Supporting Information

Stevens *et al.* 10.1073/pnas.0808927105

SI Text

Sources of Uncertainty in Results. There are several sources of uncertainty that affected our results. First, although environmental risk, demographic, and cause of death statistics in Mexico are of high quality, there was evidence of some death underreporting, especially for children. Adjustment of mortality, even by using established demographic methods, led to additional uncertainty. In addition, causes of death may be inconsistently recorded on death certificates because of factors such as limited medical information at the time of death or practices of certifying physicians. Second, we measured exposure to these risks using metrics such as fuel use or ambient particulate matter concentration rather than personal exposures to particulate matter. This was done for two reasons: The metrics used were consistent with both exposure metrics used in the existing epidemiological studies that quantify relative risks and with the available population-level exposure data. However, if the relationship between personal exposure and these metrics differs between Mexico and the epidemiological studies, results would be subject to error of an unknown direction. Third, although current epidemiological studies have not established the mortality effects of PM pollution below the low levels in some U.S. communities (1), further reductions may provide benefits, if they were technologically feasible. If mean fine particulate matter ($PM_{2.5}$) concentrations were reduced to 2.5 $\mu g/m^3$ in urban areas and mortality benefits also extend to that level, life expectancy would increase 3.0 months (compared with 2.4 months for a reduction to 7.5 $\mu g/m^3$). Fourth, our analysis of urban PM pollution focused on urban areas, because of the lack of data on rural PM concentrations in Mexico. Rural $PM_{2.5}$ concentrations of 10 $\mu g/m^3$, as observed in rural air quality monitoring in the U.S. (2), would increase the life expectancy loss due to this risk factor from 2.4 months to 2.8 months and would slightly attenuate its positive association with community SES. Fifth, we used the same the proportional increase in mortality for all SES groups. Although there is currently little evidence on whether the relative risks for the environmental exposures analyzed in this work vary across population subgroups, multicountry studies of other risk factors have generally indicated that relative risks are relatively invariable across populations (3). If relative risks were 10% larger in the low-SES *municipio*s, the number of deaths attributable to these 3 exposures would increase by 6.5% in that quintile, leading to a slightly larger disparity in attributable mortality across SES quintiles. Finally, the life expectancy estimates in this analysis are based on period life tables, which measure the mortality experience of a synthetic population that is exposed to the current age-specific mortality rates of the Mexican population. Period life expectancy combines data on current mortality rates at different ages but does not measure the expected future mortality of individual cohorts. It also does not, necessarily, reflect the current age structure of the population. Because of past dynamics in fertility and mortality, the population of the synthetic life table in Mexico is, on average, older than the actual population.

***Municipio* Socioeconomic Status Index.** We assigned a composite SES score to each *municipio* based on average wealth and educational attainment in the *municipio*. We quantified household wealth using data on housing quality and asset ownership. Asset-based measurement of household wealth is widely used for low- and middle-income countries for two reasons: (*i*) housing quality and asset ownership are indicators of long-term wealth, which is more relevant to health than income, which may fluctuate from year to year, (*ii*) asset ownership is more robust to common reporting biases that plague monetary income and expenditure data (4–6). The indicators for housing quality and asset ownership were the proportion of *municipio* population who live in a home that has (*i*) electricity, (*ii*) an earth floor, (*iii*) a television, (*iv*) a refrigerator, (*v*) a washing machine, and (*vi*) a computer. The indicators for educational attainment were (*i*) the proportion of population over age 15 who is literate, (*ii*) the proportion of population over age 15 who has completed primary school, and (*iii*) the proportion of population over age 22 who has completed secondary school. All data were obtained in the 2000 and 2005 censuses, and the average value for the 2 censuses was used in the analysis.

Following previous analyses of multifactor SES indices, we used factor analysis to combine these 9 wealth and educational variables into a single index. Factor analysis is used to find a specified number of unobservable underlying factors that can linearly reconstruct observable characteristics. Factor analysis has been used widely, and has been extensively validated, for the construction of wealth or SES indices using asset data collected in DHS surveys (4, 6–8). The underlying assumption of the method is that the factor that explains the largest proportion of variance in the observed characteristics corresponds to SES, because these characteristics are SES indicators. The index used here was adapted from the community deprivation index developed by the National Population Council (9) and used by Gakidou *et al.* (8), and was calculated by using the iterated principal-factor method (10). The results of the factor analysis, showing only the first 3 factors, are presented in Table S1. The first factor is related to all of the observed characteristics in the direction we would expect for an SES index. The second factor may be considered to be related to urbanization or connectivity (this factor would be low in moderately prosperous areas where basic amenities are had by most but not secondary education or computers). The third factor appears to load positively on wealth variables but negatively on educational variables.

Because the first factor reasonably represents community SES, we reanalyzed the data allowing only 1 factor. The final model form (given only 1 factor) is:

$$y_{ij} = z_i b_j + e_{ij} \quad [1]$$

where y_{ij} is the value of the jth community characteristic (e.g., percent literate) in the ith *municipio*, z_i is the common factor (SES index) in the ith *municipio*, b_j is the factor loading for the jth community characteristic, and e_{ij} is the error term. Factor loadings from the reanalysis are shown in Table S1.

To ensure that all analysis units had sufficient sample size to avoid unstable mortality and life expectancy estimates, we then iteratively combined each *municipio* with population <10,000 with its neighboring *municipio* in the same state with the most-similar SES score until each merged *municipio* had a population of at least 10,000 (as described in the main text). This resulted in 1,458 individual or combined "*municipio* units" created from 2,454 *municipios*. We then found the population-weighted average of SES for *municipio* units that contained >1 *municipio*, and classified the units into 5 quintiles of equal population, based on mean 2000–2005 population and estimated SES scores (Fig. S2 and Table S2). Quintile assignment was not sensitive to exclusion of specific characteristics from the SES index or to different computational methods (e.g., maximum-likelihood or principal-factor methods). Analysis of mortality

effects of risk factor exposure was conducted for each merged *municipio* unit.

Adjustment of Mortality Data. Valid and comparable data on cause-specific death rates were needed to calculate mortality attributable to each environmental risk factor. We adjusted vital registration data for the years 2001–2005 for underreporting, age misreporting, and cause of death comparability as described elsewhere (11). In summary, we first corrected the distribution of causes of death. Deaths coded to "ill-defined" ICD categories were redistributed pro-rata to other causes of death following standard Global Burden of Disease (GBD) algorithms (12). Cardiovascular deaths assigned to ill-defined codes were redistributed to ischemic heart disease by using individual-level death records as described elsewhere (13), with a similar approach used to ensure comparability of cause-of-death assignment to diabetes, cardiovascular causes, and other noncommunicable and communicable diseases (14). Because child mortality is low in many parts of Mexico, despite combining small *municipios* as described above, 65% and 54% of *municipio* units had <5 child deaths from diarrhea and lower-respiratory infections, respectively. To separate true variation in cause-specific mortality rates across *municipio* units from stochastic variability due to small number of deaths, we fit a spatial Poisson general additive model (GAM) to observed deaths (R, mgcv package version 1.3). This model spatially smoothes cause-specific death counts in proportion to uncertainty calculated from frequency of observed deaths.

We then adjusted total deaths by age, sex, and state using standard demographic techniques to account for underrecording of deaths in certain ages, misreporting of age on the death certificate, and migration (15, 16). Mortality for children <5 years of age was adjusted by using interpolated mortality values from model life tables that were selected on the basis of estimates of probabilities of dying in childhood derived from United Nations Manual X methods (17). Adult mortality was adjusted at the state level and child mortality for each *municipio* unit. Finally, we applied the adjusted distribution of disease-specific deaths to the corrected total mortality figures to obtain the corrected number of disease-specific deaths.

Calculation of the Mortality Effects of Environmental Risk Factors. The proportional reduction in disease-specific mortality that would be achieved by removing exposure to individual risk factors can be estimated using the population attributable fraction (PAF) relationship in Eq. **2** (18–20).

$$PAF = \frac{\sum_{i=1}^{n} P_i RR_i - 1}{\sum_{i=1}^{n} P_i RR_i} \quad [2]$$

P_i is the proportion of population currently in the ith exposure category, RR_i is the relative risk of disease-specific mortality for the ith exposure category, and n is the number of exposure categories; $n = 4$ for unsafe water and sanitation and $n = 2$ for solid-fuel use and urban PM pollution. In analysis of urban PM pollution, the relative risk for the exposed group was determined by the PM concentration in that *municipio*.

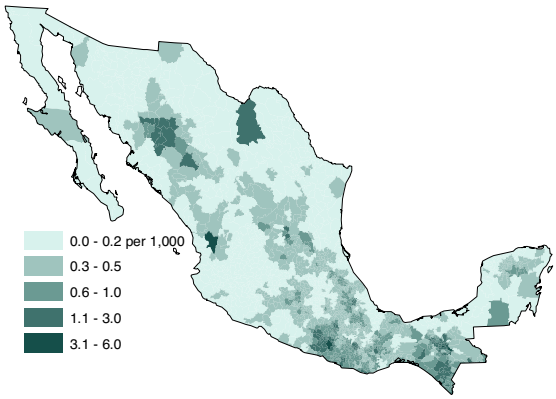
We also calculated the fraction of mortality attributed to the combined effects of these risk factors. Among those people exposed to multiple risk factors, disease-specific deaths may be caused by the simultaneous effects of multiple exposures and hence can be prevented by reducing exposure to any of the risks. For example, some deaths from acute respiratory infections or chronic obstructive pulmonary disease among urban populations who use solid fuels may be prevented by removing exposure to solid fuels or by improving ambient urban air quality through other mechanisms. As a result of multicausality, the PAFs for multiple risk factors overlap in those *municipios* with multiple exposures and cannot be combined by simple addition (21). The combined (joint) PAF that avoids double counting the overlap of multiple risk factors is given by Eq. **3** (22):

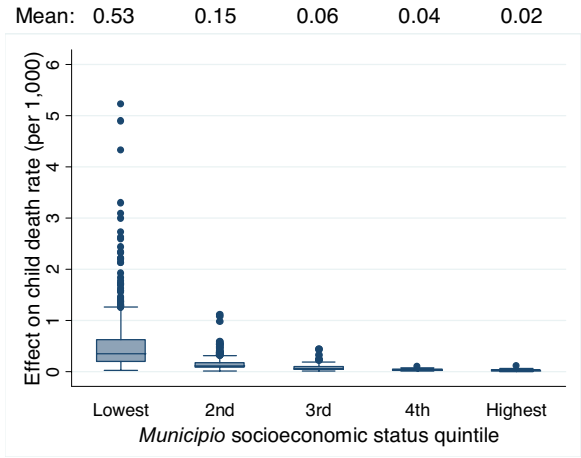
$$PAF = 1 - \prod_{i=1}^{n} (1 - PAF_i) \quad [3]$$

PAF_i is the PAF of individual risk factors and n is the total number of risk factors that affect the same disease outcome. Eq. **3** is based on specific assumptions about the correlation of the exposures to the multiple risks and the interactions of their causal effects as described elsewhere (22). Because all analyses were conducted at the *municipio* level, and because solid-fuel use is low in urban *municipios* (only 2.7% of all urban populations used solid fuels), any deviation from these assumptions would have negligible effects on the results of this analysis. Attributable deaths were calculated as the product of PAF and total disease-specific deaths for each age and sex group in each *municipio*.

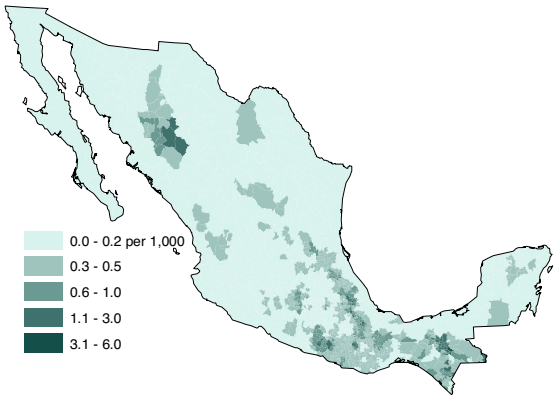
1. Pope CA, *et al.* (2002) Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *J Am Med Assoc* 287:1132–1141.
2. U.S. Environmental Protection Agency (2004) *The Particle Pollution Report: Current Understanding of Air Quality and Emissions Through 2003* (U.S. EPA, Research Triangle Park, NC).
3. Yusuf S, *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.
4. Filmer D, Pritchett LH (2001) Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. *Demography* 38:115–132.
5. Victora CG, *et al.* (2003) Applying an equity lens to child health and mortality: More of the same is not enough. *Lancet* 362:233–241.
6. World Bank (2000) *Round 1 Country Reports on Health, Nutrition, Population and Poverty* (The World Bank, Washington, DC).
7. Booysen FLR (2002) Using demographic and health surveys to measure poverty—An application to South Africa. *J Stud Econ Econometrics* 26:53–70.
8. Gakidou E, *et al.* (2006) Assessing the effect of the 2001–06 Mexican health reform: An interim report card. *Lancet* 368:1920–1935.
9. Consejo Nacional de Población (2002) *Índice de Marginación a Nivel Localidad, 2000* (Consejo Nacional de Población, Mexico, DF).
10. StataCorp (2005) *Stata Statistical Software: Release* 9 (Statacorp LP, College Station, TX).
11. Stevens G, *et al.* (2008) Characterizing the epidemiological transition in Mexico: National and subnational burden of diseases, injuries and risk factors. *PLoS Med* 5:e125.
12. Mathers CD, Lopez AD, Murray CJL (2006) The burden of disease and mortality by condition: Data, methods, and results for 2001. *Global Burden of Disease and Risk Factors*, eds Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL (Oxford Univ Press and The World Bank, Washington, DC), pp 45–240.
13. Murray CJL, Kulkarni SC, Ezzati M (2006) Understanding the coronary heart disease versus total cardiovascular mortality paradox: A method to enhance the comparability of cardiovascular death statistics in the United States. *Circulation* 113:2071–2081.
14. Murray C, *et al.* (2007) Improving the comparability of diabetes mortality statistics in the United States and Mexico. *Diabetes Care* 31:451–458.
15. Brass W (1975) *Methods for Estimating Fertility and Mortality from Limited and Defective Data* (Univ of North Carolina, Chapel Hill, NC).
16. Hill K (1987) Estimating census and death registration completeness. *Asian Pac Pop Forum* 1:8–13 and 23–24.
17. United Nations (1983) *Manual X: Indirect Techniques for Demographic Estimation* (UN, New York).
18. Miettinen OS (1974) Proportion of disease caused or prevented by a given exposure, trait or intervention. *Am J Epidemiol* 99:325–332.
19. Eide GE, Heuch I (2001) Attributable fractions: Fundamental concepts and their visualization. *Stat Methods Med Res* 10:159–193.
20. Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJ (2002) Selected major risk factors and global and regional burden of disease. *Lancet* 360:1347–1360.
21. Rothman KJ (1976) Causes. *Am J Epidemiol* 104:587–592.
22. Ezzati M, *et al.* (2003) Estimates of global and regional potential health gains from reducing multiple major risk factors. *Lancet* 362:271–280.

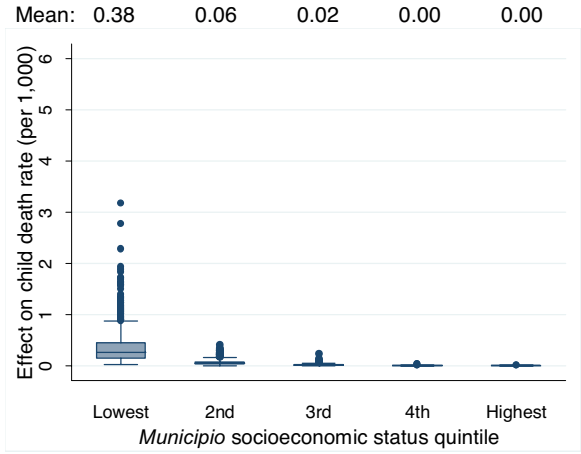
Unsafe water and sanitation



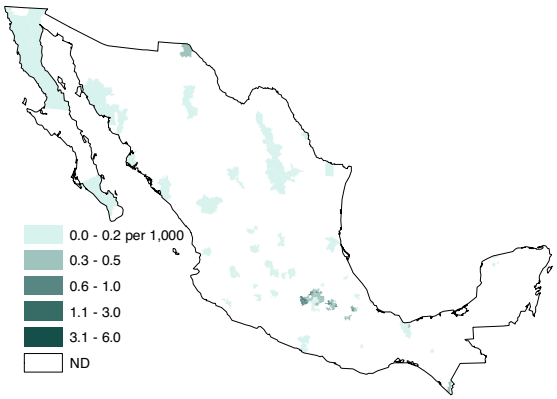


Indoor air pollution from household solid fuel use





Urban particulate matter



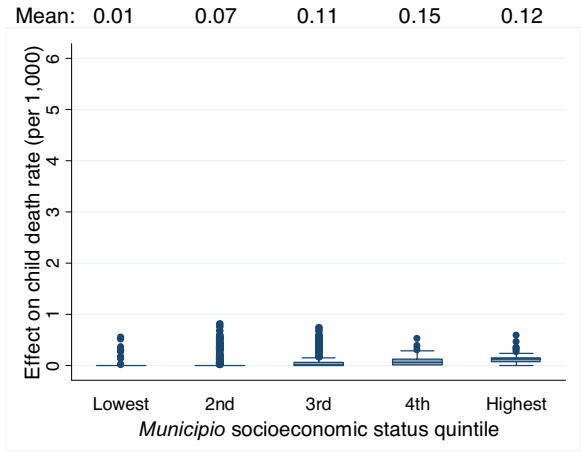


Fig. S1. Child mortality attributable to 3 environmental risk factors by *municipio* of residence and by *municipio* SES quintile.

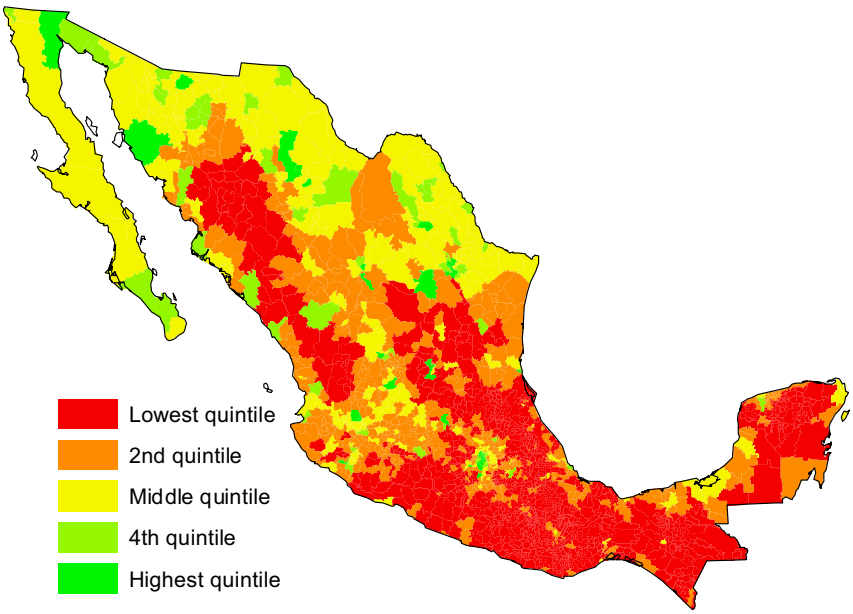


Fig. S2. *Municipios* divided into quintiles of socioeconomic status (SES). Quintiles were constructed based on cumulative population (not count of *municipios*) so there are different numbers of *municipios*, but the same number of people, in each quintile.

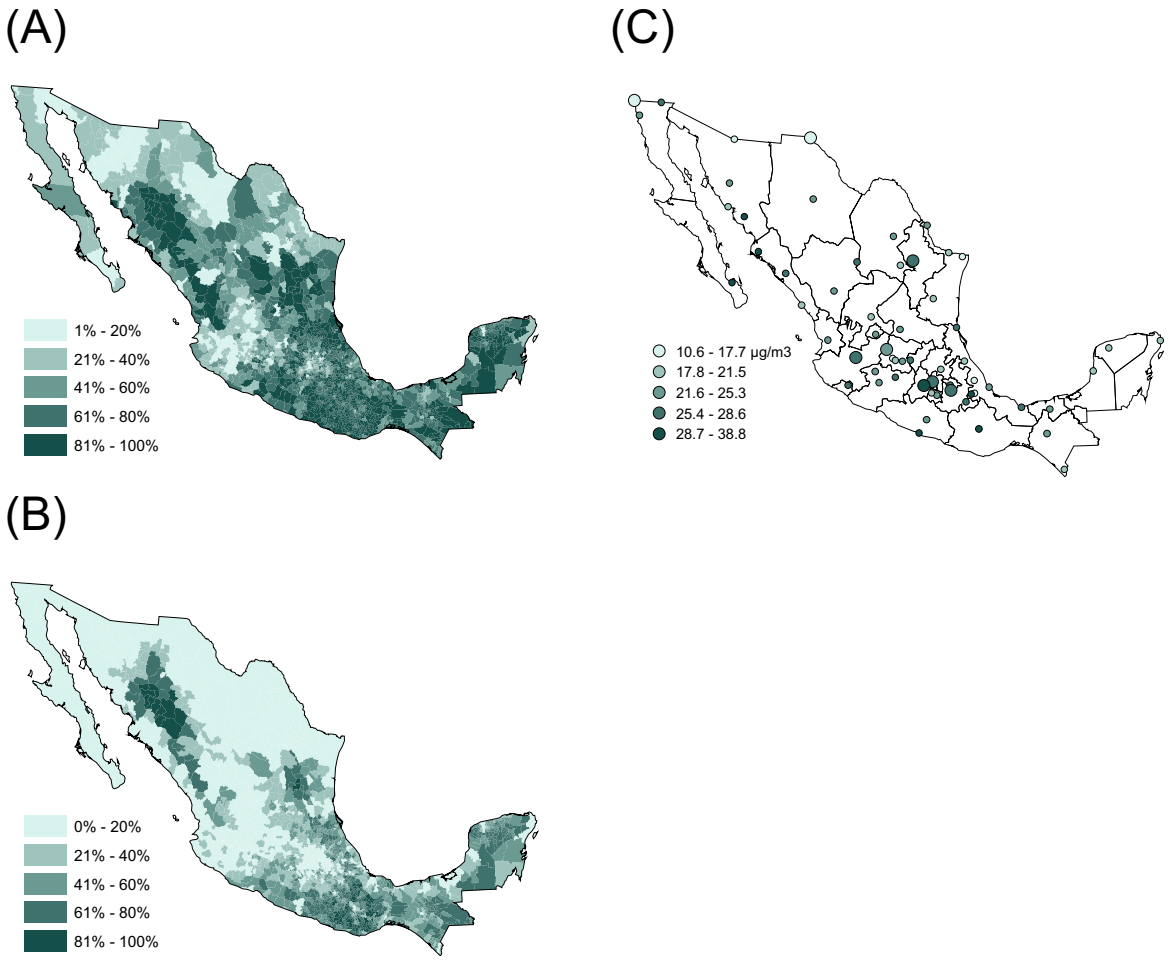


Fig. S3. Exposure to each risk factor, by *municipio*. (*A*) Percentage of population in each *municipio* unit without running water, without a flush toilet, or without either (i.e., in exposure categories ''medium,'' ''high,'' or ''very high'' in Table S4). (*B*) Percentage of population in each *municipio* unit using solid fuels for household cooking. (*C*) Measured or modeled mean annual $PM_{2.5}$ concentrations. Larger symbols indicate cities with measured values, and smaller symbols indicate modeled values.

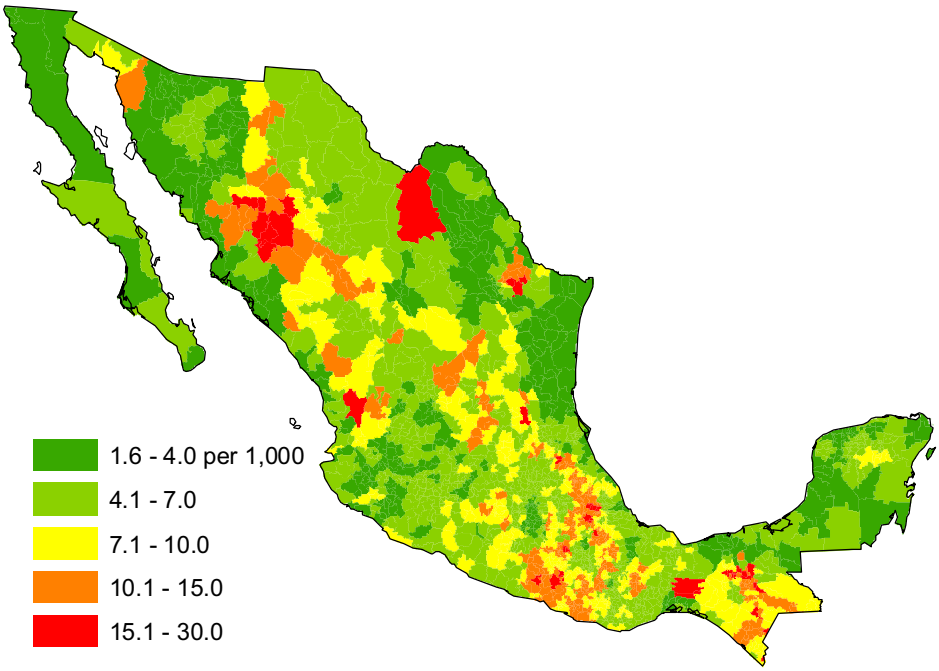


Fig. S4. *Municipio* child mortality rate.

Table S1. Community characteristic factor loadings for the first 3 factors (multiple-factor model) and for the first factor when only 1 factor was allowed (single-factor model), evaluated for 2000 and 2005

Community characteristic	Multiple-factor model			Single-factor model, 1st factor
	1st factor	2nd factor	3rd factor	
Percent > 15 years who is literate	0.89	−0.22	−0.39	0.87
Percent > 15 years who has completed primary school	0.84	−0.19	−0.46	0.82
Percent > 22 years who has completed secondary school	0.65	0.63	−0.04	0.61
Percent with electricity	0.58	−0.23	0.16	0.58
Percent without dirt floor	0.87	−0.20	0.15	0.87
Percent TV ownership	0.90	−0.22	0.24	0.89
Percent refrigerator ownership	0.91	0.01	0.21	0.91
Percent washing machine ownership	0.90	0.01	0.19	0.89
Percent computer ownership	0.74	0.58	−0.05	0.70

The single-factor model was used to determine community SES.

Table S2. Distribution of the community characteristics used in the factor analysis over quintiles of *municipio* SES

SES Quintile	Lowest	2nd	3rd	4th	Highest
Percent computer ownership	3	7	14	21	29
Percent >15 years who has completed primary school	79	89	93	96	97
Percent with electricity	88	97	98	99	100
Percent without dirt floor	61	86	94	96	98
Percent washing machine ownership	20	49	69	74	82
Percent >22 years who has completed secondary school	13	21	31	39	47
Percent >15 years who is literate	76	89	94	96	97
Percent refrigerator ownership	36	67	83	88	93
Percent TV ownership	65	90	95	97	98

Table S3. Annual deaths by disease and age group, average over 2001–2005, Mexico

Risk factor	Disease	International classification of disease (ICD-10) codes	Age			
			<5	5–29	30–59	60+
Unsafe water and sanitation	Diarrhea	A01, A03–A04, A06–A09	3,300	200	400	1,800
Indoor air pollution from household solid-fuel use	Lower-respiratory infections	J10–J18, J20–J22	4,900	—	—	—
	Chronic obstructive pulmonary disease	J40–J44	—	—	1,100	16,900
Urban PM pollution	Acute respiratory infections	J00–J06, J10–J18, J20–J22	5,300	—	—	—
	Lung-cancer mortality	C33–C34	—	—	2,500	2,000
	Cardiopulmonary diseases	I10–I13, I20–I25, I30–I33, I38, I40, I42, I60–I69, J40–J44	—	—	17,600	103,700
	All-cause mortality		55,100	34,700	109,500	268,900

The figures show total deaths for diseases and age groups affected by the three risk factors in this analysis. Only a proportion of these deaths are attributable to risk factors, calculated as described in *SI Text*.

Table S4. Unsafe water- and sanitation-exposure categories, which are ranked based on fecal–oral transmission risk by using data on household access to water and sanitation facilities

Exposure categories	Sanitation	Water
Very high	No facilities	Any water source
High	Flush toilet in the home	Water from a public tap, from a neighbor's home, or from a river, lake, or spring
	Latrine in the home	Water from a tanker truck, from a public tap, carried from a neighbor's home or from a well, river, lake or spring
Medium	Flush toilet in the home	Water from a tanker truck, from a well in the residence, or piped water in the compound but not in the home
	Latrine in the home	Piped water in the home or in the compound
Low	Flush toilet in the home	Piped water in the home

Table S5. Disease-specific relative risks used to estimate attributable mortality

Exposure	Disease	International classification of disease (ICD-10) codes	Relative risk	Source
Medium water and sanitation risk level	Diarrhea	A01, A03–A04, A06–A09	2.8	(1)
High water and sanitation risk level	Diarrhea	A01, A03–A04, A06–A09	3.5	(1)
Very high water and sanitation risk level	Diarrhea	A01, A03–A04, A06–A09	4.4	(1)
Use of solid fuels for home energy needs	Lower-respiratory infections, children <5 years	J10–J18, J20–J22	2.3	(2)
Use of solid fuels for home energy needs	Chronic obstructive pulmonary disease, adults ≥30 years	J40–J44	Men, 1.8; women, 3.2	(2)
10 $\mu g/m^3$ increase in PM_{10}	Acute respiratory infections, children <5 years	J00–J06, J10–J18, J20–J22	1.20	(3)
10 $\mu g/m^3$ increase in $PM_{2.5}$	Lung-cancer mortality, adults ≥30 years	C33–C34	1.08	(4)
10 $\mu g/m^3$ increase in $PM_{2.5}$	Cardiopulmonary diseases, adults ≥30 years	I10–I13, I20–I25, I30–I33, I38, I40, I42, I60–I69, J40–J44	1.06	(4)

1. Prüss Üstün A, Kay D, Fewtrell L, Bartram J (2004) Unsafe water, sanitation, and hygiene. *Comparative Quantification of Health Risks*, eds Ezzati M, Lopez AD, Rodgers A, Murray CJL (World Health Organization, Geneva), Vol 2, pp 1321–1352.
2. Smith KR, Mehta S, Maeusezahl-Feuz M (2004) Indoor air pollution from household use of solid fuels. *Comparative Quantification of Health Risks*, eds Ezzati M, Lopez AD, Rodgers A, Murray CJL (World Health Organization, Geneva), Vol 2, pp 1435–1493.
3. Woodruff TJ, Grillo J, Schoendorf KC (1997) The relationship between selected causes of postneonatal infant mortality and particulate air pollution in the United States. *Environ Health Perspect* 105(6):608–612.
4. Pope, *et al.* (2002) Lung cancer, cardiopulmonary mortality, and long-term exposure of fine particulate air pollution. *J Am Med Assoc* 287(9):1132–1141.