Self-healing and Underreporting of Cases of Visceral Leishmaniasis in Bihar, India: A Mathematical Modeling-based Study

Olivia F. Prosper[1], Swati DebRoy[2], Austin Mishoe[2], Cesar Montalvo[3] Niyamat Ali Siddiqui[4], Pradeep Das[4], Anuj Mubayi[3, ✉]

[1] *University of Kentucky, Lexington, USA*
[2] *University of South Carolina, Beaufort, USA*
[3] *Simon A. Levin Mathematical, Computational and Modeling Sciences Center, School of Human Evolution and Social Change, Arizona State University, Tempe, USA*
[4] *Rajendra Memorial Research Institute of Medical Sciences, Indian Council of Medical Research (ICMR), Patna, India*

✉*Corresponding Author email: amubayi@asu.edu*

Abstract

Background: Underreporting of Visceral Leishmaniasis (VL) in India remains a problem to public health controls. Effective and reliable surveillance systems are critical for monitoring disease outbreaks and public health control programs. However, in India, government surveillance systems are affected by levels of scarcity in resources and therefore, uncertainty surrounds the true incidence of asymptomatic and clinical cases, affecting morbidity and mortality rates. The State of Bihar alone contributes up to the 40% of the worldwide VL cases. The inefficiency of surveillance systems occurs because of multiple reasons including delay in seeking health care, accessing non-authentic health care clinics, and existence of significant asymptomatic self healing infectious cases. This results in a failure of the system to adequately report true transmission rates and number of symptomatic cases that have sought medical advice (thus, high underreporting of cases).

Objectives and Methods: There are several methods to estimate the extent of underreporting in the surveillance system. In this research, we use a mathematical dynamic model and two different types of data sets, namely, monthly incidence for 2003-2005 and yearly incidence from 2006-2012 from the Bihar's 21 most VL affected districts out of its 38 districts. The goals of the study are to estimate critical metrics to measure level of transmission and to evaluate the estimation process between the two data sets and 21 districts. In particularly, our focus is on (i) estimating infection transmission potential, underreporting level in incidence and proportion of self-healing cases, (ii) quantifying reproduction number of the R_0, and (iii) comparing underreporting incidence levels and proportion of self-healing cases between the two periods 2003-2005 and 2006-2012 and between 21 districts.

Results: Our research suggests that the number of asymptomatic individuals in the population who eventually self-heal may have a significant effect on the dynamics of VL spread. The estimated mean self-healing proportion (out of all infected) is found to be ~ 0.6 with only 7 out of 21 affected districts having self-healing proportion less than 0.5 for both data sets. The estimated mean underreporting level is at least 64% for the state of Bihar. The estimates of the basic reproduction numbers obtained are similar in magnitude for most of the districts, being in the range of (0.88, 2.79) and (0.98, 1.01) for 2003-2005 and 2006-2012, respectively.

Conclusions: The estimates for the two types (monthly and yearly) of temporal data suggest that monthly data are better for estimation if less number of data points are available, however, in general, using such data set results in larger variances in parameters as compared to estimates obtained through aggregated yearly data. Estimated values of transmission related metrics are lower than those obtained from earlier analyses in the literature, and the implications of this for VL control are discussed. The spatial heterogeneity in these control metrics increases the risk of epidemics and makes the control strategies more complex.

Keywords— Kala-azar, Dynamical System, Inverse Problem, Spatial Analysis

1 Introduction

Visceral Leishmaniasis characteristics: Visceral Leishmaniasis (VL) is the second deadliest neglected vector-borne disease (VBD) caused by a parasitic protozoan. Leishmaniasis is a neglected tropical disease, characterized by high levels of complexity due to the variations in transmission cycles, clinical manifestations, reactions to therapies, multiple circulating

leishmanial genus. Transmission to humans occurs through more than 30 different female 45
phlebotomine sandflies [13]. Visceral leishmaniasis (VL), also known as kala-azar (black 46
fever in Hindi) is one of the four main syndromes of Leishmaniasis and by far, the most fa- 47
tal one. There exist two mechanisms of transmission of VL: zoonotic VL, transmitted from 48
animal hosts to vector and then to humans and, anthroponotic VL when the disease is dis- 49
seminated from human to vector to human. Two species cause primarily VL: Leishmania 50
donovani and Leishmania infantum. 51

Geographical connotation of VL. VL is a disease that spreads over large geographi- 52
cal areas and it remains endemic accross tropical environments. The World Health Organiza- 53
tion (WHO) estimates around 50,000 to 90,000 new cases of VL every year . More than 90% 54
of the new cases in 2015 occurred in just seven countries: Brazil, Ethiopia, India, Kenya, 55
Somalia, South Sudan and Sudan. India, Bangladesh and Nepal have suffered the largest 56
burden of VL, accounting for 80% of the estimated annual global cases between during the 57
past decade [6]. Just within India, 90% of the cases occur in the state Bihar, which was 58
reported to have more than 10,000 cases in 2014 [16]. In addition, three other states Jhark- 59
hand, West Bengal and Uttar Pradesh suffer from VL as an endemic disease [1]. VL affects 60
other areas in the rest of Asia and several countries in South America, with Brazil being by 61
far the one with the majority of cases. 62

Clinical Epidemiology of VL in Bihar, India. The incubation period for VL lasts 63
commonly between 2 and 6 months [16]. Infected patients present several symptoms includ- 64
ing fever, loss of appetite and weight loss [16]. When infections are persistent, the parasite 65
invades the blood and endothelial system and severe manifestations affect the spleen and 66
liver. Also, anaemia is one of the most common haematological manifestation and further, it 67
may be the key factor triggering leucopenia (reduction of white cells in blood), thrombocy- 68
topenia (low levels of thrombocytes), pancytopenia (deficiency of red cells, white cells, and 69
platelets), hemophagocytosis [19]. As a directed consequence, the infected patient develops 70
fatigue and weakness. Finally, in advanced stages, hypersplenism also affect the infected 71
individual causing him permanent inflammatory problems and most commonly, bleeding. Its 72
name visceral corresponds to the fatal damages that the parasite produces in internal organs 73
and it results fatal in almost 100% of cases within two years if no treatment is provided. 74

Socio economic factors affecting VL in India. VL affects some of the poorest pop- 75
ulation in the world, more frequently, rural areas. The disease is highly associated with 76
conditions of economic stress, such as, overcrowding, malnutrition, poor housing, a fragile 77
immune system, among other characteristics affecting the poor. In India, the disease is be- 78
lieved to attack older children and young adults. The mechanism of transmission may be 79
primarily domestic, and the majority of the incidence is in the male population, in part due 80
to the higher probability of occupational needs in agriculture, cattle grazing, construction, 81
etc. VL becomes a heavy economic burden for communities with both, social and financial 82
costs. It adversely affects individuals and leaves a social stigma in the population,which 83
affects productivity and welfare of communities. Additionally, diagnosis and treatment are 84
associated with high costs for families. Malnutrition status has been considered a signifi- 85
cant factor to VL progression and infection level. In fact, malnourished individuals are more 86
likely to show prominent clinical symptoms. World Bank estimates that around 38% of the 87
population in India suffers from malnutrition in year 2015. According to the estimates of 88
the Planning Commission in India [4], poverty decreased in the country from 45% in 1993 89
to 22% in 2012 but, poverty still remains as one of the most critical issues for well being in 90
the country due to the size of its population. Bihar is the third largest State in India accord- 91
ing to population size and sixth largest according to density (inhabitants per square km). 92
Poverty at the state level decreased from 60.5% in 1993 to 34% in year 2012. The high level 93
of poverty adds further complexity to the issue of VL in the country. 94

Current Challenges in Bihar: Some of the challenges in the Bihar include high cost 95
of treatment when approximately 75% of the VL cases in Bihar live below the poverty 96
threshold of less than US $1.0 a day, available epidemiology data are scarce, access to public 97
health care clinics are difficult, as well. Estimates conducted by [9] show that there exist 98
levels of up to 90% of under reporting of cases of VL. The present work extends the analy- 99
sis to estimate self-healing levels and relations with under reporting. In Bihar, humans are 100
the only known reservoir of L. Donovani [10]. A significant number of patients in Bihar are 101
living with HIV-VL co-infection [10]. Data on the prevalence and associated risk factors on 102
malnutrition among VL patients are scarce. 103

WHO Elimination targets and Interventions: In 2005, the governments of India, 104
Bangladesh and Nepal signed a joint memorandum of understanding to eliminate VL with 105
the aim to reduce the incidence at sub-district level [8]. Because of the high mortality rate 106
of this disease, the WHO has set many deadlines for VL elimination, with the most recent 107

target set at reducing incidence to less than 1 in 10,000 cases by 2020. While this goal has been achieved in some sub-districts of Bihar because of increased surveillance of clinical cases and vector management, high levels of malnutrition and asymptomatic Leishmaniasis in endemic regions continues to present a significant challenge to achieving this target for all of Bihar.

Focus of this study: This study aims to assess the prevalence of self-healing VL individuals in a population and quantify its basic reproduction number for each of the 21 most affected districts of Indian state of Bihar for 2003-2012 via a simple mathematical model. The present study includes both practical and theoretical pieces of work, which contain relevant information about difference between time zones and difference between regions. It highlights three critical aspects in the study of VL. First, the spatial and temporal transmission risks characterized by the rate of transmission and the basic reproductive number. Second aspect of central importance is the asymptomatic prevalence, which has been under studied in the literature. Finally, the under reporting level which, still represents one of the biggest challenges for policy determination and subsequent plans for VL elimination.

2　Materials and Methods

To estimate the true incidence of the reservoir, namely the asymptomatic population and unreported symptomatic cases, we used a very simple mathematical model to describe the transmission dynamics of VL in the human population. Because our goal is to estimate (1) the prevalence of asymptomatic VL cases and (2) the degree of under-reporting in each of the districts of Bihar for which we have data, and because little is known about the sandfly population in this region, we chose to omit the vector dynamics and model transmission as a directly transmitted, rather than vector-borne disease.
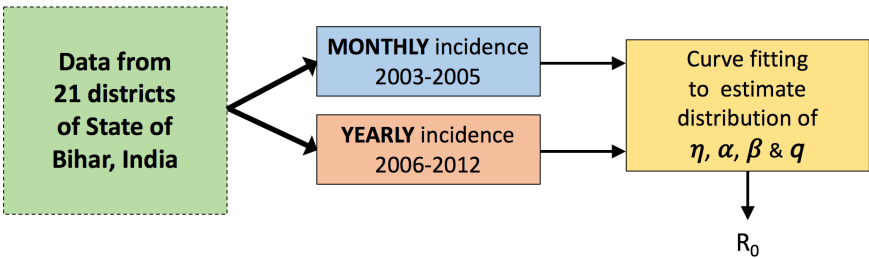

Fig 1. Visceral Leishmaniasis Dynamical Model. Proposed model to estimate and compare distributions of parameters for two different periods of incidence data.

2.1　Data Sources

Population Data

Population data from all 38 districts are obtained, however, because VL has been consistently shown in 21 districts which have high prevalence, the focus in this study is on the 21 districts. Data on population density, population size in rural and urban areas, and number of public health medical facilities in each of the 21 districts are generated from the 2001 and 2011 census [5], under the assumptions that the population of each district increases at an annual rate of 2.1% as reported in the Census. Additionally, socio-economic data are also collected from Census of India, through the Bihar state government website.

Incidence Data

Government health institutions in India collect incidence and disease mortality data. A case is reported when the patient is presented for treatment in a public health clinic. However, this passive case reporting grossly underestimates the number of cases with many cases getting treatment at private medical clinics, and thus are not reported to the state health system. The state health system (State Health Society) of Bihar is divided into 38 District Health Societies (DHS), where each DHS is further subdivided into a number of block-Primary Health Centers (PHCs). PHCs consist of Sub-Centers providing health care to a certain number of villages.

The monthly and annual number of reported VL cases from Bihar's 21 most affected districts (out of 38 districts) from 2003-2005 and 2006-2012, respectively, are used in this study. Data sets belong to the State Health Society of Bihar and the Rajendra Memorial Institute of Medical Sciences (RMRIMS) provided access. The Government of India with assistance from the WHO formed the Kala-azar Task Force in December of 2003. Through this task force, tools were made available for simple and inexpensive tests for field-level diagnosis and treatment at the community level and encouragement to VL activists for case detection [14].

- *Monthly data 2003 - 2005*

    The incidence data are illustrated in Fig 2 for the periods 2003-2005. Months with missing data are shown in white. For the purpose of data fitting, we simulated data for these months using the *inpaint_nans* function on the MATLAB File Exchange [7].
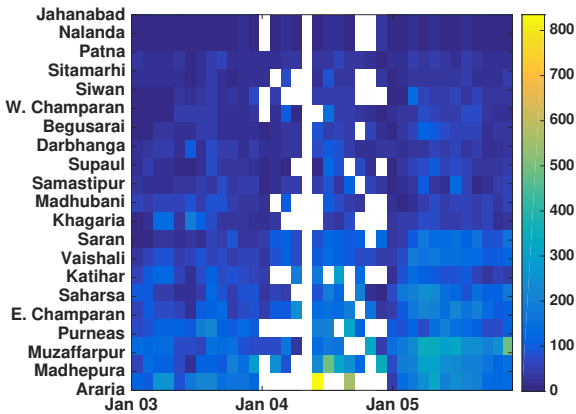

Fig 2. Incidence data. The *monthly* incidence from years 2003 to 2005 for each district is depicted. The districts are ordered according to the mean reported incidence across the respective time period. White regions indicate months with missing data.

- *Yearly data 2006 - 2012*

    Incidence data for the 2006 - 2012 period are taken on an year basis and it is depicted in Figure 3. Note that the highest incidence on this period matches also with the districts in the period 2003-2005 which takes into account monthly data.
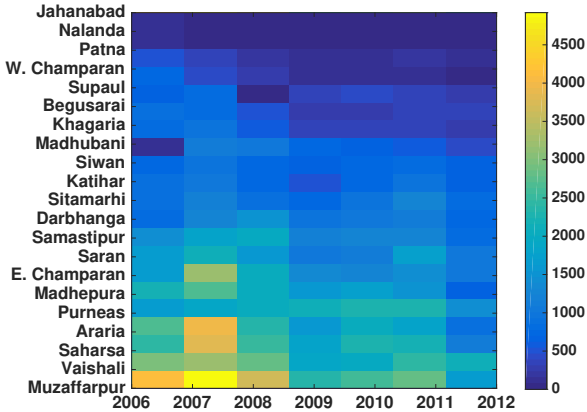

Fig 3. Incidence data. The *yearly* incidence from 2006 to 2012 for each district. Districts are ordered according to the mean reported incidence.

2.2   Model Description

The model presented does not follow the structure of a vector borne disease and there are three elements for this reasoning. First, vector-borne models commonly make use of more parameters and there is a serious limitation with data in reference to neglected tropical diseases. The current analysis favored a simple model with relatively few parameters to maintain consistency with data and avoid high levels of complexity. Also, an SIR model

has been found to present a substantially better fit than the vector-host model after model selection [12]. Finally, the time scale considerations also plays an important role since the mosquito average life span is 20 days while humans live around 70 years. The steady state of the vector, approximated by the quasi steady state, shows that vector dynamics occur very fast and this short time scale may be excluded.

The model used to compute the function $C(t, \theta)$ follows a simple SAIR (Susceptible-Asymptomatic-Symptomatic-Recovered) compartmental framework to describe the disease dynamics of VL in the human population. S denotes the number of susceptible individuals, A the number of infected individuals who are infectious but asymptomatic (or have very mild infection), I the number of infected individuals who are infectious and symptomatic (i.e. clinically ill), and R the number of recovered individuals. Define $N(t) = S(t) + A(t) + I(t) + R(t)$ to be the total population size at time t. We assumed that all individuals are born susceptible at the total birth rate Λ, and all individuals in the population are subject to the natural mortality rate μ. Thus, the total population is modeled by $N'(t) = \Lambda - \mu N$. Because both asymptomatic and symptomatic individuals are assumed to be infectious and because we are omitting the interaction between humans and sand flies, the force of infection (FOI) is given by $\lambda(t) = \beta(I + A)/N$, so that the incidence is $\lambda(t)S$. Asymptomatic individuals either develop symptoms at the rate α, or recover without ever showing symptoms (other than possibly mild symptoms) at the rate η. Symptomatic individuals either recover at a rate γ, or die due to the disease at a rate δ. All state variables and parameters are listed in Table 1 and Table 2, and the above description of the transmission dynamics leads to the system of nonlinear autonomous ordinary differential equations (see Equations (3)-(7) in Appendix 5.1). A flowchart of the model is provided in Figure 4.

Table 1. State Variables

Symbol	Description
$S(t)$	Number of susceptible humans at time t
$A(t)$	Number of asymptomatic humans at time t
$I(t)$	Number of symptomatic humans at time t
$R(t)$	Number of recovered humans at time t
$C(t)$	Cumulative number of infected humans from time 0 to t
$N(t)$	Total human population size at time t

Table 2. Parameter Estimates

Sym	Definition	Estm., Units	Ref.
Λ	Recruitment rate	$\frac{\mu N_0(p-e^{-\mu})}{(1-e^{-\mu})} \cdot \frac{1}{12}$, people·months^{-1}	Census India 2001
μ	Per capita natural mortality rate	$1/(65 \cdot 12)$, months^{-1}	Census India 2001
β	Transmission rate	persons^{-1} · months^{-1}	Estm.
α	Rate of progression to symptomatic	months^{-1}	Estm.
γ	Symptomatic recovery rate	$365/20$, months^{-1}	[17]
η	Asymptomatic recovery rate	months^{-1}	Estm.
δ	Disease induced mortality rate	$365/18$, months^{-1}	[11]
q	Proportion of new cases *not* reported	unitless	Estm.
p	Population growth rate	1.25%, unitless	Census India 2001
$I(0)$	Symptomatic cases at time 0	$1.93 \times$ cases, persons	Estm. [9]
$A(0)$	Asymptomatic cases at time 0	$8.9 \times I(t)$, persons	Estm. [9]

2.3 Linking spatial-temporal data to the Model: Parameter Estimation Technique

The model tracks the cumulative number $C(t, \theta)$, given a vector of parameter values θ, of new VL cases in the time interval $[0, t]$. This function is then fit to the cumulative incidence data for each district using maximum likelihood under the assumption of normally distributed errors in order to estimate the transmission rate, the proportion of all cases that are asymptomatic, and the proportion of cases that go unreported. Letting y_i denote the number of new cases in time interval $[t_{i-1}, t_i]$, $i \in \mathbb{N}$, and defining $c_n := \sum_{i=1}^{n} y_i$, we find the vector of parameter values $\hat{\theta}$ that minimized the sum of squared residuals (SSR)
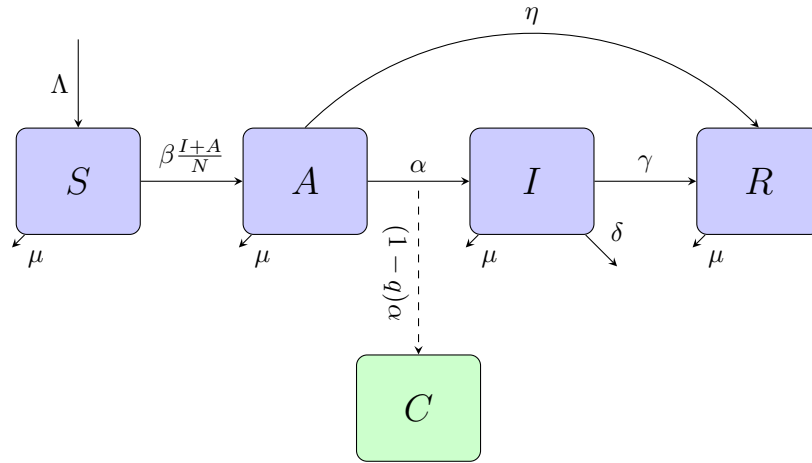
Fig 4. Flowchart of disease dynamics. The state variable $C(t)$ is used to fit to the cumulative incidence data.

$\mathcal{S}(\theta) = \sum_{n=1}^{T}[c_n - C(t_n, \theta)]^2$, where $[0, T]$ is the time span of the dataset. We define $\theta := (\beta, \alpha, \gamma, \eta, q)^t$, where β is the transmission rate, $1/\alpha$ is the average time spent asymptomatic after exposure to VL before developing symptoms, $1/\gamma$ is the average time spent symptomatic, $1/\eta$ is the average time spent asymptomatic before recovering from asymptomatic leishmaniasis, and q is the proportion of new symptomatic VL cases that are reported.

3   Results

3.1   Model mathematical analysis

Using the Next Generation Approach [18], we determine that the basic reproduction number is

$$R_0 = \frac{\beta}{\alpha + \eta + \mu} + \frac{\alpha\beta}{(\alpha + \eta + \mu)(\delta + \gamma + \mu)}.$$

The disease-free equilibrium and endemic equilibrium are $\mathcal{E}_0$ and $\mathcal{E}_1$, respectively, where their components are given in supplementary document. Following the approach in [2], we find condition when a backward bifurcation can occur at $R_0 = 1$. We found that occurrence of backward bifurcation depends on the sign of the parameter α (see Appendix 5.2 for its expression). That is, if $\alpha < 0$, then disease can be eliminated easily with existing intervention provided they are systematically implemented (since $R_0 < 10$. However, if $\alpha > 0$, then VL may persist even when $R_0 < 1$ and much higher efforts to control disease will be needed. To ascertain whether VL can be eliminated from a district, we will calculate R_0 and α for each district.

   Remark: If $R_0 > 1$, the disease is endemic in that district. If $R_0 < 1$ *and* $\alpha < 0$, then VL is moving towards elimination in that district. On the other hand, if $R_0 < 1$ and $\alpha > 0$, we do not have enough information to conclude whether or not VL is moving towards elimination in that district.

   Remark: The mathematical expression representing endemic asymptomatic and symptomatic (clinical) prevalences are given by

$$A^* = \frac{\Lambda(-(\beta - \eta - \mu)(\delta + \gamma + \mu) + \alpha(-\beta + \delta + \gamma + \mu))}{(\alpha + \eta + \mu)(\alpha(\beta - \delta) + \beta(\delta + \gamma + \mu))} \tag{1}$$

$$I^* = \frac{\alpha\Lambda(-(\beta - \eta - \mu)(\delta + \gamma + \mu) + \alpha(-\beta + \delta + \gamma + \mu))}{(\alpha + \eta + \mu)(\delta + \gamma + \mu)(\alpha(-\beta + \delta) - \beta(\delta + \gamma + \mu))} \tag{2}$$

3.2   Estimation of model parameters for 21 districts via curve fitting

3.2.1   Spatial differences for years 2003-2005

We first fit the cumulative incidence $C(t, \theta)$ from the model to the cumulative monthly incidence data from 2003-2005 for each of the 21 districts. The best fit solution curves for Araria, Nalanda, Jahanabad and Samastipur are shown in Fig 5.

High degree uncertainty is observed in the monthly reported cases as compared to an-   230
nual incidence data (see Figure 2) whereas annual data may change significantly due to   231
changes in social and environmental factors. Hence, we used both the data sets (monthly   232
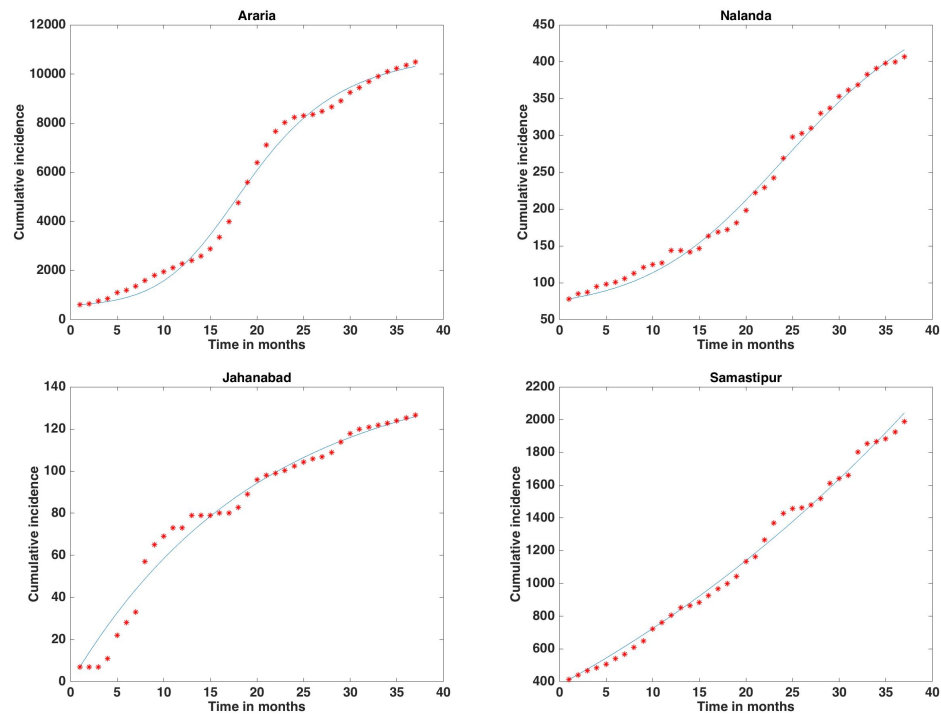and annual) to estimate key model parameters and show the differences in our estimates.   233
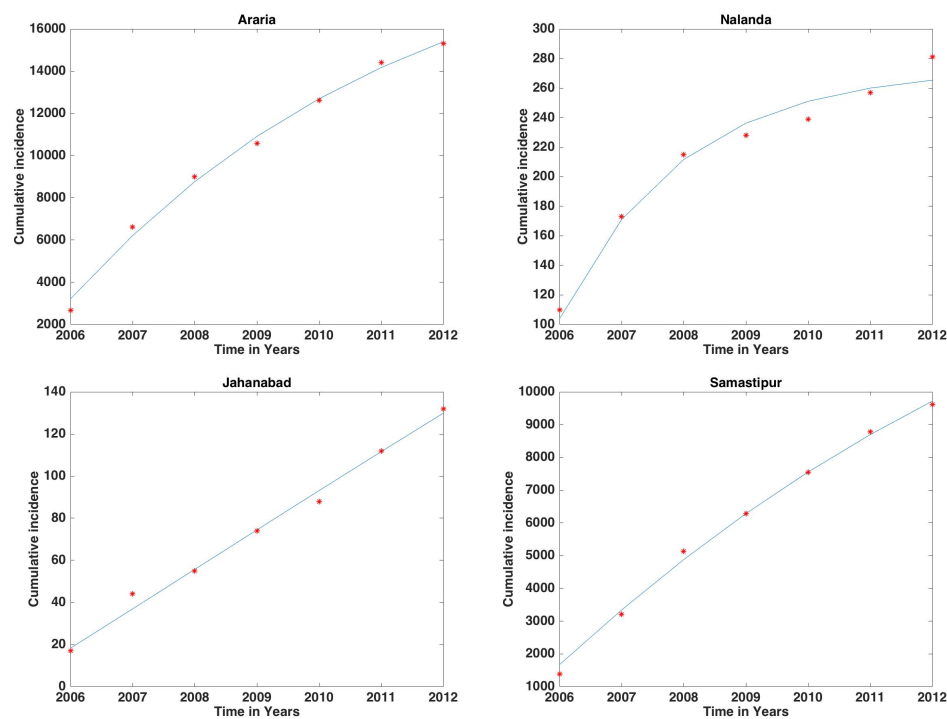

Fig 5. Best fit for Araria, Nalanda, Jahanabad and Samastipur. The best fit of the
solution curve $C(t, \theta)$ (shown as a solid blue line) is overlaid with the cumulative incidence data
(shown with red dots) for the districts with the highest and lowest mean incidence to the
monthly 2003-2005 data.

3.2.2   Spatial differences for years 2006-2012   234

Figure 6 presents the cumulative incidence $C(t, \theta)$ from the model to the cumulative yearly   235
incidence data from 2006-2012 for Araria, Nalanda, Jahanabad and Samastipur. In both   236
previous graphs, the districts were taken arbitrarily to indicate four examples of data fit   237
under the use of the model.   238

3.3   Estimation of R_0 239

The estimation of the basic reproductive number R_0 is presented in figure 7. This figure   240
include both periods and the upper graph corresponds to the 2003-2005 period and the   241
lower is for the 2006-2012 period. The upper graph has been descendantly sorted to identify   242
the districts with higher possibilities to present epidemics and the lower figure maintains   243
the same order of districts for easy comparison between the two periods. It is relevant to   244
recognize that the R_0 falls below the unit in the majority of district in Bihar during the   245
2006-2012 period, changing completely the story from past period 2003-2005 where the   246
majority of districts presented epidemics of VL. This result regarding the potential of epi-   247
demics calls for further research by exploring the activities conducted in each district to try   248
to control VL.   249

3.3.1   Spatial differences for years 2003-2005 and 2006-2012   250

Figure 8 presents the distribution of the estimated basic reproductive number R_0 for both   251
periods. Both periods maintain the big proportion of R_0 around the unit. However, during   252
2003-2005, a large number of districts of Bihar face the event of an epidemic outbreak while,   253
during the 2006-2012 period, the majority of districts had an R_0 less than one, suggesting   254
that epidemic of VL was less likely to occur.   255

Fig 6. Best fit for Araria, Nalanda, Jahanabad and Samastipur. The best fit of the solution curve $C(t, \theta)$ (shown as a solid blue line) is overlaid with the cumulative incidence data (shown with red dots) for the districts with the highest and lowest mean incidence for the annual 2006-2012 data.
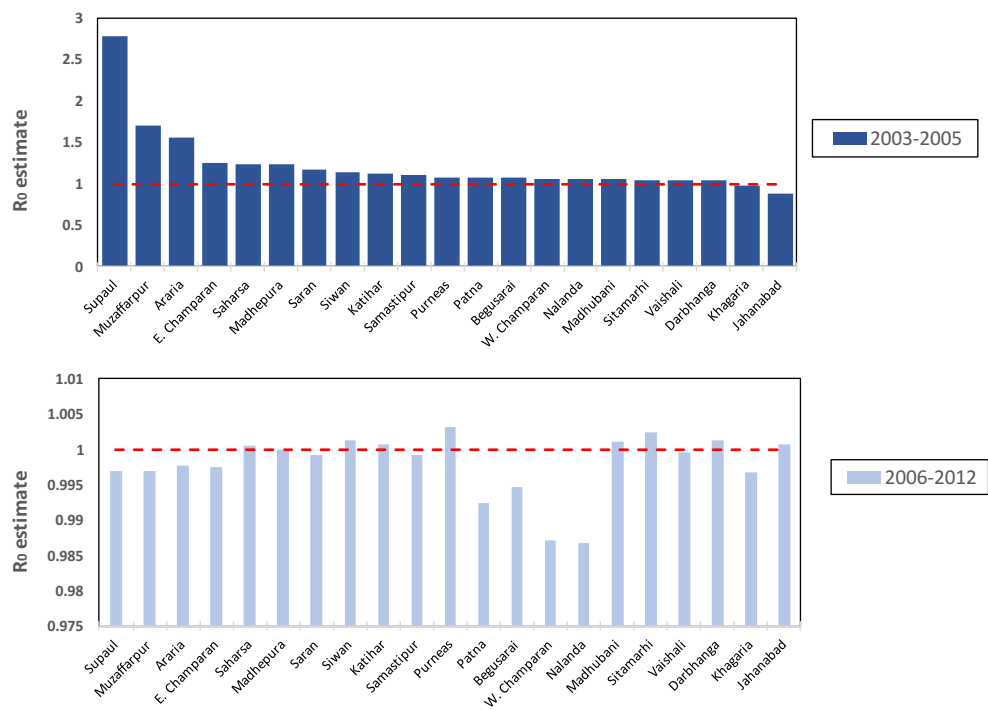

Fig 7. R_0 estimate for 2 periods. R_0 Estimate for 21 districts for period 2003-2005 and 2006-2012.

The self-healing proportion, corresponding to the percentage of population that cure themselves without any conventional medical treatment and whose recovery process is directed by the patient itself using self knowledge or instinct. Table 3 summarizes this pro-
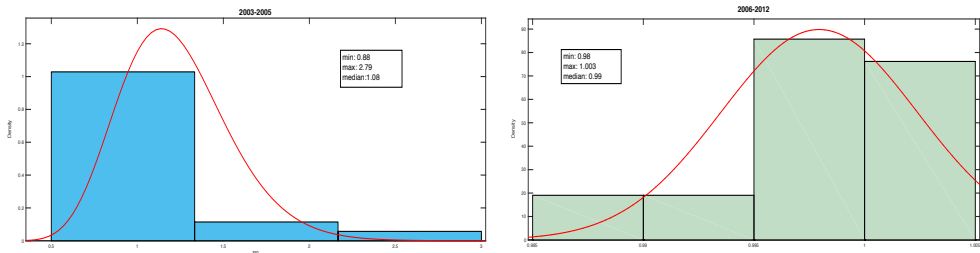
Fig 8. Estimates of R_0. Distribution of basic reproductive numbers include a fitted line using a log likelihood estimation. 2003-2005: Distribution: Gamma Log likelihood: -5.24849, 2006-2012: Distribution: Normal, Log likelihood: 84.4572

portion for every period under consideration. There exist a noticeable difference between the periods: 61% for 2003-2005 in contrast to 58% of 2006-2012. These results might arise new questions with regards to the response of individuals towards self treatment, potentially motivated by more education, higher health coverage, improved medicine availability, among other reasons.

Table 3. Temporal difference in Self-Healing Proportion ($\eta/(\eta + \alpha)$)**, transmission potential** (β)**, and underreporting level** (q)

period	param.	mean	std dev (using 21 districts)	95% confidence interval
2003-2005				
	$\eta/(\eta + \alpha)$	0.61	0.42	(0.43, 0.78)
	β	0.75	0.74	(0.43, 1.06)
	q	0.65	0.35	(0.50, 0.80)
2006-2012				
	$\eta/(\eta + \alpha)$	0.58	0.45	(0.39, 0.77)
	β	0.67	0.46	(0.47, 0.86)
	q	0.64	0.40	(0.47, 0.82)

Figure 9 depicts the goodness of fit through the squared root term of the sum of squared residuals from the estimated models per district. Nalanda and Jahanabad present the best possible fits for both periods since their SSR are close to zero. Vaishali and Araria, on the other hand, have the largest differences for fits, specially for the period 2003-2005. It is worth to note, as well, that these differences in fit are less prominent for the 2006-2012 period, which might be related to the quality of information and levels of under reporting.

Figure 10 summarizes estimates for q and η for the period 2003-2005 (left figures) and also for 2006-2012 (right figures). Recall that q corresponds to the proportion of new cases not reported for every district and η is associated to the asymptomatic recovery rate. The distribution of proportion of new cases not reported tend to be higher during the second period 2006-2012 and also presented a slight higher median. However, the level of asymptomatic recovery rate changed from a concentrated distribution (gamma) on low values during the period 2003-2005 to a higher recovery rate under a normal distribution for the period 2006-2012.

Figure 11 intends to present the level of correlation between the basic reproductive number and the self healing proportion among districts and for both periods. This correlation will be used to confirm that high levels of self healing induce higher chances of an epidemic. The figures present a different slope for every period, negative relationship for 2003-2005 while a positive relation for 2006-2012, suggesting that while self healing used to be more common and tended to reduce slightly the occurrence of and epidemic, during the the second period, a lower SHP might be associated to an increase of an epidemic event. The results from this figures might be the representation of the a behavioral change of communities threatened with kala-azar.

Finally, figure 12 presents the linear relationship connecting the Self-Healing Proportion (SHP) versus the proportion of new cases not reported, q. One important difference between the two periods is the considerable decrease in average proportion of self healing, during the period 2003-2005, a large proportion of individuals tended to use self healing but during the 2006-2012 period there is a SHP below one digit, on average for the 21 districts. If individuals were used to heal by own methods, there might be a connection with a higher level of
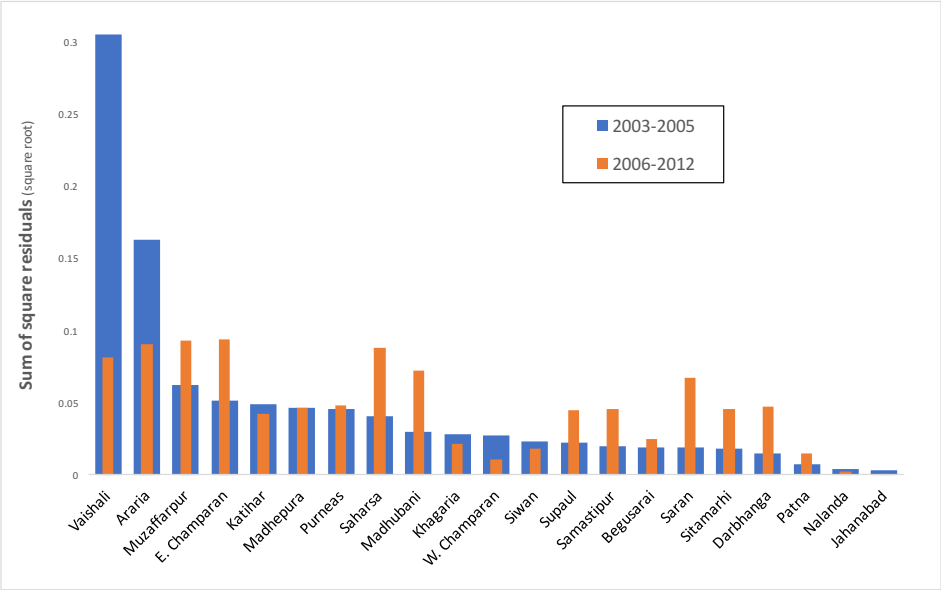
9

Fig 9. Proportion of Sum of Squared Residuals. Estimates of dispersion through the proportion of the sum of squared residuals divided by the total.
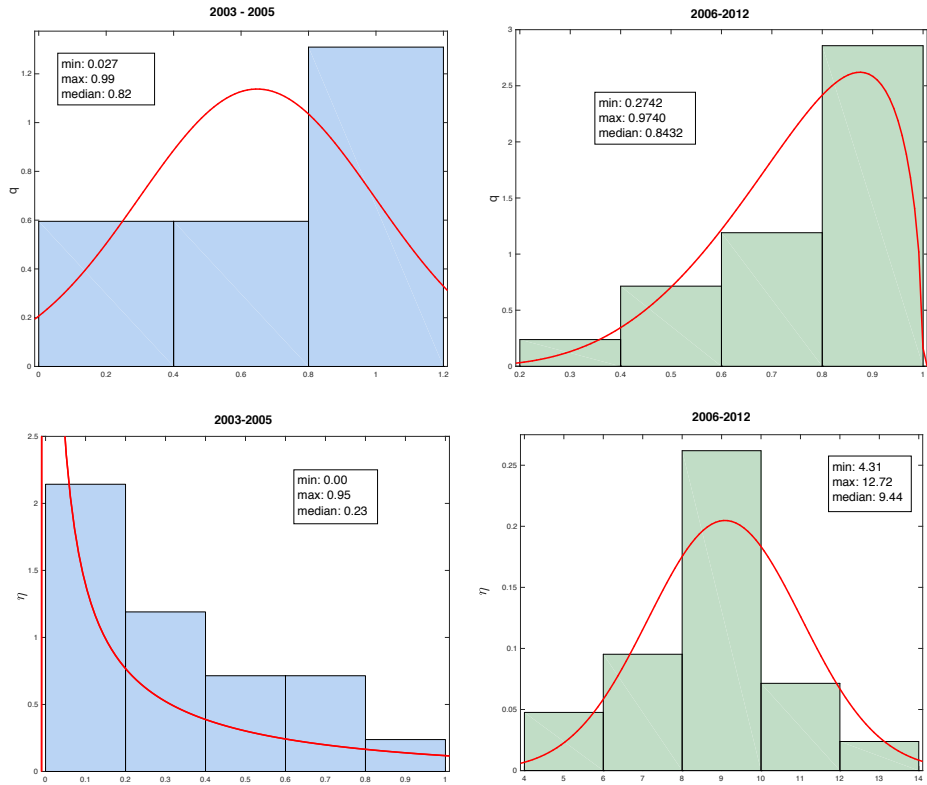

Fig 10. Estimates of q and η For q, 2003-2005: Distribution: Normal, Log likelihood: -7.29243. 2006-2012: Distribution: Beta. Log likelihood: 10.9903. For η, 2003-2005: Distribution: Gamma. Log likelihood: 27.7468. 2006-2012: Distribution: Normal, Log likelihood: -43.3025

under reporting. However, for the second period of study, it seems that it is not relevant if   293
the proportion of cases not reported increases, since population will maintain a stable behav-   294
ior towards self healing, fairly stable around 4% on average. This figure might suggest that   295
individuals tend to increase their trust in the health public system across States which can   296
be used as a measure of the improvement in this public service, relevant for further research   297
and evaluation.   298
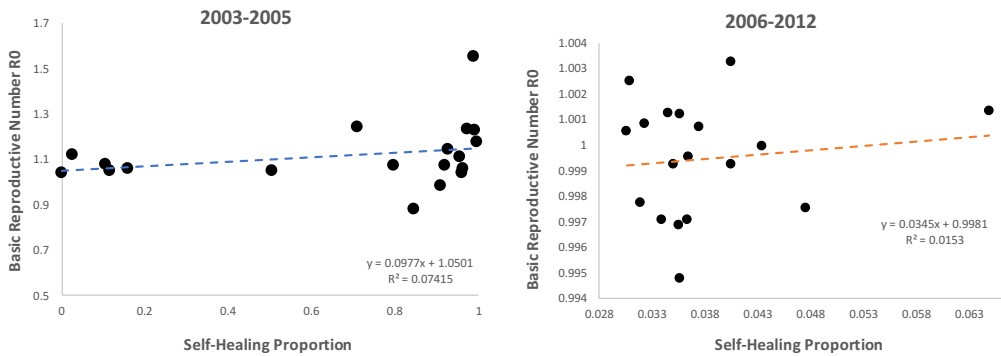
Fig 11. Scatter plot for Self-Healing Proportion (SHP) versus R_0
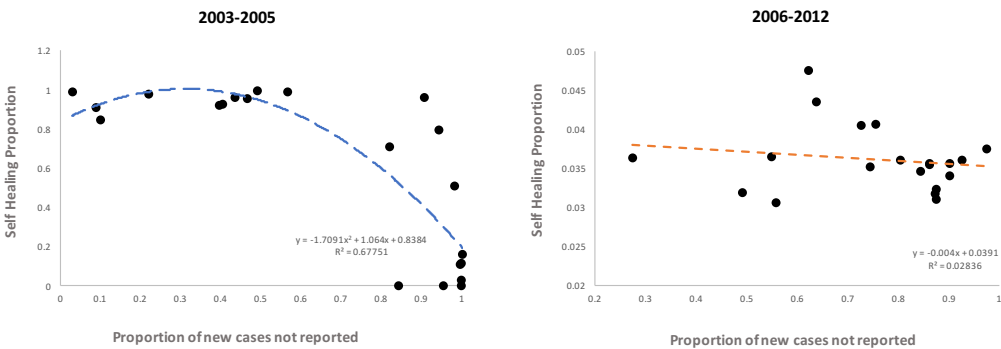

Fig 12. Scatter plot for Self-Healing Proportion (SHP) versus proportion of new cases not reported q

Table 4 presents the average number of reported cases of VL per year per district. Overall, average number or yearly cases increased by 48 %, going from 16,489 cases during 2003-2005 up to 24,467 cases on average during 2006-2012. For comparison of cases and population size, Madhepura, Araria and Saharsa have the largest number of cases per 10,000 inhabitants during the period 2006-2012 with 9.8, 6.4 and 5.9 cases, respectively. Tables 5 and 6 summarize the mean estimates obtained through the model for β , q, and η, including the square root of the SSR.

4   Discussion

Elimination of Visceral Leishmaniasis depends on an adequate knowledge of interactions among factors such as human behavior, the local environment, and the ecology of parasites and vectors. Sociocultural factors such as poverty, educational levels, malnutrition, density and accessibility to interventions in large part determine transmission and persistence of infection in the Bihar [3, 15]. Interventions designed to control or eliminate VL under any of these conditions may fail if their impact is not thoroughly understood. For example, the implementation of interventions based on true quantification of self healing cases can be significantly improve its effectiveness.

Increased advocacy efforts since the early 2000s have led to new control targets set for 2020 by the World Health Organization. However, the efforts still face long lasting success in the face of elimination goals. There are multifaceted challenges in controlling VL in resource-poor local contexts that need to be understood and quantified and the methods used are novel as it is designed for the scarce data from the region. In light of the 2020 targets, this study attempts to develop methods to measure new metrics using historical data from different districts of Indian state of Bihar. The methodology provide some of the novel ways in which social aspects can be captured effectively that can be used for designing sustainable interventions. More emphasis on measurement of non-trivial but significant aspects for VL are needed that can help VL control to broader social determinants of health, given the major social and economic inequalities that continue to underpin transmission in endemic countries such as India.

We developed a dynamic transmission model for anthroponotic Visceral Leishmaniasis in

Table 4. Average number of reported cases per year

District	period 2003-2005	period 2006-2012	Population in million (N)
Araria	1,790	2,187	2.8
Begusarai	446	479	3.0
Darbhanga	462	1,043	3.9
Jahanabad	34	19	1.1
Katihar	875	779	3.1
Khagaria	494	508	1.7
Madhepura	1,954	1,753	2.0
Madhubani	510	653	4.5
Muzaffarpur	1,899	3,144	4.8
Nalanda	96	40	2.9
W. Champaran	414	253	3.9
Patna	328	228	5.8
E. Champaran	1,348	1,704	5.1
Purneas	1,190	1,933	3.3
Saharsa	1,116	2,309	1.9
Samastipur	480	1,374	4.3
Saran	916	1,481	4.0
Sitamarhi	305	956	3.4
Siwan	349	749	3.3
Supaul	430	403	2.2
Vaishali	1,053	2,472	3.5

the Indian state of Bihar. The model incorporates proportion of self healing individuals with the aim to evaluate spatial distribution of elimination chances of VL. We used district level two different data sets to estimate model parameters. The data sets include a monthly incidence from 2003 to 2005 and annual incidence from 2006 to 2012. Owing to the availability of incidence data, the reporting interval for VL differs by origin, and is commonly monthly or annually. The number of cases that are reported can be significantly lower than the actual number of cases, and the level of under-reporting can differ widely over time reporting, even in the same location. Such challenges inherent in the data have led to a number of proposed modeling and estimation methods. An examination of the data shows high uncertainty in the monthly reported cases as compared to annual data, leading to many studies using annual data for estimating parameters. However, use of apparently long-term data may incorporate bias as we expect social and environmental factors to have changed significantly over the time horizon. A better understanding of these factors is important for improving public health policy and aiding public health officials to establish appropriate control strategies. A through investigation on the loss of efficiency in estimation when data are aggregated to form yearly data from monthly or weekly observation, is needed to better conclude model results [20]. Hence, we used both the data sets (monthly and annual) to estimate key model parameters.

In our estimation process we form a nonlinear optimization problem using least square metric, incidence data set and our dynamic model. Robust estimation from the available data sets can be a challenging task. The only disease data available are the clinical case reporting (or incidence) of VL. However, the number of asymptomatic individuals in the population who eventually self heal may have a significant effect on the dynamics of disease spread, and little quantitative data are typically available describing this population. Because most data is passively collected, the number of reported cases is often significantly lower than the true number of cases. In this study, we estimate district-wise self healing rate, proportion of case-underreporting and transmission rate using two types of incidence times series (monthly and annually). Our results show that using monthly data self healing proportion ($\eta/(\eta + \alpha)$) varied from 0.01 to 0.98 with only 7 out of 21 affected districts having self healing proportion less than 0.5.

Estimates summarized in figures 5 and 6 have a very important implication for policy considerations. Estimate of the transmission rate β might be a signal confirming the deterioration of potential elimination of VL due to the overall increase in this rate. During 2003-2005, almost no district in the Bihar state had a β higher than 2, with the exception of Nalanda and Vaishali. However, for 2006-2012, the estimates show that every district

12

has a transmission rate larger than 3. Also, asymptomatic recovery estimates increased   363
considerably for the 2006-2012 period which, indicates that the average time of infection of   364
individuals decrease.   365

Table 5. Mean estimates of 2003-2005

District	β	q	η	SSR-sqrt
Araria	0.778	0.565	0.495	1987
Begusarai	0.78	0.942	0.606	235.3
Darbhanga	1.02	0.905	0.952	185.4
Jahanabad	0.358	0.097	0.349	38.3
Katihar	1.206	0.997	0.04	600.8
Khagaria	0.408	0.086	0.38	345.5
Madhepura	0.616	0.027	0.495	569
Madhubani	1.468	0.998	0.261	361.2
Muzaffarpur	0.108	0.952	0.001	758.6
Nalanda	2.026	0.999	0.616	54
W. Champaran	0.475	0.432	0.433	336.9
Patna	0.145	0.394	0.123	92.4
E. Champaran	0.157	0.819	0.089	625.5
Purneas	1.035	0.994	0.137	555
Saharsa	0.263	0.219	0.206	497.6
Samastipur	0.199	0.463	0.17	247.1
Saran	0.788	0.488	0.665	233
Sitamarhi	0.442	0.982	0.227	218.9
Siwan	0.247	0.402	0.199	287.6
Supaul	0.055	0.841	0.001	276.1
Vaishali	3.08	0.997	0.001	3730.4

Table 6. Mean estimates of 2006-2012

District	β	q	η	SSR-sqrt
Araria	3.228	0.491	9.069	826.5
Begusarai	3.229	0.86	9.836	227.6
Darbhanga	3.236	0.843	7.384	435.7
Jahanabad	3.253	0.974	9.99	9.1
Katihar	3.246	0.874	11.333	389.5
Khagaria	3.236	0.861	9.926	194.8
Madhepura	3.271	0.637	11.694	426
Madhubani	3.251	0.901	9.954	660.9
Muzaffarpur	3.236	0.274	9.348	855.6
Nalanda	3.203	0.804	9.446	22.8
W. Champaran	3.195	0.872	9.598	97.8
Patna	3.222	0.924	9.506	138.5
E. Champaran	3.277	0.621	12.72	857.4
Purneas	3.213	0.754	4.31	443.9
Saharsa	3.229	0.558	7.988	805.4
Samastipur	3.233	0.743	7.776	421.1
Saran	3.26	0.725	11.011	614
Sitamarhi	3.23	0.874	6.983	414.6
Siwan	3.272	0.882	5.989	170.6
Supaul	3.229	0.9	8.836	408.7
Vaishali	3.238	0.548	8.105	742.4

References

1. Manjulika Addy and Ashis Nandy. Ten years of kala-azar in west bengal, part i. did post-kala-azar dermal leishmaniasis initiate the outbreak in 24-parganas? *Bulletin of the World Health Organization*, 70(3):341, 1992.

2. Carlos Castillo-Chavez and Baojun Song. Dynamical models of tuberculosis and their applications. *Mathematical biosciences and engineering*, 1(2):361–404, 2004.

3. Swati DebRoy, Olivia Prosper, Austin Mishoe, and Anuj Mubayi. Challenges in modeling complexity of neglected tropical diseases: a review of dynamics of visceral leishmaniasis in resource limited settings. *Emerging themes in epidemiology*, 14(1):10, 2017.

4. National Institution for Transforming India. State statistics. *http://niti.gov.in/state-statistics*.

5. Registrar General and Census Commissioner. `http://www.censusindia.gov.in/2011Census/pes/Pesreport.pdf`, 2011. Accessed on 2018-10-01.

6. Sarah Jervis, Lloyd AC Chapman, Shweta Dwivedi, Morchan Karthick, Aritra Das, Epke A Le Rutte, Orin Courtenay, Graham F Medley, Indranath Banerjee, Tanmay Mahapatra, et al. Variations in visceral leishmaniasis burden, mortality and the pathway to care within bihar, india. *Parasites & vectors*, 10(1):601, 2017.

7. Mathworks. `http://www.mathworks.com/matlabcentral/fileexchange/4551-inpaint-nans`, Oct 2018. Accessed on 2018-09-01.

8. Dinesh Mondal, Shri Prakash Singh, Narendra Kumar, Anand Joshi, Shyam Sundar, Pradeep Das, Hirve Siddhivinayak, Axel Kroeger, and Marleen Boelaert. Visceral leishmaniasis elimination programme in india, bangladesh, and nepal: reshaping the case finding/case management strategy. *PLoS Neglected Tropical Diseases*, 3(1):e355, 2009.

9. Anuj Mubayi, Carlos Castillo-Chavez, Gerardo Chowell, Christopher Kribs-Zaleta, Niyamat Ali Siddiqui, Narendra Kumar, and Pradeep Das. Transmission dynamics and underreporting of kala-azar in the indian state of bihar. *Journal of theoretical biology*, 262(1):177–185, 2010.

10. World Health Organization et al. Regional strategic framework for elimination of kala-azar from the south-east asia region (2005–2015). *New Delhi: WHO Regional Office for South-East Asia*, 2005.

11. World Health Organization et al. Visceral leishmaniasis rapid diagnostic test performance diagnostics evaluation series no. 4. 2011.

12. Abhishek Pandey, Anuj Mubayi, and Jan Medlock. Comparing vector–host and sir models for dengue transmission. *Mathematical biosciences*, 246(2):252–259, 2013.

13. Richard D Pearson and Anastacio de Queiroz Sousa. Clinical spectrum of leishmaniasis. *Clinical infectious diseases*, pages 1–11, 1996.

14. National Vector Borne Diseases Control Programme. `http://nvbdcp.gov.in/Doc/Annual-report-NVBDCP-2012.pdf`, 2009. Accessed on 2018-09-01.

15. Darren Sheets, Anuj Mubayi, and Hristo V Kojouharov. Impact of socio-economic conditions on the incidence of visceral leishmaniasis in bihar, india. *International journal of environmental health research*, 20(6):415–430, 2010.

16. Om Prakash Singh, Bhawana Singh, Jaya Chakravarty, and Shyam Sundar. Current challenges in treatment options for visceral leishmaniasis in india: a public health perspective. *Infectious diseases of poverty*, 5(1):19, 2016.

17. A Sud, N Varma, RK Marwaha, FM Patel, A Trehan, S Singh, and S Varma. Visceral leishmaniasis in a non-endemic area of india. *Tropical doctor*, 34(4):247–249, 2004.

18. Pauline Van den Driessche and James Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical biosciences*, 180(1):29–48, 2002.

19. Neelam Varma and Shano Naseem. Hematologic changes in visceral leishmaniasis/kala azar. *Indian Journal of Hematology and Blood Transfusion*, 26(3):78–82, 2010.

20. Christopher S Withers and Saralees Nadarajah. Estimating trend from seasonal data: is daily, monthly or annual data best? *Environmetrics*, 26(7):488–501, 2015.

5   Appendix

5.1   Model

The System of the model is given as:

$$\frac{dS}{dt} = \Lambda - \beta\left(\frac{I+A}{N}\right)S - \mu S \tag{3}$$

$$\frac{dA}{dt} = \beta\left(\frac{I+A}{N}\right)S - (\alpha + \eta + \mu)A \tag{4}$$

$$\frac{dI}{dt} = \alpha A - (\gamma + \delta + \mu)I \tag{5}$$

$$\frac{dR}{dt} = \eta A + \gamma I - \mu R \tag{6}$$

$$\frac{dC}{dt} = (1-q)\alpha A \tag{7}$$

5.2   Mathematical computation

Using the Next Generation Approach [18], we determine that the basic reproduction number is $R_0 = \dfrac{\beta}{\alpha + \eta + \mu} + \dfrac{\alpha\beta}{(\alpha+\eta+\mu)(\delta+\gamma+\mu)}$. The disease-free equilibrium and endemic equilibrium are $\mathcal{E}_0 = (S_0 = \Lambda/\mu, A_0 = 0, I_0 = 0, R_0 = 0)$ and $\mathcal{E}_1 = (S_1, A_1, I_1, R_1)$, respectively, where $S_1, A_1, I_1,$ and R_1 are defined below.

$$S_1 = \frac{\Lambda(\delta(\eta+\mu) + (\alpha+\eta+\mu)(\gamma+\mu))}{\mu(\alpha(\beta-\delta) + \beta(\delta+\gamma+\mu))} \tag{8}$$

$$A_1 = \frac{\Lambda(-(\beta-\eta-\mu)(\delta+\gamma+\mu) + \alpha(-\beta+\delta+\gamma+\mu))}{(\alpha+\eta+\mu)(\alpha(\beta-\delta)+\beta(\delta+\gamma+\mu))} \tag{9}$$

$$I_1 = \frac{\alpha\Lambda(-(\beta-\eta-\mu)(\delta+\gamma+\mu) + \alpha(-\beta+\delta+\gamma+\mu))}{(\alpha+\eta+\mu)(\delta+\gamma+\mu)(\alpha(-\beta+\delta)-\beta(\delta+\gamma+\mu))} \tag{10}$$

$$R_1 = -\frac{(\Lambda(\delta\eta+\alpha\gamma+\eta(\gamma+\mu))(-(\beta-\eta-\mu)(\delta+\gamma+\mu) + \alpha(-\beta+\delta+\gamma+\mu))}{\mu(\alpha+\eta+\mu)(\delta+\gamma+\mu)(\alpha(\beta-\delta)+\beta(\delta+\gamma+\mu)))} \tag{11}$$

Following the approach in [2], we find that whether or not a backward bifurcation occurs at $R_0 = 1$ depends on the sign of a, where

$$a = -\frac{4\alpha\beta(\alpha+\eta+\mu)(\beta(\eta-\gamma) + (\alpha+\eta+\mu)(\gamma+\mu))}{\Lambda(\alpha-\beta+\eta+\mu)^3}.$$

That is, if $a < 0$, then the system exhibits a forward bifurcation at $R_0 = 1$ and therefore, the disease-free equilibrium is the only equilibrium when $R_0 < 1$. However, if $a > 0$, then there is a locally asymptotically stable sub threshold endemic equilibrium (i.e. a backward bifurcation); that is, VL may persist even when $R_0 < 1$. To ascertain whether VL can be eliminated from a district, we will calculate R_0 and a for each district. If $R_0 > 1$, the disease is endemic in that district. If $R_0 < 1$ *and* $a < 0$, then VL is moving towards elimination in that district. On the other hand, if $R_0 < 1$ and $a > 0$, we do not have enough information to conclude whether or not VL is moving towards elimination in that district.

List of Figures