OPEN ACCESS Freely available online

PLOS | NEGLECTED TROPICAL DISEASES

Onchocerciasis: The Pre-control Association between Prevalence of Palpable Nodules and Skin Microfilariae

Luc E. Coffeng[1*9], Sébastien D. S. Pion[2,9], Simon O'Hanlon[3], Simon Cousens[4], Adenike O. Abiose[5], Peter U. Fischer[6,7], Jan H. F. Remme[8], K. Yankum Dadzie[9], Michele E. Murdoch[10], Sake J. de Vlas[1], María-Gloria Basáñez[3], Wilma A. Stolk[1¶], Michel Boussinesq[2¶]

1 Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands, **2** UMI 233, Institut de Recherche pour le Développement (IRD), University of Montpellier 1, Montpellier, France, **3** Department of Infectious Disease Epidemiology, School of Public Health, Faculty of Medicine (St Mary's Campus), Imperial College London, London, United Kingdom, **4** Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom, **5** Sightcare International, Secretariat Main Office, Ibadan, Oyo State, Nigeria, **6** Washington University School of Medicine, Infectious Disease Division, St. Louis, Missouri, United States of America, **7** Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany, **8** Independent Consultant, Ornex, France, **9** Independent Consultant, Accra, Ghana, **10** Department of Dermatology, Watford General Hospital, Watford, Hertfordshire, United Kingdom

Abstract

Background: The prospect of eliminating onchocerciasis from Africa by mass treatment with ivermectin has been rejuvenated following recent successes in foci in Mali, Nigeria and Senegal. Elimination prospects depend strongly on local transmission conditions and therefore on pre-control infection levels. Pre-control infection levels in Africa have been mapped largely by means of nodule palpation of adult males, a relatively crude method for detecting infection. We investigated how informative pre-control nodule prevalence data are for estimating the pre-control prevalence of microfilariae (mf) in the skin and discuss implications for assessing elimination prospects.

Methods and Findings: We analyzed published data on pre-control nodule prevalence in males aged ≥20 years and mf prevalence in the population aged ≥5 years from 148 African villages. A meta-analysis was performed by means of Bayesian hierarchical multivariate logistic regression, accounting for measurement error in mf and nodule prevalence, bioclimatic zones, and other geographical variation. There was a strong positive correlation between nodule prevalence in adult males and mf prevalence in the general population. In the forest-savanna mosaic area, the pattern in nodule and mf prevalence differed significantly from that in the savanna or forest areas.

Significance: We provide a tool to convert pre-control nodule prevalence in adult males to mf prevalence in the general population, allowing historical data to be interpreted in terms of elimination prospects and disease burden of onchocerciasis. Furthermore, we identified significant geographical variation in mf prevalence and nodule prevalence patterns warranting further investigation of geographical differences in transmission patterns of onchocerciasis.

Citation: Coffeng LE, Pion SDS, O'Hanlon S, Cousens S, Abiose AO, et al. (2013) Onchocerciasis: The Pre-control Association between Prevalence of Palpable Nodules and Skin Microfilariae. PLoS Negl Trop Dis 7(4): e2168. doi:10.1371/journal.pntd.0002168

Editor: Christian Bottomley, London School of Hygiene and Tropical Medicine, United Kingdom

Received June 8, 2012; **Accepted** March 1, 2013; **Published** April 11, 2013

Funding: LEC and WAS were funded by a research grant from the African Programme for Onchocerciasis Control (APOC, www.who.int/apoc) for a Health Impact Assessment of APOC (CEV/APOC/241/08/TSA). MGB and WAS acknowledge seed funding from the Institute of Health Metrics and Evaluation (Seattle, USA, www.healthmetricsandevaluation.org/) as co-disease leaders for the project on global burden of onchocercal disease. SOH was funded by an ESRC (Economics and Social Sciences Research Council, UK, www.esrc.ac.uk) PhD studentship. Surveys in Cameroon were performed with the financial supports from the Special Programme for Research and Training in Tropical Diseases (Vina; www.who.int/tdr), Helen Keller International (Lekie; www.hki.org) and Institut de Recherche pour le Développement (Mbam and Faro; www.ird.fr). Surveys in Nigeria were supported by WHO/UNDP/World Bank Special Programme for Research and Training on Tropical Diseases (Project ID Numbers 870456 and 910553), the Leverhulme Trust and Sightsavers. MGB acknowledges funding from the Wellcome Trust (http://www.wellcome.ac.uk, project grant 092677/Z/10/Z). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: l.coffeng@erasmusmc.nl

9 These authors contributed equally to this work.

¶ These authors also contributed equally to this work.

Introduction

In 1995, the World Health Organization launched the African Programme for Onchocerciasis Control (APOC). At that time, APOC aimed to control morbidity due to onchocerciasis (river blindness) in Africa, with a focus on those countries not covered by the previous Onchocerciasis Control Programme in West Africa (OCP). Since 1995, APOC has successfully coordinated mass treatment with ivermectin in sixteen onchocerciasis-endemic African countries [1]. Until recently, elimination of onchocerciasis from African foci was deemed to be not achievable by means of mass ivermectin treatment alone, considering the large size of the transmission zones, the mobility of the insect vectors and human

Author Summary

Until recently, elimination of onchocerciasis (river blindness) from Africa by mass treatment with ivermectin alone was deemed impossible. However, recent reports of elimination of onchocerciasis from various African foci have stimulated renewed interest. An important determinant of achieving elimination is the pre-control microfilarial (mf) prevalence, i.e. the percentage of people with larval stages of the *Onchocerca volvulus* worm in the skin, which can be detected in a skin snip (a small skin biopsy). Because this method is considered invasive, pre-control infection levels in Africa have been mapped mostly by means of palpation of subcutaneous nodules (protuberances under the skin where the adult worms live) in adult males, a relatively crude but non-invasive method of detecting infection. We developed a tool to derive estimates of pre-control mf prevalence from available pre-control nodule prevalence estimates. This tool can help evaluate ongoing control programs, help assess local elimination prospects, and help estimate levels of disease due to onchocerciasis by linking pre-control nodule palpation data to the large body of literature on the association between mf prevalence and disease.

populations, and poor compliance with mass treatment in some areas [2]. However, following the first reports of elimination of onchocerciasis from foci in Mali, Senegal, and Nigeria by mass treatment alone [3,4,5], there is renewed interest in elimination of onchocerciasis from Africa [6].

Pre-control infection levels are an important predictor of morbidity levels [7,8,9] and the duration of onchocerciasis control programs required to achieve elimination of infection [10,11]. High pre-control levels of infection indicate circumstances that are favorable for intense transmission in terms of vector abundance, proximity to vector breeding sites, high vectorial capacity and competence, etc. In such circumstances, mass treatment with a drug such as ivermectin, which is predominantly microfilaricidal, but has a lesser impact on adult worm survival, needs to be continued for a long time and at high therapeutic and geographical coverage before it can be stopped without considerable risk of recrudescence of infection. Progress towards elimination of onchocerciasis from APOC areas is currently being evaluated by means of ongoing skin snipping surveys that measure levels of infection in terms of presence and density of microfilariae (mf) in the skin of the general population [5]. In contrast, pre-control levels of infection in APOC areas have been quantified by the REMO method (rapid epidemiological mapping of onchocerciasis), which is based on the palpation of subcutaneous nodules containing adult *Onchocerca volvulus* worms in a sample of 30–50 males aged $\geq$20 years in villages selected using a standardized selection procedure [12,13]. Results from pre-control and ongoing surveys will have to be compared, even though the REMO method is much cruder for detecting presence and intensity of infection than skin snipping. Therefore, it is important to assess how informative pre-control nodule palpation data are, and when and whether they can be reliably translated to equivalent measures of skin microfilariae. In other words, there is need for a quantitative model describing the association between pre-control nodule prevalence and pre-control presence of skin microfilariae, which takes into account the differences between the two methods as well as other covariates. Such a model would also allow estimates of pre-control nodule prevalence to be related to the large body of literature on the correlation between mf prevalence and prevalence of onchocercal morbidity, allowing better estimation of the disease burden of onchocerciasis.

We present a statistical model describing the association between pre-control nodule prevalence in adult males and pre-control mf prevalence in the general population. Quantitative relationships for this association have been previously described, but were based on smaller number of surveys, did not provide estimates of uncertainty around parameter estimates and model predictions, and did not account for geographical variation or the relatively small sample sizes routinely used for the nodule palpation method, resulting in attenuation bias (due to measurement error in nodule prevalence) [14,15,16,17]. In this study, we analyzed original pre-control data, accounting for these factors, and using Bayesian statistical methods, well known for providing robust uncertainty estimates around model parameters.

Methods

Data and Study Sites

We analyzed original data on pre-control nodule prevalence in adult males (N = 7,525 individuals) and mf prevalence in the population aged five years and above (N = 29,775 individuals) from 148 villages in seven geographical areas including countries in the former OCP area, and foci in Cameroon, Nigeria, and Uganda, which are part of APOC (Table 1, Figure 1). Most of these data have been previously published [9,14,18,19], except for part of the data from Cameroon. The simuliid vectors responsible for transmission in each area have been described previously (Table 1) [9,19,20,21,22,23]. In all areas, data on nodule and mf prevalence had been collected simultaneously (except for Nigeria, where nodule palpation took place six to twelve months after skin snipping, though still before the start of control interventions). All data on mf prevalence were based on taking two skin snips (one from each iliac crest) from each individual examined, which were incubated in saline for 24 hours, and village-level prevalence values were age- and sex-standardized according to the reference OCP population (direct standardization, supplementary Table S1). Then, we calculated the standardized number of mf positive persons in a village by multiplying the standardized prevalence with the sample size, and rounding to the nearest integer. Nodule prevalence was based on palpation-based detection of nodules that could be attributed to onchocerciasis with reasonable certainty, similar to the methodology used for mapping of infection in APOC areas; i.e. nodules of uncertain etiology (e.g. possible enlarged lymph nodes) were excluded [12]. All data were used with permission of the authors who originally collected such data, and were analyzed anonymously.

Statistical Methods and Model Fitting

The association between village-level mf prevalence and nodule prevalence was quantified in a meta-analysis by means of hierarchical multivariate logistic regression, i.e. logistic regression where the predicted outcome is a set of correlated binary random variables rather than a single binary random variable. A hierarchical approach was taken to account for unmeasured sources of variation between geographical areas. A multivariate approach was taken to account for measurement error in each measure of infection. This approach prevents regression of model coefficients towards zero (attenuation bias) as we do not have to assume that there is no measurement error in the explanatory variable (e.g. either nodule or mf prevalence), an assumption inherent to univariate regression [24].

We extended the ordinary hierarchical logistic regression model to a multivariate model simultaneously predicting m binary

Table 1. Characteristics of data used for modeling the association between prevalence of nodules and microfilariae.

Area	Number of villages	Number of males examined for nodules	Number of individuals from general population examined for microfilariae in the skin	Bioclime	Vector responsible for transmission (*Simulium* spp)	Reference
Kigoyera Parish, Uganda	8	667	2,085	Forest	*S. neavei s.s.* [19]	[19]
Onchocerciasis Control Programme in West Africa	26	1,386	5,273	Savanna	*S. damnosum s.s.* and *S. sirbanum* [9]	[9,14,18]
Kaduna, Nigeria	33	1,822	7,274	Savanna	*S. damnosum s.s.* and *S. sirbanum* [21]	[14,18]
Lekié, Cameroon	19	806	3,430	Degraded forest	*S. squamosum B* [23]	unpublished
Mbam, Cameroon	34	1,354	6,190	Forest-savanna mosaic	*S. squamosum B* [23]	unpublished
Vina, Cameroon	19	1,122	4,266	Savanna	*S. damnosum s.s.* and *S. sirbanum* [20]	[14,18]
Faro, Cameroon	9	368	1,257	Savanna	*S. damnosum s.s.* and *S. sirbanum* [22]	unpublished

s.s.: sensu stricto.
doi:10.1371/journal.pntd.0002168.t001

outcomes:

$$\mathrm{logit}\left(\pi_{ij,m}\left(y_{ij,m}=k_{ij,m}|\mathbf{X}_{ij},\beta_m,n_{ij,m}\right)\right)=\mathbf{X}_{ij}^{T}\beta_m+\varepsilon_{ij}+\varepsilon_j,$$

where $\pi_{ij,m}$ is the probability of finding k cases of the m-th binary outcome ($m=1$: presence of microfilariae in the skin; $m=2$: presence of nodules in adult males) among n_m observed individuals from the i-th unit (village) and the j-th cluster (geographical area). The error terms ε_{ij} and ε_j (each consisting of m components) represent the variation (random effects) in infection levels within and between the j geographical areas, respectively. For each village there is a set of observed covariates $\mathbf{X}_{ij}$, and for each of the m predicted binary outcomes there is a set of parameters β_m (fixed effects), where the intercepts $\beta_{0,m=1}$ and $\beta_{0,m=2}$ represent the mean log odds of presence of mf in the general population (all those aged $\geq$5 years) and nodules in adult males. To explain possible large differences between geographical areas related to bioclime, parasite strains and clinical manifestations in onchocerciasis [25], we included a set of coefficients for bioclimatic zone in the model. Here, the parameters $\beta_{1,m=1}$ and $\beta_{1,m=2}$ represent the log odds ratio of observing microfilariae in the skin and subcutaneous nodules in forest areas (including degraded forest and forest-savanna mosaic areas), relative to savanna areas. Correlation between nodule and mf prevalence was modeled by assuming a multivariate normal distribution for the m components of the error term at each level of analysis. See supplementary Text S1, section "Model description" for a more detailed description of the model.

To account for measurement error due to misclassification of nodules (e.g. classifying lymph nodes as onchocercal nodules due to imperfect specificity; or failing to detect at least one subcutaneous onchocercal nodule when one or more are present, due to imperfect sensitivity), we added parameters to the model for specificity and sensitivity of nodule palpation, allowing these to be estimated from the data. Prior information for parameter values was based on the literature. A wide range of values is reported for specificity (60%–99%), based on various definitions [15,19,26,27]. We assumed that when performed by physicians experienced in recognizing onchocercal nodules, specificity of nodule palpation is between 98% and 100%, based on the report of finding only four non-onchocercal nodules among 312 extirpated nodules [19]. Further, we assumed that sensitivity increases with level of infection, reflecting the notion that detection of at least one nodule is more likely in a person with many onchocercal nodules than in a person with few or only one [27]. In literature, no values for sensitivity of nodule palpation as a method for detecting onchocercal nodules are reported. In the current study, sensitivity was assumed to increase linearly from some unknown minimum sensitivity (value between 60% and 100%) for nodule prevalences close to zero (when persons with nodules have few nodules) to 100% for nodule prevalence of 100%. The choice of a linearly increasing pattern was based on a simulation exercise in which we examined the association between the proportion of the nodule carriers that is detected and the 'true' nodule prevalence, given simulated true nodule counts (assuming a negative binomial distribution of counts within a village) and some probability to detect each nodule (minimum sensitivity). A sensitivity analysis showed that the model fit and model predictions did not change when assuming different values for minimum sensitivity of nodule palpation at low infection levels (60%, 80%, or 100%). This is explained by the fact that sensitivity is most important for high prevalence settings (for which we assume sensitivity is high anyway), and far less important in low prevalence settings (where

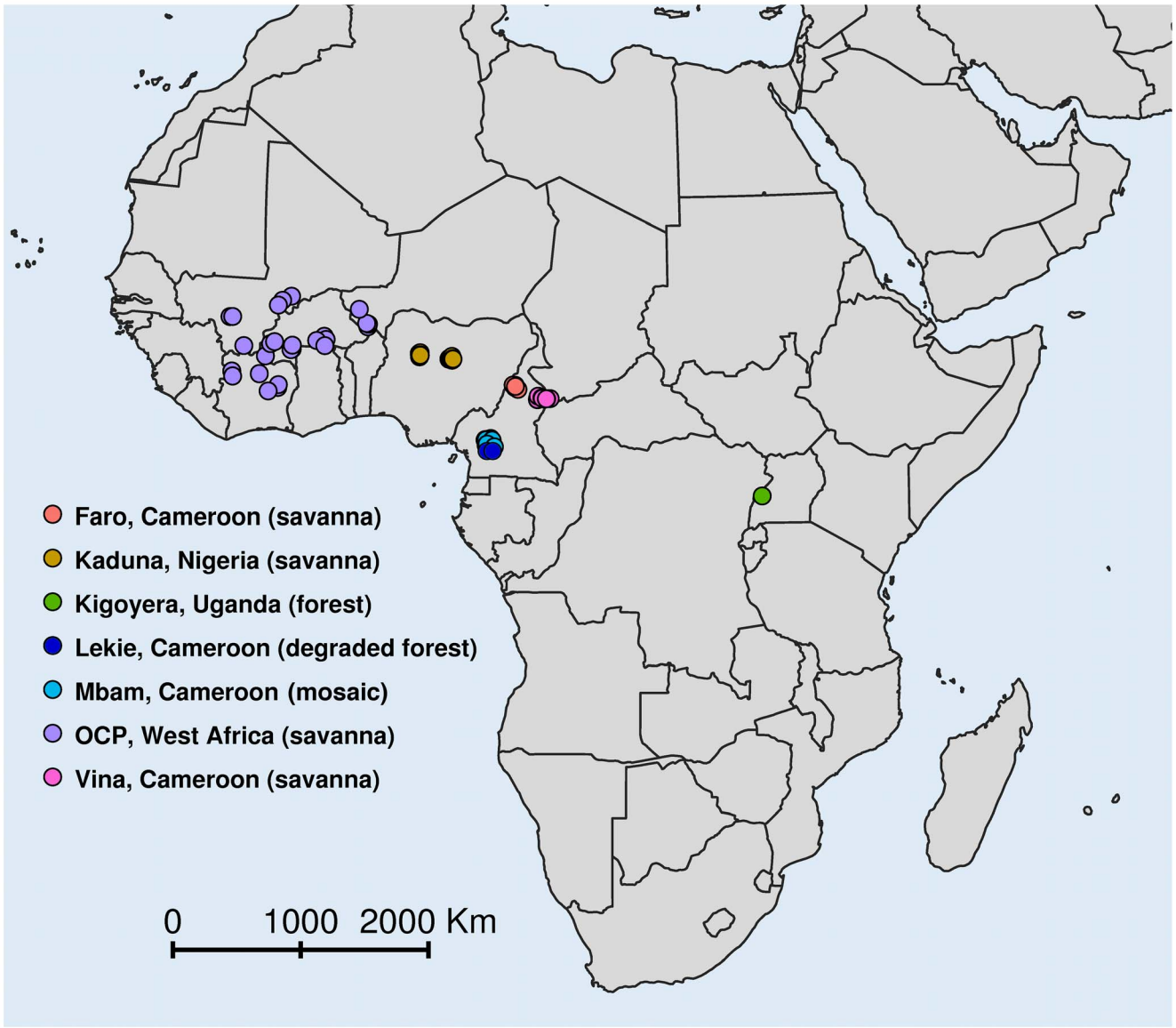


Figure 1. Locations of study sites.
doi:10.1371/journal.pntd.0002168.g001

misclassification is largely governed by specificity). Therefore, we simplified the final model by leaving out the parameter for sensitivity, effectively assuming 100% sensitivity of nodule palpation for all infection levels.

Based on the model described above, we estimated the conditional distribution of mf prevalence in a hypothetical village outside the dataset, given an estimate of the 'true' nodule prevalence in adult males (i.e. corrected for misclassification of nodules). We assumed that nodule prevalence estimates were based on a sample of 30 adult males, the minimal sample size used in REMO surveys [12,13]. See Text S1, section "Model application" for a more detailed description of the methods for predicting mf prevalences in hypothetical villages.

The model was fitted to the data in a Bayesian framework. Posterior distributions of parameters and predictions were simulated in JAGS (see Text S1, section "Model specification in JAGS" for code), a program for analysis of Bayesian models using Markov Chain Monte Carlo (MCMC) simulation based on the Gibbs sampling algorithm (version 3.2.0; Martyn Plummer, 2012, http://mcmc-jags.sourceforge.net). Simulations in JAGS were set up and analyzed in R (version 2.14.2) [28], using packages *rjags* (version 3–5, Martyn Plummer, 2011, http://CRAN.R-project.org/package = rjags) and *R2jags* (version 0.03-06, Yu-Sung Su, 2011, http://CRAN.R-project.org/package = R2jags). Improvements in model fit by addition of parameters were assessed via the deviance information criterion (DIC), a generalization of Akaike's information criterion for hierarchical models (lower values indicate better fit, taking into account model deviance and the effective number of parameters in the model) [29]. See Text S1, section "Parameter estimation" for further details about model fitting and checking of model convergence.

The final fit of the model to the data was evaluated by means of mixed posterior predictive checks [30,31]. In this procedure, the number of individuals positive for mf and nodules in each village was resampled 40,000 times from the estimated joint posterior distribution of model parameters, including resampling of all

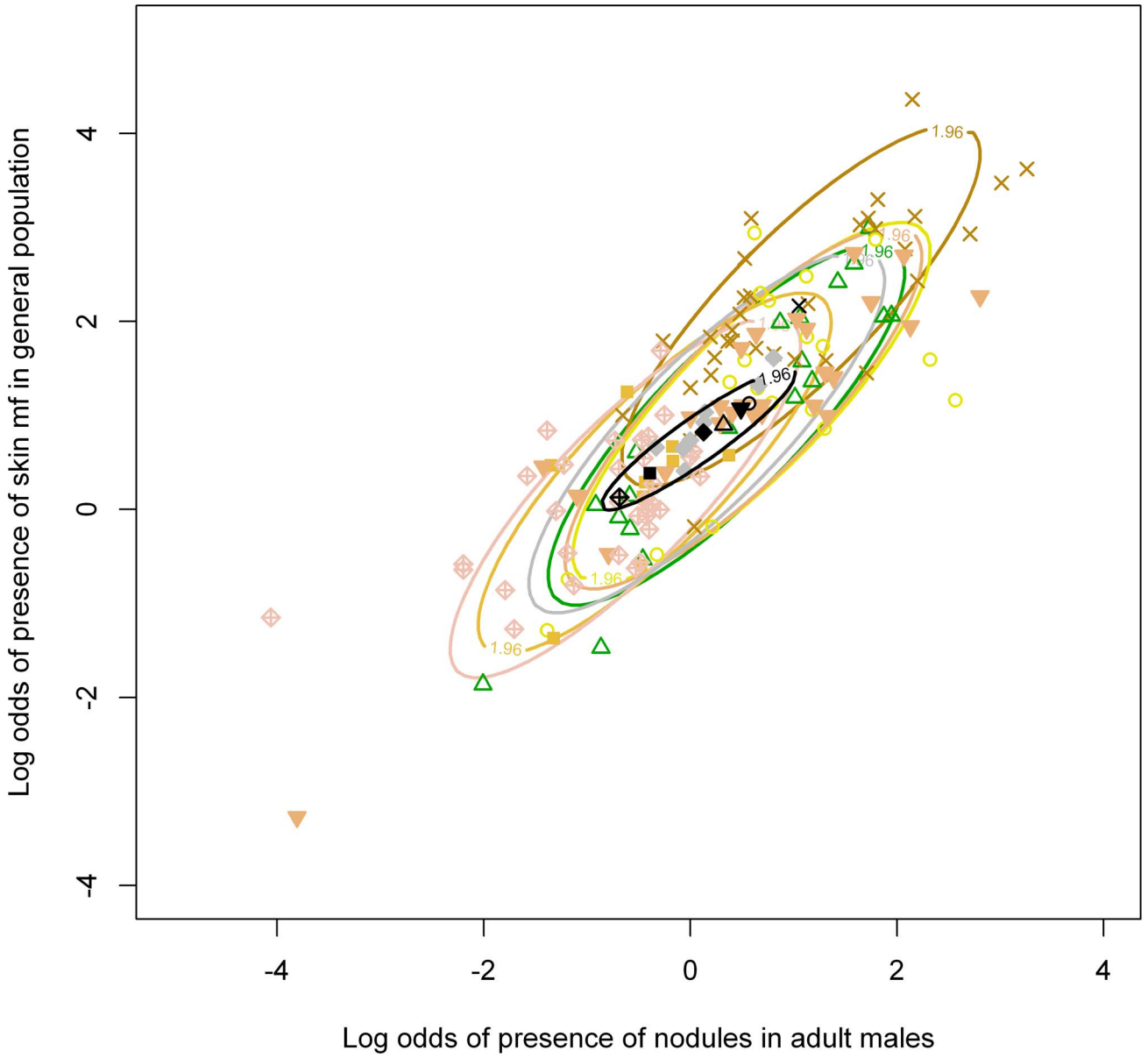


Figure 2. Association between prevalence of nodules in adult males and skin mf in the general population. Colored symbols represent data from seven geographical areas. Colored ellipses indicate the 95% percentiles (Z = 1.96) of the predicted joint distributions of infection prevalences within each geographical area, based on the estimated variances and correlation of observations within geographical areas. Black symbols represent the mean infection prevalences in each of the geographical areas. The black ellipse represents the 95% percentile of the joint distribution of mean infection prevalences in geographical areas, illustrating the deviating pattern in nodule and mf prevalence in Mbam, Cameroon (black and brown crosshairs and brown ellipse). Predictions were based on a Bayesian hierarchical multivariate logistic regression model with a fixed effect for Mbam, Cameroon, and random effects for other geographical areas.
doi:10.1371/journal.pntd.0002168.g002

random effects, and the resulting replicate dataset was compared to the original data.

Results

The median nodule prevalence in males aged ≥20 years was 58% (range: 2%–100%), and the median mf prevalence in the population aged five years and above was 74% (4%–99%). The median sample size for nodule prevalence in a village was 42 (range: 9–181). The median sample size for mf prevalence in a village was 167 (33–727).

Nodule prevalence in adult males was strongly positively correlated with mf prevalence in the general population (Table S2). There was significant geographical variation in patterns of nodule and mf prevalence; in a model without any coefficients for bioclime, the DIC increased from 1918 to 1920 when error term ε_j was omitted. Point estimates of ε_j were very similar for savanna and forest areas, with the exception of Mbam, Cameroon (forest-savanna mosaic), for which mf prevalence was relatively high compared to other areas. In line with this, the model fit did not improve when a fixed effect parameter for bioclime was added to the model. However, the model fit improved significantly when

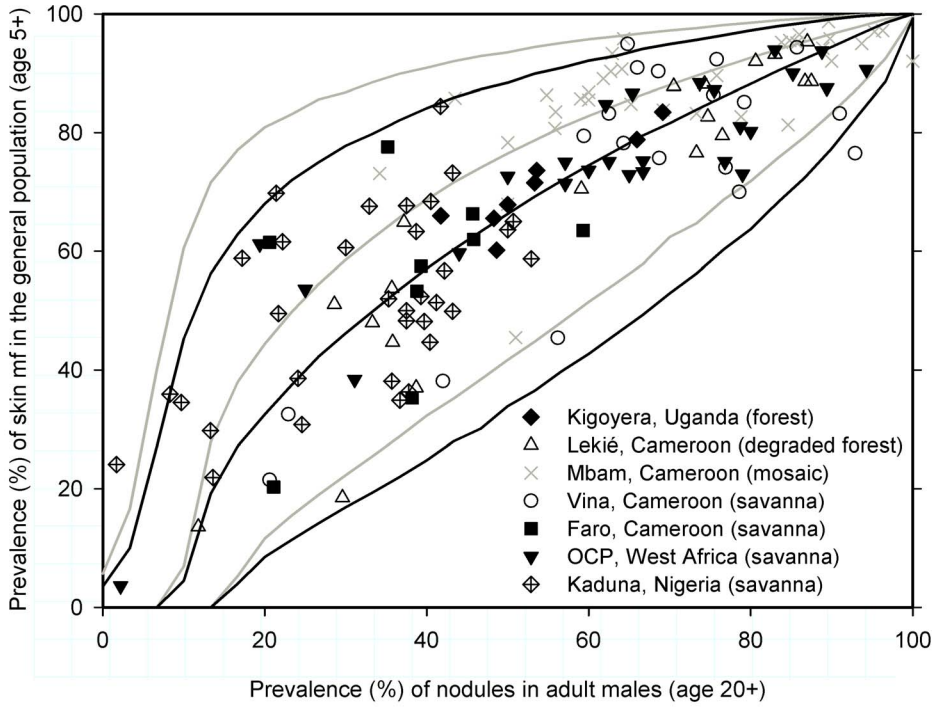


Figure 3. Predicted skin mf prevalence in the general population, given observed nodule prevalence in adult males. Symbols represent observed data by geographical area. Within each set of regression lines, the middle and outer lines relate to the median and 95% Bayesian credible intervals of the posterior predictive distribution, respectively (black set for areas all areas but Mbam; grey set for Mbam, the only forest-savanna mosaic area). Predictions were made assuming that nodule prevalence was based on a sample of 30 adult males.
doi:10.1371/journal.pntd.0002168.g003

modeling the difference between Mbam and all other areas as a fixed effect (DIC 1913 vs. DIC 1918), indicating that mf prevalences in Mbam were significantly higher than those in other areas (Table S2, Figure 2). After this adaptation of the model, there was still significant variation in patterns of nodule and mf prevalence between geographical areas due to other, unmeasured variables (the DIC increased to 1921 when error term ε_j was omitted). Further, there was considerable uncertainty in the

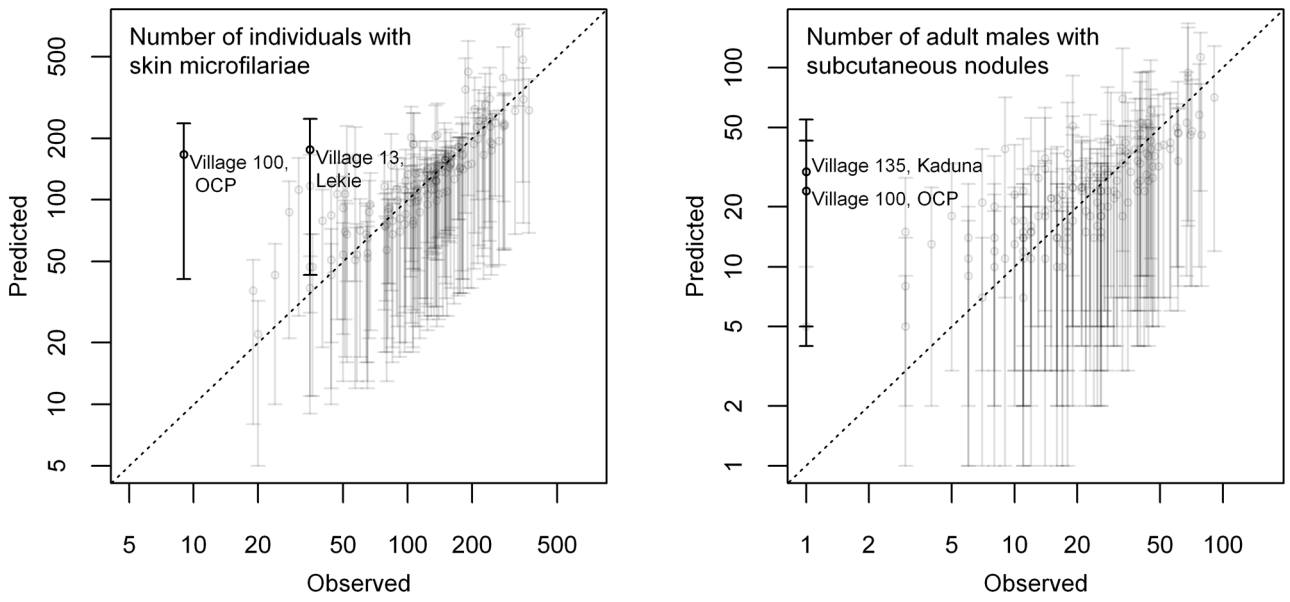


Figure 4. Comparison of observations (x-axis) versus model predictions (y-axis). The comparison was made by means of mixed posterior predictive checks of the numbers of individuals with detectable microfilariae in the skin and adult males with nodules. The dotted diagonal line represents the hypothetical perfect model fit. Error bars represent the 95% Bayesian prediction interval for the numbers of adult males with nodules and individuals with detectable microfilariae in the skin each village, and should intersect with the diagonal line if the model fit is good.
doi:10.1371/journal.pntd.0002168.g004

predictions for mf prevalence, based on nodule prevalence in a sample of 30 males from a hypothetical village outside the dataset (Figure 3).

Mixed posterior predictive checks showed that the model fitted well to the data (Figure 4). Only three villages – all from different regions, and all with relatively low infection levels compared to other villages from the same region – deviated significantly from the model predictions.

Discussion

We investigated the association between pre-control nodule prevalence in adult males (aged ≥20 years) and pre-control mf prevalence in the general population (aged ≥5 years). Our model is the first to examine geographical variation due to bioclime and other unmeasured variables, and to take account of measurement error in nodule prevalence. Our results show that there is a strong positive correlation between nodule and mf prevalence, but also significant variation between geographical regions, which should be taken into consideration when evaluating the prospects of elimination and the burden of disease.

Our analysis showed significant geographical variation in patterns of nodule and mf prevalence, though not related to bioclimatic zones according to the classic forest vs. savanna classification of onchocerciasis. In 'forest' areas – Lekié, Cameroon (degraded forest) and Kigoyera parish, Uganda (forest) – the patterns in nodule and mf prevalences did not differ much from the pattern in savanna areas. Yet, we found that mf prevalence levels in the general population were relatively higher in the only forest-savanna mosaic area (Mbam, Cameroon), while nodule prevalence in adult males levels were not significantly different. There are several possible explanations for this pattern. Most likely, the pattern in Mbam is explained by a different pattern in age-dependent exposure to black flies' bites. Both mf and nodule prevalences in individuals under the age of twenty years were relatively high in Mbam compared to the other areas in Cameroon, especially in villages with relatively low nodule prevalence in adult males (data not shown). This indicates that individuals in Mbam experience relatively high exposure levels at a young age. This might be explained by the presence of dense forest in this region with relatively few narrow open spaces, which is associated with higher dispersal of flies around the breeding sites [32]. Therefore, exposure may not be concentrated near the breeding sites, but may extend over the whole village. Related to this, exposure may be less concentrated in adults (who frequently spent time near the breeding sites, forest galleries for fishing, etc.), but may be more equally distributed over all age groups. However, dense forest may not be unique for Mbam, and may also be present in other forest areas in our data set. Therefore, we can only say that it may be important to consider age-dependent patterns in exposure to black flies' bites and their effect on transmission when translating nodule prevalence data to mf prevalence. We rule out demography and survey methods, as all mf prevalences were standardized, the mean age of the sampled men from Mbam was similar to that of men from the other Cameroonian areas, methods for skin snipping and mf enumeration were the same as in other Cameroonian areas and, in addition, even conducted by the same person (MB performed all skin snipping in Faro, Lekié, and Mbam, and 50% of skin snipping in Vina valley). Furthermore, it is also unlikely that the forest sites other than Mbam – Lekié and Kigoyera parish – harbor a savanna parasite strain (instead of the assumed forest parasite strain) as this is inconsistent with observed patterns of blindness in these areas (forest pattern) [33,34]. Lastly, variation might have been caused by parasite characteristics not related to the classic subdivision into forest and savanna strains. Herder [35] concluded that the parasite strains circulating in the Faro and Mbam areas were related but distinct from the strains from Vina and Lekié, based on phylogenetic linkage patterns. However, this pattern was not confirmed by our analysis as the association between nodule and mf prevalence in Faro was very similar to the other areas but Mbam.

Our model could be used as a tool for assessing the prospects of elimination of onchocerciasis or the burden of onchocercal disease when pre-control nodule prevalence in adult males is the only measure of infection available (as is the case for most of Africa). With our model, an estimate of pre-control mf prevalence may be derived from pre-control nodule prevalence data. Such an estimate may be helpful for program planning, providing an indication of minimum program duration (with regard to prospects of elimination), and could be helpful in the interpretation of ongoing epidemiological parasitological surveys that rely on the skin snipping method (in terms of progress towards elimination). Prospects of elimination may be evaluated by comparing the model-derived estimate of mf prevalence to known trends of infection levels in other foci with a similar history of mass treatment, or by means of dynamic modeling of the effect of mass treatments with ivermectin using onchocerciasis transmission models such as ONCHOSIM [10,11,36] and others [37,38,39]. Progress towards elimination could be evaluated by comparing current mf prevalences with model-derived estimates of pre-control mf prevalence and predicted trends in infection levels based on dynamical modeling. Likewise, the pre-control burden of ocular and dermal morbidity in endemic areas may be estimated based on literature data on the association between mf and disease prevalence [7,8,9]. This would further allow assessment of the impact of control activities on population health, especially when combined with aforementioned dynamic models. If pre-control mf prevalence were to be severely underestimated or overestimated when derived from nodule prevalence data (due to measurement error and geographical variation), this may have important repercussions for the number of treatment rounds that is thought to be required to reach elimination, or the estimated burden of disease. Therefore, it is crucial to consider variation due to sample size and geographical variation in patterns of nodule and mf prevalence when doing this kind of assessment. Given the high level of variation and consequent uncertainty in the association between nodule and mf prevalence, translations should be made carefully and critically evaluated. We recommend that translations of village-level REMO data (based on samples of about 30 adult males) to mf prevalence are made based on the black lines in Figure 3 (which include uncertainty due to measurement error and geographical variation). In case of suspected high exposure of children to flies' bites, it may be more appropriate to apply the part of the model that mimics the observations in Mbam, Cameroon (grey lines in Figure 3). For areas where infection prevalence is known to be homogeneously distributed, REMO samples from multiple villages could be pooled into a more precise estimate of pre-control nodule prevalence in the area, allowing more precise prediction of the pre-control mf prevalence. In Text S1, section "Model application", we explain in more detail how our model should be applied to convert nodule prevalence to mf prevalence (e.g. how to make predictions for a group of villages).

In conclusion, we provide a tool to convert nodule prevalence in adult males to mf prevalence in the general population, which accounts for uncertainty due to measurement error and geographical variation. This tool allows interpretation of a large amount of pre-control data on levels of infection in Africa which

may a) be combined with information on coverage of mass treatment to assess the feasibility of elimination of onchocerciasis and b) enable estimation of disease burden. Furthermore, we identified significant geographical variation in mf prevalence and nodule prevalence patterns that warrants further investigation of age-dependent transmission patterns of onchocerciasis.

Supporting Information

Table S1 Weights used to standardize mf prevalences.
(DOC)

Table S2 Parameter estimates of the model, based on Bayesian hierarchical multivariate logistic regression.
(DOC)

Text S1 Detailed description of the statistical model, and the methods used to estimate the model parameters with an explanation of how the model should be applied to data outside the current study, and providing the code that was used to specify the model in JAGS.
(PDF)

Acknowledgments

The authors would like to thank Professor Emmanuel Lesaffre (Erasmus MC, University Medical Center, Rotterdam, The Netherlands) and Martin Walker (Imperial College, London, UK) for their advice on Bayesian statistics. We also thank the peer-reviewers for their helpful feedback. We would further like to acknowledge all persons who helped in collecting the data that were used in this study.

Author Contributions

Conceived and designed the experiments: LEC WAS SJdV SOH MGB SDSP MB MEM. Performed the experiments: LEC. Analyzed the data: LEC. Contributed reagents/materials/analysis tools: SDSP MB SC PUF AOA JHFR KYD. Wrote the paper: LEC SDSP SOH SC AOA PUF JHFR KYD MEM SJdV MGB WAS MB.

References

1. Coffeng LE, Stolk WA, Zoure HG, Veerman JL, Agblewonu KB, et al. (2013) African programme for onchocerciasis control 1995–2015: model-estimated health impact and cost. PLoS Negl Trop Dis 7: e2032.
2. Dadzie Y, Neira M, Hopkins D (2003) Final report of the Conference on the eradicability of Onchocerciasis. Filaria J 2: 2.
3. Diawara L, Traoré MO, Badji A, Bissan Y, Doumbia K, et al. (2009) Feasibility of onchocerciasis elimination with ivermectin treatment in endemic foci in Africa: first evidence from studies in Mali and Senegal. PLoS Negl Trop Dis 3: e497.
4. Traoré MO, Sarr MD, Badji A, Bissan Y, Diawara L, et al. (2012) Proof-of-principle of onchocerciasis elimination with ivermectin treatment in endemic foci in Africa: final results of a study in Mali and Senegal. PLoS Negl Trop Dis 6: e1825.
5. Tekle AH, Elhassan E, Isiyaku S, Amazigo UV, Bush S, et al. (2012) Impact of long-term treatment of onchocerciasis with ivermectin in Kaduna State, Nigeria: first evidence of the potential for elimination in the operational area of the African Programme for Onchocerciasis Control. Parasit Vectors 5: 28.
6. Mackenzie CD, Homeida MM, Hopkins AD, Lawrence JC (2012) Elimination of onchocerciasis from Africa: possible? Trends Parasitol 28: 16–22.
7. Dadzie KY, Remme J, Baker RH, Rolland A, Thylefors B (1990) Ocular onchocerciasis and intensity of infection in the community. III. West African rainforest foci of the vector Simulium sanctipauli. Trop Med Parasitol 41: 376–382.
8. Dadzie KY, Remme J, Rolland A, Thylefors B (1989) Ocular onchocerciasis and intensity of infection in the community. II. West African rainforest foci of the vector Simulium yahense. Trop Med Parasitol 40: 348–354.
9. Remme J, Dadzie KY, Rolland A, Thylefors B (1989) Ocular onchocerciasis and intensity of infection in the community. I. West African savanna. Trop Med Parasitol 40: 340–347.
10. Plaisier AP, Alley ES, van Oortmarssen GJ, Boatin BA, Habbema JD (1997) Required duration of combined annual ivermectin treatment and vector control in the Onchocerciasis Control Programme in west Africa. Bull World Health Organ 75: 237–245.
11. Winnen M, Plaisier AP, Alley ES, Nagelkerke NJ, van Oortmarssen G, et al. (2002) Can ivermectin mass treatments eliminate onchocerciasis in Africa? Bull World Health Organ 80: 384–391.
12. Ngoumou P, Walsh JF (1993) A Manual for Rapid Epidemiological Mapping of Onchocerciasis (TDR/TDE/ONCHO/93.4). Geneva: UNDP/World Bank/WHO.
13. Noma M, Nwoke BE, Nutall I, Tambala PA, Enyong AP, et al. (2002) Rapid epidemiological mapping of onchocerciasis (REMO): its application by the African Programme for Onchocerciasis Control (APOC). Ann Trop Med Parasitol 96 Suppl 1: S29–39.
14. Remme JHF (2004) The Global Burden of Onchocerciasis in 1990. Geneva: WHO.
15. Vivas-Martínez S, Basáñez MG, Botto C, Villegas L, García M, et al. (2000) Parasitological indicators of onchocerciasis relevant to ivermectin control programmes in the Amazonian focus of Southern Venezuela. Parasitology 121: 527–534.
16. Kipp W, Bamhuhiiga J (2002) Validity of nodule palpation in a Simulium neavei-transmitted onchocerciasis area in Uganda. Am J Trop Med Hyg 67: 128–131.
17. Whitworth JA, Gemade E (1999) Independent evaluation of onchocerciasis rapid assessment methods in Benue State, Nigeria. Trop Med Int Health 4: 26–30.
18. WHO (1992) Methods for the community diagnosis of onchocerciasis to guide ivermectin-based control in Africa (TDR/TDR/ONCHO/92.2). Ouagadougou, 19–21 November 1991: World Health Organization.
19. Fischer P, Kipp W, Bamuhiga J, Binta-Kahwa J, Kiefer A, et al. (1993) Parasitological and clinical characterization of *Simulium neavei*-transmitted onchocerciasis in western Uganda. Trop Med Parasitol 44: 311–321.
20. Traoré-Lamizana M, Lemasson J-J (1987) Participation to a feasability study for the onchocerciasis control in the Logone Basin area. *Simulium damnosum* complex species distribution in the Cameroonian area of the project. [Participation à une étude de faisabilité d'une campagne de lutte contre l'onchocercose dans la région de bassin de Logone. Répartition des espèces du complexe *Simulium damnosum* dans la zone camerounaise du projet.]. Cah ORSTOM 25: 171–186.
21. Vajime CG, Gregory WG (1990) Species complex of vectors and epidemiology. Acta Leiden 59: 235–252.
22. Quillevere D, Hougard JM, Prud'hom JM (1990) Study of the transmission of onchocerciasis in the surroundings of a refugee camp located in the savanna zone of Cameroon [Étude de la transmission de l'onchocercose aux alentours d'un camp de réfugiés situé en zone de savane du Cameroun]. Ann Soc Belg Med Trop 70: 193–202.
23. Traoré-Lamizana M, Somiari S, Mafuyai HB, Vajime CG, Post RJ (2001) Sex chromosome variation and cytotaxonomy of the onchocerciasis vector *Simulium squamosum* in Cameroon and Nigeria. Med Vet Entomol 15: 219–223.
24. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006) Measurement error in nonlinear models: a modern perspective. London: Chapman and Hall.
25. Zimmerman PA, Dadzie KY, De Sole G, Remme J, Alley ES, et al. (1992) *Onchocerca volvulus* DNA probe classification correlates with epidemiologic patterns of blindness. J Infect Dis 165: 964–968.
26. Albiez EJ, Büttner DW, Duke BO (1988) Diagnosis and extirpation of nodules in human onchocerciasis. Trop Med Parasitol 39 Suppl 4: 331–346.
27. Duerr HP, Raddatz G, Eichner M (2008) Diagnostic value of nodule palpation in onchocerciasis. Trans R Soc Trop Med Hyg 102: 148–154.
28. R Core Development Team (2011) R: a language and environment for statistical computing.: R Foundation for Statistical Computing. Vienna, Austria.
29. Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A (2002) Bayesian measures of model complexity and fit. J R Statist Soc B 64: 583–639.
30. Green MJ, Medley GF, Browne WJ (2009) Use of posterior predictive assessments to evaluate model fit in multilevel logistic regression. Vet Res 40: 30.
31. Marshall EC, Spiegelhalter DJ (2003) Approximate cross-validatory predictive checks in disease mapping models. Stat Med 22: 1649–1660.
32. Le Berre R, Balay G, Brengues J, Coz J (1964) Biology and ecology of the female of *Simulium damnosum* Theobald, 1903, as a function of the bioclimatic zones of West Africa. Influence on the epidemiology of onchocerciasis [Biologie et l'écologie de la femelle de *Simulium damnosum* Theobald, 1903, en fonction des zones bioclimatiques d'Afrique Occidentale. Influence sur l'epid'emiologie de l'onchocercose]. Bull World Health Organ 31: 843–855.
33. Migliani R, Louis JP, Auduge A, Trebucq A, Gelas H (1993) Evaluation of visual impairment and blindness in Cameroon. A survey in a forest area. [Évaluation de la malvoyance et des cécités au Cameroun. Enquête en milieu rural forestier]. Cahier Santé 3: 17–23.
34. Babalola OE, Ogbuagu FK, Maegga BT, Braide EI, Magimbi C, et al. (2011) African programme for onchocerciasis control: ophthalmological findings in bushenyi, Uganda. West Afr J Med 30: 104–109.
35. Herder S (1994) The genetic variation of *Onchocerca volvulus* (Leukart, 1893) Relation with the eipdemiological profile of onchocerciasis [Variabilité génétique d'*Onchocerca volvulus* (Leukart, 1893). Relations avec les faciès épidémiologiques de l'onchocercose]. Montpellier: Université de Montpellier II.
36. Plaisier AP, van Oortmarssen GJ, Habbema JD, Remme J, Alley ES (1990) ONCHOSIM: a model and computer simulation program for the transmission and control of onchocerciasis. Comput Methods Programs Biomed 31: 43–56.

37. Filipe JA, Boussinesq M, Renz A, Collins RC, Vivas-Martínez S, et al. (2005) Human infection patterns and heterogeneous exposure in river blindness. Proc Natl Acad Sci U S A 102: 15265–15270.
38. Poolman EM, Galvani AP (2006) Modeling targeted ivermectin treatment for controlling river blindness. Am J Trop Med Hyg 75: 921–927.
39. Duerr HP, Raddatz G, Eichner M (2011) Control of onchocerciasis in Africa: threshold shifts, breakpoints and rules for elimination. Int J Parasitol 41: 581–589.

Onchocerciasis: the pre-control association between prevalence of palpable nodules and skin microfilariae

Luc E. Coffeng,[1,a,*] Sébastien D.S. Pion,[2,a] Simon O'Hanlon,[3] Simon Cousens,[4] Adenike O. Abiose,[5] Peter U. Fischer,[6] Jan H.F. Remme,[7] K.Yankum Dadzie,[8] Michele E. Murdoch,[9] Sake J. de Vlas,[1] María-Gloria Basáñez,[3] Wilma A. Stolk,[1,b] Michel Boussinesq[2,b]

[1] Department of Public Health, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands; [2] UMI 233, Institut de Recherche pour le Développement (IRD) and University of Montpellier 1, 911 Avenue Agropolis, BP 64501, F-34394 Montpellier cedex 5, France; [3] Department of Infectious Disease Epidemiology, School of Public Health, Faculty of Medicine (St Mary's Campus), Imperial College London, Norfolk Place, London W2 1PG, UK; [4] Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel St, London WC1 E 7HT, UK; [5] Sightcare International, P.O. Box 29771, Secretariat Main Office, Ibadan, Oyo State, Nigeria; [6] Washington University School of Medicine, Infectious Disease Division, Campus Mailbox 8051, 660 South Euclid Avenue, St. Louis, MO 63110, USA; [7] Consultant, 120 Rue des Campanules, 01210 Ornex, France; [8] Consultant, P.O. Box OS-1905, Accra, Ghana; [9] Department of Dermatology, Watford General Hospital, Watford, Hertfordshire WD18 0HB, UK

[a,b] These authors contributed equally to this work

* Corresponding author: Department of Public Health, Erasmus MC, University Medical Center Rotterdam, P.O. box 2040, 3000 CA Rotterdam, The Netherlands; l.coffeng@erasmusmc.nl, luccoffeng@gmail.com; tel. +31 10 70 38357, fax. +31 10 70 38474

Table S1: Weights used to standardize prevalence of microfilariae in the skin. Standardization weights were based on the reference population of the Onchocerciasis Control Programme.

Age	Male	Female
5–9	0.091	0.078
10–14	0.090	0.077
15–29	0.129	0.138
30–49	0.123	0.146
≥ 50	0.063	0.064

Onchocerciasis: the pre-control association between prevalence of palpable nodules and skin microfilariae

Luc E. Coffeng,[1,a,*] Sébastien D.S. Pion,[2,a] Simon O'Hanlon,[3] Simon Cousens,[4] Adenike O. Abiose,[5] Peter U. Fischer,[6] Jan H.F. Remme,[7] K.Yankum Dadzie,[8] Michele E. Murdoch,[9] Sake J. de Vlas,[1] María-Gloria Basáñez,[3] Wilma A. Stolk,[1,b] Michel Boussinesq[2,b]

[1] Department of Public Health, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands; [2] UMI 233, Institut de Recherche pour le Développement (IRD) and University of Montpellier 1, 911 Avenue Agropolis, BP 64501, F-34394 Montpellier cedex 5, France; [3] Department of Infectious Disease Epidemiology, School of Public Health, Faculty of Medicine (St Mary's Campus), Imperial College London, Norfolk Place, London W2 1PG, UK; [4] Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel St, London WC1 E 7HT, UK; [5] Sightcare International, P.O. Box 29771, Secretariat Main Office, Ibadan, Oyo State, Nigeria; [6] Washington University School of Medicine, Infectious Disease Division, Campus Mailbox 8051, 660 South Euclid Avenue, St. Louis, MO 63110, USA; [7] Consultant, 120 Rue des Campanules, 01210 Ornex, France; [8] Consultant, P.O. Box OS-1905, Accra, Ghana; [9] Department of Dermatology, Watford General Hospital, Watford, Hertfordshire WD18 0HB, UK

[a,b] These authors contributed equally to this work
* Corresponding author: Department of Public Health, Erasmus MC, University Medical Center Rotterdam, P.O. box 2040, 3000 CA Rotterdam, The Netherlands; l.coffeng@erasmusmc.nl, luccoffeng@gmail.com; tel. +31 10 70 38357, fax. +31 10 70 38474

Table S2. Model parameter estimates from Bayesian hierarchical multivariate logistic regression of infection prevalence data. The model predicts the joint distribution of prevalence of nodules in adult males (age ≥20) and presence of microfilariae (mf) in the skin of the general population (age ≥5).

Parameter[a]	Interpretation	Median[b]	Lower bound[c]	Upper bound[d]
$\text{logit}^{-1}\left(\beta_{0,m=1}\right)$	Average fraction of general population with mf in the skin (excluding Mbam)	0.68	0.55	0.78
$\text{logit}^{-1}\left(\beta_{0,m=2}\right)$	Average fraction of adult males with onchocercal nodules (excluding Mbam)	0.51	0.36	0.67
$\exp\left(\beta_{1,m=1}\right)$	Odds ratio of presence of mf in the skin in Mbam compared to other areas	4.17	1.04	16.69
$\exp\left(\beta_{1,m=2}\right)$	Odds ratio of presence of nodules in Mbam compared to other areas	2.69	0.45	14.57
$\sigma_{ij,m=1}$	Standard deviation of log odds of presence of mf within geographical areas	0.98	0.87	1.11
$\sigma_{ij,m=2}$	Standard deviation of log odds of presence of nodules within geographical areas	0.89	0.77	1.03
ρ_{ij}	Correlation of log odds of presence of nodules and mf within geographical areas	0.84	0.77	0.90
$\sigma_{j,m=1}$	Standard deviation of average log odds of presence of mf between geographical areas	0.55	0.22	1.24
$\sigma_{j,m=2}$	Standard deviation of average log odds of presence of nodules between geographical areas	0.69	0.31	1.50
ρ_{j}	Correlation of average log odds of presence of nodules and mf between geographical areas	0.88	0.28	1.00
specificity	One minus the probability of misclassifying a subcutaneous nodule as onchocercal	0.99	0.98	1.00

[a] For ease of interpretation, parameter estimates have been transformed to an intuitive scale, where possible (inverse logit transformation for intercepts and exponents for other fixed effects parameters). See Appendix A for a detailed description of the model and its parameters.
[b] Median of posterior distribution.
[c] Defined as the 2.5th percentile of the posterior distribution.
[d] Defined as the 97.5th percentile of the posterior distribution.

Onchocerciasis: the pre-control association between prevalence of palpable nodules and skin microfilariae

Luc E. Coffeng,[1,a,*] Sébastien D.S. Pion,[2,a] Simon O'Hanlon,[3] Simon Cousens,[4] Adenike O. Abiose,[5] Peter U. Fischer,[6,7] Jan H.F. Remme,[8] K.Yankum Dadzie,[9] Michele E. Murdoch,[10] Sake J. de Vlas,[1] María-Gloria Basáñez,[3] Wilma A. Stolk,[1,b] Michel Boussinesq[2,b]

[1] Department of Public Health, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands; [2] UMI 233, Institut de Recherche pour le Développement (IRD) and University of Montpellier 1, 911 Avenue Agropolis, BP 64501, F-34394 Montpellier cedex 5, France; [3] Department of Infectious Disease Epidemiology, School of Public Health, Faculty of Medicine (St Mary's Campus), Imperial College London, Norfolk Place, London W2 1PG, UK; [4] Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel St, London WC1 E 7HT, UK; [5] Sightcare International, P.O. Box 29771, Secretariat Main Office, Ibadan, Oyo State, Nigeria; [6] Washington University School of Medicine, Infectious Disease Division, Campus Mailbox 8051, 660 South Euclid Avenue, St. Louis, MO 63110, USA; [7] Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany; [8] Consultant, 120 Rue des Campanules, 01210 Ornex, France; [9] Consultant, P.O. Box OS-1905, Accra, Ghana; [10] Department of Dermatology, Watford General Hospital, Watford, Hertfordshire WD18 0HB, UK

[a,b] These authors contributed equally to this work

* Corresponding author: Department of Public Health, Erasmus MC, University Medical Center Rotterdam, P.O. box 2040, 3000 CA Rotterdam, The Netherlands; l.coffeng@erasmusmc.nl, luccoffeng@gmail.com; tel. +31 10 70 38357, fax. +31 10 70 38474

Text S1

Contents

Model description 1
Parameter estimation 3
Model application 5
Model specification in JAGS 7
References 8

Model description

The current study considers spatially clustered, bivariate binomially distributed data. For analysis of spatial data, there are geostatistical techniques that take account of the spatial correlation. These techniques require knowledge of the geographical coordinates of the data points, which were not fully available in our case. Instead, we used an ordinary hierarchical approach to model the data to account for spatial correlation.

There are several popular alternatives for modeling bivariate binomially distributed data. Multivariate probit regression models have been proposed. These models are convenient in terms of computational requirements because they only quantify the correlation between observations, leaving variance at the lowest level of the data unidentified (but accounted for) [1]. Depending on the researcher's objective, this can be considered an advantage (computationally less demanding) or a drawback (part of the model remains unidentified). Another drawback is that the interpretation of probit models may be less intuitive to some, if not many researchers. As a concession to these arguments, a

reparameterization of the multivariate logistic regression model has been proposed, which like the probit model leaves the variance at the lowest level unidentified [1,2]. However, this reparameterized logistic model requires a customized sampling algorithm for efficient Markov chain Monte Carlo sampling. In our case, we chose for intuitive interpretation, quantification of all variances, and the use of the freely available and widely implemented Gibbs sampling algorithm, leading to the choice of generalizing the familiar logistic model as described below.

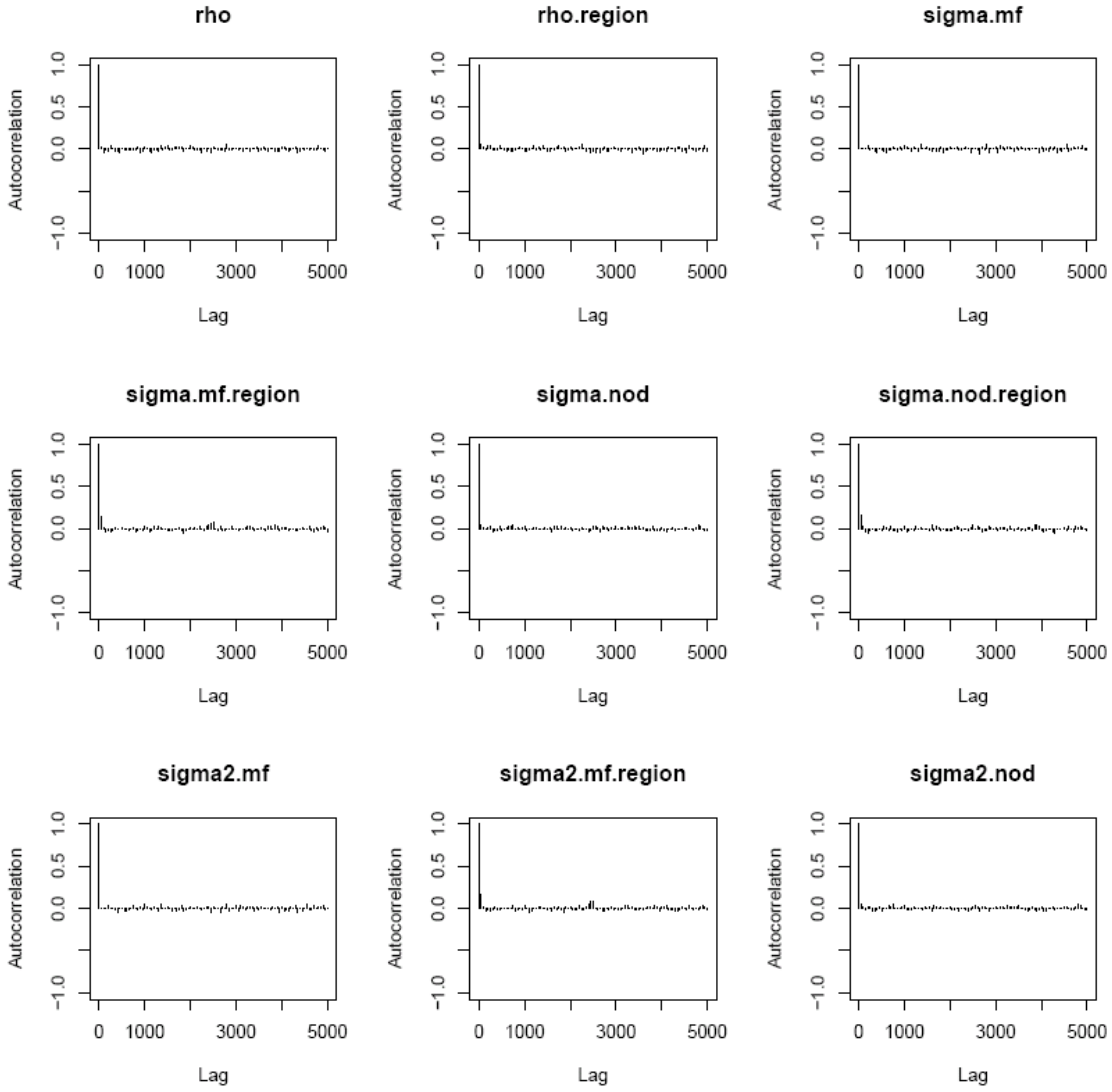
The multivariate model used in this study is an extension of the hierarchical logistic regression model $\text{logit}\left(\pi_{ij}\left(y_{ij}=k_{ij}\mid \text{X}_{ij},\beta,n_{ij}\right)\right)=\text{X}_{ij}^{T}\beta+\varepsilon_{ij}+\varepsilon_{j}$, where π_{ij} is the probability of finding k cases of the binomially distributed outcome y in n individuals from the *i*-th unit in the *j*-th cluster, conditional on a set of observed covariates X_{ij} and a set of model parameters β. The error terms ε_{ij} and ε_{j} represent the variation within and between the *j* clusters of observation, respectively. We extended this model to simultaneously predict *m* binary outcomes, leading to $\text{logit}\left(\pi_{ij,m}\left(y_{ij,m}=k_{ij,m}\mid \text{X}_{ij},\beta_{m},n_{ij,m}\right)\right)=\text{X}_{ij}^{T}\beta_{m}+\varepsilon_{ij}+\varepsilon_{j}$, where $\pi_{ij,m}$ is the probability of observing k cases of the *m*-th outcome (*m* = 1: presence of microfilariae in the skin; *m* = 2: presence of nodules in adult males) among n_{m} observed individuals from the *i*-th unit (village) in the *j*-th cluster (geographical area). Here, the error terms ε_{ij} and ε_{j} each consist of *m* components representing the variation in log odds of each of the *m* outcomes within and between the *j* clusters of observations. For each observation, there is a set of observed covariates X_{ij} (bioclime), and for each of the *m* predicted binary outcomes we have a set model parameters β_{m}. In our case, the intercepts $\beta_{0,m=1}$ and $\beta_{0,m=2}$ represent the mean log odds of presence of mf and nodule in the data, respectively. The parameters $\beta_{1,m=1}$ and $\beta_{1,m=2}$ represent the log odds ratio of observing presence of microfilariae in the skin and subcutaneous onchocercal nodules in a certain bioclime, respectively, relative to a reference bioclime (multiple sets of such parameter can be added to stratify the analysis by multiple bioclimes and/or other characteristics). Correlation between onchocercal nodule and mf prevalence was modeled by assuming multivariate normal (MVN) distributions for the error terms: $\varepsilon_{ij}\sim MVN\left(\mu_{\varepsilon_{ij}},\Sigma_{\varepsilon_{ij}}\right)$ and $\varepsilon_{j}\sim MVN\left(\mu_{\varepsilon_{j}},\Sigma_{\varepsilon_{j}}\right)$, with $\mu_{\varepsilon_{ij}}=\mu_{\varepsilon_{j}}=(0,0)$. Here, $\Sigma_{\varepsilon_{ij}}$ and $\Sigma_{\varepsilon_{j}}$ are variance-covariance matrices with size *m* x *m*, containing along the diagonal the marginal variances of the errors for the log odds of

presence of nodules and mf within ($\sigma^2_{ij,m=1}$ and $\sigma^2_{ij,m=2}$) and between ($\sigma^2_{j,m=1}$ and $\sigma^2_{j,m=2}$) the *j* clusters of observations. The off-diagonal positions of $\Sigma_{\varepsilon_{ij}}$ and Σ_{ε_j} hold the covariances $\sigma_{ij,m=1;m=2}$ and $\sigma_{j,m=1;m=2}$ of error terms within and between the *j* clusters, respectively. The correlation between log odds of presence of nodules and mf at village-level ρ_{ij} was derived by $\frac{\sigma_{ij,m=1;m=2}}{\sigma_{ij,m=1}\sigma_{ij,m=2}}$. Correlation ρ_j was derived in a similar fashion, and can be interpreted in two ways; 1) together with variances $\sigma^2_{j,m=1}$ and $\sigma^2_{j,m=2}$, ρ_j represents how the association between onchocercal nodule and mf prevalences varies between geographical regions due to e.g. environmental factors and surveys methods (analogous to linear regression models with a random intercept); 2) ρ_j is the correlation between the mean log odds of presence of nodules and mf in a geographical area (as defined for the data in this study).

Parameter estimation

Model parameters were estimated assuming non-informative prior distributions. For fixed effects parameters β_m, we assumed independent normal prior distributions $N_{\beta_m}(0,1000)$. The village-level variance-covariance matrix $\Sigma_{\varepsilon_{ij}}$ was estimated assuming a scaled Wishart prior distribution $W_m(R,k)$ for its inverse $\Sigma^{-1}_{\varepsilon_{ij}}$, where R is the *m* x *m* identity matrix I_m, and k is the number of degrees of freedom (set to 3, effectively assuming uniform prior information on ρ_{ij}). To maximize the speed of model convergence, the variance-covariance matrix Σ_{ε_j} for differences between geographical areas was hierarchically centered around fixed effects β_m, and was estimated assuming independent uniform prior distributions for the correlation ($\rho_j \sim U(-1,1)$), and standard deviations ($\sigma_{j,m} \sim U(0,10)$ or $\sigma_{j,m} \sim U(0,100)$), in line with previous recommendations for estimating hyperparameters [3]. The prior distribution for minimum sensitivity of nodule palpation (at low endemicity levels) was defined as being uniform between 60% and 100%. The prior distribution for specificity of nodule palpation was defined as being uniform between 98% and 100%.

Figure A1. Autocorrelation plots of Monte Carlo samples for nine parameters. In this study, autocorrelation was initially high for some parameters, indicating that the Gibbs sampling algorithm was slow in exploring the posterior distribution of these parameters. Autocorrelation was reduced by storing only every 20th Monte Carlo sample and running 200,000 iterations (after discarding an initial 200,000 iterations for burn-in), for which the results are shown here. After this, autocorrelation of Monte Carlo samples was similarly low for all parameters and Markov chains.



Model parameters were estimated using four Markov chains with each 400,000 Monte Carlo simulations. For each chain, the first 200,000 of the saved simulations were considered as burn-in simulations and discarded. Such a number of simulations was necessary as the Gibbs sampler explored the joint posterior distribution of parameters slowly, indicated by high autocorrelation of Monte Carlo samples. To save storage space, only every 20th Monte Carlo sample was stored,

effectively reducing autocorrelation (Figure A1). The effective amount of simulations per parameter was 40,000 (sum of four Markov chains). The point estimate for each parameter and prediction were taken to be the median of the 40,000 simulations. Ninety five percent Bayesian credible intervals were calculated as the 2.5th and 97.5th percentiles of the simulations.

Model convergence was assessed by checking whether the four Markov chains converged to the same posterior distribution for each parameter, based on Gelman and Rubin's convergence diagnostic, the potential scale reduction factor (which should be below 1.1; i.e. a posterior credible interval of a parameter estimate should not become more than 10% narrower if more Monte Carlo samples were drawn) [4]. Because this diagnostic test requires that starting values for each parameter in each Markov Chain are over-dispersed with respect to the true posterior distribution of a parameter, we assigned heavily over-dispersed initial values to each of the model parameters in each chain (e.g. the initial values 1, 10, 50, and 100 for a parameter with an uninformative normal distribution as prior, one value for each Markov chain). In our simulations, the potential scale reduction factor was at or below 1.001 for all parameters. Furthermore, we checked that all Markov chains arrived in the joint posterior distribution of parameter values, as determined by means of Geweke's test, which compares the distribution of the first 10% and last 50% of the Monte Carlo samples within a chain [5]. We also checked that Monte Carlo errors were small, relative to the point estimate of each parameter (difference of at least factor 100 – 1000).

Model application

Given estimates of β_m, $\Sigma_{\varepsilon_{ij}}$, and Σ_{ε_j}, we estimated the conditional distribution of mf prevalence $\pi^*_{ij,m=1} \mid \pi^*_{ij,m=2}$ in a hypothetical village *i* from an unspecified region *j* outside the dataset, given an estimate of the 'true' onchocercal nodule prevalence in adult males, corrected for misclassification of nodules, in the same hypothetical village (assuming this is exactly known). Given that we are working with multivariate normal distributions, the conditional distribution for $\pi^*_{ij,m=1} \mid \pi^*_{ij,m=2}$ can be described as

$$\text{logit}\left(\pi^*_{ij,m=1} \mid \pi^*_{ij,m=2}\right) \sim N\left(\mathrm{X}^T_{ij}\beta^*_{m=1} + \frac{\sigma_{ij,m=1}}{\sigma_{ij,m=2}}\rho_{ij}\left(\text{logit}\left(\pi^*_{ij,m=2}\right) - \mathrm{X}^T_{ij}\beta_{m=2}\right), \left(1-\rho^2_{ij}\right)\sigma^2_{ij,m=1}\right), \quad \text{where}$$

$\beta^*_{m=1} \sim N\left(\beta_{m=1}, \left(1-\rho^2_j\right)\sigma^2_{j,m=1}\right)$. To include uncertainty about nodule prevalence $\pi^*_{ij,m=2}$ in the

prediction of mf prevalence $\pi^*_{ij,m=1}$, we simulated values from the estimated distribution of $\text{logit}\left(\pi^*_{ij,m=2}\right)$, and fed these into the distribution for $\text{logit}\left(\pi^*_{ij,m=1} \mid \pi^*_{ij,m=2}\right)$, from which values for $\pi^*_{ij,m=1}$ were then simulated. As this was done while simultaneously estimating the values of model parameters by means of Markov Chain Monte Carlo sampling, all uncertainty in the model parameter estimates was carried through to the final predictions for mf prevalence in hypothetical villages.

It should be noted that the procedure described above produces predictions pertaining to individual villages only, and therefore will produce predictions containing a great deal of uncertainty. The amount of uncertainty would be substantially lower if predictions were made based on larger samples of adult males, or when made for the mean prevalence of infection in a group of villages. However, a prediction for a mean prevalence would ignore possible heterogeneity in infection prevalences between villages, which may lead to overly optimistic estimates when e.g. when assessing prospects of elimination (the most highly endemic village will determine the required duration of an intervention, not the mean prevalence in a region). Nevertheless, if such predictions are made (e.g. for a group of villages with known similar levels of infection), or concerning many villages (theoretically an infinite number of villages), the mean mf prevalence is described by $\mathrm{X}^T_{ij}\beta^*_{m=1} + \frac{\sigma_{ij,m=1}}{\sigma_{ij,m=2}}\rho_{ij}\left(\text{logit}\left(\pi^*_{ij,m=2}\right) - \mathrm{X}^T_{ij}\beta_{m=2}\right)$, where $\pi^*_{ij,m=2}$ is the mean nodule prevalence in the group of villages (including uncertainty related to overall sample size). However, usually the number of sampled villages is not very high (<1,000, meaning that the denominator of the standard error of the mean is <30, approx. the square root of 1,000), and one should therefore simulate the mf prevalence separately for every village, sampling village-level error independently for every village, and sampling the region-level error simultaneously for all villages. Then, for every set of many repeated simulations (i.e. a set consisting of one simulation for each village), the investigator can calculate the mean or any other summary statistic of the level of infection in the group of villages (e.g. range or variance), arriving at a distribution for the estimated mean or another summary statistic for mf prevalence in a group of villages.

Model specification in JAGS

```
for (i in 1:N) {
       # Likelihood of nodule data
       k.nod[i] ~ dbin(sens.nod.p[i] * p.nod[i] + (1-spec.nod)*(1-p.nod[i]), n.nod[i])
       sens.nod.p[i] <- sens.nod + (1-sens.nod)*p.nod[i]
       logit(p.nod[i]) <- B0[region[i],1] + e.vill[i,1]

       # Likelihood of mf data
       k.mf[i] ~ dbin(p.mf[i],n.mf[i])
       logit(p.mf[i]) <- B0[region[i],2] + e.vill[i,2]

       # Correlation of nodule and mf data with regions
       e.vill[i,1] <- e.vill.raw[i,1] * xi.nod
       e.vill[i,2] <- e.vill.raw[i,2] * xi.mf
       e.vill.raw[i,1:2] ~ dmnorm(Mu,Sigma2.inv.raw)
}

# Priors for fixed effects
sens.nod ~ dunif( [some value] ,1.0) # [some value] = 0.6, 0.8, or 1.0 (final model)
spec.nod ~ dunif(0.98 ,1)
b0.nod ~ dnorm(0,0.001)
b0.mf ~ dnorm(0,0.001)
mbam.nod ~ dnorm(0,0.001)
mbam.mf ~ dnorm(0,0.001)

# Uniform prior for correlation and marginal standard deviations
# of hierarchically centered random region effects
for (j in 1:7) {
       B0[j,1:2] ~ dmnorm(Mu.region[j,1:2],Sigma2.region.inv)
       Mu.region[j,1] <- b0.nod + mbam.nod*equals(j,2)
       Mu.region[j,2] <- b0.mf + mbam.mf*equals(j,2)
}
Sigma2.region.inv <- inverse(Sigma2.region)
Sigma2.region[1,1] <- sigma2.nod.region
Sigma2.region[2,2] <- sigma2.mf.region
Sigma2.region[1,2] <- covar.nod.mf.region
Sigma2.region[2,1] <- covar.nod.mf.region
covar.nod.mf.region <- rho.region * sigma.nod.region * sigma.mf.region
sigma2.nod.region <- pow(sigma.nod.region,2)
sigma2.mf.region <- pow(sigma.mf.region,2)
sigma.nod.region ~ dunif(0,10)
sigma.mf.region ~ dunif(0,10)
rho.region ~ dunif(-1,1)

# Scaled inverse Wishart prior for random village effects
Sigma2.inv.raw ~ dwish(R,scale)
xi.nod ~ dunif(0,100)
xi.mf ~ dunif(0,100)
Sigma2.raw <- inverse(Sigma2.inv.raw)
sigma.nod <- pow(Sigma2.raw[1,1],0.5) * xi.nod
sigma.mf <- pow(Sigma2.raw[2,2],0.5) * xi.mf
rho <- Sigma2.raw[1,2]/sqrt(Sigma2.raw[1,1]*Sigma2.raw[2,2])

Sigma2[1,1] <- pow(sigma.nod,2)
Sigma2[1,2] <- rho * sigma.nod * sigma.mf
Sigma2[2,1] <- Sigma2[1,2]
Sigma2[2,2] <- pow(sigma.mf,2)
```

```
# Predictions for REMO samples
sigma2.mf.REMO <- (1 - rho^2) * pow(sigma.mf,2)
sigma2.mf.REMO.region <- (1 - rho.region^2) * sigma2.mf.region
tau.mf.REMO <- pow(sigma2.mf.REMO,-1)
tau.mf.REMO.region <- pow(sigma2.mf.REMO.region,-1)

for (k in 1:N.REMO) {
       # Hypothetical REMO village: nodule prevalence
       k.nod.REMO[k] ~ dbin(sens.nod.p.REMO[k] * p.nod.REMO[k] +
                                          (1-spec.nod)*(1-p.nod.REMO[k]),n.nod.REMO[k])
       sens.nod.p.REMO[k] <- sens.nod + (1-sens.nod)*p.nod.REMO[k]
       logit(p.nod.REMO[k]) <- b.nod.REMO[k]
       b.nod.REMO[k] ~ dnorm(0,0.001)

       # Hypothetical REMO village: mf prevalence (non-mosaic)
       logit(p.mf.REMO.vill[k]) <- b.mf.REMO.vill[k]
       b.mf.REMO.vill[k] ~ dnorm(b0.mf.REMO.region[k],tau.mf.REMO)
       b0.mf.REMO.region[k] <- b0.mf.REMO.intercept[k] +
                                    (sigma.mf/sigma.nod) * rho * (b.nod.REMO[k] - b0.nod)
       b0.mf.REMO.intercept[k] ~ dnorm(b0.mf,tau.mf.REMO.region)

       # Hypothetical REMO village: mf prevalence (mosaic)
       logit(p.mf.REMO.vill.mosaic[k]) <- b.mf.REMO.mosaic.vill[k]
       b.mf.REMO.mosaic.vill[k] ~ dnorm(b0.mf.REMO.mosaic.region[k],tau.mf.REMO)
       b0.mf.REMO.mosaic.region[k] <- b0.mf.REMO.mosaic.intercept[k] +
                     (sigma.mf/sigma.nod) * rho * (b.nod.REMO[k] - (b0.nod + mbam.nod))
       b0.mf.REMO.mosaic.intercept[k] ~ dnorm((b0.mf + mbam.mf),tau.mf.REMO.region)
}
```

References

1. O'Brien SM, Dunson DB (2004) Bayesian multivariate logistic regression. Biometrics 60: 739-746.

2. Ovaskainen O, Hottola J, Siitonen J (2010) Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. Ecology 91: 2514-2521.

3. Gelman A (2006) Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1: 515-533.

4. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Statistical Science 7: 457-511.

5. Geweke J (1991) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernado J, Berger J, Dawid A, Smith A, editors. Bayesian statistics 4. Oxford, UK: Clarendon Press.