

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

Case study : THE LANCET – 20 November 2010

Adult and child malaria mortality in India: a nationally representative mortality survey : The - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites

Address [http://www.thelancet.com/journals/lanct/article/PIIS0140-6736\(10\)60831-8/fulltext#article_upsell](http://www.thelancet.com/journals/lanct/article/PIIS0140-6736(10)60831-8/fulltext#article_upsell) Go Links

THE LANCET

RSS Feeds

Subscribe | Register | Login

Search for in All Fields GO Advanced Search

Home | Journals | Collections | Audio | Conferences | Education | The Lancet Series | Information for | Careers

The Lancet, [Volume 376, Issue 9754](#), Pages 1768 - 1774, 20 November 2010
doi:10.1016/S0140-6736(10)60831-8 [Cite or Link Using DOI](#)
Published Online: 21 October 2010

< [Previous Article](#) | [Next Article](#) > [Get Full Text Elsewhere](#)

Adult and child malaria mortality in India: a nationally representative mortality survey

[Neeraj Dhingra](#) MD [Prabhat Jha](#) DPhil [Vinod P Sharma](#) DPhil [Alan A Cohen](#) PhD [Raju M Jotkar](#) MD [Peter S Rodriguez](#) MSA [Diego G Bassani](#) PhD [Wilson Suraweera](#) MSc [Ramanan Laxminarayan](#) PhD [Prof Richard Peto](#) FRS, for the Million Death Study Collaborators‡

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 130 trained physicians, who independently coded underlying causes, with discrepancies resolved either via anonymous reconciliation or adjudication.

Findings

Article Options

- Summary
- Full Text**
- PDF (4,597 KB)
- Cited by in Scopus (3)
- Printer Friendly Version
- Request permission
- Export Citation
- Create Citation Alert

Linked Articles

- [Comment India's invisible malaria burden](#)

Other Articles of Interest

- [Series Ranking of elimination feasibility between malaria-endemic countries](#)
- [Review Measuring malaria endemicity from intense to interrupted transmission](#)

start | Inbox - Suraweer... | 2011 | Malaria | Text analysis - an... | Malaria updated... | Adult and child m... | 2:54 PM



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)**

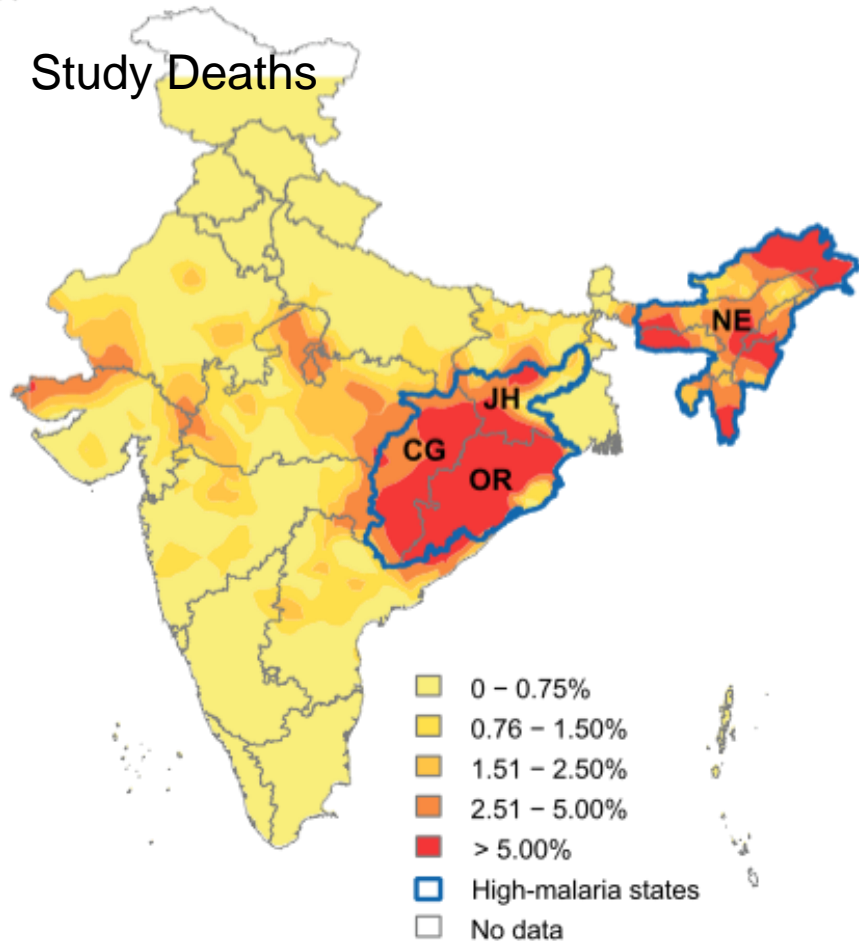
Probable causes of mortality - Verbal Autopsy (VA) method



Malaria deaths occurred where the most dangerous type (*Plasmodium falciparum*) of malaria parasite occurs

A

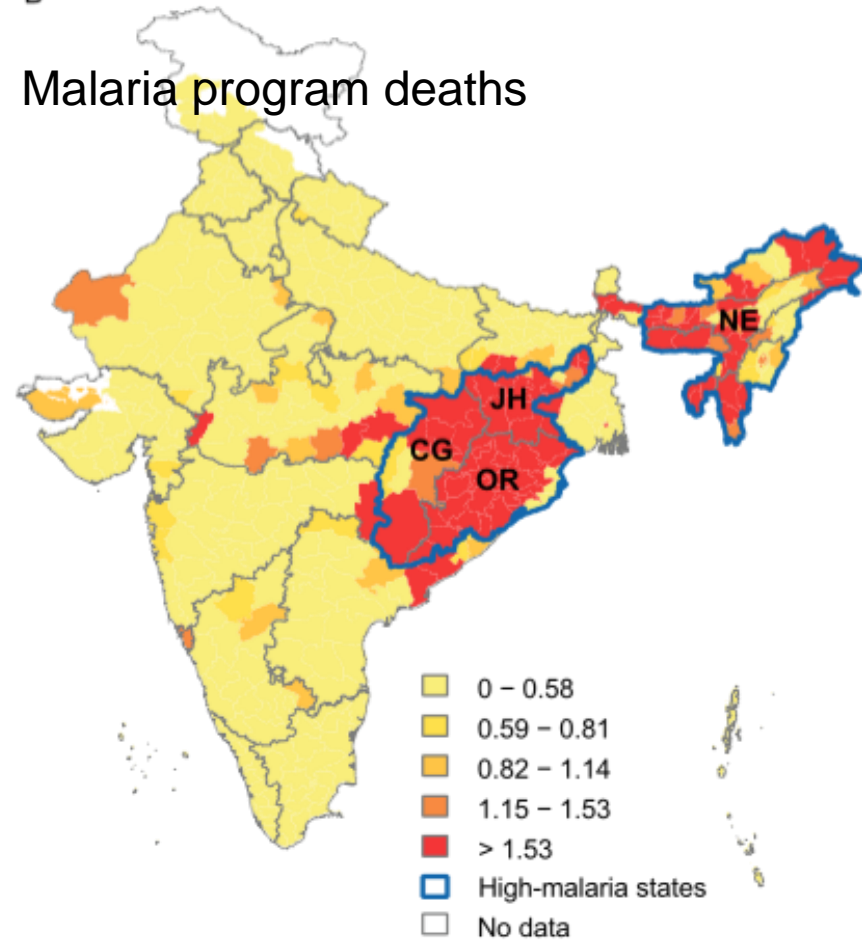
Study Deaths



Estimated 200,000 Deaths

B

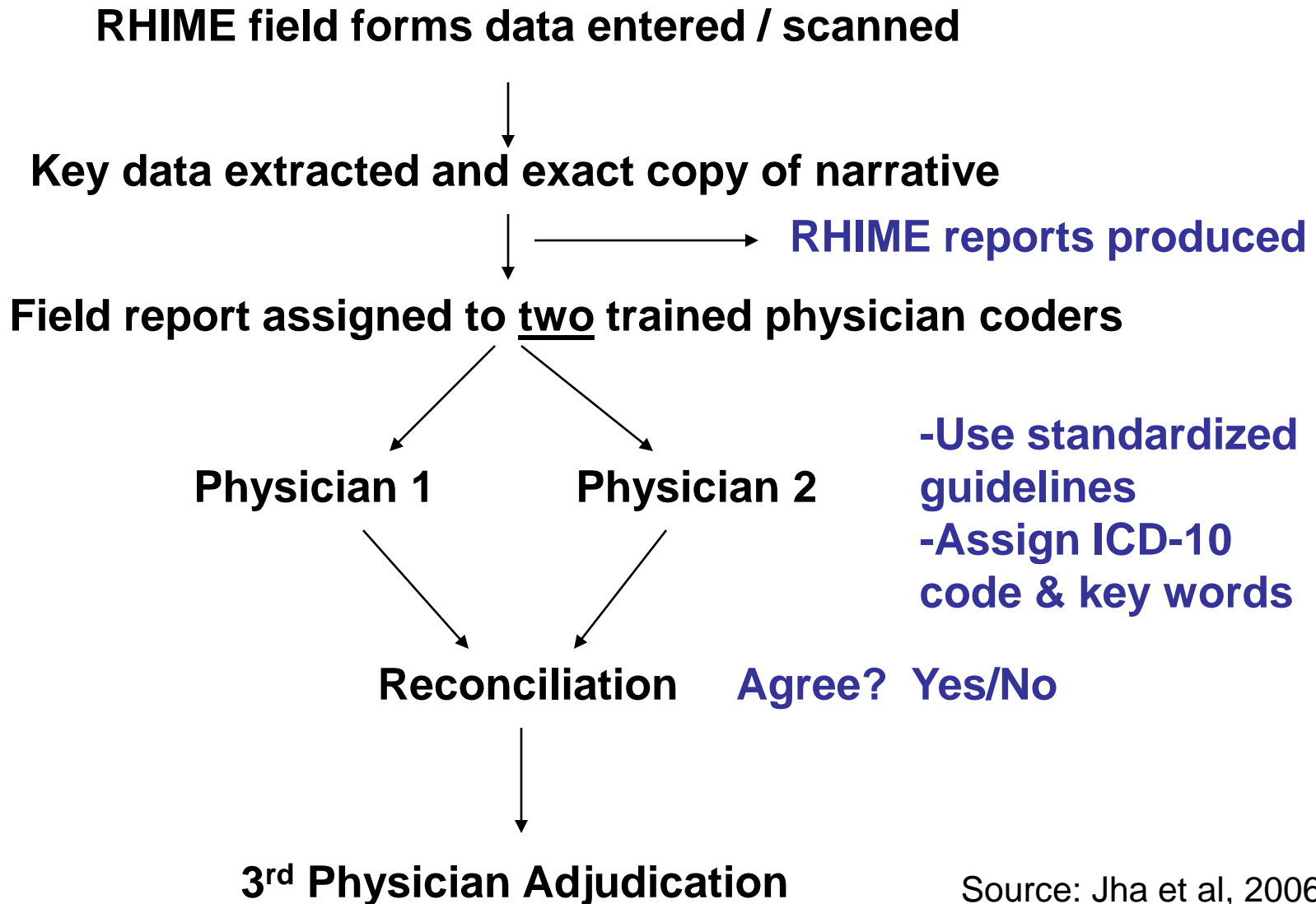
Malaria program deaths



Estimated 15,000 Deaths

Verbal Autopsy Method

RHIME method for physician coding



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

SAS data step string (text handling) functionality

Functions that Search for Strings/ Characters

ANYALNUM	NOTUPPER
ANYALPHA	FIND
ANYDIGIT	FINDC
ANYPUNCT	INDEX
ANYSpace	INDEXC
NOTALNUM	INDEXW
NOTALPHA	VERIFY
NOTDIGIT	SOUNDEX

Functions that Substitute Letters or Words in Strings

TRANSLATE
TRANWRD

Functions that Divide Strings into "Words"

SCAN
SCANQ
CALL SCAN
CALL SCANQ

Functions that Compute the Length of Strings

LENGTH	COUNT
LENGTHC	COUNTC
LENGTHM	
LENGTHN	

Method

Keyword	indexw(Kword, "FEVER")	index(up_KWord, "FEVER")	Soundex(up_KWord, "FEVER")	Malaria
Fever	p(1,1,0)	p(1,2,0)	p(1,3,0)	P(1,0)
intermit*	p(1,1,1)	p(1,2,1)	p(1,3,1)	P(1,1)
alternat*	p(1,1,2)	p(1,2,2)	p(1,3,2)	P(1,2)
chills	p(1,1,3)	p(1,2,3)	p(1,3,3)	P(1,3)
shiver*	p(1,1,4)	p(1,2,4)	p(1,3,4)	P(1,4)
rigour	p(1,1,5)	p(1,2,5)	p(1,3,5)	P(1,5)
alternat*	p(1,1,6)	p(1,2,6)	p(1,3,6)	P(1,6)
febrile	p(1,1,7)	p(1,2,7)	p(1,3,7)	P(1,7)
repeat*	p(1,1,8)	p(1,2,8)	p(1,3,8)	P(1,8)
fluctuat*	p(1,1,9)	p(1,2,9)	p(1,3,9)	P(1,9)
high	p(1,1,10)	p(1,2,10)	p(1,3,10)	P(1,10)
flutuat*	p(1,1,11)	p(1,2,11)	p(1,3,11)	P(1,11)
on & off	p(1,1,12)	p(1,2,12)	p(1,3,12)	P(1,12)
quartan	p(1,1,13)	p(1,2,13)	p(1,3,13)	P(1,13)
Headaches	p2			
Chills	p3			
Diarrhoea	p4			
Muscular pain	p5			
Lethargy	p6			
Sickness				
Coughing fits				
Abdominal pains				

Specific

Specific
**

Specific
*

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

<http://www2.sas.com/proceedings/sugj29/072-29.pdf>

SAS Macro

```
%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 stings and capture the string from where you want: Change parame
  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); /* eg. KW="FEVER" */
  S_ind=0;
  array word {50} $ 20 word1-word50;
  array S {50} $ 20 S1-S50; /* temporary indicator variable to keep KW availability */
do i=1 to 50;
  word(i)=scan( up_KWord, i, ' '); /* 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; /* probable */
else if (Indx>0 and S_ind=0) then &Var_KW_out=3; /* Less probable */
else if (Indx=0 and S_ind>0) then &Var_KW_out=4; /* Least probable*/
else &Var_KW_out=0;

%mend;
/* ----- */
Libname text 'E:\SAS Text Analysis\Presentations'; /*options mprint;*/
/*Examples:*/
%Kw_search(KW='HIGH', Var_KW_out=HIGH, Var_KW_in=KWord, Data_IN=text.t1, Data_OUT=t_out); run;
%Kw_search(KW='FEVER', Var_KW_out=FEVER, Var_KW_in=KWord, Data_IN=t_out, Data_OUT=t_out); run;
%Kw_search(KW='CHILL', Var_KW_out=CHILL, Var_KW_in=KWord, Data_IN=t_out, Data_OUT=t_out); 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

	UniqNo	codex	Final_code	KWord	Indx_W	Indx	S_ind	HIGH	FEVER	CHILL
26572	12040104	1P01	R50	<KW1> <KW2> high fever with cough, fever rose everyday, repeated attacks of chills, sweating, headache, very thirsty. <RKW1> high fever with cough, fever rose everyday, repeated attacks of chills, sweating, headache, very thirsty. <RKW2> <AKW> fever.	0	80	0	1	1	3
26573	12040160	1P01	R50	<KW1> High fever with loss of consciousness, no chills <KW2> High fever, headache, bodypain, unconscious, thirsty, weak <RKW1> High fever with loss of consciousness, no chills <RKW2> High fever, headache, bodypain, unconscious, thirsty, weak <AKW> FEVER, C	0	50	0	1	1	3
26574	12040296	1P01	R50	<KW1> FEVER NOT RESPONDING <KW2> H/O HIGH FEVER WITH CHILLS & RIGOR, SHIVERING, NAUSEA, PAIN HEADACHE, FAST BREATHING, MOUTH INFECTION, H/O SWEATING <RKW1> <RKW2> <AKW>	0	55	0	1	1	3
26575	12039721	1P01	R50	<KW1> H/O LONG DURATION FEVER NOT RESPONDING TO RX <KW2> high fever with chills coming & going with sweat, feeling cold with body pain <RKW1> high fever with chills coming & going with sweat, feeling cold with body pain <RKW2> H/O LONG DURATION FEVER NOT	0	75	0	1	1	3
26576	10053248	1P01	R50	<KW1> <KW2> High fever with chills & rigor, head ache, vomiting <RKW1> High fever with chills & rigor, head ache, vomiting <RKW2> <AKW> High fever with chills & rigor, head ache, vomiting, not improved by treatment, death, R-COD-oldage,	0	32	0	1	1	3
26577	20059493	1P01	R50	<KW1> FEVER, COUGH <KW2> fever with chills, headache, death. <RKW1> FEVER, COUGH <RKW2> fever with chills, headache, death. <AKW> INTERMITTENT HIGH FEVER FOR 1 MONTH	0	38	0	1	1	3
26578	22074515	1P01	R50	<KW1> fever high grade with chills & rigor-21d, giddiness, fall, death in 5 hr after fall, R-COD-fever <KW2> high grade fever with rigors, sweating, cough, giddiness, unconscious, died <RKW1> high grade fever with rigors, sweating, cough, giddiness, unco	0	30	0	1	1	3
26579	23073243	1P01	R50	<KW1> HIGH FEVER FOR MANY DAYS, NOT RESPONDING TO TREATMENT <KW2> fever, chills, high grade, off & on for 10 days, treated, death <RKW1> fever, chills, high grade, off & on for 10 days, treated, death <RKW2> HIGH FEVER FOR MANY DAYS, NOT RESPONDING TO TR	0	75	0	1	1	3
				<KW1> HIGH FEVER WITH CHILLS & RIGORS						

Result

State	% of fever deaths in rural areas outside health facility due to						
	Tuberculosis	Diarrhoeal diseases	Malaria	Respiratory infections	Other infectious and parasitic diseases	Fever of unknown origin	Total fever in infectious diseases
Malaria hi-prevalence states							
North-East states	14	18	20	19	24	4	100
Orissa	15	13	43	14	12	2	100
Madhya Pradesh	14	16	12	32	17	8	100
Chhattisgarh	21	11	22	28	11	8	100
Jharkhand	14	15	21	25	16	9	100
Other bigger states							
Uttar Pradesh	16	18	4	31	24	7	100
Bihar	13	15	8	38	21	6	100
Rajasthan	13	14	7	43	16	7	100
Gujarat	25	17	8	19	20	11	100
Maharashtra	23	16	12	20	25	4	100
Karnataka	16	18	5	12	33	16	100
Haryana	18	24	6	25	22	7	100
West Bengal	24	9	6	29	22	10	100
Andhra Pradesh	23	20	20	6	23	8	100
Tamil Nadu	30	15	3	12	20	20	100
Jammu & Kashmir	10	9	0	61	14	5	100
Punjab	15	16	2	23	26	19	100
Kerala	23	9	9	27	27	9	100
All other states	30	16	4	30	8	9	100
All India	16.9	16.1	11.3	27.9	20.4	7.5	100.0

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

- Zizhong Fan, Westat, Rockville, MD, Matching Character Variables by Sound: A closer look at SOUNDEX function and Sounds-Like Operator (=*) ; Paper SAS: sugi29/072-29
<http://www2.sas.com/proceedings/sugi29/072-29.pdf>
- Dhingra N, Jha P, Sharma VP, et al for the Million Death Study Collaborators. Adult and child malaria mortality in India: a nationally representative mortality survey. Lancet 2010;1016/S0140-6736(10)60831-8. published online Oct 21. [PubMed](#)
- Jha P, Gajalakshmi V, Gupta PC, et al. Prospective study of one million deaths in India: rationale, design, and validation results. PLoS Med 2006; 3: e18. [CrossRef](#) | [PubMed](#)
- SAS Inc.,(December 2010) SAS 9.1.2 documentation manual
http://support.sas.com/documentation/onlinedoc/91pdf/index_912.html

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**