CHAPTER 1
PATTERNS OF DEATH, 1800–2020
Global rates and causes

Romola Davenport

(CC-BY-NC-ND 4.0)
This OA chapter is funded by Wellcome Trust
This chapter provides a brief introduction to mortality patterns over the last two centuries from a global perspective. The period witnessed enormous and ubiquitous rises in life expectancy at birth, from a global average of perhaps 30 years in 1800 to an estimated average of 72.6 years in 2019. Remarkably, these changes took place despite a perhaps seven- to eight-fold increase of the world’s population, from around 1 billion in 1800 to 7.7 billion in 2020, and in the context of unprecedented increases in the speed and frequency of transmission of infectious diseases as a consequence of urbanization, colonization, technological change, and the globalization of trade. This chapter will outline the broad geography and chronology of changes in patterns of mortality and will introduce debates regarding the role of economic growth and those of medicine and states. It concludes with a discussion of recent and future trends.

The wide chronological and geographical reach of the chapter necessitates an immediate caveat. For much of the world, we know very little about mortality rates or levels for national populations before the mid-twentieth century at the very earliest. The United Nations began to publish estimates of mortality rates and data on causes of death for member states in 1950; however, these series are very incomplete and in many cases rely on estimates from rather fragmentary data, even now. For most of sub-Saharan Africa, scholars and public health systems remain reliant on epidemiological surveillance centers and international survey data, and even the largest states, India and China, still use sample registration and surveillance systems in lieu of complete registration of all deaths by age and cause. Given these limitations regarding national patterns and trends, we naturally know even less about health and mortality amongst subgroups within national populations, and about the heterogeneity of experiences in historical societies. This chapter therefore offers a rather impressionistic and selective view of the profound changes in mortality patterns that have occurred in the last two hundred years.

**Ages and stages**

In the broadest terms, what the economist and demographer Richard Easterlin dubbed the ‘Mortality Revolution’ was predominantly a function of enormous reductions in infectious disease mortality. These diseases had their greatest impacts on young children and young adults, and therefore mortality improved first and fastest in these age groups.
Since 1800 death has been ever-increasingly confined to the extremes of life, in early infancy and old age, and infectious diseases have given way to non-communicable diseases, especially cancers, strokes and cardiovascular diseases, and, increasingly, dementia, as more and more people live to the ages where these diseases predominate. Latterly, improvements in the prevention and treatment of non-communicable diseases have improved survival even at the oldest ages, further extending life expectancy. These patterns have, however, unfolded with very different chronologies in different parts of the world, and with some notable reversals.

Figure 1.1 presents this mortality transition in stylized form for the population of England and Wales. Panels a and b represent the situation in 1841. By 1841, as discussed below, the English population was already free of famine, and major mortality crises were extremely rare. Nonetheless, life expectancy at birth was only 42 years, and infectious diseases accounted for nearly half of all deaths. Panel a illustrates the chance of survival to each age, using a hypothetical population of 1,000 people. We have to imagine that these people were all born at the same time and lived their lives with the mortality rates of the English and Welsh population in 1841 (this is not what actually happened to those born in 1841, most of whom were subject to substantial improvements in prevailing mortality rates during their lives). This type of plot is termed a ‘survivorship curve’, and shows the survivors of our hypothetical population, or cohort, at each age. Of the 1,000 babies born in 1841, 138 would have been dead before their first birthday. For those who survived infancy, the risk lessened; however, a quarter of all those born would not survive to age 5. By age 47, half would be dead.

Panel b shows the distribution of these deaths by age, in 1841. Deaths were heavily concentrated in infancy (the first year of life) and early childhood. However, deaths in adulthood were also relatively high at all adult ages. This contrasts markedly with the situation a century later. By 1950 almost everyone born into our hypothetical cohort could expect to live to adulthood, and half would survive to 74. Deaths had become concentrated overwhelmingly in late adulthood. The absolute numbers of deaths were higher in late adulthood than in 1841, not because death rates were higher at these ages in 1950, but because a far higher proportion of the cohort now survived to these ages.

Comparison of panels a and c illustrates what has been termed the ‘rectangularization of mortality’. This is the shift in the pattern of survival from a fairly constant downward curve by age, as in panel a, to a more rectangular form, where most people survive until late age, into their 70s or 80s, and there is a compression or a concentration of mortality into these age groups. Panels e and f illustrate subsequent changes in the mortality patterns of affluent populations since the mid-twentieth century. Broadly, there has been a further ‘rectangularization’, and deaths are even more concentrated at the oldest ages. However there has also been a shift of the whole survival curve to older ages. That is, not only do most people survive to later ages (over half of our hypothetical cohort would live to 85, under the mortality rates of 2016), but the age at which most of the population has died has shifted to older ages.

Figure 1.1 also shows life expectancy measures for the population of England and Wales in 1841, 1950, and 2016. Critically, life expectancy at birth is not the age by which most of the population would be dead, as is sometimes assumed. Rather, it is the average age at death. This means that even in populations with relatively low life expectancies, a substantial proportion of adults lived into their sixties and seventies. Indeed, as
can be seen in Figure 1.1, the average age at death is usually lower than the age to which half the population could expect to survive, because the former figure takes into account how many years people lived. High death rates in childhood lower the average age at death because they contribute only very small values to the sum of ages at death. Life expectancy is calculated using hypothetical populations, as in our example, instead of actual populations because the average age at death in real populations is affected by the age structure of the population. The age structures of populations, and therefore the average age of death, can vary very considerably over time and between populations for

Figure 1.1 Distributions by age of survivors (a, c, e) and deaths (b, d, f) in a synthetic cohort subject to the death rates prevailing in the population of England and Wales in 1841, 1950, and 2016. Source: Human Mortality Database (University of California, Berkeley), and Max Planck Institute for Demographic Research (Germany), www.humanmortality.de (downloaded on January 2, 2020).
reasons other than prevailing mortality rates, including historical changes in fertility, mortality, and age-specific migration rates.

Figure 1.1 depicts the very profound changes in mortality patterns that occurred in the population of England and Wales over the last 170 years. The rest of the chapter considers the processes underlying these trends, and how widespread and uniform these processes were globally.

**Global patterns, an overview**

How ubiquitous has this ‘mortality revolution’ been? Figure 1.2a shows estimates of global average life expectancy since 1800, together with life expectancies for England and Sweden. England and Sweden were forerunners in mortality improvements and in the reporting of mortality statistics. The main pattern is clearly one of enormous improvements in survival. These started first in Europe and amongst the settler populations of North America, Australia and New Zealand in the nineteenth century (or even the mid-eighteenth century, in some cases) and extended to a small number of Latin American and Caribbean countries by the late nineteenth and early twentieth centuries. On the other hand, it is likely that mortality rose in the nineteenth century in some populations that had remained relatively isolated from global disease pools before this period, including Japan, the indigenous populations of North America, Australasia and the Pacific, and parts of tropical Africa.

There is evidence for fairly widespread reductions in mortality globally after c.1920, and by the 1940s mortality declines were apparently underway in a wide variety of settings. The evidence for this is largely indirect and rests on the assumption that the very high rates of population growth witnessed in most of the world by this date (producing an estimated global rate of c.1.5 per cent per annum by 1950) were driven at least in part by falling mortality, as well as rising fertility. These trends may have been interlinked. Where mortality declines were associated with improving adult and maternal health, then this may have resulted in increases in fecundity (the potential to conceive) and fertility.

Gains in life expectancy gathered speed in the second half of the twentieth century, especially in poorer countries. By the early 1950s, when (often rather fragile) statistics became available for many UN member states, the picture was one of enormous disparities in mortality. According to an elegant analysis by Chris Wilson, the world’s population was divided (unequally) between poor countries with low life expectancies (averaging around 35–39 years) and rich or communist Eastern bloc countries with average life expectancies of 65–69 years (Figure 1.3a). That is, the distribution of life expectancies was fairly bimodal (two-peaked), with very little overlap between the two groups. Within the space of 25 years, however, the picture had changed enormously. By the late 1970s only a very small percentage of the world’s population lived in states with average life expectancies below 40 years, and nearly 60 per cent of the global population now lived in states with life expectancies of 60 years or more (Figure 1.3b). By the twenty-first century there had been an enormous and almost ubiquitous convergence of life expectancies in rich and poor countries (Figure 1.3c). The United Nations estimated that the average life expectancy of the global population in 2019 was 72.6 years, a higher figure than that attained by any national population in 1950.
Within this general pattern of improvement and global convergence there have also been notable periods of stagnation and reversal. The most striking reversal was caused by the HIV/AIDS pandemic. The most-HIV affected countries in sub-Saharan Africa were in southern and eastern Africa and included some of those which had made the greatest prior gains in life expectancy (most notably Botswana, Figure 1.2b). Life expectancy is estimated to have fallen by 10 years in Botswana, South Africa, Namibia, and Zambia, and 20 years in Swaziland during the 1990s and early 2000s, before embarking

Figure 1.2 Long-run trends in life expectancy at birth. Data are annual estimates except in the case of England and Wales in panel a, which are five-year averages. Sources: Human Mortality Database (University of California, Berkeley), and Max Planck Institute for Demographic Research (Germany), www.humanmortality.de (downloaded on January 2, 2020); M. Roser, E. Ortiz-Ospina, and H. Ritchie, ‘Life Expectancy’, published online at OurWorldInData.org (2020), retrieved from https://ourworldindata.org/life-expectancy [Online Resource]; E.A. Wrigley, R. S. Davies, J. E. Oeppen and R. S. Schofield, English Population History From Family Reconstitution 1580–1837, (Cambridge: Cambridge University Press) (1997).
Figure 1.3 The distribution of the global population by life expectancy.

Populations were distributed according to the estimated average life expectancy of each national population (both sexes), except in the cases of India and China, where populations and life expectancies for states and provinces were used.

upon a sustained recovery from around 2005, driven by the increasing availability of retroviral therapies, as well as public health initiatives and behavioral changes.\textsuperscript{11}

Another very marked reversal is the fall in life expectancy in some parts of the former Soviet Union after the break-up of the union in 1991 (Figure 1.2b). This was preceded by a prolonged stagnation in life expectancy and infant mortality in the USSR in the decades of the 1960s–1980s. However, the collapse of communism in Europe precipitated marked increases in mortality especially of adult Russian males after 1991, a substantial proportion of which were alcohol-related.\textsuperscript{12} Another notable period of stagnation occurred amongst adult males in many high-income countries in the period c.1950–1980. The timing and extent of this slow-down varied by country and was closely related to life-time smoking patterns. In the UK and the USA, males born in the late 1890s and early 1900s had the highest lifetime smoking rates, perhaps reflecting WWI military policies to pay recruits partly in cigarettes. It was these men who experienced the largest absolute rises in lung cancers and cardiovascular diseases in the third quarter of the twentieth century, against a backdrop of steadily improving death rates for women (See Figure 1.5).\textsuperscript{13}

**Epidemiological transitions**

Can these enormous improvements in survival since 1800 be understood within a unified model of epidemiological change? The most widely used descriptive model of mortality decline is Abdel Omran’s ‘epidemiological transition’ framework.\textsuperscript{14} Omran described a three-stage process that started with the ‘Age of Pestilence and Famine’, which he regarded as the traditional state of large agrarian societies. In this state, life expectancy fluctuated between 20 and 40 years, and population growth was slow or negligible. The onset of secular mortality decline was associated with the ‘Age of Receding Pandemics’, and included reductions in the level and volatility of mortality, and a progressive shift of mortality from infectious to ‘degenerative’ diseases. This was followed by the ‘Age of degenerative and man-made diseases’, as infectious diseases diminished decisively, and most deaths occurred as a result of chronic non-communicable diseases (especially cancers and cardiovascular diseases) in late adulthood. Omran thought that gains in survival would slow or cease in this stage. However, life expectancy has continued to rise inexorably, driven largely by improvements in survival at advanced ages. This has led some scholars to advocate a fourth stage, the ‘Age of delayed degenerative diseases’.\textsuperscript{15}

Omran’s model is widely used; however, it is largely descriptive rather than analytical and may be a poor representation of the historical processes at work in many societies. In this section, we provide a very brief overview of mortality trends in England and compare these with trends elsewhere. England is the only national population for which we have very long-run and robust mortality data for the last half millennium (from the 1540s). It was also a forerunner in terms of mortality declines. We divide the discussion into four periods: (1) c.1550 to 1750, when mortality crises abated and mortality stabilized, albeit at lower levels of life expectancy; (2) 1750–1870, when some of the most lethal infectious diseases including smallpox, typhus, malaria, and cholera were brought under control in the English population; (3) 1870–1940, when an impressive range of improvements in sanitation and hygiene, preventative medicine, and living standards profoundly reduced mortality from childhood diseases and tuberculosis in European and neo-European populations; and (4) the antibiotic era post-1940, when medical
technologies including antibiotics and a range of vaccines acted in synergy with older strategies to confine infectious disease mortality largely to the oldest ages in developed countries. We conclude with a discussion of potential future trends.

**Receding pandemics (1550–1750)**

England was very precocious in escaping famine and major mortality crises. The last nation-wide subsistence crisis in England occurred in the 1590s, and mortality crises diminished progressively following the disappearance of plague after the 1660s. However, contrary to Omran’s model, life expectancy actually fell as food security improved and epidemics subsided (Figure 1.2a). The most plausible explanation for this worsening of mortality in the period 1650–1750 is that the urbanization and integration of the English economy that occurred in this period was accompanied by an epidemiological integration, that increased the circulation of infectious diseases. That is, the development of a functioning national grain market and improved transport infrastructure, which helped to reduce the impact of harvest failures, also promoted the spread of infections. Towns were also associated with high mortality, because they acted as hubs for disease importation and because poor sanitary conditions and dense populations facilitated the spread of gastrointestinal and airborne infections. As disease exposure increased, many of the diseases that had caused rare but large epidemics in the past became routine infections of childhood. This was the case for diseases that conferred lifelong immunity on survivors. Thus in historical societies, urbanization, trade, and economic growth tended to raise mortality rates and to concentrate some infectious diseases in childhood.

**Receding famine and infectious diseases (1750–1870)**

This process of worsening mortality was reversed after around 1750, in England and in other populations of north-western Europe. The reasons for this reversal remain obscure. Most puzzling is the fact that the largest improvements occurred in urban populations. Before the late eighteenth century, cities in Europe appear to have been so lethal as to require a constant flow of immigrants simply to maintain their populations. In early eighteenth-century London, a third of infants died in the first year of life, and only about 60 per cent of those born in London survived to age 15. However, after 1750 urban mortality rates improved very markedly, and cities became capable of maintaining their own populations through natural increase, although mortality remained substantially higher than in rural areas until the twentieth century. As Simon Kuznets argued, these early improvements in urban mortality rates were an essential pre-requisite for the unprecedented rates of urbanization that followed. With the death rates prevailing in early modern cities, modern levels of urbanization would have caused the decline of national populations.

Famine also receded from western Europe in the late eighteenth century (notwithstanding the Great Famine in Ireland, 1845–1849, and the Finnish famine of 1866–1868). In the twentieth century, famine reappeared briefly in western Europe in association with the extreme stresses of the two world wars. However at the global level, the twentieth century witnessed some of the largest famines ever, including the Great Leap Forward famine in China in 1959–1961 that may have killed 20–30 million
people. Nonetheless, by the late twentieth century, the global impact of famine had dwindled enormously, and even the most severe famines of recent decades (including Somalia in 1991–1992 and North Korea in 1995–2000) have been small by historical standards.

There is lively debate about the causes of famines and whether famines are primarily a function of declines in food availability or of institutional or market failures. Amartya Sen famously argued that modern famines represented a ‘failure of exchange entitlements’, rather than an absolute shortfall in food supplies. That is, certain groups lose their capacity to exchange their primary entitlement (for example, their labor) for food. This may occur, for example, because the high price of food or other factors has reduced demand for the products or services which they supply. The price of food may also bear an ambiguous relationship to actual food availability. It may be driven up by demand elsewhere or by hoarding or requisitioning in the absence of any shortfall in production. Alternatively, as Malthus noted, prices may rise only modestly despite starvation, because those starving are too poor to enter the market and to influence prices.

Explanations for the disappearance of famine often focus on advances in agriculture and food production, including the introduction of New World crops (especially potatoes and maize), and, in the twentieth century, the Green Revolution in crop productivity that started in the 1950s. However, in the English case, explanations for the early escape from famine acknowledge the importance of favorable ratios of population to food production, but also emphasize improvements in the operation of grain markets, on the one hand, and improvements in institutional responses on the other, especially the English poor laws. Recent work by Phil Slavin on the greatest famine in European history, that of 1315–1317, has revealed clear evidence of hoarding of grain. Other studies of credit have demonstrated how the impact of successive harvest failures in 1315–1317 was amplified in England by the actions of landlords and even family members who took advantage of the crisis to engross landholdings and other assets. Conversely, work by Richard Smith has demonstrated how these practices were replaced, by the sixteenth century, with a parish-based system of taxation and poor relief (before compulsory legislation). This system of communal welfare may have been sufficient to avert severe mortality in the English population during the harvest failures of the 1590s and, more assuredly, in the dearth years of the 1690s, when perhaps 10 per cent of the Scottish population died.

These works raise the intriguing question of whether the early disappearance of famine in Western Europe owed more to improvements in famine relief measures and other institutional changes than to improvements in food production and market distribution.

Another important element in the disappearance of famine (and more generally of dearth-related mortality), is the control of epidemic diseases. It is likely that the immediate cause of most famine-related mortality in historical populations (except perhaps in very isolated and small-scale societies) was infectious diseases, rather than starvation. Typhus, or ‘famine fever’, was a common and lethal correlate of harvest failure, together with typhoid, dysentery, and other diseases including smallpox and malaria. Epidemics generally accompanied major harvest failures because these failures often prompted migration in search of work and food. Such migration brought rural populations into contact with novel diseases, especially when they moved into towns, and exposed them to over-crowded and insanitary conditions at a point when they were malnourished and probably immune-compromised. These epidemics also tended to become general, and therefore the mortality associated with past famines may have been fairly socially unselective.
Reductions in epidemic disease mortality associated with famines probably occurred via several mechanisms. First, where famine relief measures were relatively local and effective (as in the English case from the late sixteenth or seventeenth century onwards), these would have reduced migration and associated disease exposure. Second, disease control measures played an increasingly important role over time. These started with measures to isolate infected individuals and to disinfect clothing and bedding and extended in the twentieth century to large-scale immunization and medical programs in concert with often international famine relief measures. A grim testament to the effectiveness of such measures is the famine associated with the siege of Leningrad. The city was blockaded for nearly two years (1941–1943), and 800,000 civilians died, mainly of starvation. However sanitary measures were so strictly adhered to that there was no major outbreak of epidemic disease. It is a peculiar achievement of the twentieth century that so many famine victims escaped infectious diseases for long enough to starve to death.

Warfare too became less disease-prone in the eighteenth century, although it was not until the early twentieth century that battle deaths exceeded those from disease during military conflicts. Indeed eighteenth-century efforts to contain and prevent major disease outbreaks in the navy; in prisons and workhouses; and on slave, convict, and immigrant ships, probably led the way in developing effective techniques for reducing the transmission of contagious (person-to-person) and vector-borne diseases. These techniques included epidemiological surveillance (first adopted in Italy against plague, and entrenched in England in the form of the London bills of mortality), the isolation of individuals suspected of infection, contact tracing, surveillance, improved hygiene and diet, and reductions in crowding. While many of these measures, such as obsessive ventilation, were based on erroneous models of disease, others, such as isolation of the infected or the issuing of new clothes to sailors, were of clear benefit regardless of their theoretical justification. Perhaps more importantly, the publication and dissemination of these ideas subjected them to wide scrutiny and replication.

Pre-eminent amongst the diseases that declined before the mid-nineteenth century was smallpox. Smallpox was probably the single most lethal infectious disease in eighteenth-century Europe and was also a major cause of death in the Americas and parts of Asia and Africa. Although smallpox is usually considered an ancient disease of humans, its origins remain obscure. Recent breakthroughs in the sequencing of ancient DNA have produced evidence that a new variant of smallpox may have spread through Europe and then globally in the seventeenth century. By the second half of the eighteenth century, smallpox was responsible for 8 per cent of all deaths in Sweden and northern England, and 20 per cent in Manchester. Attempts to control smallpox through isolation of victims and inoculation with small doses of the virus were very widespread. However, the advent of immunization with cowpox (vaccination), discovered by Edward Jenner in 1796, and published in 1798, proved a decisive turning point. Vaccination was immediately recognized as a godsend, and it spread rapidly through Europe and into European colonies and ex-colonies, although its immediate effects varied enormously. It is difficult to estimate the true impact of smallpox, and of vaccination, on global mortality trends. This is because, first, vaccination was introduced well before most states introduced the systematic measurement of cases and deaths, and second, smallpox appears to have evolved into a number of subtypes with very different rates of mortality in the course of the nineteenth century. Therefore, although smallpox is celebrated as the first disease to...
be eradicated by human efforts (declared in 1980), the true impact of smallpox on historical populations probably remains under-appreciated.

By 1820 then, mortality patterns in north-western Europe were already profoundly different from previous centuries. Famine had receded, and the most lethal epidemic diseases (plague, typhus, smallpox, and malaria) had waned decisively in impact. The mortality associated with these diseases had been fairly indiscriminate, affecting wealthy and well-nourished individuals as well as the poor. Indeed, there is little evidence for a survival advantage to wealth before the nineteenth century. This is very surprising, since it might be thought that the advantages of better diet, clothing, and housing would have been even more important in the past, when living standards were lower. However, as Stephen Kunitz and Stanley Engerman have argued, the disease environment may have been sufficient to overwhelm these advantages before the nineteenth century. As these more lethal diseases receded, then, mortality became concentrated amongst the very young, the very old, and the poor. Because we know less about mortality patterns before the nineteenth century, it is common to think of these patterns as traditional; however, they probably represented a new and ‘modern’ pattern compared with previous centuries.

In Britain further improvements in longevity were impeded after c.1830 by unprecedented rates of urbanization, and possibly by scarlet fever, which underwent an apparently autonomous increase in virulence. However, in Sweden, where urbanization remained very modest before the late nineteenth century, life expectancy improved fairly continuously from 1750 to the present (Figure 1.2a). Woods has argued that similar underlying improvements in health in Britain were masked by the rapid redistribution of population from relatively healthy but poor rural areas to wealthier but unhealthier urban ones.

The eclipse of infectious diseases (1870–1940)

Rapid secular mortality decline resumed in England from c.1870, interrupted only by the major wars and the 1918/19 influenza pandemic (Figure 1.2b). Most of the decline in mortality between 1870 and 1940 was driven by declines in infection-related mortality in children and young adults, with the most rapid improvements occurring in the period 1900–1940 (Figure 1.4).

Scarlet fever and respiratory tuberculosis mortality began to decline from the 1870s. The other major ‘childhood’ diseases (measles, diphtheria, and pertussis) all underwent very substantial and poorly understood declines in the period after 1900. In the case of scarlet fever there is reason to think that the pathogen (Streptococcus pyogenes) declined in virulence, but no similar arguments have been made for other infections in this period, with the partial exception of tuberculosis and streptococcal puerperal fever. These childhood diseases are transmitted person to person and are highly infectious, and it appears that reductions in mortality resulted from declines in case-fatality rates rather than in incidence of infection. In the case of measles, Peter Aaby and colleagues have made a strong case for the importance of crowding and dosage of infection to mortality from measles. Improvements in nutrition associated with rising incomes doubtless acted to increase host resistance. In addition, improvements in housing conditions and the progressive and substantial declines in fertility and family size from the 1870s would also have acted to reduce infection dosage and exposure to opportunistic infections.
same constellation of factors probably contributed substantially to improvements in mortality from tuberculosis (Figure 1.6), although in this case the extent to which incidence of infection was also reduced via preventative measures such as confinement of sufferers to sanatoria remains unknown.51

The most lethal waterborne diseases (cholera, typhoid, and dysentery) declined from the 1860s. These trends imply early improvements in the quality of drinking water. However, diarrhoeal diseases displayed a more complicated pattern, declining from the 1870s amongst adults and older children but remaining stubbornly high, and indeed rising in

Figure 1.4 Annual age-specific mortality rates, England and Wales, 1841–2011.
Panel a includes all age groups, panel b displays ages where mortality rates were comparatively low (and declined early).
Source: Human Mortality Database (University of California, Berkeley), and Max Planck Institute for Demographic Research (Germany), www.humanmortality.de (downloaded on January 2, 2020).
Figure 1.5 Mortality by category of cause, ages 60–64, England and Wales, 1891–1967.
Dashed vertical lines indicate dates of changes in cause of death coding schemes. ‘Respiratory’ includes respiratory tuberculosis, influenza, and pneumonia; ‘circulatory’ includes strokes; ‘external’ includes all accidental and violent deaths.
in the 1890s, amongst infants. The late decline of infant diarrhoeal mortality may reflect differences in disease transmission pathways between infants and other age groups. Infants appear to have been relatively insensitive to water quality, in England, and diarrhoeal mortality in this age group was probably more closely related to sanitary conditions (including sewerage and fecal disposal) and domestic practices (breastfeeding, handwashing, and food preparation). Conversely, diarrhoeal diseases in older children and adults were probably more closely related to the quality of drinking water.

These patterns of declines in gastrointestinal (or ‘fecal-oral’) diseases present a puzzle, because they fit very poorly with the traditional chronology of improvements in water and sanitation in England, which were especially marked after 1870 (when urban authorities began to make very major investments in water provision and sewerage). The precocious declines of the most water-dependent diseases (cholera, dysentery, and typhoid) pre-dated large-scale investments in water supplies and suggest that relatively cheap or incomplete attempts by towns to source clean water may have been sufficient to reduce transmission of these diseases, in many cases before investments in water purification or sewerage. Conversely, infant diarrhoeal mortality was apparently unaffected by very major investments in water supplies. However the installation of comprehensive urban sewerage systems and sewage treatment lagged well behind investments in clean water supplies, and the continued retention of large volumes of human excrement in many British towns until the early twentieth century may account for the continued high levels of infant diarrhoea against a backdrop of major declines in waterborne diseases.

Figure 1.6 Tuberculosis mortality, England and Wales, 1848–2000.
Death rates are directly age-standardized using the 2012 population of England and Wales. Source: R.J. Davenport, Annual Deaths By Cause, Age and Sex in England and Wales, 1848–1900 [dataset] (2007), Economic and Social Data Services (SN5705).
This pattern, of precocious improvements in the most lethal waterborne diseases and slower improvements in diarrhoeal diseases of early childhood was also evident globally in the twentieth century. The more lethal ‘Asiatic’ cholera (which killed perhaps 15 per 100 of those infected) has been largely displaced since the 1960s by the ‘El Tor’ cholera subtype, which has a case-fatality rate of around 1 death per 100 infections.\textsuperscript{56} Typhoid and dysentery have also declined enormously as causes of death. Paul Ewald argued that these changes in water-borne diseases reflected widespread improvements in water quality, which have dramatically reduced water-based disease transmission.\textsuperscript{57} The WHO and UNICEF estimated in 2019 that 90 per cent of the global population now had access to ‘improved’ water sources (a clean source within 30 minutes’ round walk).\textsuperscript{58} Nonetheless, infant diarrhoeal diseases still constitute some of the most common causes of death in young children globally.\textsuperscript{59} If, as was the case in England, most diarrhoeal diseases in young children are more commonly caused by poor hygiene and fly-borne contamination of food than by contaminated water, then much more investment in fecal disposal will be required to reduce diarrhoeal mortality further. In affluent countries it proved much harder to remove excrement from cities and to treat sewage than to provide uncontaminated water.

Mortality began to improve in older adults after c.1900 in England (with the exception of the oldest ages, 85 years and over, where substantial improvements occurred only in the second half of the twentieth century) (Figure 1.4b). However, here the patterns were more complex. Figure 1.5 shows death rates by cause for 60–64-year-old men and women in England between 1891 and 1967. Whereas improvements at younger ages reflected mainly falls in infectious disease mortality, the main gains in life expectancy at ages 40 and over from 1900 were a consequence of reductions in deaths attributed to tuberculosis, other respiratory infections (especially pneumonia and influenza), cardiovascular diseases (strokes and heart disease), and ‘other’ causes (Figure 1.5). The latter category included mainly diseases related to the stomach, liver, and kidneys and was supposed by Samuel Preston to contain a substantial cardiovascular component,\textsuperscript{60} but also included diseases of infectious origins (such as acute nephritis caused by streptococcal infection).

Some of the patterns in Figure 1.5 reflect changes in diagnostic accuracy (and especially the transfer of deaths from ill-defined categories into specific categories, including cancers). However, similar age-specific and cause-specific patterns of mortality decline have been observed in a variety of populations at different stages of economic development.\textsuperscript{61} Preston demonstrated, using long-run cause of death data for 42 countries, that falls in cardiovascular disease made a major contribution to the early stages of mortality decline, and he attributed this both to the tendency of respiratory infections in particular to trigger heart attacks, and to longer term effects of infections on susceptibility to chronic non-communicable conditions.

Infectious diseases may influence non-communicable diseases in a variety of ways, and on a variety of time scales. However, the magnitude of effects of early-life infections on health in adulthood remains hotly debated.\textsuperscript{62} There are clear examples of specific pathways from childhood infection to adult disease, as in the case of rheumatic fever and valvular heart disease or \textit{Helicobacter pylori} infection and stomach cancer.\textsuperscript{63} However, at the population level, the influences of early-life disease environments are less obvious. Life expectancy at older ages bears little relationship to levels of infection amongst the same cohorts during childhood (as gauged by childhood
mortality rates in these cohorts). Elderly cohorts in countries such as Japan and Italy with relatively high levels of historic childhood mortality now enjoy some of the highest life expectancies in the world. The synergies between different causes of death remain an under-researched area.

**The age of degenerative diseases (1940–present)**

We have divided the discussion of English mortality patterns at 1940 because the development of antibiotics in the late 1930s and early 1940s marked a turning point in the capacity to cure disease. Most health innovations up to this point were preventative – they reduced exposure to disease. Antibiotics, on the other hand, could be used to cure infections.

The discovery of the bactericidal properties of protonsil in 1931 led to the proliferation of related 'sulfa drugs' that came into wide use in Europe and the USA by the late 1930s. Sulfa drugs were effective against a narrow but important range of bacterial infections, including the streptococcal strains responsible for puerperal fever in recently delivered women and streptococcal wound infections, but not the streptococcal strains that cause scarlet fever. Their impact on maternal mortality was dramatic. Neonatal mortality, particularly deaths in the first week of life, also declined precipitously with the introduction of sulfa drugs. Anne Løkke has argued persuasively that antibiotics were key in making caesarean deliveries safe and that this meant that difficult deliveries ended more frequently in a successful live birth, rather than stillbirth and maternal infection.

Sulfa drugs also proved effective in treating pneumonia and tuberculosis. Penicillin was developed in the 1930s and widely used from 1944. It proved effective against wound infections and was widely deployed in the closing stages of WWII by the American and British armies. Of even greater significance was the discovery in 1946 that streptomycin cured tuberculosis. Nonetheless, the immediate impact of penicillin-type drugs on life expectancy trends was slight (Figure 1.2b). Even in the case of tuberculosis, the impact in Britain was much smaller than it would have been had penicillin been discovered several decades earlier. Tuberculosis mortality declined enormously in the late nineteenth century and the first half of the twentieth century, before penicillin (Figure 1.6). Nonetheless, the advent of penicillin marked a decisive break, and, within ten years of its use, most sanatoria for the treatment of tuberculosis sufferers had been converted to ski resorts, psychiatric hospitals, and for other uses.

Importantly, the apparently small contribution of the discovery of antibiotics to life expectancy trends masks their enormous contribution over the last 70 years to the treatment of non-communicable diseases. Before 1950, surgery remained a very dangerous procedure, despite very significant developments in sterile procedures and wound treatment. Many of the gains in survival from heart disease and cancers in the last 50 years depended and continue to depend on surgical interventions that would have involved substantial risk before the advent of penicillin. Chemotherapy, which reduces the patient’s resistance to infection, also relies on the prophylactic use of antibiotics to reduce the risk of opportunistic infections, as does organ transplant technology.

In European populations, infectious diseases had been controlled to a great extent by 1940 using methods that prevented transmission or raised population resistance (rather than cured infection per se). These older methods of disease prevention were further enhanced in the second half of the twentieth century by the development or wider
implementation of new vaccines against a range of infectious diseases, including measles, pertussis, yellow fever, tuberculosis, hepatitis, typhus, typhoid, polio, pneumonia, and chickenpox.

An example of this mix of preventive and curative approaches is tuberculosis. Tuberculosis was (and remains) a major cause of mortality globally. The BCG vaccine against tuberculosis was developed in the 1920s but was only widely used after WWII. Antibiotics, BCG, and surveillance were then used in a three-pronged strategy to identify and treat those infected, to reduce transmission, and to immunize the population against infection. In addition, some of the declines in tuberculosis over the latter half of the twentieth century reflect the same factors in operation in affluent countries before the advent of antibiotics and BCG vaccine, including better nutrition. However, tuberculosis resurfaced in the 1990s as an opportunistic infection in HIV-affected individuals and because of the development of strains resistant to multiple types of antibiotic treatment. BCG immunization does not face the same problem of evolved resistance, and it remains routine in most low- and middle-income countries. However, its efficacy varies. BCG is protective against tuberculosis infection in infants, but the efficacy of BCG against adult respiratory forms of the disease varies with latitude and age at immunization, with efficacy ranging from zero at low latitudes to 60–80 per cent protection at latitudes above 40 degrees. Therefore, the relative contributions of antibiotics, immunization, and improvements in living conditions to global declines in tuberculosis are very geographically as well as chronologically heterogeneous.

The globalization of mortality decline, 1950 onwards

Famines and major epidemic outbreaks began, slowly, to recede in much of the world in the late nineteenth century and the first half of the twentieth century so that, by the time the world emerged into the spotlight of global statistics in the 1950s, profound changes had already taken place in much of the world’s population. In India, for instance, large upswings in mortality caused by famine, bubonic plague, and smallpox diminished substantially after 1920. Tim Dyson and Monica Das Gupta attributed this stabilization of mortality to colonial policies that improved food distribution, monitored plague outbreaks, and increased smallpox vaccination coverage. A similar stabilization of mortality after 1920 is evident in Jamaica, where a precocious public health administration, education, and registration system facilitated rapid gains in life expectancy from this date, with the fastest gains in the period 1925–1940. In many African societies, the progressive eradication of smallpox, culminating in the WHO global campaign from 1950 to 1980, must have had an effect on populations with hitherto low rates of vaccination as profound as that in early nineteenth-century Europe.

While the early gains in life expectancy in affluent countries had increased global inequalities in health by the mid-twentieth century, the period after 1950 witnessed very substantial convergence, at the national level (Figure 1.3). In poor countries, the process of infectious disease control, so drawn out in England, was enormously accelerated in the second half of the twentieth century by the rapid availability of medical and public health technologies and knowledge. These included immunization, antibiotics, germ theory and sanitation technologies, and insecticides. The rapidity of global improvements in mortality is all the more astonishing when one considers the challenging disease environments in which some of these changes occurred. England’s slow mortality decline took
place in a temperate island setting, with a low burden of arthropod-borne infections and relatively easy means of controlling disease importation through quarantine. Populations in low-altitude tropical areas faced much higher disease burdens as a consequence of climatic and ecological conditions, especially in Africa where co-evolution between pathogens and humans and other primates was very longstanding. Moreover, these rapid mortality declines have occurred in the context of ever-increasing globalization and urbanization.

Some of the very rapid gains made in the decade after WWII must be attributed to the widespread use of DDT to destroy the insect vectors of malaria, trypanosomiasis, yellow fever, and typhus. Malaria in particular was a major cause of mortality and morbidity across tropical Africa, Central and South America, and South and South-East Asia (including southern China). Malarial infection induces only short-lived immunity, but, where malarial exposure is most intense, the burden of infection falls mainly on very young children and immigrants (because these groups lack immunity). Malarial mortality was dramatically reduced by DDT spraying of mosquito breeding sites and, more latterly, by anti-malarial drugs and the use of bed nets. However, the complexity of the malarial life cycle, the evolution of drug resistance, and regulation of insecticide use has repeatedly threatened these gains.

The huge gains in global life expectancy since 1950 were accompanied by very substantial economic development. The extent of these changes over the last 70 years remains very under-appreciated. Today over 90 per cent of the world’s population lives in ‘middle-income’ or ‘high-income’ countries, and half lives in states with below-replacement levels of fertility (that is, with total fertility rates below 2.1 children per woman). Economic growth has coincided with rapid increases in education, medical technologies, and falls in fertility, as well as enormous social and cultural changes. Given the rapidity of change, is it possible to identify the key distal drivers of mortality improvements?

The difficulties involved in teasing out the relationships between economic growth and demographic outcomes can be illustrated using so-called ‘Preston curves’. In a landmark paper in 1975, Samuel Preston produced a series of plots of national life expectancy by national income per capita (adjusted for change in nominal value of GDP over time), for 1900, 1930, and 1960. Figure 1.7 plots similar relationships for 1960 and 2018, and for the United States in 1900. As Figure 1.7 demonstrates, there is a strong relationship in any given year between the per capita income of a country and its average life expectancy at low levels of income. Conversely, above a certain threshold, income ceases to be associated with further gains in life expectancy. Crucially, however, although this relationship between income and life expectancy appears strong in any given year, the life expectancy that could be achieved for a given income has risen consistently over time. In 1900 the United States was the richest country in the world; however, it was unable to procure an average life expectancy for its citizens of more than 50 years. By 1960 countries with the same average income as the United States in 1900 had an average life expectancy of more than 60 years. By 2018 for the same income the average life expectancy was over 70 years.

Preston’s interpretation of these patterns was that rising national incomes could only account for a small fraction (less than a quarter) of improvements in life expectancy over time. He argued that health had in fact grown cheaper over time so that technologies that were not available at any price in 1900 (such as antibiotics) have become widely available since.
Figure 1.7 ‘Preston curves’; national life expectancies in relation to income per capita.

GDP per capita measures are expressed in international dollars (fixed 2011 prices). ‘Oil-rich’ states are defined as those which derive 10 per cent or more of GDP from oil rents. HIV-affected states are those with an estimated HIV prevalence of 5 per cent or more of adults aged 15–49.

Other scholars have argued that state interventions or cultural factors were more important than economic growth to improving longevity. Jack Caldwell drew on examples of ‘high-performing’ poor countries or states that managed to achieve very high levels of life expectancy before they achieved rapid rates of economic growth. These included Sri Lanka, Costa Rica, Jamaica, Kerala (in India), China, Cuba, and Vietnam. Analyzing the non-communist states, Caldwell concluded that these ‘high-performers’ were characterized by cultural traits including relatively high female status (and therefore relatively high levels of female education), institutional factors that included fairly egalitarian access to healthcare (such as immunization, rural clinics, and ‘barefoot doctor’ services), and either relatively egalitarian access to food or a social safety net that operated in times of dearth. He also drew attention to the poor performance of oil-rich states, which had relatively high incomes but unimpressive life expectancies (as is evident in Figure 1.7a).

More recently James Riley analyzed a wider set of early-achieving ‘high performers’ and concluded that there were in fact many pathways to rapid health improvements. Riley’s high performing states (relative to GDP) included a number of oil-rich populations (in particular, Oman and Venezuela), although it is only in recent decades that the generally poor performance of oil-rich countries has diminished (Figure 1.7b). Riley’s work emphasizes the heterogeneity of possible routes to low mortality.

Improvements in mortality since 1800 were associated with very profound economic, technological, and social changes and defy moncausal explanations. Explanations fall largely into two camps, those invoking economic development and rising living standards, especially improved nutrition, and those which attribute the bulk of improvement to public health measures. Predictably, the evidence is ambiguous. The ubiquity of mortality improvements in global populations at very different levels of income, development, and life expectancy points to the potential multiplicity of pathways and perhaps to the importance of fertility decline and education. Additionally, there may be bi-directional influences between health and income. Robert Fogel described a ‘techno-physio revolution’, where improvements in diet increased human productivity, which then further improved output and economic development in a virtuous circle of improving health and wealth. Easterlin, on the other hand, considered that both economic growth and mortality improvements derived from the growth of scientific knowledge but were not themselves strongly causally related. The relative importance of household income, medical technologies, public health interventions, and wider changes in fertility and education remains controversial, and this may be because their influence varied widely. This is almost inevitably the case, because local cultures, histories, and ecological conditions varied so widely.

The future

What can the past tell us about potential future trends? Perhaps the most important pattern to note is that life expectancy has generally continued to rise, despite the enormous challenges posed by rapid population growth and urbanization. These factors would have caused unsustainable increases in mortality in traditional societies, and yet now famines are rare, and most urban populations have higher life expectancies than their rural counterparts. Remarkably, the trend in ‘best case’ life expectancy (for the national population with the highest life expectancy in any given year) has continued to rise fairly linearly (at a rate of approximately three years of extra life per decade) since 1840. This suggests
that any biological limits to human life expectancy, if they exist, are yet to be encountered. Therefore, future trends will probably continue to depend on human capacity to overcome old threats and new challenges to human health.

These threats are very substantial, potentially even larger than the factors that drove the numerous setbacks to longevity gains over the last two centuries. Reasons why survival gains could falter or even reverse include the emergence of novel pathogens (through inter-species transfers and autonomous changes in pathogen properties), the multifarious effects of climate change (on environmental conditions, migration, and disease vector ranges), antibiotic resistance, and obesity. On the positive side, a concerted move towards clean energy will make cities substantially healthier, and this is where the majority of the world’s population now lives. Urban populations now generally enjoy higher life expectancies than rural populations, a striking reversal of historic norms. However, this advantage reflects, in part, the higher wealth of cities and superior access to public health measures (sewerage, clean water, immunization programs) and to clinics and hospitals, which outweigh the disamenities of crowding and poor air quality. This balance will change as air pollution reduces but also as rural populations gain greater access to health facilities. The balance of these and other trends will help to determine whether individual and global inequalities in health and longevity widen or narrow in the twenty-first century.

This chapter is Open Access with funding from Wellcome Trust Award 103322.

Notes
1 This work was funded by Wellcome Trust award 103322 (to Prof. Richard Smith, University of Cambridge). The author gratefully acknowledges this funding.
10 Roser et al., “Life Expectancy.”


31 Hoyle, “Britain”; Smith, “Dearth and Local.”


33 Walter and Schofield, “Famine, Disease.”

35 Ó Gráda, Famine, pp. 108–121.


37 Riley, The Eighteenth-Century.


45 Davenport, “Urbanisation.”


53 Ibid.


55 Davenport et al., “Cholera.”


57 Ibid, 77–86.


61 Ibid.


64 E. Barbi and J. Vaupel, “Comment on Inflammatory Exposure and Historical Changes in Human Life-Spans,” *Science*, 308 (2005), 16.

65 There were some notable curative antecedents, such as quinine, which was used against malaria from at least the sixteenth century in Bolivia and Peru; disinfectants that killed pathogens and were deployed in the treatment of wounds; anti-toxins (antibodies) against diphtheria and tetanus, developed in 1890; and salvarsan against spirochaete infections, notably syphilis (discovered in 1909).

66 Loudon, *Death*.


72 Riley, *Poverty*.


76 Ibid.


78 Ibid.


83 Floud et al., *The Changing Body*.

84 Easterlin, *Growth*.