Prospective Study of One Million Deaths in India: Rationale, Design, and Validation Results

Prabhat Jha[1*], Vendhan Gajalakshmi[2], Prakash C. Gupta[3], Rajesh Kumar[4], Prem Mony[5], Neeraj Dhingra[1], Richard Peto[6], RGI-CGHR Prospective Study Collaborators

1 Centre for Global Health Research, Public Health Sciences, St Michael's Hospital, and McLaughlin Centre for Molecular Medicine, University of Toronto, Toronto, Canada, **2** Epidemiological Research Centre, Chennai, India, **3** Healis-Seskarhia Institute of Public Health, Navi Mumbai, India, **4** School of Public Health, Post Graduate Institute of Medical Education and Research, Chandigarh, India, **5** Institute of Population Health and Clinical Research, St. John's Medical College, Bangalore, India, **6** Clinical Trial and Epidemiological Studies Unit, University of Oxford, Oxford, United Kingdom

Competing Interests: The authors have declared that no competing interests exist.

Author Contributions: All authors contributed to the design of the study, to analyses of the data, and to the writing of the paper.

Academic Editor: Mauricio Hernandez Avila, National Institute of Public Health, Mexico

Citation: Jha P, Gajalakshmi V, Gupta PC, Kumar R, Mony P, et al. (2006) Prospective study of 1 million deaths in India: Rationale, design, and validation results. PLoS Med 3(2): e18

Received: May 23, 2005
Accepted: October 18, 2005
Published: December 20, 2005

DOI: 10.1371/journal.pmed.0030018

Abbreviations: ICD-10, *International Statistical Classification of Diseases and Related Health Problems*; NFHS-2, National Family Health Survey 2; RGI, Registrar General of India; SFMS, Special Fertility and Mortality Survey; SRS, Sample Registration System; VA, verbal autopsy

* To whom correspondence should be addressed. E-mail: prabhat.jha@utoronto.ca

ABSTRACT

Background

Over 75% of the annual estimated 9.5 million deaths in India occur in the home, and the large majority of these do not have a certified cause. India and other developing countries urgently need reliable quantification of the causes of death. They also need better epidemiological evidence about the relevance of physical (such as blood pressure and obesity), behavioral (such as smoking, alcohol, HIV-1 risk taking, and immunization history), and biological (such as blood lipids and gene polymorphisms) measurements to the development of disease in individuals or disease rates in populations. We report here on the rationale, design, and implementation of the world's largest prospective study of the causes and correlates of mortality.

Methods and Findings

We will monitor nearly 14 million people in 2.4 million nationally representative Indian households (6.3 million people in 1.1 million households in the 1998–2003 sample frame and 7.6 million people in 1.3 million households in the 2004–2014 sample frame) for vital status and, if dead, the causes of death through a well-validated verbal autopsy (VA) instrument. About 300,000 deaths from 1998–2003 and some 700,000 deaths from 2004–2014 are expected; of these about 850,000 will be coded by two physicians to provide causes of death by gender, age, socioeconomic status, and geographical region. Pilot studies will evaluate the addition of physical and biological measurements, specifically dried blood spots.

Preliminary results from over 35,000 deaths suggest that VA can ascertain the leading causes of death, reduce the misclassification of causes, and derive the probable underlying cause of death when it has not been reported. VA yields broad classification of the underlying causes in about 90% of deaths before age 70. In old age, however, the proportion of classifiable deaths is lower. By tracking underlying demographic denominators, the study permits quantification of absolute mortality rates. Household case-control, proportional mortality, and nested case-control methods permit quantification of risk factors.

Conclusions

This study will reliably document not only the underlying cause of child and adult deaths but also key risk factors (behavioral, physical, environmental, and eventually, genetic). It offers a globally replicable model for reliably estimating cause-specific mortality using VA and strengthens India's flagship mortality monitoring system. Despite the misclassification that is still expected, the new cause-of-death data will be substantially better than that available previously.

Introduction

About 46 million of the estimated 60 million deaths per year worldwide occur in developing countries [1]. However, there is a dearth of reliable and accurate information on the causes and distribution of mortality in these countries. India has about 9.5 million deaths a year, or about one in six of all deaths worldwide. Over three-quarters of deaths in India occur in the home; more than half of these do not have a certified cause.

To meet these modern challenges of mortality measurement, the world's largest prospective study of the causes and correlates of mortality in India is being undertaken by the Registrar General of India (RGI)'s Sample Registration System (SRS). The study, called the Prospective Study of 1 Million Deaths in India, is implemented in close collaboration with the Centre for Global Health Research at the University of Toronto, leading Indian and overseas academic institutions, and the Indian Council of Medical Research. The study has several innovations that are relevant to other developing countries considering the measurement of mortality and to recent calls for improved health statistics [2–5]. It uses a well-validated household instrument to ascertain causes of death and dual recording methods to improve reliability and consistency. It is national in scale, representative of the population, and—by recording underlying demographics—able to quantify absolute mortality rates. Prospective and retrospective study designs, such as household and proportional mortality and nested case-control methods, permit quantification of correlates of mortality.

In this report, we present the rationale, design, results of validation studies to date, key statistical issues, and expansion to biologic measurement for the Million Death Study in India. We discuss the challenges in implementing modern mortality measurement and the implications for global health.

Rationale: Why Measure Mortality?

Historically reliable, representative, routine, low-cost, and long-term mortality measurements are the key to monitoring trends in health conditions of the population, detecting new epidemics (such as HIV/AIDS), spurring research into avoidable causes of death, evaluating the success of control programs, and improving accountability for expenditures on disease control [6,7]. Routinely collected data have helped to spur further research and public health action and contributed to the enormous increases in life expectancy in the 20th century [8].

Public health in industrialized countries was transformed when vital statistics on age, sex, and socioeconomic distribution of births and deaths became available in the late 19th and 20th centuries. Vital statistics have demonstrated major trends in fertility, child survival, and mortality. They have shown good news, such as the large declines in under-5 mortality and tuberculosis mortality during the 20th century. They have also raised alarm; in the mid-1940s, a dramatic increase in lung cancer deaths in British and American men after World War II led to much research on smoking [9]. In the early 1980s, routine mortality data from San Francisco revealed an exceptional increase in immune-related deaths among young men and signaled the start of the American HIV-1 epidemic [10].

Vital statistics need to keep up with modern patterns of disease. India and other countries have seen consistent decreases in child mortality (under-5 mortality has fallen by about 2% per year since 1971 in India [11]). Adult deaths in middle age (35–69 y) attributable partially to the effects of smoking, sedentary lifestyles, and higher saturated fat intake have been on the rise. More recently, deaths among young adults (15–34 y) have risen from HIV/AIDS. Reliable information on the diseases of adults and their causes are a large gap in global knowledge.

Recorded deaths from most communicable diseases or injuries generally correspond to their causes (for example, malaria deaths are caused by *Plasmodium* parasites), but the more "chronic" communicable diseases (such as tuberculosis) and most noncommunicable diseases can have multiple causes. For example, a myocardial infarction could be caused by smoking, elevated blood pressure, high lipids, or other factors. The age- and sex-specific importance of established risk factors, or combinations of risk factors, has only recently been reliably documented through appropriately large studies in Western populations, and with surprising results. For example, the association of blood pressure and vascular disease is twice as steep as previously believed [12] if blood pressure is measured reliably and the effects of "regression-dilution" bias [13] are properly considered. There are few epidemiological studies that document the age-, sex-, and region-specific hazards of blood pressure, blood lipids, and smoking in developing populations [14], where most of the world's vascular deaths occur [1].

Anecdotal evidence suggests that in India each disease that is common in one part of the country is relatively uncommon elsewhere, for reasons that are not understood. This means that there are important avoidable causes that still await discovery. Much more remains to be discovered about the novel genomic, proteomic, and other biochemical correlates of respiratory, intestinal, or other infections in general, and of the avoidable causes of chronic diseases such as cancer, heart attack, stroke, and lung disease[15,16] that currently account for most of the adult mortality in India. Even for infections such as HIV-1 and tuberculosis, there may well be genetic causes (such as polymorphisms in genes involved in innate [17,18] or adaptive [19] immune recognition) or environmental causes (such as other infections [20,21] other than the relevant pathogens) that make particular infections, or progression from infection to disease, more probable.

Alternative Designs for Measuring Mortality

The ideal mortality measurement system has several characteristics [7]. It is routine, reproducible, long-term, low-cost, and sustainable. It is reliable and representative of the population (implying that it avoids major selection biases in enrollment). It captures not only the death, but also reliably the cause based on the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10 [22]). The ideal system includes not only events or "numerators," but also underlying population demographics ("denominators").

Three systems measure mortality in India (http://www.censusindia.net/). The first, the Civil Registration System, is currently unreliable due to gross underregistration. While some areas have very good vital registration (Mumbai provides death registration as far back as 1848 [23]), overall, only 3.5 million of the estimated 9.5 million annual deaths were registered in India in 1999. Among registered deaths, cause-

of-death data are available for about one in three deaths, but this often merely subdivides deaths as due to accident, violence, or disease, without further details. Civil registration is the accepted mortality measurement system in Western countries where coverage nears 100%. However, access to medical care is far less common in India, and most deaths occur at home rather than in hospitals. Eventually, civil registration will increase, as has happened recently in China. But this may take decades; the United States took the better part of a century to increase death certification, and some states did not have complete coverage until the 1970s [24].

The second system is the Medically Certified Causes of Death. However, this covers only about 0.4 million deaths and is largely confined to selected urban settings that are not representative of the general population. Problems with inconsistent physician attribution of causes of death, especially for senility and ill-defined causes have been noted [6,7].

The third is the SRS (described in detail below). Only the SRS is representative of both urban and rural settings of India, covering some 6,700 to 7,600 units randomly selected from the preceding census. The SRS is much smaller, though, covering only 0.05 million deaths. Thus, its chief drawback is that it cannot yet provide district-level data for local decision making, and it lacks sufficient power to generate yearly rates for less common causes, such as maternal deaths. Until recently, the SRS did not adequately capture information on causes of death. However, we have addressed this gap by developing and implementing verbal autopsy (VA)—an innovative method to estimate cause-specific mortality.

Finally, it is worth noting that econometric models of cause-specific mortality [1] are no replacement for direct measurement. Indirect estimates are only as good as their underlying data. The econometric models have not been well tested in the presence of HIV/AIDS growth. In India the global burden of disease varies considerably depending on the assumptions used. For example, the 1994 global burden of disease estimated 0.78 million cancer deaths in 1990, but cancer registry data suggested a much lower figure of 0.43 million deaths [25]. The 1996 version of the global burden of disease projected 0.95 million deaths from tuberculosis in 2000; the 1999 version estimated 0.42 million deaths in 1998 [26].

Study Objectives

In light of the above, we have the following objectives: (a) to reliably document cause-specific mortality from 2001 to 2003 (3 y; approximately 150,000 deaths) within the SRS to establish regional-, gender-, and age-specific variation and patterns of mortality; (b) to document causes of death with routine use of VA in the new SRS sampling frame from 2004 to 2014 (10 y; approximately 700,000 deaths expected); (c) to improve our understanding of selected risk factors, most notably tobacco use, alcohol use, indoor air pollution, fertility preferences for male children and its effect on female survival, immunization, and migration through linking of mortality outcomes with exposure status, using retrospective and prospective methods; and (d) to expand the new SRS to obtain reliable epidemiological evidence about the relevance of physical (such as blood pressure and peak flow), behavioral (such as migration and HIV-1 risk), and biological (such as blood lipids and gene polymorphisms) measurements to the development of disease in individuals or disease rates in populations.

Methods

Study Setting

This study is conducted within the SRS, a large, routine demographic survey and the primary system for the collection of Indian fertility and mortality data since 1971 [27]. There are two SRS sample frames. The first SRS sample frame covers 6.3 million people (including 2.9 million adults aged 25 y or older) in all 28 states and seven union territories of India. An average of 150 households are drawn from each of 6,671 sample units (4,436 rural and 2,235 urban), which in turn are selected using 1991 census data. The new SRS sample frame covers about 7.6 million people (including 3.5 million adults aged 25 y or older) in all 28 states and seven union territories of India. Households are drawn from 7,597 sample units (4,433 rural and 3,164 urban) selected from the 2001 census.

SRS sample units are randomly selected to be representative of the population at the state level. The sample design is a unistage stratified simple random sample without replacement. The sample size for the first and new SRS sample frames are based on total fertility rates and infant mortality rates, respectively. Within the SRS, selected households are continuously monitored for vital events by two independent surveyors. The first is a part-time enumerator (commonly a local school teacher familiar with the area/village) who visits the home every month. The second is a full-time (nonmedical) RGI surveyor who visits the home every 6 mo. Each independently records the births and deaths in the household for a 6-mo period. A third staff member does a reconciliation of the two reports, arriving at a final list of births and deaths for each household, which completes each half-yearly survey. The RGI surveyors each cover about 150 households with a total average population of 900 (ranging from 700 to 1,500), and report about 50 deaths (and about 150 births) every 6 mo.

Plan of Investigation

Box 1 and Figure 1 provide an overview of the study methods. In brief, the method involves 800 trained (nonmedical) RGI surveyors implementing VA reports among enrolled populations every 6 mo. A random 10% of the VA fieldwork is repeated by an independent audit team. After data entry, field reports are sent electronically to two independent and trained physicians who assign cause of death, based on the ICD-10 [22], using a Web-based system. The two physicians have to agree on the underlying cause of death, and if they do not, such records undergo reconciliation and third-physician adjudication.

Text S1–S3 provide the details of the field collection methods, resampling, physician coding, data management, and research ethics. The full protocol, field instruments, training manuals, slide presentation of validation results, and procedures are available at http://www.cghr.org/project.htm.

Estimated Sample Size and Distribution

The primary outcomes of interest are all-cause mortality and cause-specific mortality. We expect about 150,000 cause-specific deaths determined through VA and about twice as many deaths (300,000) to study all-cause mortality in the 1998–2003 sample frame. The expected age distribution of cause-specific deaths captured with VA, based on the age-specific SRS death rates in 2001, is shown in Figure 2. Using indirect World Health Organization estimates on causes of death in India [1], we also show the approximate numbers of

Box 1. SRS VA Methods Overview

Design of verbal autopsy questionnaire. Combined open/closed format. Structured questions accompanied by an open-ended narrative. Symptom list to assist attribution of deaths.

Questionnaire layout. One-page, double-sided, scannable forms. Four age-specific forms (neonatal, child, adult, and maternal). Forms available in either English or Hindi.

Interviewers. Nonmedical RGI surveyors (mostly male) with knowledge of local language(s) and trained in VA instrument.

Interview technique. One-on-one interviews during home visits. Duration of 30–45 min.

Respondents. Family members or other informants (usually neighbors or close associates of the deceased).

Recall period. Usually less than 6 mo, but still valid up to 3 y.

Data quality. Random resample of 10% of all deaths to ensure completeness of fieldwork.

Derivation of diagnosis. Central medical review of cause by two independent physicians using modified VA reports (Physician Reports) and an Internet-based Web application. Adjudication of disagreements by an expert physician.

Mortality classification. *The International Classification System of Diseases and Related Health Problems,* Tenth Revision (ICD-10).

Sample size. One million deaths (about 0.3 million from 1998–2003, 0.7 million from 2004–2014).

major categories of deaths expected (in thousands) between ages 25 to 69 y (Table 1).

As of November 2005, 140,000 VA reports have been collected from all SRS units, and about 35,000 records have undergone double physician coding and reconciliation.

Overall we expect several thousand tuberculosis, vascular, and cancer deaths among adults. Such numbers are not excessively large, particularly if the age- and sex-specific relevance of several risk factors is to be assessed simultaneously. For example, assuming a power of 90% and two-sided α of 0.001, and assuming a 40% smoking rate among male controls, the study would have sufficient power to detect relative risks among men as low as 1.4 for lung cancer, 1.1 for all cancers or for tuberculosis, and 1.1 for cardiovascular disease [28]. Thus, the study has robust statistical power to detect small but significant increases in risk for most key variables of interest.

The main planned analyses involve simple tabulations and standard Cox proportional hazards analyses, calculating relative risks that are standardized for age, educational level, and selected covariates as relevant. Analyses of adult deaths will focus on deaths at ages 25–69 y, as these are much less likely to be misclassified than deaths occurring after age 70.

Sample-size estimates for the 2004–2014 sample frame are less certain, as the cause-specific child and adult mortality will depend on the rapidity of declines in childhood mortality and increases in HIV-1-related mortality [11]. However it is reasonable to assume that with expanding sample size (reflecting the growth of populations in the SRS households), some 700,000 deaths will occur over the 10-y period.

Reliable Measurement of Absolute Rates and Relative Risks

The ability to generate absolute rates depends on completeness of enumeration and being able to calculate underlying demographic denominators, including migration. Evaluations suggest that the SRS has a high ascertainment rate of expected events, although adjustments may be needed for certain age groups. A 1984 study [29] concluded that the system captured 90% of deaths between 1971 and 1980. The SRS, which employs continuous enumeration, is more sensitive at detecting child deaths than are single surveys, and it has recorded more child deaths than those estimated by the National Family Health Survey 2 (NFHS-2), a nationally representative demographic and health survey that interviewed nearly 0.1 million households (http://nfhsindia.org/index.html). Bhat [30] found that adult deaths were underreported by about 13% to

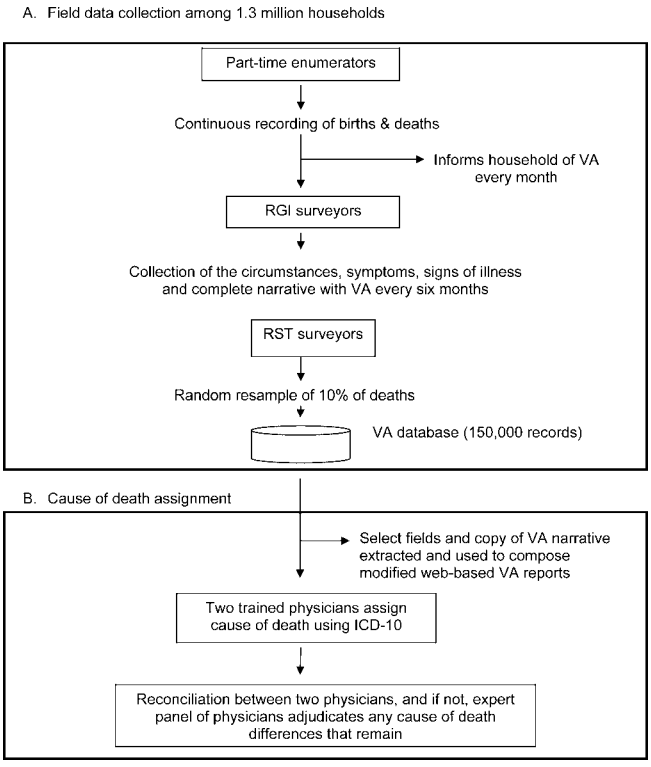

Figure 1. SRS VA Activities
DOI: 10.1371/journal.pmed.0030018.g001

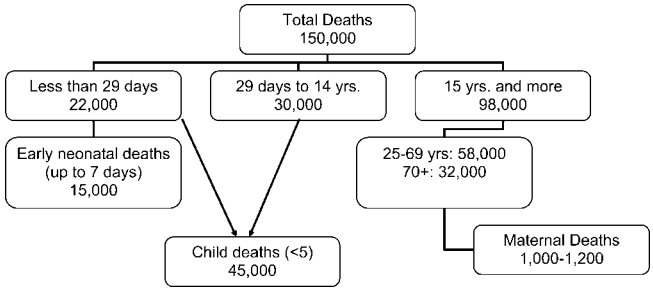

Figure 2. Expected Number of Deaths by Age Group Using VA (Thousands), 2001 to 2003
DOI: 10.1371/journal.pmed.0030018.g002

Table 1. Expected Number of Deaths between Ages 25 and 69 by Cause (Thousands), 2001 to 2003

Causes	Male	Female
All infections and maternal diseases	9.2	5.9
Tuberculosis	3.4	1.5
HIV/AIDS	1.9	0.4
Lower respiratory infections	1.3	1.0
Maternal conditions	N/A	1.1
Other infections	2.6	1.9
All noncommunicable diseases	21.2	14.6
Lung cancer	0.8	0.2
Upper aerodigestive cancer (mouth, oropharynx, esophagus)	0.8	0.3
Other cancers	1.8	2.3
Cardiovascular diseases	11.4	7.6
Chronic obstructive pulmonary disease	2.4	1.5
Other noncommunicable	4.0	2.7
Injuries	4.5	2.5
Unintentional	3.2	1.9
Intentional	1.3	0.6
Total (thousands)	34.9	23.0

The expected numbers of deaths for these causes were calculated by using the 2001 estimates of mortality from the global burden of disease for India ([1]; http://www.fic.nih.gov/dcpp/gbd.html) and multiplying the proportion of specific causes by the total age-specific number of deaths expected in the SRS study during the follow-up from 2001 to 2003.
N/A, not available.
DOI: 10.1371/journal.pmed.0030018.t001

14% (slightly higher in females) and that there is evidence to suggest that the undercount has increased slightly in particular states recently. It is expected that the new SRS sample frame should have corrected the underreporting of adult deaths noted in the first SRS sample frame. Formal demographic evaluation will be done once results from the new SRS are collected. Because household composition is updated every 6 mo, and in-migration and out-migration are traced, it should be possible to calculate absolute mortality rates in the population, including person years at risk. Deaths among in-migrating groups can be excluded, as the death report lists if the person was a usual resident in the SRS unit or not. Some periodic correction for undercount will be needed.

Loss to Follow-Up

The baseline survey for the study was called the Special Fertility and Mortality Survey (SFMS) and was confined to usual residents of households only [31]. Those living in these households in 1998 were a subselection of the SRS baseline sample frame originally done in 1993. Experience from a prospective study in Chennai (V. Gajalakshmi, personal communication) suggests that if people are resident for a few years, they are not likely to move again. Thus, the enrolled group in the current SRS sample frame is less likely to migrate than would be the general population. This is supported by data from our own small pilot examination of 389 randomly chosen SRS records in two northern states, which showed that only 6% (25/389) of the households had moved from 1998 to 2002.

Nonetheless, we do expect loss to follow-up caused by out-migration from the SRS unit. The risk ratios would be biased downward if such loss to follow-up is nondifferential between exposed and unexposed groups. Similarly, risk ratios would be biased downward if impoverished people (who are most likely exposed to smoking and other risk factors) migrate.

Results

Validation Studies of Mortality Outcomes

VA relies on the assumption that most causes of death have distinct symptoms and signs that can be recognized, recalled, and reported by household members or associates of the deceased to a trained, usually nonmedical field-worker. Further, it is assumed that deaths characterized through VA possess a distinct set of features that can be distinguished from other underlying causes of death [32]. Thus, diseases with very distinct symptoms and signs, such as tetanus, that are recognized by the local population may be more suitable for VA than systemic diseases, such as malaria, which is associated with many common symptoms and signs. Factors that influence the validity and reliability of VA include the VA instrument (mortality classification, diagnostic procedures), the data collection procedures (recall period, interviewer's characteristics, respondent's characteristics), and the underlying distribution of cause-specific mortality in a given population [33–35]. Although there is variation in the sensitivity and specificity for specific conditions, VAs are now of established value in helping to classify the broad patterns of mortality for childhood deaths in populations that are not covered by adequate medical services. VAs have also been used to assess the causes of maternal deaths [36].

Background work for this study included two validation studies of adult deaths in India [37–39]. The first study [37,38] developed and tested a VA instrument among 48,000 adult deaths in urban Chennai and 32,000 adult deaths in rural Tamil Nadu, including a 5% random resample. VA conducted by trained nonmedical field-workers resulted in 90% successful reporting on cause of deaths for middle-age adults (25–69 y). The VA instrument reduced the proportion of adult (age 25 or older) deaths attributed to unspecified or unknown causes from 54% to 23% in urban areas and from 41% to 26% in rural areas. VA yielded fewer unspecified causes (only 10%) than the death certificate (37%)—particularly for the deaths that did not occur in hospital—and often yielded somewhat more specific information, for example, about the approximate site of origin of a cancer, or about evidence of tuberculosis, stroke, myocardial infarction, or diabetes (Table 2). The urban Tamil Nadu study also compared VA results to those from a Chennai population-based cancer registry. The VA sensitivity to identify cancer was 95% in the age group 25–69 y, and VA identified 288 deaths that were not registered in the Chennai Cancer Registry.

The second validation study compared all-cause mortality determined by VA against hospital-based records for 262 adult deaths in northern India [39]. Deaths characterized with VA were compared to medically certified cause-of-death certificates for patients who had died in a hospital. Cause-specific mortality fractions assigned by the VA method were statistically similar to the causes arrived at by review of hospital records ($p > 0.05$). Specificity was high (>95%) for all broad cause groups except cardiovascular (79%) diseases. Sensitivity was highest for injuries (85%), and it was in the range of 60% to 65% for circulatory diseases, neoplasms, and infectious diseases. Sensitivity was low (20% to 40%) for respiratory, digestive, and endocrine diseases. These figures are broadly in agreement with the results from a multicenter validation study of VA for adult deaths conducted in Africa, which found a sensitivity and specificity of 82% and 78%,

Table 2. Cause of Death Based on Vital Statistics and VA of 48,000 Adult (>25 y of Age) Deaths in Chennai, India: 1995–1997[a]

Cause of Death	Cause of Death in Vital Statistics Division		Cause of Death Based on VA	
	Male (Percent)	Female (Percent)	Male (Percent)	Female (Percent)
Vascular disease	8,319 (30)	5,168 (25)	11,056 (41)	7,435 (37)
Tuberculosis (TB)	1,399 (5)	372 (2)	2,231 (8)	575 (3)
Other respiratory diseases	1,088 (4)	596 (3)	1,597 (6)	855 (4)
Neoplasm	1,163 (4)	1,002 (5)	2,344 (9)	1,999 (10)
Infection (excluding respiratory and TB)	584 (2)	303 (2)	1,034 (4)	618 (3)
Unspecified medical	12,291 (44)	11,511 (56)	4,367 (16)	5,889 (29)
Other specified medical	1,899 (7)	1,045 (5)	4,414 (16)	2,804 (14)
Cause not known	983 (4)	634 (3)	Nil	Nil
Total deaths: medical	27,726	20,631	27,043	20,175
Total deaths: external causes	Excluded	683	456	
Total deaths (medical + external)	27,726	20,631	27,726	20,631

[a] See [37,38].

DOI: 10.1371/journal.pmed.0030018.t002

respectively, for all communicable diseases, and a sensitivity and specificity of 71% and 87%, respectively, for all noncommunicable diseases [34].

Preliminary results from two states indicate that the distribution of underlying causes of death based on the random reinterview does not differ substantially from the cause of death derived from the VA reports of the original RGI surveyors (Table 3).

Validation Studies of Exposures Measured at Baseline

Baseline exposures were captured in a one-time SFMS conducted within the SRS in February 1998. Baseline exposures captured in the SFMS included socioeconomic information (education, occupation, household income, household composition), water and sanitation facilities and other living conditions, smoking and its type (bidi, cigarette, hukka, or other), age at onset, alcohol use and frequency per week, past history of various medical conditions, and the type of cooking fuel used (major and secondary, and use of a separate kitchen). The survey also recorded deaths, although not their causes, in 1997 [31].

The 2004 baseline survey in the new SRS sample frame (2004–2014) recorded similar exposures and added history of disability from various medical conditions, recent short-term illnesses, tobacco smoking and chewing, alcohol use, vegetarianism, and maternal history, including contraceptive use and pregnancies.

The reliability and representativeness of the SFMS baseline survey can be assessed, in part, by comparing the age-specific prevalence of smoking and alcohol consumption among males found in the SFMS and other standard surveys, such as the Indian Census or NFHS-2. Text S1 provides additional validation studies, such as birth order of children and measures of indoor air pollution.

Age-Specific Prevalence of Smoking in Adult Males

Smoking is currently common only among Indian males. The SFMS and NFHS-2 report very similar trends in age-specific prevalence of current male smokers. The steepest increase in prevalence of smoking is between 20–30 y of age (Figure 3). Similarly, the prevalence of smoking among males across 25 states shows a strong correlation between the two

Table 3. Cause of Death Results for RGI Surveyors and Resample Teams in Tamil Nadu, Maharashtra, and Goa Using VA (Deaths >28 d) in 2003

Causes	Tamil Nadu				Maharashtra/Goa			
	RGI Surveyors		Resample Team		RGI Surveyors		Resample Team	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Vascular	300	25.8	46	28.2	412	23.5	80	25.0
Tuberculosis (respiratory)	50	4.3	1	0.6	85	4.8	8	2.5
Other respiratory diseases	105	9.0	10	6.1	154	8.8	22	6.9
Cancer	100	8.6	18	11.0	113	6.4	15	4.7
Infection	127	10.9	24	14.7	139	7.9	44	13.8
Diabetes	42	3.6	3	1.8	56	3.2	6	1.9
Peptic ulcer	21	1.8	0	0	7	0.4	2	0.6
Undefined/inadequate information	164	14.1	23	14.1	152	8.7	45	14.1
Other specified	95	8.2	12	7.4	480	27.4	74	23.1
External causes	161	13.8	26	16	155	8.8	24	7.5
Total *(n)*	1,165	—	163	—	1,753	—	320	—

DOI: 10.1371/journal.pmed.0030018.t003

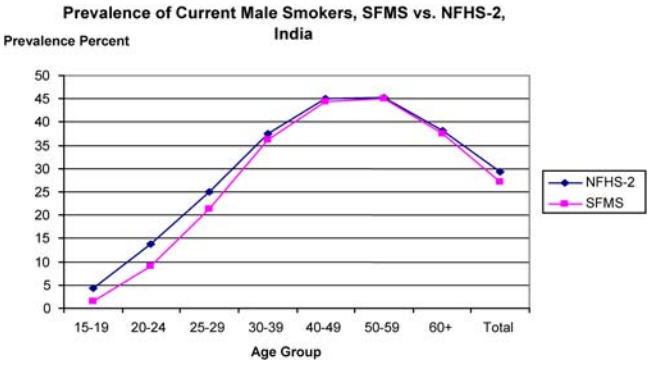

Figure 3. Prevalence of Current Male Smokers: SFMS 1998 versus NFHS-2 1998–1999

DOI: 10.1371/journal.pmed.0030018.g003

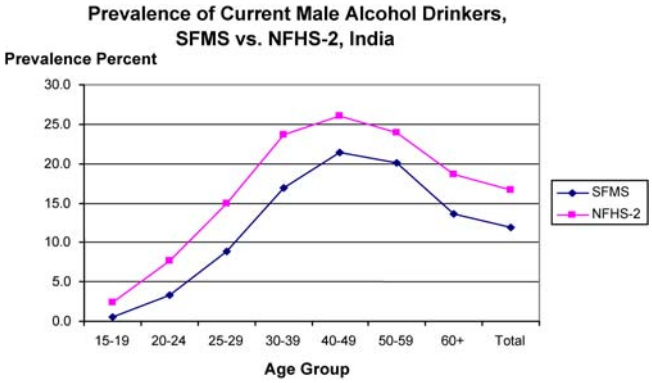

Figure 4. Prevalence of Current Male Alcohol Drinkers: SFMS 1998 versus NFHS-2 1998–1999

DOI: 10.1371/journal.pmed.0030018.g004

studies (Pearson correlation coefficient or r^2 of 0.92; data not shown). We plan to remeasure the extent of smoking by each of several thousand living adults so as to correct for regression-dilution bias [13]. Additionally, to verify and validate smoking status, we will use a simple hand-held carbon monoxide breathalyzer on about 1,000 smokers and nonsmokers in select states. These simple breathalyzers appear to be effective at detecting smoking status [40].

Age-Specific Alcohol Consumption in Adult Males

As with smoking, a similar pattern in age-specific alcohol consumption is captured by the SFMS and NFHS-2 (Figure 4). The discrepancy in absolute prevalence for each age group may be due to the difference in the sex of the respondents, where females are either overreporting male alcohol consumption in the household in the NFHS-2, or males are underreporting their own consumption in the SFMS. It should also be noted that, unlike the NFHS-2, usually the male head of the household is the respondent of the SFMS.

Retrospective Household Case-Control and Proportional Mortality Methods

The determinants of death can be identified by comparing risk factors between the dead and living. Such household case-control studies use the dead as cases and their surviving spouses or close relatives as controls. A retrospective study in Chennai of 43,000 male deaths and 35,000 living controls [41], using these methods, has documented that throughout middle age, the death rates from medical causes of smokers were double those of nonsmokers (standardized risk ratio at ages 25–69 of 2.1, with 95% confidence interval 2.0–2.2, smoking-attributable fraction 31%). A large part of this excess risk was from tuberculosis and vascular deaths. If these hazards are similar across India, then about half of all tuberculosis deaths in India could be accounted for by smoking. We will apply similar retrospective methods for smoking, tobacco chewing, and alcohol, as these exposures are gathered for dead adults and from a living household respondent.

Simply asking about the dead person's risk factors could be useful. A recent retrospective study of 1 million deaths in China compared the proportions of smokers and nonsmokers who have died of tobacco-attributable diseases versus non-tobacco-related diseases (chiefly injuries), to calculate the excess in smokers [42]. Our preliminary pilot studies among childhood deaths in Northern India found that 61% of children (16/26) who died of vaccine-preventable diseases were not immunized in comparison to 40% of control children (18/45) who died from injuries. The crude odds ratio of 2.4 suggests that half of the vaccine-preventable child deaths need not have occurred.

Using these two retrospective methods, we are conducting validation studies of selected risk factors such as childhood immunization, childhood malnutrition, alcohol, male time away from home (as a proxy of HIV-1-related sexual risk taking), and other variables.

Discussion

Key Design Challenges

Several challenges stand out in the design and implementation of this massive study. First is its sheer size. A total of 14 million people is a large sample frame, but commensurate with the needs of monitoring health status among the one billion people in India. The study builds upon the routine infrastructure that has been in place for over 30 y within the SRS. It works with the leading demographic research organization in India (for example, the decennial census involves the RGI hiring two million enumerators to survey about 150 million households within 25 d). Second is the need for simplification of methods to make such a scale of study practicable. The study is not the only epidemiological study needed in India, but its unique focus and scale raise important challenges in the development of new methods. The Million Death Study uses a "large simple" design, which places specific demands on ensuring that new procedures (such as the new VA instrument and physical and biological measurements) are rigorously tested, piloted, and simplified. The third challenge is sustainable funding. The current SRS sample frame is being followed on a total incremental budget of US$2 million (or less than 33¢ per person). This does not include the core funding of the RGI surveyors and related infrastructure. With inclusion of those, the overall study can still be done in a highly cost-effective manner at well under US$1 per person. Collection, processing, and long-term storage of biological samples is expected to cost US$2–5 per person for dried blood spots and US$10–20 per person for a 10-ml tube of blood. These costs compare extremely favorably with those existing biobanks in the United Kingdom and elsewhere [43,44]. Most of the infrastructure costs

will be sustained by the Government of India. However, we believe that redirecting some of the considerable, and often inefficient, spending on monitoring individual disease projects is required to enhance the SRS.

Key Design Issues for Blood-Based Genetic Epidemiology

Discovery of "new" risk factors should benefit from the recent and extraordinarily rapid progress by many different research groups and biotechnology companies in developing low-cost, miniaturized methods for the simultaneous assay in small volumes of blood (or, perhaps, dried blood spots) of vast numbers of nucleic acid fragments, host genetic factors, proteins, small molecules, and pathogens [45]. This rapid biotechnological progress has been backed by increasingly sophisticated computer software to help interpret the mass of numerical information that can be generated from each person's blood, yielding within the next few years many, as yet unforeseen, qualitatively different analytic capacities. Appropriately large-scale epidemiological studies that acquire blood (or other) samples from individuals and systematically link them to relevant measures of disability and future mortality are required to make such technological progress relevant to human populations. It is particularly important for such studies to address the shortcomings of the existing biological sample collections that are now being undertaken [43,44], including the particular circumstances of India and the specific infectious diseases common to developing countries.

Key issues which arise in moving to blood-based epidemiology include the choice of blood sample, long-term storage and retrieval, and statistical design issues.

Choice of Blood Sample

Biological specimen collection procedures will build on experience in China, India, and elsewhere. We will undertake pilot studies to test the feasibility and acceptability of methods, to ensure that biological samples collected can serve as a long-term source of DNA for genotyping and that material collected is suitable for biochemical, hematological, proteomic, and other assays [46]. Major options include a nonfasting blood sample collected into one 10-ml ethylenediaminetetraacetic acid Vacutainer, which has been shown previously by the University of Oxford laboratory to allow a wide range of assays [47]. This system is used by the ongoing Chinese Kadoorie Study of 500,000 adults and by the UK Biobank project of the same size [43]. Alternatively, dried blood samples on filter paper have the advantages of easy storage and transport, as well as being less intrusive (i.e., by finger prick rather than by venipuncture). Dried blood spots have been recommended by the World Health Organization for use in field HIV-1 investigations and have been used in various studies within India [48,49].

We currently plan pilot studies of dried blood spots or tubes of blood among 4,000–9,000 adults in several SRS units in four to five states. These pilots will focus on standardizing methods to obtain anthropometric, behavioral, and physiologic measurements and to evaluate the alternative methods for collection of biological specimens for their utility, cost, and practicability (dried blood spot or tube of blood). The pilots will focus on feasibility of field methods, simplification of approaches, and quality control. The results of the pilot studies, which are expected by June 2006, will help inform the design of the larger survey. A special survey will help define practicable questions on HIV-1 risk behaviors.

Long-Term Storage, Retrieval, and Analyses

We are undertaking systematic reviews of the literature and reviewing available assays in India to examine which current bioanalytic methods can be used to test current hypotheses (either for correlates of infection or chronic disease). Using the above pilot samples, we will develop testing and analytic methods for biological samples that will permit high-throughput, low-cost, and high-quality assays to be run and also permit the long-term reliable storage and retrieval of such samples. Dried blood spots may well be stable in a basic 4 °C refrigerator, with minimal storage requirements [46]. With a population of one billion, India certainly requires at least two major biorepositories (including splitting samples, which safeguards against loss at one facility). We are developing plans for long-term biorepositories with the ability to store samples for decades. Much will depend on the choice of the final assay.

Nested Case-Control Methods and Genetic Association Studies

There is a long list of biological factors in blood that might be correlates of disease-specific mortality. To be efficient, we will use a nested case-control approach in our prospective study. Biological samples are taken from all adults in a baseline survey and stored long-term; then, when sufficient numbers of cases with the disease of interest have died (based on the 6-mo follow-up of the causes of death), aliquots from those cases are retrieved from storage, plus aliquots from a few matched controls per case; and, finally, the factors of interest are assayed in these cases and these controls. This design is similar to that in the Chinese Kadoorie Study and the UK Biobank project [43]. As noted above, the number of deaths is likely to be substantial for most of the common diseases of interest to make this an efficient strategy.

Under some circumstances, studies that used population-based controls to study gene-disease associations have been biased due to underlying variation in gene frequencies between populations ("population stratification" [50]). Genetic epidemiology can provide reliable population-based estimates of disease–allele frequency, penetrance, and attributable risk, particularly if designs that account for this bias are employed, and accordingly this has led to greater emphasis recently on family-based association designs. However, there is also evidence that well-designed population-based studies are sometimes superior [51]. The SRS offers a unique opportunity to explore statistical design issues, as both general population and family-based sampling is possible. Family-based studies allow for population-specific estimates of clustering of disease and correlation (heritability). Further, families with a large number of siblings could be oversampled to increase power for genetic linkage studies. The use of "genomic controls" where anonymous markers are genotyped to test for population stratification can be employed [52].

Limitations

The study has several limitations. First, despite a very large sample size, the statistical power remains modest for less common causes such as those leading to maternal deaths. However, for most of the major public health conditions of importance, sufficient events should occur to generate plausible absolute rates and relative risks for risk factors. Second, careful attention in design means that most identified biases should be minimized and should ensure

high internal consistency of fieldwork and coding. However, periodic revalidation of mortality outcomes against external standards will be needed. Similarly, continuous improvement of exposure measurements, partially through careful pilots, will be needed. Third, VA yields broad classification of the underlying causes in about 90% of deaths before age 70. In old age, however, the proportion of classifiable deaths is lower. For some specific conditions, such as childhood pneumonia, the cause-specific fractions may be difficult to estimate below certain levels [53].

Significance to Global Health

The Million Death Study will reliably document not only the underlying cause of child and adult deaths, but also key risk factors (behavioral, physical, environmental, and eventually, genetic). It offers a globally replicable model for reliably estimating cause-specific mortality, using VA, and strengthens India's flagship mortality monitoring system. Despite the misclassification that is still expected, the new cause-of-death data will be substantially better than that available previously in India. The study builds India's capacity for research and for public health action. It provides a large, representative, low-cost and long-term system to reliably track the health status of one billion people for the next decade or longer.

Supporting Information

Text S1. Supplemental Text October 14

Found at DOI: 10.1371/journal.pmed.0030018.sd001 (80 KB DOC).

Text S2. Ethics Approval ICMR

Found at DOI: 10.1371/journal.pmed.0030018.sd002 (132 KB JPG).

Text S3. Ethics Approval PGI

Found at DOI: 10.1371/journal.pmed.0030018.sd003 (117 KB JPG).

Acknowledgments

The largest proportion of the study costs are met by the Government of India as part of the routine costs of running the SRS. External funding from the study comes from the National Institute of Health Tobacco Research Grant (R01 TW05991–01; Aron Primack), the Canadian Immunization Initiative of the International Developmental and Research Centre (Grant number 102172; Sharmila Mhatre), and the Canadian Institute of Health Research (Establishment Grant number IEG-53506; Mark Bisby). Funding also comes from unrestricted grants from the McLaughlin Centre for Molecular Medicine, University of Toronto (No. 1901643995; Duncan Stewart), St. Michael's Hospital (Arthur Slutsky), and University of Toronto (Harvey Skinner). PJ is supported by a Canada Research Chair of the Government of Canada. We thank Dr. Paul Doherty for editorial assistance. External funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

RGI-CGHR Prospective Study Collaborators
RGI-CGHR National Centre, Office of the Registrar General, RK Puram, New Delhi, India: DK Sikri[a], RC Sethi [a,b], N Dhingra[a,b], DK Dey, M Jain, S Jain, K Lal, and L Sushant.
Indian Academic Partners
Clinical Epidemiology Resource and Training Centre, Trivandarum, India: KB Leena and KT Shenoy[a].
Department of Community Medicine, Gujarat Medical College, Ahmedabad, India: DV Bala, P Seth, and KN Trivedi[a].
Department of Community Medicine, Kolkatta Medical College, Kolkatta, India: SK Roy[a].
Department of Community Medicine, Osmania Medical College, Hyderabad, India: P Bhatia[a].
Department of Community Medicine, Regional Institute of Medical Sciences, Imphal, India: L Usharani[a].
Department of Community Medicine, SMS Medical College, Jaipur, India: AK Bharadwaj[a].
Epidemiological Research Centre, Chennai, India: V Gajalakshmi[a,b].
Gandhi Medical College, Bhopal, India: R Dikshit[a] and S Sorangi.
Healis-Seskarhia Institute of Public Health, Navi Mumbai, India: PC Gupta[a,b], MS Pednekar, and S Sreevidya.
Institute of Population Health and Clinical Research, St. John's Medical College, Bangalore, India: A Kurpad, P Mony[a,b], and M Vaz.
King George Medical College, Lucknow, India: S Awasthi[a].
North Eastern Indira Gandhi Institute of Regional Medical Sciences, Shillong, Meghalaya, India: FU Ahmed[a].
Regional Medical Research Center, ICMR Institute, Bhubaneshwar, India: AS Karkettaa and K Dar.
School of Preventive Oncology, Patna, India: DN Sinhaa.
School of Public Health, Post Graduate Institute of Medical Education and Research, Chandigarh, India: N Kaur, R Kumar[a,b], and JS Thakur.
Other Partners
Clinical Trial and Epidemiological Studies Unit, University of Oxford, Oxford, England: Z Chen, R Collins, and Sir R Peto[a,b].
Hospital for Sick Children, University of Toronto, Toronto, Canada: A Patterson and S Schrier.
Indian Council of Medical Research, New Delhi, India: NK Ganguly[a].
Mt. Sinai Hospital, University of Toronto, Toronto, Canada: J McLaughlin.
McLaughlin Centre for Molecular Medicine, University of Toronto, Toronto, Canada: P Jha[a,b], K Kain and R Kaul.
United Arab Emirates University, Al Ain, UAE: N Naglekerke.
World Health Organization, Geneva, Switzerland: T Boerma[a], T Evans[a], and K Shibuyi.
World Health Organization, South East Asia Regional Office, New Delhi, India: N Singh and T Sein.
Global Coordinating Centre, Centre for Global Health Research, St. Michael's Hospital, University of Toronto, Canada: B Jacob, P Jha (Principal Investigator)[a,b], R Kadmood, C Major, J Moore[b], P Parra, S Sgaier, H Shadmand, P Vasa[b], and F Zhang.
[a]Member Advisory Committee.
[b]Writing committee for this report.

References

1. World Health Organization (2002) Reducing risks: Promoting healthy life: World health report. Geneva (Switzerland): World Health Organization. 230 p.
2. Stumbling around in the dark [editorial] (2005). Lancet 365: 1983
3. Horton R (2005) The Ellison Institute: Monitoring health, challenging WHO. Lancet 366: 179–181.
4. Stansfield S (2005) Structuring information and incentives to improve health. Bull World Health Organ 83: 562–563.
5. Murray CJ, Lopez AD, Wibulpolprasert S (2004) Monitoring global health: Time for new solutions. BMJ 329: 1096–1100.
6. Mitra B (1999) India's mortality measurement systems. In: Asma S, Jha P, Gupta PC, editors. Counting the dead in India in the 21st century. Atlanta (Georgia): US Centers for Disease Control. pp. B2–B8
7. Jha P (2001) Reliable mortality data: A powerful tool for public health. Natl Med J India 14: 129–131.
8. Jha P, Slutsky AS, Brown D, Nagelkerke N, Brunham BG, et al. (2004) Health and economic benefits of an accelerated program of research to combat global infectious diseases. CMAJ 171: 1203–1208.
9. Doll R, Peto R, Boreham J, Sutherland I (2004) Mortality in relation to smoking: 50 years' observations on male British doctors. BMJ 328: 1519.
10. Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, et al. (1981) P*neumocystis carinii* pneumonia and mucosal candidiasis in previously healthy homosexual men: Evidence of a new acquired cellular immunodeficiency. N Eng J Med 305: 1425–143l.
11. Jha P (2002) Avoidable mortality in India: Past progress and future prospects. Natl Med J India 15: 32–36.
12. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R (2002) Age-specific relevance of usual blood pressure to vascular mortality: A meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 360: 1903–1913.
13. Clarke R, Shippley M, Lewington S, Youngman L, Collins R, et al. (1999) Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. Am J Epidemiol 150: 341–353.
14. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, et al. (2004) Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): Case-control study. Lancet 364: 937–952.
15. Danesh J, Collins R, Peto R (2000) Lipoprotein (a) and coronary heart disease: Meta-analyses of prospective studies. Circulation 102: 1082–1085.
16. Knoblauch H, Bauerfeind A, Toliat M, Becker C, Luganskaja C, et al. (2004) Haplotypes and SNPs in 13 lipid-relevant genes explain most of the genetic variance in high-density lipoprotein and low-density lipoprotein cholesterol. Hum Mol Genet 13: 993–1004.
17. Gonzalez E, Bamshad M, Sato N, Mummidi S, Dhanda R, et al. (1999) Race-specific HIV-1 disease-modifying effects associated with CCR5 haplotypes. Proc Natl Acad Sci U S A 96: 1204–1209.

18. Shanmugalakshmi S, Pitchappan R (2002) Genetic basis of tuberculosis susceptibility in India. Indian J Pediatr 69: S25–S28.
19. MacDonald K, Fowke K, Kimani J, Dunand VA, Nagelkerke NJ, et al. (2000) Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. J Infect Dis 181: 1581–1589.
20. Kaul R, Kimani J, Nagelkerke NJ, Fonck K, Ngugi EN, et al. (2004) Monthly antibiotic chemoprophylaxis and incidence of sexually transmitted infections and HIV-1 infection in Kenyan sex workers: A randomized controlled trial. JAMA 291: 2555–2562.
21. Nagelkerke NJ, de Vlas SJ, MacDonald KS, Rieder HL (2004) Tuberculosis and sexually transmitted infections. Emerg Infect Dis 10: 2055–2056.
22. World Health Organization [WHO] (2003) International statistical classification of diseases and related health problems 10th rev. Volume 1. Geneva (Switzerland): World Health Organization. Available: http://www3.who.int/icd/vol1htm2003/fr-icd.htm. Accessed 18 October 2005.
23. Banthia J, Dyson T (1999) Smallpox in 19th century India. Popul Dev Rev 24: 649–680.
24. Caselli G (1991) Health transition and cause-specific mortality. In: Schofield R, Reher D, Bideau A, editors. The decline of mortality in Europe. Oxford: Clarendon Press. pp. 42–57
25. Gupta PC, Sankaranarayanan R, Ferlay J (1994) Cancer deaths in India: Is the model-based approach valid? Bull World Health Organ 72: 943–944.
26. Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC (1999) Consensus statement. Global burden of tuberculosis: Estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. JAMA 282: 677–686.
27. Office of the Registrar General (2001) Compendium of India's fertility and mortality indicators 1971–1999. New Delhi: Office of the Registrar General. 172 p.
28. Schlesselman S (1982) Case-control studies: Design, conduct, analysis. New York: Oxford University Press. 354 p.
29. Preston S, Bhat P (1984) New evidence on fertility and mortality trends in India. Popul Dev Rev 10: 481–503.
30. Bhat PN (2003) Completeness of India's sample registration system: An assessment using the general growth balance method. Popul Stud (Camb) 56: 119–134.
31. Registrar General of India (2005) Special fertility and mortality survey, 1998: A report of 1.1 million households. New Delhi: Registrar General. 302 p.
32. Anker M, Black R, Coldham C, Kalter H, Quigley M, et al. (1999) A standard verbal autopsy for investigating causes of death in infants and children. Geneva (Switzerland): World Health Organization. Report Number WHO/CDS/CRS/ISR/99.4. 83 p.
33. Quigley MA, Chandramohan D, Rodrigues LC (1999) Diagnostic accuracy of physician review, expert algorithms and data-derived algorithms in adult verbal autopsies. Int J Epidemiol 28: 1081–1087.
34. Chandramohan D, Maude GH, Rodrigues LC, Hayes RJ (1998) Verbal autopsies for adult deaths: Their development and validation in a multicentre study. Trop Med Int Health 3: 436–446.
35. Kalter HD, Gray RH, Black RE, Gultiano SA (1991) Validation of the diagnosis of childhood morbidity using maternal health interviews. Int J Epidemiol 20: 193–198.
36. Kumar R, Sharma AK, Barik S, Kumar V (1989) Maternal mortality inquiry in a rural community of north India. Int J Gynaecol Obstet 29: 313–319.
37. Gajalakshmi V, Richard P, Santhanakrishnan K, Sivagurunathan B (2002) Verbal autopsy of 48,000 adult deaths attributed to medical causes in Chennai (formerly Madras), India. BMC Public Health 2: 7.
38. Gajalakshmi V, Peto R (2004) Verbal autopsy of 80,000 adult deaths in Tamil Nadu, South India. BMC Public Health 4: 47.
39. Kumar R, Thakur J, Rao M, Singh M, Bhatia P (2005) Validity of verbal autopsy in determining causes of adult deaths. Indian J Public Health. In press.
40. Cunnington AJ, Hormbrey P (2002) Breath analysis to detect recent exposure to carbon monoxide. Postgrad Med J 78: 233–237.
41. Gajalakshmi V, Peto R, Kanaka TS, Jha P (2003) Smoking and mortality from tuberculosis and other diseases in India: Retrospective study of 43000 adult male deaths and 35000 controls. Lancet 362: 507–515.
42. Liu BQ, Peto R, Chen ZM, Boreham J, Wu YP, et al. (1998) Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths. BMJ 317: 1411–1422.
43. Biobank Project (2004) . Biobank Project (2004) Available: http://www.biobank.ac.uk. Accessed 8 May 2005.
44. Kaiser J (2002) Biobanks. Population databases boom, from Iceland to the U.S. Science 298: 1158–1161.
45. Varmus H (2003) Genomic empowerment: The importance of public databases. Nat Genet 35: 3.
46. Steinberg K, Beck J, Nickerson D, Garcia-Closas M, Gallagher M, et al. (2002) DNA banking for epidemiologic studies: A review of current practices. Epidemiology 13: 246–254.
47. Clark S, Youngman LD, Palmer A, Parish S, Peto R, et al. (2003) Stability of plasma analytes after delayed separation of whole blood: Implications for epidemiological studies. Int J Epidemiol 32: 125–130.
48. Solomon SS, Solomon S, Rodriguez II, McGarvey ST, Ganesh AK, et al. (2002) Dried blood spots (DBS): A valuable tool for HIV surveillance in developing/tropical countries. Int J STD AIDS 13: 25–28.
49. Ramakrishnan L, Reddy KS, Jailkhani BL (2001) Measurement of cholesterol and triglycerides in dried serum and the effect of storage. Clin Chem 47: 1113–1115.
50. Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, et al. (2001) Haplotype variation and linkage disequilibrium in 313 human genes. Science 293: 489–493.
51. Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. Lancet 361: 598–604.
52. Devlin B, Roeder K (2000) Genomic control for association studies. Biometrics 55: 997–1004.
53. Williams BG, Gouws E, Boschi-Pinto C, Bryce J, Dye C (2002) Estimates of world-wide distribution of child deaths from acute respiratory infections. Lancet Infect Dis 2: 25–32.

Patient Summary

Background. A few pieces of information are crucial to measure the health status of a population, which permits a sound basis for health policy. One basic question is which diseases people die from. To answer it, one needs to know the numbers and causes of death. The age at which people die is critical to estimate the burden for society and to develop strategies to prevent some of the deaths. In developed countries, doctors determine and certify the causes of most deaths that occur in hospitals. However, the majority of deaths worldwide in developing countries occur at home and without reliable recording of their causes and distribution. The present study takes place in India. About one in six deaths worldwide occurs in India, or about 9.5 million deaths per year. Only about a third of these deaths are registered.

Why Is This Study Being Done? The Indian government recognized the need for better data on the distribution and causes of deaths across the country and, together with an international group of researchers, is undertaking a large study over 16 years (from 1998 to 2014) to collect representative data for the country.

What Are the Researchers Doing? In this article, the researchers who came up with the plan for the survey and will oversee its implementation describe and discuss the details of the project. They will monitor, through regular visits and interviews by surveyors, the health status of nearly 14 million people in 2.4 million representative households from all over India. The surveyors, who will be trained for the job but not have a medical background, will collect information about the health status of all household members and about key risk factors for disease such as smoking, alcohol use, childhood immunization, and indoor air pollution. About one million deaths are expected to occur among these people in the study period. As part of their interviews, the surveyors will conduct "verbal autopsies," that is, ask specific questions about how someone died, to determine the causes of death. Each verbal autopsy will be checked by two independent physicians (and a third if there is disagreement) before the cause of death is recorded. The researchers also describe their plans to extend the survey—they plan to collect simple physical measurements such as blood pressure, height and weight, and blood samples (most likely through a simple spot of blood taken from the finger) from the participants. These biological samples contain information about the genetic makeup of the individuals and, potentially, on other markers that might be connected to the cause of death, such as exposure to environmental toxins, the presence of viral or bacterial infections, and others. This information will help them to understand what caused the diseases that the people died from.

What Does This Mean? This study will provide the most accurate picture of causes of childhood and adult death in India, as well as document key risk factors. By studying diseases that are common in one part of India but not in another, new risk factors will be identified and can be used to reduce deaths overall. It will also serve as an example for other countries.

Where Can I Find More Information Online? The following Web sites provide relevant information on this and similar large studies.
Full protocol, field instruments, training manuals, and more details for this study:
http://www.cghr.org/project.htm
Home page of the UK Biobank:
http://www.ukbiobank.ac.uk/
Kadoorie study of chronic disease in China:
http://www.ctsu.ox.ac.uk/~kadoorie/public/
The World Health Organization's Health Metrics Network:
http://www.who.int/healthmetrics/en/
Indepth Network for Continuous Demographic Evaluation of Populations and Their Health in Developing Countries:
http://www.indepth-network.net