The NHS in ‘crisis’: The role played by a shift from horizontal to vertical principles of equity

Abstract
Explanations of the state of ‘crisis’ in the English National Health Service generally focus on the overall level of health care funding rather than the way in which funding is distributed. Describing systematic patterns in the way different areas are experiencing crisis, this paper suggests that NHS organisations in older, rural and particularly coastal areas are more likely to be ‘failing’ and that this is due to the historic underfunding of such areas. This partly reflects methodological and technical shortcomings in NHS resource allocation formulae. It is also the outcome of a philosophical shift from horizontal (equal access for equal needs) to vertical (unequal access to equalise health outcomes) principles of equity. Insofar as health inequalities are determined by factors well beyond health care, we argue that this is an ineffective approach to addressing health inequalities. Moreover, it sacrifices equity in access to health care by failing to adequately fund the health care needs of older populations. The prioritisation of vertical over horizontal equity also conflicts with public perspectives on the NHS. Against this background, we ask whether the time has come to reassert the moral and philosophical case for the principle of equal access for equal health care need.

Keywords: NHS, horizontal equity, vertical equity, resource allocation, institutionalised ageism

Introduction
It has become commonplace to describe the English National Health Service (NHS) as being in a state of ‘crisis’. The last three winters (2015-17) have certainly seen numerous hospital trusts declare ‘black alerts’ about their ability to meet patient demand, with associated delays in pre-planned operations and routine outpatient appointments. There are also worrying trends in staff recruitment, retention and morale; and performance against key indicators (e.g. waiting times in A&E and for GP referrals, including for cancer) appears to have worsened.
Explanations for the current state of the NHS have tended to focus on factors such as austerity, demographic ageing, failures of local health service management and even the UK’s forthcoming departure from the European Union. Yet none of these explanations (including the local management meme, an individualising narrative that should imply random variation in performance failures) can account for systematic patterns in the extent to which different areas are experiencing crisis. Nor can they account for the relationship between indicators of organisational stress and NHS resource allocation methodologies which, as we argue in this paper, have underestimated the needs of demographically older populations for nearly two decades. It is no coincidence that NHS organisations in older, rural and particularly coastal areas are more likely to be in financial deficit, have longstanding deficits and be in ‘special measures’. They also tend to spend less per patient treating cancer and CHD, have longer than average waiting times for admission to hospital, more often fail to meet cancer referral targets, and have higher rates of delayed discharge from hospital.

Given that the NHS Constitution for England (DH, 2012) stipulates that the NHS should provide a comprehensive service, available to all, based on clinical need and not an individual’s ability to pay, what can account for the fact that the service fails to achieve ‘equal opportunity of access to health care for people at equal risk’ between different parts of the country? Why, moreover, are the difficulties facing health and social care provision in areas often described as being economically ‘left behind’ not being highlighted more widely? The aim of this paper is to try to answer these questions.

It begins by demonstrating, with reference to key indicators, that areas serving older rural populations are more likely to be experiencing signs of ‘crisis’ and more likely to be providing poorer access to services than areas serving younger urban populations. We go on to argue that this reflects historic underfunding of such areas, due to both technical failings in resource allocation formulae and, and more controversially, a philosophical shift away from support for health care equity to the argument that health equity is the more ethical objective.

Insofar as the role that health care services can play in reducing health inequalities is highly contestable, we propose that the resulting shift towards a more residual approach to distributing NHS resources is likely to be ineffective. It has, moreover, exacerbated inequalities in service provision in ways that could be described as institutionalised ageism. As this conflicts with public perceptions about the acceptable trade-off between promoting
health care equity and reducing health inequalities, we conclude by asking whether ‘we’, as academics, have lost sight of what the NHS means to the English public; whether we have inadvertently introduced ideas of ‘deserving’ and ‘underserving’ recipients of what is more generally understood to be a universal service; and whether it is time to reassert the principle of equal access for equal needs – an equity definition that has been dismissed as being both unworkable and unethical, but which nevertheless chimes with public understandings of a national health service.

**Crisis. What crisis? Geographical variations in NHS organisational stress**

Since around 2015, there have been growing references to the English NHS as a service ‘in crisis’. A Nexis search of headlines containing the words ‘NHS’ and ‘crisis’ published in UK national newspapers over the past decade finds less than ten hits per annum prior to 2012. By 2015, 105 articles referred to NHS and crisis in their headlines. In 2017, the figure had grown to 203. Further analysis using additional keywords to explore whether press coverage (headlines and lead paragraphs) associated crisis with inequality, geographical variation (by various areal characteristics) or ‘postcode lottery’ yielded nothing.

The ‘crisis’ facing the NHS is thus usually depicted as a nationwide issue and is in many accounts linked to a lack of overall funding (e.g. relating General Domestic Product (GDP) expenditure per capita against European comparators). We do not dispute the case for additional NHS and social care investment. However, the main argument of this paper is that the ‘national crisis’ meme has detracted attention away from the fact that some NHS areas in England are far more likely to be experiencing ‘crisis’ than others.

**Variations in CCG performance**

To support the Five Year Forward Plan (NHS England, 2014), NHS England introduced an assurance regime to assess the performance of the (then) 211 CCGs. The indicators used to judge performance have since been amended (NHS England, 2016; 2017a). However, by conflating various categories into a simple dichotomy between performing well (e.g. categorised as ‘Outstanding’ or ‘Good’, or given a ‘Green’ rating) and performing poorly (e.g. ‘Requires Improvement’ or ‘Inadequate’, or given an ‘Amber’ or ‘Red’ rating), it is possible to assess the extent to which poor performance varies with respect to various CCG characteristics.
One such characteristic is the extent to which CCGs are ‘urban’ or ‘rural’ according to the Office for National Statistics’ (ONS, 2011) rural/urban classification. Across almost all performance indicators, a significantly higher proportion of CCGs serving rural populations were judged to be performing poorly. In 2015/16, for instance, 52 of the 76 CCGs (68%) serving at least ‘significantly’ rural populations were judged overall to be ‘Inadequate’ or to ‘Require Improvement’, compared with 65 of the 133 urban CCGs (49%). In 2016/17, 57% of rural CCGs performed poorly compared with just 35% of urban CCGs.

Although rural and urban areas are often characterised in terms their physical environments, it is important to recognise that they also tend to contain quite different populations. Urban CCGs generally serve younger and more deprived populations than rural CCGs, and the significance of demography and deprivation can be illustrated by comparing performance across the 54 CCGs serving populations that are in both the two youngest and two most deprived quintiles with the 43 CCGs serving populations that are in both the two oldest and two least deprived quintiles. 39% of younger more deprived CCGs were judged to be performing poorly in 2015/16, compared to 70% of the older less deprived CCGs. This disparity continued into 2016/17 when the equivalent figures were 35% and 57% respectively.

**Variations in CCG deficits**

There is also a strong underlying relationship between the socio-demographic characteristics of the populations served by CCGs and their end-of-year financial outcomes. In 2015/16, nearly a quarter of the CCGs serving older less deprived populations (11 of 43) ended the year in deficit, and the cohort as a whole posted an aggregate deficit of £57.8 million (NHS England, undated, a). In contrast, only 3 of the 54 CCGs serving younger more deprived populations posted an end-of-year deficit (5.6%), and this cohort as a whole accumulated an aggregate surplus of £315.8 million. Similarly, fourth quarter accounts for 2016/17 show that by April 2017, 18 (42%) of the older less deprived CCGs were overspending relative to their spending ‘control totals’ compared to only 7 (13%) of CCGs serving younger more deprived populations. And where the aggregate over-spend for the 43 older less deprived CCGs stood at £51.7 million, the 54 CCGs serving younger more deprived populations ended 2016/17 with a £147 million net underspend.

As discussed below, we have argued for many years that the methods used by the NHS to allocate resources to CCGs (and Primary Care Trusts before them) have underestimated the
health care needs of older, less deprived populations. In view of this, the fact that commissioners serving such areas are more likely to be working under financial pressures is not particularly surprising. We now turn to whether the NHS is showing other signs of organisational stress in areas serving demographically older populations.

**Variations in provider deficits**

There are 243 NHS trusts and foundation trusts providing ambulance, hospital, community and mental health services in England. In 2010/11, just 5% providers were in deficit. By 2015/16, this had grown to 66% before falling to 44% in 2016/17 (in part due to one-off savings, temporary extra funding and accountancy changes (Gainsbury, 2017)) and rising again to 63% by the end of the 2nd quarter of 2017/18.

Provider deficits are strongly concentrated in acute trusts. Analysis by the Health Foundation (Lafond et al, 2016) found no clear regional pattern to the decline in hospital finances. Yet, the region is too large a spatial scale at which to explore potential associations between population characteristics and provider deficits and a somewhat different picture emerges by looking at the ‘pattern’ of acute hospitals that are managing large deficits or have been placed in special measures.

At the end of September 2017, 25 providers were managing deficits in excess of £20 million (NHS Improvement, 2017). According to previous quarterly performance reports, 18 of these had spent most if not all of the previous year carrying this level of deficit. Half of these have been or are in ‘special measures’, a regime predicated on the belief that a failure to balance the books is ultimately a failure of leadership:

> “Some organisations and geographies have historically been substantially overspending their fair shares of NHS funding [...] In effect they have been living off bail-outs arbitrarily taken from other parts of the country or from services such as mental health. This is no longer affordable or desirable. So going into 2017/18 it is critical that those geographies that are significantly out of balance now confront the difficult choices they have to take (NHS England, 2017b, p.53).

Yet, trusts with large deficits are not distributed randomly around the country, as one would expect if poor financial performance was simply a matter of leadership. With the exceptions of Barts Health and Kings College NHS Trust, financially struggling acute trusts do not tend to
serve metropolitan populations. Instead, they are to be found in the peripheries of cities (there are a clutch of struggling trusts serving outer London Boroughs) and on the peripheries of England itself. Coastal areas with populations that are older than the national average are over-represented in the group (trusts in Norfolk, Lincolnshire, Cumbria, Kent and Sussex having longstanding deficits), as are a number of ‘shire’ counties (Worcestershire, Leicestershire, Staffordshire). Thus, as with CCG deficits, there is evidence to suggest that demographically older areas are more likely to experience financial stress in their provider organisations.

**Variations in expenditure and service quality**

Similar patterns emerge with respect to expenditure and service quality. For instance, contemporary Quality and Outcomes Framework (QOF) disease register counts (NHS Digital, undated) can be used to contextualise the most recent (2013-14) Programme Budget Category (PBC) expenditure data on cancer and Coronary Heart Disease (NHS England, undated c). This reveals a strong association between rurality and ‘per QOF patient’ spend. Thus the 10 CCGs with the lowest per cancer patient spend (on average £1527) contain a far higher proportion of people living in rural localities (20%) than the 10 CCGs with the highest per cancer patient spend (£4230). In these CCGs, just 0.7% of people live in rural localities. A similar contrast emerges between the 10 CCGs with the lowest per CHD patient spend (£347; 22.1% of population in rural localities) and the 10 CCGs with the highest per patient spend (£927; 3.3% in rural localities). This reflects the way in which spending on both cancer and CHD patients tends to fall as the proportion of CCG populations living in rural localities increases; from an average of £3035 and £633 respectively in the most urban quintile of CCGs (n=42) to an average of £2122 and £481 respectively in the most rural quintile of CCGs (n=42).

As discussed above, this is not so much a function of geography as a reflection of the very different socio-demographic characteristics of the populations served by urban and rural CCGs, and it is the way in which spending falls with the increasing age profile of CCGs that is particularly striking. Thus, CCGs in the youngest quintile (maximum 13.7% aged 65+ in mid-2014 (ONS undated, a)) spend, on average, £3036 and £640 per cancer and CHD patient respectively, compared with just £2138 and £464 in CCGs in the oldest quintile (minimum 20.7% aged 65+). Some of this statistically significant difference in ‘per QOF patient’ spending between younger and older CCGs may be clinically appropriate in that younger patients tend
to be able to withstand longer, more aggressive and ultimately more costly treatments, though conversely a greater proportion of patients in the older CCGs are likely to be incurring substantial ‘end of life’ costs, and these are included in the ‘total’ cancer and CHD PBC expenditure calculations. The evidence of 2013/14 PBC expenditure data is, at the very least, compatible with the idea the CCGs serving older less deprived (and often rural) populations are more financially constrained than CCGs serving younger, more deprived (and largely urban) populations. This is also true of more recent data available via the 2016 Commissioning for Value (CfV) CCG data packs (NHS England, undated, d).

As with the PBC data, CfV data offer few opportunities to relate expenditure on particular conditions with denominators which capture the number of people with those conditions. This is possible, however, with respect to CHD, lung cancer and breast cancer insofar as CCG-level elective and non-elective expenditure on admissions for all three conditions is recorded, as is ‘primary care prescribing’ expenditure for CHD and breast cancer. Suitable denominator data is provided by QOF CHD register counts for 2015/16, and lung and breast cancer incidence data for 2011-14 recorded in the CfV dataset itself. The incidence data do not, of course, allow ‘per patient’ spend estimates to be calculated, but if the incidence of lung and breast cancer is taken as a reasonable proxy for the disease burden in different CCGs, then ‘expenditure divided by incidence’ should provide a useful proxy for ‘per patient’ spend in different CCGs.

As with respect to 2013/14 PBC data, ‘per CHD patient’ spend in 2015/16 is significantly higher in the more urban CCGs. Thus, primary care prescribing, elective and non-elective expenditure per QOF patient is, on average, £217.82, £146.72 and £288.04 respectively in the most urban quintile of CCGs compared to £181.35, £121.68 and £241.56 in the most rural quintile of CCGs. Much the same pattern emerges with respect to demography, where average CHD expenditure in the youngest quintile of CCGs of £218.20, £149.85 and £296.34 on primary care prescribing, elective and non-elective admissions respectively compares with £182.10, £124.25 and £240.38 in the oldest quintile of CCGs.

The picture is not quite so clear-cut with respect to lung and breast cancer. For these cancers non-elective spending (relative to incidence) is significantly higher in urban areas than rural areas: £114.48 as opposed £57.87 for breast cancer and £579.92 as opposed £486.67 for lung cancer. The scale of this disparity may partly reflect the fact that for both cancers a greater
proportion of admissions in urban areas are non-elective (emergency); 9.0% compared to 4.8% in rural areas for breast cancer, and 49.8% compared to 46.3% in rural areas for lung cancer. Yet for both cancers spending on non-elective admissions (relative to incidence) is also higher in urban areas, albeit not significantly so. With this pattern being replicated in terms of the distinction between CCGs serving relatively young and relatively old populations, the overall picture for these two cancers is broadly similar to that of CHD; i.e. less is being spent, relative to need, in older rural populations than in younger urban populations. And in view of this it is perhaps not surprising that in 2016-17 the proportion of suspected breast cancer patients not seen by a specialist following an urgent GP referral within the benchmark two-week threshold was significantly higher in CCGs in the most rural quintile (9.6%) than in CCGs in the most urban quintile (5.5%) (NHS England, undated, e). Once again, almost exactly the same contrast emerges in terms of demography; with 9.6% and 5.5% of patients in the youngest and oldest CCGs (defined in terms of percent population 65 and over (ONS undated b) not being seen within 2 weeks.

Further commissioner-level waiting time data are available via Consultant-led Referral to Treatment Waiting Times (RTT) statistics (NHS England, undated, f) and, once again, the evidence from 2016-17 is that older rural areas fare significantly worse than younger urban areas. There is no significant difference with respect to ‘non-admitted pathways’, but patients admitted to hospital wait significantly longer in the most rural quintile of CCGs (with, on average, a 10.9-week median wait) than in the most urban quintile CCGs (9.7-week median wait). This is an overall average, but the same pattern is found at speciality level. Median waiting times are, for instance, almost always higher in the oldest quintile of CCGs, although statistically significant differences are only found for cardiology (5.9 weeks compared to 3.3 weeks in the youngest quintile), gynaecology (9.9 weeks as opposed to 8.2 weeks) and Urology (8.7 weeks compared to 7.3 weeks). Waiting time data are also available for diagnostic tests and procedures (NHS England, undated g), though here there is no systematic difference in waiting times by rurality or demography, either overall or with respect to specific tests and procedures.

Other key indicators of service quality available for CCGs include Accident & Emergency (A&E) waiting times and delayed discharges from hospital (both of which are included in the CCG improvement and assessment framework 2016/17 (NHS England, undated h) and the
proportion of urgent GP cancer referrals receiving first treatment within 62 days (NHS England, undated i). Although the proportion of people attending A&E who are admitted, transferred or discharged within 4 hours shows no distinctive pattern, delayed transfers of care (proposed as a measure of the effectiveness of the interface between health and social care services) are significantly higher in the most rural quintile of CCGs (15.8 delayed days per 100,000 population) than in the most urban quintile of CCGs (10.5/100,000), and are similarly higher in the oldest CCGs than in the youngest CCGs (17.5 as opposed to 11.0 per 100,000).

The proportion of urgent GP cancer referrals which failed to meet the two-month (62-day) first treatment threshold in 2016/17, meanwhile, was also higher in rural/older CCGs than younger/urban CCGs, albeit that the difference was not statistically significant. Yet this indicator serves to illustrate the problem of depending on rather coarse classifications of CCGs based on the proportion of the population aged 65 and above and living in rural localities. Thus although 19.4% of cancer referrals of patients in the most rural quintile of CCGs were not seen within 62 days compared to 18.7% of referrals of patients in the most urban CCGs, it is perhaps notable that of the 20 worst performing CCGs, 4 were in Essex, 4 in Kent, 4 in Staffordshire, 2 in Lincolnshire, and 2 in Worcestershire. The statistical data is not able to explicitly isolate such peripheral and/or coastal areas, a limitation given recent evidence that Britain’s coastal communities are among the worst ranked parts of the country across a range of economic and social indicators (Corfe, 2017).

Nevertheless, the foregoing provides ample evidence that, across a range of measures, NHS organisations in older and more rural areas are more likely to be performing poorly and struggling financially. Such a systematic pattern belies NHS England’s individualising narrative which places responsibility for organisational failure on poor leadership and is more consistent with our longstanding argument that the health care needs of older less deprived populations have been underestimated by resource allocation methodologies (Asthana and Gibson, 2006a, 2008, 2013).

**NHS Resource Allocation: what is it aiming to achieve?**

The NHS Constitution for England (DH, 2012) stipulates that the NHS should provide a comprehensive service, available to all, irrespective of gender, race, disability, age, sexual orientation, religion or belief; and that access to NHS services (which are free, except in
limited circumstances sanctioned by Parliament) is based on clinical need, not an individual’s ability to pay. These principles provide a commitment to horizontal or ‘health care’ equity and are promoted through, among other mechanisms, the system of NHS resource allocation. Since 1976, this has stated that health care resources should be geographically distributed to ensure ‘equal opportunity of access to health care for people at equal risk’.

Health care equity has remained a key objective of resource allocation. In 1999, however, the Advisory Committee on Resource Allocation (ACRA) introduced an additional requirement; that resource allocation should ‘contribute to the reduction of avoidable inequalities in health’ (Bevan, 2009). This approach shifts the definition of equity away from horizontal principles of equal expenditure and/or access for equal needs to the principle of achieving equal health outcomes (‘health equity’). Insofar as it accepts the case for the positive targeting of resources to those with the worst outcomes (which is not the same as clinical need), it is a vertical definition. It is also enshrined in the NHS Constitution (DH, 2012) which states that the NHS “has a wider social duty to promote equality through the services it provides and to pay particular attention to groups or sections of society where improvements in health and life expectancy are not keeping pace with the rest of the population”.

On the face of it, health and health care equity may appear to be complementary goals. In practice, however, there is an inevitable trade-off between sacrificing equity in access to health care in order to redress imbalances in the achievement of health. A key difficulty is that the distribution of population ‘need’ varies according to the principle adopted (Asthana and Gibson, 2008). To promote health equity, funding (for prevention or to address unmet need) needs to be targeted so as to reduce the health gap between the most and least socio-economically advantaged groups. This tends to be measured in age-adjusted or age-standardized terms (where the effects of age and sex are designed out to allow the health profiles of different socio-economic or ethnic groups to be compared). Using such measures, health inequalities are highest in urban and declining industrial areas where social deprivation is more extreme.

In order to promote "equal opportunity of access for equal needs", on the other hand, the distribution of funding should reflect the existing burden of disease that is amenable to management or cure. For most conditions (mental health being a notable exception), age is a far more significant determinant of morbidity and mortality than deprivation (Asthana et al,
Thus, the health communities grappling with the highest burdens of chronic illness, disability and mortality in crude (or absolute) terms tend to serve the most ageing areas. Because there is a negative correlation between the geographical pattern of social deprivation and age in England (deprived areas tending to have younger populations), this means that many areas with the highest disease prevalence have good outcomes in health equity terms, while many areas with the lowest disease prevalence have poor outcomes e.g. low life expectancy.

If the distribution of ‘need’ varies depending on whether the aim is to promote equal access for equal needs or more equal health outcomes, the question arises of how the NHS has chosen to resolve this tension. Between 2002 and 2013, there was a substantial redistribution of NHS resources away from older populations with higher disease prevalence towards younger more deprived populations with lower disease prevalence. Thus, in 2012-13, 6 of the 10 most generously funded PCTs in terms of per capita allocations were in Inner London; and 3 in the metropolitan county of Merseyside. Newham and Islington PCTs topped the table with respect to funding allocations whilst having the lowest proportions of their populations aged 65 and over and 75 and over. Ranking 3rd and 14th with respect to the Index of Multiple Deprivation (IMD), these PCTs served deprived populations and had relatively high mortality rates in standardised terms (ranking 24th and 26th out of 150 PCTs). However, because of their relatively young populations, they had low levels of disease prevalence, ranking 149 and 150 with respect to QOF registered cancer prevalence; and 139 and 141 with respect to crude mortality rates.

The PCTs with the oldest populations with respect to the proportion aged 75+ (Dorset, East Sussex Downs and Weald, Hastings and Rother and Torbay) did not receive the lowest allocations. However, the fact that they ranked first, 4th, 12th and 5th with respect to QOF registered cancer prevalence, and 5th, 4th, 2nd and 3rd with respect to crude mortality, their rankings in allocation terms (83rd, 80th, 38th and 36th) suggest that, by 2012-13, the distribution of funding prioritised the principle of health as opposed to health care equity.

The mismatch between disease burden and NHS funding in part reflects methodological and technical failings in the system of NHS resource allocation. This has been dominated by an econometric approach (Carr-Hill et al, 1994, 1997; Sutton et al, 2002; Morris et al, 2007; Dixon et al, 2011; Sutton et al, 2012) which uses regression analysis to identify the variables which...
best predict past (national average) utilisation. It has been criticised on the grounds that a population’s use of services provides an inadequate measure of its need for services (Mays, 1995; Asthana et al, 2004; Asthana et al, 2011). Systematic patterns of unmet need, as well as unjustified supply are very difficult to isolate and can distort the relationship between the need for and use of services (Sheldon and Carr-Hill, 1992; 1993; Stone, 2006; Stone and Galbraith, 2006). It is, in particular, argued that utilisation-based allocation methodologies have inherent circularity, i.e. perpetuate existing patterns of service provision precisely because the allocation of resources to different client groups will reflect the use they make of services that are already differentially available (Smith, 2006).

The tendency of regression-based formulae to reflect and reinforce historic inequalities in funding was exacerbated by technical failings in the way in which the ‘AREA’ formula (which guided NHS allocations between 2002-2009) was implemented. This used a two-step procedure to model age-related and additional needs (deprivation) effects. Because, as noted above, the distribution of age and deprivation is negatively correlated in England, the sequential inclusion of deprivation indicators effectively cancelled out the effect of age (Stone, 2007; Asthana and Gibson, 2008), resulting in a significant redistribution of targeted funding towards deprived English areas. The later CARAN (Combining Age Related and Additional Needs) methodology (Morris et al, 2007) addressed this flaw by calculating age and additional needs in a one-stage model, stratified by age. If implemented without adjustment, this would have resulted in a very significant redistribution of revenue income away from the most deprived areas (Asthana et al, 2012).

In fact, the Caran formula was not implemented in a way that would have supported the health care equity objective. A ministerial decision was made to apply a ‘health inequalities’ adjustment of 15% to keep “the distribution of funding between the most and least deprived areas in line with the previous formula” (Hansard, 2009). Despite being a politically expedient decision to maintain the funding disparity between older less deprived and younger deprived populations relative to their underlying morbidity, this decision received remarkably little criticism, reflecting the wider philosophical shift among academics and policy makers in attitudes towards horizontal and vertical principles of equity.
The philosophical shift from health care equity to health equity

In the 1980s, a sustained critique developed in the UK of the equal access for equal need principle, first because various contemporary ‘horizontal’ definitions of equity were considered to be mutually incompatible – equality of expenditure for equal need not necessarily translating into equity of access or equity of treatment (Mooney, 1983; Culyer and Wagstaff, 1993); and second, for failing to consider need in terms of capacity to benefit, an approach that (a) promoted ‘efficiency’ and (b) lent itself to the more ‘ethical’ objective of achieving health equity (Culyer, 1989, 2001, 2006; Williams, 1997): “An equitable health care policy should seek to reduce the inequality in health (life expectation, self-reported morbidity, quality of life in terms of personal and social functioning) at every stage of the life-cycle. Such a policy must meet needs, but in proportion to the “distance’ each individual is from the population average . . .” (Culyer, 2001, p.281). This understanding of vertical equity, defined as the unequal but equitable treatment of unequals (Mooney and Jan, 1997) is clearly distinguishable from horizontal definitions because the primary focus is on outcomes not access. As Evans (1984) argued, “It is health, as a status, rather than health care as a commodity, which is of value to its users”.

In the late 1990s, Mooney and Jan (1997) observed that ‘vertical’ equity considerations had tended to be overlooked in the health policy literature. Reviewing both literature and policy developments (such as in NHS resource allocation) since, the reverse would appear to be true. Compared to a substantial body of literature exploring the case for distributing health care so as to secure a more equal distribution of health (e.g. Culyer, 1989; Williams, 1997; Brouwer et al, 2008), there has been a relative dearth of literature asserting the moral, philosophical and empirical case for horizontal equity (Sen, 2002; Harris, 2005). Reviewing literature from the last forty years of health economics, Wagstaff and Culyer (2012) note that, compared to articles on the determinants of health, economic evaluation and public health, articles on the supply of health services and demand for health care have low rates of citation.

As noted above, support for vertical equity has largely been made on ethical grounds. Arguing that people are often less concerned about the optimal allocation of inputs and outputs than other issues like rights and outcomes, “extra-welfarism” (so called because it departs from traditional welfare economics) makes the case for allowing gains in health to outweigh losses; using outcomes other than utility; and prioritizing social over individual values (Brouwer et al,
2008; Birch and Donaldson, 2003; Coast et al, 2008; Coast, 2009). It also claims parallels (Cookson, 2005) with Sen’s capability approach (Sen, 1993; 1999), though Sen himself has expressed reservations about prioritizing health equity, first because this is unlikely to be achieved through the distribution of health care (also acknowledged by Culyer (2012)); and second because accepting the importance of health equity does not mean that the relevance of other claims (such as non-discrimination in health care delivery) should be denied (Sen, 2002). It is worthwhile elaborating on both points.

**Can the NHS achieve health equity?**

It is broadly agreed that the unequal distribution of health outcomes reflects the unequal distribution of the social factors that influence health (Graham, 2004). Scambler (2001) proposes six capital flows through which social disadvantage and adverse health outcomes are linked: biological, psychological, social, cultural, spatial and material. This concept accommodates many of the key factors that have been identified as risk factors for health inequalities throughout the life course (Asthana and Halliday, 2006; Marmot et al, 2010), few of which lend themselves to health care interventions.

During the same period that the case was made for prioritizing the principle of health as opposed to health care equity in health policy, the distribution of capital flows (i.e. the generative mechanisms that give rise to health inequalities) became increasingly unequal in the UK. This raises the question of whether, while well-meaning, the emphasis on the role that health services should play in promoting health equity has served to medicalise the problem of health inequalities and, in so doing, diverted attention away from the fact that such inequalities are fundamentally embedded within the unequal structures of society (Asthana et al, 2012). Thus, tensions have been identified between the ‘ethical’ objective of extra-welfarism and other moral claims.

**Extra-welfarism and the institutionalisation of ageism**

Such claims include non-discrimination in the delivery of health care. Some extra-welfare theorists have been explicit in justifying the inevitable trade-off between sacrificing equity in access to health care in order to redress imbalances in the achievement of health. Take, for example, the fair innings argument (FIA), which provided a clear statement of the need to treat individuals unequally. This proposed that everyone should be given an equal chance to
have a reasonable quality-adjusted length of life. As older patients could be considered to have received their entitlement to a fair innings, younger patients should be given a higher priority for scant medical resources (Lockwood, 1988; Maynard, 1996; Williams, 1997). Even among those who object to the FIA, there is acknowledgement that the argument appeals to the instinctive assumption that it is a greater tragedy when a young person dies than when an old one dies (Harris, 1989; Dworkin, 1993).

Yet, one of the problems of treating older and younger people as if they are competing for NHS resources is that their respective requirements are so very different. Even if one proposed that each individual has a fixed lifetime entitlement to health care, it cannot be assumed that most would use this up by the time they reached three score and ten. As noted above, age is a significant determinant of most chronic diseases, younger populations tending to be characterized by low rates of disease, disability and death. Thus, just as those with the greatest needs for education are the young, the people with the highest needs for health care are older people. Against this background, denying access to the very group most likely to require NHS services runs the risk of making older people worse off without necessarily improving outcomes for the disadvantaged, leading on an overall ‘levelling down’ (Parfit, 1997; Norheim, 2009).

Against this background, it is worth noting that the lower per capita spend on cancer patients in demographically ageing areas (see above) tallies with concerns about the under-treatment of older people with cancer in England (Pritchard, 2007); and evidence that cancer outcomes for older people are poorer in the UK than in other comparable countries (Coleman et al, 2011; Maringe et al, 2013; Walters et al, 2013). Moller et al (2011) estimated that if UK survival rates matched the highest performers in Western Europe for 75-84 year olds and outcomes in the USA for those aged 85+, there would have been 15,000 fewer cancer deaths every year. Much of the explanation for these outcomes has focused on clinicians’ treatment decisions, which may be informed by a lack of clinical trial evidence on older patients and perceptions about their ability to cope with treatment (Ring, 2010, National Cancer Equality Initiative/Pharmaceutical Oncology Initiative, 2012). Is it possible that, as resources have been strongly targeted towards the principle of health equity as opposed to health care equity, the under-treatment of older cancer patients is related to the financial context of English hospital trusts?
Public perspectives on equity

The fact that vertical equity may result in ageism with respect to access to NHS services may also conflict with public perspectives on equity. There is public recognition that older people have paid national insurance contributions over their lives on the understanding that the NHS provides care from cradle to grave (Coast et al, 2002), raising the question of whether a failure to meet health needs in older age is a form of contract violation (Anand and Wailoo, 2000). More broadly, the English public is primarily supportive of a universal NHS. Thus, while public participants at deliberative events seem willing to contemplate differential access (or co-payments) for services that were perceived as luxuries as opposed to necessities, they show a practical and ideological distaste for differential access to core NHS services on the basis of income (Galea et al, 2013). Part of the rationale is a belief that, if some paid and others received NHS services for free, those who were being asked to pay could demand a higher standard of service, raising the possibility of a two-tier health service. However, even when there is an acceptance that the NHS might be under pressure, there appears to be a public unwillingness to accept a change to the fundamental principle of equal access for equal clinical need.

Against this, other studies investigating public preferences suggest that the general population is willing to give higher priority to some categories of people and lower priority to others in the provision of health care. Preference tends to be given to children and those looking after children (Charny et al, 1989; Busschbach et al, 1993; Nord et al, 1996; Bowling, 1996; Tsuchiya, 1999; Dolan et al, 2005; Brazier et al, 2007). There is also some evidence that people are less likely to give priority to those who are considered to be in some way responsible for their ill health (Neuberger et al, 1998; Jowell et al, 1998) and to give priority to those who are socioeconomically disadvantaged (Dolan et al, 2005; Brazier et al, 2007). However, the results of studies are inconsistent. For example, one survey (Anand and Wailoo, 2000) provided evidence of a strong desire to treat people equally regardless of age, and not to use the health care system as a way of compensating for socioeconomic deprivations.

One of the problems of this field of research lies in the way in which preferences are framed, respondents typically being asked to comment on one factor at a time against health gain, and not the explicit trade-offs between e.g. socio-economic status and age (Dolan et al, 2005). Public opinion may also be pluralistic. For example, respondents to Cookson and Dolan’s study
(1999) supported a broad ‘rule of rescue’ (i.e. giving priority to those in immediate need), maximising the health of the whole community and reducing inequalities in people’s lifetime experience of health, even though these principles conflict. Systematically reviewing the literature, Dolan et al (2005) note a general preference for reducing health inequalities – but not at all costs (our emphasis).

The fact that the public qualify their social value judgements is reflected in a survey which gave an explicit choice between maximising life expectancy and reducing inequalities in life expectancy between the highest and lowest social classes (Tsuchiya and Dolan, 2007). 9.2% of general public respondents gave unqualified support to targeting the worse off; 48.3% would initially target the lowest social class but would switch if the sacrifice to overall health was too great. These results are similar to those of NICE Citizen’s Council meeting in which panel members were asked whether NICE should issue guidance that concentrates resources on (1) improving the health of the whole population (which may mean improvement for all groups) even if there is a risk of widening the gap between the socio-economic groups; or (2) trying to improve the health of the most disadvantaged members of our society, thus narrowing the gap between the least and most disadvantaged, even if this has only a modest impact on the health of the population as a whole (NICE 2006). 58% of panel members (n=26) supported the second option, concluding that it would counter the discrimination experienced by disadvantaged groups; and that early intervention is more effective and, in the long run, more cost-effective. They did, however, add certain caveats: that promoting health equity should not be to the detriment of the whole population (i.e. that the gap should be narrowed upwards, not downwards); that needs should be properly identified and schemes properly monitored; and that no group should be positively excluded from access to resources (NICE, 2006). The Council’s conclusions which refer to resources being allocated in response to clinical need (i.e. health care equity) are also suggestive of a pluralistic position:

During our final discussion, one broad strategic suggestion earned the approval of many of us - and perhaps most of us . . . . One member summarised it like this: “We should concentrate on the health needs of people, not on their ethnicity, or their class, or their geographical location,” In other words, under most circumstances the process of allocating resources should be driven first and foremost by the identification of a health condition worth tackling (our emphasis); only then should we consider who suffers from it, and whether resources
need to be targeted disproportionately to particular sections of the community. If those individuals happen to belong to a particular ethnic or socio-economic group – so be it. Target them (NICE, 2006, p.15).

Is the pendulum swinging back?

2010-11 perhaps represented the apotheosis of attempts to prioritise health equity in the allocation of resources for Hospital and Community Services (HCHS) in England. Since then there have been a number of important changes to the allocation process. In 2011, a new approach was introduced, which calculated allocations for acute services using a diagnosis-based capitation model (Dixon et al, 2011). Since refreshed using more recent data and re-estimating the models to produce updated weights for different drivers of need (NHS England, Analytical Services (Finance) (2016), there has been a subtle shift in core funding targets that suggests that the pendulum might be swinging back. Thus, some of the most deprived but also most generously funded CCGs are now considered to be receiving funding above their core needs; while CCGs that have traditionally received lower allocations relative to underlying morbidity are considered to be below target. The 15% unmet need/health inequalities adjustment that was applied to the core CCG formula has been reduced to 10% (although the adjustment remains at 15% for primary medical care).

Organisational changes have also been made to the NHS. In 2013, Public Health England (PHE) was created and responsibilities for public health (including encouraging healthier lifestyles and reducing health inequalities) transferred from the NHS to local government with a view to promoting a closer relationship between public health teams and those involved in the distal determinants of health – for example, environmental health, housing and transport (Kessel and Haines, 2010). The separation of responsibilities for public health and the commissioning of health care services (which is funded by NHS England) means that there is now a more logical structure to support the potentially conflicting principles of health equity and health care equity. However, public health funding available to local authorities (£3.4 billion in 2016-17) is dwarfed by the unmet need/health inequalities adjustments applied to CCG and primary care allocations suggesting that, despite the reorganisation introduced by the Coalition Government, health equity is still linked to the distribution of health care.
The fact that the NHS is now coping with huge clinical and financial pressures (and all that entails in terms of, for example, increased waiting times and cancelled operations) may have also altered public perceptions of the acceptable trade-off between maximising health and reducing health inequalities. There are a number of elements to this. As noted above, the public seems more reluctant to prioritise improving of outcomes for the disadvantaged if this leads to an overall levelling down of health. Awareness of the scarcity of medical resources and the growing use of lifestyle-based rationing in the NHS may harden public attitudes to the treatment of disease associated with unhealthy lifestyles. Finally, it is widely recognised that older people are paying the price for what have been very significant cuts to social care. As a result, while there is much policy interest in developing better solutions for older people within their homes and communities, there seems little appetite for considering this generation to be fair game for the rationing of health care services.

**Conclusion**

Since the Conservative politician Nigel Lawson coined the phrase, it is often said that the NHS is the closest thing the English people have to a religion. There is certainly overwhelming support for the general idea of an NHS. Of those surveyed in the 2014 British Social Attitudes (BSA) Survey, 89% agreed that the government should support a national health system that is tax funded, free at the point of use and provides comprehensive care for all citizens (Gershlick et al, 2015). Yet, public attitudes are more nuanced than the religious analogy suggests and attitudes are shaped by the wider political context. During the years of the previous Labour Government (1997-2010) when spending on the NHS more than doubled, it is perhaps not surprising that voters were supportive of targeting health care resources towards the disadvantaged. During a time of austerity, a climate of *sauve qui peut* politics (everyone for themselves) may have hardened attitudes towards the redistribution of spending (Cramme et al, 2013).

In the midst of shifting public attitudes and policy responses, academia in the UK remains strongly committed to the goal of health equity. Remarkably little empirical research is conducted on inequalities in access to and use of health care compared to inequalities in health status. Since the publication of the NHS Atlases of Variation in Healthcare series (http://fingertips.phe.org.uk/profile/atlas-of-variation), there is growing recognition of
unwarranted variation in access to NHS care. Deprivation has not emerged in the Atlases as a significant predictor of variation in care. However, there are other dimensions of inequity in access – e.g. by age, sex, ethnicity and rurality - which are legitimate points for inquiry in a system that is ostensibly designed to provide equal access for equal clinical need. The relative lack of research on such variation puts British academia at odds with both public attitudes and the direction of policy making. Perhaps it is time for the research community to consider the case for rebalancing the pendulum and reasserting the case for horizontal equity.

To this end, it is important to acknowledge that a concern with socio-economic disadvantage is not inconsistent with the principle of equal access for equal needs. The key is how one recognizes and measures ‘need’. For example, it is plausible that, for patients with an identical condition, need for health care will be higher for poorer than more affluent groups. Several studies note that length of hospital stay is significantly associated with social deprivation (Hollowell et al, 2010; Cookson and Laudicella, 2011; Faiz et al, 2011; De Bruijne et al, 2013). A number of factors may account for this, such as a lack of adequate support at home for early discharge; differences in patient behaviour (e.g. with respect to adherence to medication and physical recovery regimes); differences in disease severity (e.g. due to late presentation); and differences in co-morbidity, resulting in additional treatment costs through poorer functional impairment or presenting contra-indications to standard therapies (Charlson et al, 2008; Cunic et al, 2014). The point is that such considerations are suggestive of levels of health care need and can be factored into the case for targeting greater resources to disadvantaged groups on horizontal equity grounds.

Such an approach would be more philosophically coherent than the uneasy fudge that results from claims to promote health care and health equity, an approach that has medicalised health inequalities, exacerbated inequalities in service provision, treated older, less deprived populations as somehow less ‘deserving’ than their younger, deprived counterparts and which is increasingly out of step with public understandings of a national health service.
References


Coast J, Donovan J, Litvab A, Eyles J, Morgan, M, Shepherd, M., Tacchie, J. (2002). “If there were a war tomorrow, we’d find the money”: contrasting perspectives on the rationing of health care, *Soc Sci Med* 54 1839–1851


ONS (undated, a). *Mid-year estimates for 2014 (Mid-2014: SAPE17DT5).* [Website](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/clinicalcommissioninggroupmidyearpopulationestimates)

ONS (undated, b). *Mid-year estimates for 2016 (Mid-2016: SAPE19DT5)* [Website](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/clinicalcommissioninggroupmidyearpopulationestimates)


