Treatment-seeking behaviour in low- and middle-income countries estimated using a Bayesian model

Victor A Alegana[1, 2*], Jim Wright[1], Carla Pezzulo[1,2], Andrew J. Tatem[1,2], Peter M. Atkinson[1,3,4]

[1] Geography and Environment, University of Southampton, Southampton, UK
[2] Flowminder Foundation, Stockholm, Sweden
[3] Faculty of Science and Technology, Lancaster University, UK
[4] School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, UK

Author Information
Victor A Alegana[1,2*]; V.A.Alegana@soton.ac.uk or aleganav@gmail.com
Peter M. Atkinson; pma@lancaster.ac.uk
Jim Wright; J.A.Wright@soton.ac.uk
Carla Pezzulo; C.Pezzulo@soton.ac.uk
Andrew J. Tatem, A.J.Tatem@soton.ac.uk

*Correspondence: V.A.Alegana@soton.ac.uk

Keywords: Bayesian hierarchical model – Treatment-seeking behaviour - Item Response Theory - Markov Chain Monte Carlo

Abstract

Background

Seeking treatment in formal healthcare for uncomplicated infections is vital to combating disease in low- and middle-income countries (LMICs). Healthcare treatment-seeking behaviour varies within and between communities and is modified by socio-economic, demographic, and physical factors. As a result, it remains a challenge to quantify healthcare treatment-seeking behaviour using a metric that is comparable across communities. Here, we present an application for transforming individual categorical responses (actions related to fever) to a continuous probabilistic estimate of fever treatment for one country in Sub-Saharan Africa (SSA).

Methods

Using nationally representative household survey data from the 2013 Demographic and Health Survey (DHS) in Namibia, individual-level responses (*n*=1,138) were linked to theoretical estimates of travel time to the nearest public or private health facility. Bayesian Item Response Theory (IRT) models were fitted via Markov Chain Monte Carlo (MCMC) simulation to estimate parameters related to fever treatment and estimate probability of treatment for children under five years. Different models were implemented to evaluate computational needs and the effect of including predictor variables such as rurality. The mean treatment rates were then estimated at regional level.

Results

Modelling results suggested probability of fever treatment was highest in regions with relatively high incidence of malaria historically. The minimum predicted threshold probability of seeking treatment was 0.3 (model 1: 0.340; 95% CI 0.155-0.597), suggesting that even in populations at large distances from facilities, there was still a

30% chance of an individual seeking treatment for fever. The agreement between correctly predicted probability of treatment at individual level based on a subset of data (*n*=247) was high (AUC= 0.978), with a sensitivity of 96.7% and a specificity of 75.3%.

Conclusion

We have shown how individual responses in national surveys can be transformed to probabilistic measures comparable at population level. Our analysis of household survey data on fever suggested a 30% baseline threshold for fever treatment in Namibia. However, this threshold level is likely to vary by country or endemicity. Although our focus was on fever treatment, the methodology outlined can be extended to multiple health seeking behaviours captured in routine national survey data and to other infectious diseases.

Background

Delay in seeking treatment for ill health in low- and middle-income countries (LMICs) affects disease progression, management and outcomes [1-3]. Most infectious diseases in LMICs are preventable by using cost-effective interventions and treatable at peripheral health facilities [4]. However, weak health systems affect the delivery of most interventions [5] and socio-economic and physical barriers that modify health-seeking behaviour compound this, leading to under-utilisation of health facilities [6]. Encouraging appropriate treatment-seeking behaviour for uncomplicated infections is vital to further reduce disease burden in these countries or for successful elimination. For malaria, for example, the current World Health Organisation (WHO) recommendation is for malaria treatment to be sought in the formal healthcare sector within 24 hours of fever onset and other malaria-related symptoms [7]. This is because patients who seek treatment through the formal sector are likely to receive an appropriate diagnosis and effective management [8]. However, there are many factors influencing population treatment-seeking behaviour including, but not limited to; availability of healthcare providers, proximity or travel time to healthcare facilities, condition severity and perception, and the socio-demographic profile of the population at risk [9].

Studies on treatment-seeking behaviour can be grouped into two categories of approach. The first is a qualitative description of steps undertaken by the population in different settings [10-12] while the second is a quantitative association between determinants (factors) and choice of health service use [13-18]. Although these approaches are used widely in bio-medical research, they usually do not examine the latent (i.e. theoretical) characteristics such as individual-level traits to estimate variation at population level. In addition, comparability is not simply guaranteed

with the same questionnaire because of differential item functioning problem i.e. the varying behavioural response to the same question depending on the respondent [19]. Such variation can then be translated to spatially explicit applications that can be combined with existing spatial data on populations [20] and disease incidence to inform and optimise targeting of community-based interventions.

Model-based geostatistical methods have already been used to predict and estimate disease incidence at fine spatial resolution [21, 22]. This has been aided by public health intelligence data that are increasingly becoming available across space and time from geo-located nationally representative household surveys. These include the Malaria Indicator Surveys (MIS) [23], Demographic and Health Surveys (DHS) [24], and Multiple Indicator Cluster Surveys (MICS) [25]. These nationally representative household surveys also collect information on self-reported health behaviour such as fever management [14]. However, how can responses concerning fever treatment from household surveys be compared across populations with varying access, demographics, cultures, and disease burdens? Item response theory (IRT) has been widely used to examine surveys items (questions) and person characteristics in psychology and education [26-28]. In education, for example, it has been used to estimate the person-level traits (such as ability) or item-level difficulty in an examination [29-31]. IRT concepts can be extended to health as applied previously in delirium screening [32], longitudinal data analysis [33], and interpreting medical codes from patient records [34]. IRT approaches are essentially probit models with additional regression effects used to aid estimation of item characteristics [35]. Extending this to a Bayesian framework has the advantages of incorporating uncertainty in estimating latent traits and prior distributions can be imposed on the Bayesian probability model to capture many aspects of data not

included in descriptive or quantitative frequentist approaches [36]. Although Imputation techniques can be used to handle missing data, this was beyond the current scope of this manuscript.

Here, the aim was to demonstrate the application of IRT to fever treatment-seeking modelling using data from a low malaria transmission setting, the Namibia 2013 DHS. We analyse fever treatment-seeking behaviour at a national level and derive response characteristic curves based on travel times to the nearest facilities. The rest of this paper is organised as follows. Section 2 provides an overview of household survey data in LMICs and the proposed modelling approach. We then present treatment-seeking behaviour model outputs in section 3, including evaluation of model performance. The paper concludes with a brief discussion in sections 4 and 5.

Methods

Data characteristics in low- and middle income countries

Distance or proximity to healthcare provider is an important parameter in the choice of treatment by patients in many LMICs [37-39]. In these countries, the majority of people access facilities by walking. Therefore, it is preferable to use a facility close to the place of residence because it is less costly compared to travelling greater distances requiring motorised transport [40]. Other factors that influence utilisation patterns include: age, gender, healthcare costs, socio-economic status, residence (urban or rural), familiarity with health personnel, fever severity, and quantity as well as quality of services at peripheral facilities [41, 42]. In some cases, however, the phenomenon of by-passing the nearest healthcare facility can be encountered, even for mild fever conditions [43, 44]. Empirical data are not always available to model such nuances and we therefore assume use of the nearest facility in this case study.

115

116 *Estimation of travel times to the nearest formal healthcare treatment provider.*

117 Estimating travel times between population centres and formal healthcare providers
118 has already been considered in previous research [14]. In brief, this requires a
119 combination of mode of travel (walking or motorised) and an impedance surface
120 that is constructed based on multiple data layers, including the various land use and
121 land cover characteristics, elevation, and roads [45]. Travel time to nearest
122 healthcare facility is a useful measure because it is relatively easy to estimate and to
123 relate travel times in different settings compared to estimating the actual physical
124 distance. The approach in Alegana et al. [14] shows how travel times for Namibia
125 were derived.

126 **<Insert fig 1>**

127 *Quantification of formal healthcare use based on national representative household*
128 *surveys*

129 To estimate the utilisation of healthcare facilities, this study used the reported use of
130 formal healthcare for fever treatment from the DHS. These surveys are conducted in
131 90 countries worldwide, and 44 in SSA, providing information on reproductive health,
132 fertility, population demographics and general health status, nutrition, household
133 characteristics, socio-economic status and infant and child mortality rates [46]. The
134 surveys are based on a random two-stage cluster sampling design in which clusters
135 are usually first sampled within a region on a probability-proportional-to-size basis
136 and thereafter, within each cluster, households are sampled randomly [47, 48].
137 Cluster sizes usually vary, but are typically approximately 15 to 30 households. The
138 household survey provides information on health and the socio-demographic profile
139 of consenting participants including their treatment-seeking behaviour for conditions
140 such as malaria-associated fever.

A notable feature of the fever treatment variable in the DHS is the decay in treatment with increasing travel time to nearest facility (Figure 1). The geographical barrier to utilisation, manifested as a distance decay, is a well-known phenomenon in studies of healthcare utilization [37, 38, 49] and occurs when usage of health facilities declines with increasing distance [50, 51]. This feature motivates the use of probit models to characterise treatment-seeking behaviour (section 2.4). Another feature of utilisation is that even for patients in close proximity to healthcare facilities, treatment for fever is not always 100% as some mild conditions self-resolve, are treated through informal care, or may be treated at a more distant facility [9]. Household survey data usually contain detailed information on other factors that could affect utilisation of healthcare facilities. These explanatory variables can be grouped largely into socio-economic and demographic characteristics and have been used selectively in quantitative studies of healthcare utilisation [3, 10, 17, 52, 53].

Application of Bayesian probit models to healthcare utilisation research

Item response modelling was proposed in the 1960s [54-56] and is commonly applied to studies in education and psychology to estimate item characteristics [28]. The first applications of IRT used maximum likelihood estimation [57, 58]. Bayesian extensions were proposed for one- and two-parameter models [59] and extended to the three-parameter logistic model [60]. Fitting via Gibbs sampling became popular using data augmentation (DAG) techniques in the 1990s particularly for application to the normal-ogive models [61-63]. Fu et al. [64] provided some extensions to the three-parameter model following Sahu's DAG approach [63] and compared Gibbs sampling to BILOG-MG software [65] using likelihood estimation. There have also been other innovations in parameter estimation [62], including extension to a multi-

level approach [26-28] and comparison with maximum likelihood methods [66]. Here, a unidimensional three-parameter model with a hierarchical structure was used, its parameters estimated, and prior sensitivity checked by comparing model goodness-of-fit statistics. The main objective was to estimate the probability of a positive response to choice of treatment for persons with fever associated with malaria at a household level.

In general, let Y_{ij} represent a dichotomous response variable of an individual j $(j=1,.....,N)$ on a set of questions (items) i $(i=1,.....,n)$ on use of public healthcare for treatment. $Y_{ij}=1$ represents a positive response on one item (e.g. public healthcare use), while $Y_{ij}=0$ represents a negative response (e.g., non-public healthcare use). The probability of $Y_{ij}=1$ can then be written following [64] as:

$$P\left(Y_{ij}=1|\theta_j,a_i,b_i,c_i\right)=c_i+\left(1-c_i\right)\frac{\exp\left\{\sum_{k=1}^{m}\left(a_{ik}\theta_{jk}-b_{ik}\right)\right\}}{\left[1+\exp\left\{\sum_{k=1}^{m}\left(a_{ik}\theta_{jk}-b_{ik}\right)\right\}\right]} \quad 1$$

where, $\theta_j=\theta_{j1}.....,\theta_{jk},.....\theta_{jm}$ with $-\infty<\theta_{jk}>+\infty$ for $k=1,.....m$ dimension represents the person traits (i.e. the ability parameter). a_{ik} represents item discrimination parameters between individuals separated by individual-level traits, and is positive $\left(a_{ik}>0\right)$. b_{ik} $(-\infty<b_{ik}<\infty)$ represents item difficulty (or location) parameters which for multiple items represents relationship between items and the underlying individual-traits. Lastly, c_i $(0<c_i<1)$ represents the threshold (i.e., minimum) probability for the item in question (fever treatment). This specification of threshold probability is important to this application because the estimated probability is never equal to one when θ_j is zero, due to several individual characteristics. Hence, probability of

188 treatment is constrained to be greater than zero and less than one. In many
189 applications in psychology and education, the ability parameter, for example, is
190 modelled as a latent characteristic independent of survey observations [67, 68]. In
191 this application, a predictor variable was introduced on the individual traits
192 parameter in terms of travel-time to the nearest health facility. This parameterisation
193 also enables the introduction of other variables such as residence (urban or rural),
194 socio-economic status, or educational levels. Thus, equation 1 can be simplified to:

195 $$P_{ij} = c_i + (1 - c_i)\Psi_{ij} \qquad 2$$

196 Where

197 $$\Psi_{ij} = \exp\left\{\sum_{k=1}^{m}\left(a_{ik}\theta_{jk} - b_{ik}\right)\right\} / \left[1 + \exp\left\{\sum_{k=1}^{m}\left(a_{ik}\theta_{jk} - b_{ik}\right)\right\}\right]$$
$$\theta_j = \alpha_j + \beta_{1j}X_{1j} + \ldots\ldots + \beta_{Qj}X_{Qj}$$

198 with β_{Qj} representing coefficients of dependent variables X_{Qj} exploring differences
199 in ability.

200

201 *The likelihood and posterior specification*

202 In general, let $f(\theta, a, b, c)$ denote a collection of unknown parameters, the posterior
203 can be expressed as the product of the likelihood and prior distributions for unknown
204 parameters given as:

205 $$f(\theta, a, b, c \mid y) \propto L(y \mid \theta, a, b, c) f(\theta, a, b, c) \qquad 3$$

206 where $f(\theta, a, b, c) = f(\theta) f(a) f(b) f(c)$ and the posterior density we wish to evaluate
207 is

208
$$D \times L(y \mid \theta, a, b, c) \times \left\{ \prod_{j=1}^{N} f\left(\theta \mid \mu_\theta, \sigma_\theta^2\right) \right\} \times \prod_{i=1}^{n} f\left(a_i \mid \mu_a, \sigma_a^2\right)$$
$$\times I(a_{ik} > 0) \times \prod_{i=1}^{n} f\left(b_i \mid \mu_b, \sigma_b^2\right) \times \prod_{i=1}^{n} f\left(c_i \mid \kappa_i, \tau_i\right)$$

209 where D is a proportionality constant and

$$L(y \mid \theta, a, b, c) = \prod_{i=1}^{n} \prod_{j=1}^{N} \left[P_{ij}^{y_{ij}} \left(1 - P_{ij}\right)^{1-y_{ij}} \right]$$

and

210
$$\prod_{j=1}^{N} f\left(\theta \mid \mu_\theta, \sigma_\theta^2\right) = \prod_{j=1}^{N} \exp\left\{ -\frac{1}{2} \left(\theta_j - \mu\right)^T \Sigma_\theta^{-1} \left(\theta_j - \mu\right) \right\}$$
$$\prod_{i=1}^{n} f\left(b_i \mid \mu_b, \sigma_b^2\right) \times \prod_{i=1}^{n} f\left(a_i \mid \mu_a, \sigma_a^2\right) = \frac{1}{\sigma^2} \prod_{i=1}^{n} \prod_{k=1}^{m} \frac{1}{\sigma^2} \exp\left\{ -\frac{\left(a_{ik} - \delta_a\right)^2 + \left(b_{ik} - \delta_b\right)^2}{2\sigma^2} \right\}$$
$$\prod_{i=1}^{n} f\left(c_i \mid \kappa_i, \tau_i\right) = \prod_{i=1}^{n} c_i^{k-1} (1 - c_i)^{\tau - 1}$$

211

212 *Goodness-of-fit statistics, prior specification and Markov chain Monte Carlo*
213 *implementation*

214 The same notation was used for the item discrimination parameter, with $a_i > 0$,
215 where a half-normal or truncated normal prior was used such that
216 $a_i \sim N\left(\mu_a, \sigma_a^2\right) I(a_i > 0)$ and $I(\cdot)$ is an indicator function. The rationale for this
217 specification is to ensure that the parameter estimate is positive. The probability
218 threshold parameter was constrained on $c \in (0,1]$ using a beta distribution such
219 that $\pi\left(c_k; \kappa, \tau\right) \alpha\, c_k^{\kappa - 1} \left(1 - c_k\right)^{\tau - 1}$ for suitable parameters values κ_{c_k} and τ_{c_k}.
220 The recommended procedure for selecting suitable estimates of these parameters is
221 such that the $E(c) = \kappa / (\kappa + \tau)$ and weakly informative priors may be used for
222 parameters of beta distribution.
223 Two different specifications were used for the difficulty parameter. The first was a
224 normal prior $b_i \sim N(0,10)$ (model 1) and the second a truncated normal (model 2)

restricting $b_i \sim N(\mu_b, \sigma_b^2)I(b_i > 0)$ to be positive. Thus, the difference between model 1 and model 2 was only in the prior specification for the b parameter. Figure 2a represents the overall parameter structure for Model 1 and Model 2. The rationale for using different priors for b was to evaluate the effect of constraining item difficulty to positivity $(b_i > 0)$ compared to allowing for flexible Gaussian density.

Lastly, the individual-trait parameter θ_j was modelled in a hierarchical approach following Fox and Glas [30] such that the joint distribution of θ_j parameters follows a multivariate normal distribution. Thus, in general, α and β are the intercept terms and regression coefficients, respectively, modelled as independent effects in model 1 and model 2 (Figure 2a). In extending the model to a multi-level representation, time to the nearest facility could then be used to explain individual traits. Normal priors (e.g. $\alpha \sim N(0,1)$) were used for α and β in Figure 2a. Secondly, this was extended to a random intercept in model 3 (Figure 2b) and lastly, as a random slope and intercept model including residence (urban or rural) as a centering variable (model 4, Figure 2c). For model 4, the random slope and intercept were modelled jointly as:

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim MVN\left\{ \begin{pmatrix} \beta_{\alpha,j} \\ \beta_{\beta,j} \end{pmatrix}, \sum = \begin{pmatrix} \tau_{11} & \tau_{12} \\ \tau_{21} & \tau_{22} \end{pmatrix} \right\} \quad 4$$

with a Wishart (multivariate scaled χ^2) distribution (Barnard et al. 2000) with density $f(\sum) \propto |\sum|^{-(\nu+d+1)/2} e^{-\frac{1}{2}tr\left(\wedge\Sigma^{-1}\right)}$; d dimension matrix; ν degrees of freedom; specified for covariance matrix $\sum$. Thus, the inverse is specified as $\sum^{-1} = Wishart(\Omega, p)$ where Ω is a scale matrix, usually identity, and p is the degrees of freedom equal to the number of random components. Alternative approaches

247 could employ a scaled inverse-Wishart distribution because of the large standard
248 errors associated with large variances in the use of the inverse-Wishart prior [69].
249 **< Insert fig 2>**
250 Validation was considered via a subset of 40% of the data selected randomly (*n*=247
251 of the 1,138 children) with the remaining 60% (*n*=891) used in model estimation.
252 Model 1 was then applied to the validation set and the predicted probability of
253 treatment transformed to a binary outcome. A receiver operating characteristic
254 (ROC) curve was then used to derive the specificity and sensitivity of predictions
255 when compared to observed responses from survey data. For estimation, different
256 model specifications were also used to check the sensitivity of different prior
257 specifications (i.e. models differ only on prior structure) and complexity. Model
258 outputs were evaluated and compared via goodness-of-fit statistics, for example,
259 the Deviance Information Criterion (DIC). The DIC summarises model fit based on a
260 combination of model deviance and complexity (effective number of parameters)
261 [70, 71]. This is defined as:

262 $$DIC = \overline{D} + pD \qquad 5$$

263 where $\overline{D} = E_{\theta|y}[D]$ is the mean deviance for $D = -2 \times \log\{P(y \mid \theta)\}$ with

264 $$\overline{D} = -\int 2\log\{P(y \mid \theta)\} d\theta$$
$$\hat{D} = -2\log\{P(y \mid \bar{\theta})\}$$

265 and complexity (effective number of parameters) given by $pD = \overline{D} - \hat{D}$. The two
266 parameters were monitored in the MCMC implementation using five chains in JAGS
267 version 4.2.0 and the *rjags* package in R version 3.3.1 [72]. A combination of
268 Gelman-Rubin [73] with Raftery-Lewis diagnostic [74] approaches were used to
269 check for convergence. For the former, we checked for a reduction factor of <1.05

while the latter provided estimates of burn-in and thinning factors given an accuracy of 0.0005 at quantile (0.025) and coverage probability of 0.975.

Results

We used the Namibia 2013 DHS data to estimate the probability of fever treatment in the formal sector (reported fever treatment in public and private sectors) for children under five years. There were 4,818 children under five years enumerated, of which 1,138 (23.6%) reported at least one fever episode in the preceding fortnight. Of those that reported a fever episode, 726 (63.8%) sought treatment in the formal sector (public and private sector excluding traditional healers). Overall, the proportion of children with reported fever was fairly homogeneous across all the regions surveyed but varied by estimated travel times. Estimation of probability of treatment focussed on children reporting fever (*n*=1,138) rather than all children examined in the cross-sectional survey.

<Insert Table 1>

In terms of computation, the Gelman-Rubin test was less than or equal to 1.05 for all the parameters monitored in the MCMC implementation. However, the Raftery-Lewis method showed that a minimum of 55,318 iterations were required to achieve an accuracy of 0.0005 at coverage probability of 0.999 with quantile at 0.05. More than 100,000 iterations with a burn-in of 50,000 were implemented. Table 1 shows the DIC estimates and the effective number of parameters from the four models implemented. Comparison between model 1 (M1 DIC 3615.9) and model 2 (M2 DIC 3685.1) suggests that using truncated normal priors for the b parameter did not improve model fit. Increasing DIC (for model 3 and 4) was also directly proportional to the increase in model complexity by including random intercept and slope. This also increased computational demands for M3 and M4 requiring at least 250,000

296 iterations with longer burn-in (slow convergence). The difference in DIC estimates
297 also suggested that the models were sensitive to changes in model structure. Based
298 on a binary classification of predicted probability at the individual level from model
299 1, the area under the curve (AUC) was 0.978 with a sensitivity of 96.7% and a
300 specificity of 75.3% (155 true positive, 21 false positive, 64 true negative, and 7 false
301 negative).

302 Table 2 shows posterior estimates of the parameters along with 95% equal-tailed
303 credible intervals. A plot of fever-response curves based on the fitted parameters is
304 shown in Figure 3a along with a scatterplot of α and β parameters from Model 4
305 (Figure 3b), posterior density of parameters (Figure 3c) and ROC plot (Figure 3d).
306 Different mean combinations of parameters a, b, and c resulted in response
307 characteristics based on travel time to nearest health facility (Figure 3a). Parameter
308 estimates could be compared and interpreted jointly in this manner because they
309 apply to one item (on estimating fever treatment). Comparison between model 1
310 and model 2 suggested that constraining the b parameter did not have a major
311 impact on mean estimates of the individual-level traits, a or the threshold
312 parameter c. Overall, model 4 had larger person discriminant parameter estimates
313 (mean and median) compared to all the other model specifications. The correlation
314 between mean estimates for α and β as estimated from the model was weak
315 (mean -0.011, median 0.006 scatterplot Figure 2b). The combination of correlation
316 and DIC estimates suggested a fixed prior independent specification as a better
317 choice. It also imposes less computational demand. The threshold probability was
318 >0.3 for all model estimates, suggesting this as the lower limit probability of use of
319 nearest facility for fever treatment in the four models implemented from the 2013
320 Namibia DHS.

<Insert table 2 and fig 3>

Table 3 shows the estimated mean probabilities for malaria related fever treatment at a regional level in Namibia with associated confidence intervals and population estimates. Population estimates are useful in estimating fever treatment burden based on probability estimate at regional level. For malaria, the probability of fever treatment among febrile cases was highest in endemic areas in Zambezi and Kavango (mean probability in Zambezi 0.546 (95% Credible Interval (CI): 0.369-0.671)) compared to Kunene with less than one case per 1000 population with mean probability 0.433 (95% CI: 0.364-0.614). Overall mean probability of fever treatment was greater than 0.5 in areas with malaria incidence >1 per 1000 population.

<Insert Table 3>

Discussion

Characterising treatment-seeking behaviour in LMICs is valuable because it varies by geographic location, type of disease and severity, person characteristics including age and gender, as well as health system based factors such as availability, cost among other enabling factors [9, 75, 76]. Here, the focus was on the estimation of latent parameters of a survey question on fever and estimating the probability of seeking treatment based on a dichotomous response. We used data from a nationally representative household survey from the DHS in one country to estimate fever treatment latent characteristics using a Bayesian IRT approach. By using this method, we estimated the parameters of fever response curves that characterise geographical decay in the use of formal health care based on travel time to the nearest facility. The method is particularly appealing because of the joint estimation of IRT parameters related to fever treatment with uncertainties incorporated in prior distributions and the ability to extract the full posterior distribution compared to point

estimates from maximum likelihood approaches [26, 61]. This is important because estimates from such probabilistic modelling can then be applied in estimating numbers of symptomatic infections (treatment burden) when such probabilistic estimates are transformed into gridded metrics that vary spatially [77, 78]. The modelling approach can also be extended to other items in household surveys to further understand human behaviour response to health conditions.

The lower limit probability estimated here, related to the threshold parameter (e.g. from table 2 model 1: 0.340; 95% CI 0.155-0.597), for Namibia suggests that even at large distances from health facilities, there was still a 30% chance of individuals seeking fever treatment. We suggest that this is an important property in treatment-seeking behaviour for individuals living far from health facilities in Namibia, although this threshold may be different by country or endemicity and was not explored further in this analysis. In this study, estimates of probability of fever treatment at the regional level showed that the mean probability was highest in regions with relatively high incidence of malaria historically (Table 3). Another operational application of the probability response characteristics curves, derived from the latent parameters in Figure 3a, could be in identifying areas where community health workers could be deployed [79, 80]. This, however, requires definition of a cut-off probability (*y*-axis on Figure 3a), currently not established for malaria transmission settings, to delineate areas with limited access. Constraining the b parameter (item parameter) did not influence estimates of the individual-level traits and the threshold parameters. This is primarily because only one item was used in this application resulting in similar parameter estimate for the location parameter.

In extending the model to a multilevel framework, travel times were used as predictors. Comparison between constant intercept and slope model parameters with a random parameter model showed that the former resulted in shorter MCMC runs and better model fit compared to the latter (i.e., the random slope and intercept), which experienced slow convergence as the number of effective parameters increased exponentially. We are not discouraging use of a more complex modelling approach while estimating IRT parameters, but this highlights the increasing computational demands and efficiency related to increased complexity.

MCMC techniques were used to estimate and jointly interpret IRT parameters. The three-parameter logistic model [60] was particularly useful compared to the two-parameter model [59], because, the third parameter c represents the threshold probability on the fever response curve, ensuring that probability is always greater than or equal to zero. Despite the known benefits of IRT in other fields [28], this approach has seldom been applied to modelling human behavioural aspects for treatment-seeking behaviour. The current study was confined to patients' responses to a fever question in household survey data and how latent (rather than observed) properties can be quantified in relation to patient behaviour and travel time. Dichotomous responses are common in many health surveys in LMICs and methods used here can be extended to other health conditions. Although we did not have to deal with missing data (NAs), several data imputation techniques can be used for non-ignorable NAs [81]. These may arise when there is lack of response, or, associated with refusal to participate or simply unobserved variable for survey items. When NAs are imputed into the data matrix, for example, these do not usually contribute to likelihood estimation [82] of the ability parameter and the higher the

number of missing values the more likely that there will be an increase in uncertainty for the parameter estimate.

There exist some additional limitations aside from those related to computational speed and efficiency. While fever in the Namibia 2013 DHS was associated with malaria treatment, the survey data did not include a laboratory confirmation of malaria infection [83]. Moreover, the sampling methodology for children with fever in the DHS may be inferior because the survey is not powered for fever detection [47]. Most current surveys however incorporate rapid diagnostic tests (RDTs) and future identification of febrile cases could include laboratory results as a preprocessing step in identifying malaria-related fever cases. In addition, although prior specifications introduce a measure of uncertainty in a hierarchical way, assumptions in generating input data such as use of the nearest facility may not be sufficient in understanding treatment-seeking behaviour. It has been shown in separate population surveys that patients may bypass the nearest health centre due to various individual- or supply-based factors such as quality [84]. While an obvious recommendation is to include such effects, increasing model complexity to capture such differences may have an impact on computational efficiency as seen in model 3 and model 4. More importantly, identifying measures of quality of care in public or private health sectors can be challenging [40].

Conclusion

In the context of fever treatment, we have demonstrated that there is potential to use nationally representative household data to provide a probabilistic measure of treatment using a Bayesian method. Our estimates of threshold probability apply to one low malaria transmission country and may be different in other countries with

varying malaria endemicity. Future studies will aim to conduct such comparative analysis between and within countries via spatially varying parameters. The methodology can be extended to multiple human behavioural questions (items) related to health and demographics in the routine national survey data.

Declarations

Consent for publication: Not Applicable

Availability of data and material: DHS data available in the public domain at http://dhsprogram.com/data/available-datasets.cfm

Competing interests: Authors declare no competing interests.

Ethics approval and consent to participate
University of Southampton (17263).

Authors' contributions
VA, PMA, and AJT were responsible for study design, analysis, interpretation, and production of final manuscript. CP and JW contributed to data assembly and management, interpretation and production of final manuscript. All authors have read and approved the final version of the manuscript.

Acknowledgements
We would like to thank Professor Sujit Sahu (University of Southampton) and Dr Linus Bengtsson (Flowminder co-director) for comments on the earlier version of the manuscript.

Funding
Andrew J Tatem is supported by a Wellcome Trust Sustaining Health Grant [grant number 106866/Z/15/Z] and Bill and Melinda Gates Foundation [grant numbers OPP1106427, 1032350, OPP1134076].

Abbreviations
AUC: area under curve; CHW: community health worker; DAG: data augmentation; DHS: demographic health surveys; DIC: deviance information criterion; iCCM: integrated community-case management; IRT: item response theory; LMICs: low- and middle-income countries; MCMC: Markov chain Monte Carlo; MICS: multiple indicator cluster surveys; MIS: malaria indicator surveys; ROC: receiver-operating characteristics; SSA: Sub-Saharan Africa; WHO: World Health Organization.

References

1. Rudan I, El Arifeen S, Bhutta ZA, Black RE, Brooks A, Chan KY, Chopra M, Duke T, Marsh D, Pio A *et al*: **Setting Research Priorities to Reduce Global Mortality from Childhood Pneumonia by 2015**. *PLOS Medicine* 2011, **8**(9):e1001099.

2. Herbert HK, Lee ACC, Chandran A, Rudan I, Baqui AH: **Care Seeking for Neonatal Illness in Low- and Middle-Income Countries: A Systematic Review**. *PLOS Medicine* 2012, **9**(3):e1001183.

3. Colvin CJ, Smith HJ, Swartz A, Ahs JW, de Heer J, Opiyo N, Kim JC, Marraccini T, George A: **Understanding careseeking for child illness in sub-Saharan Africa: a systematic review and conceptual framework based on qualitative research of household recognition and response to child diarrhoea, pneumonia and malaria**. *Soc Sci Med* 2013, **86**:66-78.

4. World Health Organizastion: **Integrated community-based interventions**. In., vol. TDR/BL 11.10. Geneva; 2009.

5. The malE. R. A. Consultative Group on Health Systems Operational R: **A Research Agenda for Malaria Eradication: Health Systems and Operational Research**. *PLoS Med* 2011, **8**(1):e1000397.

6. Bhutta ZA, Sommerfeld J, Lassi ZS, Salam RA, Das JK: **Global burden, distribution, and interventions for infectious diseases of poverty**. *Infectious diseases of poverty* 2014, **3**:21-21.

7. World Health Organization: **Guidelines for the treatment of malaria**. In., 3 edn. Geneva: WHO; 2015.

8. Landier J, Parker DM, Thu AM, Carrara VI, Lwin KM, Bonnington CA, Pukrittayakamee S, Delmas G, Nosten FH: **The role of early detection and treatment in malaria elimination**. *Malaria Journal* 2016, **15**(1):1-8.

9. Littrell M, Gatakaa H, Evance I, Poyer S, Njogu J, Solomon T, Munroe E, Chapman S, Goodman C, Hanson K *et al*: **Monitoring fever treatment behaviour and equitable access to effective medicines in the context of initiatives to improve ACT access: baseline results and implications for programming in six African countries**. *Malaria Journal* 2011, **10**(1):327.

10. Sundararajan R, Mwanga-Amumpaire J, Adrama H, Tumuhairwe J, Mbabazi S, Mworozi K, Carroll R, Bangsberg D, Boum Ii Y, Ware NC: **Sociocultural and Structural Factors Contributing to Delays in Treatment for Children with Severe Malaria: A Qualitative Study in Southwestern Uganda**. *The American Journal of Tropical Medicine and Hygiene* 2015, **92**(5):933-940.

11. O'Neill S, Gryseels C, Dierickx S, Mwesigwa J, Okebe J, d'Alessandro U, Grietens KP: **Foul wind, spirits and witchcraft: illness conceptions and health-seeking behaviour for malaria in the Gambia**. *Malaria Journal* 2015, **14**(1):167.

12. Hadley M: **Does increase in utilisation rates alone indicate the success of a user fee removal policy? A qualitative case study from Zambia**. *Health Policy* 2011, **103**(2â€“3):244-254.

13. Battle KE, Bisanzio D, Gibson HS, Bhatt S, Cameron E, Weiss DJ, Mappin B, Dalrymple U, Howes RE, Hay SI *et al*: **Treatment-seeking rates in malaria endemic countries**. *Malaria Journal* 2016, **15**(1):1-11.

14. Alegana V, Wright J, Petrina U, Noor A, Snow R, Atkinson P: **Spatial modelling of healthcare utilisation for treatment of fever in Namibia**. *International Journal of Health Geographics* 2012, **11**(1):6.

15. Wasunna B, Okiro EA, Webster J, Todd J, Snow RW, Jones C: **The Impact of a Community Awareness Strategy on Caregiver Treatment Seeking Behaviour and Use of Artemether-Lumefantrine for Febrile Children in Rural Kenya**. *PLoS ONE* 2015, **10**(7):e0130305.

16. Aung T, Lwin MM, Sudhinaraset M, Wei C: **Rural and urban disparities in health-seeking for fever in Myanmar: findings from a probability-based household survey**. *Malar J* 2016, **15**(1):386.

17. Matovu F, Nanyiti A, Rutebemberwa E: **Household health care-seeking costs: experiences from a randomized, controlled trial of community-based malaria and pneumonia treatment among under-fives in eastern Uganda**. *Malar J* 2014, **13**:222.

18. Manongi R, Mtei F, Mtove G, Nadjm B, Muro F, Alegana V, Noor AM, Todd J, Reyburn H: **Inpatient child mortality by travel time to hospital in a rural area of Tanzania**. *Trop Med Int Health* 2014.

19. Hays RD, Morales LS, Reise SP: **Item response theory and health outcomes measurement in the 21st century**. *Med Care* 2000, **38**(9 Suppl):li28-42.

20. **The WorldPop project** [http://www.worldpop.org.uk/]

21. Dalrymple U, Mappin B, Gething P: **Malaria mapping: understanding the global endemicity of falciparum and vivax malaria**. *BMC Medicine* 2015, **13**(1):140.

22. Alegana VA, Atkinson PM, Pezzulo C, Sorichetta A, Weiss D, Bird T, Erbach-Schoenberg E, Tatem AJ: **Fine resolution mapping of population age-structures for health and development applications**. *Journal of The Royal Society Interface* 2015, **12**(105).

23. Roll Back Malaria Monitoring and Evaluation Reference Group: **A guide to Malaria Indicator Surveys (MIS)**. In. Geneva; 2008.

24. **Demographic and Health Surveys** [http://www.measuredhs.com]

25. United Nations Children Fund (UNICEF): **Monitoring the situation of women and children: Multiple Indicator Cluster Survey manual 2005**. In. New York: UNICEF; 2006.

26. Matteucci M: **An Investigation of Parameter Recovery in MCMC Estimation for the Additive IRT Model**. *Communications in Statistics - Theory and Methods* 2014, **43**(4):751-770.

27. Yanyan Sheng, Wikle CK: **Bayesian Multidimensional IRT Models With a Hierarchical Structure**. *Educational and Psychological Measurement* 2008, **68**(3):413-430.

28. Sheng Y: **Review of Bayesian Item Response Modeling: Theory and Applications, by Jean-Paul Fox**. *Structural Equation Modeling: A Multidisciplinary Journal* 2015, **22**(3):484-488.

29. Béguin AA, Glas CAW: **MCMC estimation and some model-fit analysis of multidimensional IRT models**. *Psychometrika* 2001, **66**(4):541-561.

30. Fox J-P, Glas CAW: **Bayesian estimation of a multilevel IRT model using gibbs sampling**. *Psychometrika* 2001, **66**(2):271-288.

31. Fox J-P, Wyrick C: **A Mixed Effects Randomized Item Response Model**. *Journal of Educational and Behavioral Statistics* 2008, **33**(4):389-415.

32. Yang FM, Jones RN, Inouye SK, Tommet D, Crane PK, Rudolph JL, Ngo LH, Marcantonio ER: **Selecting optimal screening items for delirium: an application of item response theory**. *BMC Medical Research Methodology* 2013, **13**(1):8.

33. Gorter R, Fox J-P, Twisk JWR: **Why item response theory should be used for longitudinal questionnaire data analysis in medical research**. *BMC Medical Research Methodology* 2015, **15**(1):1-12.

34. Dregan A, Grieve A, van Staa T, Gulliford MC: **Potential application of item-response theory to interpretation of medical codes in electronic patient records**. *BMC Medical Research Methodology* 2011, **11**(1):168.

35. Sheng Y, Wikle CK: **Comparing Multiunidimensional and Unidimensional Item Response Theory Models**. *Educational and Psychological Measurement* 2007, **67**(6):899-919.

36. Banerjee S, Carling PB, Gelfand AE: **Hierarchical modeling and analysis for spatial data**. London: Chapman & Hall/CRC; 2004.

37. Stock R: **Distance and the utilization of health facilities in rural Nigeria**. *Soc Sci Med* 1983, **17**(9):563-570.

38. Buor D: **Analysing the primacy of distance in the utilization of health services in the Ahafo-Ano South district, Ghana**. In., vol. 18: John Wiley & Sons, Ltd.; 2003: 293-311.

39. Noor AM, Amin AA, Gething PW, Atkinson PM, Hay SI, Snow RW: **Modelling distances travelled to government health services in Kenya**. *Trop Med Int Health* 2006, **11**(2):188-196.

40. Basu S, Andrews J, Kishore S, Panjabi R, Stuckler D: **Comparative Performance of Private and Public Healthcare Systems in Low- and Middle-Income Countries: A Systematic Review**. *PLoS Med* 2012, **9**(6).

41. Tanser F, Hosegood V, Benzler J, Solarsh G: **New approaches to spatially analyse primary health care usage patterns in rural South Africa**. *Trop Med Int Health* 2001, **6**(10):826-838.

42. Noor AM, Rage IA, Moonen B, Snow RW: **Health service providers in Somalia: their readiness to provide malaria case-management**. *Malar J* 2009, **8**:100.

43. Leonard K, Mliga GR, Mariam DH: **Bypassing health centers in Tanzania: Revealed preferences for observable and unobservable quality**. In.: Columbia University, Department of Economics; 2002.

44. Akin JS, Hutchinson P: **Health-care facility choice and the phenomenon of bypassing**. *Health Policy Plan* 1999, **14**(2):135-151.

45. Ray N, Ebener S: **AccessMod 3.0: computing geographic coverage and accessibility to health care services using anisotropic movement of patients**. *International Journal of Health Geographics* 2008, **7**(1):63.

46. Corsi DJ, Neuman M, Finlay JE, Subramanian S: **Demographic and health surveys: a profile**. *International Journal of Epidemiology* 2012, **41**(6):1602-1613.

47. Aliaga A, Ren R: **Optimal sample sizes for two-stage cluster sampling in demographic and health surveys**. In: *DHS Working Papers No 30.* Calverton, Maryland, USA: ORC Macro; 2006.

48. Le TN, Verma VK: **An analysis of sampling designs and sampling errors of the demographic and health surveys**. In: *DHS Analytical Reports No 3.* Calverton, Maryland, USA: Macro International; 1997.

49. McLaren ZM, Ardington C, Leibbrandt M: **Distance decay and persistent health care disparities in South Africa**. *BMC Health Services Research* 2014, **14**(1):1-9.

50. Bailey CT, Gatrell CA: **Interactive Spatial Data Analysis**. Essex, England: Longman Scientific & Technical; 1995.

51. Cromley EK, McLafferty SL: **GIS and public health**. New York: Guilford Press; 2002.

52. Febir LG, Asante KP, Afari-Asiedu S, Abokyi LN, Kwarteng A, Ogutu B, Gyapong M, Owusu-Agyei S: **Seeking treatment for uncomplicated malaria: experiences from the Kintampo districts of Ghana**. *Malaria Journal* 2016, **15**(1):1-11.

53. Chibwana AI, Mathanga DP, Chinkhumba J, Campbell CH, Jr.: **Socio-cultural predictors of health-seeking behaviour for febrile under-five children in Mwanza-Neno district, Malawi**. *Malar J* 2009, **8**:219.

54. Rasch G: **On general laws and the meaning of the measurement in psychology**. In: *Proceedings of the 4th Berkley Symposium on Mathematical Statistics: 1961; London*: University of California Press; 1961: 321-334.

55. McDonald RP: **Numerical methods for polynomial models in nonlinear factor analysis**. *Psychometrika* 1967, **32**(1):77-112.

56. Lord FM, Novick MR, Birnbaum A: **Statistical Theories of Mental Test Scores**. MA: Addison-Wesley; 1968.

57. Darrell Bock R: **Estimating item parameters and latent ability when responses are scored in two or more nominal categories**. *Psychometrika* 1972, **37**(1):29-51.

58. Andersen EB: **The Numerical Solution of a Set of Conditional Estimation Equations**. *Journal of the Royal Statistical Society Series B (Methodological)* 1972, **34**(1):42-54.

59. Swaminathan H, Gifford JA: **Bayesian estimation in the two-parameter logistic model**. *Psychometrika* 1985, **50**(3):349-364.

60. Swaminathan H, Gifford JA: **Bayesian estimation in the three-parameter logistic model**. *Psychometrika* 1986, **51**(4):589-601.

61. Albert JH: **Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling**. *Journal of Educational Statistics* 1992, **17**(3):251-269.

62. Braeken J, Tuerlinckx F: **Investigating latent constructs with item response models: A MATLAB IRTm toolbox**. *Behavior Research Methods* 2009, **41**(4):1127-1137.

63. Sahu SK: **Bayesian Estimation and Model Choice in Item Response Models**. *Journal of Statistical Computation and Simulation* 2002, **72**(3):217-232.

64. Fu Z-H, Tao J, Shi N-Z: **Bayesian estimation in the multidimensional three-parameter logistic model**. *Journal of Statistical Computation and Simulation* 2009, **79**(6):819-835.

65. Rupp AA: **Item Response Modeling With BILOG-MG and MULTILOG for Windows**. *International Journal of Testing* 2003, **3**(4):365-384.

66. Skrondal A, Rabe-Hesketh S: **Prediction in multilevel generalized linear models**. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2009, **172**(3):659-687.

67. Hattie J: **Methodology Review: Assessing Unidimensionality of Tests and Itenls**. *Applied Psychological Measurement* 1985, **9**(2):139-164.

68. de la Torre J, Douglas JA: **Higher-order latent trait models for cognitive diagnosis**. *Psychometrika* 2004, **69**(3):333-353.

69. Barnard J, McCulloch R, Meng X-L: **Modelling covariance matrices in terms of standard deviations and correlations, with application to shrinkage**. *Statistica Sinica* 2000, **10**(4):1281-1311.

70. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A: **Bayesian measures of model complexity and fit**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002, **64**(4):583-639.

71. Plummer M: **Penalized loss functions for Bayesian model comparison**. *Biostatistics* 2008, **9**(3):523-539.

72. Plummer M: **rjags: Bayesian Graphical Models using MCMC**. In.: Plummer, M; 2016.

73. Gelman A, Rubin DB: **Inference from Iterative Simulation Using Multiple Sequences**. *Statistical Science* 1992, **7**(4):457-472.

74. Raftery AE, Lewis SM: **Implementing MCMC**. In: *Markov Chain Monte Carlo in Practice*. edn. Edited by Gilks WR, Spiegelhalter DJ, Richardson S: London: Chapman and Hall; 1996: 115-130.

75. Kizito J, Kayendeke M, Nabirye C, Staedke SG, Chandler CI: **Improving access to health care for malaria in Africa: a review of literature on what attracts patients**. *Malar J* 2012, **11**:55.

76. Lagarde M, Palmer N: **The impact of user fees on access to health services in low- and middle-income countries**. In: *Cochrane Database of Systematic Reviews.* John Wiley & Sons, Ltd; 2011.

77. Chen I, Clarke SE, Gosling R, Hamainza B, Killeen G, Magill A, O?Meara W, Price RN, Riley EM: **"Asymptomatic" Malaria: A Chronic and Debilitating Infection That Should Be Treated**. *PLoS Med* 2016, **13**(1):e1001942.

78. Sturrock HJW, Hsiang MS, Cohen JM, Smith DL, Greenhouse B, Bousema T, Gosling RD: **Targeting Asymptomatic Malaria Infections: Active Surveillance in Control and Elimination**. *PLoS Med* 2013, **10**(6):e1001467.

79. Ferrer BE, Webster J, Bruce J, Narh-Bana SA, Narh CT, Allotey NK, Glover R, Bart-Plange C, Sagoe-Moses I, Malm K *et al*: **Integrated community case management and community-based health planning and services: a cross sectional study on the effectiveness of the national implementation for the treatment of malaria, diarrhoea and pneumonia**. *Malar J* 2016, **15**(1):340.

80. Kirigia J, Asbu E: **Technical and scale efficiency of public community hospitals in Eritrea: an exploratory study**. *Health Economics Review* 2013, **3**(1):6.

81. Little R, Rubin D: **Statistical Analysis with Missing Data**, 2nd edn. New York: John Wiley & Sons, Incorporated; 2002.

82. Little RJA, Rubin DB: **On Jointly Estimating Parameters and Missing Data by Maximizing the Complete-Data Likelihood**. *The American Statistician* 1983, **37**(3):218-220.

83. Roll Back Malaria, Measure Evaluation, USAID, UNICEF, World Health Organization, MACEPA, Centre for Disease Control: **Guidelines for core population-based indicators**. In. Calverton, Maryland; 2009.

84. DiLiberto DD, Staedke SG, Nankya F, Maiteki-Sebuguzi C, Taaka L, Nayiga S, Kamya MR, Haaland A, Chandler CI: **Behind the scenes of the PRIME intervention: designing a complex intervention to improve malaria care at public health centres in Uganda**. *Glob Health Action* 2015, **8**:29067.

85. Ministry of Health and Social Services: **Namibia Health Facility Census, 2009**. In. Windhoek, Namibia: Ministry of Health and Social Services and ICF Macro; 2010: 585.

86. Alegana VA, Atkinson PM, Lourenço C, Ruktanonchai NW, Bosco C, Erbach-Schoenberg Ez, Didier B, Pindolia D, Le Menach A, Katokele S *et al*: **Advances in mapping malaria for elimination: fine resolution modelling of Plasmodium falciparum incidence**. *Scientific Reports* 2016, **6**:29628.

Appendix
Parameter notations

j	Individual/person
i	Item/survey question
k	Dimension for items
q	Dimension for dependent variables
a	Discrimination parameter
b	Difficulty parameter on items
c	Probability threshold parameter
θ	Individual trait/ability parameter
$P(Y)$	Probability that event Y occurs
$I(\cdot)$	Indicator function for event in sample space
$E(X)$	Expectation for random parameter X
μ	Mean
DIC	Deviance Information Criterion
$\overline{D}$	Mean deviance
$[\{(\cdot)\}]$	Order of brackets

Legends to figures

Fig 1: **Visualisation of malaria-associated fever treatment** from DHS data by **A**) age (Children 0-5 years) and **B**) by travel time to the nearest health facility generated from GIS methods combining spatial data (Land cover, roads), population centres and the locations of health facilities.

Fig 2: Graphical representation of the form of the models used. A) simplified fixed parameter specification used for model 1 and model 2; **B)** allowing for a random slope (model 3) on the α parameter; **C)** random slope and intercept (model 4) for the α and β parameters, respectively, centering on residence (urban and rural) with correlation estimated via the Wishart prior specification. Model 1 and Model 2 differ only in the prior specification for item difficulty (b) parameter.

Fig 3: Panel plots showing: A) Fever response decay curves from the four model parameter values from the DHS survey in Namibia for 2014. The data are from 1,138 (*n*=891 training, 247validation) children under the age of five reporting fever 2 weeks prior to survey of which 726 sought treatment in the formal sector. **B)** A scatterplot for mean estimates on α (intercept) and β (slope) parameters based on model 4

(random slope and intercept model at individual level). **C)** Posterior density for IRT parameters (a individual discriminant parameter, b item dificulty, and c probability threshold). **D**) Receiver operating characteristic (ROC) plot based on the validation dataset (*n*=247 children). The binary classification was based on the predicted probability of seeking treatment for fever (from model 1) with a cut-off at 0.65. ROC had AUC=0.978 and an accuracy measure of 0.887.

Table 1: Model comparison based on goodness-of-fit statistics. DIC is the deviance information criterion while PD is the model complexity (number of model parameters)

Model	DIC	PD	Inverse log likelihood	Number of chains
M1*	3615.9	2178.9	-0.001	3
M2*	3685.1	2256.1	-0.001	3
M3	5098.7	3693.1	-0.001	3
M4	23874.1	22754.0	-0.002	3

*Model 1 and model 2 only differ in prior specification for the b parameter

Table 2: Estimated summary statistics and the 95% Bayesian credible intervals of parameters based on all four models. M1 and M2 use a fixed parameter specification for α and β using normal priors but different priors for item parameters, M3 allows random intercepts only, and M4 is both a random slope and intercepts model. Only M4 include a measure of correlation between the multi-level regression parameters.

Model	Estimate	a	b	c	α	β	Corr(α, β)
M1	Mean	0.704	0.807	0.340	-0.084	-0.098	-
	Median	0.556	0.850	0.326	-0.087	-0.112	-
	95% CI	[0.016 - 2.194]	[-1.044 - 2.346]	[0.1554 - 0.597]	[-0.682 - 0.523]	[-0.439 - 0.394]	-
	Gelman-Rubin Convergence estimate	1.000	1.000	1.000	1.000	1.000	-
	Gelman-Rubin Convergence upper CI	1.000	1.010	1.000	1.000	1.030	-
M2	mean	0.784	1.060	0.352	-0.080	-0.123	-
	median	0.654	0.978	0.344	-0.081	-0.121	-
	95% CI	[0.042 - 2.243]	[0.100 - 2.452]	[0.172 - 0.572]	[-0.661 - 0.518]	[-0.417 - 0.218]	-
	Gelman-Rubin Convergence estimate	1.001	1.000	1.001	1.001	1.020	-
	Gelman-Rubin Convergence upper CI	1.010	1.000	1.000	1.000	1.040	-
M3	mean	0.789	0.977	0.376	-0.582	-0.140	-
	median	0.660	0.895	0.372	-0.581	-0.150	-
	95% CI	[0.046 - 2.225]	[0.055 - 2.423]	[0.176 - 0.597]	[-2.208 - 1.069]	[-0.434 - 0.248]	-
	Gelman-Rubin Convergence estimate	1.000	1.000	1.000	1.000	1.000	-
	Gelman-Rubin Convergence upper CI	1.000	1.000	1.000	1.000	1.000	-
M4	mean	0.870	1.003	0.313	-0.133	-0.008	-0.011
	median	0.768	0.912	0.311	-0.152	-0.012	0.006
	95% CI	[0.059 - 2.244]	[0.063 - 2.477]	[0.095 - 0.527]	[-0.665 - 0.501]	[-0.880 - 0.873]	[-0.957 - 0.952]
	Gelman-Rubin Convergence estimate	1.000	1.000	1.000	1.010	1.000	1.010
	Gelman-Rubin Convergence upper CI	1.000	1.000	1.010	1.040	1.000	1.050

Table 3: Estimated probability for fever treatment (mean and 95% Bayesian Credible Interval) at the nearest health facility. Data for health facilities represent public and private entities based on facility census 2009 [85] updated based on HMIS reports. Probability of use for fever treatment estimated from parameters of model 1

Region	Population estimate 2015[1]	Estimated mean malaria incidence per 1000 population in 2014[2]	Estimated Average travel time to nearest health facility (minutes)	Probability of using a dispensary or clinic for fever treatment mean (95% CI)	Probability of using a health centre for fever treatment mean (95% CI)	Probability of using a Regional or district hospital for fever treatment mean (95% CI)
Zambezi	105,804	1.612	23.0	0.546 (0.369-0.671)	0.537 (0.369-0.667)	0.531 (0.369-0.661)
Kavango	259,984	1.467	29.7	0.513 (0.368-0.649)	0.498 (0.368-0.633)	0.503 (0.368-0.638)
Ohangwena	283,188	1.426	29.3	0.522 (0.368-0.650)	0.494 (0.367-0.630)	0.497 (0.368-0.632)
Oshikoto	210,881	1.256	37.3	0.504 (0.368-0.644)	0.492 (0.367-0.633)	0.496 (0.367-0.637)
Otjozondjupa	167,186	1.227	31.8	0.504 (0.368-0.643)	0.486 (0.367-0.623)	0.499 (0.368-0.637)
Omusati	281,050	1.131	35.6	0.513 (0.368-0.650)	0.497 (0.367-0.637)	0.498 (0.367-0.638)
Omaheke	82,441	1.126	38.3	0.490 (0.367-0.631)	0.496 (0.367-0.637)	0.493 (0.367-0.634)
Oshana	207,218	1.096	17.6	0.561 (0.369-0.677)	0.545 (0.369-0.661)	0.547 (0.369-0.663)
Kunene	102,986	0.967	146.4	0.433 (0.364-0.614)	0.426 (0.364-0.608)	0.429 (0.364-0.612)
Khomas[3]	418,742	-	43.9	0.487 (0.367-0.636)	0.483 (0.367-0.631)	0.482 (0.367-0.631)
Karas[3]	88,977	-	110.2	0.447 (0.365-0.619)	0.440 (0.365-0.613)	0.446 (0.365-0.618)
Hardap[3]	93,447	-	86.7	0.471 (0.366-0.628)	0.470 (0.366-0.626)	0.461 (0.366-0.616)
Erongo[3]	180,672	-	98.4	0.443 (0.365-0.620)	0.440 (0.365-0.618)	0.440 (0.365-0.617)

[1] Population estimates derived from worldpop [20]

[2] Mean malaria incidence derived from Alegana et al. [86]

[3] Regions designated as no malaria risk with case incidence of less than 1 per 10,000 population