Text analysis

An epidemiological case study in Malaria Deaths in India

Wilson Suraweera
Centre for Global Health Research
Li Ka Shing Knowledge Institute
St Michael’s Hospital
Toronto
Adult and child malaria mortality in India: a nationally representative mortality survey

Neeraj Dhingra MD MBBS, Prof Prabhat Jha DPhil MBBS, Vinita P Sharma DPhil, Alan A Cohen PhD, Ravi M Jotkar MD, Peter S Rodriguez MSc, Diego G Bassani PhD, Wilson Suraweera MSc, Ramanathan Lawminarayan PhD, and Prof Richard Peeto FRS, for the Million Death Study Collaborator*

Summary

Background
National malaria death rates are difficult to assess because reliably diagnosed malaria is likely to be cured, and deaths in the community from undiagnosed malaria could be misattributed in retrospective enquiries to other febrile causes of death, or vice-versa. We aimed to estimate plausible ranges of malaria mortality in India, the most populous country where the disease remains common.

Methods
Full-time non-medical field workers interviewed families or other respondents about each of 122 000 deaths during 2001–03 in 6671 randomly selected areas of India, obtaining a half-page narrative plus answers to specific questions about the severity and course of any fevers. Each field report was sent to two of 120 trained physicians, who independently coded underlying causes, with discrepancies resolved either via anonymous reconciliation or adjudication.

Findings
Million Death Study (MDS) 2001-2014

Nationally representative sample
(Sample Registration System in India)

- 6,671 of these small areas randomly chosen from all parts of India (each with about 1000 people per area)

Probable causes of mortality - Verbal Autopsy (VA) method
Malaria deaths occurred where the most dangerous type (*Plasmodium falciparum*) of malaria parasite occurs.

**Study Deaths**

Estimated 200,000 Deaths

**Malaria program deaths**

Estimated 15,000 Deaths
Verbal Autopsy Method

RHIME method for physician coding

RHIME field forms data entered / scanned

Key data extracted and exact copy of narrative

Field report assigned to two trained physician coders

Physician 1

Physician 2

Reconciliation

Agree? Yes/No

3rd Physician Adjudication

- Use standardized guidelines
- Assign ICD-10 code & key words

Source: Jha et al, 2006
Rationale

- South Asia abounds in life threatening undifferentiated febrile illnesses like malaria. Typhoid, typhus (rickettsial illness), leptospirosis, and dengue are some of the common diseases which mimic malaria.

- WHO estimates Malaria causes only about 15,000 deaths
- Our study estimates 205,000 deaths, CI (195,000, 215,000)

? Whether the overall proportion of deaths ascribed to malaria was at least approximately correct.

? Whether physicians decided malaria only considering fever symptom in known Malaria hi-prevalence areas
Key symptoms

Malaria (ICD10: B50-B54)
- Fever
- Headaches
- Chills
- Diarrhoea
- Muscular pain
- Lethargy
- Sickness
- Coughing fits
- Abdominal pains

Key words: written narratives
### Functions that Search for Strings/Characters
- ANYALNUM
- ANYALPHA
- ANYDIGIT
- ANYPUNCT
- ANYSPACE
- NOTALNUM
- NOTALPHA
- NOTDIGIT
- NOTUPPER
- FIND
- FINDC
- INDEX
- INDEXC
- INDEXW
- VERIFY
- SOUNDEX

### Functions that Divide Strings into "Words"
- SCAN
- SCANQ
- CALL SCAN
- CALL SCANQ

### Functions that Substitute Letters or Words in Strings
- TRANSLATE
- TRANWRD

### Functions that Compute the Length of Strings
- LENGTH
- COUNT
- LENGTHC
- COUNTC
- LENGTHM
- LENGTHN
## Method

<table>
<thead>
<tr>
<th>Keyword</th>
<th>INDEXW(Kword, &quot;FEVER&quot;)</th>
<th>INDEX(up_KWord, &quot;FEVER&quot;)</th>
<th>SOUNDEX(up_KWord, &quot;FEVER&quot;)</th>
<th>Malaria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fever</td>
<td>p(1,1,0)</td>
<td>p(1,2,0)</td>
<td>p(1,3,0)</td>
<td>P(1,0)</td>
</tr>
<tr>
<td>intermit*</td>
<td>p(1,1)</td>
<td></td>
<td></td>
<td>P(1,1)</td>
</tr>
<tr>
<td>alternat*</td>
<td>p(1,2)</td>
<td></td>
<td></td>
<td>P(1,2)</td>
</tr>
<tr>
<td>chills</td>
<td>p(1,3)</td>
<td></td>
<td></td>
<td>P(1,3)</td>
</tr>
<tr>
<td>shiver*</td>
<td>p(1,4)</td>
<td></td>
<td></td>
<td>P(1,4)</td>
</tr>
<tr>
<td>rigour</td>
<td>p(1,5)</td>
<td></td>
<td></td>
<td>P(1,5)</td>
</tr>
<tr>
<td>alternat*</td>
<td>p(1,6)</td>
<td></td>
<td></td>
<td>P(1,6)</td>
</tr>
<tr>
<td>febrile</td>
<td>p(1,7)</td>
<td></td>
<td></td>
<td>P(1,7)</td>
</tr>
<tr>
<td>repeat*</td>
<td>p(1,8)</td>
<td></td>
<td></td>
<td>P(1,8)</td>
</tr>
<tr>
<td>fluctuat*</td>
<td>p(1,9)</td>
<td></td>
<td></td>
<td>P(1,9)</td>
</tr>
<tr>
<td>high</td>
<td>p(1,10)</td>
<td></td>
<td></td>
<td>P(1,10)</td>
</tr>
<tr>
<td>fluctuat*</td>
<td>p(1,11)</td>
<td></td>
<td></td>
<td>P(1,11)</td>
</tr>
<tr>
<td>on &amp; off</td>
<td>p(1,12)</td>
<td></td>
<td></td>
<td>P(1,12)</td>
</tr>
<tr>
<td>quartan</td>
<td>p(1,13)</td>
<td></td>
<td></td>
<td>P(1,13)</td>
</tr>
<tr>
<td><strong>Headaches</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Diarrhoea</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Muscular pain</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lethargy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sickness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Coughing fits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Abdominal pains</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**INDEXW** function - searches for strings that are words - <works only either side delimiters are blank >

**INDEX** function - searches for patterns as separate words or as parts of words - <works without concerning delimiters >

**SOUNDEX** function - searches phonetic equivalent of a text string. Often used to match some minor spelling differences

@macro kw_search(KW=, Var_KW_out=, Var_KW_in=, Data_IN=, Data_OUT=)

/* ------------------------------ */

Data_IN='Dataset name', Data_OUT='Output dataset name', KW='Keyword' string wants to search
Var_KW_in='Keyword variable name where keyword strings to be searched in data', Var_KW_out='Out put variable name'
/* ------------------------------ */

data &Data_OUT;
  set &Data_IN ;
  up_KWord=upcase(substr(&Var_KW_in,1,255)); /* Caps the Keyword string and capture the string from where you want. Change params */
  Indx_W = indexw(up_KWord, KW); /* Complete word */
  Indx = index(up_KWord, KW); /* Complete word OR word in a part of other word*/
  SOUND_KW1 = SOUNDEX(KW); /* eq. KW="FEVER" */
  S_ind=0;
  array word (50) &word1-word50;
  array S (50) $ &S1-S50; /* temporary indicator variable to keep KW availability */
  do i=1 to 50;
    word(i)=scan( up_KWord, 1, ' ' ); /* Explode the string into words */
    SOUND_KW2 = SOUNDEX(word(i));
    if SOUND_KW1 = SOUND_KW2 then S(i)=1;
    else S(i)=0;
    S_ind=S_ind+S(i); /* Keyword appearance indicator*/
  end;
  drop i up_KWord SOUND_KW1 SOUND_KW2 &word1-word50 &S1-S50 ; /* Drop temporary variables*/
/* Output variable */
  if (Indx_W>0) then &Var_KW_out=1; /* Most probable */
  else if (Indx>0 and S_ind>0) then &Var_KW_out=2; /* probably */
  else if (Indx>0 and S_ind=0) then &Var_KW_out=3; /* Least probable */
  else if (Indx=0 and S_ind>0) then &Var_KW_out=4; /* Least probable*/
  else &Var_KW_out=0;

%mend;
/* ------------------------------ */
/* Examples: */
%kw_search(KW='HIGH', Var_KW_out=HIGH, Var_KW_in=KW, Data_IN=text.d, Data_OUT=HIGH); run;
%kw_search(KW='FEVER', Var_KW_out=FEVER, Var_KW_in=KW, Data_IN=text.d, Data_OUT=FEVER); run;
%kw_search(KW='CHILL', Var_KW_out=CHILL, Var_KW_in=KW, Data_IN=text.d, Data_OUT=CHILL); run;
/* Data_IN='Input dataset name', Data_OUT='Output dataset name', Var_KW='Keyword variable name',
KW='Keyword' string wants to search, Var_KW_out='Out put variable name' */
### Output SAS Data Set

<table>
<thead>
<tr>
<th>UniqNo</th>
<th>codex</th>
<th>Final_code</th>
<th>KWWord</th>
<th>Ind_N</th>
<th>Ind_X</th>
<th>S_ind</th>
<th>HIGH</th>
<th>FEVER</th>
<th>CHILL</th>
</tr>
</thead>
<tbody>
<tr>
<td>26572</td>
<td>12040104</td>
<td>1P01</td>
<td>R50</td>
<td>high fever with cough, fever rose everyday, repeated attacks of chills, sweating, headache, very thirsty.&lt;RKW1&gt; high fever with cough, fever rose everyday, repeated attacks of chills, sweating, headache, very thirsty.&lt;RKW2&gt; fever.&lt;KW1&gt; fever.</td>
<td>0</td>
<td>80</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>26573</td>
<td>12040160</td>
<td>1P01</td>
<td>R50</td>
<td>High fever with loss of consciousness, no chills.&lt;KW2&gt; High fever, headache, body pain, unconscious, weak.&lt;RKW1&gt; High fever with loss of consciousness, no chills.&lt;RKW2&gt; High fever, headache, body pain, unconscious, weak.&lt;AKW&gt; FEVER.</td>
<td>0</td>
<td>50</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>26574</td>
<td>12040296</td>
<td>1P01</td>
<td>R50</td>
<td>FEVER NOT RESPONDING.&lt;KW2&gt; H/O HIGH FEVER WITH CHILLS &amp; RIGOR, SHIVERING, NAUSEA, PAIN HEADACHE, FAST BREATHING, MOUTH INFECTION, H/O SWEATING.&lt;RKW1&gt; &lt;RKW2&gt; &lt;AKW&gt;</td>
<td>0</td>
<td>55</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>26575</td>
<td>12039721</td>
<td>1P01</td>
<td>R50</td>
<td>H/O LONG DURATION FEVER NOT RESPONDING TO RX.&lt;KW2&gt; High fever with chills coming &amp; going with sweat, feeling cold with body pain.&lt;RKW1&gt; High fever with chills coming &amp; going with sweat, feeling cold with body pain.&lt;RKW2&gt; H/O LONG DURATION FEVER NOT</td>
<td>0</td>
<td>75</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>26576</td>
<td>10053248</td>
<td>1P01</td>
<td>R50</td>
<td>&lt;KW1&gt; &lt;KW2&gt; High fever with chills &amp; rigor, head ache, vomiting.&lt;RKW1&gt; High fever with chills &amp; rigor, head ache, vomiting.&lt;RKW2&gt; &lt;AKW&gt; High fever with chills &amp; rigor, head ache, vomiting, not improved by treatment, death, F-COD-death.</td>
<td>0</td>
<td>32</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>26577</td>
<td>2009493</td>
<td>1P01</td>
<td>R50</td>
<td>FEVER, COUGH.&lt;KW2&gt; fever with chills, headache, death.&lt;RKW1&gt; FEVER, COUGH.&lt;KW2&gt; fever with chills, headache, death.&lt;AKW&gt; INTERMITTENT HIGH FEVER FOR 1 MONTH</td>
<td>0</td>
<td>38</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>26578</td>
<td>22074515</td>
<td>1P01</td>
<td>R50</td>
<td>&lt;KW1&gt; &lt;KW2&gt; Fever high grade with chills &amp; rigor - 21d, &lt;KW2&gt; &lt;KW2&gt; &lt;KW2&gt; &lt;KW2&gt; high grade fever with chills &amp; rigor, sweating, cough, &lt;KW2&gt; &lt;KW2&gt; &lt;KW2&gt; high grade fever with chills &amp; rigor, sweating, cough, &lt;KW2&gt; &lt;KW2&gt; high grade fever with chills &amp; rigor, sweating, cough, &lt;KW2&gt; &lt;KW2&gt; high grade fever with chills &amp; rigor, sweating, cough, &lt;KW2&gt; &lt;KW2&gt; high grade fever with chills &amp; rigor, sweating, cough, &lt;KW2&gt; &lt;KW2&gt; high grade fever with chills &amp; rigor, sweating, cough,</td>
<td>0</td>
<td>30</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>26579</td>
<td>23073243</td>
<td>1P01</td>
<td>R50</td>
<td>HIGH FEVER FOR MANY DAYS, NOT RESPONDING TO TREATMENT.&lt;KW2&gt; fever, chills, high grade, off &amp; on for 10 days, treated, death.&lt;RKW1&gt; fever, chills, high grade, off &amp; on for 10 days, treated, death.&lt;RKW2&gt; HIGH FEVER FOR MANY DAYS, NOT RESPONDING TO TR</td>
<td>0</td>
<td>75</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*KW1* HIGH FEVER WITH CHILLS & RIGORS
Table: percentages attributed by physician coders to various types of infection, by area among rural fever deaths at ages 0-69 years that did not occur in a healthcare facility
END
References

Different types of Malaria

The 4 Plasmodium parasites responsible for causing malaria are called:

1. **Plasmodium falciparum** - Responsible for causing malignant Malaria, the most severe strains and the most deaths.

**benign Malaria**

2. **Plasmodium vivax** - Responsible for causing benign Malaria which is less severe, however it can stay dormant in your liver for years and lead to relapses of the Malaria disease.

3. **Plasmodium ovale** - This parasite is responsible for causing benign Malaria and can remain in the body for years without producing the symptoms of Malaria.

4. **Plasmodium malariae** - This is the rarest of the malaria parasites but when it is contracted it can cause benign Malariae