John Macnicol

The history of work-disability

Conference paper

Paper for the Social Policy Association Conference, University of Lincoln, July 2011:

The History of Work-Disability

by John Macnicol

Introduction: the present debate

Since the early 1970s, there have been marked improvements in the aggregate health status of the UK population. Between 1970 and 2009, life expectancy at birth increased from 68.72 years to 77.11 years for men, and from 74.98 years to 81.35 years for women. Life expectancy at age 65 also increased between 1981 and 2009, though by much less than is often assumed - by 4.3 years for men and 3.5 years for women: its current rate of increase is only 2 per cent per annum. Between 1968 and 2008, age-standardised mortality rates for men and women declined by 51 per cent and 43 per cent respectively, and now stand at their lowest-ever recorded level. This success story needs to be tempered by two reservations: first, there are substantial social class differentials in mortality; and second, self-reported health appears to have worsened (in part, a result of rising expectations), health care utilisation rates have risen in the long term (again, a function of increasing supply of resources, improved diagnostic techniques, earlier diagnosis and new technology) and disability-free life expectancy has become a smaller proportion of total life expectancy. There is also much debate about whether these health gains at older ages will continue in the future.[1] Nevertheless, we can conclude with reasonable confidence that, since the early 1970s, the health of the UK population has improved markedly.

[1] For an example of the 'pessimistic' viewpoint, see Jay S. Olshansky et al, 'A Potential Decline in Life Expectancy in the United State in the 21st Century', *New England Journal of Medicine*, 325, 11, 17 March 2005, pp. 1138-45.

Yet concurrent with these improvements there has occurred a remarkable counter-trend: the number of people claiming disability-related benefits of all kinds has almost tripled since the 1970s, reaching a high-point of 2,740,000 in May 2005, and then slightly declining to c.2,500,000 now. Controversially this level of claims was unaffected by the employment growth of 1992-2008, which caused other claimant groups - notably, lone parents and older people - to return to work in increasing numbers. (Of course, without this aggregate employment growth claims to disability-related benefits might have risen even higher.) On the face of it, therefore, there has been a threefold increase in the number of people considering themselves too sick to work. How is this paradox to be explained?

The causes are complex and inter-related, but essentially four principal ones can be considered. First, it has been argued that the higher monetary value and easier eligibility conditions for Incapacity Benefit (especially at the long-term rate) acted as an incentive to drop out of the labour market, particularly in areas where job offerings are scarce. Second, claimants to all disability benefits tend to be poorly qualified – some 60 per cent have no formal qualifications at all - and thus have sub-optimal employability which makes them intrinsically unattractive to employers. Third, there are issues of health: long-term joblessness undoubtedly worsens health status, and renders individuals less likely to be re-absorbed into employment once conditions improve. The fourth and final factor is lack of sectoral labour market demand, which is especially acute in certain regions. Of course, all of these broad causes interact: for example, the problem may have originated in sluggish labour market demand but it has subsequently been intensified by the damaging effects of long-term joblessness on health and motivation; again, low skills become even more disadvantaging during a period of major economic restructuring, and it is those with the lowest skills that are the first to be shaken out of the labour force during a recession.

In order to understand this, the overall socio-economic context needs to be considered. Since the early 1970s (the OPEC-led oil price shock of 1973 being the catalyst), western economies have experienced profound economic changes, amounting to no less than a second industrial revolution. There has been a massive 'redistribution of work' - from older men to women of all ages, from heavy industry to service jobs, from full-time jobs to part-time, from old industrial regions to new centres of economic growth, and so on. This is graphically illustrated if we look at the changing occupational structure between 1981 and 2006. The proportion of all jobs in manufacturing declined from 31 per cent to 17 per cent (men) and 18 per cent to 6 per cent (women); and those in banking and finance rose from 11 per cent to 21 per cent (men) and from 12 per cent to 19 per cent (women). The slow economic recovery that took place between 1992 and 2008 did little to improve the job prospects of unskilled workers. Virtually all net job growth in the UK economy has been via part-time jobs, which have increased tenfold since 1951 (from 831,000 jobs to nearly 8,000,000 now). In 2010, some 200,000 jobs were created in the UK economy, but the proportion of these that were full-time was only 3 per cent. Qualitatively, job growth has also been in the shape of an hour-glass, with expansion at the top and at the bottom. All of this has led to a polarisation between 'work-rich' households (where at least two adults have jobs) and 'work-poor' households (where nobody has a job).

These collective changes have impacted devastatingly on older men in the UK's traditional industrial heartlands, and at the same time the new service-based jobs (often part-time, feminised, low paid and insecure) have presented new social and psychological challenges. In essence, there are two distinct problems in modern labour markets: deindustrialisation and worklessness, which have left whole communities with little recourse to waged employment; and the growth of new jobs which, being low-paid and insecure, are inadequate to support a

family and symbolise the slow shift from the male breadwinner model to the adult breadwinner model.[2]

The story of work disability since the 1970s is in many ways a metaphor for these changes. It also symbolises other broad themes. One is the whole question of how far economically inactive citizens should be forced, by a withdrawal of benefits, to take any job, at any wage - or to perform, long-term, those low-paid, unattractive jobs that migrant workers are prepared to do in the short-term. This in essence is what 'employability' really means in the present economic climate. Since the recession and stagflation of the 1970s and early 1980s, successive UK governments have adopted a neoclassical economic strategy of expanding labour supply in order to achieve steady, non-inflationary economic growth by exerting downward pressure on wages. In part, this has been a rationalisation of the growth of part-time jobs: there has emerged a prescriptive ideology that all citizens of working age should support themselves through paid labour. But even more important is the fact that the control of inflation has been absolutely central to neoliberal economics ever since monetarism in the early 1980s: low inflation creates a stable, predictable world in which finance capital can flourish.[3] A major purpose of activation policies is therefore to control inflation.

Activation discourses are of course multi-layered and complex, and this underlying macroeconomic motive has been hidden beneath a rhetoric on 'rights and responsibilities', with an increasingly hostile condemnation of 'welfare dependency', and rather simplistic exhortations about the beneficial effects of paid work on an individual (with no mention of the all-important issues of job quality and remuneration). Intriguingly, the New Labour

[2] For an interesting account, see David M. Smith, *On the Margins of Inclusion. Changing Labour Markets and Social Exclusion in London* (2005).

[3] David Harvey, *A Brief History of Neoliberalism* (2007), pp. 22-6.

government (1997-2010) deployed the emancipatory, post-civil rights social model of disability to justify the labour market activation of disabled people, and a tightening-up of benefit conditionality: pushing more disabled people into jobs was presented as 'empowerment'.[4]

Another theme is that in western societies in the second half of the twentieth century there was a growing propensity to define minor health conditions as disabling - often termed 'the medicalisation of everyday life' or the 'cultural inflation' of sickness. This has affected people at every level in society, but it is most problematised when those at the bottom are under consideration. However now the opposite is taking place, and disability is being defined down. It is important to bear in mind that categories like 'retirement', 'unemployment' or 'disability' are essentially constructs of twentieth century welfare states: in the nineteenth century, there were no clear-cut distinctions between them. Now, as welfare states are being cut back, these categories are being eroded. (The move to a ‘universal credit’ is also symbolic of this conceptual change.) Welfare discourses are moving 'back to the future'.[5]

Welfare benefit caseloads in modern industrial societies are complex, reflecting the fact that postindustrial labour markets are highly variegated, with cross-cutting divisions of gender, age, region, skill, sector and so on. Before the current recession pushed formal ILO unemployment up to c.2,460,000, joblessness in the UK was the product of supply-demand mismatches between these many headings. All activation programmes thus face the problem that they have to be precisely tailor-made and personalised to fit the individual

[4] For example, Department for Work and Pensions, *A New Deal for Welfare: Empowering People to Work*, Cm. 6730, 2006, p. 15.

[5] John Macnicol, 'Anti-ageism and the Neoliberalisation of Old Age', *Paper to the ISA World Congress of Sociology*, Gothenburg, Sweden, July 2010.

circumstances of each client. Generalisations are therefore hazardous, but broadly speaking we can divide the population on all incapacity benefits into two categories: first, deindustrialised older men, concentrated in those regions which led the industrial revolution; and second, a diverse and growing group suffering mental and behavioural disorders - more feminised and located in areas where aggregate economic growth is reasonably sustained. Members of the first group are suffering because of a lack of jobs. Their age-profile is older (some 850,000 of claimants to incapacity benefits are aged 55+) and therefore they might reasonably be expected to have moved into retirement or death in ten years' time. (New Labour's 2006 target of getting 1,000,000 of them off benefit by 2016 was in many ways a rationalisation of the inevitable.) With this group, the problem is one of programme duration and lack of outflows from benefit:[6] hence the dramatic, if misleading, soundbite that, after being on incapacity benefits for two years, a person is 'more likely to die or retire than get a job'. The second group are suffering because of the nature of the new jobs and because an increasingly competitive society and labour market has created more social casualties suffering from varieties of alienation. Again, this is a development that can be observed at every level in society, and has attracted much comment. There is also the problem that younger men are now in competition with women for low-grade, service jobs - and, in consequence, women so displaced from the labour market are claiming incapacity benefits.[7] Interestingly, in its last years of office the New Labour government, anxious not to emphasise labour market demand as a causal factor, focused more on the latter group: the new trend in claimants, it argued, was 'away from the stereotype of middle-aged men in the

[6] Michael Anyadike-Danes and Duncan McVicar, 'Has the Boom in Incapacity Benefit Claimant Numbers Passed its Peak?', *Fiscal Studies*, 29, 4, 2008, pp. 415-34.

[7] Peter A. Kemp and Jacqueline Davidson, 'Gender Differences Among New Claimants of Incapacity Benefit', *Journal of Social Policy*, 38, 4, Oct. 2009, pp. 589-606; Christina Beatty, Steve Fothergill, Donald Houston, Ryan Powell and Paul Sissons, 'Women on Incapacity Benefits' (Centre for Regional Economic and Social Research, Sheffield Hallam University, 2009).

industrial heartlands and towards a new generation with manageable mental health or musculo-skeletal conditions'.[8]

Claims to disability benefits have risen in many advanced industrial societies, the actual rate of rise depending upon factors such as eligibility criteria, replacement rates, cultural attitudes towards disability, caseload age-profiles, programme duration, and so on.[9] In the UK eligibility criteria were not significantly relaxed in the 1990s, though there had been in the 1980s a deliberate government policy of pushing some of the long-term unemployed onto Invalidity Benefit (as it was then called), in order to massage the official unemployment figures.

Under New Labour, welfare reform was pursued with renewed vigour after 2006, and attention was paid to the stubborn problem of long-term disability. The new 'Pathways to Work' scheme was gradually phased in, until it became mandatory for all new claimants, Incapacity Benefit is being replaced by the Employment and Support Allowance for all new claimants and a stricter work capability assessment has been introduced. All of this has occurred against a background of growing press criticism of work-disabled people, with claims that a majority of them are capable of work.[10] Conditionality has intensified under the new Conservative-Liberal Democrat government, which aims to push as many of the long-term disabled as possible onto Jobseeker's Allowance (where they will be pressurised much more to accept any job or lose their right to benefit). A major driver behind this change is the perceived need to reduce public expenditure.

---

[8] Department for Work and Pensions, *Reducing Dependency, Increasing Opportunity: Options for the Future of Welfare to Work* by David Freud (2007), p. 28.

[9] Leo J. M. Aarts, Richard V. Burkhauser and Philip R. de Jong, *Curing the Dutch Disease. An International Perspective on Disability Policy Reform* (1996)

[10] Christopher Hope, 'Three in Four Sickness Benefit Applicants "Fit to Work" or Stop Claiming Due to Medicals', *Daily Telegraph*, 28 July 2010.

Defining work-disability

How should work-disability be defined?[11] It is a truism that all disability is an interaction between person and environment. It is equally a truism that sickness is a multidimensional phenomenon, influenced by many contextual and psycho-social factors - of which, labour market demand will be an important one. Work-disability therefore involves a complex interaction between an individual's self-defined state of health and their working environment. It is a phenomenon often dismissed pejoratively as 'malingering' (originally used by military doctors to denote the use of feigned sickness to evade combat duties).[12] However, malingering is better understood as a conscious and calculated attempt at deception, whereas the phenomenon under study here is a much more subtle and complex process whereby reduced labour market demand and/or greater job insecurity causes individuals to take a more pessimistic view of their own health status. Poor health becomes a socially-acceptable reason for joblessness. Whether or not this amounts to a deliberate act of deception is a moot point. Individuals may or may not be aware that the perceived severity of their work-disabling condition is relative to prevailing job opportunities: unravelling the complex layers of conscious and unconscious motivation involved would be a wellnigh impossible task.

Work-disability can therefore be envisaged as having a hierarchy of component parts on each side of the equation, and the interaction between these component parts is complex:

[11] Curiously, work-disability has been relatively unresearched by UK academics, despite its enormous importance. One interesting American study is Edward H. Yelin, *Disability and the Displaced Worker* (1992).

[12] An interesting guide to doctors is Sir John Collie, *Malingering and Feigned Sickness* (1913).

Individual	**Workplace**
Prevailing health expectations	New technology
Doctors' definitional thresholds	Stress of work
Self-referral thresholds	Demand for product
Medical technology	Local economy
Diagnostic techniques	Globalisation
Timing of medical intervention	Workforce downsizing
Functional ability	and the existence of a job

With this in mind, we need to look backwards at how this problem was discussed in the past. From this historical gaze, we can see that the boundaries between 'unemployment' and 'sickness' have always been blurred and that there have been two previous versions of today's debate.

From the late nineteenth century to the First World War

The 1880s marked a turning-point in the economic and social development of the UK. There occurred a number of important structural changes, which interacted in a highly complex way: the amalgamation of firms into larger units of production, increasing international economic competition, a growing emphasis on individual workplace productivity, an intensification of work for all urban workers and, as a consequence, an increasing displacement of older male workers from the workforce. Modern 'jobless' retirement began in the 1880s: whereas in 1881 73.6 per cent of UK males aged 65+ had been recorded by the census as 'gainfully occupied', by 1931 this had fallen to 47.9 per cent, by 1971 it was 23.5 per cent and now it is just under 11 per cent. Before the 1880s, retirement had been

intimately associated with disability, as the Poor Law category 'aged and infirm' demonstrated. Most workers moved to progressively lighter tasks as they aged and became less physically capable, their diminishing incomes often supplemented by Poor Law outdoor relief. Retirement in its modern sense had little meaning: it tended to be confined to a wealthy elite.[13]

By the 1890s, these displaced older workers were attracting much attention. The social literature of the time is littered with concerns that older workers were not sharing in the general improvement in living standards and real wages: economic progress appeared to be passing them by. As one commentator (a prominent friendly society member) argued in 1896, it was now 'recognised by all students of industry' that

> improved methods of production, the introduction of machinery, the competitive stress involved in the fight for the world's markets, have placed old age, inconvenienced already by its natural disadvantages, at an ever-increasing discount.[14]

Modern urban industry appeared to 'age' workers more rapidly than did traditional rural society. Charles Booth's extensive researches discovered that urban workers appeared to 'break down' earlier than did agricultural workers. As he graphically put it,

> In one way or another effective working life is ten years longer in the country than in the town, or, speaking generally, is as seventy to sixty.[15]

The most interesting source of contemporary evidence is to be found in the oral sessions held in front of the 1893-5 Royal Commission on the Aged Deserving Poor (the Aberdare

---

[13] John Macnicol, *The Politics of Retirement in Britain, 1878-1948* (1998), ch. 2.

[14] George Turner, 'State Pensions in Old Age', *Oddfellows' Magazine*, XXVII, 261, Sept. 1896, p. 271.

[15] Charles Booth, *The Aged Poor in England and Wales* (1894), p. 321.

Commission). Witness after witness testified to the fact that older male workers were 'worn out' at progressively earlier ages compared with the rural economy. Although this evidence was anecdotal and impressionistic, it did come from those with direct experience of the late-Victorian labour market. For example, a wire-worker from Finsbury, London, testified that, in his own trade, new technology (in the form of tools) had increased the pace of work compared to his father's day; another witness, a carpenter from Birmingham, claimed that in large towns a man of only 55

> is looked upon upon as almost played out, and the competition of younger men is so great that he has very little chance if he gets out of employment at that period of life of ever getting on again at his own trade.[16]

The main sources for systematic evidence on the sickness experience of workers in the nineteenth century are the records of the friendly societies. These mutual self-help bodies drew their membership predominantly from the skilled, male working class, and their membership was growing in the late nineteenth century, reaching 4,200,000 in registered societies in 1898, with perhaps another 4,000,000 in unregistered ones. The two big federal organisations, the Manchester Unity of Oddfellows and the Ancient Order of Foresters, consisted of many branches spread all over the UK. Coverage for old age per se (via superannuation schemes) was uncommon, but the societies paid benefits to those of their members unable to work through sickness or infirmity. By the end of the nineteenth century sickness benefit was increasingly becoming a surrogate old age pension for older members in response to rises in a range of medical conditions. Sickness benefit claims rose at all ages, but particularly so for older members, who also experienced a higher duration of claims ('protracted sickness').

---

[16] *Royal Commission on the Aged Poor*, 1895, C-7684-II, Vol. III, *Minutes of Evidence*, pp. 742, 880

The finances of friendly societies were closely monitored and regulated by the state: their financial soundness was in the hands of their consultant actuaries, who were not loath to dispense rigorous advice. However, various factors inclined the societies towards an increasingly elastic definition of sickness in the last decades of the nineteenth century. Modern scholars have discussed these possible factors behind this, such as age compositional changes in membership, more liberal definitions of sickness, the survival of members with more health-impaired lives, different administrative practices, changes in sickness recording, or the fact that the societies' driving-down of doctors' fees (in the interests of cost-cutting) resulted in more perfunctory medical assessments.[17] One factor - stressed elsewhere by this author[18] - was the changing labour market. The friendly society definition of sickness was 'inability to work' or 'incapacity from labour'; conversely, as James Riley aptly puts it, 'wellness....was the ability to work'.[19] It will here be argued that the societies were forced to become more generous in their interpretation of sickness when faced with growing job insecurity experienced by their older members.

There were several reasons for this. First, the fraternal ethos of the societies, plus the fact that their internal workings were closely monitored by their members, especially in the smaller branches, disinclined them to be anything less than generous with their older members who had long records of contributions. Blatant malingering was held in check by the observations of fellow-members and quite strict medical assessments, but the adverse

[17] James C. Riley, *Sick, Not Dead. The Health of British Workingmen During the Mortality Decline* (1997), p. 75. For a good general discussion, see Martin Gorsky, Aravinda Guntupalli, Bernard Harris and Andrew Hinde, 'Ageing, Sickness and Health in England and Wales During the Mortality Transition', *Paper to the Annual Meeting of the Social Science History Association*, 25 October 2008.

[18] Macnicol, *The Politics*, ch. 5.

[19] Riley, *Sick, Not Dead*, p. 127.

effects of labour market changes at a local level would be well known and would encourage a more elastic definition of what constituted work-disabling sickness. Again, competition for new members meant that the societies had to appear to be generous in dispensing benefits: it would look bad if older society members had to have recourse to the Poor Law, and the societies prided themselves on how few actually did. Providing that sickness was certified by a society doctor and a sick visitor (who conducted home inspections), payment of benefits was a contractual right. Doctors would have had intimate knowledge of local labour markets and would be sympathetic towards those long-serving older society members whose sickness masked de facto unemployment. As one well-informed observer put it,

> When the period of loss of wages arising from the disability of old age and worn-out working powers arrives, the society's doctor in many cases feels compelled to stretch a point and, rules notwithstanding, to judge cases brought to his notice by the heart rather than the head, lest the old folk become altogether destitute and fall on the poor-rate.[20]

As is well known, the friendly societies faced something of a financial crisis at the end of the nineteenth century owing to this rise in protracted sickness. Bentley Gilbert's famous verdict that the unfolding of the epidemiological transition was bankrupting the societies, compounded by their tendency to use outmoded life tables,[21] has been substantially moderated by subsequent historians: the 'crisis' was manageable, and was dealt with by a number of means, notably by raising weekly subscriptions. A steady influx of new members kept contribution income buoyant. Nevertheless, the rise in protracted sickness aroused

---

[20] J. Frome Wilkinson, 'Friendly Society Finance', *Economic Journal*, II, 8, Dec. 1892, pp. 725-6.

[21] Bentley B. Gilbert, *The Evolution of National Insurance in Great Britain* (1966), pp. 165-80.

much comment at the time and it is well worth considering because in many ways it was a rehearsal for the debate one hundred years later.

The apparent rise in sickness was, of course, counter-intuitive, in that it had occurred alongside great improvements in real wages and the built environment, reductions in mortality, and so on. In short, it seemed puzzling to many contemporaries that one section of society should be falling behind the overall rise in material prosperity. The rise in long-term sickness was also occurring among the labour aristocracy - exactly those men most deeply imbued with Smilesean virtues, and the least likely to be workshy malingerers.

The most systematic and important contemporary investigation was by Alfred Watson, consultant actuary to the Manchester Unity of Oddfellows. It showed that sickness benefit claims had risen at all ages in the Manchester Unity between 1846-8 and 1893-7, but the highest rate of increase had occurred among those members aged 65+. Thus the 'weeks of sickness' experienced by those aged 65+ had risen from 1.8 per cent of the total weeks of sickness to 31.6 per cent over that period. The most striking feature was the increase in protracted sickness (that is, more than two years' duration) and permanent sickness.[22] Watson's classic study was methodologically imperfect and has been criticised by modern historians,[23] but his evidence regarding protracted sickness stands largely unchallenged.

From this brief exploration we can see that the late nineteenth century debate on work-disability was extraordinarily similar to that of today - for example, the suspicion that both friendly society members and their doctors had inflated their definitions of what constituted

[22] Alfred W. Watson, *An Account of an Investigation of the Sickness and Mortality Experience of the I.O.O.F. Manchester Unity During the Five Years 1893-1897* (1903).

[23] James Riley, *Sick, Not Dead. The Health of British Workingmen During the Mortality Decline* (1997), p. 173.

work-disabling sickness. What was less frequently mentioned - except by a few contemporaries[24] - was the fact that both increasing job insecurity and the disappearance of jobs was the major driver behind the rise in long-term sickness claims by older friendly society members.

An interesting and very prescient commentary on this phenomenon was provided by the economic historian T. S. Ashton in 1916, based on his study of the sickness records of Amalgamated Society of Engineers members before and during the First World War. It might be intuitively assumed, suggested Ashton, that 'when earnings are high, and employment good, the numbers on sick benefit would grow, because men would be able to afford an illness'. Conversely, in times of economic depression recorded sickness rates would fall: workers would try to remain in work, for fear of losing their jobs to others. In fact the reverse was true, and Ashton's research showed that there was a correlation between unemployment among engineers and sickness benefit claims by them: when the former rose, the latter also rose.[25]

From the First World War to the 1950s

Between 1912 and the launching of the National Health Service in mid-1948, the principal source of health care for the majority of the population in the UK was National Health Insurance (NHI). Originally introduced as something of a partial measure, NHI appeared more radical than it actually was because of the controversies that attended its passage

[24] For example, Turner, 'State Pensions', p. 271.

[25] T. S. Ashton, 'The Relationship Between Unemployment and Sickness', *Economic Journal*, XXVI, Sept.1916, pp. 396-400.

through Parliament in 1911. Indeed, such was the opposition of the British Medical Association and other vested interests that the Act nearly proved unworkable. Only an immense amount of hard work by the specially convened and exceptionally able team of civil servants charged with framing the Act, plus the political courage of the politicians in charge, saved it.

Given that they were breaking new ground, the planners of the new scheme could not predict the likely adverse consequences. The official most closely involved with the framing of the Act, W. J. Braithwaite, recorded in his famous contemporary account that there were three risks: 'the risk of discovering that sickness was much greater than supposed, the risk of malingering, and the Parliamentary risk of greater and greater demands'.[26] Braithwaite considered that outright malingering was unlikely, given the low monetary level of sickness benefit. However, he made an interesting conceptual distinction between 'conscious swindling' on the one hand, and a more subtle kind of psychosis, comprising, on the other, an unconscious tendency to sickness better described by the German word 'rentenhysterie'.[27] Some opponents of the Act played upon this point. 'I fear the growth of malingering', wrote Beatrice Webb in her diary.[28] Although Sidney Webb was not so implacably opposed, the Webbs both disliked the Act for falling too far short of the fully preventive and curative state medical service they sought and had recommended in the famous Minority Report of the 1905-9 Royal Commission on the Poor Laws. Convinced that the German scheme of health insurance had encouraged fraudulent claims, they pointed to the late-nineteenth century paradox of falling death rates yet rising friendly society sickness benefit claims and argued

---

[26] H. N. Bunbury (ed.), *Lloyd George's Ambulance Wagon: Being the Memoirs of William J. Braithwaite, 1911-1912* (1957), p. 127

[27] Ibid, pp. 94-5.

[28] Beatrice Webb, *Our Partnership* (1948), p. 474.

that an insurance-based scheme inevitably encouraged moral hazard among its users via 'the half-conscious determination to get value for their money by drawing out in benefits the full measure of their own contributions'.[29] Similar fears were occasionally expressed during the Parliamentary debates on the Act: most notably, Sir Thomas Whittaker MP warned that 'malingering and slackness' were the great dangers and would increase, especially as there would henceforth be included people of 'a less satisfactory character'. Another MP, Theodore Taylor, also viewed malingering as the greatest danger but was confident that the friendly societies' experience in dealing with this problem would minimise it.[30] Once the scheme commenced, the higher-than-expected level of sickness benefit claims raised questions of lax certification procedures by doctors. Others argued that this merely revealed a submerged mass of hidden, untreated sickness.

Estimates vary, but one authoritative source records that, two years after its inception (in late 1914), NHI covered 13,689,000 manual workers and other employees earning less than £160 per annum (the income tax limit). This number had risen to 19,706,000 by late 1938 (with overall population growth, and a raising of the eligibility income limit in 1920 to £250 per annum).[31] Initially, sickness benefit of 10s0d (50p) per week was paid to men and 7s6d (37p) per week to women for six months, followed by disablement benefit of 5s0d (25p) per week for an indefinite duration. These benefit rates were raised in 1920 to 15s0d (75p), 12s0d (60p) and 7s6d (37p) respectively, and then reduced slightly (in the case of married women) as an economy measure from January 1933. Maternity benefit was also paid. The scheme was funded by state-supervised contributory insurance (since it covered those below

---

[29] Sidney and Beatrice Webb, *The Prevention of Destitution* (1911), pp. 160-7.

[30] *H of C Deb.*, 5s, vol. XXVI, 24 May 1911, cols. 331, 551.

[31] *Social Insurance and Allied Services*, Cmd. 6404, 1942, p. 213.

the income tax limit), and administered through the 'approved societies' (basically, friendly societies and industrial insurance companies).

Notoriously, NHI had several glaring faults: the dependants of a wage-earner were not covered (which meant that the majority of working class women and children had no guaranteed right to health care until 1948); self-employed working people were also not covered; there was no routine access to specialist treatment in a hospital; and additional benefits (mainly ophthalmic and dental) were patchily provided (depending on the solvency of the individual approved society). These deficiencies and the administrative complexities of the scheme exacerbated all the usual problems of measuring the 'true' level of morbidity. NHI sickness records were therefore a less-than-perfect reflection of the health of the nation. For example, claim levels by region were not routinely available.

Nevertheless, something interesting can be learned from the course of NHI sickness benefit claims in the inter-war years. For the purposes of this paper, the focus will be on the way that unemployment and increased job insecurity pushed up claims, against a background of falling mortality and overall health improvements.[32] Then, as now, this was a paradox to many contemporaries, especially those in the health-related professions who prided themselves on the improvements in public health that had taken place since the late nineteenth century.[33]

---

[32] This issue has already been interestingly explored in: Bentley B. Gilbert, *British Social Policy, 1914-39* (1970), ch. 6; Noelle Whiteside, 'Private Agencies for Public Purposes: Some New Perspectives on Policy Making in Health Insurance Between the Wars', *Journal of Social Policy*, 12, 2, 1983, pp. 165-94; Noel Whiteside, 'Counting the Cost: Sickness and Disability among Working People in an Era of Industrial Recession, 1920-39', *Economic History Review*, 40, 2, May 1987, pp. 228-46; Noel Whiteside, 'Unemployment and Health: An Historical Perspective', *Journal of Social Policy*, 17, 2, 1988, pp. 177-94.

[33] See, for example, speech by Sir Kinglsey Wood, *H of C Deb.*, 5s, vol. 238, 29 April 1930, col. 83.

The inter-war depression engendered an enormously wide-ranging debate on the effects of unemployment and low incomes on physical and mental health. After a brief postwar economic boom, unemployment started to rise in 1921, reaching 1,751,000 in June 1926, and falling slightly to 1,059,000 in May 1927; it then rose again, reaching just under 3,000,000 in January 1933. There then occurred a slow fall and stabilisation for the rest of the 1930s; but even on the eve of the Second World War unemployment totalled 1,232,000. Rarely in the inter-war years did unemployment fall below 10 per cent of the insured population, and at its highest it reached 23 per cent. As is well known, inter-war mass unemployment was a product of a worldwide recession bringing about a slump in demand for the products of the old 'staple' industries, exacerbated by changes in world markets induced by the First World War.

The human impact of unemployment was, of course, greatly exacerbated by its regional concentration in those 'depressed areas' that were centres of the recession-hit heavy industries (most notably, coal mining, shipbuilding, iron and steel production and heavy manufacturing) located in South Wales, the North West and North East of England and central Scotland - precisely those regions that have high levels of claims to long-term sickness and disability benefits today. The inter-war economy experienced growth in some sectors and regions, and decline in others. In the 1930s, these depressed areas became the focus of numerous investigations into the effects of unemployment on child nutrition, maternal and infant mortality, life expectancy, psychological well-being, and so on. The public health controversy in the 1930s was a bitter and protracted one, not least because it was a cardinal principle of the National Government's neoclassical, deflationary economic strategy to keep wage and benefit levels low in order to cut production costs and render British-made goods more competitive on world markets. There were frequent accusations

that this strategy was condemning the unemployed and their families to extreme poverty and even malnutrition.

The NHI scheme was tested to breaking point in the inter-war years. During the First World War, full employment increased NHI contribution income and decreased benefit claims, which boosted approved society funds and dispelled for a time concerns over malingering. But recorded sickness began to rise as unemployment rose from 1921 onwards. An official analysis of claims between 1921 and 1927 by Sir Alfred Watson (now Government Actuary) showed that claims to sickness benefit had risen by 41 per cent for men, 60 per cent for unmarried women and 106 per cent for married women; for each of these groups, claims to disablement benefit had risen by 85 per cent, 100 per cent and 159 per cent respectively.[34] Expenditure on sickness benefit increased from £8,010,000 in 1914-5 to £13,153,000 in 1921-2 and £20,482,000 in 1926-7 (partly owing to the General Strike); it then stabilised at just under £19,000,000 per annum.[35] While a funding crisis never actually materialised, there was enough concern to cause some cuts and a tightening-up of administration in the early 1930s. NHI in the inter-war years was the subject of some controversy - admittedly, not as bitter as the controversies that bedevilled unemployment insurance - over subjects like the administrative efficiency of the approved societies, the financial surpluses they had built up by the end of the 1930s, the question of what to do with contributors who became unemployed and fell into arrears, the treatment of married women (basically, whether maternity constituted 'sickness'), and so on. But for the purposes of this paper, only one of these will be considered - the effect of unemployment on self-defined health, and the related accusation that a significant number of NHI claimants were malingering.

[34] *National Health Insurance. Report by the Government Actuary on an Examination of the Sickness and Disablement Experience of a Group of Approved Societies in the Period 1921-27,* Cmd. 3548, 1930, pp. 5-6.

[35] *Social Insurance and Allied Services,* p. 214.

The NHI scheme contained within itself some contradictory incentives. On the one hand, sickness benefit was generally easier to claim than unemployment benefit, and was subject to less conditionality (for example, the 'genuinely seeking work' test in the latter). Unlike unemployment benefit, it could be received during a trade dispute, and sickness benefit claims rose temporarily during the General Strike of May 1926 and the subsequent miners' lock-out.[36] Again, after January 1935, about 1,000,000 unemployed were transferred to the Unemployment Assistance Board and subject to the controversial household means test, which had been introduced in 1931. Sickness benefit had no equivalent, being a contributory entitlement.[37] It was possible - and often desperately necessary - to supplement it with other sources of income.

On the other hand, there is powerful testimonial evidence that some sickness benefit claimants with families pressurised health insurance GPs ('panel doctors') to certify them 'fit for work' so that they could move to unemployment benefit which was higher in value for a family man, since it carried dependants' allowances - a situation welcomed by the approved societies, since it saved them money. For example, in 1936 a man with a wife and three children would have been able to claim 35s0d (£1.75p) per week in unemployment benefit, but only 15s0d (75p) per week if on sickness benefit. Unemployment benefit was on the margins of subsistence (as defined by the many inter-war poverty surveys), so sickness benefit was drastically below it: such a family would have required £2 per week to reach an agreed 1936 subsistence level - nearly three times the level of sickness benefit and six times the level of long-term disablement benefit. The illogical variation in benefit levels had been

[36] Helen Fisher Hohman, *The Development of Social Insurance and Minimum Wage Legislation in Great Britain* (1933), p. 172.

[37] Whiteside, 'Unemployment and Health', pp. 187-8.

discussed by the 1924-6 Royal Commission on National Health Insurance: the Minority recommended an equalisation of benefit levels, but the Majority could only support dependants' allowances. Many critics agreed that this only resulted in an unemployed person putting off seeking treatment for sickness for fear of being moved to the lower benefit - quite contrary to any rational public health policy.[38]

It was clear, therefore, that the monetary level of sickness benefit was far too low to act as much of an incentive in anything other than the most desperate financial circumstances. An individual or family could only survive on it by utilising other sources of income - from savings, trade union benefit schemes, relatives, neighbours, moneylenders, pawnbrokers, the Poor Law or Public Assistance Committees, and so on. Such households would have to reduce their expenditure to the lowest level possible in order to survive.[39] Even more serious was the reduction of disablement benefit by half after six months, which was designed to sift out all but the most genuine cases. Critics argued that this reduction occurred exactly at that point when a claimant needed more financial help, not less, to help them through a period of convalescence; disablement benefit also carried no provision for occupational therapy or rehabilitation.[40] As one commentator put it, these low benefit levels could only be justified 'on the ground of deterrence'.[41]

Were there other administrative factors that might have caused an upward rise in sickness benefit claims? Some argued at the time that the problem was compounded by lax

[38] *Report of the Royal Commission on National Health Insurance*, Cmd. 2596, 1926, pp. 144, 318-9.

[39] Hermann Levy, *National Health Insurance. A Critical Study* (1944), pp. 77-8.

[40] Joan Simeon Clarke, 'National Health Insurance', in William A. Robson (ed.), *Social Security* (1943), pp. 96-7.

[41] Hohman, *The Development of Social Insurance*, p. 180.

administration in certain approved societies, the societies' reluctance to countenance any criticism of their workings, the tendency of would-be claimants to 'shop around' until they found a panel doctor willing to certify them as unfit to work, panel doctors being too anxious to attract new patients (and the related capitation fee) by appearing to be generous with sick notes, perfunctory diagnoses by panel doctors owing to overwork, or a general cynicism shown by panel doctors towards a system that they had never fully accepted.[42]

However, these possible inflationary factors were more than counterbalanced by deterrent devices. Approved societies operated a system of domestic surveillance by 'sick visitors' who tested the authenticity of claims, and after 1930 these visits were increased. In general, societies did everything they could to get claimants back to work quickly, so that benefit expenditure was minimised: the more they amassed in their accumulated reserves, the more could they offer in the way of additional benefits and thereby attract new members.[43] Finally, any case of diagnostic disagreement could be referred to Regional Medical Officers (fulltime salaried doctors employed by the Ministry of Health). At the time, much was made of the fact that only a small proportion of such referrals were allowed to continue on benefit: this was 34 per cent in 1930 (for England and Wales), with 66 per cent being judged 'fit for work'.[44] However, then, as now, being judged 'fit for work' was not the same thing as having a job. All in all, therefore, the financial and administrative incentives worked *against* the claiming of sickness benefit.

Intriguingly, in the inter-war years there were exactly the same allegations as there are today regarding the tendency of panel doctors who worked in the high-unemployment depressed

[42] Ibid, p. 165.

[43] Clarke, 'National Health Insurance', pp. 101-2.

[44] R. W. Harris, *National Health Insurance in Great Britain 1911-1946* (1946), p. 115.

areas to dispense sick notes too readily, out of misguided kindness. Such doctors would be aware of the devastation caused to local economies by the world recession, and would feel considerable sympathy for men thrown out of work as a result. A charge of malingering was therefore as much a charge against a doctor, who made the diagnosis, as against a patient.

It is clear that many contemporaries in the inter-war years realised that the labour market was having a profound effect on self-defined work-disability, even if they expressed this in guarded terms. Bentley Gilbert recounts a 1926 speech to the annual conference of panel doctors by Walter Kinnear, Controller of the Insurance Department of the Ministry of Health, expressing deep concerns about this and imploring them not to turn the NHI system into a form of the dole.[45] There were also periodic warnings from the Ministry of Health about 'the doctor as relieving officer instead of physician'.[46] However, the Ministry of Health tended to see the problem largely as a decline in health on the part of the long-term unemployed and emphasised administrative factors such as lax certification.[47]

The whole question of incentives internal to the benefit system must therefore be placed in the much bigger context of the economic depression. As today, lack of labour market demand encouraged the more infirm to self-classify themselves as too sick to work, rather than as unemployed (and fit enough to take any suitable job offered). Only a small part of the increase in recorded sickness can be attributed to improved medical diagnostic techniques, higher health expectations or a deterioration in the health of the long-term unemployed. Those who would be first displaced from the labour market by economic recession would be

[45] Gilbert, *British Social Policy*, pp. 289-90.

[46] Hohman, *The Development of Social Insurance*, p. 174.

[47] See, for example, *Eleventh Annual Report of the Ministry of Health 1929-1930*, Cmd. 3667, 1930, p. 199.

those with the most health-impaired lives. Once unemployed, with little prospect of regaining work, they would perceive their symptoms as work-disabling. It is clear, therefore, that sickness was being used to mask de facto unemployment. As in every recession, the line between 'sickness' and 'unemployment' became blurred - especially in the case of mental disorders, which in 1934-5 amounted to over one-third of 'chronic' disability cases (off work for one year or more).[48] Interestingly, in the 1920s, short-term claims to sickness benefit predominated; but in the 1930s these claims stabilised, and it was the long-term claims to disablement benefit that continued to rise[49] - exactly what one would expect to find in a prolonged recession. As Noel Whiteside observes,

> When unemployment rose in the 1930s, numbers of 'impaired lives', with no hope of finding work, settled into semi-retirement on disability benefits and public assistance.[50]

Even Bentley Gilbert, whose somewhat stern account is entitled 'unemployment and malingering', nevertheless acknowledges that

> the line between sickness and well-being for a man suffering economic deprivation was likely to be unclear even to the individual.[51]

In other words, in a recession both employers and employees become more stringent in their assessment of what constitutes 'employability'.

Finally, it is possible to view the inter-war rise in NHI claims as a much-exaggerated problem. To be sure, Watson's 1930 report did reveal a significant rise in claims between 1921 and 1927, but only for women were they markedly higher than the actuarial

---

[48] Whiteside, 'Unemployment and Health', p. 188.

[49] Whiteside, 'Counting the Cost', pp. 233-5.

[50] Ibid., pp. 240-1.

[51] Gilbert, *British Social Policy*, pp. 285, 292.

'expectation' set by Watson himself on the basis of his 1903 study of the Manchester Unity - and arguably Watson had set the expectation for women too low, basing it on the sickness experience of the small number of women friendly society members. If the increase for men had been caused by outright malingering, then claims to long-term disablement benefit (the most likely refuge of the malingerer) would have been much higher than the expectation, but they were not. In addition, there would have been a marked increase in claim duration as compared with the number of claims - but both had risen equally. Only in the 1930s, once the economic recession had settled in, did claim duration increase significantly.

In the late 1930s, as today, there was a growing concern in governmental circles that long-term unemployment had eroded the will to work in the depressed areas and that the 'dole habit' might even become intergenerational. The problem was of course resolved, in somewhat spectacular fashion, by the Second World War. Unemployment fell significantly from mid-1940, with the enormous economic stimulus created by a war economy (assisted by military call-up removing from the civilian labour market many men and women of prime working age). During the War, many groups who had previously been marginal to the labour force (including those judged to be 'unemployable') found jobs in a very tight labour market. Interestingly, sickness benefit claims rose for a time, probably because of the entry into the labour market of more health-impaired people (the long-term unemployed, older workers and less healthy women), plus the stresses and strains of wartime.[52] However, from 1946 claims fell. In the full-employment 1950s and 1960s, sickness benefit claim levels remained roughly constant, at just under or just over 1,000,000 claims per annum.[53] From the 1970s, however, things began to change once again.

---

[52] R. M. Titmuss, *Problems of Social Policy* (1950), pp. 527-9.

[53] Central Statistical Office, *Annual Abstract of Statistics No. 97 1960* (1960), p. 50, and *Annual Abstract of Statistics No. 108 1971* (1971), p. 50.

Two important conclusions can be drawn from this. First, were the rise in sickness claims mainly the result of a 'cultural inflation' or a change in the propensity to claim benefits, then there would not have been a low level of claims in 1915-21 and in the 1950s and 1960s. Instead, recorded sickness would have inexorably risen in both good times and bad. Second, had health worsened because of the adverse effects of heavy industrial working then sickness benefit claims would have been high in the full-employment 1950s and 1960s. In fact, the fluctuations in claim levels correlate closely with economic conditions.

Conclusion

The history of work-disability has much to teach us regarding the present. From this necessarily-brief and tentative exploration, we can see that over the past hundred and thirty years there have been three periods in which trends in mortality and recorded morbidity have followed divergent paths and levels of work-disability have risen. Despite overall improvements in population health, sickness has appeared to increase for a minority. In each period, there has been a vibrant debate over the many factors that might cause levels of recorded sickness to rise, and a recognition that sickness is a continuum which is subject to many contextual influences. What light does the historical evidence throw on the four possible explanations outlined at the start of this paper? First, the incentive effect of benefits was only minimal. Nineteenth century friendly society benefit payments were heavily policed by fellow members and society officials, who seem to have sanctioned the increasing subsidy of their older members in recognition of their increasing difficulty in holding on to jobs. Again, NHI sickness benefit was markedly lower in value than unemployment benefit, and was therefore only used as a desperate last resort (often as a supplement to other meagre sources of income). Second, the work-disabled certainly suffered poor employability and

were the first to be shaken out during major economic restructuring, but this was primarily a consequence of diminished sectoral labour market demand. Third, deteriorating health among the long-term unemployed was a serious problem in the inter-war years, but it was much more a consequence of unemployment than a cause of it: once the UK labour market tightened up after 1940, even the most health-impaired found jobs. Finally, the one factor common to all three peaks in work-disability has been labour market restructuring, causing jobs to disappear and/or producing lower-quality replacement jobs. Conversely, sickness benefit claims went down when tight labour markets existed in 1914-21 and the 1950s and 1960s.

The crisis in the late nineteenth century was ameliorated by the introduction of old age pensions and a growing acceptance of the inevitability of retirement. The crisis in the inter-war years was abruptly resolved by the Second World War and the subsequent twenty-five year postwar boom in manufacturing. It remains to be seen whether supply-side policies and a withdrawal of benefits will be able to resolve this latest crisis.