UNIVERSITY *of York*

Discussion Papers in Economics

No. 25/01

Healthcare, Instability Risk and Cost Increases.

David Mayston

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

HEALTHCARE, INSTABILITY RISK AND C0ST INCREASES

David Mayston

1. Introduction

In this paper we examine conditions under which instability risk and sustained labour cost increases can arise in a healthcare system that seeks to combine public sector healthcare provision that is free at the point of delivery with the availability of elective private healthcare at a positive user price. A key role is played by potential labour market shortages for public sector providers in impacting on their quality of healthcare, and by the density function of consumers' net income to finance alternative private healthcare. Variables and parameters which influence the tipping point between stable and unstable outcomes can then be identified. The extent to which macro-economic increases in the rate of economic growth in the wider economy can overcome potential difficulties will also be examined.

2. The Model

We will assume that the inputs into healthcare are labour inputs, with different grades of labour used in fixed proportions, and an associated weighted wage level of w in the public sector. The wage level in the private sector in comparison is $(1+\zeta_{00})w$, where $\zeta_{00} \geq 0$ reflects a small wage premium that the private sector pays to attract staff to ensure that it is fully staffed at its desired level of unit resourcing β_1 per unit per unit of its treatment, so that it can service its resultant healthcare demands without significant delays or other quality failures.

Whilst healthcare treatments are provided free of charge to patients in the public sector, the private sector levies on its patients a user charge that reflects its unit costs, plus a profit mark up of $\zeta_0 \geq 0$ per unit of its treatment costs that reflects any additional market power that it has in setting its price for providing private healthcare. The price charged to patients per unit of elective healthcare treatment in the private sector is then:

$$p_P = \zeta_1 w \;\; where \;\; \zeta_1 = (1+\zeta_0)(1+\zeta_{00})\beta_1 > 0 \qquad (1)$$

The public sector healthcare provider seeks to provide its elective healthcare using a level of resources per unit of treatment of $\alpha_1 < \beta_1$ that is less than the private sector considers it needs to attract consumers into the private sector. Resource pressures in the public sector may in comparison lead the public sector to engage in less elaborate testing procedures and medication than the private sector, where such additional resourcing levels per patient may boost its financial profitability. While α_1 represents the governmental prescribed level of resourcing per unit of treatment for an elective patient in the public sector, its receipt by the patient may be subject to delays which reduce the effective quality of care provided by the public sector and the patient's health gain from the treatment compared to it being available without delay. The extent of these delays and any other quality of care impacts will depend upon the availability of healthcare staff to the public sector, which in turn depends upon a number of factors which we will explore.

Consumers with elective healthcare needs face a decision of whether or not to opt for private healthcare, with an associated price of $\zeta_1 w$ per unit of treatment to be provided without delay, or obtain it for free in the public sector, though after a possible delay on a public sector waiting list. Consumers' willingness to pay for private healthcare then reflects these considerations and their net income to pay for the higher cost to them of private healthcare. In this paper we will focus on consumer net income above a critical hurdle level as the prime driver of the demand for private elective health care. Since this higher cost may be significant compared to their current income, the relevant net income involves their income from salaries, pensions and investments net of the other expenditure items they wish to protect, such as food and housing. Consumer income was indeed found by Propper (2000) to be a major determinant of the demand for private healthcare, though with other demographic, attitudinal and past use variables also found to be empirically significant.

For simplicity we will assume initially that elective healthcare consumers' net income per capita is normally distributed, with a frequency density function $f(y)$ around a mean level of μ, with a standard deviation of σ. Negative values of individual available net income to spend on private healthcare are then possible in the population at large if some individuals have limited total income and savings, and limited budgets to pay for other commodities. The associated total frequency function of y is given by $g(y) = mf(y)$, where m is the total population size of those needing elective healthcare treatment. The level of net income of the marginal consumer of private healthcare will be designated by y_0, such that consumers in need of elective healthcare opt for private healthcare provision if their net income equals or exceeds y_0, but rely upon the public healthcare system for it if their available net income falls short of this hurdle level.

In this paper, we will assume for simplicity that healthcare needs for elective treatment from one source or another are themselves independent of income and predictable in advance, with competition between private healthcare insurers resulting in an equality between private insurance premiums and the price that the individual would have to pay for the treatment in the private sector if they paid for it directly themselves. We will also assume initially that the need for emergency healthcare is independent of the availability of elective treatment, and is met without excessive delay by the public sector as a matter of priority.

The total level of desired demand for healthcare labour by the public sector is then:

$$D_N = \alpha_0 + D_{NE} \; where \; D_{NE} = \alpha_1 m \int_{-\infty}^{y_0} f(y)dy \tag{2}$$

where D_{NE} is the volume of healthcare labour needed to treat all of its patients who are in clinical need of elective healthcare, at its target standard unit of per capita resourcing α_1. α_0 reflects the additional resource needs of the public sector for providing non-elective emergency healthcare, and the healthcare training and education which they provide.

The total level of desired demand for healthcare labour by the private sector is:

$$D_P = \beta_1 \int_{y_0}^{\infty} g(y)dy = \beta_1 m \int_{y_0}^{\infty} f(y)dy \quad (3)$$

The critical level, y_0, of net income which influences consumers' decisions on whether to pay for private healthcare or opt for free public healthcare is assumed to depend up the price, p_P, per unit of treatment charged by the private sector, and the quality of elective healthcare that public sector patients face on a potential waiting list for treatment, as reflected in the quality of timely treatment index Q_N, with:

$$y_0 = \gamma_0 + \gamma_1 Q_N + \gamma_2 p_P \quad where\ \gamma_1 > 0, \gamma_2 > 0 \quad (4)$$

The higher the price per unit of treatment that consumers would have to pay for private healthcare, the higher the critical level of net income at which consumers are willing to opt for private health care. In addition, the higher is the quality of timely treatment that is available in the public sector the higher is the critical level of net income at which consumers are willing to opt for private health care, and the lower is the demand for private healthcare in (3).

At any wage level w, the total supply of healthcare labour is assumed to be given by:

$$S_H = \theta_1 w + \theta_{00} \quad \text{with } \theta_1 > 0, \theta_{00} < 0 \quad (5)$$

As noted above, the private sector is assumed to offer a small wage premium to ensure that the healthcare provision for its level of desired demand D_P is fully staffed without delays in its treatment. The public sector is then able to recruit a total amount of healthcare labour of $S_N = S_H - D_P$ at the wage level w. This differs from the amount of healthcare labour needed to promptly service the desired demand D_N for healthcare which it faces by:

$$S_H^N = S_N - D_N = S_H - D_P - D_N = \theta_1 w + \theta_0 - D_P - D_{NE} \quad where\ \theta_0 = \theta_{00} - \alpha_0 \quad (6)$$

which, if it is negative, represents a labour shortage that will impair the ability of the public sector to service all of its desired demand without delay and with an adequate level of staffing. If we define $\hat{S}_H^N = -S_H^N$, the magnitude of the *labour shortage* $\hat{S}_H^N$ *facing the public sector* is here equal to the *total excess demand for healthcare*, given by:

$$X_H = D_H - S_H = D_N + D_P - S_H = D_N - S_N = -S_H^N = \hat{S}_H^N, where\ D_H = D_N + D_P \quad (7)$$

with the private sector able to secure its desired labour supply, and the public sector experiencing the impact of an excess demand for healthcare through a shortfall of its available healthcare labour, S_N, at the prevailing wage rate w compared to the volume required to fully service without delay the desired demand, D_N, for elective healthcare which the public sector faces. The extent of the shortfall $\hat{S}_H^N = D_N - S_N = X_H$ will then be reflected in the size of the waiting list that arises in the public sector when its available labour supply falls short of that needed to treat all of its patients who are assessed to be in clinical need of elective healthcare. The magnitude, and persistence, of such waiting lists confirm the existence of continuing positive excess demand for elective healthcare, as a symptom of a *persistent disequilibrium* between the available healthcare labour supply at the prevailing wage level and that needed to service the desired demand for elective healthcare. There can therefore be no necessary assumption that equilibrium will prevail between such desired demand for elective healthcare and the available labour supply.

The perceived quality of elective healthcare provided by the public sector is assumed to be a decreasing function of the extent of its labour shortage $\hat{S}_H^N$, such that:

$$Q_N = (\phi_0 - \phi_1 \hat{S}_H^N (1+\phi_2)) \ where \ \phi_1 > 0, \ \phi_2 \geq 0 \tag{8}$$

The resultant pressures on public sector healthcare labour availability, compared to its demand, from positive labour shortages are likely to be compounded by their effect on staff workloads. Greater values to $\hat{S}_H^N$ imply increased staff workloads that may increase staff sickness and absence rates, that in turn may have a damaging multiplier effect on the quality of health care index Q_N , as reflected in the ϕ_2 term in the Q_N index in (8). Conversely, positive values to the labour availability $S_H^N = -\hat{S}_H^N$ compared to demand will ease such workload pressures, and tend to boost the perceived quality of healthcare provided by the public sector.

In equation (4) above, what matters in the determination of the hurdle level of consumer income y_0 , at which consumers opt for private healthcare, is the *perceived* level of healthcare quality in the public sector. This may be influenced by the size of the waiting list for elective healthcare in the public sector that receives widespread publicity from readily available nationally published statistics. Such is the case, for instance, in the UK's public sector National Health Service (NHS), where the total NHS waiting list for consultant-led elective care in England is published monthly, and has risen from 4.19 million in August 2007 to 7.64 million in August 2024 (BMA, 2024). A positive association between the decision to opt for private healthcare insurance and the length of NHS waiting lists was indeed found by Besley, Hall and Preston (1996).

Key determinants of the size of waiting lists are likely to be the extent of the public sector healthcare labour shortage and the excess demand for healthcare which it faces. Moreover, labour shortages for public sector healthcare are likely to manifest themselves not just in the size of waiting lists, but also in difficulties experienced in accessing GP appointments, and in making initial and continuing primary care contact (see, for example, Demos (2023)), that add to reductions in the perceived quality of healthcare that is available in the public sector.

3. Implications

From equations (1) – (8), we can derive the impacts of an increase in the healthcare wage level upon the quality index, Q_N , and the critical level of net income, y_0 , to be:

$$\partial Q_N / \partial w = \phi_1 (1+\phi_2)[\theta_1 + (\beta_1 - \alpha_1) m f(y_0) \partial y_0 / \partial w] \tag{9}$$

and $\partial y_0 / \partial w = \gamma_1(\partial Q_N / \partial w) + \gamma_2\zeta_1 = \Phi + Z_1(\partial y_0 / \partial w)$ (10)

$where\ \Phi \equiv (\gamma_1\phi_1(1+\phi_2)\theta_1) + \gamma_2\zeta_1 > 0\ \ and\ \ Z_1 \equiv \gamma_1\phi_1(1+\phi_2)(\beta_1 - \alpha_1)mf(y_0)$ (11)

Increasing the level of healthcare wages will increase the total supply of healthcare labour, as reflected in the parameter θ_1 in equations (5) and (9). If the higher wage level also results in a higher critical level of net income at which consumers opt for private healthcare, the supply of labour to the public sector will also be boosted by less labour being absorbed by the private sector than the public sector uses per patient, as reflected in the $(\beta_1 - \alpha_1)mf(y_0)$ term in (9), where $mf(y_0)$ is the number of patients who are on this margin. If the overall effect is to increase the supply of healthcare labour to the public sector, the extent of any labour shortage and of workload pressures will decline, and the quality of care in the public rector will rise in equation (9).

However, the impact of a rise in the healthcare wage level on the critical value, y_0, of their net income at which individuals opt for private healthcare in (10) itself depends on how this wage rise impacts on the public sector quality index, Q_N, which in turn depends in (9) on the magnitude and sign of the impact of the wage rise on the critical value y_0, thereby creating a *feedback effect* of $\partial y_0 / \partial w$ upon itself in equation (10) . The size of this feedback effect is reflected in the value of Z_1 in (10) and (11), which imply:

$\partial y_0 / \partial w = \Phi / (1 - Z_1)\ \ for\ Z_1 \neq 1\ and\ hence\ (\partial y_0 / \partial w) <,> 0\ as\ Z_1 >,< 1$ (12)

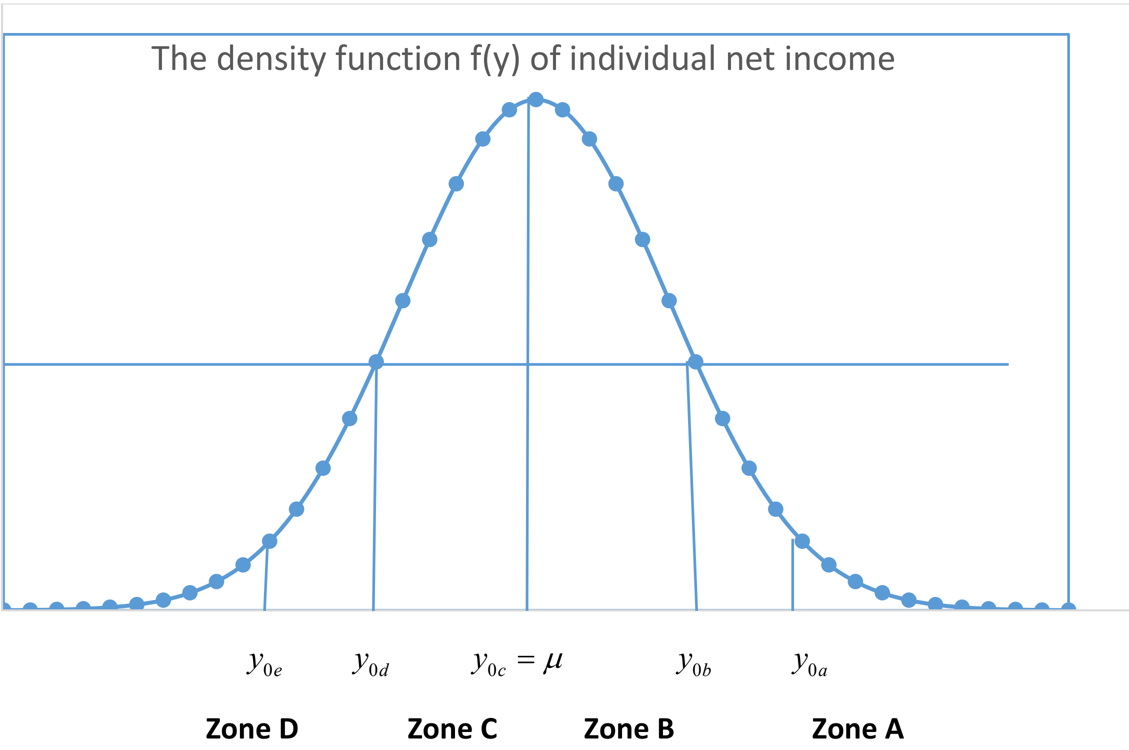

FIGURE 1

A rise in healthcare wages will therefore result in a lowering of the critical value y_0 of patients' net income, and an increase in the numbers opting for private elective healthcare, whenever $Z_1 > 1$. Whether or not $Z_1 > 1$ does prevail in turn depends in (11) upon where in the distribution $f(y)$ per capita of net income the critical value y_0 falls. It follows from (11) that:

$$Z_1 >,=,< 1 \; as \; f(y_0) >,=,< Z_0, where \, Z_0 \equiv 1/[\gamma_1\phi_1(1+\phi_2)m(\beta_1 - \alpha_1)] > 0 \quad (13)$$

so long as $\alpha_1 < \beta_1$. If instead $\alpha_1 \geq \beta_1$, $Z_1 \leq 0$ in (11) and $\partial y_0 / \partial w > 0$ in (10). However, since $f(y)$ obtains its maximum value of $(1/\sqrt{2\pi\sigma})$ when $y = \mu$, if

$$0 < Z_0 < (1/\sqrt{2\pi\sigma}) \, and \, hence \; \gamma_1\phi_1(1+\phi_2)m(\beta_1 - \alpha_1) > \sqrt{2\pi\sigma}) \quad (14)$$

there is a non-trivial range of values of y_0 for which $f(y_0) > Z_0$ and hence $Z_1 > 1$ and therefore $(\partial y_0 / \partial w) < 0$. In Figure 1, this is illustrated by the range of values of y_0 between the points y_{0b} and y_{0d} at which $f(y_0) = Z_0$.

Around the mean value $y_{0c} = \mu$ of $f(y)$, these values define a number of alternative Zones and intermediary points that are relevant to the consideration of the implications of different values of y_0. Specifically, we have in Figure 1 from (12) and (13):

Zone A:	$y_0 > y_{0b}$	$f(y_0) < Z_0$	$Z_1 < 1$	$\partial f / \partial y_0 < 0$
Point b:	$y_0 = y_{0b}$	$f(y_0) = Z_0$	$Z_1 = 1$	$\partial f / \partial y_0 < 0$
Zone B:	$y_{0c} < y_0 < y_{0b}$	$f(y_0) > Z_0$	$Z_1 > 1$	$\partial f / \partial y_0 < 0$
Point c:	$y_0 = y_{0c} = \mu$	$f(y_0) > Z_0$	$Z_1 > 1$	$\partial f / \partial y_0 = 0$
Zone C:	$y_{0d} < y_0 < y_{0c}$	$f(y_0) > Z_0$	$Z_1 > 1$	$\partial f / \partial y_0 > 0$
Point d:	$y_0 = y_{0d}$	$f(y_0) = Z_0$	$Z_1 = 1$	$\partial f / \partial y_0 > 0$
Zone D:	$y_0 < y_{0d}$	$f(y_0) < Z_0$	$Z_1 < 1$	$\partial f / \partial y_0 > 0$

TABLE 1

4. Stability of the Healthcare Labour Market

Since healthcare often involves a labour intensive production process using relatively large amounts of labour, with healthcare costs a significant item of interest for many governments, important questions arise as to whether the healthcare labour market is a stable one. In their well-established text *Introduction to General Equilibrium Theory and Welfare Economics*, Quirk and Saposnik (1968, p. 149) stressed that questions of stability arise even in the standard theory of competitive markets as a way of organising the interface between producers and consumers:

> Study of the stability properties of equilibrium positions is essential to the theory of the competitive mechanism because it is only when stability has been verified that the competitive mechanism can be regarded as a workable device for

generating optimal solutions to the problems of allocating resources to producers and distributing output among consumers.

These remarks imply that, if the relevant competitive market is unstable, such a market cannot be regarded as a workable mechanism for generating optimal solutions to the relevant resource allocation and distribution problems. Central to the analysis of healthcare labour market stability is how the *excess demand* for healthcare labour:

$$X_H = D_H - S_H, \, where \, D_H = D_N + D_P \tag{15}$$

varies with the wage level. From (2), (3), (12), (13) and (15):

$$\partial D_H / \partial w = -(\beta_1 - \alpha_1) m f(y_0)(\partial y_0 / \partial w) \tag{16}$$

with $\partial D_H / \partial w >,< 0 \; as \, (\partial y_0 / \partial w) <,> 0 \; i.e. \; as \, f(y_0) >,< Z_0 \; for \, \beta_1 > \alpha_1$ (17)

Hence from (11) and (12):

$$\partial X_H / \partial w = (\partial D_H / \partial w) - (\partial S_H / \partial w) = Z_2 / (Z_1 - 1) \tag{18}$$

$$where \; Z_2 \equiv [\gamma_2 \zeta_1 (\beta_1 - \alpha_1) m f(y_0)] + \theta_1 > 0 \tag{19}$$

with $\partial X_H / \partial w < 0 \;\; if \;\; f(y_0) < Z_0 \; and \; hence \; Z_1 < 1$ (20)

and $\partial X_H / \partial w > 0 \;\; if \;\; f(y_0) > Z_0 \; and \; hence \; Z_1 > 1$ (21)

	$\partial D_H / \partial w$	$\partial Q_N / \partial w$	$\partial y_0 / \partial w$	$\partial X_H / \partial w$	$\partial^2 D_H / \partial w^2 = \partial^2 X_H / \partial w^2$	$\partial^2 Q_N / \partial w^2$
Zone A: ($y_0 > y_{0b}$)	-	+	+	-	+	-
At point b: ($y_0 = y_{0b}$)	$\hat{\infty}$	$\hat{\infty}$	$\hat{\infty}$	$\hat{\infty}$	$\hat{\infty}$	$\hat{\infty}$
Zone B: ($y_{0c} < y_0 < y_{0b}$)	+	-	-	+	-	+
At point c: ($y_0 = y_{0c}$)	+	-	-	+	0	0
Zone C: ($y_{0d} < y_0 < y_{0c}$)	+	-	-	+	+	-
At point d: ($y_0 = y_{0d}$)	$\hat{\infty}$	$\hat{\infty}$	$\hat{\infty}$	$\hat{\infty}$	$\hat{\infty}$	$\hat{\infty}$
Zone D: ($y_0 < y_{0d}$)	-	+	+	-	-	+

TABLE 2

Further light is shed on the behaviour of the excess demand for healthcare labour and the quality of care available in the public sector by examining their corresponding second derivatives with respect to a rise in the wage level.

$$(\partial^2 X_H / \partial w^2) = (\partial^2 D_H / \partial w^2) = -(\beta_1 - \alpha_1)\Phi m(\partial f / \partial y_0)(\partial y_0 / \partial w) / (1 - Z_1)^2 \quad (22)$$

$$(\partial^2 y_0 / \partial w^2) = \gamma_1(\partial^2 Q_N / \partial w^2) = \gamma_1\phi_1(1+\phi_2)(\beta_1 - \alpha_1)\Phi_0^2 m(\partial f / \partial y_0) / (1 - Z_1)^3 \quad (23)$$

The sign pattern across the different Zones, and at the mid-point c, of these rates of change, and of their acceleration or deceleration, is shown in Table 2. At points b and d, where $Z_1 = 1$, the absolute value of these marginal rates of change tends towards infinity, and is denoted here by $\hat{\infty}$.

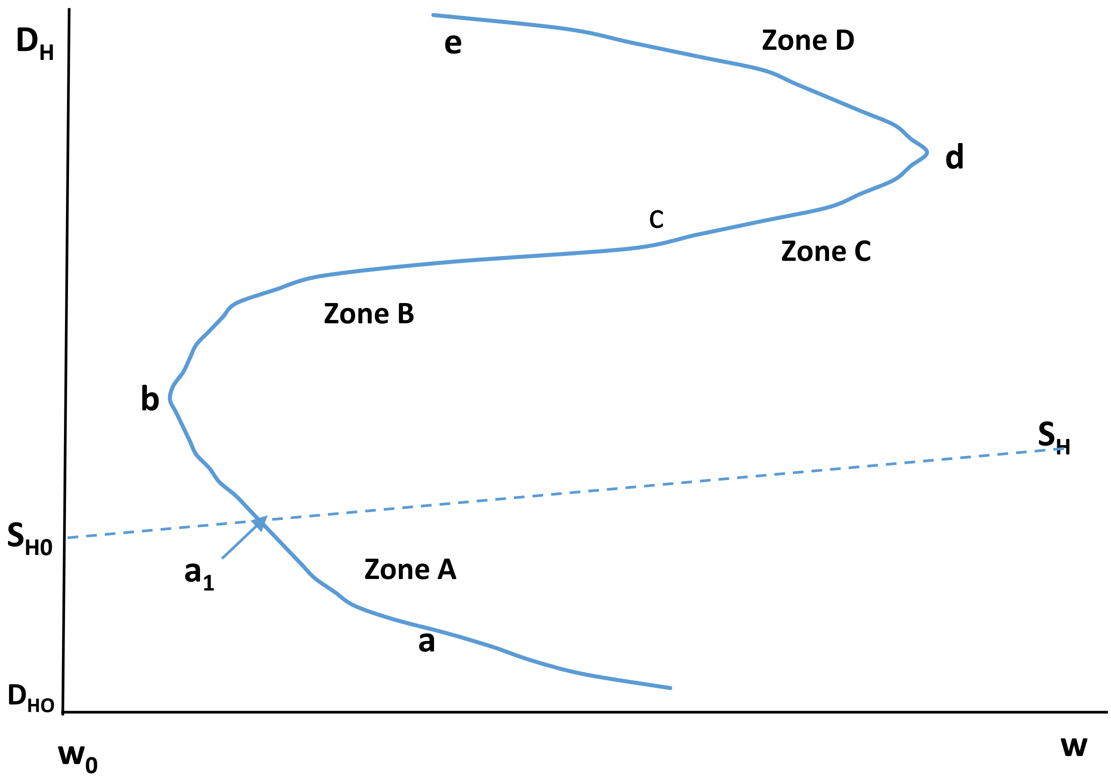

FIGURE 2

The implied pattern of these rates of change in the total demand for healthcare as the wage level increases is illustrated in Figure 2. Within Zone A, the critical level, y_0, of net income at which consumers are willing to opt for private healthcare is relatively high in Figure 1, so that the number of willing private sector patients with net income above this level is relatively low. The total demand for healthcare labour is therefore lower than it would be with a large private sector, with D_{HO} the level of demand for healthcare labour in the absence of private sector demand. A rise in healthcare wages within Zone A increases the total supply of healthcare labour. Within Zone A this has the *self-reinforcing* effects that the quality of healthcare the public sector is able to offer rises, fewer consumers want to opt for private healthcare and in the process absorb more healthcare labour in the private sector than they would in the public sector, and as a result the *net supply* of healthcare labour to the public sector increases. Within Zone A, the density function, $f(y_0)$, of consumers who are on the margin of opting for private healthcare is sufficiently low that $Z_1 < 1, (\partial y_0 / \partial w) > 0$, and $(\partial Q_N / \partial w) > 0$ in equations (9), (12) and (13), so that within Zone A these interacting effects pull in the same direction and are self-reinforcing.

However, if in contrast the wage level declines from a point such as **a** that is well inside Zone A towards a point such as point **b** that is on the boundary with Zone B, there will be a decline in the total supply of healthcare labour, S_H, as wages decline, as reflected in the parameter θ_1 in equations (5) and (9). The decline in the net supply of healthcare labour to the public sector will be reinforced in equations (10) and (11) by the $\gamma_2\zeta_1$ term, which reflects the greater affordability of private healthcare in equations (1) and (4) when the wage level drops. This is reinforced in equation (9) by the positive value to $\partial y_0 / \partial w$ within Zone A in Table 2 that implies a resultant fall in the critical level, y_0, and a further boost to private sector demand when w declines. There will then be a further decline in the net supply S_H^N of healthcare labour to the public sector and in its associated quality of care, Q_N, in equations (9) – (11) when $\beta_1 > \alpha_1$, so that the private sector absorbs more labour than it releases in the public sector when consumers switch, to an extent that depends upon how many consumers, i.e. $mf(y_0)$, are at this critical boundary. Table 2 confirms that the boost to overall total healthcare demand for labour, and the decline in the quality of public sector health care, will occur at an increasing rate within Zone A if the wage level declines.

Figure 2 illustrates a situation in which the total supply of healthcare labour curve intersects with the total demand for healthcare labour curve, D_H, at a point such as **a₁** within Zone A. The Hicksian condition for *perfect (local) stability* (see Quirk and Saposnik, 1968, p. 153) that the excess demand declines as wages rise, and rises as wages decline, i.e. $\partial X_H / \partial w < 0$ in Figure 2 will then prevail if there is an equilibrium between the supply and demand for healthcare labour at a point, such as **a₁**, within Zone A. This point may also be a *locally dynamically stable* equilibrium point under a Walrasian tâtonnement adjustment mechanism (see Quirk and Saposnik, 1968, pp. 160-161) in which the quoted wage level rises if there is excess demand for healthcare labour and falls if there is positive excess supply in the neighbourhood of any such equilibrium point. Table 2 also implies that within Zone A the excess demand for healthcare labour will increase at an increasing rate if the wage level declines away from its equilibrium position, potentially then leading to a greater wage rise back towards the equilibrium of zero excess demand if such a wage rise is funded.

The Walrasian tâtonnement adjustment mechanism itself assumes the existence of an auctioneer who makes price adjustments when the existing price does not equate supply and demand, and market participants only make transactions if the price on offer is an equilibrium price (see Arrow and Hahn, 1971, p. 264-5). In the healthcare labour market, participants make supply and demand decisions even in the presence of excess demand. Dynamic behaviour can then be incorporated into the model if we assume that labour market pressures imply that the rate at which healthcare wages rise is positively related to the extent of the excess demand for healthcare, as in the case where:

$$dw / dt = \psi_0 X_H \quad where\ \psi_0 > 0 \tag{24}$$

We then have from equations (18) and (19), which include the impact on X_H of variations in y_0 and Q_N in equations (9)-(13) and (15)-(19):

$$dX_H / dt = \psi_1 X_H \quad where\ \psi_1 = (Z_2 \psi_0 / (Z_1 - 1)) >,< 0\ as\ Z_1 >,< 1 \tag{25}$$

In Zones A and D where $Z_1 < 1$, we have $\psi_1 < 0$ and hence stabilising behaviour in the form of reductions over time in the excess demand for healthcare labour when the excess demand is positive. From (7) and (8), we also have:

$$dQ_N / dt = -\psi_2 X_H \quad where\ \psi_2 = \phi_1 (1 + \phi_2)(Z_2 \psi_0 / (Z_1 - 1)) >,< 0\ as\ Z_1 >,< 1 \tag{26}$$

In Zones A and D we have $\psi_2 < 0$ if condition (24) holds, with the associated stabilising reduction in the excess demand for healthcare labour in (25) when $X_H > 0$ driving up the quality of healthcare Q_N in the public sector. From (11) and (12), we also have:

$$dy_0 / dt = -\psi_3 X_H \quad where\ \psi_3 = \Phi \psi_0 / (Z_1 - 1)) >,< 0\ as\ Z_1 >,< 1 \tag{27}$$

In Zones A and D we have $\psi_3 < 0$, so that the critical hurdle level of net income at which consumers are on the margin of whether or not to opt for private healthcare rises, reducing private sector healthcare demand and reinforcing the fall in the excess demand for healthcare labour. However, the adjustment mechanism in equation (24) assumes a willingness by the public sector to sufficiently fund healthcare wage increases until the excess demand for healthcare labour is eliminated, a condition which may not hold in practice. If instead healthcare wage levels are allowed to decline below the equilibrium level at $\mathbf{a_1}$, Table 2 implies an accelerating rate of increase in the excess demand for healthcare, and increase in the total demand for healthcare, as the private sector expands and the quality of care offered by the public sector declines at an increasing rate within Zone A.

5. Instability in the Healthcare Labour Market

If the healthcare wage level is allowed to decline further towards the level at point **b** in Figure 2, the associated fall in the hurdle level of income, y_0, in Table 2 will encounter a rising value of the density function $f(y_0)$ in Figure 1. Once the relative number of consumers who are on the margin of opting for private healthcare, as reflected in the density function, $f(y_0)$, has risen to exceed the critical value Z_0, the healthcare labour market, however, is no longer in Zone A but instead falls into Zone B in Figures 1 and 2, and Tables 1 and 2. The positive feedback effect of $\partial y_0 / \partial w$ upon itself in equation (10) now dominates, with $Z_1 > 1$. Rather than a rise in the wage level increasing the critical value of y_0, it now reduces it in equation

(12), further boosting private healthcare labour demand and reducing the net supply of healthcare labour S_H^N that is available in the public sector compared to the level needed to service its demand for healthcare in equations (2), (6) and (8). Since from equations (1), (4) and (10):

$$\partial Q_N / \partial w = [(\partial y_0 / \partial w) - \gamma_2 \zeta_1] / \gamma_1 \textit{ where } \gamma_1 > 0, \gamma_2 > 0, \zeta_1 > 0 \tag{28}$$

such a negative value to $\partial y_0 / \partial w$ will be associated with a declining quality of care in the public sector as wages rise, with its rate of decline accelerating in Zone C in equation (23) and Table 2.

If there is a point at which the total supply of healthcare labour curve intersects the total demand for healthcare labour curve at a point, such as **b₁** in Figure 3, within Zones B or C, it will be dynamically unstable if the above adjustment mechanism prevails. An increase in the wage level beyond this point then reduces the critical value of y_0 and increases the total demand, and the excess demand, for healthcare labour, putting further upward pressure on wages in an upward spiral of healthcare costs. However, a fall in the wage level, below the wage level at any intersection point where the excess demand for healthcare labour is initially zero, will result here in an excess supply of healthcare labour when $\partial X_H / \partial w > 0$ in Zones B and C. There will then be further downward pressure on healthcare wages.

Rather than moving back to the wage level at which the supply and demand for healthcare labour functions intersect, the healthcare wage level will move progressively away from it, contrary to the requirements for local and global dynamic stability (see Quirk and Saposnik, 1968, p. 162; Balasko, 1988, p. 245). Thus, under a value of ψ_0 in (24) that is positive though not necesssarily constant, once $Z_1 > 1$ in Zones B and C, we have $\psi_1 > 0$ in (25), given that $Z_2 > 0$ from (19). Positive excess demand will then lead to further destabilising positive increases in excess demand, and positive excess supply will lead to further destabilising increases in excess supply.

Whether healthcare wages are under labour market pressure to spiral upwards or downwards within Zones B and C then depends critically upon where any such intersection point is. If the total supply of healthcare labour is insufficient to meet a large proportion of the additional total demand for healthcare labour that arises within Zones B and C, an upward spiral of labour costs is predicted at many wage levels. However if the total supply of healthcare labour is sufficient to meet a large proportion of this additional total demand, a downward spiral of labour costs is predicted at many wage levels. Nevertheless, where the total demand for healthcare labour exceeds the total supply, progressively rising healthcare wage levels, and declining quality of healthcare in the public sector, are again implied within Zones B and C. The expanding size of the private healthcare sector within Zones B and C will also reduce any monopsony power of the public sector to hold down healthcare wages. If the private healthcare sector market is competitive, the above adjustment process may then prevail with healthcare wages responding positively to positive healthcare excess demand, and dynamic instability existing within Zones B and C.

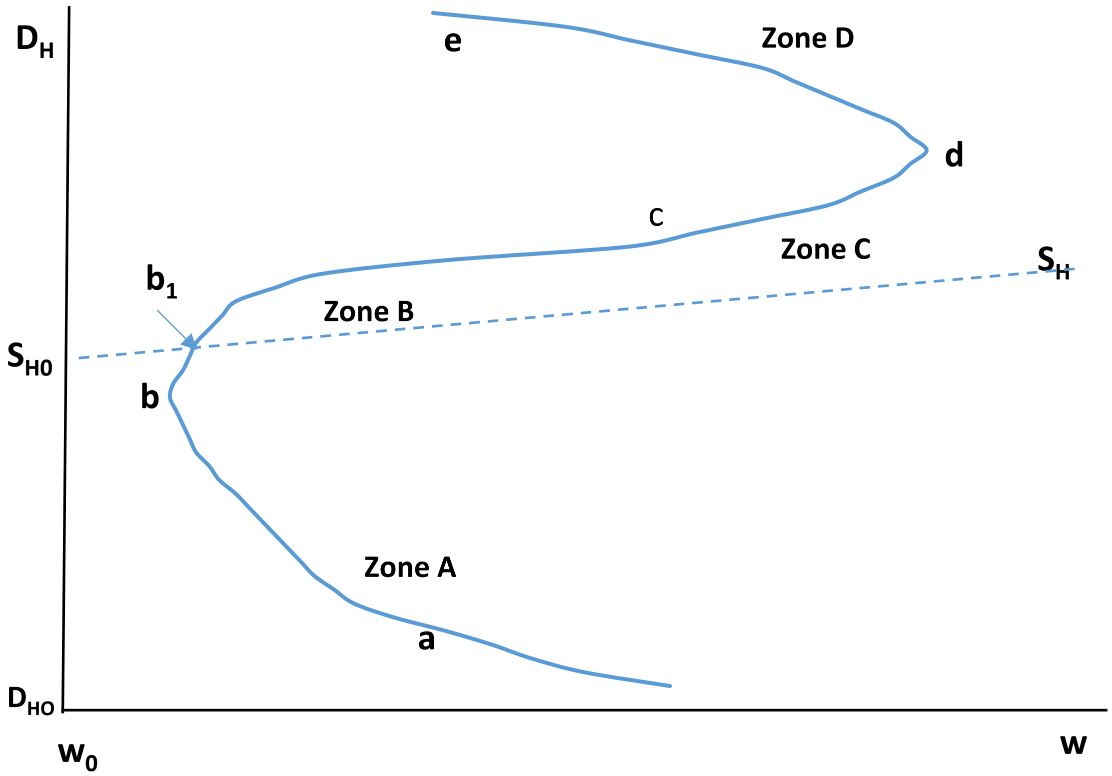

FIGURE 3

Within Zone D, the quality of elective healthcare that is available free of user charges in the public sector has fallen to a level that the majority of the population, i.e. those with net income above the relatively modest level y_{0d} in Figure 1, find unattractive compared to opting for private healthcare with its positive user charges. There is then a strongly two-tier healthcare system in which the majority of the population find it necessary to pay for their elective healthcare privately and the public healthcare service provides a significantly lower quality of elective healthcare to those with relatively low levels of net income, but still substantial elective healthcare needs.

If the total supply of healthcare labour is sufficiently great that the supply of labour curve intersects with the total demand for healthcare labour at a point, such as $\mathbf{d_1}$ in Figure 4, within Zone D, then the finding in Table 2 that $(\partial X_H / \partial w) < 0$ within Zone D implies that such a point will be locally dynamically stable under a Walrasian tâtonnement adjustment process. The same implication follows from equation (25). However, the level of healthcare wages and unit costs facing the public sector will be potentially much higher at an intersection point such as $\mathbf{d_1}$ in Zone D in Figure 4 than would have prevailed at an intersection point such as $\mathbf{a_1}$ in Zone A in Figure 2.

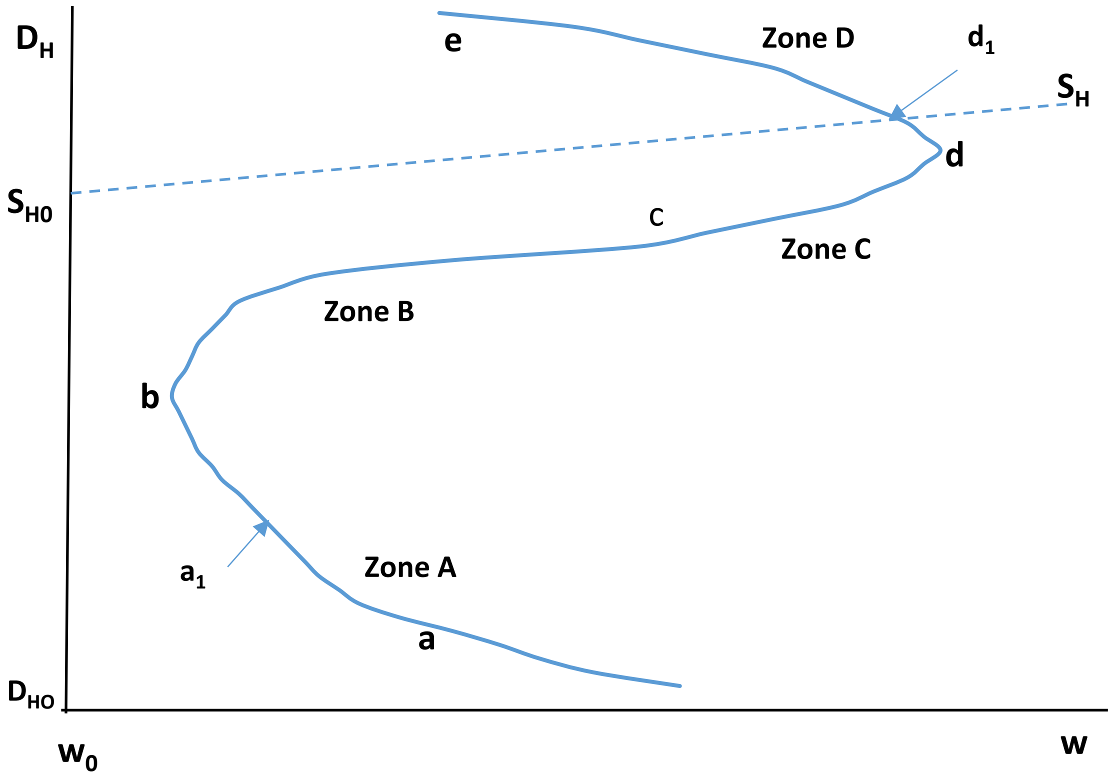

FIGURE 4

However, if the total supply of healthcare labour is not sufficiently great that the supply of labour curve intersects with the total demand for healthcare labour at a point within Zone D, it may instead intersect the total demand curve at a point, such as $\mathbf{b_1}$ in Figure 3, within Zones B or C. The implication in Table 2 that $(\partial X_H / \partial w) > 0$ within Zones B and C in turn implies sustained upward pressure on healthcare wages and unit costs within Zones B and C once the total level of healthcare demand exceeds that at the intersection point within Zones B or C. However, this upward pressure on healthcare wages will not lead to an equilibrium between the total demand for healthcare labour and the available total supply in a competitive labour market. Instead the implication of increasing excess demand for healthcare labour as wages rise, $(\partial X_H / \partial w) > 0$, within Zones B and C is of a persistent and growing positive gap between healthcare labour demand and the available supply over this range.

Such persistent positive excess demand would imply a disequilibrium in a competitive labour market. However, for profit maximising firms, it provides the opportunity for substantial excess profits if they have positive market power in hiring healthcare labour at the relatively low competitive wage level and selling it at a higher price, either as agency staff or through their direct provision of private healthcare at a higher price than that implied by the relatively low competitive wage level.

We can consider here the private sector healthcare price p_P that will maximise private sector profits $\Pi_P = (p_P - w)D_P$ at any given wage level w. From (3) – (8) we have:

$$\partial\Pi_P / \partial p_P = D_P + (p_P - w)(\partial D_P / \partial p_P) = D_P - (p_P - w)\beta_1 m f(y_0)(\partial y_0 / \partial p_P) \quad (29)$$

$$\partial y_0 / \partial p_P = \gamma_1(\partial Q_N / \partial y_0)(\partial y_0 / \partial p_P) + \gamma_2 \quad where\ \partial Q_N / \partial y_0 = \phi_1(1+\phi_2)(\beta_1 - \alpha_1) m f(y_0) \quad (30)$$

Hence from (11):

$$\hat{y}_0 \equiv \partial y_0 / \partial p_P = \gamma_2 / (1 - Z_1) \ \ for\ Z_1 \neq 1 \ and\ hence\ \hat{y}_0 <,> 0 \ as\ Z_1 >,< 1 \quad (31)$$

In Zones B and C, we have $Z_1 > 1$ and therefore $\hat{y}_0 < 0$. So long healthcare labour can be recruited at a wage that is less than p_P, we then have $\partial\Pi_P / \partial p_P > 0$ in (29) within Zones B and C. Increasing the private sector price of healthcare throughout Zones B and C is therefore implied here by profit-maximising behaviour.

Within Zones B and C, relatively large numbers of consumers in Figure 1 and (30) are on the margin of opting for private healthcare. Any fall in the level y_0 of net income that defines this margin will then lead to a relatively large increase in private sector demand for healthcare labour, net of the labour which it releases in the public sector when individuals move to the private sector, as reflected in the $(\beta_1 - \alpha_1) m f(y_0)$ component of (30). Such a boost to the net private sector demand for healthcare labour will then lead to a fall in the quality of care Q_N that is available in the public sector, with $\partial Q_N / \partial y_0 > 0$ and $(\partial y_0 / \partial p_P) < 0$ in (30) and (31), further reinforcing this boost to private sector healthcare demand and its profit level.

However, once the level of y_0 drops to enter Zone D in Figure 1, we now have $Z_1 < 1$ and $\hat{y}_0 > 0$ in (29) and (31), with a private sector profit-maximising price of:

$$p_P^* = w + \{D_P(1 - Z_1) / [\beta_1 m f(y_0)\gamma_2]\} \quad (32)$$

above the cost of its labour input. The second-order condition for a private sector profit maximum that

$$\partial^2\Pi_P / \partial p_P^2 = -2\beta_1 m f(y_0)\hat{y}_0 - (p_P - w)\beta_1 m[(\partial f(y_0) / \partial y_0)\hat{y}_0^2 + f(y_0)(\partial^2 y_0 / \partial p_P^2)] < 0 \quad (33)$$

is satisfied within Zone D, given that $(\partial f(y_0) / \partial y_0) > 0$ and from (11) and (31):

$$(\partial^2 y_0 / \partial p_P^2) = [\gamma_2 / (1 - Z_1)^2][\gamma_1\phi_1(1+\phi_2)(\beta_1 - \alpha_1) m(\partial f / \partial y_0)\hat{y}_0] > 0 \quad (34)$$

both prevail within Zone D.

Within Zone A, $Z_1 < 1$ and $\hat{y}_0 > 0$ also prevail, with the first-order condition $\partial\Pi_P / \partial p_P = 0$ again implying (32). However, the second-order condition (33) for a private sector profit maximum may not hold within Zone A, with $(\partial f(y_0) / \partial y_0) < 0$ in Figure 1 and (33) – (34), and the prospect of significantly greater private sector profits prevailing in Zone D.

6. Policy Considerations

Zone A is characterised by a relatively high quality, Q_N, of elective healthcare provided by the public sector to a relatively large proportion of the overall population, where Q_N reflects quality of healthcare factors that are sensitive to the availability of healthcare labour to service the demand for its elective healthcare treatments. Zones B and C are characterised by cumulative deteriorations in the quality of elective care that is provided by the public sector and an increasing proportion of the total population who opt for private elective healthcare, despite its higher direct financial cost to them. However, as noted above, points within Zones B and C are dynamically unstable, with the excess demand for healthcare labour increasing as healthcare wages rise, and positive feedback effects between the proportion of the population who opt for private healthcare and the cumulative deterioration on the quality of care that is available in the public sector at a given wage.

One main context in which significant policy considerations arise is the prevailing demographic trends, such as an ageing population and long-term effects of pandemics such as Covid-19, that may increase the population's need for elective healthcare, as reflected in the parameter m. From equations (2) – (8):

$$\partial X_H / \partial m = \beta_1(1 - N(y_0)) + \alpha_1 N(y_0) - (\beta_1 - \alpha_1) f(y_0)(\partial y_0 / \partial m) \quad (35)$$

$$(\partial y_0 / \partial m) = \gamma_1 \phi_1 (1 + \phi_2)(\partial \hat{S}_H^N / \partial m) = -\gamma_1 \phi_1 (1 + \phi_2) S_H / m^2 < 0 \quad (36)$$

so that $\partial X_H / \partial m > 0$. An increased value of m therefore increases the extent of the positive excess demand for healthcare demand whenever it arises, and reduces the extent of the excess supply of healthcare elsewhere.

Offsetting reductions in the value of m may to some extent be achieved by policy measures that are effective at improving population health through improved diets, lifestyle, reduced socio-economic disadvantage and preventive healthcare (see NHS England, 2024). The extent of the challenge to improve population health has been emphasised, for example, by Marmot (2017), Mayston (2000) and Zhou and Shen (2024).

Any net increase in the value of m will reduce the prevailing value of y_0 in equation (36) and Figure 1. It thereby increases the risk that y_0 will fall short of the critical *tipping point* level, y_{0b}, of the net income y_0 at which consumers are on the margin of switching between the public and private sectors, and which divides Zone A from the unstable Zones B and C.

The tipping point in Figure 1 is determined by the level of y_0 at which the value of the density function $f(y_0)$ equals Z_0. If $f(y_i)$ is here a normal density function for the distribution of individual net income, with mean μ and standard deviation of σ, we have:

$$y_{0b} = \mu + \sigma[2\ln(1/\sigma Z_0 \sqrt{2\pi})]^{0.5} \text{ with } \partial y_{0b} / \partial \mu > 0,\ \partial y_{0b} / \partial \sigma > 0 \text{ and } \partial y_{0b} / \partial Z_0 < 0 \quad (37)$$

under condition (14) that $\sigma Z_0 \sqrt{2\pi} < 1$. A similar result would indeed be implied if $f(y_i)$ were instead a log-normal distribution, but with μ then designating the expected value of

$\ln y_i$, σ the standard deviation of $\ln y_i$, $\ln y_{0b}$ in place of y_{0b} on the LHS of (37), and $\ln y_i$ in place of y_i on the horizontal axis of Figure 1, though now with the restriction that net income $y_i > 0$.

A key policy consideration raised by the recent 2024 UK General Election is whether achieving a higher level of economic growth will necessarily be consistent with easing the challenges facing the UK public sector healthcare system, the National Health Service (NHS) (see e.g. Labour Party, 2024). If such stronger economic growth results in a boost to consumer net income, and possibly also greater inequality in their distribution, equation (37) implies that the benefits for the NHS will not all be positive. Instead, a larger mean value of individual net income, and a greater inequality in their distribution, will increase the tipping point y_{0b} of values of y_0 below which the unstable Zone B comes into effect. An increase in the value of the tipping point y_{0b} will mean that that there will be values of y_0 for which previously $y_0 > y_{0b}$, and which therefore fell within the stable Zone A in Figure 1, but for which now $y_0 < y_{0b}$, and which therefore now fall into the unstable Zone B in Figure 1. Ironically, cost of living pressures that limit the growth in consumer net income will tend to lessen this risk.

However, as Baumol (1967) emphasised, the labour intensive nature of healthcare and similar public services makes them particularly exposed to long-run factors, such as the *relative price effect*, in which increasing real incomes with economic growth both increase the general level of wages and reduce the supply of labour to such public services at previous wage levels, thereby reducing the associated value of θ_0. From (4) - (8) and (11):

$$\partial y_0 / \partial \theta_0 = \gamma_1 (\partial Q_N / \partial \theta_0) = \gamma_1 \phi_1 (1+\phi_2)[1+(\beta_1-\alpha_1) m f(y_0)(\partial y_0 / \partial \theta_0)] \quad (38)$$

and hence $\partial y_0 / \partial \theta_0 = \gamma_1 \phi_1 (1+\phi_2)/(1-Z_1) <,> 0 \; as \; Z_1 >,< 1$ (39)

with $\partial X_H / \partial \theta_0 = -(\beta_1-\alpha_1) m f(y_0)(\partial y_0 / \partial \theta_0) - 1 = [1/(Z_1-1)] >,< 0 \; as \; Z_1 >,< 1$ (40)

Since a decrease in θ_0 shifts downwards the supply of healthcare labour curve for elective care in the public sector in (5) and (6), the direct effect of this is to lower the value of y_0 in (39) and increase the excess demand for healthcare labour in (40) in Zones A and D, where $Z_1 < 1$.

The extent of the decrease in θ_0 will be greater if there are additional *positive feedback effects* on θ_0, such as from the increased waiting lists that are associated with declines in the quality of care, Q_N, in the public sector resulting in a greater demand for emergency admissions in (2) and (6) further reducing the value of θ_0. Further positive feedback effects on θ_0 may arise if a lower quality of public sector healthcare results in a continuing decline in the health and productivity of the labour force (see Health Foundation, 2024a), and in the taxation revenue which it can contribute towards funding improved public sector healthcare. Such positive feedback effects nevertheless imply cumulative advantages if the direction of change can be reversed by adequate policy initiatives.

Positive increases in the value of θ_{00} and θ_0 in equations (5) and (6) may arise from planned increases in the domestically trained healthcare labour force and in the recruitment of healthcare staff from abroad. However, such increases need to be large enough to offset increases in the quantity of healthcare labour diverted to non-elective accident and emergency care, and to education and training activities with increased staff turnover, as reflected by increases in the parameter α_0 in equations (2) and (6). In addition, decreases in the value of θ_{00} and θ_0 may arise from demographic changes such as an ageing workforce that may result in more retirements than new recruits to the healthcare labour force, from pension scheme changes or in taxation that make the early retirement of healthcare staff more attractive, and from increased competition from higher wages available in other countries or in other occupations.

Many of the long-term demographic trends which are now impacting on the UK's public sector healthcare system have indeed been known for several decades (see for example Mayston, 1990). Given the substantial lags that are involved between recruiting more medical and nursing students and their availability to increase the healthcare labour force, it is therefore not evidence of good governmental policy that it was only recently in the year 2023 that a Long Term Workforce Plan was initiated for NHS England (2023). The Health Foundation (2024b) has estimated that achieving sustained reductions in NHS England's waiting lists and improvements in GP availability will require an additional £38 billion a year funding up to 2029/30 above earlier spending plans.

Significant other sources of positive increases in the value of θ_{00} and θ_0 may come from the more efficient use of the available healthcare workforce to increase its effective supply to improve patient throughput and reduce waiting lists, and therefore boost the quality of care variable Q_N. The finding by LaingBuisson (2024) that over £10 billion was spent by the NHS in 2022/23 on temporary agency and bank staff, typically incurring additional fees, may imply scope for reducing unit costs and expanding the permanent workforce. Greater capital investment in more rapid and more automated scanning and testing procedures may further reduce patient delays. So too would removing major bottlenecks, such as bed-blocking by patients who are well enough to be discharged but for which sufficient residential or community care is not currently available for their onward care. A governmental failure to effectively tackle the mounting demands on the social care sector to provide such care at a reasonable cost has, however, also persisted over many decades, despite the availability of well-informed analysis and proposals, such as in the Dilnot Report (2011).

Reducing excess demand in the social care sector requires significant increases in the supply of social care staff, who are currently paid relatively low wages. Again demographic factors, the relative price effect and the current need to recruit staff from abroad pose considerable challenges. The Department of Health and Social Care has estimated that by 2038 about 29 per cent more adults aged 18 to 64, and 57 per cent more adults aged 65 and over, will require social care than in 2018, with the total costs of social care projected to rise by 90 per cent for adults aged 18 to 64, from £9.6 billion to £18.1 billion, and 106 per cent for adults aged 65 and over from £18.3 billion to £37.7 billion between 2018 and 2038 (NAO, 2021).

However, the National Audit Office (2023) estimates that the total number of vacancies in adult social care in England increased by 173 per cent from 60,000 to 164,000 between 2012-13 and 2021-22, though in the year to 2022-23, it fell 7 per cent to 152,000 (a vacancy rate of 9.9 per cent), supported by the recruitment of 70,000 staff from overseas. Moreover, in 2021 one in seven individuals over 65 were expected to face lifetime care costs above £100,000 (NAO, 2023), with the Competition and Markets Authority (2017) finding that in 2017 self-funders paid on average 41 per cent more than local authorities for care home places.

Questions then arise as to how far more volunteer support staff might be able to improve the availability of social care, and the extent to which there are age groups, such as those between ages 50 and 70, that contain able-bodied individuals who might be incentivised by a system of care credits to currently provide more volunteer hours of social care to others, in return for a later reduction in their own care cost when they are themselves in greater need.

If, in contrast, a net reduction in θ_0, and therefore a downward shift in the supply of healthcare labour curve in (5) and (6), is associated with an increase in the wage level of $w' = (-dw/d\theta_0) > (1/\theta_1)$, the overall impact on y_0 from (11), (12) and (39) will be:

$$dy_0 / d\theta_0 = \{[\gamma_1\phi_1(1+\phi_2)(1-\theta_1 w') - \gamma_2\zeta_1 w'] / (1-Z_1)\} <,> 0 \ as \ Z_1 <,> 1 \qquad (41)$$

The resultant net increase in y_0 in Zone A will then tend to move the value of y_0 away from the tipping point value of y_{0b}, if $w' > (1/\theta_1)$. However, without such a wage increase or other counter measures, the direct effect of a downward shift in the supply of healthcare labour curve for elective care will be a lowering of the prevailing value of y_0 in (35), with $Z_1 < 1$ in Zone A. This will bring the prevailing value of y_0 in Zone A closer to the tipping point value of y_{0b} in Figure 1. Avoiding this tipping point may then require a substantial increase in the level of healthcare wages in the public sector in (41), with θ_1 low in value if the total supply of healthcare labour is relatively insensitive to wages. Unless there are significant sources of labour productivity growth and/or capital-labour substitution, this will in turn require substantial increases in public healthcare expenditure, with Hartwig (2008) confirming Baumol’s (1967) prediction that wage increases in excess of productivity growth are key determinants of long-term increases in healthcare expenditure.

Within Zones B and C, the direct effect of a downward shift in the supply of healthcare labour curve will be to increase the prevailing value of y_0 in (39). However, if it is accompanied by an increase in the wage level of $w' = (-dw/d\theta_0) > (1/\theta_1)$, the net result from (41) is to lower the prevailing value of y_0 in (39), boosting private sector demand and overall excess demand, and adding to the instability of Zones B and C in which wage rises increase the excess demand for healthcare, as in equations (12) and (21) and Table 2.

There are therefore good reasons to avoid falling into Zones B and C. In equation (37) and Figure 1, the tipping point y_{0b} at which the unstable Zone B begins is a decreasing function of Z_0. As in equation (13), Z_0 is itself a declining function of the parameters γ_1, ϕ_1 and ϕ_2, which reflect the importance of the quality of care variable Q_N in consumer decisions in equation (4), and of shortages of healthcare labour in reducing Q_N in equation (8). In (13), Z_0 is also a declining function of the gap, $(\beta_1 - \alpha_1)$, between the unit level of resourcing of elective healthcare treatments in the private and public sectors. Lower values to the parameters γ_1, ϕ_1 and ϕ_2, and $(\beta_1 - \alpha_1)$ will therefore imply a higher value to Z_0 and lower the value of the tipping point y_{0b} in Figure 1. If they are feasible, policy measures that reduce the values of the parameters γ_1, ϕ_1 and ϕ_2, and of $(\beta_1 - \alpha_1)$ will therefore tend to expand the values of y_0 that fall within the stable Zone A and reduce the extent of the unstable Zones B and C.

Investment in capital equipment to automate diagnostic and treatment procedures, and in innovations to reduce bed blockages, cancelled appointments and other impediments to an increased patient throughput, may, for example, mitigate the impact ϕ_1 of healthcare labour shortages on elective healthcare delays and other quality shortfalls in the public sector. Reductions in the overall demands on the available healthcare labour, such as the time taken to find available beds, poor access to IT support, and excessive administrative tasks, may also reduce workload stress, as reflected in the parameter ϕ_2 in equations (8) and (13).

Maintaining a high level of elective healthcare resourcing per patient in the public sector, as reflected in the parameter α_1, that is closer to that offered by the private sector will reduce the gap $(\beta_1 - \alpha_1)$ and further contribute to an increase in Z_0, a reduction in the value of the tipping point y_{0b}, and an increase in the extent of the stable Zone A. However, if β_1 is initially significantly greater than α_1, questions may arises as to how far such a variation in clinical practice for elective healthcare between the private and public sectors includes some degree of supplier-induced demand. A public sector healthcare provider facing strong budgetary pressures may have little incentive to unnecessarily use additional scarce resources if they do not significantly improve patient health. A private sector provider might have a greater financial incentive to increase their revenue through a greater use of their capital equipment and staff. The possible link between high levels of healthcare expenditure per capita in the United States private healthcare sector, disappointing health outcomes and supplier-induced demand was highlighted by Mulley et al (2009). This implies setting α_1 at a level which is consistent with NICE (2024) Guidance based on evidence of the clinical effectiveness of healthcare expenditures, and once achieved directing additional resources at improving patient outcomes by reducing the delays patients face to receive the recommended level of care.

7. The Prevalence of Excess Demand

The approach adopted by this paper differs from several earlier studies of the impact of delays in the provision of healthcare elective treatments in a number of important ways:

a. It explicitly recognises the central role played by the healthcare *labour market* and of healthcare *labour shortages*, and of associated variables such as healthcare *wage levels*, in influencing the extent of delays in the provision of healthcare elective treatments by the public sector.

b. In doing so, it explicitly recognises that private healthcare provision *competes with the public sector* within the healthcare labour market, and that there are *important feedback relationships* between healthcare labour shortages in the public sector, the quality of healthcare offered by the public sector and the demand for healthcare labour by the private sector.

c. The current paper does not assume that equilibrium will necessarily prevail between the available supply and the desired demand for elective healthcare, but instead makes the existence and behaviour of the *excess demand* for healthcare labour a key focus of its analysis. In doing so, it explicitly recognises that there are risks of instability in public sector healthcare provision, and in healthcare labour costs, once a key *tipping point* is passed in the proportion of the relevant population opting for private healthcare. Any equilibrium that does temporarily exist will then not necessarily be stable, with the excess demand for healthcare labour potentially escalating over time.

Several earlier studies (see Lindsay and Feigenbaum, 1984; Martin and Smith, 1999; Gravelle, Smith and Xavier, 2003; Siciliani, 2006) of hospital waiting time have assumed a utility maximising hospital manager whose performance is judged on patient waiting times, with Martin and Smith (1999) and Siciliani (2005) assuming that waiting lists do achieve an equilibrium through hospital managers increasing inpatient admissions to match demand increases. The implied estimates of the elasticity of demand for public sector healthcare with respect to waiting times, however, can themselves conceal important differences in the sources of the demand response, namely whether it is due to patients dying whilst waiting, patients or their GPs giving up trying to obtain treatment, with potentially adverse long-term health consequences, or to patients switching to private sector healthcare.

The detailed empirical study of waiting times for elective surgery by Martin et al (2003) did find “some evidence that better access to private healthcare provision may depress both the demand for NHS services and also NHS supply. These results must be viewed in the light of the rudimentary measures of private supply we had available, but they do suggest that interactions with private sector provision may be quite subtle and require careful examination before drawing policy conclusions”(ibid, p 170). They also note that while Propper (2000) found “no association between the length of either waiting lists under a year or over a year and the use of public and/or private in-patient health care... it should be noted that Propper’s waiting time variables are constructed at the regional level…Given that waiting times vary considerably across both DHAs [District Health Authorities] and wards within any given region, then the apparent insignificance of waiting times on the demand for private health care in Propper’s model might be due more to the way in which the waiting time variable has been constructed rather than the nature of the underlying relationship.” (ibid, p. 165).

In their study of the relationship between hospital activity, waiting times and population characteristics for the NHS using aggregate times-series data over the period 1952- 2005, Iacone et al (2007) assume that if policy makers observe an increase in waiting time in one period, they will be more willing to fund increases in supply in subsequent periods. As a result, "In the long-run equilibrium, waiting times do not vary over time...so that demand for and supply of treatment have reached equilibrium (ibid, p. 9)". Using a dynamic model, Smith and Van Ackere (2002) assume more generally that "the policy maker is often interested not only in the equilibrium predictions arising from an economic model, but also in the path taken by policy variables as they move towards that equilibrium".

The 82.3 per cent rise in NHS waiting lists for consultant-led elective healthcare in England since 2007 (BMA, 2024), however, suggests that such a stable long-run equilibrium does not necessarily prevail, with policy variables not guaranteed to be chosen to move towards a stable long-run equilibrium, even if one exists within Zone A. Within the decades since 1952, there have instead been different degrees of political willingness to fund increases in NHS healthcare, and associated wage pressures, to keep waiting times stable over time and to fully match the substantial increases in healthcare demand from an ageing population and from life-style related chronic illnesses.

The desired demand for healthcare labour in our above model is from those in clinical need for elective healthcare. For this desired demand to be equated to the available supply of healthcare labour in Zone A depends upon a willingness and ability of the public sector to fund healthcare wages at an equilibrium level, which itself will tend to rise with demographic trends in Zone A. Downward budgetary pressure on healthcare wages below an equilibrium level in Zone A then results in a reversal of the self-reinforcing effects of a wage rise in Table 2. Instead the total supply of healthcare labour now declines, the marginal income at which consumers opt for private healthcare falls, private sector healthcare absorbs more labour than it releases in the public sector, the available healthcare labour supply to the public sector reduces, and the perceived quality of care that is available in the public sector declines. While the resultant increase in the excess demand for healthcare labour might stimulate an offsetting rise in wages in a simpler market, sustained downward budgetary pressure on public sector healthcare wages may in contrast result in a steady drift away from an equilibrium and towards an unstable tipping point on the boundary of Zone B.

It is notable that the Secretary of State for Health and Social Care in the incoming Labour Government of 2024 has publicly declared that "The NHS is broken" following the above increases in NHS waiting lists under the previous Government (Secretary of State for Health and Social Care, 2024). While this has led to a 4.7 per cent increase in NHS annual revenue funding for the NHS in England in 2024-25 ahead of general inflation, in the view of the King's Fund (2024), "the health spending announced today is unlikely to be enough for patients to see a real improvement in the care they receive… because the £22 billion for two years allocated for day-to-day spending will also need to cover existing commitments for new staff pay deals and rising costs of delivering care".

There is then no guarantee that an equilibrium will necessarily prevail, even in the long-run, between the desired demand for healthcare and the available healthcare labour supply which can provide it. Attention therefore needs to be directed more closely at assessing the risks and implications of moving beyond a tipping point into unstable zones driven by a persistent disequilibrium between healthcare labour demand and supply, and associated sustained healthcare cost escalation. The desirability of avoiding future unstable healthcare cost escalation strengthens the case for greater current investment to safeguard the quality of care that is available in the public sector before reaching such a tipping point.

8. Conclusion

In our above model, the tipping point relates to the percentage of the relevant population of those in need of elective healthcare that opts for private provision, to an extent determined by underlying parameters. Even in a relatively simple model in which individual consumer income and the quality of public sector healthcare are the prime determinants of this decision, non-linearities in the distribution of income imply that rising numbers of consumers who are on the margin of opting for private healthcare can result in cumulative instability in the response of the excess demand for healthcare labour to rising wage costs. A basic stability condition that the excess demand for healthcare labour will decline when wages rise is not everywhere guaranteed.

Achieving improvements in the quality of public sector healthcare in the face of prevailing demographic trends is likely to involve significant further increases in funding and taxation revenue, if public sector healthcare is to remain free at the point of need. However, the alternative of lower taxation and further declines in public sector elective healthcare may be an unstable downward spiral in which the quality and availability of healthcare in the public sector deteriorate further. Those with substantial net income will receive superior access to healthcare in the private sector, albeit at a significantly higher user cost that may more than outweigh any financial gain that they receive through lower taxation.

The alternative of higher unit costs of healthcare in Zone D will itself pose a substantial challenge to public expenditure control while achieving lower value for money and poorer quality of care for many patients, compared to remaining in Zone A. The risk of cumulative deteriorations in public sector healthcare quality and availability is underlined by the recent decline in the state of NHS dentistry in England, which the Nuffield Trust (2023) has found “is at its most perilous point in its 75-year history”, with the British Dental Association (BDA) concluding that NHS dentistry is facing a “genuine crisis” of access, with many patients lacking access to an NHS dentist, or “forced to pay to see one privately if they can afford to do so” (House of Commons Health and Social Care Committee, 2023).

References

Arrow, K. and F. Hahn (1971), *General Competitive Analysis*, Oliver and Boyd, Edinburgh.

Balasko, Y. (1988), *Foundations of the Theory of General Equilibrium*, Academic Press, Boston.

Baumol, W. (1967), “Macroeconomics of Unbalanced Growth”, *American Economic Review*, vol. 57, pp. 415-426.

Besley, T., J. Hall and I. Preston (1996), P*rivate health insurance and the State of the NHS. Institute for Fiscal Studies.*

British Medical Association (2024), *NHS Backlog Data Analysis*, https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/pressures/nhs-backlog-data-analysis

Competition and Markets Authority (2017), *Care Homes Market Study: Final Report*, CMA.

Dilnot, A. (2011), *Fairer Care Funding: The Report of the Commission on Funding of Care and Support*, The Stationery Office, London.

Hartwig, J. (2008), "What Drives Health Care Expenditure? – Baumol's Theory of Unbalanced Growth Revisted", *Journal of Health Economics*, vol. 27 (3), pp. 603-623.

Health Foundation (2024a), *How Can the Next Government Improve the Health of the Workforce?* *https://www.health.org.uk/publications/long-reads/how-can-the-next-government-improve-the-health-of-the-workforce*

Health Foundation (2024b), *How Much Funding Does the NHS Need Over the Next Decade?* https://www.health.org.uk/publications/long-reads/how-much-funding-does-the-nhs-need-over-the-next-decade

House of Commons Health and Social Care Committee (2023), *NHS Dentistry*, HC 964.

Iacone, F., S. Martin, L. Siciliani, and P.C. Smith (2007), *Modelling the Dynamics of a Public Health Care System: Evidence from Time-Series Data.,*Centre for Health Economics Research Paper No. 29, University of York.
https://www.york.ac.uk/media/che/documents/papers/researchpapers/rp29_the_dynamics_of_a_public_health_care_system.pdf

Labour Party (2024), *Labour's Manifesto: Mission-Driven Government*, https://labour.org.uk/change/mission-driven-government/.

LaingBuisson (2024), *NHS Spending on Agency Staff Continues to Skyrocket*, https://www.laingbuisson.com/press-releases/nhs-spend-10bn-on-temp-staff/#:~:text=The%20NHS%20is%20by%20far,hospitals%20and%20on%20GP%20locums.

Lindsay, C. and B. Feigenbaum (1984), "Rationing by Waiting Lists", *American Economic Review*, vol. 74 (3), pp. 404-417.

Marmot, M. (2017), "The Health Gap: The Challenge of an Unequal World", *International Journal of Epidemiology*, 46(4), pp. 1312-1318.

Martin, S., R. Jacobs, N. Rice, and P. Smith (2003), *Waiting Times for Elective Surgery: A Hospital-Based Approach - Project Report, https://www.york.ac.uk/media/che/documents/reports/wait2.pdf*

Martin S. and P. Smith (2003), Rationing by Waiting Lists: an Empirical Investigation", Journal of Public Economics, vol. 71, pp. 141-164.

Mayston, D. J. (1990), "NHS Resourcing: A Financial and Economic Analysis", in *Competition in Healthcare*, eds A. Culyer, A. Maynard and J. Posnett, Macmillan Press, London, pp.67 – 109.

Mayston, D.J. (2000), "The Economic Determinants of Health Inequalities", *University of York Discussion Papers*, No. 2000/49, 43 pp. https://www.york.ac.uk/media/economics/documents/discussionpapers/2000/0049.pdf

Mulley, A. et al (2009), "Inconvenient Truths About Supplier Induced Demand and Unwarranted Variation in Medical Practice", *British Medical Journal*, 339 (7728), pp.1007-1009. https://www.bmj.com/content/339/bmj.b4073

National Audit Office (2021), *The Adult Social Care Market in England*, HC1244, https://www.nao.org.uk/wp-content/uploads/2021/03/The-adult-social-care-market-in-England.pdf

National Audit Office (2023), *Reforming Adult Social Care in England*, https://www.nao.org.uk/wp-content/uploads/2023/11/Report-reforming-adult-social-care-in-England.pdf

NHS England (2023), *NHS Long Term Workforce Plan*, https://www.england.nhs.uk/wp-content/uploads/2023/06/nhs-long-term-workforce-plan-v1.2.pdf

NHS England (2024), *NHS Prevention Programme*, https://www.england.nhs.uk/ourwork/prevention/

National Institute for Health and Clinical Excellence (2024), *NICE Guidance*, https://www.nice.org.uk/guidance

Nuffield Trust (2023), *Bold Action or Slow Decay? The State of NHS Dentistry and Future Policy Actions*, Policy Briefing, Nuffield Trust, London.

Quirk, J. and R. Saposnik (1968), *Introduction to General Equilibrium Theory and Welfare Economics*, McGraw-Hill, New York.

Propper, C. (2000), "The Demand for Private Health Care in the UK", *Journal of Health Economics,* vol. 19(6), pp, 855 – 876.

Secretary of State for Health and Social Care (2024), *The NHS is broken: Health and Social Care Secretary statement,* https://www.gov.uk/ government/speeches/statement-from-the-secretary-of-state-for-health-and-social-care .

Siciliani, L. (2006), "A Dynamic Model of Supply of Elective Surgery in the Presence of Waiting Times and Waiting Lists", *Journal of Health Economics*, vol. 25 (5), pp. 891-907, Smith, P. C and van Ackere, A. (2002), "A note on the integration of system dynamics and economic models", *Journal of Economic Dynamics and Control,* vol. 26, pp. 1-10.

Zhou M. and H. Shen (2024), "Forecasting the Global Burden of Disease to 2050", *Lancet*, 403 (10440), pp. 1961-1963.