Estimation of under-reported visceral Leishmaniasis (VL) cases in Bihar: a Bayesian approach

Ranjan A, Bhatnagar T

1Department of Community and Family Medicine, All India Institute of Medical Sciences, Patna, 2National Institute of Epidemiology, Indian Council of Medical Research, Chennai

Abstract

Background: Visceral leishmaniasis (VL) is a major health problem in the state of Bihar and adjoining areas in India. In absence of any active surveillance mechanism for the disease, there seems to be gross under-reporting of VL cases. Objective: The objective of this study was to estimate extent of under-reporting of VL cases in Bihar using pooled analysis of published papers. Method: We calculated the pooled common ratio (RRMH) based on three studies and combined it with a prior distribution of ratio using inverse-variance weighting method. Bayesian method was used to estimate the posterior distribution of the “under-reporting factor” (ratio of unreported to reported cases). Results: The posterior distribution of ratio of unreported to reported cases yielded a mean of 3.558, with 95% posterior limits of 2.81 and 4.50. Conclusion: Bayesian approach gives evidence to the fact that the total number of VL cases in the state may be nearly more than three times that of currently reported figures.

Key Words

Visceral Leishmaniasis (VL); Bayesian; Under-Reporting

Introduction

Visceral Leishmanias (VL), known as Kala-azar in the Indian Subcontinent, is one of the major health problems for the state of Bihar, and adjoining states like West Bengal, Jharkhand and Uttar Pradesh for many decades (1). Focal and sporadic cases of VL have regularly been occurring in many districts of Bihar since 1977 (2). At present, 31 out of 38 districts in Bihar are endemic at different levels and nearly 67.5 million populations are at the risk of disease.

On an average, more than 90% of VL cases in India are reported from Bihar alone (1, 3, 4).

There is no active surveillance mechanism for estimating the incidence of VL in affected areas. Based on the passive case detection at public health facilities such as Primary Health Centres (PHCs), and the district and medical college hospitals, the state health department has been compiling annual incidence of VL since 1977. Majority of VL cases in the endemic areas seek diagnosis and treatment offered by private health facilities such as private clinics.
and NGOs. These cases are not taken into account for compilation of annual incidence of VL in the state leading to gross under reporting of VL cases. It has been reported earlier that the total number of estimated cases would be 2 to 2.5 times more than the actual incidence and may be even 5 times higher than the officially reported figures (1, 5).

Few studies have been conducted to quantitatively estimate the extent of under-reporting of VL cases in different endemic areas of kala-azar (6, 7, 8). The extent of under-reporting of VL cases varies from 1.31 to 7.1 times the officially reported cases over different time periods in high endemic areas like Muzaffapur and Vaishali districts. The extent of under-reporting could vary from one endemic area to another because of various known and unknown factors which could influence health seeking behaviour of individuals in each community. These influencing factors can severely affect estimation of parameter of interest i.e. ratio of unreported to reported VL cases (9, 10). It is understood that ln (RR) follows a log-normal distribution (13).

On the basis of various discussions held with the program managers and experts of VL control in India, it was expected that the extent of under-reporting of VL at the public health facilities could vary in different endemic areas depending upon various factors influencing the health seeking behaviours. We made a priori by placing 95% probability on RRUR for the range of priors in order to get the best prior in terms of amount of information and a reasonable average risk ratio. We performed sensitivity analysis taking several sets of prior distributions taking the upper and lower bound of ratio of unreported to reported VL cases. Among these prior distributions, the best prior distribution having the minimum variance and reasonable average risk ratio was selected for pooling with the common summary estimate based on three studies for getting posterior distribution of parameter of interest using inverse-variance technique.

Our objective was to estimate the extent of under-reporting of VL cases using Bayesian method.

**Methods**

The results of three previous studies carried out in different endemic areas of VL in Bihar were used (Table 1) (6, 7, 8). We used Woolf’s method and Mantel-Haenszel method for estimating the “under-reporting factor (UR)” as a common ratio of unreported to reported VL cases (9, 10). It was preferred to use Mantel-Haenszel method because of the assumption of homogeneity of stratum-specific parameters, more common and valid for small data (11, 12). We considered the “under-reporting factor” as a risk ratio (RRUR) measuring the ratio of not reported to the reported among VL cases in the same population. It is understood that ln (RR) follows a log-normal distribution (13).

Statistical techniques are available to get a common estimate of parameter of interest by pooling the results of studies estimating similar parameter of interest. In order to account for uncertainty due to known and unknown factors, Bayesian method was used taking a prior distribution of parameter of interest and the pooled estimate based on three published studies for estimating a posterior distribution of parameter of interest using inverse-variance technique.
Out of several set of priors in the range of 10 to 1/10, the best priors that would follow from a non-normal prior for the ln(RR_{UR}) and satisfied the following equations as given below:

=> \exp (\text{prior mean} - 1.96 \times \text{prior SD}) = 3

………………………………………...[Eq. 1]

=> \exp (\text{prior mean} + 1.96 \times \text{prior SD}) = 9

………………………………………...[Eq. 2]

Prior standard deviation of ln(RR_{UR}) = (width of interval in ln(RR) units) / (width of interval in SD units)……………………………………………[Eq. 3]

Inverse of variance provides information obtained from the data and the prior assuming both are adequately approximated by normal distributions (12). Assuming a normal distribution of the “under-reporting factor” we ignored the distinction between mean, median and mode.

We applied inverse-variance weighted method to approximate the Bayesian posterior distribution for the “under-reporting factor (UR)” (14, 15, 16). We calculated the posterior mean for ln(RR_{UR}) as the average of the prior mean ln(RR_{UR}) and the data estimate, using the information weights.

Posterior mean for ln(RR_{UR}) = expected ln(RR_{UR}) given data

\[
{(\text{prior mean ln(RR_{UR})}/ \text{prior variance of ln(RR_{UR})}) + (\text{ln(RR_{MH})}/\text{variance of ln(RR_{MH}))}}
\]

= \frac{1}{(1/ \text{prior variance of ln(RR_{UR})}+(1/ \text{variance of ln(RR_{MH}))}}

The compatibility of the data and prior were also checked using an analogue of the frequentist two-stratum homogeneity statistic (13, 16, 17).

**Bayesian Analysis:** According to the available data, the “under-reporting factor” (RR_{MH}) i.e., the average Mantel-Hanszel risk ratio of unreported to reported cases was 3.27, and the variance of ln(RR_{MH}) was 0.01761 with 95% lower and upper limits of 2.52 and 4.24, respectively.

Out of the several sets of priors, we obtained the best prior by placing 95% probabilities on the prior between 3 and 9.

Solving equations [Eq. 1] and [Eq. 2],

prior mean of ln(RR_{UR}) = average of the limits = \[ \text{ln}(9) + \text{ln}(3)/2 = 1.647 \]

prior RR_{UR} = \exp[1.647] = 5.19

Prior standard deviation of ln(RR_{UR}) =

Solving equation [Eq. 3],

\[
[\text{ln}(9) –\text{ln}(3)]/2 \times 1.96 = 0.2802
\]

Prior variance of ln(RR_{UR}) = (0.2802)^2 = 0.0785.

Thus, the normal prior distribution of ln(RR_{UR}) that would produce the stated bet has mean 1.647 and variance 0.0785.

Combining this prior with the pooled Mantel Hanszel estimate of RR_{MH} by information weighting yields the posterior mean for ln(RR_{UR}) = 1.269 and posterior variance of ln(RR_{UR}) = 0.01439. Thus, the posterior distribution of RR_{UR} had mean of 3.558, with 95% posterior limits of 2.81 and 4.50. The posterior distribution of RR_{UR} had lower variance as compared to both the pooled i.e. data and prior. Also, the amount of information of the pooled data was nearly four times more than the prior as the weight of pooled data was nearly four times of the prior information.

The homogeneity statistic between the data and the prior was 1.1497 with a corresponding p-value of 0.1271.
Discussion

Using Bayesian analysis, we estimated the extent of under-reporting that was nearly three and half times of the currently reported official statistics. Previous studies had shown a wide range of figures for under-reporting of cases making it difficult to generalize the parameter of interest. There were many factors which could influence the extent of under-reporting in an endemic area. VL cases usually occur in clusters in the endemic areas. Within one high endemic district, some PHCs/blocks have higher incidence as compared to other PHCs/blocks. Also, under-reporting of VL cases could be an outcome of synergistic effects of many factors such as socio-cultural factors, economic status, availability and utilization of health facilities, attitude towards government health facilities, awareness etc. These factors may vary from one endemic area to another and hence can lead to severe uncertainty in the estimation of under-reporting of VL cases.

Bayesian analysis assigns a subject-matter-based prior distribution to the parameter of interest, and combines this prior distribution with data probabilities to obtain a posterior distribution for parameter of interest (13). We performed Bayesian analysis by using information weighting of prior estimates with the estimates obtained from the meta-analysis of three previous studies after pooling the results. By taking a broad range of ratio of unreported to reported cases as prior distribution, we accounted for uncertainty due to various factors mentioned above. Thus, the prior and the data appeared compatible. The amount of information provided by the data (1/0.01761 = 56.76) was nearly more than four times the amount of information provided by the prior (1/0.0785 = 12.73), showing that the data was given more emphasis than the prior in estimating the posterior.

Conclusion: Bayesian approach used in this paper shows that the results of various studies can be used to get a common or more generalized estimate of parameter of interest, and uncertainty due to various factors such as confounding, selection biases, and other unknown factors in estimation of parameter can be accounted for. This approach is computationally simple and can be implemented in practice for estimating parameter of interest and at the same time gauging the uncertainty factors, usually not discernable in the observational studies due to confounding, selection bias or misclassifications.

References

7. Singh SP, Reddy DCS, Rai M, Sundar S. Serious underreporting of visceral leishmaniasis through passive case reporting in Bihar, India. Tropical Medicine and International Health 2006;11(6):899-905.


### Tables

**TABLE 1: ESTIMATED RATIO OF UNREPORTED TO REPORTED CASES OF VL IN BIHAR**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year of Study</th>
<th>Study Population</th>
<th>Study Area</th>
<th>Total VL cases</th>
<th>Risk Ratio (UR/R) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bora D, 1994</td>
<td>1991</td>
<td>9172</td>
<td>Vaishali District</td>
<td>74</td>
<td>1.31 (0.83 – 2.07)</td>
</tr>
<tr>
<td>Singh SP et.al,2006</td>
<td>2003</td>
<td>27566</td>
<td>Muzaffarpur District</td>
<td>65</td>
<td>7.13 (3.40 – 14.93)</td>
</tr>
<tr>
<td>Singh VP et.al,2010</td>
<td>2007</td>
<td>31324</td>
<td>Vaishali District</td>
<td>177</td>
<td>4.17 (3.75 – 4.63)</td>
</tr>
</tbody>
</table>