Chronic illness: a revisionist account

David Armstrong

Abstract This article challenges the generally accepted thesis that the emergence and dominance of chronic illness over the last half century is due to the receding tide of acute infectious diseases and an ageing population. Instead, through an analysis of contemporary reports in the Journal of the American Medical Association, it is argued that the construct of chronic illness emerged as part of a new focus on the downstream consequences of disease and as a means of transferring what had been seen as the natural processes of ageing and senescence into an explanatory model based on pathological processes. The widely accepted idea of an epidemiological transition in illness prevalence has served to conceal the ways in which medicine has extended its remit and suppressed alternative explanatory frameworks.

Keywords: chronic illness, chronic disease, epidemiological transition, ageing, death

In 1971 Omran advanced the concept of the epidemiological transition to explain the changing spectrum of diseases to be found in populations over historical time. The Age of Pestilence and Famine was followed by The Age of Receding Pandemics and finally by The Age of Degenerative and Man-Made Diseases. For western countries this transition was marked by the decline of acute and infectious diseases during the 19th and early 20th centuries, which, together with an ageing society, replaced mortality as the primary index of health with morbidity in the form of ‘degenerative and chronic disease’. This change can be illustrated by examining the frequency of use of the term ‘chronic illness’ (and chronic disease) in books during the 19th and 20th centuries (Michel 2011) (see Figure 1): it appeared with increasing frequency during the 1920s and 1930s, becoming, by the 1950s, as Omran showed, the main health problem of the post-war western world. In effect, the tide of acute and infectious diseases had receded just as the population had become older resulting in a fundamental change in the patterning of illness.

Omran’s thesis has proved influential (with over 1700 citations) as a means of explaining the emergence of chronic disease/illness during the inter-war period of the 20th century (Weisz 2011) and its growing prevalence during the second half of the century (Weisz and Olszynko-Gryn 2010). The sociology of health and illness, with its frequent investigations of the personal and social impact of chronic illness, has accepted Omran’s argument, implicitly or explicitly, to explain the emergence of this new form of sickness during the 20th century. Taylor and Bury (2007), for example, in a highly cited overview of chronic illness, identified ‘the demographic and epidemiological transitions that have increased the prevalence of chronic illness throughout the world’ (Taylor and Bury, 2007: 40). Or, as May’s opening essay in the new journal Chronic illness argued, it was the ‘epidemiological explosion in chronic illness’ that had ‘disturbed traditional forms of clinical practice and brought about new opportunities for involvement by patients in managing their own problems’ (May 2005: 16). Indeed, the high prevalence of chronic illness in the second half or the 20th century and beginning of the 21st century has offered sociologists a
platform from which to examine and theorise a wide range of aspects of personal and social engagement with illness.

Reliance on the epidemiological transition to explain the emergence of chronic illness, however, implicitly endorses its central assumption; that the diagnostic labels of any period reflect an underlying biological reality and that data in the historical record result from a medical perception largely uncontaminated by contemporary world-views or theoretical frameworks. This article, however, explores these morbidity changes through reports of clinical observers who described what they could perceive of disease and illness at various times in the 20th century. It is argued that clinical practice at the beginning of the 20th century recognised a separate domain of age-related physical and mental changes alongside those of disease processes; over the subsequent half century this world of inevitable decline and decay was reconfigured and incorporated into a model of pathological disease. This transformation engendered a new diagnosis of chronic illness as well as a reconstruction of the relationship between life and death.

**A methodological note**

This article draws on material published in a leading US medical journal, JAMA. JAMA was first published in 1883 and all issues have now been digitised and made word searchable. The method used in this article involved first searching for the terms ‘chronic illness’ and ‘chronic disease’. The broad historical patterning of these terms was confirmed using Google’s Ngrams (see Figure 1). Individual texts were then scrutinised to identify the context of the term’s use and its implied meaning. This initial exploration also identified a number of related constructs (such as ‘disabling illness’ and ‘degenerative disease’) which were treated in the same way. Finally an historical narrative was constructed using illustrative quotations from the JAMA database. The trustworthiness of the resulting account can easily be checked (or challenged) as the JAMA database is publicly accessible.

**Pathological medicine in the 19th century**

According to Foucault, in the 18th century the physician’s task was to identify particular diseases in the ‘garden of species’ (Foucault 1973). Just as the gardener could identify flowers
by their outward appearance – such as their shape, colour and smell – so the physician could use the symptom picture to identify disease types. The emergence of the new medical paradigm or episteme of pathological medicine at the end of the century, however, introduced a new way of seeing disease that localised it to a pathological lesion inside the body. This lesion was not an abstract ‘essence’ but an organic form with its own (pathological) life. At times this pathology could be directly visualised (for example, in dermatology, in surgery or in the post mortem) but mostly its character was inferred through the symptoms it induced in the patient and the signs which the attentive physician could elicit. Sometimes the pathological process might be brief and self-limiting; at other times it might unfold over months or even years. It therefore became useful to describe diseases as either acute, having a relatively short life, or chronic, having a longer one. Bright’s disease (of the kidney), for example, could present in either acute or chronic form (Flint 1886).

When faced by an acute disease the clinical task was relatively straightforward as the symptoms and signs could be read in a cross-sectional manner. For chronic disease, on the other hand, the physician had to identify the disease without necessarily knowing at what stage in its pathological life it was presenting to the clinical gaze. A discussion of chronic disease was often concerned with this clinical dilemma, frequently contrasting acute and chronic presentations of the same disease. There was no direct correspondence between chronicity and severity: the duration of a disease could give no indication whether it was serious or minor – indeed a chronic disease could be ‘essentially harmless’ (Watson 1886). Nor did the extended timeline of a chronic disease indicate it was permanent: ‘In no sense of the word can chronic disease be distorted into meaning incurable’ (Pope 1898: 1009). A chronic disease could remain stationary or quiescent, not endangering life, while an acute one could kill, and vice versa. Indeed, when a chronic disease was serious it needed to be described as such, perhaps a ‘serious chronic disease’ (Manley 1892: 186) or a ‘grave chronic disease’ (Cutter 1894: 457).

In summary, for the pathological medicine of the Clinic (Foucault 1973) the depiction of the time course of a disease did not reflect anything more than another characteristic of the disease in question: ‘the only difference between the acute and the chronic disease is in its duration’ (Paine 1900: 1058). To be sure, chronic diseases could be grouped together but no more so than could diseases which caused pain or those which presented with fever. In the 20th century, however, this limited use of chronicity was transformed as a new ways of perceiving disease and practising medicine began to undermine the old Clinic.

The emergence of disabling illness

The construct of chronic illness appeared once in JAMA in the 19th century and nine times in the first decade of the 20th century (whereas there were over 200 mentions of chronic disease in the same period). Thereafter use of chronic illness grew rapidly, particularly from the 1930s. But why, while chronic disease was in common use, was there any need for a new term? The key difference was that chronic illness incorporated a new attribute that can be identified in the contemporary emergence of another construct, disabling illness. The descriptor disabling had been in use in the 19th century but usually in the context of injury (‘bruises and sprains may be prevented by prompt and intelligent treatment from developing into serious and disabling injuries’ [Pilcher 1897: 820]). In the new century, however, disabling illness was for 30 years, between about 1925 and 1955, an important focus of medical attention.

Whereas chronicity referred to a property of the pathological lesion, the descriptor disabling applied to an effect of the lesion on the patient. Symptoms, such as the duration implied by chronicity, had been primarily used by the physician to make a diagnosis in that pain,
chronicity, nausea or fatigue and could be used to try and differentiate between possible underlying pathologies. But the new 20th century focus on disabling symptoms indicated a shift in pathological direction; symptoms might indeed continue to provide valuable diagnostic information but they might also indicated the downstream consequences of the lesion. Here then were two separate analytic spaces: it was the particular consequences of symptoms that made them disabling but it was the underlying pathological lesion that underpinned symptom expression. The condition that had a disabling effect could be described either by summarising the patient-experienced symptoms as an illness or in terms of the underlying lesion as a disease. In those early decades of the 20th century, therefore, the constructs of chronic disease and chronic illness became interchangeable. Although each reflected a different emphasis on patients’ experience or the characteristic of the underlying lesion producing that experience, they both were bound together through their association with disabling illness and disabling disease.

Engagement with how the pathology (the disease) affected the patient (the illness) was reflected in a widening use of the disabling descriptor. ‘Disabling symptoms’, ‘disabling complications’, ‘disabling pain’, ‘painful and disabling affections’ and ‘disabling, distressing consequences’ were all characteristics of the patient who could be described as having a ‘disabling condition’, a ‘disabling illness’ or a ‘disabling disease’. It was but a small step for chronic illness (and chronic disease) to assume the characteristics and patient focus of disabling illness. Indeed, the term chronic disabling illness appeared in the 1930s but the elements of the emerging concept had been accumulating over the previous three decades. There was a clear association between an illness being disabling and having long duration, so when Evans discussed the disease of chronic brucellosis in 1934 she noted that there was little information in the literature ‘about the disabling chronic form of the disease’ (Evans 1934: 666). And when Perrott and Holland reported the overall impact of chronic disease in 1937 they explicitly linked both chronic disease and chronic illness to disabling illness (Perrott and Holland 1937).

Once the attribute of disabling had been added to that of duration in chronic illness and chronic disease the use of the separate disabling descriptor became redundant (though the notion of disability would resurface in the 1970s as a key element in later medical classification systems). The label of chronic disease in the inter-war years therefore implied a significant impact on the patient:

Chronic organic disease usually necessitates a change in the patient’s mode of life, makes life more complicated and often less pleasant, and thus forces on him an adjustment to lessened activity and interests usually quite at variance with his former life. (Casamajor 1920: 471)

**Spatialising illness across the population**

Chronic diseases of the 19th century, like all pathological lesions, were distributed within the three-dimensional confines of the patient’s body; they might have long duration but they were still diseases with pathological form. Chronic disease/chronic illness of the 20th century, however, began to present itself not by its pathological type but by its disabling qualities and these appeared beyond the body, in social interaction and in population spaces. If its disabling impact was one of its defining characteristics this could not easily be identified or assessed using the usual tools of clinical medicine such as a stethoscope, a blood test or an X-ray. Instead new technologies emerged to capture this novel medical construct and, given the population framework of this new medical object, it propelled epidemiology and public health
from their former sanitary realm to engage with these new illnesses/diseases with disabling attributes.

The new way of practising medicine did not involve the close examination of the body of the patient in the Clinic but of the body of the population through the survey. And just as the 19th century physician reported the patterning of signs and symptoms accompanying a pathological lesion, so the 20th century physician reported the population distributions of chronic illness/chronic disease/disabling disease. Numerous health surveys of the population were carried out in the inter-war years, from the small United Charities of Chicago survey of 184 families under treatment for an extended period of time (Bedford 1926) to the celebrated National Health Survey conducted in 1935–1936 by the US Public Health Service (Weisz 2011) which estimated that 20 million cases of disabling illness would occur in the population (National Health Conference report, 1938). Giving evidence to a Congressional hearing in 1939 Leland claimed to know of 4,000 such surveys (Leland 1939).

By the middle of the 20th century the old chronic disease of the 19th century Clinic had almost disappeared. In its place was a new form of morbidity that was based less on pathology and more on a patient’s capacity to function. Chronic disease, it was observed, ‘may be defined as any deviation from health that affects a person’s total life pattern in a significant way because of either duration or prolonged after-effects’ (JAMA 1950: 466). It could be called chronic illness to respect the ways in which collections of symptoms impacted on patients’ lives or it could be called chronic disease to capture the pathological forms which had these effects. The emphasis on functional capacity, however, marked the beginning of the end for a medical regime that equated health with the absence of (pathological) disease. ‘Let us redefine health’ argued Stieglitz, ‘as having quantitative attributes and perfect health as that state of being in which all the functional capacities of the organism have maximum reserves’ (Stieglitz 1952: 482), a position that found echo in the contemporary World Health Organization’s definition of health. But this redefinition of a population’s morbidity burden was accompanied by a parallel change in the conceptualisation of ageing, which extended and deepened the reconfiguration of chronic disease in the 20th century.

The pathology of ageing

The medical revolution at the end of the 18th century, which localised disease to an intra-corporal lesion, also reconstructed death (Foucault 1973). When diseases existed in the garden of species there was no necessary connection with mortality: death was a natural event that came from outside life (and the body). Even had a post mortem been carried out there was no more chance of identifying this cause of death than there would be in dissecting the corpse of a patient today to identify the cause of their suicide or depression. The new pathological medicine, however, linked death inextricably with life and disease. The localised physical lesion was pathological because it caused death; natural death was therefore replaced by pathological death which came from inside the body. By opening up a corpse it became possible to give death a name, and subsequently enter this cause of death on the newly introduced death certificate.

Throughout the 19th century, despite the increasing penetration of pathological medicine, elements of the old form of natural death continued to haunt clinical practice. This could be seen in the frequent use of natural causes such as ‘old age’, ‘senility’, ‘atrophy’ and ‘debility’ on death certificates up until the end of the century, as well as the continuing belief that, especially for the older patient, pathology was interpolated into a ‘natural’ ageing process characterised by decay and involution. This idea of in-built senescence meant that clinical medicine had to address two domains: one was the diagnostic puzzle of the pathological and
the other was the natural and inexorable degenerative processes leading to senility and decrepitude. For the younger patient, death was usually the consequence of pathology, but for the older one it was more often the result of the inevitability of natural decline and decay.

Over the first half of the 20th century the line that separated the domains of pathological processes and natural ageing became blurred. And in the same way as disabling illness had performed the role of vector in transferring new meaning to chronic illness so a new transition concept, degenerative disease (and degenerative illness), facilitated the incorporation of the ageing process into the pathological framework. Figure 2 shows its rapid growth in *JAMA* in the early 20th century.

The significance of degenerative disease lay in the fact that the exact status of such a disease was uncertain and the label could therefore mediate between ideas of natural ageing and the pathological paradigm. Degeneration implied processes of decay that placed a natural limit on human life:

> The statements that old age can be deferred have no more scientific truth in them than the widely advertised promises of rejuvenation. Senescence is a normal process of involution as necessary to the progress of life as is the normal process of growth. It is intrinsic, inheritable, fixed in the germ plasm through the action of all of the forces of the Universe. (*JAMA* 1928a: 2067)

To relabel these ageing changes as degenerative disease simply redescribed what they really were but, at the same time, it introduced a new tension between pathological processes which could be addressed by medicine through accurate diagnosis and appropriate treatment (the disease part) and an in-built ageing process that was regressive and irreversible (the degenerative part).

During the early decades of the 20th century the new degenerative label opened up the possibility of the reclassification of all those processes that had belonged to the natural course of ageing and decay. Osteoarthritis, for example, was identified as largely if not entirely a degenerative disease (Rolleston quoted in *JAMA* 1928, Swaim 1934). Degenerative disease, moreover, engendered a vulnerability to pathological disease as ‘over 135,000 lives are taken by pneumonia, chiefly as a result of weakened resistance from degenerative disease’ (Rittenhouse, quoted in *JAMA* 1912: 2087). Yet medicine could also pluck pathological disease from this sea of decay: poliomyelitis ‘for years was considered to be a degenerative disease of the nervous system,

![Figure 2 Occurrence of ‘degenerative disease’ and ‘senescence’ in JAMA](image-url)
and which recent researches have proved to belong among the infectious diseases’ (JAMA 1912: 2157).

The process through which the senescence of ageing was wrested from nature and located in a degenerative space and the further transformation of that space into a pathological one can be illustrated by the great 20th century health problem of heart disease. For the medicine of the Clinic, failure of the heart mainly belonged to the ageing process:

The statement to the effect that everyone will die from arteriosclerosis if he lives long enough is probably true, and by the same token many deaths in the later decades of life are caused by arteriosclerotic heart disease. It is intended to include in this category only those cases in which the heart disease results from the arteriosclerosis of senescence – i.e. the degenerative type – and to exclude those in which an existing arteriosclerosis is secondary to vascular hypertension. (Stone 1934: 153)

In effect, in the inter-war years two types of heart disease could be identified: ‘one occurs as a phase of a general arteriosclerosis; it is part of the process of senescence. It is encountered in the elderly’; the other occurred in younger patients for whom the ‘strain of modern business and professional life leads to an early degenerative changes in the coronary arteries’ (Boas and Donner 1932: 2186). In terms of laboratory results, clinical testing and post-mortem findings these types of heart disease were indistinguishable; their difference lay solely in their putative cause. Indeed, after successfully treating all heart disease that had a pathological basis, especially in the young, it would still leave the degenerative form to affect later life:

[If prior to the age of 60 there is a decrease in deaths from all causes, there will be a proportional increase in the number of deaths from arteriosclerosis and consequently from arteriosclerotic heart disease. Arteriosclerosis, as we know it, is bound by the laws that govern the average age of a species. It is ultimately inevitable, therefore, and not subject to change; (there could) be no biologic antidote. (Stone 1934: 153)]

By the early 1930s degenerative disease of the coronary arteries had become a significant cause of death. But if degenerative disease was just another name for changes associated with natural ageing then an increase might be expected as the population aged:

[Increase in life expectancy in the past few decades results in a larger number of the population reaching the age at which degenerative disease is to be expected. This raises the question as to whether the increase in heart disease is a relative or absolute increase. (JAMA 1937: 154)]

The question was not easily answered as the extent of degenerative heart disease depended on how frequently the label was applied:

Among peoples in whom heart disease takes a somewhat lower rank it is noteworthy that a large number of deaths are reported as due to ‘senility’ or ‘decrepitude’. If these deaths were more accurately classified, heart disease would have to be accepted as the leading cause of death in the Western world and a serious aspirant for leadership in the East. (JAMA 1928a: 2068)

In other words, the reported increase in the pathologies of old age ‘such as cancer and cardiovascular disorders, especially degenerative disease of the myocardium’ (Staehelin, quoted
in *JAMA* 1938: 60), could simply be an artefact of diagnostic practice. The reason for this confusion lay with the degenerative disease label:

The blame must be borne rather by the degenerative diseases, whose obscure etiology makes uncertain any measures of prevention. The fact that it is not the acute heart diseases of early and middle age but rather the degenerative diseases of senescence that are making the mischief is forcibly demonstrated in a recent analysis made by the Pennsylvania Department of Health. (*JAMA* 1928b: 2068)

Towards mid-century the degenerative label became less acceptable and increasingly it was prefaced by ‘so-called’:

The major chronic disorders associated with senescence, both in incidence and in pathogenesis, are the so-called ‘degenerative diseases’ ... the so-called ‘degenerative diseases’ are characteristically progressive, leading slowly to functional depreciation. (Stieglitz 1952: 481–2)

As the label was applied less so the prevalence of degenerative disease declined. The causes of sudden death reported in 1942 had already relegated degenerative disease of the heart muscle to third place behind coronary artery disease and acute coronary occlusion (*JAMA*, 1942). But even this residuum began to yield to a pathological perspective as it was challenged by the logic of causal analysis.

**The causes of disease**

There was no external cause for the process of ageing and decay as it was a result of the in-built obsolescence of the human body; this was simply the innate destiny of all human cells, organs and tissues. If, however, degenerative disease was perceived as having a variable cause then its predetermined status was at once undermined. So while diseases such as hypertension, myocardial degeneration and arteriosclerosis might be labelled as ‘of unknown origin and of the chronic degenerative type’ (Langstroth 1929: 1607), the observation that there were variations in the resilience of individuals to these afflictions that might be due to, say, living habits (Ferguson 1928) marked the appearance of a crack in the edifice of natural ageing. Diet, for example, might be able to ‘prevent and alleviate much of the so-called degenerative disease’ (Sherman 1931: 1426). In other words, as degenerative diseases extended their reach they also began to metamorphose into preventable and/or treatable, even curable, pathological diseases.

By the mid-century mention of degenerative disease reached a peak in *JAMA* and then began to decline. The degenerative label implied inevitability and there were emerging indications that degenerative conditions could be treated. Indeed, the identification of potentially modifiable causal factors implied they were not part of some fatalistic process of decay. As Wynder and Day noted, ‘Chronic diseases are not likely to be due to one specific factor, but rather are a result of multiple factors usually operating over a long period of time’ (Wynder and Day 1961: 997). Claims to be able to treat or prevent degenerative disease were at first tentative but the very fact that they were being considered marked the final separation of degeneration from its ‘natural’ origins. ‘There are no satisfactory data’, noted Darling and his colleagues to predict that the physically fit subject will have a longer life span, a lower incidence of degenerative disease or a greater resistance to infection than his unfit confere, although these relationships might well be studied further. (Darling *et al*. 1948: 766)
In the second half of the 20th century there was no need for a degenerative disease label as the ageing process had been colonised by pathology:

It has become apparent in recent years that atherosclerosis, particularly of the coronary arteries, can no longer be regarded as the natural result of old age or as a medical curiosity when it occurs in the early decades of life. (Glendy et al. 1937: 1775)

At the beginning of the century ageing and pathology/chronic disease had been in opposition, components of alternative explanatory frameworks, but by mid-century chronic disease had replaced ageing. At the same time ageing itself was reconfigured as simply another risk factor in the overall explanation for why this had occurred:

We no longer consider the chronic diseases – heart disease, hemiplegia from cerebrovascular accidents, cancer, diabetes, arthritis, and the many other long-term illnesses – as inevitable concomitants of age. The former notion that these represented some mysterious ‘degenerative’ process about which the physician could do very little is giving way to a more hopeful and a more realistic view. (Breslow 1956: 1364)

The very fact that the ageing process encoded in the concept of degeneration was perceived to have a cause was the key to mapping the construct of chronic illness onto the space of inevitability, decay and death. But this mapping was further facilitated by the emergence of a multi-causal property space in which to embed chronic illness. This new causal framework infiltrated the rest of medicine in the second half of the 20th century but both chronic illness and the reframed chronic disease were the main vehicles for mapping the contours of these ‘risk factors’ and ‘multifactorial aetiology’. Just as many of the causes of chronic disease materialised in an external environment far removed from the ‘milieu interieur’ of pathological medicine, so the effects of chronic illness were located further downstream in patients’ disability (and, a few years later, in their activities of daily living and health-related quality of life). Moreover, the hospital, especially the hospital bed, which had been the temple of pathological medicine, was usurped by a population surveillance perspective that saw the increasing decline of the hospital in-patient as the prime locus of health care delivery. The plea for a Department of Continuing Care to be attached to every hospital (Bluestone 1947) signalled the new direction of healthcare delivery.

Throughout the history of JAMA its obituary columns recorded the deaths of physicians together with an analysis of the causes of those deaths. The rise and fall of myocardial degeneration as a cause of death is shown in Figure 3. Dr John Connor was the last physician to die from this cause, in 1972; thereafter it was no longer possible to die from natural causes.

The age of chronic illness

By the second half of the 20th century the transformation of chronic disease – and of medicine – was almost complete. The 19th century medicine of the Clinic had focused on a pathological lesion. It was located in the patient’s body; it was caused by a pathological process such as inflammation and irritation; it produced, fortuitously, the symptoms and the signs that would enable it to be identified and, classically, it was analysed in the neutral space of the hospital. The problem of chronic illness was completely different. It was not located in a corporal space but a population one, its causes were more likely to be external than internal and multiple
rather than single and its locus of treatment was not the in-patient bed but the ongoing care offered the ambulatory patient.

The advent of chronic illness stabilised a temporal space for clinical practice in which illnesses moved and physicians worked, a space through which disease and its effects could be surveyed and managed during its temporal course. This space allowed the deployment of technologies undreamed of in the Clinic: the survey, case detection, screening, prevention in its primary and secondary forms, continuity of care and the long-term surveillance of illness (Breslow 1956, Levin 1951, Roberts and Wylie 1956). This reconstruction of medicine had occurred over barely 50 years as ephemeral constructs such as disabling illness and degenerative disease facilitated the transference of meaning to chronic disease and chronic illness.

As late as 1935 Hedley could note that ‘Deaths previously attributed to senility are now listed as heart disease’ (Hedley 1935: 1407) but by the second half of the 20th century a new narrative emerged that dominated the subsequent analysis of chronic illness. This narrative stressed the medical conquest of acute illnesses, the dramatic decline in the mortality of infectious diseases, the ageing of the population and the subsequent rise of chronic illness as a major health problem (Breslow 1956). ‘Particularly notable’, observed Stieglitz:

[1]s the fact that in 1900 the three most important causes of death were diseases of infectious origin, but in 1951 the three major sources of death were non-infective, so-called degenerative disorders. (Stieglitz 1952: 481)

A few years later this explanation for the shift in medical paradigms was formalised by Omran (1971) in his theory of epidemiological transition which, in locating the new challenge of chronic disease in medicine’s triumphant past (which had extended life expectancy though its defeat of acute illnesses), offered a rhetorical device to ensure a smooth transition from one framework of medical perception to another.

Chronic illness might have replaced acute disease as the leitmotif of the new medicine but this new pattern of morbidity reflected less a change in the nature of the problems presenting to medicine and more a transformation in the classification of those problems. Chronic illness marked a new target for clinical practice as the old Clinic disappeared and a new gaze took its place. It might therefore seem ironic that the medical narrative of continuity that has served to

![Figure 3](#)
obscure the ways in which meanings shifted has also been readily endorsed by medical historians and social scientists for, in accepting the epidemiological transition, a biological rather than sociological account of illness has been prioritised. If, however, the progenitor of chronic illness was not a receding tide of acute and infectious disease but a relabelling of ageing and senility this at once makes visible the social and historical forces that have reconfigured medical reality.

At the heart of the epidemiological transition lies the belief that the material basis of the changing spectrum of diseases can readily be identified from contemporary morbidity statistics, in the diseases reported daily in healthcare settings, and from the experiences of millions of patients. Yet the biological reality of disease cannot be separated from its contemporary description. In the inter-war years myocardial degeneration was a major cause of death (see Figure 3), yet in the 21st century it no longer exists. There was no treatment breakthrough that can account for its disappearance; perhaps, as some commentators have suggested (Spiliopoulou et al. 2005), it never really existed. Or it may have been redescribed (perhaps as coronary heart disease) – but this is simply to replace one label with another. In 50 years’ time it is possible that coronary heart disease will have disappeared and commentators may be tempted to say it, too, never really existed. The ‘reality’ of diseases and illnesses over the last two centuries, in their great variety of forms, might frequently be stabilised by linking the label to a biological referent but it is contemporary medical perceptions and explanatory frameworks that create the analytic space for their first emergence and continued existence.

The major illness problems of the early 21st century, then, did not have their origin in some biological shift in the nature of disease in the preceding century but in the way in which medicine both began to focus on the downstream consequences of disease and asserted the superiority of the pathological paradigm over the ageing process. Medicine used chronic illness to overwhelm the last remaining elements of natural life and natural death and establish a pathological death for everyone. An older explanatory framework was therefore lost as the colonisation of ageing by medicine in the 20th century changed the ways in which getting old, dying and death were construed and analysed. The transformation is not quite complete, as evidenced by the continued growth of an anti-ageing medicine (Fishman et al. 2010, Mykytyn 2010) and recurrent attempts to control the ageing body (Brooks, 2010 Kinnunen 2010); indeed, the emotional, mental and physical changes associated with ageing continue to be transformed into illnesses (Joyce and Loe 2010). Yet it seems clear that the new configuration of life, disease and death has already established a pathological lens to transform the natural ageing process and ensured that the threats of chronic illness have enabled clinical strategies which have opened all of life (and death) to intensive surveillance.

Address for correspondence: David Armstrong, Department of Primary Care and Public Health Sciences, King’s College London, 42 Weston Street, London, SE1 3QD, UK e-mail: david.armstrong@kcl.ac.uk

Acknowledgements

This article is based on a presentation to a workshop on chronic disease in the 20th century held in Paris in 2011 and organised by Isabelle Baszanger, Nicolas Henckes and George Weisz. I am grateful for helpful comments from workshop participants.
Notes

1 As this article is about the changing meaning of chronic illness and chronic disease, reference to the terms hereafter relate to their contemporary usage.

2 The difference between disease and illness was only formalised in the 1970s (Eisenberg, 1976, 1977) but the debates described in this article show the origins of this distinction.

References


Evans, A.C. (1934) Chronic brucellosis, JAMA, 103, 9, 665–667.


Flint, A. (1886) Elements of prognosis in Bright’s disease, JAMA, 6, 2, 29–32.

Ferguson, C.L. (1928) Health in industry, JAMA, 91, 12, 868–70.


JAMA (1950) Editorial. Chronic illness, JAMA, 144, 6, 466.

JAMA (1937) Queries and minor notes. Death rates from arteriosclerosis and heart disease, JAMA, 109, 2, 151–5.


Watson, B.A. (1886) Fibro- or spindle-celled sarcomatous tumors, with the report of a case and presentation of the specimen, *JAMA*, 7, 16, 421–5.

