Industrialization and health

Simon Szreter
St John’s College, Cambridge, UK

Throughout history and prehistory trade and economic growth have always entailed serious population health challenges. The post-war orthodoxies of demographic and epidemiological transition theory and the Washington consensus have each encouraged the view that industrialization necessarily changes all this and that modern forms of rapid economic growth will reliably deliver enhanced population health. A more careful review of the historical demographic and anthropometric evidence demonstrates that this is empirically false, and a fallacious oversimplification. All documented developed nations endured the ‘four Ds’ of disruption, deprivation, disease and death during their historic industrializations. The well-documented British historical case is reviewed in detail to examine the principal factors involved. This shows that political and ideological divisions and conflict—and their subsequent resolution in favour of the health interests of the working-class majorities—were key factors in determining whether industrialization exerted a positive or negative net effect on population health.

Industrialization refers to a process which has occurred in the history of all economically ‘developed’ nation states and which remains an aspiration for most of the governments of those many populations which remain today relatively undeveloped. Through industrialization the economy of a country is dramatically transformed so that the means whereby it produces material commodities is increasingly mechanized since human or animal labour is increasingly replaced by other, predominantly mineral sources of energy in direct application to the production of useful commodities. Industrialization is a special case of the near-universal phenomenon of human trade and economic change. It refers to a period of marked intensification of such activity, which in all known cases has resulted in an irreversible change in a country’s economy, after which the production and international trading of commodities remains permanently at a much higher level of intensity. This is largely because the factorial increase in productive capacities made possible by the technological shift in power supply simultaneously entails a wide range of accompanying transformations in the social relations of work, trade, communications, consumption and human settlement patterns and so, inevitably, also implies profound cultural, ideological and political change.
It would be extraordinary if such a thoroughgoing process did not have a range of significant health implications. Two of the oldest, most well-established relationships between economic activity, or trade, and population health are recognized to be mediated through the epidemiological implications of, firstly, regular social interaction between populations previously not exposed to each other’s disease ecology, and, secondly, the increasingly dense permanent settlement of populations, which occurs in the form of towns occupying nodal or strategic points in trading networks. Both of these relationships have always been understood to be negative, in terms of the health of the populations exposed\textsuperscript{2-4}. It has always been realized that the lure and the material benefits of economic exchange between peoples possessing different resources and producing different commodities carry enhanced risks of the accompanying exchange of potentially fatal diseases. The historical records of the early modern city-states of Italy, for instance, demonstrate their governments’ attentions to a range of public health issues to do with the sanitary problems of packed, urban living and the periodic threats of imported epidemics\textsuperscript{5}. The gradual expansion of international and intercontinental trade, including of course in persons themselves, throughout the subsequent centuries was characterized by a sequence of extraordinarily lethal epidemics of infectious disease, most tragic of all for the indigenous populations of the Americas. Thus one of France’s most eminent historians has famously written of the era of rising world trade from the 14th century to the 17th as the era of ‘l’unification microbienne du monde’\textsuperscript{6}.

However, despite these well-understood, long-standing negative health risks associated with urbanization and with trade, by contrast the process of industrialization has in general been considered to have a much more positive relationship with human health. There is of course a very obvious intuitive reason for this. It is widely understood that industrialization was a necessary initiating historical process experienced by all today’s ‘successful’, high per capita income societies. These are generally among the populations with the highest life expectancy at birth in the world today. This has been made possible by the advanced medical technology, better food supply, and increased material living standards as a result of the continuous process of economic growth they have all experienced ever since industrialization. The apparently compelling logical inference is that industrialization has improved human welfare and health. This conclusion has been repeatedly supported during the course of the 20th century by a succession of research-based interpretations of the relationship between health and the kind of sustained economic growth made possible by industrialization\textsuperscript{7-13}. The study of British economic history has played a particularly crucial role in informing this generally positive evaluation, partly because it was the first nation-state ever to industrialize but also because of the exceptionally high quality and quantity of its historical
medical, epidemiological and demographic as well as economic data. This is due principally to the fact that the British nation-state, as a record-creating and preserving entity, has maintained its integrity throughout many centuries, resulting in the survival of a relative abundance of evidence.\footnote{47}

The preponderant importance of a secular fall in mortality as being the first and foremost welfare dividend to flow from industrialization has been a central feature of the orthodox consensus throughout the last century. By the beginning of the 20th century, it was obvious that rapid population growth had accompanied the process of industrialization in each modern country’s history. In Sweden, the only country whose official vital statistics reliably reached back to the 18th century, it was also evident that the population growth of the 19th century had been principally as a result of falling mortality, reflecting improving population health.\footnote{48} In 1926, two independent research monographs on Britain appeared\footnote{7,8}, each documenting all the important advances in medical knowledge and institutions, which occurred from the late 17th through to the early 19th century. These were portrayed as the health-enhancing first fruits of the same burgeoning spirit of rational scientific enquiry which had produced concomitant advances in technology and industry. By 1929, a grand general theory of ‘demographic transition’ had been sketched, which was to become the dominant international ‘development’ orthodoxy throughout the post-war era\footnote{14–16}. This envisaged all industrializing countries necessarily passing through a linear evolutionary pattern of three stages. The \textit{primum mobile} of economic growth directly caused a fall in the high mortality rates characterizing stage one, by raising living standards and through the society’s enhanced ability to benefit from medical science, hygiene and sanitation. Consequently, during the transitional stage two, population growth rates increased rapidly until, in the final stage three, parents adjusted their traditional pro-fertility behaviour by reducing their birth rates to reflect the new circumstances of much higher survival rates for their offspring.

In the 1970s, transition theory was apparently further elaborated by two influential contributions. Firstly, Omran’s concept of the epidemiological transition specified three types of epidemiological regime typical of the three stages of demographic transition\footnote{17}. Famines and pestilence dominated the pre-industrial high mortality stage, followed by ‘receding pandemics’ as transitional societies industrialized, became wealthier and their medical technology advanced. Finally, the most developed, high life expectancy societies of stage three were afflicted primarily by a residual of ‘degenerative and man-made diseases’. Secondly, Thomas McKeown’s widely-read \textit{The Modern Rise of Population} argued that the principal cause of the mortality decline consequent on industrialization, as specified in the transition model, was not medical science and technology but primarily rising living standards\footnote{10}. The beneficial effect of economic
growth on population health was initially transmitted primarily through a gradually rising per capita nutritional intake made possible by a better food supply and rising real incomes (purchasing power). McKeown founded this conclusion on his pioneering epidemiological analysis of the historical series of detailed cause-of-death data available for the whole population of England and Wales since the mid-19th century.

Although McKeown’s thesis, to the extent that it was evidence-based, applied only to the epidemiological history of one country, his findings were nevertheless taken to be broadly generalizable. This was partly because of McKeown’s persuasive skills and his impressively detailed epidemiological data. It was also the result of a widespread assumption, which pervaded the post-war era and which continues to be influential, that the demographic or epidemiological transition is itself a singular, generic process, which has occurred repeatedly following industrialization in all developed countries’ histories. It follows from this assumption that it can therefore be adequately studied through a single well-documented example. It also followed that the currently non-industrialized countries of the 1970s might profitably learn from such a model and fashion their development policies accordingly.

The 1970s also witnessed the emergence of a resurgent, monetarist and neo-classical economics, which, during the course of the 1980s, replaced the social democratic ‘Keynesian’ with the neo-liberal ‘Washington’ consensus as the dominant programmatic set of policy prescriptions informing the macro-economic and lending policies of western governments and banks and the major Bretton Woods institutions of the World Bank and the IMF, located in Washington. The existence of McKeown’s well-publicized work made it much easier to press forward the neo-liberal economic agenda in the course of the 1980s, with its focus on maximizing capitalist, free market economic growth, not only in the ‘First World’ but also in the world’s least developed countries, since McKeown had apparently proved that the rising living standards facilitated by industrialization had been the principal cause of epidemiological transition in the past.

There had always been important dissenting voices, which disputed the general validity of McKeown’s work, notably, Sam Preston’s important cross-national statistical research. This indicated that during the course of the 20th century rises in societies’ overall investments in health-promoting technology and services—much of it state-organized and funded—was a more significant source of gains in average life expectancy than their rising per capita incomes. However, this was not the message that neo-liberal economists wanted to hear, intent as they were on ‘rolling back’ the state and freeing-up the market. Furthermore, during the 1980s, McKeown’s emphasis on the importance of nutrition also caught the eye of the highest profile practitioner of economic history. The Nobel prizewinner Robert...
Fogel published a series of research papers during the late 1980s and early 1990s which presented a new source of long-run historical health data—the anthropometric evidence of American military recruits’ heights and weights. He argued, along McKeownite lines, that this also showed that nutritional inputs were the most important driver of population health during the initial stages of industrialization. Thus, in the important World Development Report for 1991, compiled under the general direction of the leading neo-liberal, Lawrence Summers, Fogel’s work was given prominence and McKeown was cited but there was no reference to Preston’s alternative analyses.

However, in Britain the 1980s also saw the publication of a major new work of long-term historical demographic reconstruction, which radically undercut the crucial assumptions of ‘transition’ theory and so, also of McKeown’s interpretation of the British epidemiological data from the mid-19th century onwards. The Cambridge Group for the History of Population and Social Structure succeeded in reconstructing the population history of England, including national trends in mortality and fertility, on the basis of a 4% sample of the data held in the 10,000 parish registers of England back to their instigation by Henry VIII in 1538. Their work demonstrated, firstly, that England before industrialization was not a regime of high famine and pestilence mortality as envisioned in transition thinking. Secondly, the quadrupling of English population, which took place during industrialization between 1700 and 1870, was driven principally by the increased fertility of marriage and only to a relatively slight extent by a modest fall in mortality. Around 1700, expectation of life at birth had been approximately 36 years and by 1871 it still stood at no more than 41 years. Following this pioneering effort, there has been an enormous flow of further primary research exploiting Britain’s parish registers and much other relevant evidence, which has confirmed these two principal findings.

McKeown had supposed, from within the perspective of modernization and transition thinking, that in addressing the epidemiological patterns of falling mortality, which he could track from the Registrar-General’s official cause of death data from 1851 onwards, he was analysing a single secular trend, which would have started during the late 18th century when it was believed that the British industrial revolution had begun. However, one of the further important conclusions to emerge from the research of the demographic historians was that McKeown’s data series began in the middle of a strange, half-century-long period of stasis in the nation’s mortality. The national average expectation of life at birth had improved fitfully and gradually during the 18th century to reach a level of about 41 years by 1811 but thereafter it failed to register any further improvement above that level until the 1870s. This meant that during the whole of the period when the British economy experienced its
historically unprecedented, sustained economic growth rates, while its steam-driven economy powered its way to global trading predominance during the long mid-Victorian boom, overall mortality rates failed to improve at all. Although health had apparently improved moderately during the initial phases of slow economic growth in the 18th century, when full-scale industrialization arrived with the diffusion of steam technology, factories and rail transport, there were then no further net gains in health for two generations. This is despite the fact that workers’ average real wages, which showed no overall improvement before 1811, now began definitely to rise throughout the rest of the 19th century. This chronology is all wrong for the McKeown thesis. Mortality fell in the 18th century without the benefit of increased purchasing power for food (the fluctuating cost of food was the major budgetary item influencing the reconstructed average real wage trend), whereas overall health failed to improve between 1811 and 1871, despite enhanced purchasing power.

Further research on an independent body of evidence, British anthropometric data, has confirmed that late 18th-century improvements in height attainments were curtailed and then even reversed during the second quarter of the 19th century. It is now clear from this and from other detailed demographic research on urban patterns of mortality during this period, that the principal reason for the failure of the national average life expectancy to register any further gains between 1811 and 1871 was due mainly to deteriorating health conditions in Britain’s industrializing towns and cities (Szreter and Mooney). All the available evidence for a variety of towns of very different sizes, from a Carlisle or a Wigan to Glasgow, exhibits the same patterns and trends. Urban life expectancies, though they had probably improved during the late 18th century, were well below the national average by the end of the first quarter of the 19th century. Thereafter they experienced a particularly deep crisis persisting for two decades during the 1830s and 1840s, followed by a return to the pre-crisis levels (i.e. still well below the static national average) in the 1850s and 1860s. From the 1870s onwards, urban life expectancies finally began to climb above the levels of the early 19th century and, in so doing, pushed the national average onto an upward trend, too (Britain by this time having become a predominantly urban society).

Thus, quite to the contrary of the dominant 20th-century consensus, the only abundantly documented historical case, Britain, shows that industrialization had a powerfully negative direct impact on population health, concentrated particularly among the families of the relatively disempowered, displaced migrants who provided a large part of the workforces in the fast-growing industrial towns and cities. According to this viewpoint, industrialization is not a special case, but conforms to the more general pattern, throughout human history, that periods of increasing economic activity, because they are associated with increasing
trade and urban settlement, are also intrinsically productive of increased health risks. Indeed, industrialization, because it is so extensive in its economic scale of transformation, may well exert its negative health effects more dramatically and rapidly than any of the historically earlier forms of more moderate increases in trade and economic activity.

There are a number of ways of seeking to explain these findings about 19th-century Britain in such a way as to reject this conclusion and to preserve instead the conviction that industrialization is, still, a special case and has been a positive influence on health. However, each of these collapses on closer examination. It is, for instance, not the case that such negative health effects as Britain’s towns experienced in the 1830s and 1840s were ‘merely’ the result of urban size or speed of growth or inadequate knowledge of health-preserving technology at that time. Towns of all different sizes from just 20,000 to over 100,000 inhabitants were affected. Most cities grew no faster in these two decades than any two of the previous six or seven decades. Nor was there an inevitable knowledge or ‘learning’ deficit. The technology for constructing urban water supply and the importance of sanitation and sewering was well-understood, as Edwin Chadwick’s summation of knowledge published in 1842 shows29; the importance of personal hygiene, good food and cleanliness of the personal environment was also well-understood as Haines et al have ingeniously demonstrated30.

The heterodox thesis is that industrialization itself, like all forms of economic growth, exerts intrinsically negative population health effects among those communities most directly involved in the transformations which it entails. The case for this apparently paradoxical proposition grows much stronger when it is realized that in virtually all known cases of the industrialization of today’s successful developed economies, their historical demographic or anthropometric trends exhibit the same ‘trademark’ pattern of a negative inflection in the health trends during the decades in which industrialization most affected their populations. This is true, for instance of studies which have been published on populations in USA, Germany, France, Holland, Japan, Australia, Canada and Sweden31 (Sweden has sometimes been considered an exception, but the most recent research has shown that the landless Swedish rural populace did suffer significant health consequences during the second quarter of the 19th century when their agricultural economy was first exposed to commercial pressures necessitating raised productivity, whereas later in the century it was the crucial role played by advanced government public health measures in the 1870s in anticipating the health problems of industrial urbanization, which minimized such negative effects when Sweden experienced its own industrialization)32,33.

However, it is also true that in each of these cases, as in Britain, a period during which the health of the population was compromised by
industrialization was ultimately resolved, so that continuing economic growth came eventually to be accompanied by generally rising health—even in the largest most densely populated cities—resulting in the high life expectancy societies of the present day. The crucial analytical point, of enormous policy relevance, is that this potential capacity of post-industrial economic growth to provide the material basis for generally enhanced population health is not intrinsic to the process of industrialization or of economic growth in itself.

As careful attention to the historical relationship between industrialization and health in the case of Britain and most other countries shows, the direct consequences of rapid economic growth on health are likely to be negative, for a set of long-understood epidemiological reasons. In fact the kind of dramatic transformation associated with the industrialization of an economy is especially likely to be negative in its immediate impact on health and welfare because of the profoundly disruptive nature of this change. The disruption is simultaneously multi-dimensional: social and familial relations, moral codes, ethical standards of behaviour, the physical and the built environments, forms of government, political ideologies and the law itself are all thrown into flux and tumult when a society experiences industrialization and the consequent population movements that are entailed. Such disruptions tend to cause forms of social deprivation to arise, which can lead to disease and ultimately to death for the most unfortunate and marginalized individuals—often children, migrants or ethnic minorities. These are the ‘four Ds’ of rapid economic growth: disruption, deprivation, disease and death. They can only be addressed through political mobilization of the society to devise new structures, which can respond to the forces of disruption and remedy their consequences. This typically requires, at a minimum, massive investment in urban preventive health infrastructure, and an accompanying regulatory and inspection system, along with a humane social security system.

The classic, catch-22 problem, for societies experiencing the disorienting transformations of industrialization is that politics itself is profoundly disrupted, since the process throws up, by definition, a variety of newly powerful commercial and business ‘interest’ groups, typically very divided among themselves on ethnic, regional, industrial or religious lines, to challenge the incumbent governing classes. In British society and its industrial towns, an effective paralysis of the political will occurred for two generations between approximately 1830 and 1870 as successive national and local governments doggedly dodged the expensive issue of investment in urban preventive health infrastructure, even in the face of recurrent cholera visitations. The default ideology of this era, ‘laissez-faire, laissez passer’, reflected the political wisdom that in such a socially fissured society of vigorous competing interests, ‘every man for himself’ was the only general proposition which could command assent. In an as
yet undemocratic ‘shopocracy’, dominated by the votes of those precariously trying to keep their heads above water in a roller-coaster market economy, the only elective governments were those which promised to keep national income tax or local rates to an absolute minimum—the most common electoral battle cries were ‘retrenchment’ and ‘economy’\textsuperscript{35}. As a result, whereas the ‘winners’ in this society invested and gambled huge amounts of capital in the railway mania, there was no adequate collective investment in even the basic urban health infrastructure of sewers and clean water and street paving (crucial for health in a horse-drawn economy)\textsuperscript{36}. Whereas the paternalistic landed governing class had presided in the late 18th century over an increasingly generous national social security system, the Old Poor Law, spending was slashed under the deterrent ‘workhouse’ system of the New Poor Law of 1834, reflecting the evaporation of social trust between the classes in this disrupted and divided society\textsuperscript{37,38}.

After delaying for as long as they dared, from 1867 to 1928, in response to organized male working-class and subsequent feminist political pressure, the British propertied governing class passed a sequence of four major enfranchisement acts which ultimately granted the vote to all adults of both sexes on an equal basis. From 1867 onwards, this began to transform the electoral arithmetic and the politics of the health and social security needs of the wage labour class in society. The shift in political economy occurred first at municipal level. Under its visionary Mayor Joseph Chamberlain, an industrial magnate, the city of Birmingham pioneered a programme of ‘gas and water socialism’ as its opponents vilified it\textsuperscript{39,40}. Local monopoly services were bought, built and run by the city to provide revenue for an expanding preventive health and social services infrastructure. Once Chamberlain had proved both the electoral and the practical viability of this new political economy, all other major cities and eventually smaller towns, too, followed suit over the next three decades\textsuperscript{41}. The towns were beautified but also, crucially, the urban death-rates came tumbling down as local authorities’ expenditure on the health and environmental needs of their mass electorates multiplied to the point where in 1905 the total amount spent by vigorous local governments actually exceeded (for the only time in Britain’s recorded history) the total spent by central government\textsuperscript{42}. In December 1905, the ‘New Liberal’ administration won a landslide general election victory and ushered in an entirely new era of state activism with a host of centrally-organized and funded measures, such as old age pensions, labour exchanges, a school medical inspection service, free school meals for the needy, and national insurance against sickness and unemployment for workers. The politics of working-class interests had thus transmuted from the municipal to the national stage in Britain, something which would ultimately lead to the enactment of the welfare state.
The lessons of history, therefore, are that all economic exchange entails health risks and that industrialization typically results in a particularly concentrated cocktail of such health risks. From a policy point of view, it is particularly important that currently non-industrialized societies are neither encouraged nor forced to enter the industrialization process without a clear understanding of the difficult prospects which they face for at least a generation while undergoing this profoundly disruptive process. It may well be possible to avoid the undesirable fourth ‘D’ of death and possibly even the third ‘D’ of disease, given a sufficiently careful and thoroughgoing effort to manage and respond to the forms of deprivation which rapid economic growth produces as it transforms communities and relationships—something which Sweden may well have achieved during the last quarter of the 19th century. Like the Swedish case, the British historical case also suggests that extremely committed, well-informed, well-funded, devolved and democratically responsive forms of local government may be more important than the central state in effectively managing the immediate negative health consequences of industrialization. However, ultimately, the redistributive resources and authority of the central state in a democratic society will undoubtedly become important in ensuring that long-term sustained economic growth continues to be a benefit to the health and welfare of the whole population, rather than merely a source of ever-increasing private wealth to a small proportion of individuals favoured by birth and by chance, which is a tendency inherent in the normal working of unregulated, free market capitalism.

The apparently intuitively obvious notion that the economic growth of industrialization must be straightforwardly beneficial for health has, thus, been shown to be based on a misleading simplification of economic and demographic history, though one which was apparently supported by now-obsolete historical and epidemiological interpretations of history. It is now increasingly emphasized by historical researchers that politics and government have played an all-important role in ensuring that the wealth accumulated by the socially divisive and competitive processes of market economic growth is recycled and redistributed throughout a society to ensure that it contributes more equitably to the overall population health and welfare of the vast majority of the citizens involved in the process as producers and consumers\textsuperscript{43,44}. Unfortunately there is insufficient sign as yet that this understanding is informing the strategy of the most important international institutions which influence the future course of world development, notably the IMF and the WTO (the World Bank has been notably more ambivalent in its approach since the World Development Report of 1997). Policy prescriptions for the world’s poorest countries need to recognize that their state and local government capacity has been dangerously decimated during the last two decades of neo-liberal, free market fundamentalism\textsuperscript{45,46}.
Notes

a Such transition thinking is an integral part of a more general, encompassing ‘modernization’ ideology, a set of ideas which trace their genealogy to the post-Enlightenment project to spread liberty, scientific reason and democracy to the world, which remains a profoundly influential motivating force in contemporary global history, in particular providing the ethical rationale for the project of international ‘development’.

b Fogel had shot to fame in the 1970s with his co-author Stanley Engerman through their pioneering quantitative econometric history of slavery which startlingly concluded that slavery was an efficient economic system and that most black southern slaves had enjoyed a higher standard of living than freed wage-earners in the industrial north in the pre-civil war era: Fogel RW, Engerman SL, *Time on the Gross*. London: Wildwood House, 1974.

References

1 Wrigley EA. *Continuity, Chance and Change*. Cambridge: Cambridge University Press, 1988
7 Griffith GT. *Population Problems of the Age of Malthus*. Cambridge: Cambridge University Press, 1926
13 Pritchett L, Summers LH. Wealthier is healthier. *J Hum Resour* 1996; 31: 841–68
42 Szreter S. Health, class, place, and politics: social capital and collective provision in Britain. *Contemp Br Hist* 2002; 16: 27–57