Appendix

Table of contents

Model description 2
- Overview 2
- Sampling details 3
- Hazard calculation 4
- Palm oil tax effects 5
- Food insecurity 8

Appendix Table 1: Population distribution of systolic blood pressure 10

Appendix Table 2: Population distribution of total cholesterol 11

Appendix Table 3: Population distribution of tobacco smoking 12

Appendix Table 4: Diabetes prevalence 13

Appendix Table 5: Coronary heart disease prevalence 14

Appendix Table 6: Cerebrovascular disease prevalence 15

Appendix Table 7: Correlation matrix among risk factors in Appendix Tables 1-6 16

Appendix Table 8: Secular trends in risk factor levels 17

Appendix Table 9: Relative risk estimates 19

Appendix Table 10: Mortality rates 22

Appendix Table 11: Secular trends in mortality rates 24

Appendix Table 12: Secular trends in mortality rates 25

Appendix Figure 1: Validation figures 26

References 28

1 Model description

1.1. Overview

The model used in this study is a discrete-time microsimulation, which projects myocardial infarction and stroke deaths at the level of the individual. The model creates a series of individual life histories for members of each population cohort being studied within India, spanning the decade from 2014 to 2023. Unlike a typical Markov model, the microsimulation approach can capture the impact of interventions on individual risk factor profiles, not just the average population effect of an intervention—allowing for complex relationships among multiple co-morbid risk factors to be incorporated into the experiment. This is important because reducing total cholesterol in an individual with many other co-morbid risk factors (tobacco smoking and hypertension, for example) will have a different impact on ultimate mortality risk than reducing total cholesterol in someone without any other co-morbid risk factors. The modeling method adopted here was previously devised by the Institute for Health Metrics and Evaluation [1].

The simulation begins by generating 10,000 individuals for each of 24 cohorts, where each cohort is defined by age (categorized into 10-year age clusters from 20-29 through 70-79 years old, with exact age distributed within the cluster according to Indian Census demography estimates as of 2011, [2]), gender, and residence location (urban or rural). The choice of 10,000 people per cohort was an arbitrary large number conventionally used in heart disease models [1,3]; we found through repeated replication that this number was large enough to provide stable estimates of outcomes and changes in outcomes to the nearest 100 persons in the overall population. To account for individuals graduating from one cohort to the next, the age of each simulated individual determines his/her age-specific risk of mortality from myocardial infarction and stroke (Appendix Table 10). To account for demographic shifts over the 10 years, individuals enter the youngest cohort (20-29 years old) and leave all cohorts (mortality) at different rates based on their age, gender and urban or rural location; these rates of entry and exit have been calculated and projected into the future by the Indian census [2] and World Health Organization [4], respectively (Appendix Table 11).

Individual-level risk factor profiles are generated for each of the 10,000 individuals in each cohort. The risk factor profiles consist of systolic blood pressure, total cholesterol, tobacco smoking (current), diabetes status, coronary heart disease history and cerebrovascular disease history. The risk profiles are generated by Monte Carlo sampling from the distributions of risk factors specific to each cohort (Appendix Tables 1 through 6) estimated by the World Health Organization [5]. The sampling procedure uses the correlation matrix between the risk factors (Appendix Table 7, [1]). The distributions of each risk factor are updated in each year of the simulation to reflect secular changes in risk (Appendix Table 8) [6].

1.2 Sampling details

To generate each individual's risk factor profile, a random number r is sampled from a normal distribution of mean 0 and standard deviation 1. For each continuous risk factor i (systolic blood pressure, total cholesterol), the individual's risk factor value (i.e., their individual systolic blood pressure in mmHg or total cholesterol in mmol/L) is determined by the following function, which was found to correct for the right-skewed distributions of these risk factors in a prior analysis [1]:

(1) $$x_i = e^{(r\sigma_i + \mu_i)}$$

where x is the continuous risk factor value (e.g., the systolic blood pressure) for the individual for risk factor i, σ is the transformed standard deviation of the risk factor in the individual's cohort that year, and μ is the transformed mean value of the risk factor in the individual's cohort that year. Transformations for the right-skewed nature of the risk factor distributions, derived previously [1], are as follows:

(2) $$\mu_i = \ln(\omega_i^2) - \frac{\ln(\omega_i^2 + \delta_i^2)}{2}$$

and

(3) $$\sigma_i = \ln\sqrt{\ln(\delta_i^2) + e^{2\ln(\omega_i)}} - 2\ln(\omega_i)$$

where ω is the mean and δ is the standard deviation of risk factor i's distribution for the individual's cohort during the specific year of the simulation.

For dichotomous risk factors (tobacco smoking, diabetes, coronary heart disease, and cerebrovascular disease), an individual is assigned to have that risk factor with a probability *r* equal to the prevalence of the risk factor in the individual's cohort that year. To capture dependence among the risk factors (e.g., to capture the fact that individuals with diabetes are also more likely to have high cholesterol), a multivariate normal distribution is sampled from, using the covariance matrix given in Appendix Table 7.

To update the risk factor profiles between years of the simulation, pre-existing conditions (tobacco smoking, diabetes, coronary artery disease, and cerebrovascular disease) are carried over from one year to the next and individuals are tracked over time for consistency (e.g., an individual with high blood pressure will continue to have high blood pressure rather than a blood pressure randomly resampled from the population distribution each year), updating individual profiles for age-related and secular trends (Appendix Table 8). To achieve this consistency between years, a rank variable is recorded to capture the rank of each individual's risk in the cohort (e.g., the person with highest systolic blood pressure has rank #1 in the systolic blood pressure rank list). Then the individual with the highest risk factor value in one year will get the highest value sampled for that risk factor in the next year, and the individual with the second highest risk factor value will get the second highest sample, etc. This technique prevents survival bias during the subsequent mortality calculation described below, as individuals who are high risk are less likely to survive to later years.

1.3 Hazard calculation

An individual's risk of myocardial infarction death, stroke death or other cause of death is calculated each year as a function of the individual's risk profile. The individual's relative hazard λ, the hazard of death from disease *j* in relation to the typical hazard in the cohort that year, is defined by:

(4) $$\lambda_j = e^{\sum_i \beta_i x_i}$$

where β is the log of the relative risk of each disease contributed by each risk factor i (Appendix

Table 9) and x is the value of the risk factor for the individual that year.

To determine individual mortality risk from myocardial infarction, stroke, or other causes of death for a particular year, the population-level cohort- and year-specific mortality rate ρ for each disease j is multiplied by the ratio of the individual's relative hazard λ and the mean relative hazard ψ in that individual's cohort that year for that disease:

(5) $$\kappa_j = \rho_j \frac{\lambda_j}{\psi_j}$$

where κ is the mortality rate for the given disease j (coronary disease, cerebrovascular disease, or other-cause mortality) for the individual that year (Appendix Tables 10 and 11). The probability of an individual's death in a given year is given by the sum of the individual's yearly mortality risk from myocardial infarction, stroke and other causes. The simulations were repeated 10,000 times to estimate mean mortality rates and 95% confidence intervals around each mortality rate estimate. All calculations and simulations were performed in MATLAB version R2012a (Cambridge, The MathWorks, Inc.).

1.4 Palm oil tax effects

To estimate the reduction in total cholesterol in each individual's risk factor profile that might result from a palm oil tax, a standard Quadratic Almost Ideal Demand System (QUAIDS) was employed to estimate own- and cross-price elasticities among the vegetable oils used for domestic consumption [8]. Data from the Indian National Sample Survey (NSS) was used. The NSS is a repeated annual cross-sectional survey that collects household food consumption data from populations in multiple social and economic strata in India. Years 2009 through 2010 of the survey (the most recent data available) reported vegetable oil prices paid and amount consumed based on surveys of 100,855 households, from which per-capita consumption (kilocalories per capita) and price (in 2010 Indian Rupees, adjusted through GDP price deflators) were estimated. The following demand system equations were used to then estimate the own-price elasticity of palm oil and cross-price elasticities for other oils.

The equations specify that the share of expenditures for a given good i in an n-good system is:

(6) $$\omega_i = \left\{ \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln p_j + \beta_i \ln\left[\frac{x}{a(p)}\right] + \frac{\lambda_i}{b(p)} \left[\ln\left(\frac{x}{a(p)}\right) \right]^2 \right\} \Phi + \kappa_i \phi + \tau_i \quad \text{for } i=1,\ldots,\text{k},$$

where ω_i is the share of expenditure associated with the *i*th oil, α_i is is the constant coefficient to be estimated for the *i*th share equation, γ_i and κ_i are slope coefficients to be estimated for the *j*th good in the *i*th share equation, p_j is the price on the *j*th good, β_i is a parameter to be estimated that will enter into the elasticity estimate, x is the total expenditure on oils, τ_i is the error term, and *a, b,* and λ are transformations of price specified below. Φ and ϕ are the univariate standard normal cumulative distribution function and the probability density function, respectively, estimated from equation 7 below. Equation 7 is a standard probit regression to estimate the probability that a household will consume good *i* (to account for censoring and zero consumption):

(7) $$d_{ih} = \theta_0 + \sum_j \theta_{ij} \ln p_j + \theta_x \ln x_h + \sum_k \theta_{nk} n_{kh} + \mu_i,$$

where d_{ih} =1 if the *h*th household consumes the *i*th good and zero otherwise, and *n* are demographic variables (mean per-capita total expenditure as a marker of wealth, season in which the household reported consumption, household size, district of residence, household type that corresponds to major source of income, caste, and land ownership status). The corresponding Φ and ϕ are computed as instruments for the second-stage estimation of equation 6 [9]. The price transformations include the transcendental logarithm function:

(8) $$\ln a(p) = \alpha_0 + \sum_{k=1}^{n} \alpha_k \ln p_k + \frac{1}{2} \sum_{k=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \ln p_k p_j$$

and a Cobb-Douglas price aggregator:

(9) $$b(p) = \prod_{k=1}^{n} p_{k=1}^{\beta_k}$$

and

(10) $$\lambda(p) = \sum_{i=1}^{k} \lambda_i \ln p_i$$

The demand system is constrained by the following restrictions on the parameters to ensure the parameters properly add up and follow homogeneity and Slutsky symmetry impositions:

(11) $$\sum_{i=1}^{k} \alpha_i = 1,$$

(12) $$\sum_{i=1}^{k} \beta_i = 0,$$

(13) $$\sum_{i=1}^{k} \lambda_i = 0$$

(14) $$\sum_{i=1}^{k} \gamma_{ij} = 0$$ for all i (homogeneity), and

(15) $\gamma_{ij} = \gamma_{ji}$ for all i and j (symmetry).

From this demand system, uncompensated price elasticities are estimated as:

(16) $$\varepsilon_{i,j} = \frac{1}{\omega_i}\left\{ \gamma_{ij} - \left[\beta_i + \frac{2\lambda_i}{b(p)}(\ln x - \ln a(p)) \right]\left(\alpha_j + \sum_{k=1}^{n} \gamma_{kj} \ln p_k \right) - \frac{\beta_i \lambda_i (\ln x - \ln a(p))^2}{b(p)} \right\} - \delta_{ij},$$

where δ_{ij} is the Kronecker delta function equal to 1 if $i=j$ (own-price elasticity) and equals zero (cross-elasticity) otherwise. The demand system was estimated in Stata version MP12.1 (StataCorp, College Station, Texas), and the resulting parameters are listed in Appendix Table 12.

Mustard/rapeseed, groundnut and coconut oils were separately catalogued in the NSS, while the fourth oil category is "other oils"; palm oil was extracted from an "other oils" category using FAO estimates of the proportion of domestic oil consumption for food (other than mustard/rapeseed, groundnut and coconut oil) that was made up by palm oil consumption [10]. As detailed in the main text, we simulated two scenarios to deal with the aggregate "other oil" category in NSS that also contains soybean, cottonseed, and sunflowerseed oils. In a "pessimistic" case, cottonseed oil would be the "other oil" substituted for palm oils (with perfect 1.0 elasticity) after accounting for the proportion substituted for mustard/rapeseed, groundnut and coconut oils, given that cottonseed oil has the next highest ratio of saturated fat to polyunsaturated fat per gram. In the "optimistic" case, soybean oil would be the “other oil”, given that it has the lowest ratio of saturated fat to polyunsaturated fat per gram.

The updated Keys formula was used to convert changes in consumption of the various oils into estimated changes in serum cholesterol level [11]. The formula is:

(17) $$\Delta C = 1.35(2\Delta S - \Delta P)$$

where C refers to serum cholesterol, S for change in saturated fat, and P for change in polyunsaturated fat. The equation calculates change in cholesterol in units of mg/100ml, which was converted to mmol/L by dividing by 38.7. The changes in fats were computed from the change in each oil consumed, and the saturated and polyunsaturated fat content per unit of each oil starting from an Indian average total consumption of 2.13 tablespoons of oil per day. The saturated and polyunsaturated fat content of each oil is given by the USDA Nutrient Database for Dietary Studies [12]. Further expressions were not included in the formula to reflect potential changes in dietary cholesterol concentration per kilocalorie or stearic acid content, given the absence of data regarding these properties of the oils, and limited or no added predictive value from including these oil components [11].

1.5 Food insecurity

The risk of food insecurity was estimated using data from the UN Food and Agriculture Organization [10]. The FAO estimates price per ton of the above-mentioned oils, which was converted into consumer price per day based on the Indian consumption patterns described above. The differential prices between palm oil and other oils were multiplied by the changes in consumption to reflect the new “price burden” among the percent of the population who defer palm oil consumption due to the tax. The price burden was inflated by the percent increase in demand for each oil to reflect higher prices from higher demand. This price burden was then compared to the price burden of the 20% palm oil tax among those who remain consuming palm oils, to find the maximum price burden. Using the income elasticity estimate of 0.65 (S.E. = 0.12) among the “very poor” (the change in kilocalorie consumption for each change in income among households less than 75% of the poverty level) [13], the maximum reduction in kilocalories per day from the maximum price burden was estimated to be 13.1 kcal/person/day (95% CI: 9.4 to 16.7 kcal/person/day). The FAO estimates that mean kilocalorie consumption per day in India is 2,323 kcal/person/day, and is approximately normally distributed with a coefficient of variation of 0.25 [10]. The cumulative distribution function (cdf) of the normal distribution of kilocalorie consumption was computed before and after the 13.1 kcal reduction; this cdf estimates the percent of the population that falls below a particular kilocalorie level of consumption. The cdf was calculated at 1780 kcal/person/day, which is defined by the FAO as the threshold for food insecurity.

Appendix Table 1: Population distribution of systolic blood pressure

Age (years)	Male urban		Female urban		Male rural		Female rural	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
20-29	119.3	13.1	111.4	13.8	110.1	13.3	110.2	14.0
30-39	121.9	20.3	114.3	16.3	112.5	19.5	113.1	16.0
40-49	121.0	24.3	118.6	17.1	111.6	22.5	117.4	17.0
50-59	125.2	21.0	123.6	15.2	115.4	20.8	122.4	15.1
60-69	127.2	17.7	128.0	12.1	117.4	17.8	126.6	12.2
70-79	130.5	23.8	129.9	12.6	120.3	22.7	128.5	12.5

Systolic blood pressure is described in mmHg from a population-representative study across Indian districts. Source: [5]. SD: standard deviation. For all Appendix Tables, estimates are given for the year 2013, and for subsequent years the secular trends listed in Appendix Table 8 are applied.

Appendix Table 2: Population distribution of total cholesterol

Age (years)	Male urban		Female urban		Male rural		Female rural	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
20-29	5.45	1.01	5.33	0.96	4.82	0.93	4.70	0.52
30-39	5.51	1.01	5.46	1.00	5.08	0.92	4.89	0.66
40-49	5.67	1.06	5.75	1.12	5.13	0.74	5.10	0.89
50-59	5.80	1.15	5.94	1.15	4.80	0.52	4.96	0.75
60-69	5.81	1.19	5.97	1.09	4.68	0.47	4.81	0.52
70-79	5.81	1.19	5.97	1.09	4.68	0.47	4.81	0.52

Total cholesterol is described in mmol/L from a study across Indian districts. Source: [6]. SD: standard deviation. For all Appendix tables, estimates are given for the year 2013, and for subsequent years the secular trends listed in Appendix Table 8 are applied.

Appendix Table 3: Population distribution of tobacco smoking

Age (years)	Male urban		Female urban		Male rural		Female rural	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
20-29	11.9%	0.9%	0.7%	0.0%	8.7%	1.2%	2.2%	0.1%
30-39	22.7%	2.5%	1.7%	0.1%	15.5%	3.3%	3.1%	0.2%
40-49	32.1%	3.3%	3.3%	0.2%	26.6%	5.3%	4.1%	0.5%
50-59	36.6%	3.4%	3.9%	0.2%	32.9%	5.9%	6.8%	0.6%
60-69	32.6%	2.0%	3.4%	0.2%	24.1%	3.9%	7.3%	0.9%
70-79	32.6%	1.6%	3.4%	0.2%	24.1%	3.8%	7.3%	0.9%

Smoking includes both cigarette and bidi smoking prevalence rates based on 2013 estimates from a population-representative survey of adults in Indian districts. Source: [14]. SD: Standard deviation. For all Appendix tables, estimates are given for the year 2013, and for subsequent years the secular trends listed in Appendix Table 8 are applied.

Appendix Table 4: Diabetes prevalence

Age (years)	Male urban		Female urban		Male rural		Female rural	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
20-29	5.5%	1.1%	1.2%	1.1%	3.8%	0.9%	0.8%	0.9%
30-39	10.6%	1.1%	2.2%	1.1%	4.2%	0.9%	0.9%	0.9%
40-49	14.0%	1.1%	2.9%	1.1%	6.4%	0.9%	1.3%	0.9%
50-59	14.8%	1.1%	3.1%	1.1%	6.0%	0.9%	1.3%	0.9%
60-69	18.2%	1.1%	3.8%	1.1%	15.7%	0.9%	3.3%	0.9%
70-79	18.2%	1.1%	3.8%	1.1%	15.7%	0.9%	3.3%	0.9%

Diabetes prevalence is estimated from a random multistage cross-sectional population survey [15], updated to 2013 using secular trend estimates from a Bayesian analysis of diabetes prevalence trends [16]. SD: standard deviation. For all Appendix tables, estimates are given for the year 2013, and for subsequent years the secular trends listed in Appendix Table 8 are applied.

Appendix Table 5: Coronary heart disease prevalence

Age (years)	Male urban		Female urban		Male rural		Female rural	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
20-29	1.38%	0.82%	1.65%	1.37%	1.56%	0.27%	1.24%	0.28%
30-39	3.01%	1.85%	3.57%	1.80%	1.56%	0.45%	1.34%	0.37%
40-49	6.53%	2.55%	8.56%	2.79%	1.93%	0.20%	3.10%	1.10%
50-59	12.78%	4.91%	13.23%	2.89%	2.89%	0.31%	5.04%	0.07%
60-69	17.37%	6.67%	17.54%	3.83%	7.11%	0.77%	6.74%	0.10%
70-79	17.37%	6.67%	17.54%	3.83%	7.11%	0.77%	6.74%	0.10%

Coronary heart disease prevalence is from a prior WHO meta-analysis of Indian district surveys [9], updated to the year 2013 based on WHO estimates of secular trends [6]. SD: standard deviation. For all Appendix tables, estimates are given for the year 2013, and for subsequent years the secular trends listed in Appendix Table 8 are applied.

Appendix Table 6: Cerebrovascular disease prevalence

Age (years)	Male urban		Female urban		Male rural		Female rural	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
20-29	0.00%	0.00%	0.02%	0.51%	0.00%	0.00%	0.02%	0.51%
30-39	0.00%	0.00%	0.02%	0.02%	0.00%	0.00%	0.02%	0.02%
40-49	0.05%	0.32%	0.05%	0.32%	0.05%	0.32%	0.05%	0.32%
50-59	0.58%	0.42%	0.58%	0.42%	0.58%	0.42%	0.58%	0.42%
60-69	0.88%	0.18%	0.88%	0.18%	0.88%	0.18%	0.88%	0.18%
70-79	0.91%	0.02%	0.91%	0.02%	0.91%	0.02%	0.91%	0.02%

Cerebrovascular disease prevalence is from a prior WHO meta-analysis of Indian district surveys [9], updated to the year 2013 based on WHO estimates of secular trends [6]. SD: standard deviation. For all Appendix tables, estimates are given for the year 2013, and for subsequent years the secular trends listed in Appendix Table 8 are applied.

Appendix Table 7: Correlation matrix among risk factors described in Appendix Tables 1-6

	Systolic blood pressure	Cholesterol	Tobacco exposure	Diabetes	Coronary heart disease	Cerebrovascular disease
Systolic blood pressure	1.000	0.174	-0.096	0.087	0.037	0.045
Cholesterol	0.174	1.000	-0.107	0.098	0.014	0.012
Tobacco exposure	-0.096	-0.107	1.000	-0.034	-0.003	-0.003
Diabetes	0.087	0.098	-0.034	1.000	0.037	0.031
Coronary heart disease	0.037	0.014	-0.003	0.037	1.000	0.200
Cerebrovascular disease	0.045	0.012	-0.003	0.031	0.200	1.000

Correlation coefficients between risk factors in the model, as provided by the Institute for Health Metrics and Evaluation from a prior assessment of risk factors based on population surveys [1].

Appendix Table 8: Secular trends in risk factor levels (% change in prevalence per year). These are relative increases in the prevalence rate (not absolute increases).

Age	Gender	Location	SBP	Chol	Tob exp	DM	IHD	Stroke
20-29	Male	Urban	0.42%	0.21%	-0.01%	0.17%	0.05%	0.00%
30-39	Male	Urban	0.58%	0.07%	-0.02%	0.23%	0.05%	0.00%
40-49	Male	Urban	0.32%	0.04%	0.00%	0.30%	0.03%	-0.27%
50-59	Male	Urban	0.29%	0.06%	0.02%	0.07%	0.04%	-0.01%
60-69	Male	Urban	0.29%	0.07%	0.11%	0.29%	0.05%	0.01%
70-79	Male	Urban	0.29%	0.07%	-0.50%	0.15%	0.04%	-0.03%
20-29	Male	Rural	0.42%	0.21%	-0.01%	0.17%	0.03%	0.00%
30-39	Male	Rural	0.58%	0.07%	-0.02%	0.23%	0.05%	0.00%
40-49	Male	Rural	0.32%	0.04%	0.00%	0.30%	0.02%	-0.27%
50-59	Male	Rural	0.29%	0.06%	0.02%	0.07%	0.02%	-0.01%
60-69	Male	Rural	0.29%	0.07%	0.11%	0.29%	0.01%	0.01%
70-79	Male	Rural	0.29%	0.07%	-0.50%	0.15%	0.03%	-0.03%
20-29	Female	Urban	0.28%	0.25%	-0.15%	1.00%	0.05%	-0.18%
30-39	Female	Urban	0.48%	0.18%	-0.19%	0.37%	0.05%	-0.47%
40-49	Female	Urban	0.25%	0.08%	-0.01%	0.66%	0.02%	-0.11%
50-59	Female	Urban	0.18%	0.21%	0.22%	0.09%	0.01%	0.00%
60-69	Female	Urban	0.20%	0.13%	0.19%	0.16%	0.05%	0.00%
70-79	Female	Urban	0.20%	0.13%	0.19%	0.07%	0.04%	0.05%
20-29	Female	Rural	0.28%	0.25%	-0.15%	1.00%	0.04%	-0.18%
30-39	Female	Rural	0.48%	0.18%	-0.19%	0.37%	0.05%	-0.47%
40-49	Female	Rural	0.25%	0.08%	-0.01%	0.66%	0.03%	-0.11%
50-59	Female	Rural	0.18%	0.21%	0.22%	0.09%	0.04%	0.00%
60-69	Female	Rural	0.20%	0.13%	0.19%	0.16%	0.04%	0.00%
70-79	Female	Rural	0.20%	0.13%	0.19%	0.07%	0.03%	0.05%

Numbers listed describe percentage change in prevalence rates per year in each cohort from the sources described in Appendix Tables 1-7. SBP = systolic blood pressure; Chol = total cholesterol; Tob exp = tobacco exposure; DM = diabetes; CHD = coronary heart disease; Stroke = cerebrovascular disease. Sources: [5,6,14–17].

Appendix Table 9: Relative risk per unit increase in each risk factor

Estimates are from a prior series of meta-analyses using international data [1,18]. A "unit increase" is defined as a 1mmHg increase for systolic blood pressure, a 1mmol/L increase in total cholesterol, and for the dichotomous variables (any kind of tobacco use and diabetes), a unit increase is defined as going from not exposed to exposed (e.g., becoming newly diabetic, or newly starting tobacco smoking).

(A) of coronary heart disease

Age	Gender	Location	SBP	Chol	Tob	DM
20-29	Male	Urban	1.08	3.65	2.43	2.03
30-39	Male	Urban	1.07	3.65	2.43	2.03
40-49	Male	Urban	1.06	2.08	2.43	2.03
50-59	Male	Urban	1.05	1.55	1.84	2.03
60-69	Male	Urban	1.03	1.42	1.70	2.03
70-79	Male	Urban	1.02	1.42	1.38	2.03
20-29	Male	Rural	1.08	3.65	2.43	2.03
30-39	Male	Rural	1.07	3.65	2.43	2.03
40-49	Male	Rural	1.06	2.08	2.43	2.03
50-59	Male	Rural	1.05	1.55	1.84	2.03
60-69	Male	Rural	1.03	1.42	1.70	2.03
70-79	Male	Rural	1.02	1.42	1.38	2.03
20-29	Female	Urban	1.08	3.65	2.18	2.54
30-39	Female	Urban	1.07	3.65	2.18	2.54
40-49	Female	Urban	1.06	2.08	2.18	2.54
50-59	Female	Urban	1.05	1.55	2.12	2.54
60-69	Female	Urban	1.03	1.42	1.70	2.54
70-79	Female	Urban	1.02	1.42	1.31	2.54
20-29	Female	Rural	1.08	3.65	2.18	2.54

30-39	Female	Rural	1.07	3.65	2.18	2.54
40-49	Female	Rural	1.06	2.08	2.18	2.54
50-59	Female	Rural	1.05	1.55	2.12	2.54
60-69	Female	Rural	1.03	1.42	1.70	2.54
70-79	Female	Rural	1.02	1.42	1.31	2.54

SBP = systolic blood pressure; Chol = total cholesterol; Tob = smoking; DM = diabetes.

(B) of cerebrovascular disease

Age	Gender	Location	SBP	Chol	Tob	DM
20-29	Male	Urban	1.10	1.48	2.43	2.00
30-39	Male	Urban	1.09	1.35	2.43	2.00
40-49	Male	Urban	1.08	1.42	2.43	2.00
50-59	Male	Urban	1.07	1.35	1.84	2.00
60-69	Male	Urban	1.05	1.25	1.70	2.00
70-79	Male	Urban	1.03	1.09	1.38	2.00
20-29	Male	Rural	1.10	1.48	2.43	2.00
30-39	Male	Rural	1.09	1.35	2.43	2.00
40-49	Male	Rural	1.08	1.42	2.43	2.00
50-59	Male	Rural	1.07	1.35	1.84	2.00
60-69	Male	Rural	1.05	1.25	1.70	2.00
70-79	Male	Rural	1.03	1.09	1.38	2.00
20-29	Female	Urban	1.10	1.48	2.18	2.04
30-39	Female	Urban	1.09	1.35	2.18	2.04
40-49	Female	Urban	1.08	1.42	2.18	2.04
50-59	Female	Urban	1.07	1.35	2.12	2.04
60-69	Female	Urban	1.05	1.25	1.70	2.04
70-79	Female	Urban	1.03	1.09	1.31	2.04
20-29	Female	Rural	1.10	1.48	2.18	2.04
30-39	Female	Rural	1.09	1.35	2.18	2.04
40-49	Female	Rural	1.08	1.42	2.18	2.04
50-59	Female	Rural	1.07	1.35	2.12	2.04
60-69	Female	Rural	1.05	1.25	1.70	2.04
70-79	Female	Rural	1.03	1.09	1.31	2.04

SBP = systolic blood pressure; Chol = total cholesterol; Tob = smoking; DM = diabetes.

Appendix Table 10: Mortality rates from coronary heart disease, cerebrovascular disease, and other causes (per year). The denominator is total population (data reflect the % of population per year dying from each cause).

Age	Gender	Location	Coronary heart disease	Cerebrovascular disease	Other
20-29	Male	Urban	0.137%	0.115%	0.235%
30-39	Male	Urban	0.137%	0.115%	0.235%
40-49	Male	Urban	0.286%	0.486%	3.071%
50-59	Male	Urban	0.286%	0.486%	3.071%
60-69	Male	Urban	0.435%	0.857%	5.907%
70-79	Male	Urban	0.435%	0.857%	5.907%
20-29	Male	Rural	0.137%	0.115%	0.235%
30-39	Male	Rural	0.137%	0.115%	0.235%
40-49	Male	Rural	0.286%	0.486%	3.071%
50-59	Male	Rural	0.286%	0.486%	3.071%
60-69	Male	Rural	0.435%	0.857%	5.907%
70-79	Male	Rural	0.435%	0.857%	5.907%
20-29	Female	Urban	0.009%	0.154%	0.123%
30-39	Female	Urban	0.009%	0.154%	0.123%
40-49	Female	Urban	0.053%	0.498%	2.540%
50-59	Female	Urban	0.053%	0.498%	2.540%
60-69	Female	Urban	0.097%	0.842%	4.957%
70-79	Female	Urban	0.097%	0.842%	4.957%
20-29	Female	Rural	0.009%	0.154%	0.123%
30-39	Female	Rural	0.009%	0.154%	0.123%
40-49	Female	Rural	0.053%	0.498%	2.540%
50-59	Female	Rural	0.053%	0.498%	2.540%
60-69	Female	Rural	0.097%	0.842%	4.957%
70-79	Female	Rural	0.097%	0.842%	4.957%

Mortality rates are from World Health Organization estimates [4]. For all Appendix tables, estimates are given for the year 2013, and for subsequent years the secular trends listed in Appendix Table 11 are applied.

Appendix Table 11: Secular trends in mortality rates (% change in mortality rate per year)

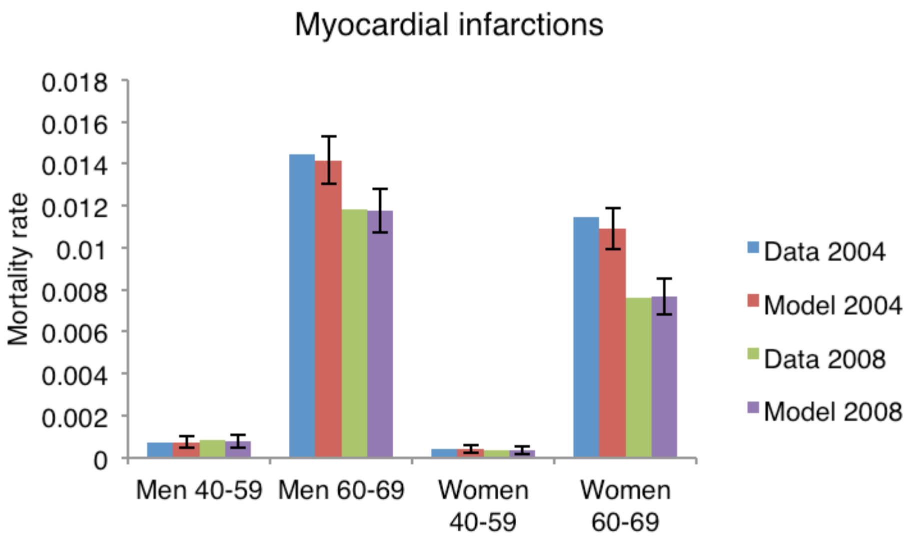
Age	Gender	Location	Coronary heart disease	Cerebrovascular disease	Other
20-29	Male	Urban	2.7%	8.0%	-2.5%
30-39	Male	Urban	2.7%	8.0%	-2.5%
40-49	Male	Urban	-0.9%	4.3%	-0.3%
50-59	Male	Urban	-0.9%	4.3%	-0.3%
60-69	Male	Urban	-4.5%	0.6%	2.0%
70-79	Male	Urban	-4.5%	0.6%	2.0%
20-29	Male	Rural	2.7%	8.0%	-2.5%
30-39	Male	Rural	2.7%	8.0%	-2.5%
40-49	Male	Rural	-0.9%	4.3%	-0.3%
50-59	Male	Rural	-0.9%	4.3%	-0.3%
60-69	Male	Rural	-4.5%	0.6%	2.0%
70-79	Male	Rural	-4.5%	0.6%	2.0%
20-29	Female	Urban	-5.6%	6.4%	-1.0%
30-39	Female	Urban	-5.6%	6.4%	-1.0%
40-49	Female	Urban	-7.0%	12.1%	-3.9%
50-59	Female	Urban	-7.0%	12.1%	-3.9%
60-69	Female	Urban	-8.5%	0.6%	1.9%
70-79	Female	Urban	-8.5%	0.6%	1.9%
20-29	Female	Rural	-5.6%	6.4%	-1.0%
30-39	Female	Rural	-5.6%	6.4%	-1.0%
40-49	Female	Rural	-7.0%	12.1%	-3.9%
50-59	Female	Rural	-7.0%	12.1%	-3.9%
60-69	Female	Rural	-8.5%	0.6%	1.9%
70-79	Female	Rural	-8.5%	0.6%	1.9%

Appendix Table 12: Indian vegetable oil demand elasticities, 2009-2010. The element in row *i*, column *j* of the table contains the elasticity of good *i* with respect to changes in the price of good *j*. Standard errors in parentheses. Asterisks refer to significant elasticities at the $p<0.05$ level.

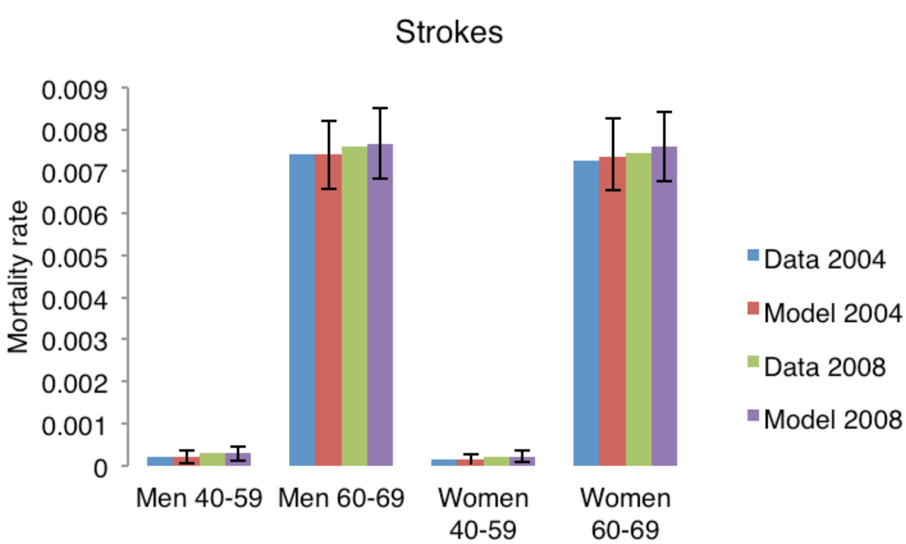
	Uncompensated price elasticities			
Oil	Palm	Mustard/rapeseed	Groundnut	Coconut
Palm	-0.71 (0.14)*	0.09 (0.06)	0.64 (0.21)*	0.60 (0.20)*
Mustard/rapeseed	0.70 (0.21)*	-0.18 (0.02)*	0.13 (0.04)*	0.88 (0.20)*
Groundnut	0.67 (0.15)*	0.02 (0.03)	-0.09 (0.04)*	0.30 (0.18)
Coconut	0.45 (0.48)	0.59 (0.45)	0.26 (0.12)*	-0.31 (0.02)*

Appendix Figure 1. Face validity of the model against historical data from 2004 and 2008, when inputting year 2000 data into the model.

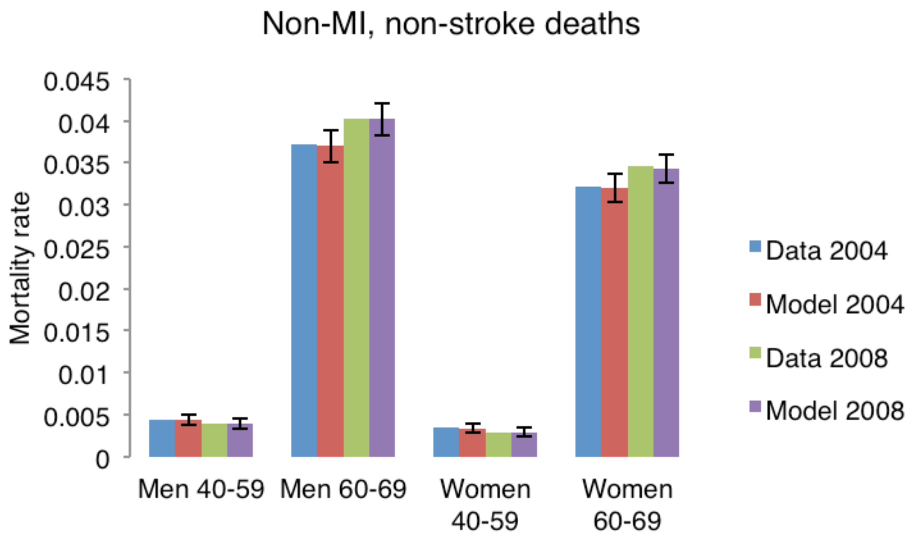
(A)



(B)



(C)



Data are from independent estimates [19].

References

1. Lim SS, Gaziano TA, Gakidou E, Reddy KS, Farzadfar F, et al. (2007) Prevention of cardiovascular disease in high-risk individuals in low-income and middle-income countries: health effects and costs. Lancet 370: 2054–2062.

2. Registrar General & Census Commissioner (2011) Census of India. Delhi: Ministry of Home Affairs.

3. Wang YC, Coxson P, Shen YM, Goldman L, Bibbins-Domingo K (2012) A Penny-per-ounce tax on sugar-sweetened beverages would cut health and cost burdens of diabetes. Health Affairs 31: 199–207.

4. World Health Organization (2008) Disease and injury country estimates. Geneva: WHO.

5. World Health Organization (2011) Study on Global AGEing and Adult Health (SAGE). Geneva: WHO.

6. World Health Organization (2012) WHO Global InfoBase. Geneva: WHO.

7. Andreyeva T, Long MW, Brownell KD (2010) The impact of food prices on consumption: a systematic review of research on the price elasticity of demand for food. American journal of public health 100: 216–222. doi:10.2105/AJPH.2008.151415.

8. Banks J, Blundell R, Lewbel A (1997) Quadratic Engel curves and consumer demand. Review of Economics and Statistics 79: 527–539.

9. Shonkwiler JS, Yen ST (1999) Two-step estimation of a censored system of equations. American Journal of Agricultural Economics 81: 972–982.

10. Food and Agricultural Organization (n.d.) FAOSTAT database. Rome: United Nations.

11. Keys A, Anderson T, Grande F (1965) Serum cholesterol response to change in the diet. Metabolism 14: 776–787.

12. Ahuja J, Montville J, Omolewa-Tomobi G, Heendeniya K, Martin C, et al. (2012) USDA Food and Nutrient Database for Dietary Studies, 5.0. Beltsville, MD: U.S. Department of Agriculture, Agricultural Research Service, Food Surveys Research Group.

13. Kumar P, Kumar A, Parappurathu S, Raju S (2011) Estimation of Demand Elasticity for Food Commodities in India. Agricultural Economics Research Review 24: 1–14.

14. Tobacco Free Initiative (2011) Global Adult Tobacco Survey (GATS) India Report 2009-2010. Geneva: WHO.

15. Sadikot SM, Nigam A, Das S, Bajaj S, Zargar AH, et al. (2004) The burden of diabetes and impaired glucose tolerance in India using the WHO 1999 criteria: prevalence of diabetes in India study (PODIS). Diabetes Res Clin Pract 66: 301–307.

16. Danaei G, Finucane MM, Lu Y, Singh GM, Cowan MJ, et al. (2011) National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants. Lancet 378: 31–40. doi:10.1016/S0140-6736(11)60679-X.

17. Shah B, Kumar N, Menon G, Khurana S, Kumar H (2010) Assessment of burden of non-communicable diseases. Delhi: WHO India.

18. Teo KK, Ounpuu S, Hawken S, Pandey MR, Valentin V, et al. (2006) Tobacco use and risk of myocardial infarction in 52 countries in the INTERHEART study: a case-control study. Lancet 368: 647–658. doi:10.1016/S0140-6736(06)69249-0.

19. World Health Organization (2011) The global burden of disease, Projected deaths for MDG regions, 2004 and 2008. Geneva: WHO.