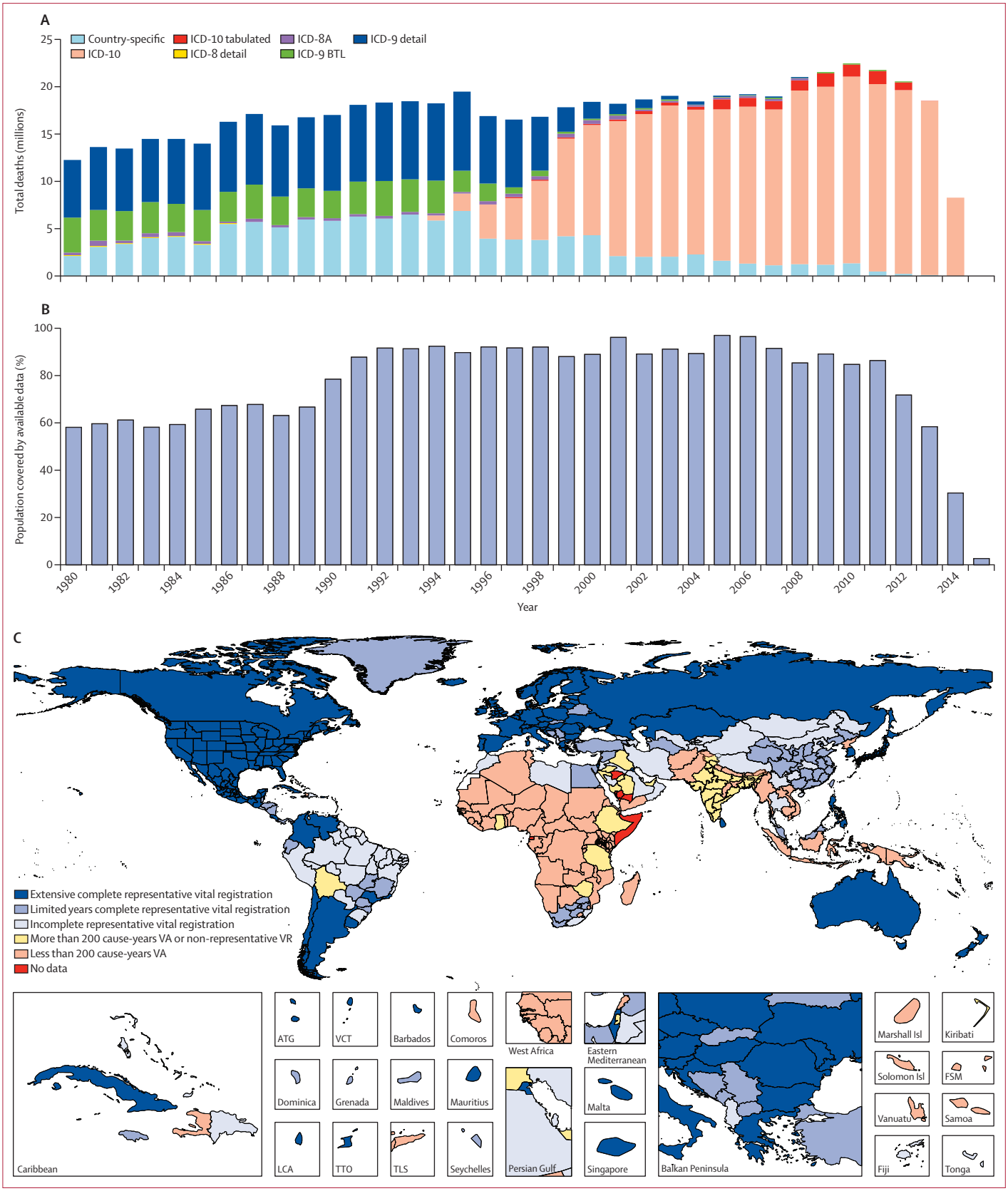
A
Country-specific
ICD-10
ICD-10 tabulated
ICD-8 detail
ICD-8A
ICD-9 BTL
ICD-9 detail
Total deaths (millions)
25
20
15
10
5
0
B
Population covered by available data (%)
100
80
60
40
20
0
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
Year
C
Extensive complete representative vital registration
Limited years complete representative vital registration
Incomplete representative vital registration
More than 200 cause-years VA or non-representative VR
Less than 200 cause-years VA
No data
Caribbean
ATG
VCT
Barbados
Comoros
Dominica
Grenada
Maldives
Mauritius
LCA
TTO
TLS
Seychelles
West Africa
Persian Gulf
Eastern Mediterranean
Malta
Singapore
Balkan Peninsula
Marshall Isl
Kiribati
Solomon Isl
FSM
Vanuatu
Samoa
Fiji
Tonga

through vital registration systems are available from governments and coded to different variants of the ICD including various national ICD variants. Multiple sources were used in addition to vital registration data, including verbal autopsy data, cancer registries, maternal mortality surveillance, census and survey data on maternal death, census and survey data on selected injuries, and police records for some injuries. Figure 2A shows how each type of data were processed to deal with the challenges of different coding schemes, different age group reporting, variation in certification, misclassification of HIV/AIDS deaths, misclassification of maternal HIV/AIDS deaths, and incorporation of population-based cancer registry data. The first and second steps in the cause of death database development were standardisation of multiple data formats to a single GBD standard, then the mapping of each ICD or verbal autopsy variant to the GBD cause map. Figure 7A shows the number of deaths captured for each year in the GBD causes of death database by coding version. In step 3, we split a small subset of data reported in non-GBD-standard age formats into GBD age categories using the global relative age pattern of mortality for each cause as estimated from the pooled data that provide full age detail. In step 4, based on expert judgment, some causes were not allowed for certain age–sex groups, for example, male uterine cancer.

In step 5, deaths assigned to causes that cannot be underlying causes of death (ie, garbage coded) were reassigned to their likely underlying cause of death.[4,7] These redistribution algorithms are based on three approaches. For some garbage codes, such as senility or old age, deaths were proportionately reassigned to all causes that are not garbage codes for a country–age–sex–year. For HIV/AIDS in many countries, deaths from HIV/AIDS have been misclassified as opportunistic infections, tuberculosis, cancer, digestive diseases, and immune deficiencies. In step 6, using methods developed by Birnbaum and colleagues,[41] these deaths were identified and reclassified as HIV/AIDS in select countries with evidence of misclassification. In step 7, data from the China Center for Disease Control and Prevention (CDC) vital registration system were re-weighted to take into account potential selection bias caused by a larger fraction of deaths being captured in hospital than out of hospital in some locations.[14] Step 8 ensured that the process of redistributing garbage codes or identifying misclassified HIV/AIDS deaths would not assign deaths to causes in an age–sex–country–year that violated age–sex or other restrictions.

Step 9 excluded vital registration sources that were less than 50% complete in a given geography from the database, because of the potential for selection bias in highly incomplete sources. Sources estimated to be 50–70% complete were identified as non-representative, which was information that we used in the building of the cause of death statistical model to increase the estimated data variance for these datapoints. All included sources were corrected to be 100% complete by multiplying the cause fraction in a source for a country–age–sex–year by the estimate of all-cause mortality for that country–age–sex–year. Step 10 aggregated causes of death from most to least detailed levels of the GBD hierarchy, ensuring deaths for a given cause were representative of all branches of the hierarchy that fall beneath it. In step 11, deaths due to HIV/AIDS and various types of fatal discontinuities were removed before cause fractions were computed. Because of the very large effects of fatal discontinuities, such as wars and natural disasters in some cases, and the impact of HIV/AIDS in countries with large epidemics, we converted cause fractions to be cause fractions excluding HIV/AIDS and fatal discontinuities in the denominator. Deaths from HIV/AIDS and the fatal discontinuities were added back during the final stages of the modelling process. Because many sources on maternal mortality identify deaths during pregnancy and the post-partum period and not maternal deaths, the separation of HIV/AIDS deaths during pregnancy and HIV/AIDS deaths aggravated by pregnancy was more complicated (methods appendix p 66).

Figure 7B provides information about the fraction of the 519 geographies in the analysis for which cause of death data were available in each year from 1980 to 2015 for any cause, including maternal death and injuries. Data availability by geography–year by cause is shown in

Figure 7: **Availability and quality of cause of death data in the GBD 2015 database**

(A) Total deaths with a WHO-standard death certificate available in the GBD 2015 cause of death database classified by the variant of the International Classification of Diseases used for reporting. Cause of death data have been reported in national variants of ICD-8, ICD-9, and ICD-10 during the interval 1980–2015. Because of lags in reporting of both vital registration data and the release of household survey or census data, the availability of data was much lower for 2014 than for previous years and no data existed for 2015. (B) Percentage of global population covered by cause-specific data in the cause of death database for GBD 2015, 1980–2015; the percentage of available data was calculated by dividing the population of locations covered by available cause-specific data by the total global population. This figure is computed using vital registration, verbal autopsy, maternal, cancer, and injury sources. (C) Overall classification of each GBD subnational level 1 geography by availability and quality of cause of death data for the period 1980 to 2015. Countries have been assigned on the basis of the available time series of data into one of six categories. The figure uses GBD subnational level 1 geographies because subnational level 2 cannot be easily seen on a map. Extensive complete representative vital registration was defined as 25 total years or more of vital registration data with an estimated 95% completeness or above. All geographies that do not meet the threshold for extensive complete representative vital registration are classified as one of the following: limited years of complete representative vital registration, defined as 5 years or more of vital registration data with an estimated 95% completeness or above; incomplete representative vital registration, defined as at least 1 year of vital registration data with an estimated 70% completeness or above; more than 200 cause-years VA or non-representative VR, defined as more than 200 cause-years of verbal autopsy or at least 1 year of vital registration with an estimated 50% completeness or above; less than 200 cause-years of VA; or no data. Cause-years are defined as the number of years for each cause for which data are available. GBD=Global Burden of Disease. ICD=International Classification of Diseases. BTL=basic tabulation list. VA=verbal autopsy. VR=vital registration. ATG=Antigua and Barbuda. VCT=Saint Vincent and the Grenadines. LCA=Saint Lucia. TTO=Trinidad and Tobago. TLS=Timor-Leste. FSM=Federated States of Micronesia.

the methods appendix (pp 318–401). To facilitate understanding of the range of quality and availability of data for each geography, we classified geographies into six categories: extensive complete representative vital registration (vital registration data that are 95% complete and cover more than 25 years); moderate data (vital registration data that are 95% complete but cover fewer than 25 years); incomplete representative vital registration (all other geographies with some representative vital registration data); extensive verbal autopsy and other sources (covering more than 20% of cause-years); limited verbal autopsy or other data (all others with some data available); and no data for any cause (methods appendix pp 691–710). Figure 7C shows this designation for each class of country.

CODEm

Figure 2B shows the analytical flow chart for modelling different causes of death and combining them into internally consistent estimates of cause-specific mortality that sum to all-cause mortality with uncertainty levels. 167 individual causes of death were modelled using CODEm. Developed for GBD 2010,[5] CODEm tests a large number of model specifications, comparing different functional forms and permutations of relevant covariates for each cause of death. Models that met requirements for direction and significance of the regression coefficients were then evaluated for out-of-sample predictive validity through multiple iterations of cross-validation testing. We then combined these models into an ensemble, weighting them such that top performing models (in terms of out-of-sample prediction error on levels and trends) contributed the most to the final prediction. Out-of-sample predictive validity testing was also used to select the psi parameter that determines the number of models and their weight in the final ensemble (figure 8).

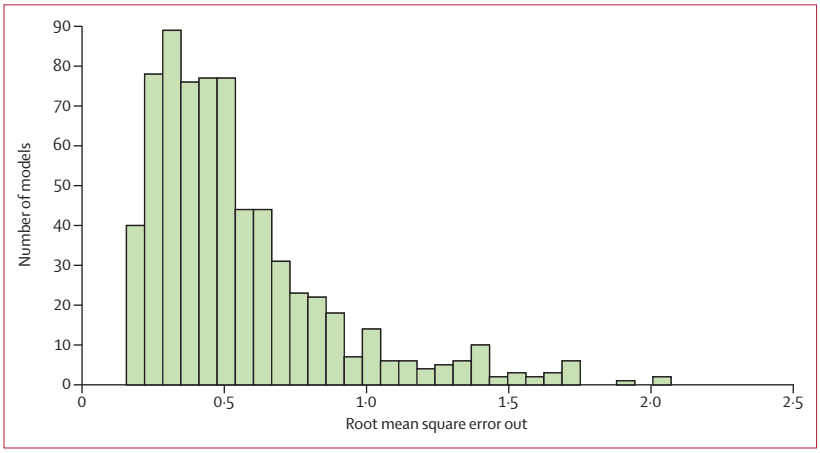

Figure 8: **Distribution of out-of-sample model performance for CODEm models used for GBD 2015**
Model performance was assessed by use of the root mean square error of the ensemble model predictions of the log of the age-specific death rates for a cause assessed with 15% of the data held out from the statistical model building. The figure shows the distribution of root mean square error across the set all models for all causes. Model performance varies substantially across causes. GBD=Global Burden of Disease. CODEm=cause of death ensemble modelling.

For each cause of death, we ran independent CODEm models by sex and for countries with extensive complete vital registration representation and all other countries. We included all datapoints for the other categories of geographies, whereas for countries with extensive complete vital registration representation, we included only datapoints from those countries, so that heterogeneous data from other countries did not inflate the uncertainty interval.

Negative binomial models

For ten causes of death, the number of events are so low, including many zero counts in countries with high income per capita or high educational attainment, that CODEm out-of-sample predictive validity testing was unstable. For these rare causes of death, which included other intestinal infectious diseases, upper respiratory infections, diphtheria, varicella and herpes zoster, malaria, schistosomiasis, cysticercosis, cystic echinococcosis, ascariasis, and iodine deficiency, we used negative binomial regression to develop simple models to predict deaths. More details are available in the methods appendix (pp 185–200; negative binomial models).

Natural history models

For some causes, deaths are rarely recorded in either vital registration data or verbal autopsy data. Partly, this is because of the geographical location of the deaths or because of the potential for systematic bias in vital registration data or verbal autopsy data. For 14 causes, we have developed natural history models in which incidence and case-fatality rates are modelled separately and combined to yield estimates of cause-specific mortality. We developed natural history models for typhoid fever, paratyphoid fever, whooping cough, measles, visceral leishmaniasis, African trypanosomiasis, yellow fever, syphilis (congenital), and acute hepatitis A, B, C, and E. Additionally, for malaria in sub-Saharan Africa, we have used a natural history model based on the incidence estimated by the Malaria Atlas Project and age–sex-specific case-fatality rates estimated from available data. Further details on the development of these natural history models are available in the methods appendix (pp 201–26; natural history models).

Subcause proportion models

For meningitis, maternal disorders, liver cancer, cirrhosis, and chronic kidney disease, we estimated detailed causes for each of these cause groupings by modelling the proportion of the cause grouping (parent cause) due to each of the component causes. We used this approach because the available data on the specific causes can come from sources other than vital registration, such as end-stage renal disease registries, or from too few places to model the death rates directly. For these causes, the parent cause was first estimated with CODEm and the fraction of the parent due to each component cause for

each age–sex–geography–year was generally estimated with DisMod-MR 2.1, a Bayesian meta-regression method developed for the GBD studies.[42,43] Details for each cluster of causes analysed in this way are shown in the methods appendix (pp 233–52; subcause proportion models).

Prevalence-based models

For Alzheimer's disease and other dementias and atrial fibrillation and flutter, there is evidence of marked changes over time in the propensity of individuals who completed death certificates to list these causes as underlying causes of death.[44,45] These changes created increases in the reported death rates. Conversely, prevalence surveys do not show a matching increase in age-specific disease prevalence. Garbage code redistribution algorithms used in the development of the cause of death database have so far not accurately captured this shift over time in the certification of underlying causes of death. For these two causes, we based our estimates on prevalence surveys and estimates of excess mortality based on deaths certified in countries with the greatest proportion of deaths allocated to the correct underlying cause of death in recent years. In both cases, more detail is available in the methods appendix (pp 227–32; prevalence-based models). We developed models for prevalence and excess mortality using DisMod-MR 2.1.

CodCorrect

Depending on the specific data availability and details of individual causes, we adopted different modelling strategies for each cause. We generated a set of underlying cause of death estimates, with uncertainty intervals, that equalled all-cause mortality, with uncertainty intervals, for each age–sex–year–geography and cause and all-cause mortality at the individual draw level.[24] In CodCorrect, for each draw from the posterior distribution of each cause, the sum of cause-specific estimates is rescaled to equal the draw from the all-cause distribution (methods appendix p 285).

Pathogen counterfactual analysis

We used a counterfactual analysis approach to estimate aetiology-specific population attributable fraction for mortality due to lower respiratory infections and diarrhoeal diseases. This approach involved analysing changes in mortality on the basis of the estimated prevalence of each pathogen and relative risk of developing disease given pathogen exposure.

The prevalence of each pathogen in diarrhoeal cases was extracted from a systematic literature review and modelled with DisMod-MR 2.1. The odds ratios of an episode of diarrhoea given exposure to the pathogen were estimated from a reanalysis of the Global Enteric Multicentre Study (GEMS) that used the TaqMan Array Card (TAC), which is based on a quantitative polymerase chain reaction diagnostic (qPCR).[46,47] We attributed mortality to all pathogens, even if the odds ratio was not significant in all age groups. We corrected the estimated prevalence for each pathogen on the basis of conventional laboratory techniques, such as bacterial culture or enzyme-linked immunosorbent assay (ELISA), to be consistent with the new qPCR method. Cholera mortality was estimated by modelling the under-reporting to the WHO cholera case notification system and applying this correction factor to estimate the number of cholera cases and deaths (methods appendix p 281). The incidence and mortality of *Clostridium difficile* was modelled with natural history and incidence data in DisMod-MR 2.1.

We estimated attributable mortality due to respiratory syncytial virus and influenza with a similar approach to that for diarrhoea. We used a counterfactual approach whereby the prevalence in patients with lower respiratory infection was extracted from a systematic literature review and modelled with DisMod-MR 2.1. The odds ratios of lower respiratory infections given pathogen presence were obtained from a meta-analysis by Shi and colleagues.[48] We adjusted the population attributable fraction for lower respiratory infection mortality due to respiratory syncytial virus and influenza for the relative case-fatality rate of viral to bacterial pneumonia episodes by age. *Haemophilus influenzae* type b and pneumococcal pneumonia (*Streptococcus pneumonia*) were estimated with a vaccine probe approach whereby the attributable fraction was calculated as the ratio of vaccine efficacy against non-specific pneumonia to vaccine efficacy against pathogen-specific and serotype-specific pneumonia. Studies that report vaccine efficacy against vaccine-type invasive pneumococcal disease were adjusted for the relative efficacy against vaccine-type clinical pneumococcal pneumonia using a uniform distribution of uncertainty around this ratio.[49,50]

Socio-demographic Index and epidemiological transition analysis

In this Article, we built on GBD 2013[51] concepts by improving the interpretability of sociodemographic status and characterising and describing this relationship in more detail for years of life lost due to premature mortality (YLLs), as well as highlighting changes in age-standardised death rates, population age structure, and YLL rates. We have made two important changes to the GBD 2013 computation. First, we have used only lag-dependent income per capita, average educational attainment in the population over age 15 years, and the total fertility rate. We excluded the mean age of the population because it is directly affected by death rates. Second, we have applied the methods used to compute the Human Development Index to generate an interpretable scale, resulting in the Socio-demographic Index (SDI).[52] The Human Development Index method weights each component equally and rescales each component on a zero-to-one scale with zero being the lowest value observed in the time period 1980 to 2015 and 1 being the highest value observed.

The final composite SDI value is the geometric mean of each of the components. The SDI ranges from 0·060 in Mozambique in 1987 to 0·978 in Washington, DC, USA, in 2015. The correlation of the SDI with the socio-demographic status principal component analysis used in GBD 2013 was 0·982. The very high correlation is because the principal component analysis yields weights that are nearly equal across components. The advantage of the index is that 1 can be interpreted as the level of SDI at which a geography has the highest observed log income per capita and educational attainment and lowest fertility rate. We tested whether alternative lags of the components of SDI would provide a better predictor of outcomes such as life expectancy and age-specific probabilities of death. Using lag distributed income per capita, educational attainment, and the total fertility rate in the current year was the most predictive of these mortality outcomes (methods appendix p 286).

To report on aggregate results, we divided geographies into SDI quintiles in 2015. Quintile cutoffs were based on the entire distribution of geography–years from 1980 to 2015, excluding populations smaller than 1 million. Figure 9 shows a map of the SDI level in 2015 categorised into five groups including subnational geographies. Because SDI includes educational attainment and the total fertility rate, some countries which have very high income, such as Saudi Arabia, are classified in the second quintile of SDI because of lower educational attainment and higher fertility rates.[53]

To capture the average relationships for each age–sex–cause group, we used spline regression of death rates on SDI (methods appendix pp 285–86). To ensure a coherent set of estimated death rates for Levels 1, 2, and 3 in the GBD cause hierarchy for each level of SDI, the Level 2 death rates were rescaled such that for each age–sex–cause bin, the sum of Level 2 death rates equalled the Level 1 death rate. This procedure was repeated for Level 3 and Level 2 causes. These rates were used as the expected death rates by age–sex–cause and SDI. Various summary measures have been computed on the basis of the age–sex–cause-specific predictions based on SDI, including age-standardised death rates, age-standardised YLL rates, and life expectancy at birth.

To further characterise how patterns of crude death rates and death numbers change with SDI, we have computed the average population age structure associated with each SDI level. These population age structures have then been used to estimate how crude death rates and death numbers by cause are expected to change with rising SDI.

Decomposition of changes in global deaths

To analyse the drivers of change in the numbers of deaths by cause or geography, we decomposed change from 2005 to 2015 into three explanatory components: change due to growth of the total population; change in the population structure by age or sex; and change in age-specific, sex-specific, and cause-specific rates. We refer to all changes in age-specific, sex-specific, and cause-specific death rates not explained by demographic change (population growth and ageing) as the epidemiological change. The observed change in the total number of deaths equals the net change of these three components.

Decomposition analyses for 1980 to 2015 and 2000 to 2015 are shown in the results appendix (pp 6–7). The decomposition analysis uses methods developed in demographic research by Das Gupta.[54] As an example, we describe our approach to decomposition for the 2005 to 2015 period. We used counterfactual scenarios to calculate two different sets of numbers for death. In the first scenario, for population growth, the number of deaths in 2015 was the number expected if the total population increased from 2005 as observed, but the age–sex-specific population structure and rates of death were the same in 2015 as in 2005. In the second scenario, for population growth and ageing, the number of deaths in 2015 was the number expected according to the 2015 age–sex-specific population structure, but with the age–sex-specific rates of death held constant to 2005. The difference between the number of deaths observed in 2005 and those estimated for 2015 with the population growth scenario is the change in the number of deaths exclusively from population growth. The difference between the scenario for population growth alone and the scenario for population growth and ageing is the change in the number of deaths exclusively attributable to population ageing.

Attribution of changes in life expectancy to changes in causes of death

When considering the estimated levels and changes in all-cause and cause-specific mortality rates for each geographical area covered by GBD 2015, it is important to understand the relative contribution of changes in mortality due to each cause to the overall changes in life expectancy at birth during the same period. To examine the changes in life expectancy at birth between 2005 and 2015, we have applied the state-of-the-art life expectancy cause-specific decomposition method developed by Beltran-Sanchez, Preston, and Canudas-Romo.[55]

YLL computation

We computed YLLs using the standard GBD methods whereby each death is multiplied by the normative standard life expectancy at each age. The normative standard life expectancy at birth is 86·59 years, which is based on the lowest observed death rates for each 5-year age group in populations larger than 5 million. For GBD 2015, we computed age-standardised mortality rates and YLL rates from the updated world population age standard developed for GBD 2013.[7] Details of the GBD world population age standard are available in the methods appendix (pp 286–287 and 314).