (dots in Figure 1) shows that the discrepancy between levels of prevalence is on average 1.7 percentage points for 'any STH prevalence'. The model can also predict the curves of each individual STH parasite and for each category of intensity of infection, and in this case the discrepancy between predicted and observed prevalence was between 0 and 5 percentage points (data not shown).

In all cases, the differences between estimated and observed levels of prevalence were constantly non-significant from the statistical point of view. We consider also that a difference of 1.7 percentage points would not be significant from a programmatic point of view, as it would not result in a change of treatment approach.

Discussion

The minimal discrepancy between real data observed in Vietnam and data estimated by the Markov model implies that such a model can aptly be used to predict the changes produced by preventive chemotherapy (large-scale treatment) on STH epidemiology. In addition to not being statistically significant, the difference between the observed and the estimated prevalence was only 1.7% and, as such, non-significant even from a programmatic point of view.

The model is considered valid also for the following reasons: 1. The Markov method enables us to represent changes in class of intensity that are logically plausible; after treatment, most individuals in the class 'infected at low intensity' (CS2) move to the class 'not infected' (CS1), while most of those in the class 'infected at high intensity' (CS3) move to a class for infection of low or moderate intensity (CS2 or CS3). 2. The transition possibilities for the Markov models can be estimated with a minimum of two observations (baseline and 1-year follow up).[17]

In addition, in our opinion, the proposed model can be used when only a single set of data is available (e.g. baseline only), which is the position in which most programme managers find themselves. The model can be used by applying the transition probabilities calculated in Vietnam, on the basis that such probabilities are valid in different geographical settings (e.g. other countries) where STH are transmitted. For example, a group of individuals infected at heavy intensity with hookworms and treated at the same frequency as in Vietnam will move to other classes of infection intensity in a similar way to individuals in Vietnam and elsewhere.

The estimation of the Markov transition probability on the basis of Vietnam data can be regularly improved if more data are brought into the system; data are especially needed for individuals heavily infected with *T. trichiura*, a group not represented in the present dataset. The accuracy of transition probability predictions can also be improved retrospectively, with the help of existing historical data, through a regression analysis with underlying parametric distribution.

This paper presents a methodology based on a Markov model to predict the changes in prevalence of any STH infection under control pressure. Changes in prevalence of STH infections are described as a transition probability among discrete conditions represented by the classes of intensity of infection referred to in the WHO treatment recommendations. Model estimations were compared with real changes in prevalence observed during 5 years of programme implementation; the small range of discrepancy in the comparisons indicated the usefulness of the proposed methodology. The general good fit of the data predicted by the Markov model with those observed demonstrates that the public health impact of the preventive chemotherapy intervention on each class of infection intensity is constant and therefore can be predicted using a Markov model.

The model presents two limitations that can probably be improved in the future.

The first limitation is related to the assumption that transmission remains constant over time. As can be seen from Figure 1, predicted prevalence for each of the three STHs reaches a steady state after 30 months of intervention. This is because the elimination of worms as a result of treatment is compensated by reinfection and therefore a new equilibrium state is reached in the model. This is not the case for the observed data, as they show that after a period of stability (30–50 months), prevalence of any STH infection starts again to decrease. In our opinion this discrepancy occurs because the model is based on data for the first year of implementation, when transmission is more intense as the environment is more intensively contaminated by eggs expelled by infected individuals; the model therefore applies the same, unchanged transmission rates to the following years. This is, however, unlikely to be correct, as after 5 years of intervention, multiple deworming interventions have significantly reduced the number of expelled eggs and the rate of reinfection is therefore significantly lower. It would be necessary to correct the model by taking into account this circumstance and the survival time of the different eggs and larvae in the soil.

The second limitation of the present model is linked to the small number of individuals infected at high intensity for *T. trichiura*; this could have resulted in an imprecise estimation of the transition probability for this group.

We conclude that: 1. Use of the Markov model is appropriate to predict the changes in STH prevalence occurring in a country; 2. The transition probability estimated on Vietnam data can potentially be generalised and applied to any STH control programme to predict the impact of implemented activities; in this case it would be appropriate to refine and further validate the model with data from other geographical and epidemiological settings.

References

1 Hall A, Hewitt G, Tuffrey V, de Silva N. A review and meta-analysis of the impact of intestinal worms on child growth and nutrition. Matern Child Nutr 2008;4(Suppl 1):118–236.

2 WHO. Preventive chemotherapy in human helminthiasis: coordinated use of anthelminthic drugs in control interventions: a manual for health professionals and programme managers. Geneva: WHO; 2006.

3 Gabrielli AF, Montresor A, Chitsulo L et al. Preventive chemotherapy in human helminthiasis: theoretical and operational aspects. Trans R Soc Trop Med Hyg 2011;105:683–93.

4 Anderson RM, May RM. Population dynamics of human helminth infections: control by chemotherapy. Nature 1982;297:557–63.

5 Bundy DAP, Thompson DE, Cooper ES et al. Population dynamics and chemotherapeutic control of *Trichuris trichiura* infection of children in Jamaica and St Lucia. Trans R Soc Trop Med Hyg 1985;79:759–64.

6 Anderson RM, May RM, Anderson B. Infectious Diseases of Humans: Dynamics and Control. New York: Oxford University Press, USA; 1992.

7 Luz PM, Struchiner CJ, Galvani AP. Modeling transmission dynamics and control of vector-borne neglected tropical diseases. PLoS NTD 2010;4:e761.

8 ter Kuile FO, Terlouw DJ, Kariuki SK et al. Impact of permethrin-treated bed nets on malaria, anemia, and growth in infants in an area of intense perennial malaria transmission in western Kenya. Am J Trop Med Hyg 2003;68:68–77.

9 WHO. Prevention and control of schistosomiasis and soil-transmitted helminthiasis. Report of a WHO Expert Committee. Geneva: WHO; 2002.

10 Kay R. A Markov model for analysing cancer markers and disease states in survival studies. Biometrics 1986;42:855–65.

11 Casey GJ, Phuc TQ, Macgregor L et al. A free weekly iron-folic acid supplementation and regular deworming program is associated with improved hemoglobin and iron status indicators in Vietnamese women. BMC Public Health 2009;9:261.

12 Casey GJ, Montresor A, Cavalli-Sforza LT et al. Effectiveness of weekly iron + folic acid in eliminating iron-deficiency anaemia: evidence from a 54-month weekly supplementation and deworming program. PLoS NTD. Forthcoming 2013.

13 Katz N, Chaves A, Pellegrino J. A simple device for quantitative stool thick-smear technique in *Schistosomiasis mansoni*. Rev Inst de Med Trop Sao Paulo 1972;14:397–400.

14 Howard RA. Dynamic Programing and Markov Processes. Cambridge, MI: Technology Press of the Massachusetts Institute of Technology, 1960.

15 Lancaster T. The Econometric Analysis of Transition Data. Cambridge: Cambridge University Press; 1990.

16 de Silva N, Hall A. Using the prevalence of individual species of intestinal nematode worms to estimate the combined prevalence of any species. PLoS NTD 2010;4:e655.

17 Gill J. Bayesian Methods: A Social and Behavioral Sciences Approach. 2nd edition. CRC Press (Taylor and Francis Group); 2006.