A socio-biological explanation for social disparities in non-communicable chronic diseases: the product of history?

C M Schooling, G M Leung

ABSTRACT
This study places social disparities in the major non-communicable chronic diseases within their global economic and historical contexts. Rapid economic transition outside the developed world provides a unique opportunity to re-examine the origins of, and biological mechanisms driving, social disparities. Gaps in prevailing theories focusing on material resources, civic infrastructure and social structure are identified. Using longstanding experimental evidence and epigenetic theories, it is suggested that exposure to economic development over generations (ie, improved living conditions over historical time) could by acting on different biological axes (somatotropic and gonadotropic) generate specific patterns of social disparities. Moreover, these same processes could initially generate a transient epidemic of diabetes as well as a permanent increase in male risk of premature ischaemic heart disease. As such, this study demonstrates the importance of context, and implies that current evidence from the developed world may be largely uninformative for preventing or mitigating social disparities in non-communicable chronic diseases elsewhere, suggesting research efforts should be focused on developing countries.

INTRODUCTION
Social disparities in health are pervasive and persistent. Impoverished environments indisputably lead to ill health, emphasising the importance of environmental conditions, encompassing material resources, civic infrastructure and social structure. Key to mitigating the effects of social disparities on health, indeed of any environmental threats, is to understand completely the aetiologic pathways through which adverse consequences are generated. Although the literature is populated with explanations for the multiple pathways by which the embodiment of lack of resources, social exclusion and discrimination may damage health, there are important gaps.

Social disparities have been most extensively documented in developed countries, albeit from different perspectives. In the USA, emphasis was initially on racial disparities in health, perhaps because of the civil rights movement and ideal of a classless society. In contrast, in the UK, greater emphasis has been placed on socio-economic disparities in health, at least since the 1980 Black Report in the present era. Despite political opposition, the underlying explanation is recognised as deprivation, although which of the multiple dimensions of deprivation are most important in generating disparities is difficult to distinguish from inevitably observational studies. Nevertheless, lack of personal resources and public infrastructure throughout life are recognised as important factors, with a heated debate as to whether a neo-material explanation is sufficient or whether depletion of social capital, stresses of low status or prestige, lack of power or invidious comparisons also make a difference. This debate is difficult to resolve because the processes and pathways by which social disparities generate diseases are not fully elucidated. However, it is increasingly important for non-communicable chronic diseases in the developing world. With economic development, rising national incomes and the associated changes in life expectancy, nutrition and lifestyle, chronic diseases are expected to be responsible for an increasing share of global deaths in developing countries over the next 20 years.

Most current theories concerning disparities in chronic disease originate from observations in long-term developed Caucasian populations. Evolution of disparities over time in parallel with economic development, either considered as historical time or as epidemiological time relative to the start of economic development, is not usually considered. This missing dimension may result in empirically driven hypotheses, which are inappropriate for, or easily disproved in, populations with different histories of economic development. Current rapid economic development and industrialisation in the developing world provides a unique opportunity to re-examine the origins and biological mechanisms driving social disparities. Below, social disparities in the major non-communicable diseases (ie, cardiovascular diseases (ischaemic heart disease (IHD) and stroke) and diabetes) are placed within their global and historical contexts, which may provide clues as to how the emergence of these disparities can be prevented and/or mitigated most effectively in transitioning economies.

GLOBAL CONTEXT
Disparities between countries
Figure 1 shows the current association between national income (gross domestic product (GDP) per head) and mortality (age-standardised mortality per 100 000) from stroke and IHD, where available (153 countries). Figure 2 shows the current association between national income and morbidity from diabetes (age-standardised disability adjusted life years per 100 000) for these same countries. In all cases, assessed from model fit (data not shown),
curvilinear relationships (as shown), obtained using fractional polynomials, captured the associations better than purely linear relationships. Morbidity is shown in preference to mortality for diabetes, because diabetes is not usually recorded as the cause of death.

Stroke mortality is consistently negatively associated with national income, with possibly a stronger association at higher levels of national income. In contrast, low levels of national income are positively, albeit weakly, associated with IHD and diabetes. Similarly, biological risk factors, such as cholesterol or obesity are positively associated with national income at low income levels, but negatively at high levels, whereas blood pressure has little clear association with national income.14

Disparities within countries

Somewhat similar to between country disparities, within countries at an individual level there is often a reverse social gradient at low levels of national income and the usual negative association between disease and socio-economic position at high levels of national income for IHD16–23 and diabetes.24–37 Moreover, there often appears to be a more marked reverse gradient for men, which is evident at higher levels of national income for IHD.16 38–41 There are also similar patterns for biological risk factors, such as obesity15 42 43 and hypertension,44 45 with again a more marked reverse gradient among men for obesity.15

Overall, it appears that the pattern of disparities may vary by disease and sex as well as changing over epidemiological time, with stroke perhaps having the most consistent association with poverty. However, global disparities between and within countries are limited by being inevitably contemporaneous, hence the need to consider trends over time.

HISTORICAL CONTEXT

Secular trends

Congruent with the disease-specific patterns of disparities between countries by national income, there appear to be disease-specific secular trends accompanying economic development over historical time, that is, with epidemiological time. First, the epidemiological transition of stroke usually involves two opposite trends; haemorrhagic stroke declines, whereas ischaemic stroke increases. This trend was observed for stroke mortality over the 20th century in a long-term developed country, the UK,46 and more recently for stroke morbidity in one of the most developed areas of a rapidly developing country, Beijing, China.47 Second, recent economic development (via economic growth or migration) appears to be associated with diabetes in men and women, with little change in the sex ratio over time.48 This is currently the case in Asia, with its history of recent economic development.49 It is also evident for migrants from economically less to more developed locations, such as Hispanics in the USA50 or southern Europeans in Australia.51 Whether such an epidemic is a permanent attribute of that population or the product of history is difficult to ascertain. It is perhaps noteworthy that following the Jewish exodus from Russia in the late 19th century to the north-eastern US cities, Jews were in the early 20th century transiently more prone to diabetes.52 53 Third, the increase in IHD mortality only occurred after many decades of economic development. In the USA and the UK economic growth started in the late 18th century,11 54 but only in the early 20th century did an epidemic of IHD take place.55 56 Similarly, in sub-Saharan Africa, in some Chinese populations with a short history of economic development and in first generation migrants from economically less to more developed locations, IHD has not risen strongly, although...

Figure 1  Mortality from stroke (left) and ischaemic heart disease (IHD) (right) by national income (gross domestic product (GDP) per capita).

Figure 2  Morbidity from diabetes (disability adjusted life years (DALYs)) by national income (gross domestic product (GDP) per capita).
diabetes has. Moreover, when this epidemic of premature IHD does occur, it is usually limited to men, for reasons that are unclear.

Coincident with these secular trends in disease patterns with epidemiological time, there are also long-term, though not simultaneous, secular trends towards greater height, larger frame size, and earlier age at sexual maturity. Greater height, and particularly longer legs, is usually associated with lower risk of common non-communicable diseases, such as stroke, diabetes and IHD. However, again these associations appear to vary with disease and epidemiological time. Congruent with the consistent negative association between stroke and national income, this negative association is most consistent for stroke, and less consistent for diabetes, IHD or its component risk factors, particularly in populations with a short history of economic development. Height also has incoherent associations with IHD at individual and population levels.

Disparities over historic time

In contrast to the secular trends observed over epidemiological time since the start of economic development, information concerning changes over historical time is less well documented and may relate to periods preceding economic development when the infrastructure to collect such information was not in place. Health disparities are usually considered to have been consistent over historical time, although their extent in Europe before the start of the Industrial Revolution is unclear, and virtually unknown elsewhere. However, disease-specific secular trends alone would be expected to produce varying patterns of disparities. Historical evidence concerning disparities is limited. Nevertheless, consistent with the change in the pattern of disparities with epidemiological time, there is some evidence that racial and socio-economic disparities have changed over historic time. Diabetes was well known in Europe and the USA in the 19th century. At the end of the 19th century black people in the USA had lower rates of diabetes than white people, a reversal of the current situation. In the 20th century, the pattern of social disparities in male IHD reversed from an association with higher social position to an association with lower social position, although this initial positive association with higher social position to an association with cardiovascular disease or diabetes may be restricted to the advantaged, giving a reverse social gradient, which later reverses as the benefits of economic development spread to less advantaged sections of populations. Similarly, in low-income countries larger, and fatter, body sizes may be valued when restricted to the more advantaged, but less valued when more widely accessible. Moreover, these social constructs and social norms concerning body size may be different for men and women, such that the social gradient changes more quickly for women than men. Finally, lifestyle changes in diet and physical activity with economic development may have a more immediate impact on glucose metabolism and diabetes than on IHD. The weaknesses with an explanation essentially premised on access to resources and social constructs are fivefold. First, the pattern of these diseases is heterogeneous over epidemiological time, necessitating a specific, but as yet unknown, sequence of lifestyle changes. Second, the pattern of change with epidemiological time varies with sex, necessitating sex-specific changes in lifestyle perhaps driven by changing social constructs of body shape. However, in long-term developed countries some of these changes took place largely before the current epidemic of obesity. Third, these diseases may share risk factors, but that does not imply they share causes. Fourth, this explanation in terms of lifestyle gives no insight into why there have been extended secular trends over time. Fifth, a lifestyle explanation may also imply that a common infection does not underlie social disparities, because the same group of people, most likely, are the first ones in their societies to experience socio-economic development with access to associated amenities (including running water) and to experience diabetes and heart disease, suggesting that infection is an unlikely causative agent, although more plausibly so for haemorrhagic stroke.

Other possible reasons for the incoherence of individual-level and between-population disparities globally and historically may provide insight into how disparities have originated and more importantly can be prevented, particularly in the developing world. It is becoming increasingly evident that there may be non-genomic biological mechanisms that transmit environmental influences on health across generations, that is, intergenerational effects. Such mechanisms provide a framework that would generate changes over many generations, with patterns of health varying with the number of generations since the start of economic development (eg, the Industrial Revolution or other similar watershed developmental milestones in more recently developed or developing economies), because of the long-range effect of living conditions in earlier generations. Such intergenerational effects would also generate effects contingent on the stage of the epidemiological transition. Below, it is tentatively suggested how framing within a historical context, including intergenerational effects, may provide fresh insights concerning disparities in some major non-communicable chronic diseases.

EXPLANATIONS

There are several explanations commonly given for the incoherence of individual-level and between-population disparities in these diseases, globally and historically with economic development. Global disparities may be explained by genetic differences, the thrifty genotype or the thrifty phenotype. To date, the genetics of diabetes and IHD are still being explored. Promising candidate genes have been found, mainly among European populations, although whether these also explain global disparities is not yet known. In contrast, there is less genetic evidence for the thrifty phenotype hypothesis, which has also been challenged from anthropological and racial perspectives. Moreover, genetic explanations do not provide any insight into disparities within countries. It is also unlikely that the disadvantaged are also genetically disposed towards these diseases. Arguably the disadvantaged might be more liable to environmental change over their lifetimes, making them more vulnerable to environmental ‘mismatch’ with its hypothesised consequences. However, evidence for that hypothesis mainly comes from small animals. In humans, social mobility, which equates to life course mismatched environments, carries no risk for cardiovascular disease beyond that of its component parts.

Reversals in the social gradient of disease with economic development over historical time are usually explained in terms of inequalities in access to the benefits of economic development and changing social constructs of the value of body size. In low-income countries, access to the lifestyle (more food, smoking, physical inactivity) associated with cardiovascular disease or diabetes may be restricted to the advantaged, giving a reverse social gradient, which later reverses as the benefits of economic development spread to less advantaged sections of populations.
A socio-biological thesis

Well-documented and extended changes over epidemiological time in height, frame size and build suggest that consideration of intergenerationally and environmentally driven changes during growth to the associated somatotropic and gonadotrophic axes might provide a starting point for a socio-biological explanation of long-range changes in chronic diseases. Moreover, any associated differences in chronic disease risk factors at the termination of growth would be expected to track into adult life. The biological mechanisms underling intergenerational changes to these axes are not yet entirely clear, although the recent rediscovery of non-Mendelian heritable changes in gene expression driven by the environment in previous generations, that is, epigenetics, provides a candidate. Moreover there also appears to be some preliminary evidence that intergenerational growth conditions may permanently upregulate offspring hormone levels. Similarly, hormone levels appear to increase with epidemiological time. Very speculatively, figure 3 shows schematically how the same exposure (improved living conditions over generations or historical time) might act on different biological axes (somatotropic vs gonadotrophic) to generate specific patterns of diseases and disparities over epidemiological time.

Somatotrophic axis

Epigenetic mechanisms evolved to maximise maternal-offspring coadaptations, including regulation of postnatal growth. Different mechanisms control prepubertal and pubertal linear growth. When conditions allow, constrained precubertal growth is usually followed by greater pubertal linear growth. In a developing or recently developed population, pubertal growth may respond immediately to better living conditions but leg length only responds over generations. Relatively or absolutely more pubertal linear growth is associated with diabetes and/or insulin resistance, most likely due to changes in the somatotropic axis, that is, on insulin-like growth factor/growth hormone. These effects could be exacerbated by lack of compensating muscle mass from longer legs and larger frames.

In this socio-biological historical framework, the first generation to experience better living conditions during growth could be more vulnerable to diabetes because of compensatory growth at puberty. That first generation will have a reverse social gradient in diabetes, as those who did not experience improved conditions would have no compensatory growth at puberty. Such a pattern has been seen in populations with a very recent history of economic development over one generation, such as China, India, Bangladesh, Pakistan or Egypt. However, this reverse gradient ‘wears off’ with generations of economic development, as maternal constraint wears off, as for example seen in the Chinese population of Hong Kong, who have a slightly longer history of economic development over one or two generations, or in some countries in South America with again a slightly longer history of economic development.

When most people in the population have several generations of economic development behind them the usual social gradient in diabetes is seen. Moreover, all else being equal, rates of diabetes would be lower in long-term developed populations than in populations in transition. However, the countervailing forces of ageing/survivorship and contemporaneous environment in long-term developed countries are currently conspiring to keep population rates of diabetes persistently high. Nevertheless, currently, diabetes prevalence in the recently transitioned populations of Asia outstrips rates seen in long-term developed Western populations. Thus, epigenetic constraint combined with varying levels and histories of economic development may potentially generate disparate patterns of disparities within countries and a disparate pattern of diabetes with national income, because of higher rates of diabetes in rapidly transitioning (and usually middle income) populations.

Gonadotrophic axis

Sex-specific pubertal changes have previously been suggested as contributing to the earlier development of IHD in men, because pubertal sex-steroids act on key IHD risk factors. Pubertal sex-steroids reduce lower HDL cholesterol and increase android fat patterning in boys, with the opposite effects in girls. Whether sexual dimorphism in blood pressure also arises at puberty, similarly due to the action of sex-steroids is less clear. These pubertal changes have generally been assumed to be a universal environmentally independent feature of growth and development. However, sex-differences in HDL cholesterol are contextually specific, suggesting environmental influences. Moreover, experimental evidence suggests that pubertal sex-steroids are dietary and possibly intergenerationally driven; in animals underfeeding reduces testosterone and oestrogen at puberty. Even in universally well-fed human populations in the West, slight changes in diet affect sex-steroids in girls but not in boys.
although larger changes between environments may make a difference. In rats, maternal food restriction during late pregnancy and lactation drastically reduces testicular size, with potential consequences at puberty. As would be expected from a dietary-driven process, the pattern of pubertal development is different in developed environments as opposed to very poor environments, and has changed over historical time.

In this alternative, less Western-centric, conceptualisation, economic development permanently upregulates the gonadotrophic axis via pubertal nutrition with corresponding inter-generational effects. Higher levels of pubertal sex-steroids have sex-specific effects on IHD mediated by central obesity and lipids. Hence, a sex-specific pattern of disparities in IHD would be expected to emerge with economic development. As with diabetes, in the early stages of economic development it is likely to be the advantaged who are initially at increased risk of IHD, hence the reverse gradient in IHD observed for men but less so in women. This sexually dimorphic pattern of social disparities in IHD, obesity and/or lipids, but not diabetes, in the first few generations to experience economic development has been observed in a recent study from a developing population, where the generally protective effect of social advantage was obscured for these risk factors in men. Similarly, there are many studies from long-term developed countries showing a usually unexplained or weaker social gradient for men than women in obesity, lipids and IHD, with particularly marked differences in disparities by sex for populations at an earlier stage of economic development or with a shorter history of economic development. Finally, the implication of this conceptualisation is that higher IHD mortality in men is partly an epidemiologically stage-specific product of growth in an economically developed environment. Seminal epidemiological studies in the 1970s of Japanese men who had migrated to the USA found that ‘Westernisation’ during childhood was a key exposure driving IHD. More recent studies of migration within China have replicated these observations and indicated that they are specific to men. Vulnerability to premature IHD in men may not be an immutable sex difference, but a disparity resulting from socio-economic development.

LIMITATIONS

Although this conceptualisation uses largely known physiological evidence to provide a socio-biological explanation for the observed trends in aggregate and disaggregated disparities in some major non-communicable chronic diseases, key gaps and limitations remain. First, the biological basis for the epigenetic mechanisms that constrain infant and childhood growth have yet to be mapped out, although there is epidemiological evidence of such constraint. Second, it is a supposition that these epigenetic mechanisms act via the somatotrophic axis; however, growth hormone is more important to prepubertal than pubertal growth. Third, it is not clear in this framework why improved living conditions in the first generation should have less effect on IHD, obesity and lipids than on diabetes. However, in that first generation there would not be any additional contribution to the gonadotrophic axis from maternal hormone levels and it is possible that differences in the somatotrophic axis may alter the tempo of pubertal development. Fourth, it is also largely unknown how maternal sex-steroids may affect pubertal sex-steroids in her offspring. Fifth, this framework provides no insight about a key risk factor for non-communicable chronic diseases, that is blood pressure, because it is not clear how blood pressure fits into this framework. There is no clear global pattern to blood pressure. Biologically and epidemiologically it is not clear how blood pressure is affected by puberty. Sixth, this conceptualisation has not considered the effects of interplays between the somatotrophic and gonadotrophic axes during growth, such as upregulation of the gonadotrophic axis increasing peak pubertal mass muscle mass and thereby reducing the long-term risk of diabetes. Seventh, inter-generational upregulation of the gonadotrophic axis with epidemiological time would be expected to have many other impacts on population health and disparities because of the pleiotropic effects of pubertal sex-steroids. Intergenerational increases in pubertal sex-steroids could underlie other features of the epidemiological transition, such as higher rates of female hormonally related cancers mediated by oestrogen. Similarly, higher levels of pubertal sex-steroids would be expected to have sex-specific immunomodulatory effects possibly making men more vulnerable to inflammation and infection, possibly with lifelong effects. Eighth, it is quite possible that with economic development and the associated socially patterned reductions in infectious diseases and increases in life expectancy, the relationship between acute and chronic diseases may itself drive some of the patterns of disparities in chronic diseases. However, it is not obvious why socially patterned survival should produce disease or sex-specific disparities, nevertheless such a possibility cannot be ruled out. Ninth, the present theoretical framework does not provide direct insight into the processes that are currently generating associated epidemics of obesity and diabetes in long-term developed countries, because the focus is aimed at explaining the pattern of social disparities and disease that emerges with economic development. As such the present framework may be most relevant to developing countries, especially those undergoing rapid change. Finally, this conceptualisation in no way negates the multidimensional role of deprivation in generating disparities. Instead, this conceptualisation simply seeks to add a dimension across time and to draw attention to the off-diagonals to generate new hypotheses. For example, it suggests that one exposure (improved living conditions over generations) has long-term effects on two different biological axes. Coincident changes to these two axes may generate a myriad of non-causal associations.

Several of the limitations concerning the hypothesised biological mechanisms could be addressed by a research programme using animal models supplemented by observations of adolescents. To the authors’ knowledge, no previous study has monitored in small animals, taken from a constrained environment, what changes in size, build, metabolic parameters, immune function and to the gonadotrophic and somatotrophic axes occur with each successive generation of exposure to a more ‘Westernised’ environment for a varying number of generations. Such an experiment would test the present hypothesis by mimicking different trajectories of economic development over a fairly short timescale. Alternatively, tracking size, muscle mass, metabolic parameters, immune function and sex-steroids through puberty in a well-characterised birth cohort whose members had different family histories of economic development would also provide insights within a reasonably short timescale. Furthermore, evidence from such studies could be used to inform policies to reduce disparities.

Public health implications

This conceptualisation represents a move away from black box risk factor epidemiology, with its tangled web of associations, towards a more aetiological approach and thus the prospect of more focused interventions with potentially a better understanding of any unintended consequences. Most obviously, it
focuses attention on puberty as a key window of intervention to prevent disparities arising. However, it also suggests that one generic intervention will not necessarily have similar effects across all non-communicable chronic diseases. Instead, any intervention should be grounded in a contextually specific assessment of which non-communicable chronic disease is forecast to be most salient for the current generation of children, backed up by a detailed assessment of their current living conditions and any potential impacts on long-term health. Specifically, a comparative examination of the effect of pubertal nutrition on body composition and chronic disease risk factors in children growing up in the developing world would be informative. Moreover, it is also quite possible that such an assessment may indicate that sex or disease-specific policies are warranted, for example, a focus on boys or lifelong diabetes prevention may have the greatest impact. Systematic surveillance programmes monitoring children’s lifestyle and health in developing countries, alongside focused investigation of specific hypotheses, by global organisations, such as the WHO, would contribute to the global evidence base for development. As always, countries supporting the economic development of developing populations should carefully assess, a priori, the anticipated impact of their policies on long-term health and tailor their interventions to generate an environment that promotes optimal child growth. Although it may be tempting or even justifed after thorough deliberation that, in the short-term, economic growth and associated gains in life expectancy could or should be traded off against disparities and the long-term health of today’s children, such calculus must be made explicitly and honestly based on a full understanding of potential environmental impact during growth. Finally, the authors caution that one must be mindful that the theoretical construct as presently proposed requires further epidemiologic confirmation and laboratory-based validation, before it should be relied on to formulate large-scale interventions or policies. Given these considerations and the changes in nutrition and physical activity that have already occurred in some developing countries, immutable changes in risk may have already taken place, which perhaps need to be addressed by healthcare provision as well as prevention.

CONCLUSION

Despite its tentative and speculative nature, taking a broad brush approach, this study demonstrates the value of framing non-communicable chronic disease disparities within a global and historical perspective. It also implies that evidence from the Western world at its particular epidemiological stage is insufficient to interpret the global pattern of disparities, because patterns of disparities are epidemiologically stage-specific. As such, mounting evidence from the developed world is largely uninformative for the developing world, and instead there is an ethical imperative to focus research efforts on developing countries. In particular, examination of south—rather than north—south, similarities and contrasts, may be uniquely informative in identifying how economic transition can be achieved while minimising social disparities. Perhaps most importantly from a practical point of view, further investigation of the role of pubertal exposures in the development of IHD and diabetes is urgently needed, so that the long-term consequences of the nutrition transition across life can be assessed and any potential risks averted before irrevocable changes have taken place throughout the developing world.


