

# THE UNIVERSITY OF WARWICK

**Original citation:**

Crump, Ron E. and Medley, Graham F.. (2015) Back-calculating the incidence of infection of leprosy in a Bayesian framework. *Parasites & Vectors*, 8 (1). 534.

**Permanent WRAP url:**

<http://wrap.warwick.ac.uk/77707>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 4.0 International license (CC BY 4.0) and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by/4.0/>

**A note on versions:**

The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: [publications@warwick.ac.uk](mailto:publications@warwick.ac.uk)

warwick**publications**wrap  
  
highlight your research

<http://wrap.warwick.ac.uk>

RESEARCH

Open Access



# Back-calculating the incidence of infection of leprosy in a Bayesian framework

Ronald E. Crump<sup>1\*</sup> and Graham F. Medley<sup>2</sup> 

## Abstract

**Background:** The number of new leprosy cases reported annually is falling worldwide, but remains relatively high in some populations. Because of the long and variable periods between infection, onset of disease, and diagnosis, the recently detected cases are a reflection of infection many years earlier. Estimation of the numbers of sub-clinical and clinical infections would be useful for management of elimination programmes. Back-calculation is a methodology that could provide estimates of prevalence of undiagnosed infections, future diagnoses and the effectiveness of control.

**Methods:** A basic back-calculation model to investigate the infection dynamics of leprosy has been developed using Markov Chain Monte Carlo in a Bayesian context. The incidence of infection and the detection delay both vary with calendar time. Public data from Thailand are used to demonstrate the results that are obtained as the incidence of diagnosed cases falls.

**Results:** The results show that the underlying burden of infection and short-term future predictions of cases can be estimated with a simple model. The downward trend in new leprosy cases in Thailand is expected to continue. In 2015 the predicted total number of undiagnosed sub-clinical and clinical infections is 1,168 (846–1,546) of which 466 (381–563) are expected to be clinical infections.

**Conclusions:** Bayesian back-calculation has great potential to provide estimates of numbers of individuals in health/infection states that are as yet unobserved. Predictions of future cases provides a quantitative measure of understanding for programme managers and evaluators. We will continue to develop the approach, and suggest that it might be useful for other NTD in which incidence of diagnosis is not an immediate measure of infection.

**Keywords:** Leprosy, Hansen's disease, Back-calculation, Forecast, Surveillance, Thailand, Elimination

## Background

Despite a large reduction in prevalence over the last half century, leprosy remains a public health issue in many countries [1]. In 2013 there were 215,000 new cases of leprosy detected and reported worldwide, with 96 % of these being detected in the 18 countries with the highest burden [2]. Leprosy is caused by the bacterium *Mycobacterium leprae* and this infection can be cured through the use of multidrug therapy (MDT). The drug regimen is dependent upon the classification of the case: paucibacillary cases receive six months of treatment with two-drug MDT while multibacillary cases are treated for 12 months with three-drug MDT. Multibacillary cases

are identified as those with more than five skin lesions; paucibacillary cases have five skin lesions or fewer. Outside of the Americas there is little published evidence of transmission to humans from non-human reservoirs of infection [3]. Taking this as evidence that non-human reservoirs are not an important source of transmission, then removing clinical, or sub-clinical, infection from people by treatment should reduce the incidence of infection.

In 1991, the World Health Assembly resolved to eliminate leprosy as a public health problem by the year 2000 [4]. The practical definition of this target was a prevalence of less than 1 registered case per 10,000 of the global population. Progress towards the target was helped by the free provision of multidrug therapy (MDT) from 1995 through the World Health Organisation

\* Correspondence: r.e.crump@warwick.ac.uk

<sup>1</sup>Warwick Infectious Disease Epidemiology Research, School of Life Sciences, Gibbet Hill Campus, The University of Warwick, Coventry CV4 7AL, UK  
Full list of author information is available at the end of the article

(WHO), initially by the Nippon Foundation and since 2000 by Novartis. The global target was achieved, but leprosy remains an important public health issue at a national level for a number of countries and in some regions of other (previously high burden) countries [5, 6].

Early diagnosis and use of MDT are the principle tools used to combat leprosy [1]. The social stigma associated with leprosy remains a barrier to early diagnosis, and education is seen as an important step to overcome this. The bacillus Calmette-Guérre (BCG) tuberculosis vaccine also has a protective effect against *M. leprae* infection [7], although the relative contribution of these interventions to current control has not been quantified.

The latest WHO target for leprosy is a reduction in the number of new cases with grade 2 disability to below one per million of the global population by 2020 [5]. Grade 2 disability involves deformity of the hands or feet, or severe visual impairment. The national number of new cases with grade 2 disability is one of the variables reported annually to the WHO, along with the total number of new cases and the number of new multibacillary cases.

We consider elimination as the process of reducing the incidence of infection until transmission is zero. The aim of modelling in the context of elimination is two-fold. First, to provide quantitative forecasts of future cases based on data and current understanding. Such forecasts (made public) are essential as a test of programme management and evaluation, and are an early warning of changes in programme effectiveness. Given the long and variable time periods between infection and disease in leprosy, the incidence of disease is a biased measure of the incidence of infection (unless the infection rate is constant). For example, when incidence is exponentially decreasing, there is a bias towards individuals who have longer infection times [8]. Consequently, changes in incidence of disease require interpretation before they can be used to infer changes in incidence of infection. The second, and most common, use of modelling for elimination is the prediction of impacts of changes to control programmes, this type of modelling has been applied to leprosy [9]. Here we concentrate on the first of these aims: forecasting and interpretation. The method we have chosen to use is back-calculation [10], which provides estimates of past incidence of sub-clinical infections, current prevalence of (undiagnosed) sub-clinical and clinical infections and short-term forward predictions of the numbers of new detected and reported cases. Note that even if transmission is halted, cases will still arise in the future as infected individuals progress through disease stages.

Back-calculation [10] has been used to study the infection dynamics of a variety of infectious diseases with long and variable incubation period distributions, i.e. the

time between infection and diagnosis, for example HIV/AIDS, BSE, vCJD, and also cancer. Each observed diagnosis is the result of a previous infection. Therefore, given information on diagnoses and the incubation period distribution (IPD), empirical estimates can be made of the number of infections per time interval. The diagnoses made during one interval are a mixture of individuals infected at different times in the past; for example, individuals diagnosed in 2015 might include people infected in the years 1990 to 2010, depending on the infection rates in those years, and the IPD.

For discrete time periods, the basic back-calculation equation is:

$$D_i = \sum_{j=1}^i I_j f_{i-j}$$

where  $D_i$  is the number of new diagnoses in the  $i^{\text{th}}$  time interval,  $I_j$  is the expected number of infections in the  $j^{\text{th}}$  time interval and  $f_{i-j}$  is the probability that the time between infection and diagnosis is  $i-j$  time intervals. In the simplest case  $f_{i-j}$  comes directly from the IPD. The logic of back-calculation is that given  $D$  and  $f$ , the convolution above allows estimation of  $I$ . Here, we adopt a Bayesian approach to estimate posterior densities, with a concentration on estimation of variability in  $I$  as much as central locations, and so include some variability in the time period distributions,  $f$ , and estimate some relative factors. It is impossible to estimate both  $I$  and  $f$  simultaneously because of non-identifiability, but we can use prior knowledge and external data to provide estimates of  $I$  given uncertainty in  $f$ .

This basic model can be modified to reflect the natural history of the disease under study. For example, AIDS diagnosis was originally modelled as the end-point for disease [10] but this has subsequently been refined to incorporate both AIDS and HIV diagnoses as well as a number of other intermediate disease states based on CD4 count [11].

Previous uses of back-calculation have concentrated on epidemic or endemic situations (i.e. where  $I$  is increasing or constant). In contrast, leprosy is globally decreasing, and back-calculation might be more generally useful in disease control situations approaching elimination. In the case of leprosy, the time to diagnosis also depends on the public health system in that more active case finding will result in faster diagnosis. Consequently, we will consider time-varying hazards associated with the distribution of time from onset of symptoms to diagnosis, as well as time-varying rates of infection.

## Methods

The endemic nature of leprosy and its long incubation period mean that data on individuals' time of infection

are difficult to obtain. Consequently, information on the IPD and the distribution of the time period from onset of clinical symptoms to detection, the detection delay distribution (DDD), must be obtained independently and then incorporated into the back-calculation. Both the fitting of distributions to time periods and the back-calculation were implemented in a Bayesian Markov Chain Monte Carlo (MCMC) setting. The program Just Another Gibbs Sampler (JAGS) [12] was used, with the R statistical environment [13] acting as the front end.

### Fitting time period distributions

#### Data on time of infection, onset and diagnosis

Data on 49 individual cases of leprosy with short exposure periods were extracted from the literature [14–20]. These were mostly military service personnel who spent known, limited periods of time in endemic countries before returning to live in non-endemic communities and subsequently being diagnosed.

#### Modelling time period distributions

For each individual, the observed endpoints for each time period (onset of symptoms and diagnosis) were modelled as  $t_o = t_i + p_{i;o}$  and  $t_d = t_i + p_{i;o} + p_{o;d}$ , respectively, where  $t_i$  was the unobserved time of infection and  $p_{i;o}$  and  $p_{o;d}$  were the periods from incubation to onset and from onset to diagnosis. Infection time was a uniformly distributed variable between the start and end of exposure and the two time periods were gamma distributed random variables. The observed values,  $t_o$  and  $t_d$ , were both left and right censored, with onset and diagnosis dates taken as being known to within some time interval. The Gibbs Sampler was used to obtain samples from the posterior probability distributions of the shape and rate parameters of the gamma distributions.

### Back-calculation

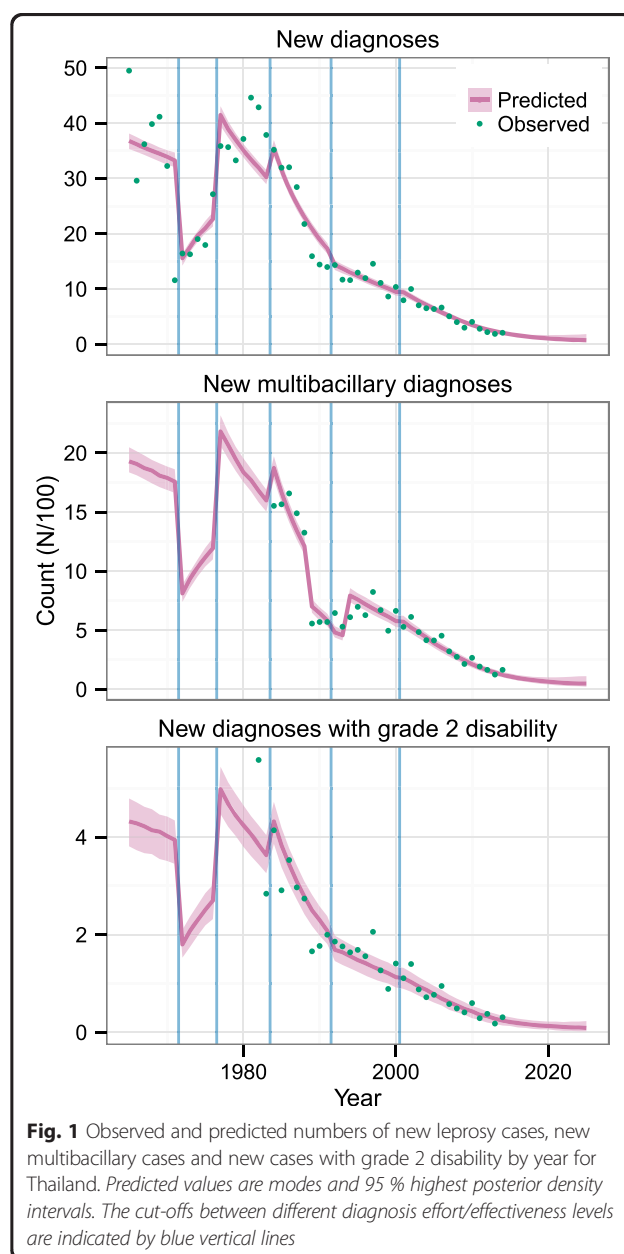
#### National incidence of new leprosy cases

The data used consisted of annual numbers of new leprosy cases, new multibacillary leprosy cases and new leprosy cases with grade 2 disability reported at the national level for Thailand. These data were taken from published reports [2, 21–28]. The variables reported to the WHO have changed over time, resulting in the number of new cases detected, number of new cases with grade 2 disability and the number of new multibacillary cases being available from 1965, 1982 and 1984, respectively. Data on the three variables were available up to 2014. Since 1993, cases of multibacillary leprosy have been identified via case classification (with >5 skin lesions indicating a multibacillary case) rather than a bacterial index score. The transition from the use of the bacterial index to case classification took place from 1989 to 1993. From 1984 to

1988 the number of new lepromatous leprosy cases identified under the Madrid classification system [29] was used as the count of new multibacillary cases [21]. The observed values are plotted against year in Fig. 1.

#### Back-calculation model

The progression from sub-clinical to clinical infection and subsequent diagnosis of leprosy was described by a discrete time period model. Calendar years were used as the discrete time step, in keeping with the annual reporting of national summary information to the WHO. The time period of the analysis was set to start 30 years before the first observed record, to



**Fig. 1** Observed and predicted numbers of new leprosy cases, new multibacillary cases and new cases with grade 2 disability by year for Thailand. Predicted values are modes and 95 % highest posterior density intervals. The cut-offs between different diagnosis effort/effectiveness levels are indicated by blue vertical lines

allow the system to generate an appropriate pattern of infection to explain the early observations, leprosy being an endemic disease.

The expected incidence of infections in year  $b$  was simulated with a random walk process. The observed number of infections in year  $b$  was modelled as:

$$I_b \sim \text{Poisson}(\lambda_b)$$

where prior distributions for the  $\lambda_b$  were

$$\lambda_b \sim \begin{cases} \text{Uniform}(0, 10000) & \text{if } b = 1 \\ \mathcal{N}(\lambda_{b-1}, \sigma^2), \lambda_b > 0 & \text{if } b > 1 \end{cases}$$

The standard deviation,  $\sigma$ , used in the random walk process was set to 30 for the analyses reported in the paper. This was chosen from a number of test runs of the model to provide a smooth downward curve of infections over time while still maintaining some variation. There is no information in the data to allow  $\sigma$  to be estimated within the analyses.

Progress from sub-clinical infection to the onset of symptoms was controlled by the IPD, a gamma distribution with constant parameters within a single MCMC chain. The parameters of the IPD were taken from the analyses reported here. From the numbers of sub-clinical infections, the expected number of new clinical infections in any year can be calculated using the IPD:

$$E[O_c] = \sum_{b=1}^c I_b f_{c-b}$$

Where  $f_{c-b}$  was the probability that time between sub-clinical infection and onset of clinical symptoms was  $c-b$  years. The observed number of clinical infections in year  $c$  was then  $O_c \sim \text{Poisson}(E[O_c])$ .

There is no test for sub-clinical leprosy. Therefore, diagnosis occurs subsequent to the onset of symptoms in the model. The probability of being diagnosed in any year given that onset occurred in some known previous year was derived from the onset to diagnosis time period distribution; again using a gamma distribution with constant parameters within a single MCMC chain with the parameters taken from the separate time period fitting reported in this paper. We additionally added a parameter reflecting the effort being put into diagnosis at that time (i.e. scaling the diagnosis hazard for that year).

Variations in the effort put into diagnosis programmes, or the effectiveness of these programmes, is handled by scaling the hazard of being diagnosed in year  $d$  given that onset of clinical symptoms occurred in year  $c$  ( $h_{dc}$ ). The probability of an individual being diagnosed in year  $d$  given that the onset of clinical symptoms was in year  $c$  is:

$$g_{dc} = \gamma_k h_{dc} (1 - G_{dc})$$

Where  $G_{dc} = \sum_{i=1}^{d-1} g_{ic}$ , diagnosis effort is modelled by  $\gamma_k$  and  $h_{dc}$  comes from the DDD:

$$h_{dc} = \frac{\text{Prob}(c-1 \leq X < c)}{1 - \text{Prob}(X < c-1)}$$

For the Thailand data, six different effort parameters were included to correspond to periods in which the diagnosis effort or effectiveness may have differed [21], as set out in Table 1. The effort parameters,  $\gamma_k$ , were modelled using a random walk, such that effort in period  $k$  was a change from the effort in the previous period. The prior distributions for  $\gamma_k$  were:

$$\gamma_k \sim \begin{cases} \text{Uniform}(1, 0.25^2) & \text{if } k = 1 \\ \mathcal{N}(\gamma_{k-1}, 0.25^2) & \text{if } k > 1 \end{cases}$$

The expected number of diagnoses in year  $d$  was then:

$$E[D_d] = \sum_{j=1}^d E[O_j] g_{dj}$$

The observed number of new diagnoses in year  $i$  was a Poisson variable with parameter  $E[D_i]$ ,  $D_i \sim \text{Poisson}(E[D_i])$ . The probability of a multibacillary diagnosis being made in year  $d$  given that onset of clinical symptoms happened in year  $c$  was assumed to be unrelated to both the length of time since onset and calendar time;  $f_{dc}^M = \phi_{Mi} f_{dc}$ , where  $\phi_{Mi}$  was the proportion of multibacillary cases expected in the  $i^{\text{th}}$  period ( $d < 1989$ ,  $1989 \leq d \leq 1993$  and  $d \geq 1994$ ) and  $f_{dc}$  was the probability of being diagnosed in year  $d$  given that the onset of clinical symptoms occurred in year  $c$ . The same prior was used for all three  $\phi_{Mi}$ ,  $\phi_{Mi} \sim \text{Beta}(2, 2)$ .

The existence of a relationship between the proportion of grade 2 disability cases and detection delay has been demonstrated previously [30]. This relationship was incorporated into the back-calculation model as a linear regression on detection delay. The probability of a new case with grade 2 disability being made in year  $d$  given that onset of symptoms happened in year  $c$  was taken to be  $f_{dc}^D = \min(\phi_D + (d - c + 1)\kappa_D, 1)f_{dc}$ . Uniform priors were assumed for  $\phi_D$  and  $\kappa_D$ , ranging from zero to 0.25 for both parameters.

The uncertainty about the parameters associated with the IPD and DDD was incorporated into the back-calculation via multiple imputation [31] to avoid problems arising from the lack of information in reported case data on the parameters of the IPD and DDD. Multiple MCMC chains were run, with each one using a randomly sampled set of parameters from a multivariate

**Table 1** Periods for which the effort or effectiveness of diagnosis may have varied in Thailand, corresponds to parameters fitted

Diagnosis year, <i>d</i>	Why
$d < 1972$	In 1971 leprosy treatment moved into general health-care services
$1972 \leq d < 1977$	On-going training of health-care workers, so possible loss of effectiveness
$1977 \leq d < 1984$	In 1984 WHO introduced multidrug therapy (MDT), so this period uses general health-care services but no MDT
$1984 \leq d < 1992$	Both general health-care services and treatment by MDT
$1992 \leq d < 2001$	1991: global target set of 1 new case per 10,000 of population by 2000
$d \geq 2001$	Past the cut-off for the global target, possible change in effort

normal approximation to the posterior probability distribution of the IPD and DDD parameters. A total of 100 chains were run, with 50 samples being retained from each. Each chain was run for 250,000 iterations; after 25,000 adaptation and 150,000 burn-in samples one out of every 1,500 samples was retained. The resulting 5,000 samples were merged to provide an overall posterior probability distribution for the parameters of interest. The chains were given equal weight in the merging as they were expected to fit the data equally well.

**Results**

**Time period distributions**

The IPD and DDD were fitted as gamma distributions with a common rate parameter. In this way the distribution of the sub-clinical infection to detection period is also a gamma distribution with the same rate parameter and a shape parameter equal to the sum of the shape parameters of the two constituent distributions. Table 2 contains a summary of the posterior probability distributions of the three parameters. The parameters are strongly correlated with each other and these relationships were taken into account in the back-calculation. This was done through the use of a multivariate normal approximation to the joint posterior probability distribution of these parameters from which samples were drawn for the multiple imputation runs.

The posterior samples of the time period distribution parameters were used to generate samples of the gamma density for each curve at each time point. These were used to create Fig. 2, which shows the mode and 95 % highest posterior density interval (HPD95) of the two distributions. The modal value of the IPD was 3.8 years (HPD95: 2.10 to 5.64 years). For 95 % of infected individuals, onset occurs by 17.8 years after infection (modal value, HPD95: 15.2 to 22.3 years). Detection occurred for 95 % of clinical infections by 11.4 years post-onset (modal value, HPD95: 9.4 to 14.9 years). The mode of the sub-clinical infection to diagnosis period distribution was at 7.9 years (HPD95: 5.9 to 9.7 years), and 95 % of clinical infections were detected by 23.6 years after sub-clinical infection (modal value, HPD95: 20.4 to 29.2 years).

**Trends in cases**

Figure 3 shows the predicted trends in the numbers of new sub-clinical and clinical infections and reported cases for 1950–2025. The trends in sub-clinical infections and clinical infections are smooth with the clinical infection curve lagging the sub-clinical infection curve slightly.

The diagnosis hazard in any year is scaled by a diagnostic effort parameter, and this enables the trend in new diagnoses to follow the observed data points reasonably well (Fig. 1). The posterior probability distributions of the six effort parameters are in Fig. 4. During the period where leprosy care was being incorporated into the general health care system and workers were being trained (1972–1977), fewer diagnoses were being made than would be predicted based on the DDD. The modes of the parameter distributions for the other five intervals are above 1, indicating that more diagnoses are made than predicted just using the DDD estimated previously. In particular, the highest diagnostic effort/effectiveness was before 1972, when leprosy care joined the general health care system.

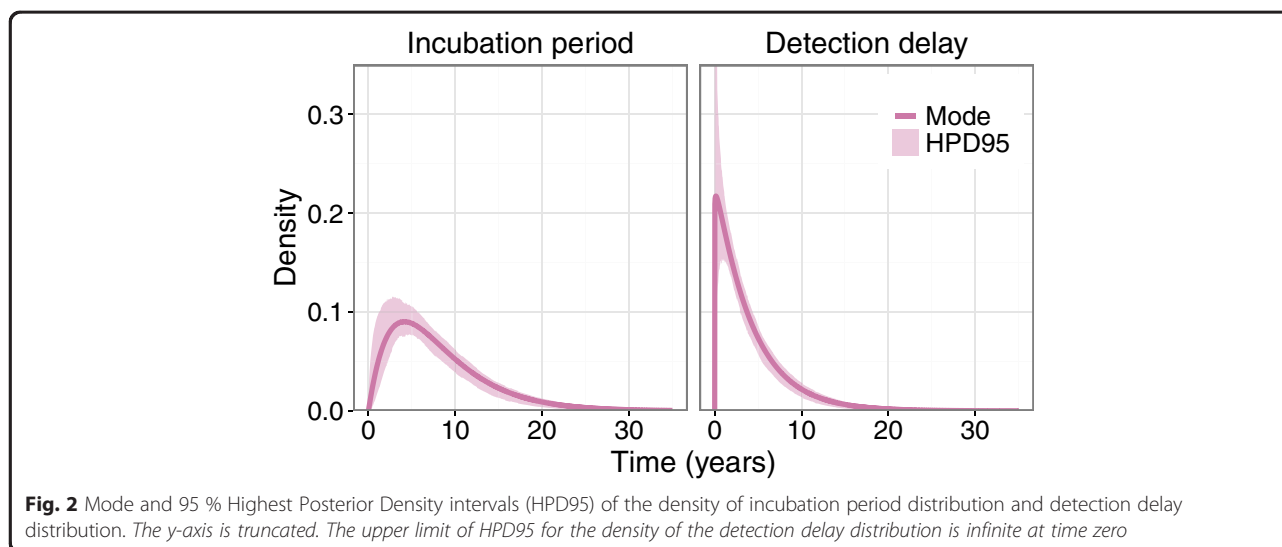
Figure 1 shows the predicted numbers of new cases, new multibacillary cases and new cases with grade 2 disability along with the observed numbers and the cut-offs for different periods of diagnostic effort. The observed values are quite well predicted by the analysis.

The scale which results from the large reduction in the number of new cases over recent decades prevents the variation of the posterior probability distribution

**Table 2** Summary of posterior probability distributions of the shape parameters for the incubation period and detection delay distributions ( $a_{i,o}$  and  $a_{o,d}$ ) and the common rate parameter ( $\beta$ )

Parameter	Mode	95 % Highest Posterior Density Interval	Correlation with	
			$a_{i,o}$	$a_{o,d}$
$a_{i,o}$	2.03	1.37 to 2.71		
$a_{o,d}$	1.03	0.73 to 1.37	0.59	
$\beta$	0.25	0.18 to 0.34	0.80	0.72

The incubation and detection delay periods are assumed to be Gamma distributed with a common rate parameter such that the interval from sub-clinical infection to diagnosis is also Gamma distributed with shape parameter  $a_{i,o} + a_{o,d}$  and rate parameter  $\beta$



being as clear as it would be in close-up, but there is an increase in variation after 2014, when the observations end.

#### Prevalence of infection

In line with the observed decline in new cases, the cumulative burden of undiagnosed infections, both sub-clinical and clinical, has also reduced over time (Fig. 5). As the incidence of diagnoses falls, so the uncertainty in numbers of undiagnosed sub-clinical and clinical infections reduces. The uncertainty in incidence of sub-clinical infection in the very recent past (see Fig. 3) contributes to slightly increasing uncertainty in prevalence in the very recent past. However, the uncertainty in the recent past is largely determined by the choice of  $\sigma$  in the random walk.

#### Forecasting of diagnoses

The back-calculation analysis was run excluding data from 2010–2014, in order that these could be predicted (Fig. 6). As with the analysis of the full data, 100 samples were taken from a multivariate normal approximation to the posterior probability distribution of the parameters of the IPD and DDD and included via multiple imputation. The predictions of new cases, new multibacillary cases and new cases with grade 2 disability show the correct trend, but a number of observations lie outside of the HPD95 of their predicted values.

#### Discussion

We have developed a back-calculation approach to reconstruct the past incidence of infection with leprosy in Thailand. This approach also predicts the numbers of individuals with sub-clinical and clinical leprosy who are yet to be diagnosed, and also future predictions of

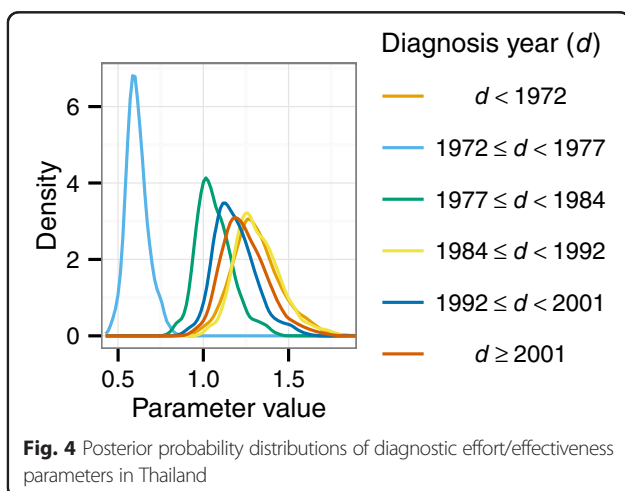
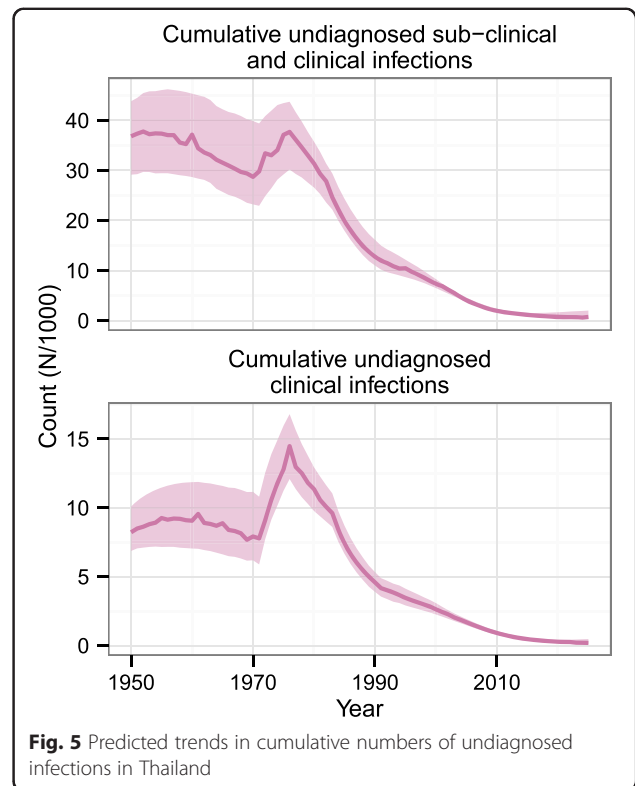
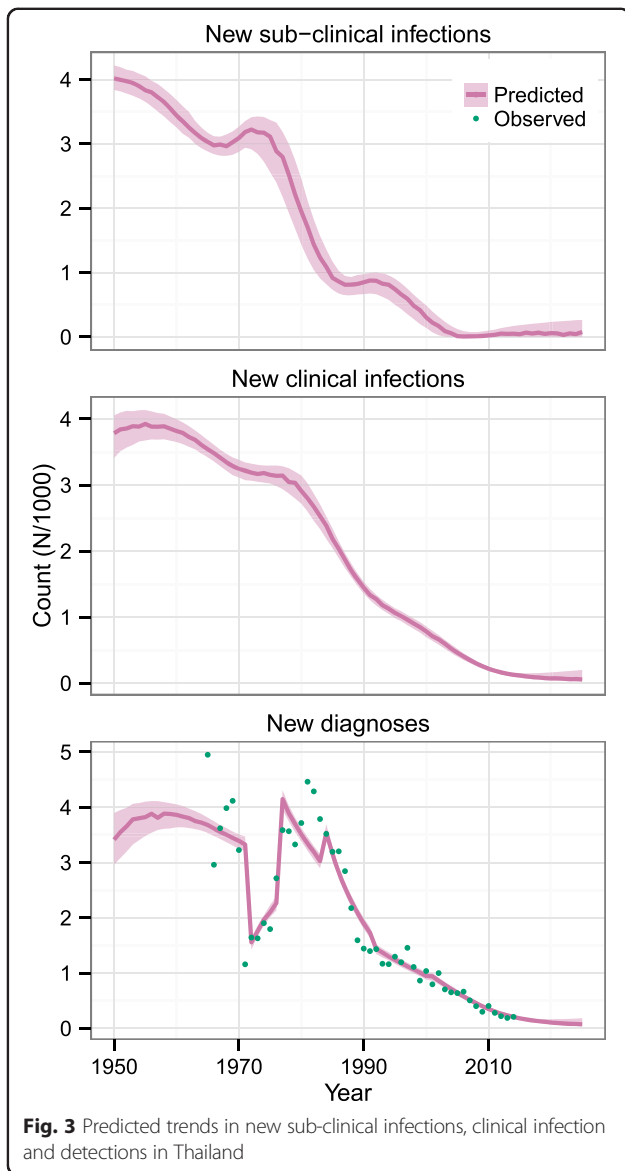
diagnoses of different types. We believe that these results demonstrate the value of this approach.

#### Time period distributions

The use of our estimated time period distributions (IPD and DDD) in the back-calculation analysis is a major source of assumptions. It is assumed that the time period distributions are invariant over time, between sexes and between ages of infected individuals. This is reflected in the transition from sub-clinical infection to clinical infection being completely mediated by the IPD. The assumptions around the DDD are even greater, assuming that detection delays will be far more specific to a country/region and time period. The assumptions about the DDD are mitigated by scaling the hazard associated with diagnosis years to reflect the effort put into, and effectiveness of, case detection. Future work will consider how to extract additional information on time period distributions, particularly as infection rates drop to zero then the last cases are more reflective of the distribution.

#### Trends in cases

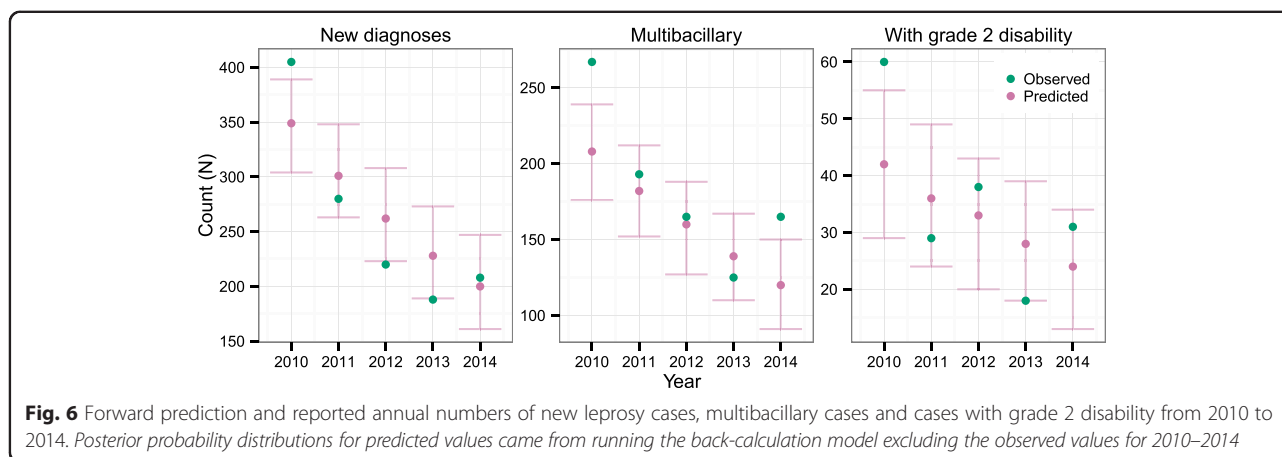
The annual trends in numbers of new sub-clinical infections and clinical infections were smooth curves, with the new clinical infection trend lagging the sub-clinical infection trend slightly. This is expected from the use of an invariant IPD, with nothing else contributing to the hazard of onset of clinical symptoms in any year. The variation in the number of new infections per year is constrained by the choice of standard deviation associated with the random walk. A number of test runs were performed and a value of 30 was chosen which gave a smooth curve without allowing the new sub-clinical infections to be over-reactive to the observations. It might be more realistic to scale the standard deviation of the



random walk in accordance with the number of new infections and this will be investigated in the future. One suggestion is the use of a gamma distribution random walk in which the expected number of new sub-clinical infections in year  $b$  would be  $\lambda_b \sim \text{Gamma}\left(k, \frac{k}{\lambda_{b-1}}\right)$ . The choice of standard deviation in the current random walk largely controls the amount of uncertainty in the incidence of sub-clinical infection (and hence prevalence of sub-clinical infection) in the very recent past. This is essentially unobservable (i.e. sub-clinical infections yesterday will have no impact on diagnoses for a number of years), so the bounds are effectively 0 to the total population.

The highest estimated diagnostic effort/effectiveness parameter corresponded to the period before leprosy care joined the general health care system, which took place after 1971 [21]. This may reflect a relatively lower level of expertise of diagnostic clinicians in the general health care system leading to longer average detection delays. The diagnostic effort parameters express the effort/effectiveness of the programme relative to a value of one, when the DDD would be the sole determining factor. The effort parameters are therefore dependent upon the specific time period distribution used, which in our case would reflect the predominant health care system at the time of diagnosis of our small sample of individuals with information on the time of sub-clinical infection, onset of clinical symptoms and detection.





The models relating to the occurrence of multibacillary and grade 2 disability cases are very simple, but appear to fit quite well. Additional data would be required to allow investigation of more sophisticated relationships between the incidence of multibacillary or grade 2 disability cases and the detection delay, incubation period and other possible risk factors. The results of analyses of these additional data would then be incorporated into the back-calculation rather than trying to estimate the parameters within the back-calculation from little or no information.

#### Prevalence of infection

Note that we are only able to estimate cases of infection that are eventually diagnosed in Thailand. If a significant number of infections resolve spontaneously or do not progress to clinical disease, then our results will underestimate the true numbers of infections. Similarly, if there is significant migration of infections between countries, then our results are less interpretable. In our current model, the cumulative burden of undiagnosed clinical infections is not related to the number or rate of new sub-clinical infections, which might be expected in reality as leprosy is transmissible. This is inherent in the current approach to the analysis; in which infection leads to onset and subsequent diagnosis with no mechanism by which the burden of infection impacts upon the number of future new infections. Future work will consider whether it is feasible to make the incidence of sub-clinical infection a function of the prevalence of clinical, or sub-clinical, infection.

#### Forecasting of diagnoses

Our preliminary attempts to forecast cases from 2010–2014 reproduced the decreasing trend but do not fully reflect the variation in the observed values, with several observations falling outside the HPD95 of their predicted values. Given the simplicity of our model, in

particular for the prediction of the numbers of new multibacillary cases and new cases with grade 2 disability, this appears to have been quite successful. Future work will include refining the model, including with respect to predicting the numbers of new multibacillary cases and grade 2 disability cases reported. This may require additional data to enable the investigation of parameters associated with some aspects of the model external to the back-calculation, as has been done in other studies [32].

#### Reaching the 2020 goal

The goal for 2020 is a reduction in the annual number of new cases with grade 2 disability to below 1 per million of the global population. United Nation predictions of the population of Thailand in 2020 have a 95 % probability interval from 67.8 to 69.4 million, with a median value of 68.6 million [33]. The cut-offs for 1 new case with grade 2 disability per million people are therefore from 67.8 to 69.4. Our observations are that this has already been achieved *within Thailand*, and that the predicted 95 % highest probable density interval of the number of new grade 2 disability cases in 2020 goes from five to 22, with a mode of 13. While it is not possible to extrapolate from predictions for a single country to the global leprosy situation, more meaningful observations about the 2020 goal will be possible when equivalent analyses are run on data from the countries reporting the majority of the world's new cases for leprosy. This will constitute a major part of the next phase of our research programme.

#### Conclusions

Based on this initial analysis the Bayesian back-calculation approach shows much promise as a tool to provide insights into numbers of individuals with sub-clinical or clinical leprosy infections but as yet undiagnosed. Predictions of the expected number of future

diagnoses may help to provide some understanding of the elimination process.

Our model remains under development. As well as making improvements around the current analysis, infection incidence by age and gender remain to be incorporated. Differences associated with the infection to onset and onset to diagnosis time period distributions associated with time, age and gender are also being considered.

We suggest that there may be value of using this method of analysis for other neglected tropical diseases in which the incidence of diagnosis is not an immediate measure of infection.

#### Abbreviations

DDD: Detection delay distribution; HPD95: 95 % highest posterior density interval; IPD: Incubation period distribution; MCMC: Markov Chain Monte Carlo; WHO: World Health Organisation.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

REC performed the analyses and drafted the manuscript. GFM conceived the study, assisted with methodology and revised the manuscript. Both authors read and approved the final manuscript.

#### Acknowledgements

REC and GFM gratefully acknowledge funding of the NTD Modelling Consortium by Novartis Foundation. The views, opinions, assumptions and conclusions or any other information set out in this article are solely those of the authors and not of Novartis Foundation. Novartis Foundation was not involved in the establishment of this article and this article should not be attributed to Novartis Foundation or any person connected with Novartis Foundation. We are very grateful to Dr Paul Birrell and his group for providing an invaluable kickstart to the development process by generously providing access to their back-calculation code and expertise. Earlier versions of our model were tested using data kindly provided by the Anandaban Hospital, Kathmandu, Nepal. We acknowledge the help of Preethi George, Savannah Watts and João França.

#### Author details

<sup>1</sup>Warwick Infectious Disease Epidemiology Research, School of Life Sciences, Gibbet Hill Campus, The University of Warwick, Coventry CV4 7AL, UK.

<sup>2</sup>Public Health & Policy, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK.

Received: 28 August 2015 Accepted: 5 October 2015

Published online: 22 October 2015

#### References

- Rodrigues LC, Lockwood DNJ. Leprosy now: Epidemiology, progress, challenges, and research gaps. *Lancet Infect Dis*. 2011;11:464–70.
- World Health Organisation. Global leprosy update, 2013; reducing disease burden. *Wkly Epidemiol Rec*. 2014;89(36):389–400.
- Bratschi MW, Steinmann P, Wickenden A, Gillis TP. Current knowledge on *Mycobacterium leprae* transmission: A systematic literature review. *Lepr Rev*. 2015;86(2):142–55.
- World Health Assembly. Leprosy resolution WHA 44.9 : Forty-fourth world health assembly, 13 may 1991. [http://www.who.int/neglected\\_diseases/mediacentre/WHA\\_44.9\\_Eng.pdf](http://www.who.int/neglected_diseases/mediacentre/WHA_44.9_Eng.pdf); 1991.
- World Health Organisation. Accelerating work to overcome the global impact of neglected tropical diseases: a roadmap for implementation. 2012. [http://www.who.int/neglected\\_diseases/NTD\\_RoadMap\\_2012\\_Fullversion.pdf](http://www.who.int/neglected_diseases/NTD_RoadMap_2012_Fullversion.pdf)
- Brook CE, Beauclair R, Ngwenya O, Worden L, Ndeffo-Mbah M, Lietman TM, Satpathy SK, Galvani AP, Porco TC. Spatial heterogeneity in projected leprosy trends in India. *Parasit Vectors*. 2015;doi:10.1186/s13071-015-1124-7.
- Merle CS, Cunha SS, Rodrigues LC. BCG vaccination and leprosy protection: Review of current evidence and status of BCG in leprosy control. *Expert Rev Vaccines*. 2010;9(2):209–22.
- Cox DR, Medley GF. A process of events with notification delay and the forecasting of AIDS. *Philos Trans R Soc Lond, B, Biol Sci*. 1989;325(1226):135–45.
- Blok DJ, de Vlas SJ, Richardus JH. Global elimination of leprosy by 2020: Are we on track? *Parasit Vectors*. 2015;doi:10.1186/s13071-015-1143-4.
- Brookmeyer R, Gail MH. A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *J Am Stat Assoc*. 1988;83(402):301–8.
- Birrell PJ, Gill ON, Delpech VC, Brown AE, Desai S, Chadborn TR, et al. HIV incidence in men who have sex with men in England and Wales 2001–10: a nationwide population study. *Lancet Infect Dis*. 2013;13:313–8.
- Plummer M. JAGS Version 3.4.0 user manual. 2013. <http://mcmc-jags.sourceforge.net/>
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2015. <http://www.R-project.org/>
- Brubaker ML, Binford CH, Trautman JR. Occurrence of leprosy in U.S. veterans after service in endemic areas abroad. *Public Health Rep*. 1969;84(12):1051–8.
- Hasseltine HE. Leprosy in men who served in United States military service. *Int J Lepr Other Mycobact Dis*. 1940;8:501–8.
- Doyle JO. Case of leprosy seen in a venereal disease clinic in Britain. *BMJ*. 1953;2:261–2.
- Rogers J, Adamson DG. Leprosy: Report on four cases. *BMJ*. 1953;2(4830):259–60.
- Medford FE. Leprosy in Vietnam veterans. *Arch Intern Med*. 1974;134:373.
- Rose HD. Leprosy in Vietnam returnees. *J Am Med Assoc*. 1974;230(10):1388.
- Brickell K, Frith R, Ellis-Pegler R. Leprosy in a backpacker. *J Travel Med*. 2005;12(3):161–3.
- World Health Organisation. Leprosy control in Thailand: Trends in case detection, 1965–2005. *Wkly Epidemiol Rec*. 2007;82(30):261–71.
- World Health Organisation. Global leprosy situation, 2007. *Wkly Epidemiol Rec*. 2007;82(25):225–32.
- World Health Organisation. Global leprosy situation, beginning of 2008. *Wkly Epidemiol Rec*. 2008;83(33):293–300.
- World Health Organisation. Global leprosy situation, 2009. *Wkly Epidemiol Rec*. 2009;84(33):333–40.
- World Health Organisation. Global leprosy situation, 2010. *Wkly Epidemiol Rec*. 2010;85(35):337–48.
- World Health Organisation. Leprosy update, 2011. *Wkly Epidemiol Rec*. 2011;86(36):389–400.
- World Health Organisation. Global leprosy situation, 2012. *Wkly Epidemiol Rec*. 2012;87(34):317–28.
- World Health Organisation. Global leprosy: Update on the 2012 situation. *Wkly Epidemiol Rec*. 2013;88(35):365–79.
- Santos VS, de Mendonça Neto PT, Falcao Raposo OF, Fakhouri R, Prado Reis F, Corrêa Feitosa VL. Evaluation of agreement between clinical and histopathological data for classifying leprosy. *Int J Infect Dis*. 2013;17:e189–92.
- van Veen NH, Meima A, Richardus JH. The relationship between detection delay and impairment in leprosy control: A comparison of patient cohorts from bangladesh and ethiopia. *Lepr Rev*. 2006;77(4):356–65.
- Plummer M. Cuts in Bayesian graphical models. *Statistical Computation*. 2015;25:37–43.
- Birrell PJ, Chadborn TR, Gill ON, Delpech VC, Angelis DD. Estimating trends in incidence, time-to-diagnosis and undiagnosed prevalence using a CD4-based Bayesian back-calculation. *Stat Commun Infect Dis*. 2012;4(1):Article 6.
- United Nations. Probabilistic population projections based on the world population prospects: The 2015 revision [Internet]. Population Division, DESA; 2015. <http://esa.un.org/unpd/ppp/>. Accessed 17 Aug 2015.