Prediction of community prevalence of human onchocerciasis in the Amazonian onchocerciasis focus: Bayesian approach

Hélène Carabin,[1] Marisela Escalona,[2] Clare Marshall,[3] Sarai Vivas-Martínez,[2, 4] Carlos Botto,[2, 5] Lawrence Joseph,[6] & María-Gloria Basáñez[2, 7]

Objective To develop a Bayesian hierarchical model for human onchocerciasis with which to explore the factors that influence prevalence of microfilariae in the Amazonian focus of onchocerciasis and predict the probability of any community being at least mesoendemic (>20% prevalence of microfilariae), and thus in need of priority ivermectin treatment.

Methods Models were developed with data from 732 individuals aged ⩾ 15 years who lived in 29 Yanomami communities along four rivers of the south Venezuelan Orinoco basin. The models' abilities to predict prevalences of microfilariae in communities were compared. The deviance information criterion, Bayesian *P*-values, and residual values were used to select the best model with an approximate cross-validation procedure.

Findings A three-level model that acknowledged clustering of infection within communities performed best, with host age and sex included at the individual level, a river-dependent altitude effect at the community level, and additional clustering of communities along rivers. This model correctly classified 25/29 (86%) villages with respect to their need for priority ivermectin treatment.

Conclusion Bayesian methods are a flexible and useful approach for public health research and control planning. Our model acknowledges the clustering of infection within communities, allows investigation of links between individual- or community-specific characteristics and infection, incorporates additional uncertainty due to missing covariate data, and informs policy decisions by predicting the probability that a new community is at least mesoendemic.

Keywords Onchocerciasis/epidemiology/drug therapy; Onchocerca volvulus; Ivermectin/therapeutic use; Prevalence; Risk factors; Bayes theorem; Models, Statistical; Venezuela (*source: MeSH, NLM*).

Mots clés Onchocercose/épidémiologie/chimiothérapie; Onchocerque volvulus; Ivermectine/usage thérapeutique; Prévalence; Facteur risque; Théorème Bayes; Modèle statistique; Venezuela (*source: MeSH, INSERM*).

Palabras clave Oncocercosis/epidemiología/quimioterapia; Onchocerca volvulus;Ivermectina/uso terapéutico; Prevalencia; Factores de riesgo; Teorema de Bayes; Modelos estadísticos; Venezuela (*fuente: DeCS, BIREME*).

الكلمات المفتاحية: داء كُلَّابيَّة الذَّنَب ، وبائيات داء كُلَّابيَّة الذَّنَب، المعالجة الدوائية لداء كُلَّابيَّة الذَّنَب، كُلَّابيَّة الذَّنَب الملتوية، إيفيرميكتين، الاستخدام العلاجي للإيفيرميكتين، معدل الانتشار، عوامل الخطر، نظرية بايبس، تحليل الموضوعات، نماذج، نماذج إحصائية، فنزويلا (*المصدر: رؤوس الموضوعات الطبية، المكتب الإقليمي لشرق المتوسط*).

Bulletin of the World Health Organization 2003;81:482-490.

Voir page 488 le résumé en français. En la página 489 figura un resumen en español. يمكن الاطلاع على الملخص بالعربية على الصفحة ٤٨٩ .

Introduction

Data produced from public health research often are organized in a hierarchical structure, with clustering within units. For example, individuals "cluster" in the same community, and communities cluster within regions. Individuals who belong to the same "unit" may share common genetic, behavioural, or social risk factors of disease. They may also have similar exposures to environmental factors or, in the case of communicable diseases, infectious agents. The health outcomes of two individuals within the same unit, therefore, will correlate more highly than those of two individuals from different units. This correlation structure must be accounted for irrespective of whether data on chronic diseases or on communicable diseases are being analysed.

Hierarchical or random effect models acknowledge the nested form of such data and allow for appropriate modelling of the correlation structure (*1–4*). The advantages of hierarchical models are not exclusive to the Bayesian framework; nevertheless, Bayesian hierarchical models are unique in that they provide a single coherent framework that allows the incorporation of multiple sources of variability (including variability that arises from missing outcomes or exposures) and subsequent quantification of within- and between-unit variability in outcome through the investigation of potential risk factors at each "level" of the model. The appropriate

[1] Department of Biostatistics and Epidemiology, Oklahoma University Health Sciences Center, Oklahoma City, USA.
[2] Centro Amazónico para Investigación y Control de Enfermedades Tropicales 'Simón Bolívar' (CAICET), Estado Amazonas, Venezuela.
[3] Department of Epidemiology and Public Health, Faculty of Medicine (St Mary's Campus), Imperial College London, London, England.
[4] Departamento de Medicina Preventiva y Social, Escuela Luis Razetti, Facultad de Medicina, Universidad Central de Venezuela, Caracas, Venezuela.
[5] Instituto de Medicina Tropical 'Dr Felix Pifano C.', Universidad Central de Venezuela, Caracas, Venezuela.
[6] Department of Epidemiology and Biostatistics, McGill University, and Division of Clinical Epidemiology, Montreal General Hospital, Quebec, Canada.
[7] Department of Infectious Disease Epidemiology, Faculty of Medicine (St Mary's Campus), Imperial College London, Norfolk Place, London W2 1PG, England (email: mbasanez@imperial.ak.uk). Correspondence should be addressed to this author.

Ref. No. **02-0399**

pooling of information across units means that hierarchical Bayesian models also overcome problems associated with small sample sizes and thus produce more reliable estimates (or predictions) of individual- and unit-specific parameters.

A fully Bayesian approach to inference requires the specification of a full probability (likelihood) model for the data, together with a prior distribution for all the unknown parameters. Once data are available, inference is made on the basis of the posterior distribution. The posterior represents what is known currently, including the prior information and that contained in the data. By Bayes' theorem, this joint posterior distribution is proportional to the product of the likelihood function and the prior distribution.

In practice, interest lies typically with the marginal posterior distributions of a subset of parameters. In realistically complex applications, evaluation of these marginal posterior distributions requires high-dimensional integration and rarely is possible analytically. One powerful technique (and the approach taken in this paper) is the implementation of a Markov chain Monte Carlo algorithm, such as the Gibbs sampler, to obtain samples from the marginal posteriors. These sampled values are then used to describe the complete distributions for the parameters of interest or to provide summaries, such as point and interval estimates. It also is possible to estimate the posterior distribution of any arbitrary function of the parameters; this is particularly useful when estimating quantities needed to inform decision making. For example, several WHO guidelines for the control of parasitic infections use threshold prevalence values to guide priority interventions (*5–6*). Often, definitive diagnosis at an individual level is difficult to acquire, in which case, interest lies with the probability that, conditional on easily observable characteristics, the (unknown) prevalence of infection in a given community is above a pre-defined threshold.

This study aimed to show the use of Bayesian methods in the analysis of the type of clustered data often encountered in public health research. In particular, we developed a Bayesian hierarchical model for human onchocerciasis to explore a variety of factors thought to influence the prevalence of infection. Onchocerciasis is caused by the parasitic nematode *Onchocerca volvulus* and is transmitted from person to person by the bite of river-breeding blackfly vectors of the genus *Simulium* (*7*). Onchocerciasis is the second most common worldwide cause of infectious blindness, and it also causes severe and incapacitating skin disease. More than 90% of the people afflicted live in Africa, but smaller foci are found in Latin America (*7*). In the latter, the Amazonian focus between Venezuela and Brazil is one of the most remote and severe (*8–12*). In addition to investigating risk factors for onchocerciasis, therefore, we estimated the probability of any community being "at least mesoendemic" and thus in need of priority ivermectin treatment (*13*), on the basis of information given by predictive variables that do not require invasive parasitological procedures. Such a method would provide an advantage over those that rely solely on skin biopsies, as well as providing valuable information for the setting of treatment priorities in areas of difficult access. A given community is at least mesoendemic when the infection prevalence surpasses 20% in the Americas (*14*) and 35% in Africa (*15*). We used the former criterion for the Amazonian focus and a method that allowed us to acknowledge the clustering of infection within communities, investigate links between individual- or community-specific characteristics and infection, incorporate additional uncertainty due to missing covariate data, and inform policy decisions by predicting the probability that a new community is at least mesoendemic.

Data and methods

Study area

From April 1995 to August 1999, 46 Yanomami communities were visited as part of an ongoing onchocerciasis epidemiological survey and control programme coordinated by Centro Amazónico para Investigación y Control de Enfermedades Tropicales. We included 29 of these communities in our study because they had not been treated with ivermectin; were situated near to or along the rivers Ocamo, Orinoco, Padamo, and Mavaca; and were located less than 250 m above sea level (Fig. 1). Above this elevation, altitude has been shown to have no impact on the prevalence of microfilariae and all communities are considered to be hyperendemic — that is, to have a prevalence $\geqslant 60\%$ (*11, 15*). From the 29 selected villages, 732 Yanomami individuals aged $\geqslant 15$ years were examined for the presence of *O. volvulus* microfilariae in the skin; this age group was chosen because it had been identified as the indicator age group in the Amazonian focus (*12*). Investigators reached the communities by first flying to the closest rural medical dispensary and then by travelling on foot or by boat, or both. Twelve villages were situated along the Ocamo river, eight along the Orinoco, four along the Padamo, and five along the Mavaca.

Measurement of microfilarial prevalence and ethical issues

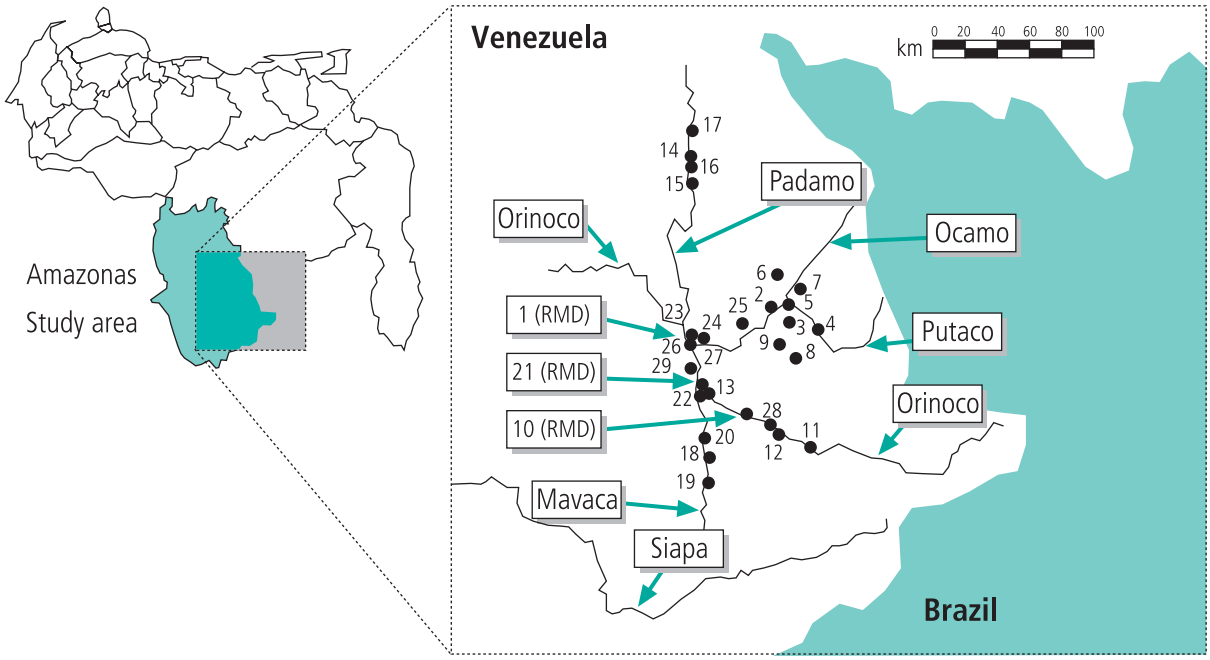
Two iliac skin snips from each participating individual were taken with a disinfected Holth-type corneoscleral punch; after 24 hours' incubation, emerging microfilariae were confirmed morphologically as *O. volvulus* (*9, 10*). The outcome was dichotomized as positive or negative for microfilariae. Informed consent for the parasitological evaluation was obtained from each individual. All eligible members of the village were treated after the parasitological examination, irrespective of their infectious status.

Measurement of factors potentially associated with prevalence of microfilariae

Potential risk factors for infection with microfilariae were measured at both individual and community levels. At the individual level, age and sex were recorded. At the community level, altitude, accessibility, and presence of a missionary post were recorded. Other variables, such as level of clothing and type of housing, correlated almost exactly with the presence of a mission and so were not considered further (data not shown). Altitude was measured in metres above sea level, as described previously (*11*). Accessibility was entered in the model as a categorical variable: "near" (<5 hours to reach the community), "intermediate" (5–24 hours), and "remote" (>24 hours), where the number of hours to reach the community started from the nearest rural medical dispensary (Fig. 1).

We hypothesized that the prevalence of onchocerciasis is influenced by two main pathways in addition to individual-level variables. First, altitude may affect infection status through its influence upon entomological determinants (*11*). Second, the presence of a mission may influence behavioural patterns, which in turn influence exposure to vectors.

Fig. 1. **Location of the 29 Yanomami study communities along the Ocamo, Orinoco, Padamo, and Mavaca rivers of the south Venezuelan Orinoco basin.** The positions of the rural medical dispensaries (RMD) are also shown



Statistical analysis

Descriptive statistics

Summary statistics for all variables — including means, medians, standard deviations, and proportions — were generated first, so we could perform preliminary investigations of the association between each variable and the outcome of interest (microfilarial status) and of the potential for confounding between variables. Subsequent analyses used age as a categorical variable (15–19 years, 20–39 years, and ⩾40 years), because its effect was not linear on a logit scale.

Bayesian hierarchical model (*16, 17*)

At the first level, we assumed a logistic regression model, in which the logit probability of infection of each individual was modelled as an additive function of possible individual-level risk factors (for example, age and sex) and of a random community-specific intercept. The latter reflected the underlying prevalence level (on a logit scale) in each community after individual-level characteristics were adjusted for. A linear regression model was then considered at the second level, and the community effects were modelled as a function of community-level risk factors (e.g. altitude and mission). This model was later extended to a third "river" level by introducing a random river-specific intercept at the second level (see Appendix A, web version only, available at: http://www.who.int/bulletin). The third model acknowledged the presence of unmeasured river-specific effects that might influence an individual's probability of infection and so induce correlation in prevalence among communities along the banks of one river. This consideration was motivated by the observation in the two-level model (and even after altitude was adjusted for) that the posterior estimates of the community-specific intercepts were higher in communities along the Ocamo and Orinoco rivers than in those along the Padamo and Mavaca.

All regression coefficients and associated 95% Bayesian credible intervals (95% BCI) were computed via the Gibbs sampler, which was implemented using WinBUGS software (*18*). The exponential of these coefficients was taken to obtain estimates of prevalence odds ratios and their 95% BCI. Risk factors were retained in the model if the associated 95% BCI excluded one or was borderline.

The results were based on two runs of 10 000 iterations each, after a burn-in of 1000 iterations. Convergence was assessed using the Gelman and Rubin statistic (*19, 20*).

Missing data

Covariate data were imputed for two individuals with missing age and for a community for which no altitude measurement was available. The missing ages were imputed at each iteration of the sampler in relative proportion to the age distribution in the remaining population. To reflect a priori knowledge, the missing altitude was imputed from a uniform distribution with range 165–200 m above sea level. The posterior estimates of all parameters in this model fully incorporated the additional uncertainty in these imputed data.

Model selection

After we identified the important predictors of infection, we used the deviance information criterion to compare the fit of: a naive flat model that assumed independence of disease status across all individuals after the already identified risk factors were adjusted for; a two-level hierarchical model that acknowledged clustering of infection within communities; and a three-level model that acknowledged additional clustering of communities along rivers. The deviance information criterion is a measure of model "support" that aims to balance the fit (deviance) and complexity (effective number of parameters) of a model (*21*). In this way, it can be viewed as a generalization of the Akaike

information criterion (*22*). Lower values of the deviance information criterion indicate higher model support, and a difference $\geqslant 4$ was used to discriminate between models. This threshold, which was proposed for Akaike information criterion (*23*), also was regarded as appropriate for deviance information criterion (*21*).

We then assessed the candidate models on the basis of their ability to predict the observed community-specific prevalences of microfilariae. Assessment with a Bayesian predictive model such as this is in the spirit of classical hypothesis testing (*24, 25*), in that each candidate model can be criticized without explicit consideration of an alternative. An ideal approach involves cross-validation, which contrasts the observed prevalence in each community with its corresponding predictive distribution, given the data from all remaining communities. Because this can quickly become computationally prohibitive, we adopted an approximation that allowed us to estimate the cross-validatory predicted prevalences in all of our communities in a single fit of the model (*26, 27*).

The comparison between observed and predicted prevalence is summarized via the tail area probability, or the Bayesian *P*-value: *P*(predicted $\geqslant$ observed) (*24, 25*). A probability close to 0% or to 100% would mean that prediction is nearly always lower or higher, respectively, than that observed, thus casting doubt on the model. In our context, it is less acceptable not to treat a meso- or hyperendemic community than to treat a hypoendemic community, so particular concerns would surround a model that led to estimates of *P* close to 0%.

In addition to the deviance information criterion and *P*(predicted $\geqslant$ observed) values, we calculated posterior estimates of the residuals (predicted – observed) and their 95% BCI, as a third criterion of "goodness of fit". Finally, for the model that seemed to be the most accurate according to these three criteria, we calculated the probability that each community was at least mesoendemic (*P*(predicted prevalence of microfilariae >20%)).

Results

Table 1 gives the characteristics of the study population. The overall observed prevalence of microfilariae was 32.8% (240/732), although the variation between communities was large, ranging from 0% to 100% (Fig. 2).

Accessibility to the community was associated strongly and negatively with the presence of a mission. A decision was made not to include these variables in the same model, therefore, as this introduced near-multicollinearity (data not shown). The inclusion of the presence of a mission or of accessibility did not provide significant improvements in fit, as defined above. Host age and village altitude thus were confirmed as important predictors of individual infection status. The variable "sex" was kept in all models to control for its possible confounding effect. The effect of altitude on an individual's infection status was modified clearly according to which river their community was located along (Fig. 2). The coefficient for altitude was assumed, therefore, to be river-specific. The small numbers of communities along each river meant that we achieved greater precision in our estimates of river-specific coefficients by assuming that they are exchangeable a priori.

The values of the deviance information criteria for the three models tested were highly supportive of the two models that acknowledged clustering of infection within communities. The deviance information criterion for model 1 was 669, for model 2 was 588, and for model 3 was 587 — a difference of more than 80 was seen between the flat and hierarchical models.

Table 1. **Characteristics of and prevalence of infection with *Onchocerca volvulus* microfilariae in 732 people aged $\geqslant$ 15 years from 29 Yanomami communities in the Amazonian focus of onchocerciasis, southern Venezuela**

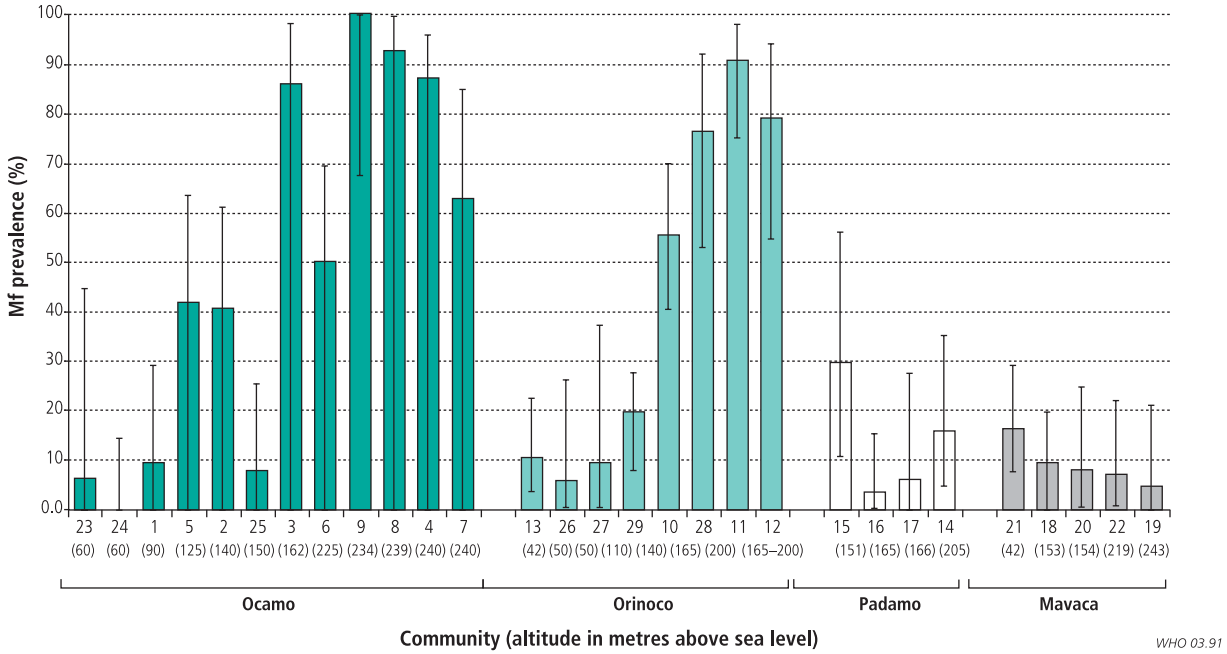
Factor	Average[a]	No.[b]
Individual level		
Prevalence of microfilariae	NA[c]	240 (32.8)
Age (years)[d]		
15–19	NA	138 (18.6)
20–39	NA	391 (52.7)
$\geqslant$40	NA	201 (27.1)
Sex ratio (male:female)	NA	384:348 (52.5:47.5)
Community level		
Altitude (m above sea level)[e]	150.7 (67.5)	NA
Ocamo river	163.8 (70.8)	NA
Orinoco river[f]	108.1 (63.0)	NA
Padamo river	171.8 (23.2)	NA
Mavaca river	162.2 (78.0)	NA
Accessibility[g]		
Near	NA	12 (41.4)
Intermediate	NA	12 (41.4)
Remote	NA	5 (17.2)
Presence of a mission	NA	10 (34.5)

[a] Values in parentheses are standard deviations.
[b] Values in parentheses are percentages.
[c] NA = not applicable.
[d] Two values missing for age.
[e] One value missing for altitude, but known to be between 165 and 200 m.
[f] One value missing for the Orinoco river.
[g] Near, <5 hours to reach community from nearest rural medical dispensary; intermediate, 5–24 hours; and remote, >24 hours.

Fig. 3 shows the predicted prevalences of microfilariae and their 95% BCIs, along with observed prevalences. The predictions from model 3 were markedly better than those from model 2. Table 2 gives Bayesian *P*-values and residuals. Model predictions seem to be reasonably good for all communities except for community 3 (Aweitheri), which is at a relatively low altitude (162 m) but has a very high recorded prevalence of infection (86%).

Model 3 showed that prevalence odds ratios increased with age: prevalence odds ratios were 2.44 (95% BCI 1.29–4.17) times higher for 20–39 year olds than for individuals aged 15–19 (reference group) and 4.18 (2.00–7.76) higher for those aged $\geqslant$40 years. The prevalence odds ratio also was slightly higher for men (1.40, 0.88–2.14) than women (reference group). The effect of altitude varied according to which river the community was situated along. For communities situated along the Ocamo and Orinoco rivers, the prevalence odds ratio increased as the altitude measured in metres above sea level increased: 1.03 (1.02–1.04) vs 1.04 (1.02–1.05), respectively. In contrast, altitude had a negligible effect on the prevalence odds ratio for communities along the Padamo river (1.01, 0.96–1.06), and even slightly reduced the prevalence odds ratio for communities along the Mavaca river (0.99, 0.97–1.00). These prevalence odds ratios indicate the increase in the prevalence odds for a community only one metre higher than its reference.

Fig. 2. **Observed community prevalence of *Onchocerca volvulus* microfilariae (mf) in the indicator age group (15 ≥ years) in the 29 Yanomami study communities, ordered by ascending altitude within river systems** (figures in parentheses show height, in metres above sea level). The community numbers correspond to those in Table 2. Dark green: Ocamo; pale green: Orinoco; white: Padamo; grey: Mavaca. Whisker plots show 95% Bayesian credible intervals



In model 3, the residual variability in (adjusted) prevalence between communities on the logit scale was decomposed into the variability between communities situated along the same river and the variability between rivers. After age, sex, and the differential effect of altitude on probability of infection were adjusted for, approximately 77% of the between-community variability was attributed to differences between rivers. This information is crucial to the reliable prediction of community-specific prevalence (Fig. 3).

To demonstrate the usefulness of model 3 in identifying communities that warrant mass ivermectin treatment, the approximate cross-validatory probability that the model would predict a prevalence of microfilariae >20% was assessed for each community. Twenty-five of the 29 communities were classified correctly (Fig. 4). One of the incorrectly classified communities (Purima) had an observed prevalence of 19.4%, but its probability of being classified as at least mesoendemic was 65%. Toothothopiwei had an observed prevalence of 39%, but the probability that we classified this as at least mesoendemic was 45%, marginally below our decision threshold. Haruri is on the banks of the Padamo river, yet unlike the other three communities alongside this river, its prevalence of onchocerciasis was mesoendemic (observed prevalence, 29%). Finally, Yepropë had a very low recorded prevalence (only 8%) compared with other communities lining the Ocamo at the same altitude.

Discussion

This paper shows the broad applications and usefulness of Bayesian random-effects models in dealing with covariates measured at various levels, clustering effects, and missing data. Another advantage of this approach is its ability to obtain estimates of arbitrary functions of model parameters, which automatically and coherently take into account all sources of uncertainty. We were able to predict correctly the outcome variable of interest for control planning purposes (that is, the probability that a new community has a prevalence larger than a predetermined threshold) in 25 out of 29 Yanomami communities of the Amazonian focus of onchocerciasis.

In this focus, it had already been established that communities situated along the Ocamo and Orinoco rivers and higher than 200 m above sea level were hyperendemic (prevalence of microfilariae ⩾60%) (*12*). Below this altitude, however, prevalence had been shown to range from hypoendemic to hyperendemic, so that other criteria on which a decision on whether a community should receive priority treatment could be made needed to be identified. We extended the analysis to two additional river systems (Padamo and Mavaca) and to other community-level variables.

The study population represented approximately 40% of the Yanomami communities located lower than 250 m above sea level in the Venezuelan Amazonian focus of onchocerciasis; this in turn represented 30% of the total communities (*28*). The study communities varied widely in their degree of contact with mainstream culture because their selection was not biased in favour of the most accessible communities, and this allowed us to explore their contribution to the predictive model. In addition to the four rivers investigated, other river systems in the region (for example, Siapa) hardly have been studied. Ideally, the same exercise should be run for a sample of communities located on all rivers.

Our results show that most variability in prevalence is attributable to differences between rivers, but that important between-community variations remained, even within river systems. At the community level, and for a chronic infection such as onchocerciasis that has average latent and duration periods of one year and >10 years, respectively, the patterns of

Fig. 3. **Observed (triangles) and predicted (circles) microfilariae (mf) prevalence for (a) model 2, and (b) model 3** (see text for details). Error bars show 95% Bayesian credible intervals

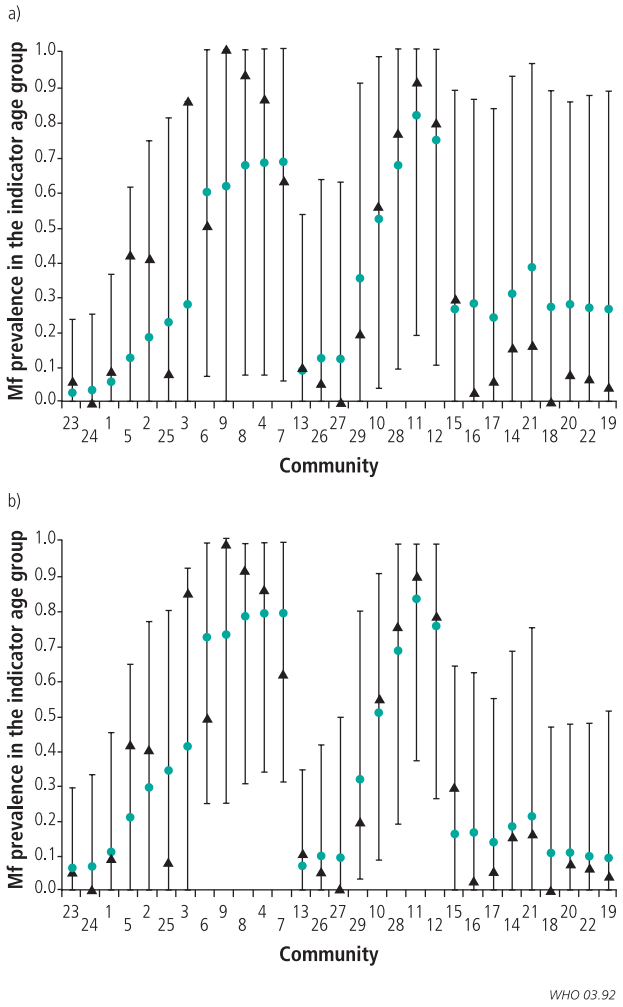


Table 2. **Observed, predicted, and residual prevalences of microfilariae in 29 communities in the Amazonian onchocerciasis focus of southern Venezuela and probability that predicted values ⩾ observed values**

Community	Prevalence			
	Observed	Predicted[a, b]	Residual[b]	*P*-value[c]
1 Ocamo	9.1	9 (0–45)	0 (–9–41)	43.3
2 Maweti	40.7	26 (0–78)	–14 (–41–37)	26.8
3 Aweitheri	85.7	43 (0–92)	–43 (–86–37)	4.2
4 Pashopeka	86.8	84 (34–100)	–3 (–55–13)	43.7
5 Toothothopiwei	41.7	17 (0–65)	–22 (–39–26)	15.3
6 Arata	50.0	79 (25–100)	25 (–29–50)	86.7
7 Potomawei	62.5	86 (31–100	19 (–38–38)	83.3
8 Warapawei	92.3	85 (31–100)	–8 (–69–8)	29.5
9 Wareta	100.0	75 (25–100)	–25 (–83–0)	7.2
10 Mahekoto	55.3	51 (8–91)	–4 (–47–36)	44.0
11 Hasupiwei	90.6	88 (37–100)	0 (–50–9)	44.3
12 Hapokashita	79.0	79 (26–100)	0 (–53–21)	53.9
13 Shakita	10.2	4 (0–35)	–6 (–10–24)	21.9
14 Buena Vista	15.4	12 (0–69)	–3 (–15–58)	43.7
15 Haruri	29.4	12 (0–65)	–18 (–29–35)	19.0
16 Tacamare	2.9	11 (0–63)	8 (–3–60)	84.8
17 Yahanamaña	5.6	11 (0–56)	6 (–6–50)	65.0
18 Mavaquita	0.0	6 (0–47)	6 (0–47)	82.9
19 Mrakapiwei	4.0	4 (0–52)	0 (–4–48)	54.9
20 Sipoi	7.4	7 (0–48)	0 (–7–44)	48.1
21 Warapana	16.0	14 (0–76)	–2 (–16–60)	47.6
22 Washewe	6.5	6 (0–48)	0 (–6–45)	45.2
23 Iyewei	5.9	0 (0–29)	–5 (–6–23)	34.4
24 Kashora	0.0	4 (0–33)	4 (0–38)	80.3
25 Yepropë	7.7	30 (0–81)	23 (–7–73)	90.5
26 Yoohopë	5.3	5 (0–42)	0 (–5–42)	53.2
27 Shashana	0.0	0 (0–50)	0 (0–50)	71.2
28 Cerrito	76.2	71 (19–100)	–4 (–57–24)	47.5
29 Purima	19.4	29 (3–81)	10 (–16–61)	45.0

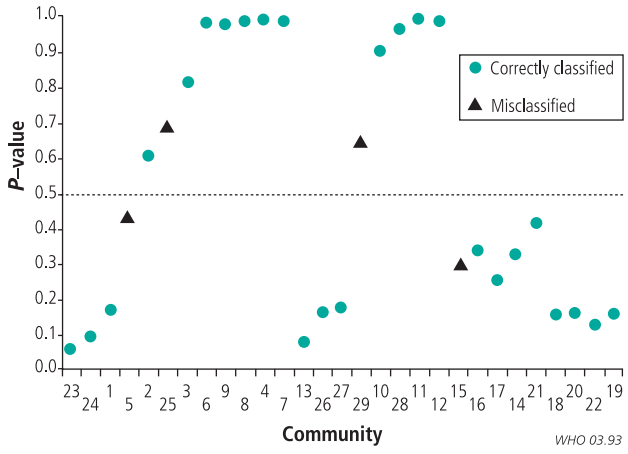
[a] Predicted by model 3 (a three-level Bayesian hierarchical model).
[b] Values in parentheses are 95% Bayesian credible intervals.
[c] Probability (predicted ⩾ observed).

micro-movements (every 2–3 years) and macro-movements of Yanomami communities (*29*) may contribute towards the variation of prevalence of microfilariae among villages. Over the last 200 years, the Yanomami people have tended to migrate from the higher altitudes of the Parima mountains to lower riverine locations (*29*). Some communities now found at similar altitudes may have had different geographical origins within the region, and geographical proximity between villages is not necessarily a good reflection of their contact and exposure patterns.

Another factor that could explain the large variation in prevalence of microfilariae among communities may be the slope of the terrain. We showed that within the Orinoco and Ocamo river systems, prevalence of microfilariae increased with altitude. Along these two rivers, altitudinal gradient was a strong determinant of the presence, species composition, and abundance of blackfly vectors, which also differed in vectorial competence and capacity (*10, 30, 31*). Altitude itself, therefore, had a strong biological effect on infection prevalence. In contrast, the effect of altitude was negligible for communities located along the Padamo and Mavaca rivers. The rate at which altitude varies with distance is very different between these rivers, and slope (rather than just altitude) could influence the distribution of sites suitable for the immature stages of the different vector species.

At the river level, micro-movements of communities usually take place along the same river, and some river-specific variables might explain why prevalence of microfilariae is higher along the Ocamo and Orinoco rivers. The variability among rivers could be due to ecological factors that in turn determine the availability and productivity of the breeding sites of the different vector species, as well as the abundance and age-structure of the biting population (*32*). As an example, previous work conducted between 1995 and 1999 showed that, although the three vector species *Simulium guianense*, *S. incrustatum*, and *S. oyapockense* were present in communities along the Ocamo river, *S. incrustatum* consistently was absent along the Orinoco (*32*). Preliminary entomological surveys along the Padamo river showed that *S. oyapockense* (a less successful vector) was present but that *S. guianense* (the species with highest vector competence) was absent. Other, as yet unrecognized, characteristics of the rivers — or characteristics common to communities situated along the same rivers — may be important for the transmission of onchocerciasis.

Fig. 4. **Community-specific probabilities that mf prevalence > 20% (at least mesoendemic).** All communities with $P \geq 0.5$ are classified as at least mesoendemic and would be candidates for priority ivermectin treatment



Investigation of the pattern of estimated community- and river-specific intercepts may help generate new hypotheses about these characteristics.

Given our current knowledge about factors that precisely determine prevalence of onchocerciasis, accurate prediction of the prevalence of infection remains difficult, and even our best model is imperfect. The problem is of obvious importance, however, as decisions must be made about which communities should be included in control programmes. As more data become available, our model can be updated to provide increasingly accurate predictions.

Although we considered variables other than altitude in our analyses, we were unable to disentangle behaviour-related pathways from vector-related pathways associated with the prevalence of microfilariae. This was because most missions were located at a low altitude (and therefore were easier to access) in our dataset. *S. oyapockense* had already been shown to be the predominant species at sites lower than 150–200 m above sea level (*11, 31*). A larger dataset, with more altitudinal variation in the location of missionary posts, would be needed to disentangle the possibly independent effects of altitude (our indicator of prevailing vectors) and presence of a mission (our indicator of cultural changes). Well-established mission posts are found at 250 and 950 m (in villages located in the Sierra Parima, which were not analysed here therefore), and these have a predominance of the highly competent *S. guianense* (*30, 31*) and a high (pre-control) prevalence of microfilariae despite long-lasting missionary influence (*8–11, 33*).

Although we have shown the application of a Bayesian approach to the selection of communities with prevalence of microfilariae >20%, a different threshold for mesoendemicity could be set according to the epidemiological patterns of a given focus of onchocerciasis. In the Amazonian focus, previous research highlighted that prevalence of microfilariae increases with age, the threshold prevalence could be 30% instead of 20% when the indicator age group is used (*12, 34*). In Africa, the threshold prevalence would be set at 35–40% (*13, 15*). The approach presented here would need information to be collected on the age and sex distribution, as well as on the size of Yanomami communities not yet evaluated. This is becoming possible, as several health care programmes are being implemented, and communities are visited regularly by trained personnel (*35*). In this way, Bayesian modelling could play an important role in the planning of community-based interventions of onchocerciasis in general and of control programmes in particular. ■

Acknowledgements

We thank all the Yanomami communities that participated in this study. William Bourgeon, Hortensia Frontado, Mayila García, Miliam Pacheco, Nátali Vásquez, and Néstor Villamizar from Centro Amazónico para Investigación y Control de Enfermedades Tropicales helped process the parasitological and entomological samples. We also wish to thank the Onchocerciasis Elimination Program for the Americas, the health personnel of the Upper Orinoco Health District, the authorities of the Regional Health Directorate, the missionaries, and the Venezuelan Air Force for logistical support. Roberto Barrera kindly helped prepare the maps in Fig. 1. Finally, we thank David Balding and two anonymous referees for their helpful comments on an earlier draft of this paper.

Funding: H.C. and M.-G.B. received funding from the Wellcome Trust, M.E. and M.-G.B. from the British Council, S.V.-M. and C.B. from World Bank (grant no. BM-VEN-96002), M.E. and C.B. from Onchocerciasis Elimination Program for the Americas, and L.J. from the Canadian Institute for Health Research.

Conflicts of interest: none declared.

Résumé

Prévision de la prévalence communautaire de l'onchocercose humaine dans le foyer amazonien : approche bayésienne

Objectif Mettre au point un modèle hiérarchique de type bayésien applicable à l'onchocercose humaine permettant d'étudier les facteurs qui influent sur la prévalence des microfilaires dans le foyer amazonien d'onchocercose et de prévoir la probabilité que l'onchocercose sévisse au moins sur le mode mésoendémique dans une communauté donnée (prévalence des microfilaires >20 %) et nécessite par conséquent en priorité un traitement par l'ivermectine.

Méthodes Mise au point de modèles à partir des données recueillies auprès de 732 personnes de 15 ans au moins habitant dans 29 communautés Yanomami situées le long de quatre fleuves du bassin méridional de l'Orénoque au Venezuela. La capacité des divers modèles à prévoir la prévalence des microfilaires dans la communauté a été comparée. Le meilleur modèle a été sélectionné par approximation croisée en utilisant le critère d'information bayésien, les valeurs de p des modèles bayésiens et les résidus.

Résultats Le meilleur modèle est un modèle à trois niveaux qui tient compte du regroupement des cas dans les communautés, avec, au niveau individuel, la prise en compte des caractéristiques d'âge et de sexe de l'hôte, au niveau communautaire, la prise en compte de l'effet de l'altitude fleuve-dépendant et, au troisième niveau, la prise en compte de l'agrégation des communautés le

long des fleuves. Ce modèle a permis de classer correctement 25 des 29 villages (soit 86 %) quant à la priorité du traitement par l'ivermectine.

Conclusion Les méthodes de Bayes sont une approche souple et utile pour la recherche en santé publique et la planification de la lutte contre les maladies. Notre modèle tient compte de l'agrégation des cas au sein des communautés, permet d'étudier le lien entre d'une part les caractéristiques particulières aux individus ou aux communautés et d'autre part l'infection, tient compte de l'incertitude supplémentaire due aux données manquantes concernant les covariables, et permet d'étayer les décisions politiques grâce à des variables prédictives de la probabilité que l'onchocercose soit au moins mésoendémique dans une nouvelle communauté.

Resumen

Predicción de la prevalencia comunitaria de la oncocercosis humana en el foco amazónico: un enfoque bayesiano

Objetivo Desarrollar un modelo jerárquico bayesiano de la oncocercosis humana para estudiar los factores que influyen en la prevalencia de microfilarias en el foco amazónico de oncocercosis, y estimar la probabilidad de que una comunidad sea como mínimo mesoendémica (prevalencia de microfilarias > 20%) y necesite por tanto tratamiento prioritario con ivermectina.

Métodos Se desarrollaron modelos con datos de 732 individuos ⩾ 15 años que vivían en 29 comunidades yanomami a lo largo de cuatro ríos de la cuenca meridional del Orinoco venezolano, y se comparó la capacidad de cada modelo para predecir la prevalencia de microfilarias en las comunidades. Se seleccionó el mejor modelo por aproximación cruzada utilizando el criterio de información bayesiano, los valores *P* de los modelos bayesianos y los residuos.

Resultados El modelo que mejor funcionó fue uno de tres niveles que tenía en cuenta el agrupamiento de los casos en las comunidades. El modelo incluía la edad y el sexo del huésped a nivel individual, un efecto de altitud río-dependiente a nivel de la comunidad y, en el tercer nivel, el agrupamiento adicional de las comunidades a lo largo de los ríos. Este modelo permitió clasificar correctamente 25/29 (86%) aldeas en lo referente a su necesidad de tratamiento prioritario con ivermectina.

Conclusión Los métodos bayesianos brindan un criterio flexible y útil para las investigaciones de salud pública y la planificación de la lucha contra las enfermedades. Nuestro modelo reconoce el agrupamiento de la infección en las comunidades, permite investigar la relación entre la infección y características particulares de los individuos y las comunidades, incorpora la incertidumbre adicional por falta de datos de covariables, y puede informar las decisiones de política mediante variables predictivas de la probabilidad de que una nueva comunidad sea como mínimo mesoendémica.

ملخص

التنبؤ بمعدل انتشار داء كلابية الذنب بين الناس في بؤرة أمازونية للداء بستخدام أسلوب بايس

الغرض: إعداد النموذج التراتبي المنسوب لبايس لداء كُلَّابِيَّة الذَّنَب بين الناس بقصد استكشاف العوامل المؤثرة في معدلات انتشار المكروفيلاريات في بؤرة توطُّن داء كُلَّابِيَّة الذَّنَب في الأمازون والتنبؤ باحتمال إصابة أي مجتمع آخر بتوطُّن متوسط الكثافة (يزيد فيه معدل انتشار المكروفيلاريات عن عشرين بالمئة)، مما يترتب عليه اعتبار الاحتياج للمعالجة بالإيفيرمكتبن من الأولويات.

الطريقة: تم إعداد نماذج مستندة على معطيات جمعت من ٧٣٢ شخصاً ممن تزيد أعمارهم عن ١٥ عاماً ويعيشون في ٢٩ مجتمعاً من مجتمعات بانومامي على طول أربعة أنهار من حوض أورينوكو الفنزويلي الجنوبي، ثم قورنت قدرات تلك النماذج على التنبؤ بانتشار المكروفيلاريات في المجتمعات، وقد استخدم معيار الانحراف للمعلومات (قيم P المنسوبة لبايس والقيم الثُّمالية لاختيار أفضل النماذج مع عملية تصحيح تصالبيه.

الموجودات: لقد كان أداء النموذج الثلاثي المستويات في التعرُّف على تجمعات العدوى ضمن المجتمعات هو الأفضل. وتشمل المستويات عمر الثوي وجنسه في المستوى الخاص بالأشخاص. وتأثير الارتفاع عن سطح البحر والذي يعتمد على النهر في المستوى الخاص بالمجتمعات. ثم التجمعات الإضافية للمجتمعات على طول النهر. وقد نجح هذا النموذج في تصنيف ٢٥ قرية من أصل ٢٩ قرية (٨٦٪) تصنيفاً صحيحاً يتعلق باحتياجاتها للمعالجة بالإيفيرمكتين.

الخلاصة: تُعَدُّ طريقة بايسان طريقة مرنة ومفيدة في بحوث الصحة العامة والتخطيط للمكافحة. وقد كان بوسع النموذج الذي درسناه تحديد تجمعات العدوى ضمن المجتمعات، وجعل بالإمكان الاستقصاء عن الروابط بين الخواص الفردية والخواص النوعية للمجتمعات بالنسبة للعدوى، مع إدماج المزيد من انعدام الثقة الناجم عن غياب معطيات خاصة بالمتغيرات المصاحبة، وأتاح لأصحاب القرار السياسي الفرصة للتنبؤ باحتمال إصابة مجتمع جديد بتوطن متوسط الكثافة على الأقل.

References

1. Laird NM, Warem JH. Random effects models for longitudinal data. *Biometrics* 1982;38:963-74.
2. Bryk AS, Raudenbush SW. *Hierarchical models*. Newbury Park (CA): Sage; 1992.
3. Longford NT. *Random coefficient models*. Clarendon Hills (IL): Clarendon Press; 1993.
4. Goldstein H. *Multilevel statistical models*. London: Edward Arnold; 1995.
5. World Health Organization. *Report of the WHO informal consultation on hookworm infection and anaemia in girls and women*. Geneva: WHO; 1996.
6. World Health Organization. *Guidelines for the evaluation of soil-transmitted helminthiasis and schistosomiasis at the community level*. Geneva: WHO; 1998.
7. *Onchocerciasis and its control. Report of a WHO Expert Committee on Onchocerciasis Control*. Geneva: WHO; 1995 (WHO Technical Report Series, No 852).
8. Rassi E, Monzón H, Castillo M, Hernández I, Ramírez-Pérez J, Convit J. Discovery of a new onchocerciasis focus in Venezuela. *Bulletin of the Pan American Health Organization* 1977;11:41-64.
9. Yarzábal L, Botto C, Arango M, Raga LM, Wong F, Allan R, et al. Epidemiological aspects of onchocerciasis in the Sierra Parima, Federal Territory of Amazonas, Venezuela. In: Yarzábal L, Botto C, Allan R, editors. *La oncocercosis en América*. Caracas: Centro Amazónico para la Investigación y Control de Enfermedades Tropicales; 1985. p. 43-63.

10. Basáñez M-G, Yarzábal L. Onchocerciasis in the Sierra Parima and Upper Orinoco regions, Federal Territory of Amazonas, Venezuela. In: Miller MJ, Love EJ, editors. *Parasitic diseases: treatment and control.* Boca Raton (FL): CRC Press; 1989. p. 231-56.
11. Vivas-Martínez S, Basáñez M-G, Grillet M-E, Weiss H, Botto C, García M, et al. Onchocerciasis in the Amazonian focus of southern Venezuela: altitude and blackfly species composition as predictors of endemicity to select communities for ivermectin control programmes. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1998;92:613-20.
12. Vivas-Martínez S, Basáñez M-G, Botto C, Villegas L, García M, Curtis CF. Parasitological indicators of onchocerciasis relevant to ivermectin control programmes in the Amazonian focus of southern Venezuela. *Parasitology* 2000;121:527-34.
13. Richards Jr FO, Boatin B, Sauerbrey M, Sékétéli A. Control of onchocerciasis today: status and challenges. *Trends in Parasitology* 2001;17:558-63.
14. Onchocerciasis Elimination Program for the Americas. *Evaluaciones epidemiológicas de la oncocercosis en América. Taller Operativo de Epidemiología*. Ecuador: OEPA; 1996.
15. Prost A, Hervouet JP, Thylefors B. Les niveaux d'endémicité dans l'onchocercose. *Bulletin of the World Health Organization* 1979;57:655-62.
16. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian data analysis*. London: Chapman and Hall; 1995.
17. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov chain Monte Carlo in practice*. London: Chapman and Hall; 1996.
18. Spiegelhalter D, Thomas A, Best N. WinBUGS Version 1.3. 2000. Available from: URL: http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml
19. Gelman, A, Rubin, D. Inference from iterative simulation using multiple sequences. *Statistical Science* 1992;7:457-511.
20. Cowles MK, Carlin BP. Markov chain Monte Carlo convergence diagnostics — a comparative review. *Journal of the American Statistical Association* 1996; 91:883-904.
21. Spiegelhalter DJ, Best NG, Carlin BP. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* 2002;64:583-616.
22. Lindsey JK. *Parametric statistical inference*. Oxford: Oxford Science Publications; 1996.
23. Burnham KP, Anderson DR. *Model selection and inference. A practical information theoretic approach*. New York (NY): Springer-Verlag; 1998.
24. Gelman A, Meng X-L, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* 1996;6:733-807.
25. Gelfand AE, Dey DK, Chang H. Model determination using predictive distributions with implementation via sampling based methods. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, editors. *Bayesian statistics 4*. Oxford: Oxford University Press; 1992. p. 147-67.
26. Marshall EC, Spiegelhalter DJ. Approximate cross-validatory predictive checks in disease mapping models. *Statistics in Medicine* (forthcoming).
27. Marshall EC, Spiegelhalter DJ. *Identifying outliers in Bayesian hierarchical models: a simulation-based approach. Technical report. 2003*. London: Department of Epidemiology and Public Health, Imperial College; 2003.
28. Ministerio del Ambiente y Recursos Naturales Renovables, Servicio Autónomo de Desarrollo Ambiental Amazonas. *Proyecto reserva de la biósfera Alto Orinoco-Casiquiare, Venezuela. Informe Técnico no. FT/93/09*. [Project for the biosphere reserve of the Upper Orinoco-Casiquiare area. Technical report no. FT/93/09.] Caracas: Ministerio del Ambiente y Recursos Naturales Renovables; 1998. In Spanish.
29. Chagnon NA. *Yanomamö*. Fort Worth (TX): Harcourt Brace College Publishers; 1997.
30. Takaoka H, Suzuki H, Noda S, Tada I, Basáñez M-G, Yarzábal L. Development of *Onchocerca volvulus larvae in Simulium pintoi* in the Amazonas region of Venezuela. *American Journal of Tropical Medicine and Hygiene* 1984;33:414-9.
31. Basáñez M-G, Yarzábal L, Takaoka H, Suzuki H, Noda S, Tada I. The vectorial role of several blackfly species (Diptera: Simuliidae) in relation to human onchocerciasis in the Sierra Parima and Upper Orinoco regions of Venezuela. *Annals of Tropical Medicine and Parasitology* 1988;82:597-611.
32. Grillet M-E, Basáñez M-G, Vivas-Martínez S, Villamizar N, Frontado H, Cortez J, et al. Human onchocerciasis in the Amazonian area of southern Venezuela: spatial and temporal variations in biting and parity rates of black fly (Diptera: Simuliidae) vectors. *Journal of Medical Entomology* 2001;38:520-30.
33. Yarzábal L, Arango M, Botto C, Jaimes JL, Sánchez-Beaujon R, Raga LM. Nuevas observaciones sobre la endemia oncocercósica de la Sierra Parima, Territorio Federal Amazonas, Venezuela. [New observations on endemic onchocerciasis in the Sierra Parima, Amazonas Federal Territory, Vanezuela.] In: Yarzábal L, Holmes R, Basáñez M-G, Petralanda I, Botto C, Arango M, et al., editors. *Las filariasis humanas en el Territorio Federal Amazonas, Venezuela*. [Human filariases in the Amazonas Federal Territory, Venezula.] Caracas: PROICET-Amazonas; 1983. p. 3-19. In Spanish.
34. Basáñez M-G. *Report of a short-term epidemiology consultancy for the Onchocerciasis Elimination Program for the Americas*. Puerto Ayacucho: Centro Amazónico para la Investigación y Control de Enfermedades Tropicales, Onchocerciasis Elimination Program for the Americas; 1999.
35. Ministerio de Salud y Desarrollo Social, Dirección Regional de Salud de Amazonas, Distrito Sanitario del Alto Orinoco, Laboratorios de Salud Pública, Centro Amazónico de Investigación y Control de Enfermedades Tropicales. *Plan de salud para el pueblo Yanomami*. [Health plan for the Yanomami population]. Puerto Ayacucho: Ministerio de Salud y Desarrollo Social, Dirección Regional de Salud de Amazonas, Distrito Sanitario del Alto Orinoco, Laboratorios de Salud Pública, Centro Amazónico de Investigación y Control de Enfermedades Tropicales; 2000. In Spanish.

Appendix A. Detailed descriptions of models

Model 3

Stage 1

The observed status of infection with microfilariae (0/1), Y_{icr}, of individual i ($i = 1,.., K_{cr}$) living in community c ($c = 1, ... C_r$) along river r ($r = 1, 2, 3, 4$) is modelled as a Bernoulli variate with mean θ_{icr}. That is $Y_{icr} \sim$ Bernoulli(θ_{icr}), where:

$$\text{logit}(\theta_{icr}) = \delta_{cr} + \beta_{age1} \times \text{age1}_{icr} + \beta_{age2} \times \text{age2}_{icr} + \beta_{sex} \times \text{sex}_{icr} \quad (1)$$

The parameter θ_{icr} corresponds to the underlying probability of infection for the given individual; β_{age1} and β_{age2} represent the regression coefficients for the age groups 1 (20–39 years) and 2 ($\geqslant$ 40 years), respectively; β_{sex} represents the effect of being male; and δ_{cr} represents the underlying community-within-river specific intercept.

Stage 2

The community-specific intercepts are modelled in stage 2, where $\beta_{alt[r]}$ represents the specific regression coefficients for altitude varying by river r, and ϕ_r represents the underlying river-specific intercept. The parameter σ_c^2 reflects the variability in prevalence (on a logit scale) between communities located along the same river after adjustment for altitude.

$$\delta_{cr} \sim \text{Normal}(\mu_{cr}, \sigma_c^2) \quad (2)$$

$$\mu_{cr} = \phi_r + \beta_{alt[r]} \times \text{altitude}_{cr} \quad (3)$$

Stage 3

The river-specific intercepts and regression coefficients for altitude are assumed to be normally distributed, where λ_r = global model intercept and ν_r^2 = measure of variability in prevalence between rivers.

$$\beta_{alt[r]} \sim \text{Normal}(\lambda_{alt}, \nu_{alt}^2) \quad (4)$$

$$\phi_r \sim \text{Normal}(\lambda_r, \nu_r^2) \quad (5)$$

Stage 4

At the fourth and last stage of the model, prior distributions are set for all unknown parameters, including σ_c^2, λ_r, ν_r^2, and all regression coefficients. We used diffuse prior distributions, so that *a priori* all values in the feasible range have approximately equal values.

Fig. 5 gives a graphical representation (*1A*).

Model 1

In contrast to the above, model 1 assumes common underlying prevalence of infection in all communities after taking into account sampling variability and differences in known individual- and community-level factors. Thus, model 1 assumes simply that $\sigma_{cr} = \lambda_r + \beta_{alt[r]} \times \text{altitude}_{cr}$ in equation (1) and ignores the additional complexity represented in equations (2)–(5).

Diffuse priors are specified for all unknown parameters in model 1.

Model 2

Model 2 goes a step further than model 1 to acknowledge explicitly that, after adjustment for known risk factors, prevalence of infection still varies between communities (due to the effect of unmeasured or unmeasurable factors). Communities are, however, assumed to be fully exchangeable, and so, in the above notation, model 2 assumes $\delta_{cr} \sim N(\mu_{cr}, \sigma_c^2)$, where now $\mu_{cr} = \lambda_r + \beta_{alt[r]} \times \text{altitude}_{cr}$.

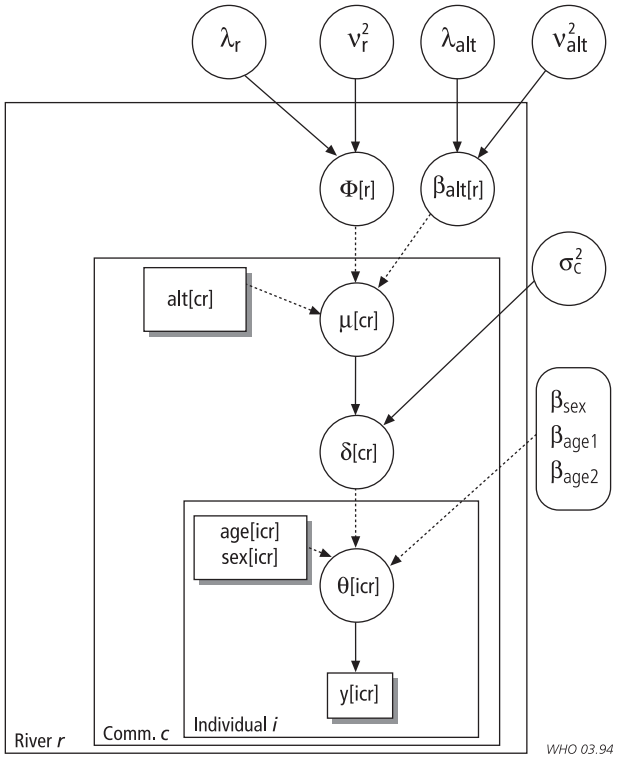
Diffuse priors are specified for all unknown parameters in model 2.

Model implementation

Convergence of the Gibbs sampler was assessed informally by examining trace and auto-correlation plots, and, more formally, via Gelman and Rubin's criterion (see Cowles and Carlin (*2A*) and incorporated references). Auto-correlation plots (not shown) illustrated negligible within-chain correlation.

In the approximate cross-validation, we drew replicate community-specific random effects and used these to predict replicate individual-level outcomes (0 or 1). The mean of these predicted outcomes for each community formed a realization from the posterior distribution of the predictive prevalence that was conditional on the known characteristics of the individuals in that community and the community itself. We then compared the predicted prevalence with the observed prevalence in each community, by computing the posterior probability that the former was higher than or equal to the latter (that is, P(predicted $\geqslant$ observed)). These Bayesian P-values were computed with a Markov chain Monte Carlo algorithm, by introducing a dummy indicator variable for each community that took the value 1 at a given iteration if, at that iteration, the value of the predicted prevalence was greater than or equal to the observed prevalence, and 0 otherwise. The average of the indicator variables over all iterations for a given community gave the required P-value.

Fig. 5. **Graphical representation of model (3).** Each quantity (parameters and data) in the model appears as a node in the graph. Rectangular nodes represent data and covariates; circular nodes represent model unknowns. Solid arrows correspond to stochastic dependencies, while those represented by dashed lines are deterministic. The overlaid "plates" represent the levels of the model



References

1A. Whittaker, J. *Graphical models in applied multivariate analysis*. Chichester: Wiley; 1992.

2A. Cowles MK, Carlin BP. Markov chain Monte Carlo convergence diagnostics — a comparative review. *Journal of the American Statistical Association* 1996;91:883-904.