School of Economics and Finance Discussion Papers

# The Drivers of the Nursing Workforce Gap: a Theoretical Framework

David Ulph

University of St Andrews

School of Economics and Finance Discussion Paper No. 2005
3 Jun 2020 (revised 11 Sep 2020)
JEL Classification: D11; I1; J20; J45
Keywords: Workforce Gap, nurses' labour supply; overwork; patient care; unstable equilibria

School of Economics and Finance
Online Discussion Paper Series
issn 2055-303X
http://ideas.repec.org/s/san/wpecon.html
info: econ@st-andrews.ac.uk

# The Drivers of the Nursing Workforce Gap: a Theoretical Framework

David Ulph[1]

## Abstract

This paper presents a new model of nursing labour supply that reconciles key features of the current state of nursing in the NHS: a large, persistent and potentially growing workforce gap; staff surveys that show that nurses gain great satisfaction from the job they do, but feel they don't have enough colleagues to enable them to do it well; a large fraction of nurses working paid and unpaid overtime; nurses leaving citing as significant factors both workload and an inability to deliver the quality of care that they would like to give. The paper examines both the intensive and extensive margins of nurse labour supply by extending the standard model of labour supply to include the intrinsic value nurses attach to providing care, which depends on the time a nurse can spend with each patient. This leads nurses to overworking in a well-defined sense. Choosing to work as a nurse involves balancing off: relative pay; the disutility from overworking as a nurse; the intrinsic benefit derived from nursing. For given levels of demand and a given relative wage for nursing, there may be multiple "equilibrium" levels of the workforce gap, some of which are unstable. However calibrating the model to recent data for the UK shows that the current workforce gap seems to be characterised as a unique stable equilibrium, and that increasing the pay of nurses is an effective means of reducing the gap.

JEL Classification: D11, I1, J20, J45

Key Words: Workforce gap; patient care; overwork; nurses' labour supply; unstable equilibrium

**September 2020**

---

[1] Professor of Economics, University of St Andrews and Member of the NHS Pay Review Body since 2015. This paper is written purely in my capacity as an academic and should not be taken to represent the views of the NHS Pay Review Body. I am grateful to David Bell, Margaret Leighton, Carol Propper, Alistair Ulph and Jonathan Wadsworth for comments and discussions on earlier versions of this paper, and to Ryan Trusler, undergraduate Economics and Mathematics student at St Andrews, for very able research assistance.

# Introduction

*Background*

The background to this paper is the following set of features of the current nurses' labour market in England:

(i) There is widespread agreement that there is a large, persistent and potentially growing workforce gap of around 41,000 nurses (1 in 8 of the workforce)[2].
(ii) Estimates suggest that around 80% - 90% of this gap is filled by a combination of temporary (agency) staff and staff overtime (both paid and unpaid)[3].
(iii) Staff surveys indicate that around 1/3 of staff work paid overtime – an increasing trend - while a fluctuating number averaging around 55% work unpaid overtime[4].
(iv) Staff surveys also show that less than half of staff feel that they are able to meet all the conflicting demands on their time, and just under 1/3 feel that there are enough staff to enable them to do their job properly[5].
(v) This raises concerns, not least amongst NHS Employers, that these workforce pressures created by vacancies may impact on retention[6].
(vi) Indeed evidence suggests that of those leaving the profession voluntarily for reasons other than retirement, 44% cite workload pressures and 27% cite disillusionment with the quality of care provided to patients[7].
(vii) Furthermore over recent years, work-life balance in particular has increased as a reported driver, with more than two-and-a-half times as many people citing it as a reason for leaving the NHS in 2018/19 than in 2011/12[8].
(viii) Nevertheless, despite all these pressures, nurses derive great satisfaction from the work they do, with just under 60% saying that they look forward to going to work and around 75% saying that they are enthusiastic about their job, and time passes quickly when they are working - figures which have been pretty steady or slowly rising over time[9].

*The Aim of the Paper*

A key idea that emerges from this snapshot is the possibility that vacancies create a vicious circle, whereby the workforce pressures they create lead to difficulties of retention and recruitment because staff feel over-worked and unable to deliver the quality of care to which they aspire, which leads to higher vacancies, further intensified pressure on staff who remain, and so on.[10] This suggests that, for given wages in nursing and alternative professions, there might be multiple equilibrium workforce gaps, some of which are unstable.

[2] See Beech et.al. (2019), p5; Hird et.al. (2019), p (vii), paras 3.64 and 4.72-4.78 and Table 4.5; Wollaston et.al. (2018), p6; Clayton-Hathway et. al. (2020), p12.
[3] See Hird at. al (2019) para 3.64; Wollaston et. al. (2018), p6; Clayton-Hathway et.al. (2020), p12.
[4] See Hird et. al. (2019) Table 4.12.
[5] Again see Hird et. al. (2019) Table 4.12.
[6] See Hird et. al. (2019) para 3.102.
[7] See Wollaston et.al. (2018), pp 12-13.
[8] See Buchan et. al. (2019) p.43.
[9] See Hird et.al. (2019), Table 4.11.
[10] See Hird et/al. (2020) paras 4.155 – 4.161 for an articulation of this idea. Recruitment difficulties may arise through student nurses doing placements either experiencing these pressures or hearing about them from experienced nurses and deciding not to join the NHS.

The first aim of this paper is to develop a theoretical framework which both captures many of the features set out above and allows this idea of multiple workforce gap equilibria to be rigorously tested.

The paper presents a new model of nurses' labour supply in which a population of individuals can choose to work as nurses or in some non-caring profession. Each individual has a utility function depending on consumption and leisure and so captures all the standard labour supply considerations in which individuals give up valuable leisure time to earn income that can be spent on consumption which they also value. However if an individual chooses to work as a nurse they derive an additional source of utility from the care they give their patients – feature (viii). This additional benefit depends on the quality of care – features (v) and (vi) - which in turn depends on the amount of time they can spend with each patient and so is negatively affected by the number of patients per member of staff, $p$. Individuals choose both how long to work in whatever job they do, and which job to do. The fact that nurses care about patient quality leads them to work longer hours than they would do if they cared only about consumption and leisure leading to a well-defined utility loss from excess workload – features (v), (vi) and (vii). The choice of occupation depends on balancing off three factors: pay differences; the disutility of workload pressures from nursing; the direct benefit derived from providing care to patients. This leads to both a "demand" and "supply" relationship between the number of patients per member of staff and the number of people choosing to work as nurses. Both are downward-sloping, which leads to the conclusion that for a given level of demand and for given wages in the two professions it is theoretically possible for there to be multiple "equilibrium" values for the number of individuals who choose to work as nurses (and associated numbers of patients per member of staff, $p$) and so potentially multiple workforce gaps, some of which might be unstable and so exhibit the vicious circles described above.

The second aim is to calibrate the model to recent UK data to see whether there are indeed such multiple equilibria.

*Links to the Literature*

There is of course an extensive literature on nurses' labour supply. Studies have looked at both the intensive (hours of work/participation ) margin – see for example Rice (2005) and, - and the extensive (job choice margin) – see for example Hamel, Kalb and Scott (2014), and Crawford, Disney and Emmerson (2015). For surveys see Antonazzo, Scott, Skatun and Elliott (2003), and Shields (2004). The evidence suggests that while the intensive own-wage elasticities can be positive or negative, in absolute terms they are low. However allowing for the extensive margin produces positive and larger elasticities – though this might vary regionally depending on the state of the local labour market[11].

While studies allow for the possibility that factors other than pay might affect nurse labour supply – see, for example, Shields and Ward (2001) – there has been little systematic investigation of the drivers of non-pay utility differences between nursing and non-nursing professions.

[11] Indeed Propper and van Reenen (2010) argue that having a uniform pay spine across all of England under the Agenda for Change pay framework, can make it very difficult to recruit sufficient high quality staff in certain areas, with consequent implications for patient care.

Evers et.al. (2004) provides a survey of labour supply studies covering all professions.

The papers that come closest to this one are those by Heyes (2005), (2007), and Taylor (2007). Heyes allows for the possibility that nurses might be motivated by both pay and a desire to deliver high quality care, and argues that an increase in nursing pay will disproportionately attract those motivated by pay , so resulting in a reduction in the average quality of nursing care. The study by Fedele (2018) comes to the opposite conclusion. While the model in this paper considers similar factors driving the decisions to be a nurse, it is assumed that all nurses are of the same quality.

*Structure of the Paper*

Section 1 sets out all the elements of the model. Section 2 derives nurses' labour supply looking in turn at the intensive and extensive margins. Section 3 derives the equilibrium levels of the workforce gap for given levels of demand and for given wages in nursing and non-nursing professions, and shows that there might be multiple equilibria, some of which are unstable. Section 4 calibrates the model to recent UK data and shows that this is consistent with there being a unique and stable workforce gap equilibrium. Section 5 concludes.

## Section 1: The Model

To focus on the essential issues discussed in the introduction, the model of nurse labour supply developed here is deliberately rather simple. In particular, the following assumptions are made:

i. Individuals have no unearned income and so have no option but to work in order to have any consumption. So there is no participation decision, and the only job-choice they face is whether to work as a nurse (caring job) or in a non-caring job. Consequently the only pay considerations relate purely to pay in nursing relative to that in a non-caring job.
ii. To capture the idea that a concern for patient care could lead to nurses working longer hours than they might otherwise wish to and hence work-life balance considerations, the number of hours of work undertaken in each type of job is endogenous, and freely chosen by the worker. However all the complexities relating to shift/overtime pay are ignored and it is assumed that in whatever job they do individuals face a constant gross hourly wage rate. So this paper does not capture the phenomenon of unpaid overtime.
iii. All the complexities of the income tax system are ignored and it is assumed that all income is taxed at a constant rate, so the net hourly wage rate earned in each job is also constant and independent of the number of hours worked.
iv. Finally all complexities relating to skill and grading are ignored and, in each sector, all individuals earn the same hourly wage rate – though, as noted in (i) this can vary across sectors..

Formally, the model has the following components.

There is a population of individuals that can work either as nurses or in some other non-caring job. The mass of this population is normalised to 1.

All individuals in this population have the same standard consumption/leisure utility function $u(c,\ell)$ where $c \geq 0, \ell, 0 \leq \ell \leq 1$ are individual consumption and leisure. This utility function satisfies all the standard properties of being strictly increasing and strictly concave. To guarantee that all individuals always do some work, but do not spend their entire time working, it also assumed that both consumption and leisure are necessities. Formally, the assumption is:

A1) $\forall c > 0; \ell, 0 < \ell \leq 1, \quad u_c(c,\ell) \to \infty$ as $c \to 0$ and $u_\ell(c,\ell) \to \infty$ as $\ell \to 0$.

If $\omega > 0$ generically denotes the net hourly wage rate – hereafter just the wage rate – and $h = 1 - \ell$ denotes hours of work, let $h(\omega) = \underset{h,\, 0 \leq h \leq 1}{\arg\max}\, u(\omega h, 1-h)$ denote the associated labour supply function, and $v(\omega) \equiv \underset{h,\, 0 \leq h \leq 1}{\max}\, u(\omega h, 1-h)$ the associated indirect utility function. It follows from the assumptions that consumption is a necessity and that there is no unearned income that individuals will always choose to work, so $h(\omega) > 0$. However the model is sufficiently general that the labour supply function can be increasing or decreasing in the wage rate. Given that labour supply is always positive, indirect utility is always strictly increasing in the wage rate. Formally, from Roy's Identity we have $v'(\omega) = u_c\left[\omega h(\omega), 1-h(\omega)\right].h(\omega) > 0$.

Individuals have a choice between working in a caring profession or in a non-caring profession. For the purposes of this paper the caring profession is nursing and the people for whom nurses care are patients. But the ideas developed here will generalise to many other contexts.

Let $w > 0$ denote the wage rate that an individual who works in the non-caring profession will be paid. It is assumed that an individual who works in the non-caring profession derives no intrinsic satisfaction from the work they do, so the only source of utility is that from individual consumption and leisure, and consequently the number of hours they work is given by $h(w)$ and the utility from working in the non-caring sector is $v(w)$.

An individual who works as a nurse gets paid a wage rate $W > 0$ and, in addition to the flow of utility from consumption and leisure, derives a direct utility benefit from providing patient care. This utility benefit may capture a number of factors – self-esteem, a concern for the well-being of others, a sense of doing something that is of intrinsic value and valued by society at large. Let $p > 0$ be the number of patients that a member of staff is expected to look after, so, if a nurse works $H$ hours, then $H/p$ is the amount of time the nurse can spend with each patient and serves as a measure of the quality of care that the nurse can provide. In general it would be reasonable to assume that the flow of benefit a nurse derives from the care they give their patients will be positively related to both $p$ and $H/p$, but, given the importance that nurses seem to attach to the quality of care they can provide, for simplicity in this paper it is assumed that the utility benefit depends solely on the quality of care. Consequently the flow of benefit that a nurse derives from caring is captured through the function $B\left(H/p\right)$ which is assumed to be strictly increasing and strictly concave.

Crucially, however, it is assumed that $B\left({}^{H}\!/\!_{p}\right)$ can be negative if the amount of care per patient is sufficiently low. This captures the idea of moral distress.

Since it is assumed that $u(c,\ell)$, $w$, and $W$ are the same for everybody it follows that if the function $B\left({}^{H}\!/\!_{p}\right)$ were also the same for everybody then either everybody or nobody would work as a nurse. So finally it is assumed that $B\left({}^{H}\!/\!_{p}\right)=\beta\left({}^{H}\!/\!_{p}\right)+\gamma$ where the strictly increasing and strictly concave function $\beta\left({}^{H}\!/\!_{p}\right)$ is the same for everyone but the shift parameter γ varies across the population[12]. It is assumed that the support of the distribution of γ in the population is the real line $\mathbb{R}$, and that the distribution function is $\Gamma(\gamma)$ with all the usual properties.

## Section 2: Nurse Labour Supply

2.1 *The Intensive Margin: Hours of Work*

Disregarding the constant shift parameter, γ, the number of hours of work chosen by a nurse as

$$H(W,p)=\underset{H,\,0\le H\le 1}{\arg\max}\; u(WH,1-H)+\beta\left({}^{H}\!/\!_{p}\right). \tag{1}$$

Because of the presence of the benefit derived from caring for patients that increases with the amount of care that a nurse can give each patient, it is clear that the number of hours worked will be higher than those that would have been worked had the nurse been concerned solely with individual consumption and leisure. Formally

$$H(W,p)>h(W). \tag{2}$$

which implies that

$$u\left[WH(W,p),1-H(W,p)\right]<v(W). \tag{3}$$

So, even though nurses choose to work these additional hours because they want to provide high quality care for their patients, they are aware that they are working longer than they strictly need to do in order to balance the needs of earning income and having time to spend for themselves and/or partners and children.

In what follows, $H(W,p)-h(W)$ will be identified as the amount of overwork that nurses do and $v(W)-u\left[WH(W,p),1-H(W,p)\right]$ the loss of utility that nurses suffer because of this overwork.

[12] This means that the critical quality of patient care at which the intrinsic value of being a nurse drops to zero will also vary across the population.

To determine the comparative statics of the supply of hours of work by nurses, notice that, given the assumption A1, the first order condition characterising the labour supply choice is:

$$Wu_c\left(W\hat{H},1-\hat{H}\right)-u_\ell\left(W\hat{H},1-\hat{H}\right)+\frac{\beta'\left(\hat{H}/p\right)}{p}=0. \qquad (4)$$

where, for ease of notation, $\widehat{H}=H(W,p)$.

If we start with the effect of the wage rate, the following results follow from the standard comparative static theory of optimisation.

**Lemma 1**

(i) $sign\frac{\partial H}{\partial W}=sign\frac{dh}{dW}$ ;

(ii) $\left|\frac{\partial H}{\partial W}\right|<\left|\frac{dh}{dW}\right|$.

**Proof**: If we let $D_1=-\left[W^2u_{cc}-2Wu_{c\ell}+u_{\ell\ell}\right]>0$, $D_2=D_1-\frac{1}{p^2}\beta''>D_1$ and $N=u_c+H\left(Wu_{cc}-u_{c\ell}\right)$, then standard comparative static theory implies:

$\frac{dh}{dW}=\frac{N}{D_1}$; $\frac{\partial H}{\partial W}=\frac{N}{D_2}$, which in turn implies (i) $sign\frac{\partial H}{\partial W}=sign\frac{dh}{dW}=signN$; and

(ii) $\left|\frac{\partial H}{\partial W}\right|=\frac{|N|}{D_2}<\frac{|N|}{D_1}=\left|\frac{dh}{dW}\right|$.QED

So working as a nurse produces an hours of work labour supply response to increased wages that is of the same sign as that which would arise if working at the same wage in the non-caring sector but of a smaller absolute magnitude.

Turning to the effect of an increase in the number of patients per nurse, notice that an increase in *p* has two effects on the third term on the LHS of (4): (i) it increases the denominator and so reduces the value of the term; (ii) it reduces the amount of care each patient receives, $\hat{H}/p$, and so, because of diminishing marginal benefit, increases the numerator in this term. Provided marginal benefit falls sufficiently fast this second effect will dominate the first and an increase in *p* will cause an individual to work harder. Formally, we have:

**Lemma 2**

$$\frac{\partial H}{\partial p}\gtrless 0 \text{ as } -\frac{\left(\hat{H}/p\right)B''\left(\hat{H}/p\right)}{B'\left(\hat{H}/p\right)}\gtrless 1 . \qquad (5)$$

**Proof**: From the standard comparative static theory of optimisation we have

$\frac{\partial H}{\partial p} = \frac{\tilde{N}}{D_2}$ , where $\tilde{N} = -\frac{1}{p^2}\left[\beta'\left(\hat{H}/p\right) + \left(\hat{H}/p\right).\beta''\left(\hat{H}/p\right)\right]$ , from which the result follows immediately. QED

Associated with the maximisation problem in (1) is the indirect utility function

$$V(W,p) = \max_{H,\,0\le H\le 1} u(WH, 1-H) + \beta\left(H/p\right) = u\left(W\hat{H}, 1-\hat{H}\right) + \beta\left(\hat{H}/p\right). \qquad (6)$$

The comparative statics of this function are given by

**Lemma 3**

(i) $\frac{\partial V}{\partial W} = u_c\left(W\hat{H}, 1-\hat{H}\right).\hat{H} > 0$**;**

(ii) $\frac{\partial V}{\partial p} = -\frac{\hat{H}}{p^2}.\beta'\left(\hat{H}/p\right) < 0$ .

**Proof:** Follows immediately from the Envelope Theorem. QED

Finally, taking account of the constant shift parameter that varies across the population, the utility that a typical individual derives from working as a nurse is:

$$V(W,p) + \gamma = u\left(W\hat{H}, 1-\hat{H}\right) + \beta\left(\hat{H}/p\right) + \gamma = u\left(W\hat{H}, 1-\hat{H}\right) + B\left(\hat{H}/p\right) . \qquad (7)$$

2.2 *The Extensive Margin: job-choice*

An individual will choose to work as a nurse if and only if

$$V(W,p) + \gamma \ge v(w) . \qquad (8)$$

From (7) it therefore follows that an individual will work as a nurse if and only if

$$\left[v(W) - v(w)\right] - \left\{v(W) - u\left[W\hat{H}, 1-\hat{H}\right]\right\} + B\left(\hat{H}/p\right) \ge 0 . \qquad (9)$$

Taking the three components of the expression on the LHS of the inequality in (9) in turn, reveals that decision to work as a nurse involves balancing off three considerations:

(i) the pay of a nurse relative to that in the non-caring profession;
(ii) the disutility from the high workload in nursing;
(iii) the satisfaction derived from working as a nurse, which depends on the amount of care that a nurse can give each patient.

Notice that this implies that an individual may choose to work in the non-caring profession even if they would have derived positive satisfaction from being a nurse - i.e. even if $B\left(\hat{H}/p\right) > 0$ - provided this is not sufficient to compensate for any negative net loss of utility from the other two components. Conversely, an individual who is so concerned about the quality of care that they are able to provide their patients that $B\left(\hat{H}/p\right) < 0$ would only ever choose to work as a nurse if there was a sufficiently large wage premium.

Finally from (8) we see that the fraction of the population who are willing to work as nurses – the supply of nurses - is

$$F\left(W, w, p\right) = 1 - \Gamma\left[v(w) - V(W, p)\right] \; . \qquad (10)$$

Assuming that, for a given value of $v(w) - V(W, p)$ the density of the distribution of $\gamma$ is positive, it follows from Lemma 2 that this supply function has the following comparative static properties:

**Lemma 4** (i) $\frac{\partial F}{\partial W} > 0$; (ii) $\frac{\partial F}{\partial w} < 0$; (iii) $\frac{\partial F}{\partial p} < 0$ .

**Proof**: Follows in an obvious way from previous results, especially Lemma 2. QED

So the supply of nurses is a strictly increasing function of the wage rate in nursing, but a strictly decreasing function of the wage rate in the non-caring profession and in the number of patients that each nurse has to look after[13].

## Section 3. Demand, Supply and the Workforce Gap

Recalling that the mass of the population that might potentially work as nurses has been normalised to 1, assume that there is a given demand, $D_0$, for nursing hours. Assume also that there is a standard number of contracted hours, $h_0 \le h(W) < H(W, p)$.

Now the way the NHS vacancy/workforce gap is typically calculated is as the difference between the number of nurses needed if nurses work the contracted number of hours and the number of nurses actually working in the NHS. So the workforce gap is:

$$G = D_0/h_0 - F\left(W, w, p\right). \qquad (11)$$

But since nurses are actually working $H(W, p)$ hours, the amount of the demand for nursing hours that is not being met is $D_0 - H(W, p)F(W, w, p)$ while the amount that is being met by overtime is $\left(H(W, p) - h_0\right)F(W, w, p)$, so we can re-write the workforce gap as:

$$G = \left[\frac{D_0 - H(W, p)F(W, w, p)}{h_0}\right] + \left[\frac{H(W, p)}{h_0} - 1\right]F\left(W, w, p\right), \qquad (12)$$

and in this way the gap can be decomposed into an amount that represents work not being done – the first term - and an amount that is being filled by overtime – the second term.

[13] This latter effect has not so far been systematically investigated in the existing empirical literature.

Consistent with the definition in (11), for any given fraction of the population that works as nurses, $F, \quad 0 \le F \le 1$, we can define the number of patients that each nurse is expected to treat as

$$p = \frac{p_0}{F}, \qquad (13)$$

where $p_0 = D_0 / h_0$. So $p_0$ is the patient/staff ratio that would prevail if the entire population chose to work as nurse, and serves as a measure of the demand pressures facing the service. We can think of (13) as defining a "demand" curve relating $p$ and $F$, and is illustrated in Figure 1[14]. By definition this curve takes the form of a rectangular hyperbola.

On the other hand, for given wages in nursing and in the non-caring profession, equation (10) defines a potentially downward-sloping supply relationship between $p$ and $F$. At this level of generality it is hard to be definitive about what precise shape it takes. However, suppose that there exists some critical value of $p, \underline{p} > 0$ such that for all $p \le \underline{p}$ everyone in the population chooses to work as a nurse – so $F(p) \equiv 1$ - while for $p > \underline{p}$ some choose not to work as a nurse, though there is always some who do. So for all $p > \underline{p}, \quad 0 < F(p) < 1$ and, from Lemma 4 (iii), is strictly decreasing in *p*. Beyond that the shape of the supply curve is not well determined at this level of generality. Such a supply curve is illustrated in Figure 2.

To see what possible outcomes could arise by bringing together these two concepts, consider first of all Figure 3. Here there is a unique combination of *p* and *F* - that represented by point A, and designated $(p_A, F_A)$ - that is consistent with both (10) and (13) in the sense that, (a) given the fraction of the population working as nurses, $F_A$, the patient/staff ratio determined by the demand through (13) is $p_A$, and (b) given this patient/staff ratio the fraction of the population willing to work as nurses, as given by the supply function (10) is $F_A$. So in that sense $(p_A, F_A)$ is an equilibrium.

To see whether it is stable, consider the following dynamic process. Start with some initial value of *F*, call it $F_0, 0 < F_0 \le 1$. Use (13) to determine the corresponding patient/staff ratio $p_0$. Given $p_0$ use the supply function (10) to determine the fraction of the population willing to work as nurses, $F_1, 0 < F_1 \le 1$, and so on. The equilibrium $(p_A, F_A)$ is locally stable if for all $F_0, 0 < F_0 \le 1$ in a neighbourhood of $F_A$ this process converges to $F_A$ - and the corresponding values of *p* as defined by (13) converge to $p_A$. It is globally stable if the convergence happens for any $F_0, 0 < F_0 \le 1$.

Given this, it is easy to see that the unique equilibrium A in Figure 3 is stable.

Now consider Figure 4. Here there are two equilibria given by points A and B. Equilibrium B is locally stable in the sense that for any value of $F_0, F_A < F_0 < 1$ the dynamic process considered above will converge to B. Here there is a virtuous circle at work: given the

[14] All figures are contained in the Appendix

relatively high number of people who chose to work as nurses (relatively low workforce gap) patient loads are relatively light, causing more people to want to work as nurses, further reducing the workload and so on.

This argument also demonstrates that equilibrium A is locally unstable, for if only a slightly higher fraction of the population than $F_A$ people are willing to work as nurses then the dynamic process will lead away from A and result in everyone working as nurses.

The instability of A is also clear if we now suppose that a lower fraction of the population than $F_A$ wanted to work as nurses, so $F_0, 0 < F_0 < F_A$. Here there will be a vicious circle at work whereby too few people being willing to work as nurses – too high a workforce gap - creates high patient/staff loads causing fewer people to want to work as nurses because of poor worklife-balance and dissatisfaction with the quality of care they can deliver, thus driving up the workload on those that remain and so on.

So a clear implication of this analysis is that high workforce gaps may be more problematic than low workforce gaps.

Given the generality of the underlying functions and the consequent difficulty of pinning down the exact shape of the supply function - other than being downward sloping - there are many other configurations of supply and demand that can in principle arise. The fact that both the demand and supply curve are downward-sloping in *p* suggests that the problem of multiple equilibria some of which are unstable is potentially pervasive.

To get a better understanding of how important this problem might be in practice, in the next section I adopt some explicit, but still quite general, functional forms and calibrate the model to recent UK data, and, in particular, some of the salient background facts set out at the start of the paper.

## Section 4 Calibrating the Model

The nursing labour market is described by two variables, $F$ and $p$. The aim of this section is to calibrate the model so that the actual position in the NHS $\left(\tilde{F}, \tilde{p}\right)$ is an equilibrium of the model, i.e. lies on both the supply and demand curve. The question of interest is whether this equilibrium is (i) unique, (ii) stable.

I start by trying to pin down the values of $\left(\tilde{F}, \tilde{p}\right)$. Starting with $\tilde{F}$ Crawford et al, they say that "At any one time on average just under 70% of potential NHS nurses work as NHS nurses, with just over 10% working as nurses outside the NHS. The remainder are found working in other occupations." In the model there are just two occupations – working as a nurse in NHS and working in non-nursing job. So to try to get at the "right" value for $\tilde{F}$ using our model one could use $\tilde{F} = 0.7$ or $0.8$, or alternatively the intermediate value $\tilde{F} = 0.75$. In what follows I set $\tilde{F} = 0.75$.

Now use the fact that the workforce gap is 1/8 of the workforce – i.e.

$$\tilde{G} = \frac{\tilde{F}}{8}. \qquad (14)$$

Plugging this into equation (11) we get

$$p_0 = \left( D_0 \big/ h_0 \right) = \frac{9\tilde{F}}{8} \approx 0.84. \qquad (15)$$

But then, since we want the actual position to be an equilibrium, and so lie on the demand curve, it follows from (13) that

$$\tilde{p} = \frac{p_0}{\tilde{F}} = 1.125 \ . \qquad (16)$$

Now use the fact that 80% of the workforce gap is filled by overtime. Using the decomposition of the workforce gap given in equation (12) in the paper this implies

$$\left[ \frac{\tilde{H}}{h_0} - 1 \right] \tilde{F} = 0.8\tilde{G} \qquad (17)$$

But then from (14) we have:

$$\tilde{H} = 1.1 h_0 \qquad (18)$$

So, on average, the amount of overtime nurses do is around 10% of their contractual hours.

Since the contracted hours are 37.5, so, from (18), this would imply that nurses work 41.25 hours a week, doing, on average, just over 4 hours of overtime. Figures from the RCN say that nurses work on average 45 hours per week, so this is a reasonable approximation.

Data from Statista 2000[15] shows that in the UK in 2019 women on average work around 34.5 hours – which we can take to be the figure for those in the non-caring sector This figure varies from year to year but, since 2000 has fluctuated between just under 34 to just under 35. Given the above calculation that nurses work 41.25 hours per week, this implies that nurses work around 20% more than those in the non-caring sector. So it is assumed that

$$\tilde{H} = 1.2\tilde{h} \ . \qquad (19)$$

Now in the model hours of work are measured as a fraction of some total amount of time available for work or leisure, and so $H$ and $h$ are pure numbers lying between 0 and 1.

Although there are 168 hours in the week, one has to allow for time needed for sleeping, eating, essential household chores etc over which individuals exercise no choice. So it is plausible that in fact the realistic amount of time available for work or leisure is something like 100 hours, so in what follows take it that

$$\tilde{h} = 0.35; \quad \tilde{H} = 0.42. \qquad (20)$$

[15] www.statista.com

Assume now that individual utility takes the form of a CES utility function. In particular, if an individual works in a non-caring profession then utility from consumption and leisure is:

$$u(c,\ell)=\begin{cases}\dfrac{\alpha c^{1-\varepsilon}}{1-\varepsilon}+\dfrac{(1-\alpha)\ell^{1-\varepsilon}}{1-\varepsilon}, & 0<\alpha<1;\quad \varepsilon>0, \varepsilon\neq 1\\ \alpha\ln(c)+(1-\alpha)\ln(\ell), & 0<\alpha<1;\quad \varepsilon=1\end{cases} \tag{21}$$

where, (a) $\sigma=\dfrac{1}{\varepsilon}$ is the elasticity of substitution; (b) the case where $\varepsilon=1$ corresponds to the Cobb-Douglas utility function. This generates the intrinsic labour supply function in the non-caring profession:

$$h(w)=\frac{\eta w^{\sigma-1}}{1+\eta w^{\sigma-1}},\quad \text{where } \eta=\left(\frac{\alpha}{1-\alpha}\right)^{\sigma}. \tag{22}$$

By differentiating this w.r.t the wage rate it is straightforward to see that the expression for the own-wage elasticity is

$$\frac{dh}{dw}.\frac{w}{h}=(\sigma-1)(1-h)\Rightarrow \sigma=1+\frac{\frac{dh}{dw}.\frac{w}{h}}{(1-h)}. \tag{23}$$

Evidence from the studies cited in the introduction suggested that, for non-nursing jobs, the hours of work own-wage elasticity was somewhere around 0.4 – 0.48. So if $h=0.35$ this implies that σ lies somewhere in the range 1.62 – 1.74. So let us set the value at

$$\sigma=\frac{5}{3}\Rightarrow \varepsilon=0.6. \tag{24}$$

Turn now to the hours of work supplied by nurses. Consistent with the CES utility function, assume that $\beta\left(H/p\right)=\dfrac{\delta\left(H/p\right)^{1-\varepsilon}}{1-\varepsilon}$, then it is straightforward to show that the hours of work labour supply function is:

$$H(W,p)=\frac{\eta\left(W^{\frac{\sigma-1}{\sigma}}+\kappa p^{\frac{1-\sigma}{\sigma}}\right)^{\sigma}}{1+\eta\left(W^{\frac{\sigma-1}{\sigma}}+\kappa p^{\frac{1-\sigma}{\sigma}}\right)^{\sigma}},\quad \text{where } \kappa=\frac{\delta}{\alpha}. \tag{25}$$

Differentiate w.r.t. the nurse wage rate, *W*, and the formula for the own-wage elasticity is:

$$\frac{\partial H}{\partial W}.\frac{W}{H}=(\sigma-1)(1-H).\frac{W^{\frac{\sigma-1}{\sigma}}}{W^{\frac{\sigma-1}{\sigma}}+\kappa p^{\frac{1-\sigma}{\sigma}}}<(\sigma-1)(1-H), \tag{26}$$

and, given that nurses work longer hours than those in the non-caring profession, this shows that the hours of work own-wage elasticity of nurses has to be lower than that in the non-caring profession. Indeed, empirical studies cited in the introduction suggest that the hours of work own-wage elasticity for nurses is 0.2, so plugging that plus the values of σ and $H$ given by (24) and (20) into (26) we get:

$$\frac{W^{\frac{\sigma-1}{\sigma}}}{W^{\frac{\sigma-1}{\sigma}}+\kappa p^{\frac{1-\sigma}{\sigma}}}=\frac{0.2}{\frac{2}{3}.0.58}=0.52\,. \qquad (27)$$

This implies

$$\kappa=\frac{0.48}{0.52}\left(pW\right)^{\frac{\sigma-1}{\sigma}}\ . \qquad (28)$$

We have a degree of freedom to choose the units of money in which we measure consumption etc. Use this to normalise the nurses wage rate to

$$\tilde{W}=1\ . \qquad (29)$$

In other words, 1 unit of consumption is equated with the amount that a nurse would earn in a week if she took no leisure and spent the total amount of time available (which we take to be 100 hours) in work. Given this normalisation, the fact that, from (16), $\tilde{p}=1.125$, and the fact that, from (20) $\sigma=\frac{5}{3}$, this implies that

$$\kappa=\frac{0.48}{0.52}(1.125)^{0.4}\approx 0.97\,. \qquad (30)$$

Now we can re-arrange the expressions for hours of work labour supply

$$\eta\left(W^{\frac{\sigma-1}{\sigma}}+\kappa p^{\frac{1-\sigma}{\sigma}}\right)^{\sigma}=\frac{H}{1-H}\ \text{ and }\ \eta w^{\sigma-1}=\frac{h}{1-h}\ ^{16}. \qquad (31)$$

Take the ratio, and use (27) to get:

$$\left(\frac{w}{W}\right)^{\sigma-1}=\frac{1}{(0.52)^{\sigma}}\frac{h}{1-h}\frac{1-H}{H}\ . \qquad (32)$$

Plugging in the values for $\sigma, \tilde{h}, \tilde{H}$ and $\tilde{W}$ we get:

$$\tilde{w}\approx 3.29\,. \qquad (33)$$

However this implies that $\tilde{w}\tilde{h}\approx 2.74\tilde{W}\tilde{H}$ , and so gross weekly pay in the non-caring profession is just under 2.75 times that in nursing.

Now substitute (27) into the first equation in (31) and we get

[16] Essentially these are just re-statements of the first-order condition that MRS = relative price.

$$\left(\frac{\alpha}{1-\alpha}\right)^{\sigma}=\eta=\frac{H}{1-H}\frac{(0.52)^{\sigma}}{W^{\sigma-1}}. \tag{34}$$

Plugging in the values for $\sigma, \tilde{h}, \tilde{H}$ and $\tilde{W}$ and we get:

$$\frac{\alpha}{1-\alpha}\approx 0.43 \Rightarrow \alpha \approx 0.30. \tag{36}$$

But then, since $\kappa=\frac{\delta}{\alpha}$ , it follows from (30) and (34) that

$$\delta=\kappa\alpha=0.29. \tag{37}$$

So we now have the equilibrium values of $\tilde{F}, \tilde{p}, \tilde{w}, \tilde{W}, \tilde{h}, \tilde{H}$ , and associated values of the crucial parameter, $p_0$ of the demand curve and the 3 parameters $\varepsilon$ (equivalently $\sigma$) , $\alpha$ and $\delta$ of the utility function.

From this we can calculate $V\left(\tilde{W},\tilde{p}\right)$ and $v\left(\tilde{w}\right)$.

To complete the analysis it is necessary to pick a distribution function for γ. Since in principle $v\left(\tilde{w}\right)-V\left(\tilde{W},\tilde{p}\right)$ can be positive or negative, assume that γ is normally distributed, with mean μ and standard deviation, φ[17]. That is:

$$\Gamma(\gamma)=\mathrm{N}\left(\gamma;\mu,\varphi^2\right). \tag{38}$$

Since we have no reason to think the mean of γ is anything other than 0, set $\mu=0$ . The final parameter we need to set is the standard deviation. To do this pick φ so that

$$\tilde{F}=1-\mathrm{N}\left[v\left(\tilde{w}\right)-V\left(\tilde{W},\tilde{p}\right);0,\varphi^2\right], \tag{39}$$

which guarantees that the demand curve cuts the supply curve at our equilibrium values $\left(\tilde{F},\tilde{p}\right)$. It turns out that the value of the standard deviation that does this is

$$\varphi=0.24 \ . \tag{40}$$

Finally in order to check whether this equilibrium is unique it is necessary to plot the demand and supply curves. To do this requires keeping constant the parameters $p_0, \varepsilon\ \alpha,\ \delta$ and $\varphi$ and the wage rates $\tilde{w}$ and $\tilde{W}$, letting $p$ vary across a suitable range of positive values and plotting out the associated demand and supply curves:

$$F=\frac{p_0}{p}; \quad F=1-\mathrm{N}\left[v\left(\tilde{w}\right)-V\left(\tilde{W},p\right);0,\phi\right]. \tag{41}$$

[17] To check that nothing much depends on this particular choice of distribution function I repeated the analysis using a uniform distribution with zero mean, but this gives no detectable difference to the shape of the supply function. I also used different scalings of the utility function, again with no discernible difference.

The resulting demand and supply curves are shown in Figure 5, and as can be seen the current situation in the UK corresponds to a unique stable equilibrium.

As mentioned above I checked the robustness of this conclusion by assuming a uniform distribution of the parameter $\gamma$ which has a very different pattern of weights in the tail and this made no discernible difference to the supply curve. What is driving the shape of the curve is just the way in which the utility function $V(W, p)$ varies with *p*. Of course that in turn depends o how we choose to measure utility, but I checked for this by using different transformations of utility, and again the supply function showed no discernible difference.

## Section 5: Final Thoughts and Conclusions

In this paper I have shown that the concern about possible vicious circles driving up the nursing workforce gap can be given some rigorous foundations in economic theory. Specifically I have built a very simple but general labour of the intrinsic and extrinsic labour supply of nurses that captures many of the observed features of nurses' attitudes and behaviour and shown that, holding the wage rates in nursing and in an alternative non-caring profession constant there could theoretically be multiple equilibrium workforce gaps, some of which are unstable. Moreover if the actual workforce gap is above such an unstable equilibrium, and so, is "too high" then the natural dynamics would be those captured by the narrative of the vicious circle and would drive the gap up.

However there are clearly some problems in expecting this to account for the observed position in the NHS. First it is extremely unlikely that the existing position would be such an unstable equilibrium, because, by definition, small perturbations would drive the situation away from such an equilibrium. Secondly, even if the actual situation were close to an unstable equilibrium then there is just as much reason to think that it is on the virtuous circle side of the equilibrium as on the vicious circle side.

So it is important to understand what the supply and demand curves in the UK look like and where the existing situation lies. In relation to these curves. To this end I have calibrated the model so that the existing situation is an equilibrium and the aim was to learn whether this is unique (and hence stable) or if there were multiple such equilibria and if the current situation corresponds to a stable or unstable equilibrium.

As shown in Section 4 the conclusion that emerges is that the current situation in UK seems to correspond to a unique stable equilibrium. As indicated in the previous section this conclusion seems robust to some significant perturbations. Nevertheless it is important to stress that the model is very simple, and ignores all the complexity that arises from a very heterogeneous workforce serving a geographically dispersed and equally heterogeneous patient population. While I have used a relatively generally class of utility functions – CES – this another source of simplification. The high (relative) wage rate and earnings in the non-caring profession that emerges in equation (33) indicates that this model is failing to capture some important elements of this complex reality.

With these caveats, the message that emerges is that of the *persistence* of the workforce gap. Unless there is some significant change to the exogenous features of the model the gap will remain. The obvious policy that one would think of using to reduce the gap is that of an increase in the relative pay of nurses. So I undertook the comparative static exercise of increasing the pay of nurses by 5%. This shifted the equilibrium values of $(F, p)$ from (0.75,1.125) to (0.78,1.077). With the demand parameter $p^0 = 0.84$ remaining constant this implies that the workforce gap drops from 0.09 to 0.06 - a drop of 33%. Expressed as a fraction of the workforce, the gap falls from 1/8 to 1/13.

This conclusion is somewhat sensitive to the underlying assumptions. So, if the distribution of $\gamma$ is uniform rather then the reduction in the gap is around 22%. Nevertheless it is worth noting that the recent 3 year pay deal increased pay by around 6.5% (ignoring pay progression) and that it was recently announced that between 2019 and 2020 the number of nurses in NHS had increased by 12,000 – equivalent to a 30% drop in a gap of 40,000.

## References


Antonazzo, E., Scott, A., Skatun, D. and Elliott, R. F. (2003), "The Labour Market for Nursing: A Review of the Labour Supply Literature", *Health Economics*, 12, 465–478.

Beech J., et al. (2019), "Closing the gap: Key areas for action on the health and care workforce" *Joint Report of The Health Foundation, The King's Fund and Nuffield Trust,*

Buchan, J., et. al. (2019), "Falling Short: the NHS Workforce Challenge", *The Health Foundation Report.*

Clayton-Hathway, K. et.al. (2020), "Gender and Nursing as a Profession", *Royal College of Nursing Report*.

Crawford, R., Disney, R. and Emmerson, C. (2015), "The short-run elasticity of NHS nurses' labour supply in Great Britain." IFS Working Paper W15/04.

Evers, M., De Mooij, R. and Van Vuuren, D. (2008). "The Wage Elasticity of Labour Supply: a Synthesis of Empirical Estimates", *De Economist*, **156**, 25-43

Fedele, A. (2016), "Well-paid nurses are good nurses". *Health Economics*, **27**, 663-674.

Hamel, B., Kalb, G. and Scott, A. (2014), "Nurses Labour Supply Elasticities: The Importance of Accounting for Extensive Margins." *Journal of Health Economics*, **33**, 94-112.

Heyes, A. (2005), "Economics of Vocation or "Why is a Badly Paid Nurse a Good Nurse?", *Journal of Health Economics*, **24**, 561-9.

Heyes, A. (2007), "Optimal wages in the market for nurses", *Journal of Health Economics,* **26**, 1031-1033.

Hird et. al. (2019), "Thirty-Second Report*", NHS Pay Review Body*, CCS0419071156 07/19

Hird et. al. (2020), "Thirty-Third Report*", NHS Pay Review Body*, CCSO320353028 07/20

Propper, C. and Van Reenen, J. (2010) 'Can pay regulation kill? Panel data evidence on the effect of labor markets on hospital performance', *Journal of Political Economy,* 118, 2, pp. 222-273.

Rice, N. (2005), "The labour supply of nurses in the UK: evidence from the British Households Panel Survey", *University of York HEDG* Working Paper 05/10.

Shields, M. (2004), "Addressing nurse shortages: what can policy makers learn from the econometric evidence on nurse labour supply?", *Economic Journal*, **114,** F464-F498.

Shields, M. and Ward, M. (2001), "Improving nurse retention in the National Health Service in England: the impact of job satisfaction on intentions to quit". *Journal of Health Economics*, **20**, 677-701.

Taylor, L. (2007), "Optimal wages in the market for nurses: An analysis based on Heyes' model", *Journal of Health Economics*, **26**, 1027-1030.

Wollaston, S., et. al. (2018), “The Nursing Workforce”, *Health Committee, House of Commons Report*, HC353.

Appendix: Diagrams

Figure 1: Demand Curve

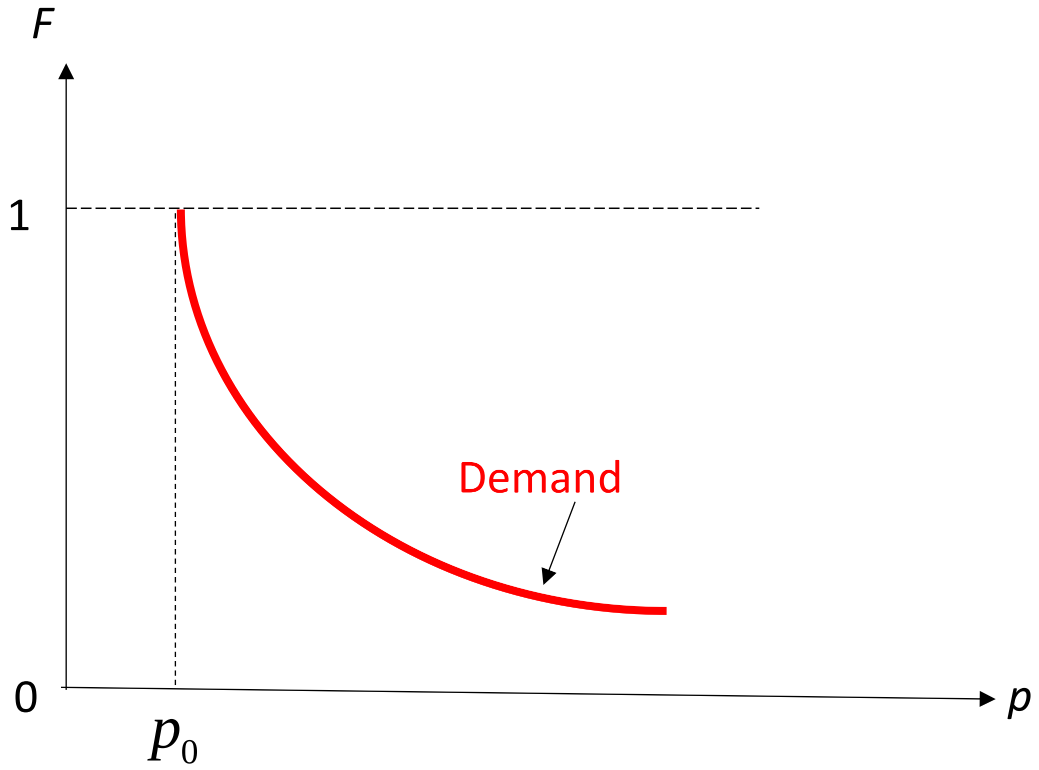

Figure 2:  Supply Curve

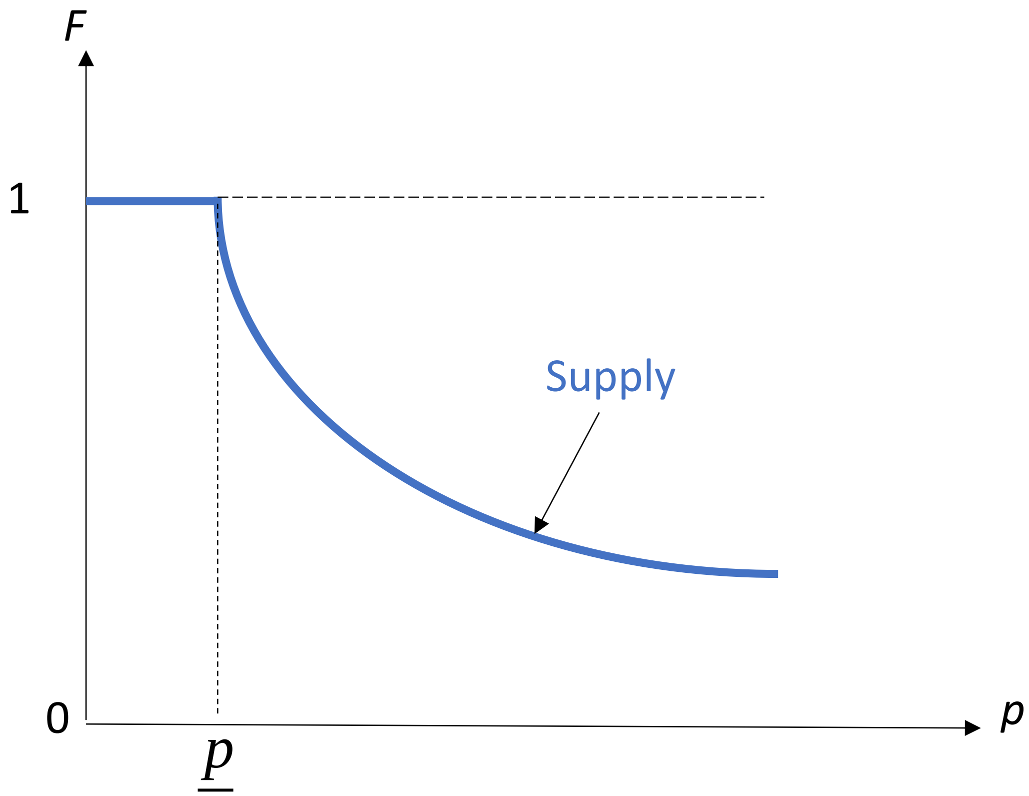

Figure 3: Unique stable equilibrium

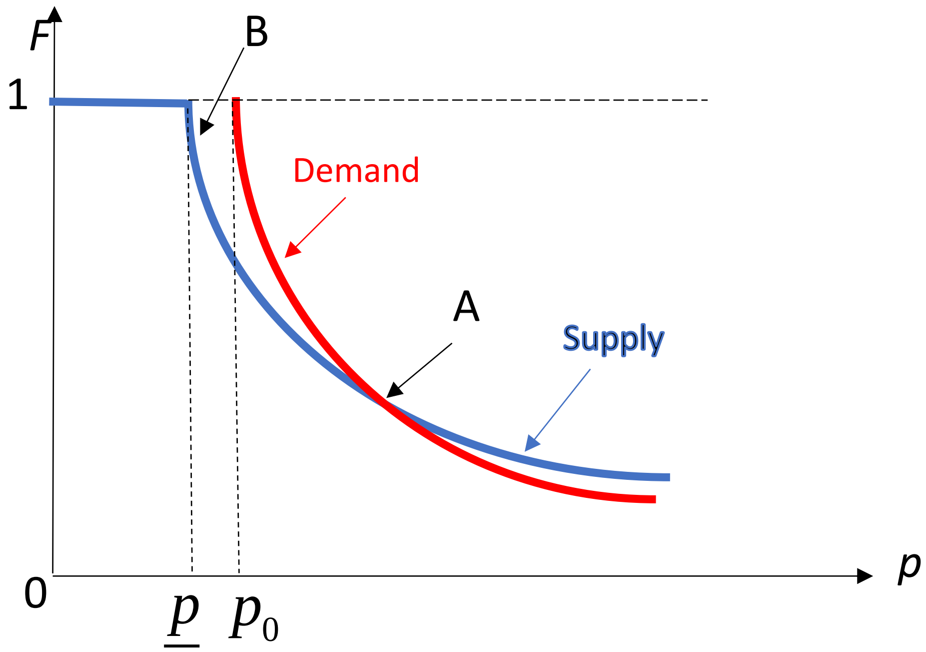

Figure 4: Two equilibria

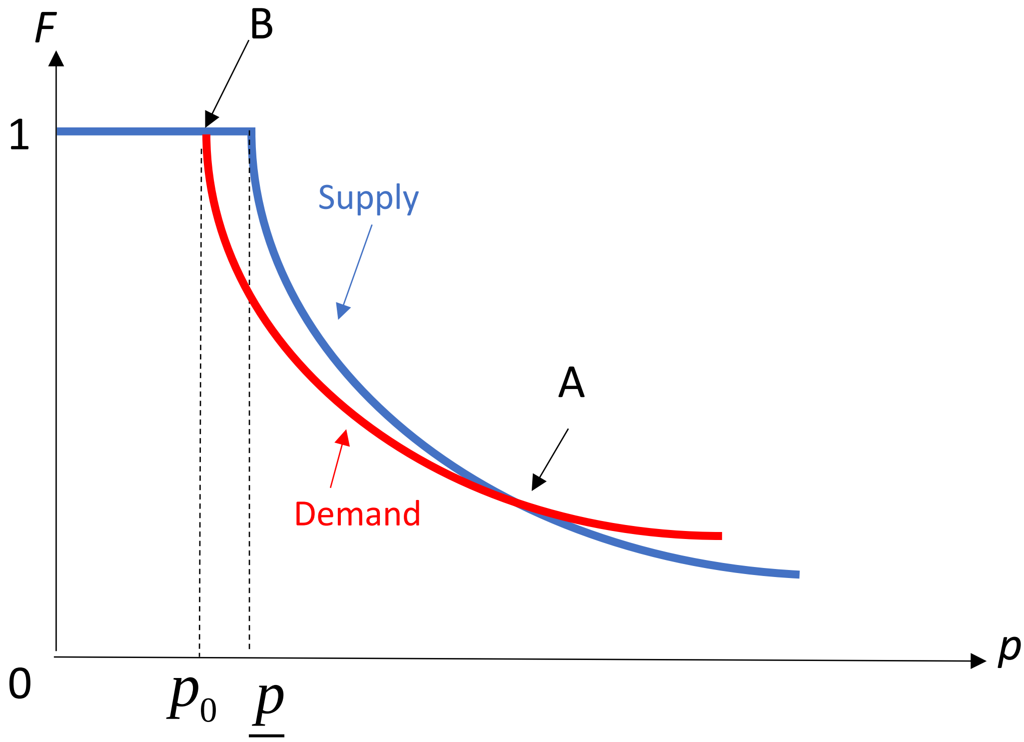

Figure 5 Calibrated UK Equilibrium

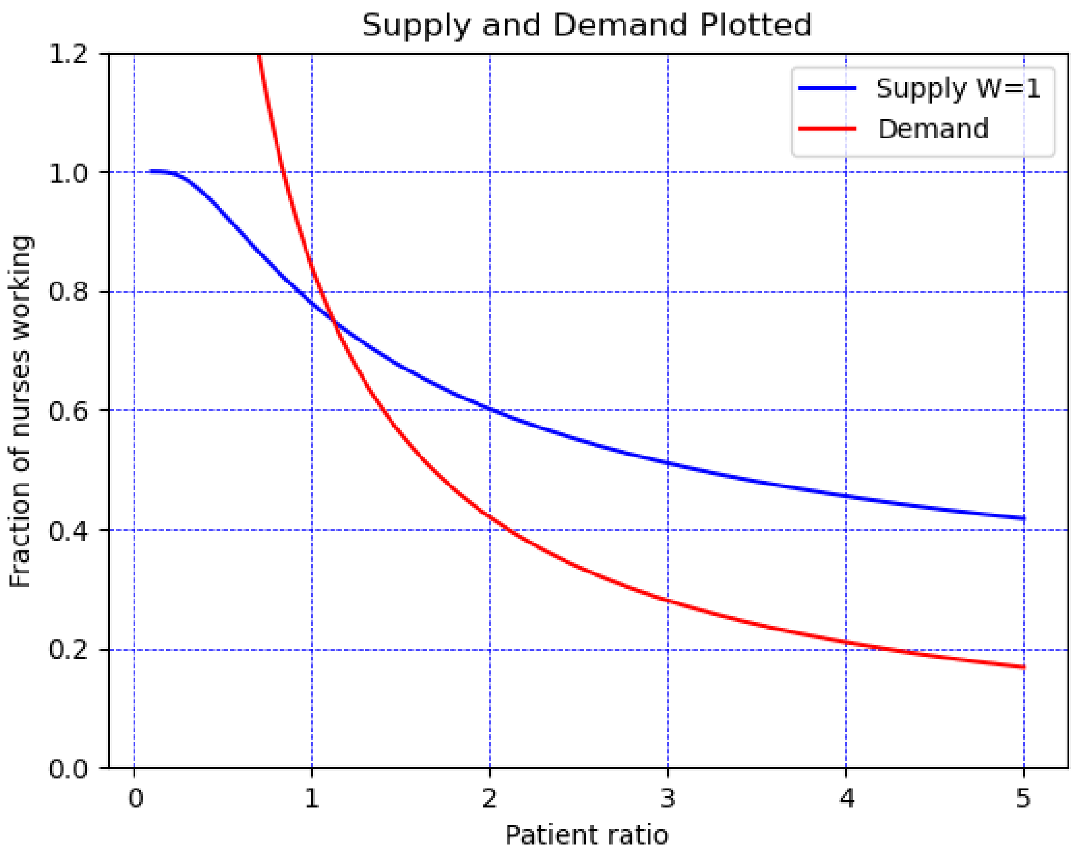

Figure 6. Comparative Static Effect of 5% increase in *W*.