Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Jha P, Jacob B, Gajalakshmi V, et al. A nationally representative case–control study of smoking and death in India. N Engl J Med 2008;358:1137-47. DOI: 10.1056/NEJMsa0707719.

Part I. Additional details of methods

Nationally representative study sample

The Registrar-General of the Government of India has, in every decade since the 1970s, conducted a nationwide Sample Registration System (SRS) in which several thousand small areas are chosen, based on the preceding census, to be randomly representative of the whole population (10). The 1993 SRS sample, which is the subject of the present study, included 1.1 million homes in 6671 areas (4436 rural, 2235 urban: mean 165 homes [and about 1000 persons] per area). For the next decade (1994-2003), a part-time enumerator living in each area recorded all births and deaths in that area every month, and 800 full-time SRS (non-medical) supervisors independently collected the same information every six months (3, 10-11).

Each home in the 1993 SRS sample where a death had been recorded in 2001-3 was to be visited in 2004-5 by an SRS supervisor to collect narrative information about cause of death, smoking, drinking and educational status. This study of the 2001-3 deaths is the first phase of the "Million Death Study", which will eventually study about 0.35 million childhood and adult deaths during 1997-2003 retrospectively and 0.65 million during 2004-2014 prospectively (12). The underlying causes of all deaths were sought by an enhanced version, "RHIME" (13-14), of a verbal autopsy. Verbal autopsy is a structured investigation of the train of events, circumstances, symptoms and signs of illness leading to death through an interview of relatives or associates of the deceased. This particular verbal autopsy technique had initially been developed for adult mortality in the southern Indian state of Tamil Nadu (6, 14). Nearly 900 (mainly non-medical) field interviewers (800 SRS supervisors and 100 re-sampling staff) were trained to record a narrative of the symptoms, signs and key circumstances leading to death, including a brief descriptive narrative in the local language of the main events, and of any relevant treatments. The mean time between the 2001-3 death and the RHIME interview was 2.1 years (range 0.5-3.9 years). A random sample of about 10% of the areas was re-surveyed independently, generally with consistent results, and the RHIME methods, quality control checks and validation results have been published (12-14, 16-17).

Medical assignment of underlying causes of death

Causes of death were assigned independently by two physicians who examined scanned images of the written field reports in the local language (eg, Telugu), using a web-based system (12-13) without access to the history of tobacco or alcohol use. If they did not agree on the three-digit ICD-10 (15) code of the underlying cause, their forms were re-assigned to each other (anonymously) for reconciliation. Continuing disagreements were adjudicated by a senior physician. As of October 2007, 82% of the 74 123 deaths had been coded by at least two physicians and 18% by one. The second coding changed the cause in only 4% of the cases, so we have analyzed all single- or double-coded deaths together. A probable cause (ie, excluding ill-defined causes and "senility" [ICD-10 code R54]) was assigned to 91% of the deaths at ages 20-69 and to 71% of those at older ages.

Definitions of smoking and of cases and controls
The field teams asked a respondent (typically a household member) if the dead person had been a smoker within the last 5 years and, if so, the usual number of bidis and/or cigarettes per day. Cessation of smoking is uncommon in India (5, 18), except when illness leads to quitting. Thus, the key comparisons are of ever versus never smoking within the last 5 years. Questions were also asked about other tobacco smoked, quid chewing, alcohol drinking and years of education. Adult respondents were asked similar questions about themselves.

Potential cases were all adults aged 20 years or more whose death during 2001-3 had been recorded in the earlier SRS field work. Of these enumerated deaths, approximately 12.1% could not be studied, chiefly because of out-migration. We had encoded a total of 36 779 female and 47 591 male deaths by October 31, 2007. Of these, 2685 female and 5338 male deaths were excluded from the analyses as they were from non-medical or from maternal causes, a further 1014 women (3%) and 1187 men (3%) were excluded because of missing information on whether they had smoked in the past 5 years and 11 women and 12 men were excluded because of missing address, yielding 33 069 female and 41 054 male cases.

Potential controls (who were not individually matched to the cases) were the adult respondents who were living in the same house as a person whose death they were describing, irrespective of whether that death qualified as a case, and provided information on whether they had smoked in the past 5 years. Controls were not matched within households to cases. Thus, they do not control for household. However, they were approximately matched within areas (as the total number of deaths in an area determined the number of controls from that area). There were 36 076 women and 44 560 men available as controls, of whom 1219 women (3%) and 1482 men (3%) were excluded due to missing smoking information, leaving 34 857 female and 43 078 male unmatched controls. Of these, only 11 486 (33%) of the females and 4912 (11%) of the males had been married to the person who had died in 2001-3. The age-standardized, sex-specific prevalences of smoking, drinking and education were similar among controls who had and who had not been married to such a person (data not shown).

Statistical methods
Women and men were analyzed separately. Controls are not matched to particular cases, so the control group remains the same when studying different underlying causes of death. Controls are, on average, younger than cases, but the major analyses use logistic regression to adjust for age (five-year age groups), education (no education [defined as no formal education or not completing primary school]; primary or middle; grade 10 or higher; unknown) and alcohol use (yes; no; unknown). For each age range, the number of deaths associated with smoking was calculated by multiplying the overall number of deaths among smokers by 1-1/RR, where RR is the risk ratio (adjusted for age, education and alcohol use). Additional adjustment for other variables did not change the results significantly: among women aged 30-69, the smoker: non-smoker RR for deaths from all medical causes [and its 99% confidence interval] was 2.2 [1.9-2.5] with adjustment for age and alcohol, 2.0 [1.8-2.3] after further adjustment for education and 1.95 [1.73-2.21] after further adjustment for chewing. The comparable smoker: non-

smoker RRs for deaths from all medical causes among men aged 30-69 years were 1.8 [1.7-1.9] with adjustment for age and alcohol, 1.7 [1.6-1.8] after further adjustment for education and 1.66 [1.57-1.74] after further adjustment for chewing. For both sexes, further adjustment for geographic location or religion did not materially alter the RRs.

For each RR, attributable fractions were calculated by the formula P (RR-1)/RR, where P is the prevalence of smoking among cases and RR is the age, education and alcohol use adjusted relative risk. The 99% CIs for these attributable fractions are based only on the variance in the RR, as the contribution from the variance in the proportions of smokers was negligible (data not shown). The terms risk ratio and relative risk are used interchangeably. All reported p-values are 2-sided.

Deaths at ages 30-69 years are presented separately because: (i) deaths at these ages involve substantially more years of life lost than deaths at older ages; (ii) the assignment of underlying causes of death is substantially more reliable at ages 30-69 years than at older ages (12, 14-16) and; (iii) the main effects of smoking on mortality occur after age 25 years (the approximate median age of onset of smoking among men in India; 4).

The SRS enumeration of individual deaths slightly underestimates overall Indian adult mortality rates (19-20) and, in addition, our survey missed about 12% of the enumerated deaths. Thus, to estimate absolute death rates and cumulative mortality risks at ages 30-69 from various groups of underlying causes (unaffected by under-enumeration and by the deaths we missed), we estimate the absolute rates of death from those causes in India as a whole from the World Health Organization (WHO; 2) estimates of age-specific mortality (irrespective of cause), applying to them the age-specific, cause-specific proportions for the major groups of underlying causes of death in our study.

To help estimate absolute hazards, we used WHO estimates of the mortality rates and numbers of deaths, sub-divided by 5-year age groups. Within each age group, our study provides the proportion of deaths from all medical causes (eg, about 90%); the smoker vs non-smoker relative risk for all medical deaths; and the control smoking prevalence. Combination of these with the WHO data yields for each age group: (i) the overall smoker and non-smoker death rates, adjusted for any differences in education, alcohol or non-medical mortality, and (ii) the national number of deaths avoidable if smokers had the same medical mortality as nonsmokers of similar age, education and alcohol use. In any one particular 5-year age group (eg, 45-49 years), the non-smoker death rate from a particular disease category or categories (as above) is obtained by dividing the national death rate from that disease category/categories by 1+p(RR-1), where p is the prevalence of smoking among the controls in that age group and RR is the risk ratio (ever versus never smoker, adjusted for education and alcohol use) in that age group; the smoker death rate is then RR times this. Let d denote the sum of the all-cause mortality rates (per 1000 person-years) in each of the 5-year age groups from the start to the end of a given age range (eg, 30-69 years). The probability of dying from medical and non-medical causes in that age range is then 1-exp(-5d/1000). This yields graphs of all-cause mortality versus age for smokers and nonsmokers. The same formula can be used for mortality from specific causes in the hypothetical absence of other diseases. These are presented in Part II of this Supplementary Appendix for mortality at ages 30-69 from three

major groups of causes of death: heart disease, tuberculosis and respiratory disease. (If cumulative risks over the 40-year age range of 30-69 years were to be presented instead, the results would be similar. For example, a 40-year mortality risk of 20% would correspond to a 40-year cumulative mortality risk of 22%, etc.)

The error bars in figure 3 (and in Part II of this Supplementary Appendix) combine conservatively the uncertainties in the relative risk and in the prevalence of smoking. As an upper bound they combine the upper confidence limits for the RR and for the control prevalence of smoking, and as a lower bound they combined the corresponding lower confidence limits. Although this exaggerates the purely statistical uncertainty, it does so only slightly as the contribution from the variation in the control prevalence of smoking is relatively unimportant.

Part II. Additional Results (Webtables 1-3, Webfigures 1-2):

Webtable 1. Effects of standardization to ages 30-69 years* on characteristics on the subjects according to sex:

Characteristic	WOMEN				MEN			
	Cases n=33 069		Controls n=34 857		Cases n=41 054		Controls n=43 078	
	%	(*)	%	(*)	%	(*)	%	(*)
Age group								
20-29	6.3		18.3		4.5		20.5	
30-39	6.1		24.6		6.8		25.3	
40-49	7.4		23.1		10.9		23.0	
50-59	12.5		17.3		17.7		15.8	
60-69	23.6		11.6		26.2		9.4	
70+	44.2		5.1		33.9		6.0	
Residence								
Rural	81.8	(82.6)	78.1	(78.7)	80.6	(79.7)	83.8	(83.6)
Urban	18.2	(17.4)	21.9	(21.3)	19.4	(20.3)	16.2	(16.4)
Educational level								
None completed	82.8	(80.8)	56.0	(68.9)	56.3	(52.8)	30.5	(37.8)
Primary or middle school	10.1	(11.2)	18.1	(15.3)	22.0	(23.1)	23.4	(24.4)
Secondary school or higher	6.1	(7.0)	24.5	(14.4)	20.5	(22.9)	44.6	(36.2)
Unknown	1.0	(1.0)	1.4	(1.6)	1.2	(1.2)	1.5	(1.7)
Religion								
Hindu	82.4	(81.9)	81.6	(82.3)	82.2	(82.5)	83.1	(82.8)
Muslim	9.9	(10.6)	9.7	(9.1)	9.9	(9.7)	9.8	(9.7)
Others	6.6	(6.5)	7.5	(7.6)	6.5	(6.5)	5.7	(6.3)
Unknown	1.2	(1.0)	1.1	(1.1)	1.4	(1.3)	1.4	(1.2)
Chewing tobacco								
Yes	15.2	(14.8)	7.8	(10.4)	25.2	(26.4)	21.2	(22.9)
No	82.6	(82.9)	89.7	(87.2)	71.1	(69.8)	75.8	(74.2)
Unknown	2.2	(2.3)	2.5	(2.4)	3.7	(3.8)	2.9	(2.9)
Alcohol consumption								
Yes	3.6	(4.7)	2.6	(2.9)	26.9	(33.0)	15.5	(16.4)
No	93.9	(92.8)	95.2	(94.7)	69.5	(63.5)	81.7	(80.7)
Unknown	2.5	(2.5)	2.3	(2.4)	3.6	(3.5)	2.9	(2.9)
Smoking prevalence in various subgroups								
All adults	8.5	(9.3)	3.5	(4.5)	50.5	(55.4)	31.9	(37.0)
- Rural only	8.9	(9.8)	3.8	(4.8)	51.4	(56.4)	33.3	(38.7)
- No education completed	9.2	(10.2)	4.5	(5.3)	52.9	(58.2)	41.1	(43.4)
- Hindu only	8.3	(8.9)	3.5	(4.6)	51.1	(55.9)	32.1	(37.6)
- Tobacco chewers only	12.9	(15.3)	11.0	(10.4)	54.0	(58.2)	37.4	(40.5)
- Alcohol drinkers only	40.1	(41.6)	32.1	(33.3)	74.4	(75.7)	62.4	(65.9)

* The numbers in brackets are prevalences standardized (in 5-year age groups) to the age distribution of cases aged 30-69 years.

Webtable 2. Smoking and death from any medical cause among women and men in India, 2001-3: age-specific numbers of cases and controls, risk ratio (ever/never smoked) and smoking-associated excess deaths

Age range (years)	Medical deaths (% ever smoked)	Living controls (% ever smoked)	RR * (99% CI)	Smoking-associated excess deaths	
				No.	%
		WOMEN			
20-29	2070 (2.9)	6388 (1.8)	1.5 (1.0-2.4)	20	1 (0-2)
30-39	2029 (5.8)	8567 (2.7)	2.1 (1.5-2.8)	61	3 (2-4)
40-49	2437 (8.3)	8052 (3.4)	2.3 (1.8-3.0)	115	5 (4-6)
50-59	4126 (9.9)	6027 (4.1)	2.3 (1.9-2.9)	231	6 (5-7)
60-69	7794 (10.2)	4032 (5.6)	1.9 (1.5-2.3)	376	5 (3-6)
70+	14613 (8.3)	1791 (6.8)	1.3 (0.9-1.7)	281	2 (1-3)
Subtotal 30-69	**16386 (9.3)**	**26678 (4.5 †)**	**2.0 (1.8-2.3) ‡**	**783**	**5 (4-6)**
		MEN			
20-29	1840 (28.5)	8816 (17.8)	1.2 (1.0-1.5)	88	5 (0-10)
30-39	2805 (50.3)	10883 (31.9)	1.5 (1.3-1.7)	470	17 (12-21)
40-49	4458 (57.4)	9920 (36.8)	1.6 (1.5-1.8)	961	22 (19-25)
50-59	7278 (58.8)	6792 (39.6)	1.7 (1.5-1.9)	1762	24 (20-28)
60-69	10749 (53.5)	4066 (36.7)	1.8 (1.5-1.9)	2558	24 (18-25)
70+	13923 (44.6)	2601 (32.7)	1.6 (1.4-1.9)	2328	17 (13-21)
Subtotal 30-69	**25290 (55.4)**	**31661 (37.0 †)**	**1.7 (1.6-1.8) ‡**	**5751**	**23 (21-24)**

* Risk ratio (and 99% confidence interval) adjusted for education and alcohol use

† Prevalence among controls aged 30-69 years standardized to the age distribution of cases

‡ Additional adjustment for tobacco chewing did not materially alter the relative risks at ages 30-69 years in women (RR=2.0) or men (RR=1.7); see Methods

Webtable 3. Relevance of type of smoking and daily amount smoked to mortality from medical causes among women and men aged 30-69 years

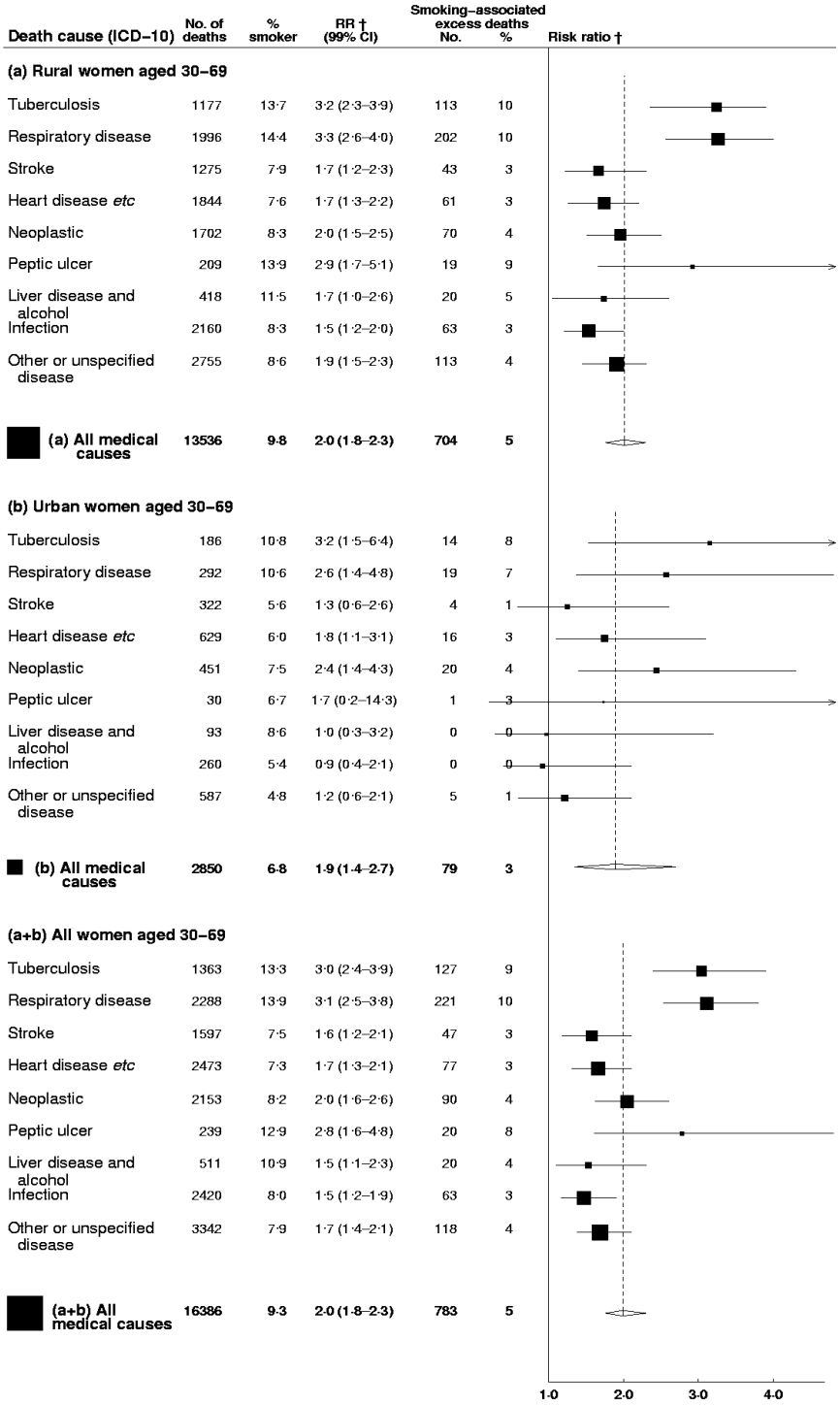
Type of smoking (and mean no./day smoked by controls)	Number of smokers		RR (99% CI) *
	Cases	Controls	
WOMEN			
Bidi only (11.1)	1106	632	2.2 (1.9-2.5)
Cigarette only (6.9)	47	39	2.1 (1.1-4.0)
Other †	370	304	1.6 (1.3-1.9)
Any type	1523	975	**2.0 (1.8-2.3)**
MEN			
Bidi only, by no./day ‡			
1-7 (4.4)	2691	2572	1.3 (1.2-1.4)
8-14 (10.2)	3088	2775	1.5 (1.4-1.6)
15+ (23.9)	3186	1863	2.2 (2.0-2.4)
Unknown	1041	834	1.6 (1.4-1.8)
Any bidi (12.9)	10006	8044	1.6 (1.5-1.7)
Cigarette only, by no./day ‡			
1-7 (4.0)	574	588	1.8 (1.6-1.9)
8+ (13.7)	689	385	2.9 (2.5-3.4)
Any cigarette (8.6)	1263	973	2.2 (2.0-2.5)
Other †	2731	2290	1.6 (1.5-1.8)
Any type	14000	11307	**1.7 (1.6-1.8)**

* Adjusted for age, education, and alcohol use

† Includes those who smoked both bidis and cigarettes, those who smoked anything other than bidis and cigarettes and (the largest contributor) those whose type of smoking was unspecified

‡ Chi-squared test for trend (on 1 degree of freedom) for amount smoked among men was 90.0 ($p<0.0001$) for bidis (nonsmoker, 1-7, 8-14, 15+) and 49.4 ($p<0.0001$) for cigarettes (nonsmoker, 1-7, 8+). Note that nonsmokers with (zero consumption) also contribute to these trends. If, instead, the trend tests were restricted only to the relevant smokers, the Chi-squared statistics would be smaller.

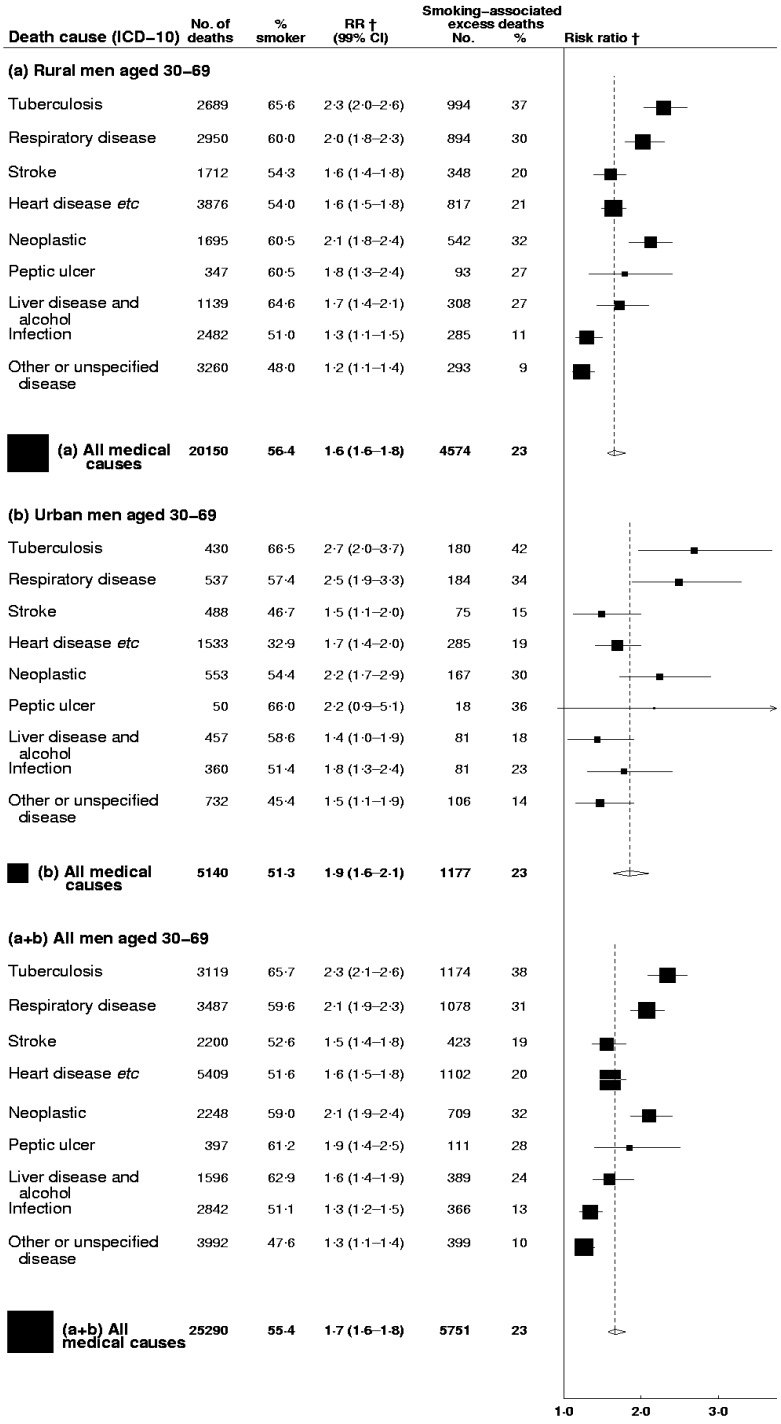
Web figure 1a: Death from medical causes, by smoking: Women aged 30–69 years in rural and urban India, 2001–2003



† Risk ratio (and 99% confidence interval) adjusted for age, education and alcohol use.

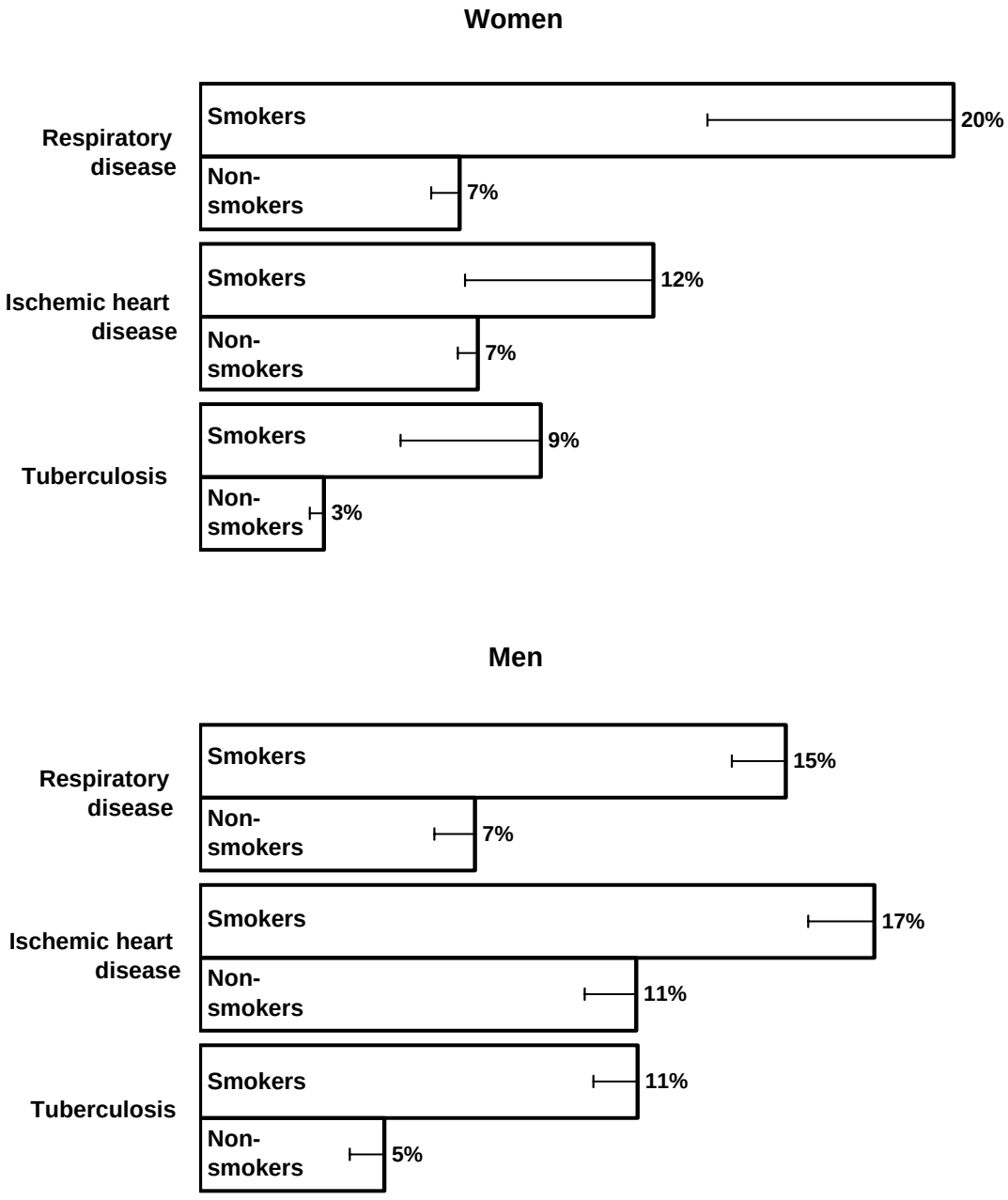
10:32:25 6 November 2007
Not for publication or citation

Web figure 1b: Death from medical causes, by smoking: Men aged 30–69 years in rural and urban India, 2001–2003



† Risk ratio (and 99% confidence interval) adjusted for age, education and alcohol use.

10:33:28 6 November 2007
Not for publication or citation



Web figure 2. Risks of death* at ages 30-69 years from respiratory disease, ischemic heart disease and tuberculosis among Indian women and men

* Each risk is calculated in the hypothetical absence of all other causes of death (so the cumulative mortality rates at ages 30-69 years would be similar to these risks of death, e.g. 22% instead of 20% for respiratory disease in women, etc). For simplicity, only the lower uncertainty limits are shown.

Part III. RGI-CGHR MILLION DEATH STUDY: COLLABORATORS *

Indian Academic Partners (in alphabetical order):

1. Clinical Epidemiology Centre, Trivandarum: KB Leena, KT Shenoy (until 2005)
2. Dept. of Community Medicine, Gujarat Medical College, Ahmedabad: DV Bala, P Seth, KN Trivedi
3. Dept. of Community Medicine, Kolkatta Medical College, Kolkatta: SK Roy
4. Dept. of Community Medicine, Osmania Medical College, Hyderabad: P Bhatia
5. Dept. of Community Medicine, Regional Institute of Medical Sciences, Imphal: L Usharani
6. Dept. of Community Medicine, S.C.B. Medical College, Cuttack, Orissa: B Mohapatra
7. Dept. of Community Medicine, SMS Medical College, Jaipur: AK Bharadwaj, R Gupta
8. Epidemiological Research Center, Chennai: V Gajalakshmi, CV Kanimozhi
9. Gandhi Medical College, Bhopal: RP Dikshit, S Sorangi
10. Healis-Seskarhia Institute of Public Health, Mumbai: PC Gupta, MS Pednekar, S Sreevidya
11. Institute of Health Systems Research, Hyderabad: P Mahapatra (until 2004)
12. St. John's Research Institute, St. John's Academy of Health Sciences, Bangalore: A Kurpad, P Mony, M Vaz, R Jotkar, S Rao-Seshadri, S Shrihari
13. King George's Medical College, Lucknow: S Awasthi
14. Najafgarh Rural Health Training Centre, Ministry of Health, Government of India, New Delhi: N Dhingra, J Sudhir, I Rawat
15. National Institute of Mental Health and Neurosciences, Bangalore: G Gururaj (until 2004)
16. Nizams Institute of Medical Sciences, Hyderabad: V Lakshmi, S Sudha
17. North Eastern Indira Gandhi Regional Institute of Health and Medical Sciences, Shillong: FU Ahmed (until 2005), DK Parida
18. Regional Medical Research Center-ICMR, Bhubaneshwar: AS Karketta, SK Dar
19. School of Preventative Oncology, Patna: DN Sinha
20. School of Public Health, Post Graduate Institute of Medical Education and Research, Chandigarh: N Kaur, R Kumar, JS Thakur
21. Tata Memorial Hospital, Mumbai: RP Dikshit

Lead Partners:

1. Office of the Registrar-General of India (RGI), New Delhi, India: DK Sikri (RGI since Sept 2004), RC Sethi, S Jain, DK Dey, AK Saxena, MK Thapa, JK Banthia (RGI until Sept 2004)
2. Million Death Study Coordinating Centre, Centre for Global Health Research (CGHR), St. Michael's Hospital, University of Toronto, Canada: D Bassani, D Corsi, B Jacob, P Jha, R Jotkar, D Kam, R Kamadod, P Mony, J Moore (until 2005), S Jayakumar (until 2007), S Rao-Seshadri, W Suraweera, S Sgaier, D Thiruchelvam (until 2005), P Vasa (until 2005)

Affiliated Partners:

1. Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), University of Oxford, England: J Boreham, Z Chen, R Collins, J Godwin, R Peto
2. Indian Council of Medical Research, New Delhi, India: NK Ganguly (Director-General until 2008), L Kant, SK Bhattacharya (Acting Director-General from 2008)
3. School of Population Health, The University of Queensland, Australia: AD Lopez, C Rao
4. World Health Organisation, Geneva and New Delhi: T Boerma, T Evans, A Fric, S Khanum, CD Mathers, K Shibuya, N Singh, DN Sinha

* This list is an update of that published in Jha P, et al, PLoS Med 2006; 3: e18. **Disclaimer:** The results of this paper do not necessarily represent the official views of any of the Government of India, the Office of the Registrar General, the study partners or the study sponsors.